

LIZIANE CRISTINA CAMPOS BRUSAMARELLO

**SEQÜENCIAMENTO E ANÁLISE DE SEQÜÊNCIAS EXPRESSAS (ESTs)  
DE ARROZ (*Oryza sativa*) INOCULADO COM *Herbaspirillum seropedicae***

**Dissertação apresentada como  
requisito à obtenção do grau de  
Mestre em Genética, Programa de  
Pós-Graduação em Genética,  
Departamento de Genética, Setor  
Ciências Biológicas, Universidade  
Federal do Paraná**

**Orientadora: Roseli Wassem**

**Co-Orientador: Emanuel M. Souza**

CURITIBA

2007

Dedico este trabalho

A minha família, alicerce da minha vida, em especial aos meus pais pelo exemplo de amor e dedicação e por sempre valorizarem a educação. E a Alcides J. F. Santos pelo amor, companheirismo, paciência e dedicação.

## **AGRADECIMENTOS**

A Deus pela força espiritual, pelos momentos de reflexão e por guiar sempre meu caminho.

Aos meus orientadores, Roseli Wassem e Emanuel M. Souza, pela oportunidade e pelos ensinamentos científicos que contribuíram muito para o meu crescimento profissional, sem os quais este trabalho não poderia ser realizado.

Ao Centro Nacional de Pesquisa de Arroz e Feijão (EMBRAPA-CNPAP) pelas sementes de arroz, em especial ao pesquisador João Fowler.

A Coordenadoria de Aperfeiçoamento Pessoal de Ensino Superior (CAPES) e ao Conselho Nacional de Pesquisa e Desenvolvimento (CNPq) pelo apoio financeiro.

A Vanessa Kava Cordeiro, Lygia Vitoria Galli Terasawa e Humberto J.O. Ramos pela leitura e correções.

A Rose Adele Monteiro pela amizade, auxílio no laboratório, leitura e correções dessa dissertação.

Ao Leonardo Magalhães Cruz pela amizade, ajuda, auxílio na análise bioinformática e correções deste trabalho.

Aos amigos do Departamento de Genética em especial ao Prof. Elias Karam, Vanessa Sallai Romeiro e Anilda pelo pela amizade e carinho.

A Coordenação da Pós-Graduação em Genética e a Luciana Marques pelo apoio e carinho.

Aos funcionários Valter Baura, Julieta Pie e Roseli Prado pela amizade e apoio técnico. Vocês são exemplos de competência, profissionalismo e grandes amigos.

Aos professores do Grupo de Fixação de Nitrogênio do Departamento de Bioquímica e Biologia Molecular pela confiança e por abrir as portas do laboratório para que este trabalho pudesse ser concretizado.

Aos meus colegas de mestrado em especial aos meus amigos Marcos E. Maciel, Márcio M. Moraes, Kelly Nunes e Jociney de Lima. Obrigada pela amizade, momentos de alegria e pela ajuda nos momentos difíceis.

Aos amigos do Anexo, Rafael Ioris, Helisson Faoro, Anelis Marin, Giovani Pisa, Maria Augusta, Micheli Tadra, Daniela Seixas, Marcelo Müller, Stefania e Eduardo. Obrigada por sempre me ajudarem quando foi preciso e obrigada pelos vários momentos de descontração tornando o trabalho no laboratório ainda mais agradável.

Aos amigos do Grupo de Fixação de Nitrogênio que direta ou indiretamente contribuíram para a realização deste trabalho. Em especial a Fernanda Pacheco pela ajuda no seqüenciamento e a André Fachini pelas correções, discussões filosóficas e científicas. A Juliana Inaba e Giovana Magnani pelo carinho e ajuda nos momentos difíceis. A amizade de todos vocês foi um grande presente durante esses anos.

Ao time de vôlei Genética-Bioquímica pelos momentos de lazer e descontração.

A Família Souza, Leda e Rodrigo, Ligia, Lilian, Ana e Leandra pela amizade, carinho e por estarem comigo nos momentos difíceis e alegres.

A minha amiga Carolina Zabini pelo auxílio fotográfico e por sua amizade.

A Família Santos, Alcides, Eunice e Hallen pela amizade, carinho e apoio. Vocês são para mim um exemplo e motivação para seguir em frente.

A minha família, Sra Dejanira, Sr Benedito, Ronaldo, Rodrigo, Cristiane, Lucas, Lucieli e Alfredo, pela imensurável ajuda e pelo carinho sem os quais esta dissertação não seria possível.

A todos que de alguma forma contribuíram para a realização desta dissertação de mestrado.

## SUMÁRIO

SUMÁRIO.....	v
LISTA DE FIGURAS.....	vi
LISTA DE TABELAS.....	vii
LISTA DE SIGLAS.....	viii
LISTA DE ABREVIATURAS.....	ix
RESUMO.....	x
1 – INTRODUÇÃO.....	1
2 – REVISÃO DE LITERATURA.....	3
2.1 <i>Herbaspirillum seropedicae</i> .....	3
2.2 FIXAÇÃO BIOLÓGICA DE NITROGÊNIO (FBN) EM ARROZ ( <i>Oryza sativa</i> ).....	5
2.3 A UTILIZAÇÃO DE ESTs PARA COMPARAÇÃO DE PERFIL DE EXPRESSÃO DE GENES EM ARROZ.....	7
2.4 O SEQÜENCIAMENTO DO GENOMA DE ARROZ.....	9
3 – OBJETIVOS.....	14
3.1 OBJETIVO GERAL.....	14
3.2. OBJETIVOS ESPECÍFICOS.....	14
4 – MATERIAL E MÉTODOS.....	15
4.1 BACTÉRIAS E VETOR.....	15
4.2 MEIOS DE CULTURA.....	15
4.3 PREPARO DE MATERIAL LIVRE DE RNase.....	18
4.4 MANIPULAÇÕES DE DNA.....	18
4.4.1 Preparo de Células eletrocompetentes.....	18
4.4.2 Transformação bacteriana.....	19
4.4.3 Extração de DNA plasmidial em placas de 96 poços.....	19
4.4.4 Reação de Seqüenciamento.....	20
4.5 OBTENÇÃO DE EST DE <i>Oryza sativa</i> .....	20
4.5.1 Cultivar utilizada.....	22
4.5.2 Desinfecção superficial das sementes.....	22
4.5.3 Condições de cultivo das plantas de arroz.....	22
4.5.4 Controle de contaminantes e eficiência de infecção.....	23
4.5.5 Extração de RNA.....	23
4.5.6 Quantificação de RNA total.....	23
4.5.7 Síntese de cDNA.....	24
4.5.8 Reação de PCR.....	25
4.5.9 Digestão com proteinase K.....	25
4.5.10 Restrição com Sfil.....	25
4.5.11 Fracionamento por tamanho dos fragmentos de cDNA.....	26
4.5.12 Ligação dos fragmentos de cDNA ao vetor.....	26
4.6 ANÁLISE COMPUTACIONAL DE ESTS.....	26
4.6.1 Nomenclatura utilizada para nomear as amostras dos cromatogramas.....	28
4.6.2 <i>Scripts</i> para edição e alinhamento das seqüências.....	29
4.6.3 Banco de dados público utilizado para a comparação.....	30
4.6.4 <i>Scripts</i> para a comparação das seqüências com o banco de dados público.....	31
5 – RESULTADOS E DISCUSSÃO.....	33
5.1 CULTIVO DE <i>Oryza sativa</i> .....	33
5.2 EXTRAÇÃO DE RNA TOTAL E CONSTRUÇÃO DE BIBLIOTECAS DE cDNA.....	35
5.3 ANÁLISE DAS SEQÜÊNCIAS DA BIBLIOTECA RRCH7SN.....	37
6 – CONCLUSÃO.....	48
7 – REFERÊNCIAS BIBLIOGRÁFICAS.....	49
ANEXOS.....	57

## LISTA DE FIGURAS

FIGURA 1 – FLUXOGRAMA DE OBTENÇÃO DE EST DE <i>Oryza sativa</i> .....	21
FIGURA 2 – OBTENÇÃO DE BIBLIOTECAS DE cDNA A PARTIR DE RNA TOTAL .....	24
FIGURA 3 – FLUXOGRAMA GERAL DA ANÁLISE COMPUTACIONAL DE EST .....	27
FIGURA 4 – FLUXOGRAMA DE ANÁLISE DE SEQÜÊNCIAS EST DE ARROZ.....	28
FIGURA 5 – NOMENCLATURA UTILIZADA PARA NOMEAR CROMATOGRAMAS .....	29
FIGURA 6 – SISTEMA DE CULTIVO E PLANTAS DE <i>Oryza sativa</i> COM 14 DIAS DE CRESCIMENTO .....	34
FIGURA 7 – PERFIL ELETROFORÉTICO DE RNA TOTAL, cDNA E RESTRIÇÃO DE CLONES .....	36
FIGURA 8 – DISTRIBUIÇÃO DAS SEQÜÊNCIAS OBTIDAS DE ACORDO COM COMPRIMENTO E QUALIDADE .....	38
FIGURA 9 – DISTRIBUIÇÃO DO <i>E-value</i> DAS SEQÜÊNCIAS ANALISADAS .....	39
FIGURA 10 – ONTOLOGIA USANDO SWISS-PROT .....	42

## LISTA DE TABELAS

TABELA 1 – BACTÉRIAS E VETOR UTILIZADO .....	15
TABELA 2 – LISTA DOS "SCRIPTS" CRIADOS E UTILIZADOS .....	30
TABELA 3 - PROGRAMAS DE BIOINFORMÁTICA UTILIZADOS .....	32
TABELA 4 – DADOS GERAIS DAS BIBLIOTECAS DE cDNA PRODUZIDAS.....	35
TABELA 5 – RESULTADO DA PESQUISA BLAST PARA OS "SINGLETS" .....	43
TABELA 6 – RESULTADO DA PESQUISA BLAST PARA OS "CONTIGS" .....	47

## LISTA DE SIGLAS

IRGSP	–	<i>International Rice Genome Sequence Project</i>
IRRI	–	<i>International Rice Research Institute</i>
SUCEST	–	<i>The sugar Cane EST Project</i>
BGI	–	<i>Beijing Genomics Institute</i>
RAP-DB	–	<i>Rice Annotation Project DataBase</i>
TIGR	–	<i>The Institute for Genomic Research</i>



## LISTA DE ABREVIATURAS

BAC	–	cromossomo artificial de bactéria : do inglês <i>bacterial artificial chromosome</i>
cDNA	–	DNA complementar
DO <sub>600</sub>	–	densidade óptica em $\lambda$ 600 nm
DEPC	–	dietil pirocarbonato
DNA	–	ácido desoxirribonucléico
EST	–	Etiqueta de seqüências expressas, do inglês <i>Expressed sequence tag</i>
FBN	–	Fixação biológica de nitrogênio
GC	–	Guanina, citosina
GUS	–	$\beta$ - glucuronidase
mRNA	–	RNA mensageiro
ORF	–	fase aberta de leitura, do inglês <i>open reading frame</i>
PAC	–	cromossomo artificial baseado no bacteriófago P1, do inglês <i>P1-derived artificial chromosome</i>
pb	–	pares de bases
PCR	–	Reação em cadeia da Polimerase, do inglês <i>Polymerase chain reaction</i>
RNA	–	ácido ribonucléico
RNAse	–	ribonuclease
RYMV	–	virus do arroz amarelo mosqueado, do inglês <i>rice yellow mottle virus</i>
SDS	–	dodecil sulfato de sódio
STR	–	repetições curtas em tandem, do inglês <i>short tandem repeats</i>
TR	–	transcriptase reversa
U	–	unidade de atividade enzimática
YAC	–	cromossomo artificial de levedura do inglês, <i>yeast artificial chromosome</i>

## RESUMO

O arroz (*Oryza sativa*) é o principal alimento para metade da população mundial. Somente na Ásia, mais de dois bilhões de pessoas obtêm aproximadamente dois terços da energia da dieta diária através da ingestão do arroz. Depois da água, o nitrogênio é o fator mais limitante para a planta de arroz e o de maior impacto econômico. Assim, o desenvolvimento de tecnologias para melhorar a fixação biológica de nitrogênio de bactérias associadas ao arroz é essencial. Quando associado à planta de arroz *Herbaspirillum seropedicae* pode fixar 31-54% do nitrogênio total acumulado pelo vegetal em ambiente controlado. O seqüenciamento de ESTs de arroz inoculado com fitopatógenos tem se revelado uma importante ferramenta para entender a interação planta-bactéria. Com esse objetivo foram construídas quatro bibliotecas (8544 clones) de cDNA de raiz de arroz inoculadas ou não com *H. seropedicae*. Os insertos de aproximadamente 900 clones da biblioteca inoculada com *H. seropedicae* foram seqüenciados com o *primer* reverso, que permite determinar a seqüência da região equivalente à extremidade 3' do mRNA. As seqüências foram analisadas e comparadas com as seqüências depositadas no banco RAP-DB do *International Rice Genome Sequencing* (IRGSP) utilizando ferramentas de bioinformática desenvolvidas neste trabalho. Foram obtidos 482 ESTs únicos com leitura média de 333pb (Phred maior ou igual a 15), sendo que destes 446 são *singlets* e 36 são *contigs*. A comparação destas seqüências com o banco de dados revelou que 46% delas não possuem similaridade com seqüências depositadas (*e-value* > que  $10^{-6}$ ) o que pode indicar prováveis genes desconhecidos. Neste grupo de genes novos podem estar incluídos genes específicos da interação *Oryza-H. seropedicae*. A biblioteca obtida deve conter genes importantes para o processo de infecção planta-bactéria e deverá ser utilizada para rastrear genes envolvidos neste processo.

Palavras-chave: *Oryza sativa*, EST, *Herbaspirillum seropedicae*, cDNA.

## 1. INTRODUÇÃO

A Organização das Nações Unidas declarou 2004 o Ano Internacional do Arroz para ressaltar sua importância como alimento mundial e no combate a pobreza. O arroz é o principal alimento de aproximadamente metade da população mundial e constitui em média um quinto da energia alimentar total de cada indivíduo. Isso é ainda mais marcante na Ásia, onde mais de 2 bilhões de pessoas obtêm aproximadamente dois terços da energia da dieta diária através da ingestão do arroz (*International Rice Research Institute, 2006.*)

A produção global de arroz bateu recorde de 628 milhões de toneladas em 2005, sendo que o Brasil está entre os 10 maiores produtores do mundo. Esta grande produção pode, entretanto, ser dramaticamente afetada por mudanças climáticas globais. No *Climate Change and Rice Planning Workshop* promovido pelo *International Rice Research Institute* (IRRI) (Filipinas, março de 2006) afirmou-se que a mudança climática já está afetando a capacidade da Ásia para produzir arroz. Conseqüentemente, isto poderia comprometer os esforços que tem sido feitos para reduzir a pobreza na região habitada por uma grande parcela da população mais pobre do mundo. Estimou-se que para superar essas mudanças climáticas a produtividade de arroz terá que dobrar nos próximos 50 anos. O IRRI investirá US\$ 2 milhões em pesquisas para buscar novas variedades de arroz e novas tecnologias para minimizar o efeito das mudanças climáticas na produção de arroz (IRRI, 2006).

O aumento previsto na produção de arroz implicará em um aumento proporcional no consumo de fertilizantes nitrogenados uma vez que, depois da água, o nitrogênio é o nutriente que mais limita a produtividade de arroz (LADHA; REDDY, 2003). Na agricultura, a produtividade de cereais é suportada principalmente pelo uso de fertilizantes nitrogenados inorgânicos, os quais tem alto custo de produção e causa danos ao meio ambiente, principalmente poluição da água (FISCHER, 1999). Uma alternativa ao uso de fertilizantes químicos é explorar e melhorar a capacidade do arroz para obter nitrogênio através da Fixação Biológica de N<sub>2</sub> (FBN) (LADHA; REDDY, 1995). Sendo assim, a FBN é mais vantajosa em relação a aspectos ecológicos e econômicos (MUTHUKUMARASAMY et al., 2005).

*Herbaspirillum seropedicae* é uma bactéria fixadora de nitrogênio, capaz de colonizar arroz e cuja eficiência para transferir nitrogênio fixado para as plantas tem sido descrita em vários trabalhos (BALDANI V.L.D; BALDANI J.I.; DÖBEREINER, 2000; CANUTO et al., 2003). Neste contexto, o presente trabalho tem como principal objetivo contribuir para a melhor compreensão dos mecanismos de colonização por *H. seropedicae* em *Oryza sativa* L. (*japonica*), através da identificação de genes envolvidos nessa interação. Os resultados poderão ser importantes para trabalhos que visam melhorar a eficiência na interação de *H. seropedicae* com arroz e outras gramíneas.

A construção de bibliotecas de cDNA obtidos a partir de raízes inoculadas com *H. seropedicae* poderá contribuir para a compreensão da interação planta - bactéria através da identificação de genes que são expressos somente nesta condição. Esta estratégia (construção de bibliotecas e seqüenciamento em grande escala) tem sido utilizada com grande sucesso para investigar genes expressos sob condições específicas (NOGUEIRA et al. 2001; LAMBAIS, 2001; JANTASURIYARAT et al. 2005, VENDELON-DEBOUT et al. 2003; REDDY et al. 2002; RAUYAREE et al. 2001).

## 2 - REVISÃO DE LITERATURA

### 2.1 *Herbaspirillum seropedicae*

*Herbaspirillum seropedicae* é uma bactéria diazotrófica endofítica, Gram-negativa, vibrióide, membro da subdivisão  $\beta$  das Proteobactérias e foi primeiramente isolada de milho, arroz e sorgo (BALDANI et al., 1986). Posteriormente, *H. seropedicae* foi também isolado de outras gramíneas como cana de açúcar, trigo <sup>1</sup> (KENNEDY; CHOUDHURY; KECSKE'S, 2004) e também de plantas não pertencentes ao grupo das gramíneas (CRUZ et al., 2001). A colonização das plantas hospedeiras por *H. seropedicae* ocorre através de rupturas no tecido radicular ou ainda através dos pontos de emergência das raízes laterais. Após a entrada, *H. seropedicae* rapidamente coloniza espaços intercelulares, aerênquima e xilema das raízes e continua a colonização até alcançar tecidos aéreos da planta, sendo freqüentemente encontrado em caules e folhas (OLIVARES et al, 1997; OLIVARES, JAMES, 2000; JAMES et al., 2002, RONCATO-MACCARI et al., 2003). Por outro lado, a estirpe B501 de *Herbaspirillum* spp. parece não ser capaz de iniciar colonização de arroz selvagem (*Oryza officinalis*) através das raízes, sendo observada apenas na superfície radicular. Esta estirpe aparentemente inicia a colonização através do caule e passa para as partes aéreas via espaços apoplásticos, uma vez que não foi observada em tecidos vasculares (ELBELTAGY et al., 2001).

JAMES e colaboradores (2002) utilizaram a estirpe Z67 de *H. seropedicae* marcada com  $\beta$ -glucuronidase (GUS) para inocular plantas de arroz e mostraram que esta coloniza preferencialmente o aerênquima e espaços intercelulares das raízes. Estes autores observaram também colonização de vasos xilemáticos de partes aéreas, e em menor frequência vasos xilemáticos de raízes, especialmente na cultivar IR72 de *Oryza sativa*. RONCATO-MACCARI e colaboradores (2003) também observaram *H. seropedicae* colonizando espaços intercelulares e

---

<sup>1</sup>James, E.K.; Gyaneshwar, P.; Barraquio, W.L.; et al.. Endophytic diazotrophs associated with rice. In: Ladha, J.K.; Reddy, P.M. (Eds.), The Quest for Nitrogen Fixation in Rice. **International Rice Research Institute**, Los Baños, p. 119–140. 2000.

confirmaram sua capacidade de fixar nitrogênio *in planta*, através da detecção de  $\beta$ -glucuronidase expressa a partir do promotor *nifH*.

Em condições de campo, a inoculação com *H. seropedicae* pode aumentar o crescimento de raízes, brotos (ARANGARASAN; PALANIAPPAN; CHELLIAH, 1998) e a taxa de germinação de sementes de arroz (PEREIRA et al., 1988). Todavia, a inoculação simultânea de vários diazotrofos, como mostra OLIVEIRA et al. (2002), pode contribuir ainda mais para o aumento do conteúdo de nitrogênio da planta. Estes autores observaram que o melhor tratamento foi à mistura de cinco linhagens (*Gluconacetobacter diazotrophicus*, *Herbaspirillum seropedicae*, *Herbaspirillum rubrisubalbicans*, *Azospirillum amazonense* e *Burkholderia* sp. (originalmente isolada de cana de açúcar)), seguida pelo tratamento com *Herbaspirillum* ssp. Sendo assim, o autor sugere que a combinação de espécies é a melhor estratégia para o melhoramento de cultivo de cana de açúcar. CANUTO et al, 2003, ao comparar a contribuição da fixação biológica de nitrogênio por *H. seropedicae* e *G. diazotrophicus* em cana-de-açúcar, cultivada em casa de vegetação, mostrou que os melhores resultados de fixação biológica de N<sub>2</sub>, foram de *H. seropedicae*, sendo as raízes o local preferencial para colonização e estabelecimento destas duas espécies.

Em cana de açúcar, além de *H. seropedicae*, *H. rubrisubalbicans* também é encontrado colonizando raízes, caules e folhas (REIS et al., 2000). Estes diazotrofos podem aumentar o conteúdo de nitrogênio das folhas e a produção de cana de açúcar significativamente, mas não podem substituir completamente o uso de fertilizantes químicos (MUTHUKUMARASAMY; REVATHI; LAKSHMINARASIMHAN, 1999).

Alguns cultivares de cana de açúcar, como a B-4362, desenvolvem a patologia estria mosqueada ao serem inoculadas com *H. rubrisubalbicans* (OLIVARES et al., 1997). A inoculação de *H. seropedicae* em tal variedade, bem como da variedade SP 70-1143, resistente à doença estria mosqueada, induziu uma resposta hipersensitiva em ambas. Entretanto, os sintomas não evoluíram e as plantas não desenvolveram a estria mosqueada (OLIVARES et al., 1997). Estes resultados indicam que *H. seropedicae* não atua como um fitopatógeno, e sim estabelece uma interação planta-bactéria com vantagens para ambas às espécies. O acúmulo de polissacarídeos e taninos no parênquima celular ao redor do

metaxilema de cana de açúcar Ja60-5 inoculada com *G. diazotrophicus* foi observado, sugerindo que o sistema de defesa da planta é ativado durante a interação com a bactéria (DONG; MCCULLY; CANNY, 1997). Estudos de interação planta-patógeno mostraram que a ativação do sistema de defesa da planta é dependente do reconhecimento de sinais moleculares por receptores da planta. Estes sinais moleculares gerados no sítio de infecção resultam na ativação/repressão de fatores de transcrição que reconhecem seqüências específicas da região promotora do DNA de genes relacionados à defesa (MACDOWELL; DANGL, 2000).

## 2.2 FIXAÇÃO BIOLÓGICA DE NITROGÊNIO (FBN) EM ARROZ (*Oryza sativa*)

Uma maior diversidade de bactérias fixadoras de nitrogênio é encontrada na superfície e no interior das raízes de arroz quando comparada com o solo ao redor destas. As raízes de arroz cultivado em solo irrigado podem oferecer condições mais favoráveis ao crescimento de bactérias heterotróficas fixadoras de N<sub>2</sub>, uma vez que estas dependem de substâncias orgânicas excretadas pelas raízes. Geralmente observa-se um número consideravelmente maior de bactérias diazotróficas em raízes de arroz cultivado em solo irrigado quando comparado com raízes de solo seco, provavelmente decorrente da prolongada estabilidade do ambiente irrigado (UEDA et al. 1995a).

BALDANI (1996) mostrou que a inoculação de plântulas de *Oryza sativa* com *H. seropedicae* leva a um aumento de seu peso seco em até 50%. Além disso, é observado um aumento do conteúdo de nitrogênio da planta, induzido pela fixação de nitrogênio realizada por *H. seropedicae*, mas este varia de 31 a 54%, dependendo da estirpe utilizada (BALDANI, V.L.D.; BALDANI, J.I.; DOBEREINER, 2000; <sup>2</sup>MIRZA et al 2000). Os benefícios observados nas plantas são também altamente variáveis (0 a 36% de nitrogênio total presente na planta, obtido do ar através de fixação biológica) em decorrência da variedade de arroz utilizada

---

<sup>2</sup> MIRZA, M.S.; RASUL, G.; MEHNAZ, S.; et al. Beneficial effects of inoculated nitrogen-fixing bacteria on rice. In: Ladha, J.K., Reddy, P.M. (Eds.), The Quest for Nitrogen Fixation in Rice. **International Rice Research Institute**, Los Baños, p. 191–204. 2000.

(SHRESTHA; LADHA, 1996; MALARVIZHI; LADHA, 1999), sendo estes afetados pela interação entre o genótipo da planta e o meio ambiente (ROGER; LADHA, 1992; MALARVIZHI; LADHA, 1999). De forma geral, quando a associação entre diazotrofos e planta é pouco eficiente, a taxa da fixação biológica é mais suscetível a fatores ambientais, produzindo resultados variáveis. Outro fator limitante é a disponibilidade de carbono na rizosfera e/ou no interior da planta. GYANESHWAR e colaboradores (2002) mostraram que variedades de arroz tolerantes a presença de alumínio no solo, quando inoculadas com *Herbaspirillum* spp, possuem um maior nível de fixação de nitrogênio que variedades sensíveis. Isso se deve provavelmente à habilidade do cultivar tolerante de liberar ácidos orgânicos que são utilizados como fonte de carbono para o crescimento e convertido em energia para a fixação de nitrogênio pelo *Herbaspirillum* spp (GYANESHWAR et al, 2002). JAMES e colaboradores (2002) também obtiveram resultados semelhantes. Ao avaliarem a fixação de nitrogênio em plantas de arroz inoculadas com *H. seropedicae*, observaram atividade de nitrogenase somente quando malato foi adicionado como fonte de carbono ao meio de cultivo das plantas.

Baseado em análises de seqüências parciais dos genes *nifH* ou *nifD* obtidas de DNA amplificado da superfície de raízes, uma grande diversidade molecular de fixadores de nitrogênio foi encontrada (UEDA et al., 1995b). No entanto, essas bactérias fixadoras de nitrogênio, associadas externamente ao arroz, são ineficientes em sustentar o desenvolvimento da planta e alta produtividade (LADHA et al., 1998), enquanto que, aquelas que colonizam o interior da planta, com menor competição por fontes de carbono e com um ambiente mais adequado para a fixação de nitrogênio, poderiam estabelecer associações mais eficientes (REINHOLD-HUREK; HUREK, 1998a, 1998b).

Com o objetivo de identificar novas espécies de microrganismos do interior da raiz, muitas bactérias endofíticas fixadoras de nitrogênio têm sido isoladas de tecido de arroz após esterilização superficial (BARRAQUIO; REVILLA; LADHA, 1997; ENGELHARD; HUREK; REINHOLD-HUREK, 2000; STOLTZFUS et al., 1997; GYANESHWAR et al., 2001). Possivelmente todas ou algumas dessas bactérias poderiam estar contribuindo para o balanço de nitrogênio nas culturas de arroz, o que tem sido confirmado por estudos de FBN, comparando plantas não inoculadas (BODDEY et al., 1995; WU et al., 1995; SHRESTHA; LADHA, 1996; MALARVIZHI;



LADHA, 1999) e inoculadas (BALDANI, V.L.D.; BALDANI, J.I.; DÖBEREINER, J., 2000; MEHNAZ et al., 2001; NIEUWENHOVE et al., 2001).

Acredita-se que diazotrofos endofíticos, principalmente *G. diazotrophicus* e *Herbaspirillum* spp, possam ser responsáveis por disponibilizar uma quantidade significativa de nitrogênio fixado em algumas variedades de cana de açúcar cultivadas no Brasil (*Saccharum* spp.) (OLIVARES et al, 1997). Entretanto, é possível que variedades de arroz com maior potencial para aproveitar a fixação de nitrogênio possam se beneficiar da associação com endófitos (SHRESTHA; LADHA, 1996).

### 2.3 A UTILIZAÇÃO DE ESTs PARA COMPARAÇÃO DE PERFIL DE EXPRESSÃO DE GENES EM ARROZ

Os processos que envolvem a associação entre gramíneas e bactérias diazotróficas são muito complexos, principalmente devido à ausência de uma estrutura visível, como os nódulos formados na interação de rizóbio com as leguminosas (BALDANI, 1996). Em função disso, diferentes estratégias são utilizadas para estudar a interação entre bactérias associativas e plantas. Uma das estratégias utilizadas é a análise de expressão diferencial para identificar genes envolvidos em tais associações.

A disponibilidade de um grande número de ESTs fornece uma oportunidade de identificar genes que são expressos em vários tecidos, tipos celulares ou estágios do desenvolvimento (LAMBAIS, 2001; MICHALEK et al., 2002; JANTASURIYARAT et al., 2005; OGIHARA et al., 2003; RONNING et al., 2003). Análises dessa magnitude podem ser úteis na comparação dos níveis de expressão gênica em diferentes tecidos submetidos a condições diversas através da utilização de análise computacional em um banco de dados de ESTs. Isso porque, geralmente, a frequência de um determinado EST está correlacionada ao acúmulo de transcritos nos tecidos dos quais as bibliotecas de cDNA foram construídas. Essa abordagem tem sido muito utilizada em estudos de expressão gênica em vários organismos (NOGUEIRA et al. 2001; LAMBAIS, 2001).

NOGUEIRA e colaboradores (2001) identificaram diversos genes de cana de açúcar expressos somente na presença de bactérias endofíticas diazotróficas. Os

genes diferencialmente expressos foram agrupados em três processos principais: metabolismo de nitrogênio, crescimento da planta e interação planta microrganismo. Esta análise foi feita através de ESTs de tecidos distintos de cana de açúcar inoculadas com *G. diazotrophicus* e *H. rubrisubalbicans*, utilizando banco de dados de EST do SUCEST (<http://sucest.lad.ic.unicamp.br/en/>).

JANTASURIYARAT et al. (2005), utilizaram EST para estudar a interação entre o fungo fitopatogênico *Magnaporthe grisea* e arroz, obtendo uma grande coleção de ESTs de arroz a partir de tecido inoculado e não inoculado. Estes dados produziram uma base sólida para a caracterização do sistema de defesa de arroz, além de produzir dados genômicos públicos para outros estudos de genoma funcional.

VANTELON-DEBOUT e colaboradores (2003) também utilizaram a construção de bibliotecas de cDNA de tecido de arroz em diferentes condições para observar o padrão de expressão dos genes. O cDNA foi obtido a partir de folhas de plantas de arroz não estressadas, de folhas feridas e infectadas com RYMV (*rice yellow mottle virus*) de duas variedades: IR64 (*Oryza sativa indica*) e Azucena (*Oryza sativa japonica*). A análise de EST mostrou que para cada subespécie há um padrão diferente de expressão gênica. DEBOUT e colaboradores detectaram 2.342 ESTs sem similaridade significativa com qualquer seqüência conhecida, o que os caracteriza como prováveis candidatos a genes ainda desconhecidos.

Diferentes padrões de expressão de EST são observados em cana de açúcar, quando infectada com *G. diazotrophicus* e *H. rubrisubalbicans*. Esse dado sugere que o controle de infecção e/ou processo de colonização é dependente do endófito em questão (LAMBAIS, 2001).

WU et al. (2002) realizaram mapeamento de ESTs para determinar a posição cromossomal de genes expressos em arroz. A utilização de *primers* específicos obtidos de bibliotecas de cDNA em combinação com análise genética permitiu a construção de um mapa genômico. Nesse estudo, os autores descreveram um mapa transcricional contendo 6.591 sítios EST que cobrem 80,8% do genoma de arroz. Os cromossomos 1, 2 e 3 possuem densidades relativamente altas de EST, aproximadamente duas vezes maior que a dos cromossomos 11 e 12, e contém 41% dos sítios de EST analisados. As regiões de EST, em sua maioria, estão distribuídas na região distal de cada braço dos cromossomos e em menor densidade

nos centrômeros. A maioria (40%) dos ESTs mapeados ocupa somente 21% de todo o genoma.

A caracterização de ESTs contribui não somente para a construção de um mapa físico dos cromossomos como também permite a investigação dos mecanismos de expressão de várias famílias gênicas (NOGUEIRA et al., 2001). O Programa de Pesquisa do Genoma de Arroz (RPG) isolou e seqüenciou parcialmente mais de 29.000 clones de cDNA de vários tecidos de arroz (cv. Nipponbare, var. *japonica*). Ao traduzir esses dados de seqüência em aminoácidos, foi observado que aproximadamente 25% dos clones tinham similaridade significativa com proteínas conhecidas. Alguns desses clones mostraram distribuições específicas nas bibliotecas, indicando que a composição dos clones em cada biblioteca reflete, até certo ponto, a regulação de genes específicos de expressão para diferenciação, condições de crescimento ou estresse ambiental. A montagem foi realizada com 24.728 clones diferentes e estes se agruparam em 10.000 grupos independentes (*contigs*, *singlets* e *singletons*), sugerindo que metade dos genes expressos em arroz foram obtidos. Esses resultados fornecem informações importantes sobre a expressão e regulação gênica em arroz (YAMAMOTO; SASAKI, 1997).

Atualmente, as espécies com maior número de seqüências de ESTs depositadas nos bancos de dados públicos são *Homo sapiens*, *Mus musculus*, *Danio rerio*, *Bos taurus* e *Oryza sativa*, que aparece em quinto lugar, com mais de 1.211.078 ESTs disponíveis, sendo a planta com maior número de ESTs ([http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html); fev/07). A maioria das ESTs de *Oryza sativa* é derivada da subespécie *japonica*.

## 2.4 O SEQÜENCIAMENTO DO GENOMA DE ARROZ

O genoma de arroz foi estimado em 400-430 milhões de pares de bases (Mb), o menor genoma entre os cereais (SASAKI; BURR, 2000). O seu seqüenciamento é importante para explorar todo o potencial do arroz, visando principalmente seu melhoramento. O uso de marcadores moleculares durante o melhoramento tem permitido que características de interesse fossem rastreadas em uma ampla população, necessitando cada vez menos de testes de progênie. O conhecimento da

localização de todos os genes do genoma aumenta a utilidade de marcadores moleculares porque permitem a identificação de genes candidatos envolvidos no controle de uma característica específica de interesse (SASAKI; BURR, 2000). Em função de sua importância, quatro grupos independentes realizaram grandes projetos de seqüenciamento do genoma do arroz. Estes foram: *International Rice Genome Sequencing Project* (IRGSP), que consiste de um grande consórcio internacional de instituições de pesquisa, a empresa de biotecnologia Monsanto; Syngenta, outra empresa privada; e o *Beijing Genomics Institute* (BGI). O IRGSP, Syngenta e Monsanto utilizaram a sub-espécie *japonica*, enquanto o BGI utilizou *indica* para seus projetos. Outra diferença observada entre os grupos foi na estratégia utilizada; enquanto a Monsanto e o IRGSP realizaram o seqüenciamento utilizando uma estratégia conhecida como “*clone-by-clone*”, onde sub-bibliotecas de pequenas regiões mapeadas (clonadas em BACs, PACs e YACs) são seqüenciadas, a Syngenta e BGI fizeram seqüenciamento utilizando genoma total (conhecido como seqüenciamento *shot-gun*). Antes de iniciar o seqüenciamento propriamente dito, o IRGSP e a Monsanto construíram bibliotecas, mapas genômicos físicos com marcadores moleculares definidos, que são utilizados tanto na montagem do genoma quanto para estratégias de melhoramento genético.

A Monsanto realizou uma análise no banco de dados de seqüência do genoma de arroz e produziu um banco de dados contendo todos os STR (*short tandem repeats*) de no mínimo 24 pb na seqüência preliminar obtida. Estas seqüências STR são a parte do genoma prevista para ser a mais útil como marcadores em melhoramento de arroz e estão disponíveis no site <http://www.riceresearch.org> e no GenBank (BARRY, 2001).

Em 2000 a Monsanto anunciou a obtenção de uma versão preliminar não finalizada do genoma de arroz e posteriormente compartilhou seus dados de mapeamento com o IRGSP para auxiliar na obtenção da seqüência completa do genoma de arroz cultivar Nipponbare. O projeto de Seqüenciamento de arroz da Monsanto produziu aproximadamente 399 Mb de seqüências de 3.391 BACs. As seqüências contaminantes foram removidas do banco de dados e as seqüências redundantes foram reunidas em grandes *contigs*. Este processo produziu uma versão preliminar do genoma, totalizando 259 Mb (BARRY, 2001).

No fim de 2002 o IRGSP completou o seqüenciamento do genoma da cultivar Nipponbare após a etapa de finalização. Mais de 500 Mb dessas seqüências estão disponíveis em um banco de dados público no portal da IRGSP (<http://rgp.dna.affrc.go.jp/cgi-bin/statusdb/seqcollab.pl>)(<http://www.rice-research.org>). As seqüências disponíveis da subespécie *japonica* cobriam 93% das 420 Mb estimadas do genoma desta variedade, que segundo previsões continha 32.000 a 50.000 genes (GOFF et al, 2002). Nesse mesmo ano YU et al. (2002), também mostraram versões preliminares do genoma de arroz, entretanto, do cultivar *indica*.

YU et al. (2005), realizaram uma otimização através de análise computacional dos genomas seqüenciados de *indica* e *japonica*, e conseguiram agrupar as seqüências obtendo *contigs* maiores. Utilizando dados de ESTs para ajustar erros de previsão, foram estimados pelo menos 38.000 – 40.000 genes. Comparações do conteúdo de genes entre *indica* e *japonica*, revelaram poucas diferenças entre eles, sendo que somente 2 a 3 % dos genes são únicos para uma das duas subespécies. Entretanto, no que diz respeito a regiões intergênicas observou-se uma enorme variação entre as subespécies.

Duplicações do genoma de arroz, propostas desde 1930 por Bárbara McClintock têm sido confirmadas (YU et al., 2005). Os mesmos autores relataram 18 distintos pares de segmentos duplicados que cobrem 65,7% do genoma, sendo que 17 dessas duplicações ocorreram antes da divergência das gramíneas.

MATSUMOTO et al. (2005), finalizaram o seqüenciamento do genoma de arroz, cobrindo 95% das 389 Mb estimadas, o que inclui todas as regiões de eucromatina e dois centrômeros. Esta versão do genoma de arroz possui apenas pequenas porções ainda não seqüenciadas, sendo que estas contêm regiões centroméricas e outras regiões repetitivas. A análise das seqüências permitiu a identificação de 37.544 ORFs (*open reading frames*) com similaridade a seqüências codificadoras de proteínas. O genoma de arroz apresentou também uma média de um gene a cada 9,9 kb e uma média de comprimento de gene de 2.699 pb. Os cromossomos 1 e 3 possuem a maior densidade de genes com 1 gene por 8,9 kb e 1 gene por 8,7 kb, respectivamente. A menor densidade de genes está nos cromossomos 11 e 12, que possuem 1 gene por 10,7 kb e 1 gene por 11,6 kb, respectivamente (MATSUMOTO et al, 2005).

*Oryza* e *Arabidopsis* são representantes das duas maiores classes de plantas, monocotiledôneas e dicotiledôneas. Sendo assim uma análise comparativa das seqüências expressas entre essas espécies pode colaborar para o entendimento da evolução dessas espécies que divergiram há aproximadamente 200 milhões de anos.

O genoma de arroz é aproximadamente três vezes maior e possui quase 50% a mais de genes que o de *Arabidopsis*. Aproximadamente 90% dos genes de *Arabidopsis* possuem similaridade com os de *Oryza*, enquanto 71% dos genes de *Oryza* possuem similaridade com os de *Arabidopsis* (MATSUMOTO, 2005). Entretanto, apenas uma pequena região sintênica, contendo 119 genes, foi observada entre os cromossomos 5 de *Arabidopsis* e 4 de arroz. Os genes de arroz que não possuem equivalentes em *Arabidopsis* exibem características gerais (% de GC, tamanho médio, número de íntrons) diferentes dos outros genes de arroz e podem conter ainda erros de anotação e mesmo não serem genes verdadeiros. Boa parte destes genes não possui ESTs identificados em arroz e nem em milho, mas poderão vir a ser confirmado por novos estudos de genômica funcional (MATSUMOTO, 2005).

SALSE et al. (2004) ao compararem ESTs de milho com o genoma de arroz através de alinhamento, observaram que 74% dos *contigs* de milho possuem uma região homóloga no genoma de arroz e 47% mostraram relação ortóloga. Os autores realizaram também, análises de colinearidade entre alguns dos cromossomos de milho e arroz e mostraram rearranjos internos e também regiões colineares. Análises detalhadas do cromossomo 1 e 5 de arroz mostraram que 11% dos *loci* anotados no cromossomo 1 estão duplicados no cromossomo 5, sugerindo um alto grau de reorganização interna dessas seqüências.

O conteúdo de GC em genomas de gramíneas é maior quando comparado a dicotiledôneas (CAREL; BERNARDI, 2000). A quantidade de GC no genoma de *Arabidopsis* corresponde a 34,7% com os exons tendo 44,1% e os íntrons 32,7% (*The Arabidopsis Genome Initiative*, 2000). O genoma de arroz apresenta uma média de 43,6% de GC, sendo que os exons possuem 54,2% e os íntrons 38,3% (Matsumoto et al, 2005). Outra diferença é a presença de um gradiente do conteúdo de GC nos genes de arroz, com uma quantidade 25% maior de GC na ponta 5' do que a 3'. Essa característica não é observada em *Arabidopsis* (Yu et al, 2002).

O conhecimento do genoma de arroz tem crescido exponencialmente nos últimos anos principalmente devido a grandes investimentos em pesquisa neste tema e a formação do consórcio internacional de grupos de pesquisa. Essas informações permitem a realização de estudos de genômica funcional com cada vez maior precisão e maior potencial de aplicação no melhoramento. Todas essas informações obtidas sobre o genoma de arroz, são de fundamental importância também para o desenvolvimento do presente trabalho, uma vez que esses dados serviram como base para comparação dos resultados obtidos.

### 3 – OBJETIVOS

#### 3.1 OBJETIVO GERAL

O objetivo deste trabalho é a obtenção de seqüências EST de arroz inoculado com *H. seropedicae* para rastrear genes expressos na presença deste diazotrofo e analisar o perfil de expressão gênica nessa condição.

#### 3.2. OBJETIVOS ESPECÍFICOS

- Obter raízes de arroz (*Oryza sativa* L. ssp *japonica*) inoculadas com *Herbaspirillum seropedicae*, estirpe SMR1, após 5 e 7 dias de incubação, raízes controle sem microrganismos e raízes crescidas na presença de fonte de nitrogênio (NH<sub>4</sub>Cl);
- Sintetizar cDNA a partir do mRNA total das raízes obtidas;
- Construir bibliotecas com o cDNA obtido das raízes inoculadas e não inoculadas com *Herbaspirillum seropedicae*;
- Seqüenciar os insertos das bibliotecas de cDNA obtidas;
- Analisar as seqüências obtidas e comparar com as seqüências do genoma de arroz depositadas por outros pesquisadores em um banco de seqüência público, utilizando ferramentas de bioinformática.



## 4 – MATERIAL E MÉTODOS

### 4.1 BACTÉRIAS E VETOR

As bactérias e vetor utilizados neste trabalho estão descritos na Tabela 1.

TABELA 1 – BACTÉRIAS E VETOR UTILIZADO

Bactéria	Estirpe	Fenótipo/Genótipo	Referência
<i>Herbaspirillum seropedicae</i>	SMR1	selvagem, Nif <sup>+</sup> , Sm <sup>R</sup>	SOUZA et al, 1995
<i>Escherichia coli</i>	TOP10	F- <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\phi$ 80/ <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74 recA1 araD139 <math>\Delta</math>(<i>araleu</i>) 7697 <i>galJ galK rpsL</i> (Str<sup>R</sup>) <i>endA1 nupG</i></i>	Invitrogen <sup>TM</sup>
Vetor	Propriedades		
pDNR-LIB	Cm <sup>R</sup> Permite clonagem direcionada, mediada pela recombinase Cre e susceptibilidade a sacarose (SacB)		Clontech

### 4.2 MEIOS DE CULTURA

Para o crescimento de *Escherichia coli* os meios de cultivo utilizados estão descritos abaixo.

#### Meio SOB

Bacto-triptona	20 g.L <sup>-1</sup>
Extrato de levedura	5 g.L <sup>-1</sup>
NaCl	0,584 g.L <sup>-1</sup>
KCl	0,186 g.L <sup>-1</sup>

#### Meio SOC

Bacto-triptona	20 g.L <sup>-1</sup>
Extrato de levedura	5 g.L <sup>-1</sup>
NaCl	0,6 g.L <sup>-1</sup>
KCl	0,19 g.L <sup>-1</sup>

MgCl <sub>2</sub>	0,94 g.L <sup>-1</sup>
MgSO <sub>4</sub>	1,2 g.L <sup>-1</sup>
Glicose	3,6 g.L <sup>-1</sup>

#### Meio Terrific Broth (TB)

Bacto-triptona	12 g.L <sup>-1</sup>
Extrato de levedura	24 g.L <sup>-1</sup>
Glicerol	4 mL.L <sup>-1</sup>

#### Meio Luria-Bertani Agar (LA)

Triptona	10 g.L <sup>-1</sup>
Extrato de levedura	5 g.L <sup>-1</sup>
NaCl	10 g.L <sup>-1</sup>
Ágar	15 g.L <sup>-1</sup>

O pH foi ajustado para 7,0 com NaOH 5 mol.L<sup>-1</sup>

Para o crescimento de *Herbaspirillum seropedicae* utilizou-se o meio de cultura NFbHP (Machado *et al.* 1991) modificado, com malato substituindo lactato como fonte de carbono. O meio NFbHP-malato possui a composição a seguir.

	g. L <sup>-1</sup>
KH <sub>2</sub> PO <sub>4</sub>	7,97
K <sub>2</sub> HPO <sub>4</sub>	8,9.10 <sup>-1</sup>
MgSO <sub>4</sub> .7H <sub>2</sub> O	2,0.10 <sup>-1</sup>
NaCl	1,0.10 <sup>-1</sup>
CaCl <sub>2</sub>	2,0.10 <sup>-2</sup>
Ácido nitrilo triacético	5,6.10 <sup>-2</sup>
FeSO <sub>4</sub> .7H <sub>2</sub> O	2,0.10 <sup>-2</sup>
Biotina	1,0.10 <sup>-4</sup>
Na <sub>2</sub> MoO <sub>4</sub> .2H <sub>2</sub> O	2,0.10 <sup>-3</sup>
MnSO <sub>4</sub> .H <sub>2</sub> O	2,35. 10 <sup>-3</sup>
H <sub>3</sub> BO <sub>3</sub>	2,8.10 <sup>-3</sup>

CuSO <sub>4</sub> .5H <sub>2</sub> O	8,0.10 <sup>-5</sup>
ZnSO <sub>4</sub> .7H <sub>2</sub> O	2,4.10 <sup>-4</sup>
Malato	5,0

Como fonte de nitrogênio utilizou-se 20 mmol.L<sup>-1</sup> de NH<sub>4</sub>Cl. A solução de NH<sub>4</sub>Cl 1mol.L<sup>-1</sup> e a solução de fosfatos foram preparadas separadamente, autoclavadas e acrescentadas ao meio NFbHP-malato no momento do uso. Para o cultivo de SMR1 utilizou-se como antibiótico de seleção, estreptomicina 80 µg.mL<sup>-1</sup>. As culturas foram incubadas à 30 °C com agitação de 120 rpm.

Para o cultivo de *Oryza sativa* foi utilizado o meio Hoagland que possui a seguinte composição:

### **Solução de Hoagland semi-sólido, sem fonte de nitrogênio**

KH <sub>2</sub> PO <sub>4</sub>	0,136 g.L <sup>-1</sup>
K <sub>2</sub> HPO <sub>4</sub>	0,174 g.L <sup>-1</sup>
CaSO <sub>4</sub> .2H <sub>2</sub> O	0,172 g.L <sup>-1</sup>
MgSO <sub>4</sub> .7H <sub>2</sub> O	0,25 g.L <sup>-1</sup>
Solução de micronutrientes <sup>a</sup>	1,0 mL
Solução de ferro <sup>b</sup>	1,0 mL
Ágar	0,2 %

#### **a- Solução de micronutrientes**

	mg.L <sup>-1</sup>
H <sub>3</sub> BO <sub>3</sub>	2,86
MnSO <sub>4</sub> .4H <sub>2</sub> O	1,81
ZnSO <sub>4</sub> .7H <sub>2</sub> O	0,22
CuSO <sub>4</sub> .5 H <sub>2</sub> O	0,08
Na <sub>2</sub> MoO <sub>4</sub> .2 H <sub>2</sub> O	0,02

#### **b- Solução de ferro**

	g.L <sup>-1</sup>
Na <sub>2</sub> H <sub>2</sub> EDTA	1,21
FeCl <sub>3</sub>	0,60

O pH foi ajustado entre 6,5 a 7,0 com NaOH 0,1 M.

### 4.3 PREPARO DE MATERIAL LIVRE DE RNase

O material utilizado para o isolamento e manuseio de RNA (vidraria e material plástico) foi esterilizado e tratado para eliminar RNases exógenas. Todos os procedimentos foram feitos com luvas, que foram trocadas quando houve contato com a superfície e/ou material não tratado com DEPC (dietil pirocarbonato). As vidrarias foram incubadas em DEPC 0,1 % (v/v), enxaguadas com água previamente tratada com DEPC e esterilizadas por autoclavação. Os tubos de microcentrífuga de 1,5 mL usados no manuseio de RNA eram certificados como livres de nucleases pelo fabricante. Todas as soluções, com exceção de Tris-HCl e solventes orgânicos foram preparadas com água previamente tratada com DEPC e esterilizada.

### 4.4 MANIPULAÇÕES DE DNA

A purificação de plasmídeos baseada em lise alcalina, preparo de células eletrocompetentes e eletroforese em gel de agarose foram feitas como descrito (SAMBROOK, FRITSCH, MANIATIS, 1989; MILLER, 1990). A digestão com endonuclease de restrição e ligação DNA inserto:vetor foram feitas como recomendadas pelo fabricante.

#### 4.4.1 Preparo de Células eletrocompetentes

A estirpe de *E. coli* TOP10 (Invitrogen) foi cultivada em 100 mL de meio SOB até atingir  $DO_{600nm}$  entre 0,5 e 0,7. As células foram submetidas a centrifugação a 2.800 g por 5 minutos a 4°C. Após centrifugação, o sobrenadante foi descartado e as células foram lavadas com 30 mL de água ultra pura estéril gelada. Esta etapa foi repetida 2 vezes e então as células foram lavadas 1 vez com 10 mL de solução de glicerol 15% estéril gelada. O sedimento foi ressuspenso em 500  $\mu$ L de glicerol 15% e alíquotas das células (100  $\mu$ L) foram distribuídas em tubos de 1,5 mL e estocados a -70° C.

#### 4.4.2 Transformação bacteriana

As células competentes de *E. coli* TOP10 foram transformadas por eletroporação com 0,5  $\mu\text{L}$  da reação de ligação, utilizando Gene Pulser II (BioRad) configurado com os parâmetros adequados para *E. coli* e sugeridos pelo fabricante. Após transferência para 1 mL de meio de cultura SOC e incubação por 30 minutos à 37 °C, as células foram semeadas em placas de petri contendo meio de cultura LA com 30  $\mu\text{g.mL}^{-1}$  de cloranfenicol. As colônias foram analisadas após aproximadamente 16 horas de crescimento a 37 °C.

#### 4.4.3 Extração de DNA plasmidial em placas de 96 poços

Após obtenção de colônias transformantes, estas foram coletadas e inoculadas em 1,25 mL de meio Terrific Broth contendo 30  $\mu\text{g.mL}^{-1}$  de cloranfenicol em blocos de 96 poços e crescidas por 16 horas sob agitação a 37 °C. As células foram coletadas por centrifugação (2.057 *g*, 7 minutos), lavadas em 180  $\mu\text{L}$  de solução GET (Tris-HCl 25  $\text{mmol.L}^{-1}$ , pH 8,0; glicose 50  $\text{mmol.L}^{-1}$  e EDTA 10  $\text{mmol.L}^{-1}$ ). Novamente as células foram centrifugadas a 2.057 *g* por 7 minutos e o sobrenadante descartado. Em seguida as células foram ressuspensas em 80  $\mu\text{L}$  de GET com 0,125  $\text{mg.mL}^{-1}$  de RNase e lisadas com 80  $\mu\text{L}$  de solução de lise (NaOH 0,2 N e SDS 1 %). As proteínas, o DNA cromossomal desnaturado, os restos celulares e o SDS foram precipitados com a adição de 80  $\mu\text{L}$  de acetato de potássio 3  $\text{mol.L}^{-1}$  pH 5,5. Em seguida o lisado foi incubado por 10 minutos em temperatura ambiente, 30 minutos a 90 °C e finalmente 15 minutos no gelo. Após centrifugação a 2.057 *g* por 10 minutos sob temperatura de 4 °C o sobrenadante foi coletado (~120 $\mu\text{L}$ ) e transferido para placa de filtro Millipore (MAGV N22) contendo placa de coleta acoplada. As amostras foram, então, centrifugadas por 5 minutos a 2.057 *g* sob temperatura de 20 °C. O DNA plasmidial foi precipitado com 0,6 volumes de isopropanol, centrifugado por 45 minutos a 2.057 *g* e 20 °C, lavado com 150  $\mu\text{L}$  de etanol 70 %, seco a 37 °C por 1 hora e dissolvido em 30  $\mu\text{L}$  de água ultrapura estéril. Para avaliar o tamanho dos insertos obtidos, coletou-se 5  $\mu\text{L}$  dos plasmídeos obtidos na extração plasmidial, acrescentou-se tampão universal para todas as enzimas (OPA) 1X e 0,5  $\mu\text{L}$  de *HindIII* e 0,5  $\mu\text{L}$  de *EcoRI* (10 unidades/ $\mu\text{L}$ ) num total de 20  $\mu\text{L}$ .

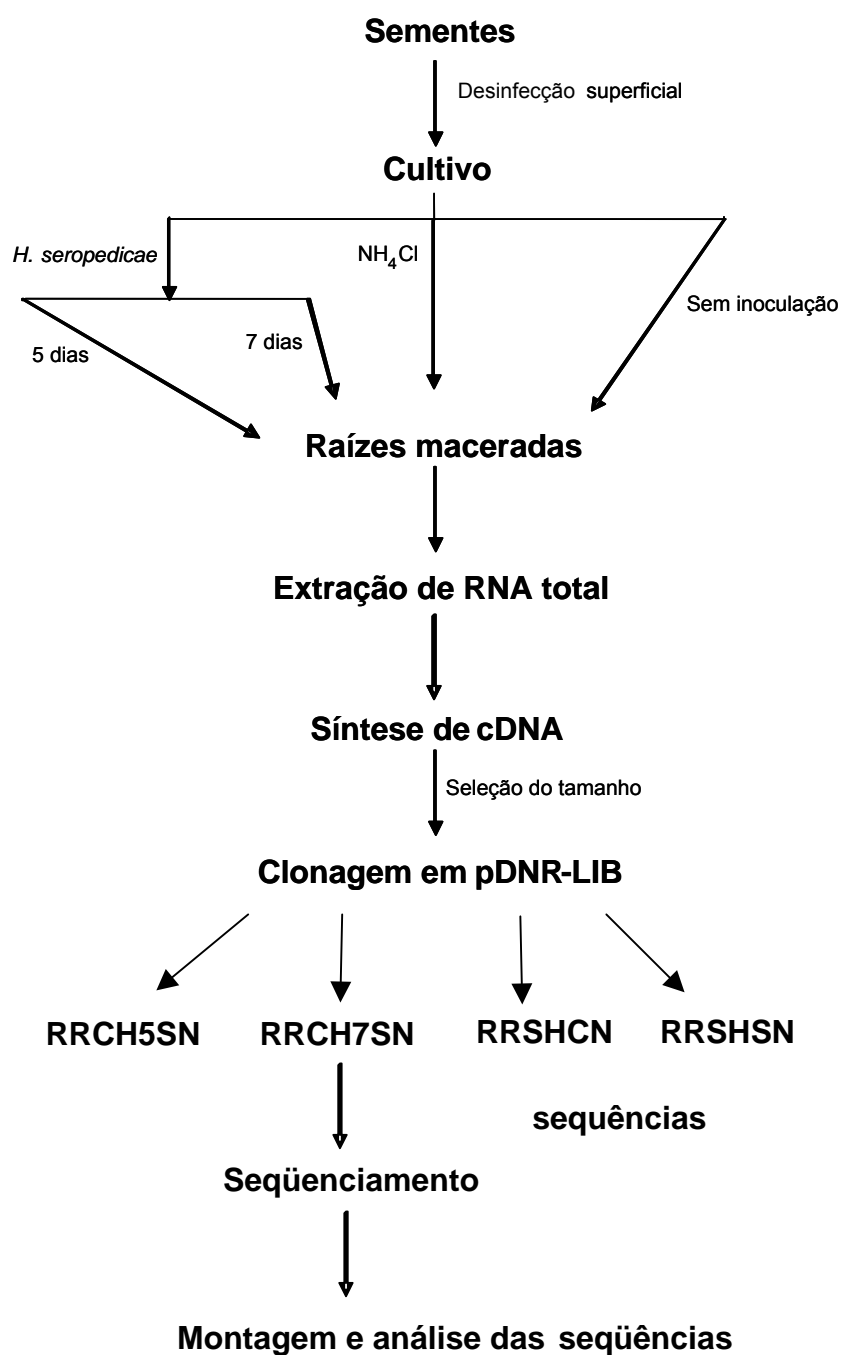
Após incubação a 37 °C por 2 horas as amostras foram aplicadas em gel de agarose 1 %, coradas com brometo de etídeo e visualizadas em transiluminador de ultravioleta (SAMBROOK et al., 1989).

#### 4.4.4 Reação de Seqüenciamento

O método de seqüenciamento utilizado tem como princípio a incorporação de dideoxynucleotídeo marcados com material fluorescente. Para esta reação foi utilizados cerca de 100 a 400 ng de DNA purificado, 3,75 pmol do oligonucleotídeo iniciador (*primer reverso*: TTTCACACAGGAAACAGCTATGAC; *primer universal*: GACGTTGTAAAACGACGGCCAGT), 3 µL de reativo de seqüenciamento (ET terminator) num volume final de 7,5 µL, utilizando placas de 96 poços. O *primer reverso* permite a leitura da extremidade 5' do cDNA, correspondente a cauda poli A do mRNA. Os parâmetros para o termociclador foram os seguintes: 1 ciclo de 1 minuto a 95 °C, seguido por 35 ciclos de 94 °C por 20 segundos e 60 °C por 1 minuto e 30 segundos.

#### 4.5 OBTENÇÃO DE EST DE *Oryza sativa*

Para a obtenção de EST de arroz (*Oryza sativa*) as sementes sofreram uma desinfecção superficial, descritas posteriormente, seguida pelo seu cultivo em meio semi-sólido Hoagland. Após 7 dias de crescimento as plantas foram submetidas a quatro tratamentos (inoculação com *H. seropedicae* por 5 ou 7 dias, adição de NH<sub>4</sub>Cl e controle sem inoculação). Após os tratamentos, o RNA total das raízes foi extraído e utilizado para a síntese de cDNA, que foi clonado no vetor pDNR-LIB (Figura 1). Os insertos foram parcialmente seqüenciados e as seqüências obtidas foram editadas e analisadas utilizando ferramentas de bioinformática (Figura 4).

FIGURA 1 – FLUXOGRAMA DE OBTENÇÃO DE EST DE *Oryza sativa*

#### 4.5.1 Cultivar utilizada

Em todo o delineamento experimental foi utilizado arroz (*Oryza sativa*) da subespécie *japonica*, cultivar Cateto Zebu, gentilmente cedido pelo Centro Nacional de Pesquisa de Arroz e Feijão (EMBRAPA-CNPAP).

#### 4.5.2 Desinfecção superficial das sementes

Inicialmente, o tegumento das sementes foi removido, uma vez que este possui muitos microrganismos e dificulta a eficiência da desinfecção (BALDANI, 1996). As sementes foram em seguida imersas em solução de etanol 70 % por 4 minutos e então incubadas por 30 minutos em solução de hipoclorito 11 % contendo 0,1 % de Triton X-100. Após esta etapa as sementes foram lavadas várias vezes com água ultrapura (Milli-Q) estéril com incubações intercaladas, durante várias horas.

#### 4.5.3 Condições de cultivo das plantas de arroz

As sementes esterilizadas foram colocadas em tubos de ensaio contendo 20 mL de solução de Hoagland semi-sólido e então incubadas em câmara de cultivo sob condições de fotoperíodo de 14h luz/10h escuro, temperatura de 24 °C e umidade superior a 70 % por 7 dias. No sétimo dia foram inoculadas aproximadamente  $10^7$  células.mL<sup>-1</sup> (0,1 mL da cultura DO<sub>600</sub>=1, equivalente a  $10^8$  célula.mL<sup>-1</sup>) de *H. seropedicae* SMR1 no meio de cultura das plantas e incubadas por cinco dias, sendo a biblioteca gerada a partir desse tecido nomeada RRCH5SN. A biblioteca obtida a partir de plantas (aproximadamente 3) incubadas por 7 dias foi nomeada de RRCH7SN. Plantas controle sem nitrogênio adicionado ao meio de cultura e sem o diazotrofo também tiveram seu RNA extraído, e a partir deste foi gerada a biblioteca RRSN0SN. A biblioteca RRSN0CN foi gerada a partir de tecido de arroz onde no sétimo dia de crescimento foi adicionado 4 mmol.L<sup>-1</sup> de cloreto de amônio ao meio de cultura.

Com 14 dias de idade as plantas tiveram suas raízes cortadas e congeladas em nitrogênio líquido para posterior extração de RNA. Um total de 12 plantas controle e 12 plantas inoculadas com *H. seropedicae* foram mantidas em cultivo e



tiveram seu peso seco determinado após 4 dias (18° dia), conforme metodologia descrita por BALDANI V.L., BALDANI J.I., DÖBEREINER (2000).

#### 4.5.4 Controle de contaminantes e eficiência de infecção

Para certificar que as plantas não estavam contaminadas com microrganismos, foi feito um macerado das plântulas. No dia da inoculação com *H. seropedicae* uma planta de cada tubo foi pesada e macerada em 1 mL de meio Nfb-malato. A partir deste, diluições seriadas foram feitas e plaqueadas em meio NFbHP-malato sólido e incubadas a 30 °C por 3-5 dias para observação. Quando foram observados contaminantes, as plantas contaminadas foram descartadas. O mesmo procedimento foi realizado no momento da coleta das raízes e o número de bactérias presentes no tecido radicular foi determinado.

#### 4.5.5 Extração de RNA

A extração de RNA total foi realizada com o *kit* comercializado pela Qiagen, denominado *Rneasy Plant Mini Kit*, seguindo as recomendações do fabricante. Neste processo as raízes foram maceradas em nitrogênio líquido e transferidas para tubo livre de Rnase ao qual adicionou-se tampão de extração. O lisado foi transferido para uma coluna que tem como função separar os restos celulares do RNA. O filtrado contendo o RNA é então transferido para outra coluna de afinidade pelo RNA. Esta segunda coluna é importante para que o RNA possa ser lavado, eliminando outros componentes celulares, auxiliando na purificação da amostra. Todo o processo foi realizado com material livre de Rnase.

#### 4.5.6 Quantificação de RNA total

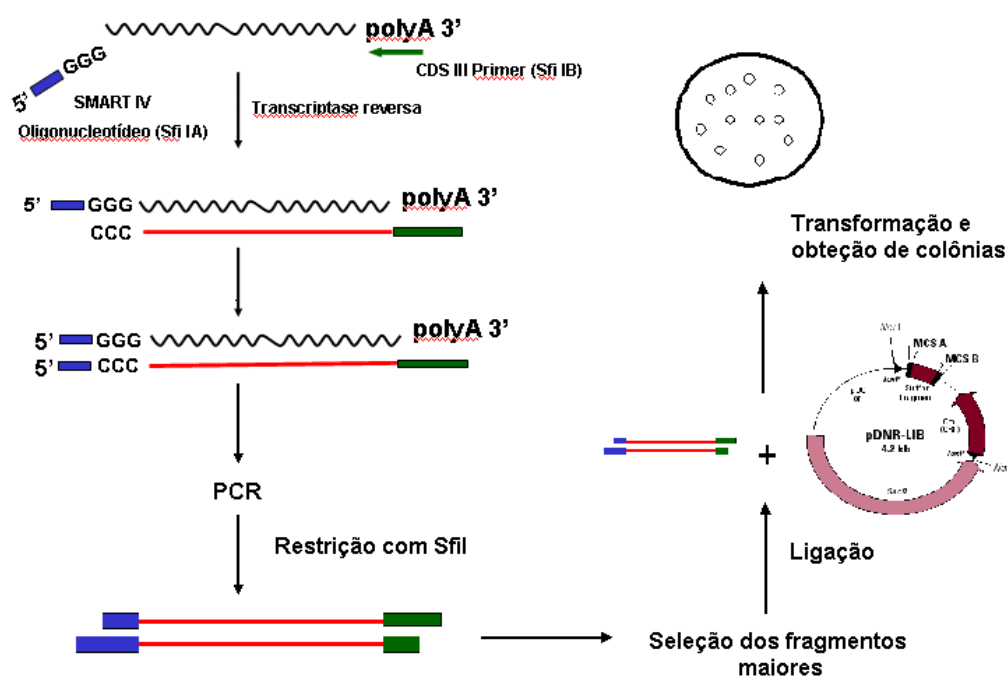
A concentração de RNA total foi determinada em uma amostra diluída em água tratada com 0,1 % de DEPC (dietilpirocarbonato) e lida em espectrofotômetro Beckman DU 7400 à 260 e 280 nm e considerando que 1 unidade de absorbância a 260 nm equivale a 40  $\mu\text{g.mL}^{-1}$  de RNA. A pureza da amostra de RNA foi analisada através da proporção  $A_{260}/A_{280}$  onde valores entre 1,7 e 2 foram considerados adequados. Alternativamente, foi utilizado o Kit Qubit comercializado pela Invitrogen,

que utiliza moléculas fluorescentes para detectar pequenas quantidades de moléculas de RNA.

#### 4.5.7 Síntese de cDNA

Para a construção das bibliotecas, desde a síntese de cDNA até a reação de ligação, foi utilizado o *kit Creator SMART cDNA Library Construction* da Clontech TAKARA BIO Company (Figura 2). A síntese da fita de cDNA é realizada pela transcriptase reversa a partir de um *primer* de oligo-dT modificado (CDSIII/3' PCR Primer) que anela na cauda poli-A dos RNAs mensageiros. Ao atingir a extremidade 5' do mRNA molde a transcriptase reversa adiciona uma cauda de desoxicitidina criando um molde estendido. Um segundo oligonucleotídeo, chamado SMART IV, possuindo uma seqüência oligo (G) na sua extremidade 3' e sítio de restrição para a enzima *SfiI*, pareia-se com tal cauda de desoxicitidina. Ao encontrar tal extremidade, a transcriptase reversa (TR) converte-a em DNA, produzindo uma molécula com a seqüência completa do RNAm, além do oligo SMART IV, que pode ser utilizado para amplificação do fragmento. Dessa maneira obteve-se cDNAs completos a partir de RNA total.

FIGURA 2 – OBTENÇÃO DE BIBLIOTECAS DE cDNA A PARTIR DE RNA TOTAL



#### 4.5.8 Reação de PCR

A reação de PCR foi realizada utilizando 1  $\mu\text{L}$  de cDNA, obtido a partir de aproximadamente 30 ng de RNA total, num volume final de 50  $\mu\text{L}$  e os seguintes parâmetros: um ciclo de 1 minuto à 95 °C, seguido por diversos ciclos de 5 segundos à 95 °C e 6 minutos à 68 °C. Para a biblioteca RRCH7SN foram realizados 20 ciclos e para as demais bibliotecas, 23 ciclos.

#### 4.5.9 Digestão com proteinase K

Após a amplificação do cDNA, a Taq DNA polimerase foi inativada pela adição de 2  $\mu\text{L}$  de proteinase K (20  $\mu\text{g}/\mu\text{L}$ ) e o sistema foi incubado a 45 °C por 20 minutos. Em seguida o cDNA foi extraído com 100  $\mu\text{L}$  de fenol:clorofórmio:álcool isoamílico (25:24:1), que após homogeneização no vórtex e incubação na bancada por 15 minutos foi centrifugado a 12.100  $g$  por 5 minutos. O sobrenadante foi transferido para novo tubo e a fase orgânica foi lavada com 50  $\mu\text{L}$  de água tratada com 0,1 % de DEPC (dietilpirocarbonato) e centrifugou-se a 15.000  $g$  por mais 5 minutos. Os dois sobrenadantes foram unidos e o cDNA precipitado na presença de 10  $\mu\text{L}$  de acetato de sódio 3  $\text{mol}\cdot\text{L}^{-1}$  pH 4,8, 1,3  $\mu\text{L}$  de glicogênio (20  $\mu\text{g}\cdot\mu\text{L}^{-1}$ ) e 260  $\mu\text{L}$  de etanol 96%. Após 30 minutos a amostra foi centrifugada por 20 minutos, o sobrenadante foi removido e a amostra foi lavada com 100  $\mu\text{L}$  de etanol 80 %. Após descarte do sobrenadante o sedimento foi seco sob vácuo por 15 minutos e dissolvido em 50  $\mu\text{L}$  de água tratada com DEPC.

#### 4.5.10 Restrição com *Sfi*I

Os fragmentos de cDNA foram digeridos com a enzima de restrição *Sfi*I, para permitir sua posterior ligação ao vetor pDNR-LIB. Para isto combinou-se 39,5  $\mu\text{L}$  de cDNA tratado com proteinase K a 5  $\mu\text{L}$  de tampão para *Sfi*I 10X, 0,5  $\mu\text{L}$  de BSA 100X e 5  $\mu\text{L}$  de *Sfi*I. A mistura foi incubada a 50 °C por 3 horas.

#### 4.5.11 Fracionamento por tamanho dos fragmentos de cDNA

Para selecionar os maiores fragmentos de cDNA utilizou-se a coluna CHROMA SPIN- 400. Primeiramente a matriz da coluna foi ressuspensa em tampão fornecido pelo *kit* e eluído observando se a coluna seguia os padrões esperados. Em seguida, 50  $\mu\text{L}$  de cDNA contendo 1  $\mu\text{L}$  de xileno cianol 1 % foi aplicado à superfície da resina, que após absorção completa foi lavada com 100  $\mu\text{L}$  do mesmo tampão e então se iniciou a coleta das frações, utilizando 600  $\mu\text{L}$  de tampão. Diversas frações foram coletadas e unidas de acordo com seu tamanho, após análise eletroforética.

As amostras foram precipitadas adicionando-se 0,1 volume de acetato de sódio  $3\text{mol.L}^{-1}$  pH 4,8, 1,3  $\mu\text{L}$  de glicogênio ( $20\ \mu\text{g}.\mu\text{L}^{-1}$ ) e 2,5 volume de etanol 96 %. Após incubação a  $-20^\circ\text{C}$  durante a noite as amostras foram centrifugadas a  $15.000\ \text{g}$  por 20 minutos, o sobrenadante descartado e o precipitado foi secado a vácuo por 15 minutos. As amostras foram dissolvidas em 3  $\mu\text{L}$  de água deionizada e 0,5  $\mu\text{L}$  de cada amostra foi aplicado em gel de agarose 1,5% para eletroforese a 50 V por aproximadamente 3 horas.

#### 4.5.12 Ligação dos fragmentos de cDNA ao vetor

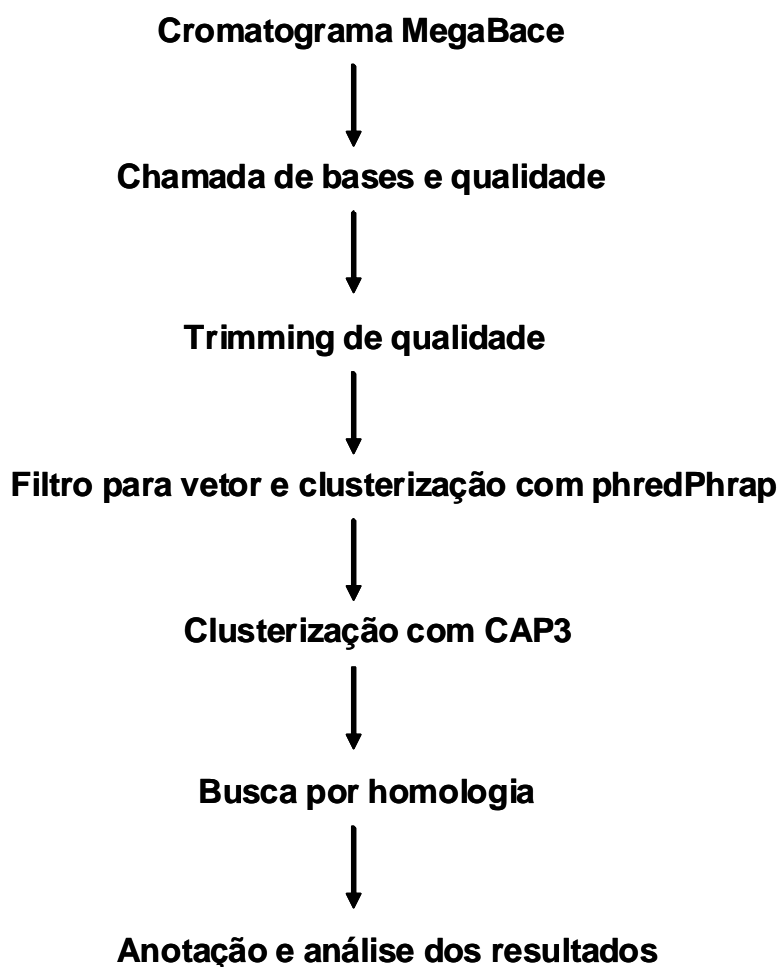
Para a reação de ligação dos fragmentos de cDNA, utilizou-se o vetor pDNR-LIB digerido com *SfiI* (fornecido pelo *kit*). Uma vez que tanto os fragmentos quanto o vetor possuem dois sítios de restrição diferentes para *SfiI* em suas extremidades, a ligação ocorra somente entre fragmentos e vetor. A reação de ligação foi feita utilizando 2,5  $\mu\text{L}$  de cDNA, 1  $\mu\text{L}$  de pDNR LIB digerido com *SfiI*, 0,5  $\mu\text{L}$  de tampão  $T_4$  DNA ligase 10X, 0,5  $\mu\text{L}$  de ATP ( $10\text{mmol.L}^{-1}$ ) e 0,5  $\mu\text{L}$  (200 unidades) de  $T_4$  DNA ligase. A reação foi incubada na geladeira à  $8^\circ\text{C}$  por 2 a 3 dias para então ser transformada em *E. coli*.

#### 4.6 ANÁLISE COMPUTACIONAL DE ESTs

As análises das seqüências foram feitas através de vários *scripts* desenvolvidos em linguagem BASH. Os *scripts* automatizam desde a criação de diretórios e nomeação dos arquivos de seqüência até a análise e identificação das seqüências contíguas obtidas na etapa de agrupamento (“clusterização”). O

fluxograma de análise é mostrado nas Figuras 3 e 4 e a Tabela 2 apresenta a lista dos *scripts* desenvolvidos. Os *scripts* estão ainda integrados a alguns programas de bioinformática que executam tarefas específicas (Tabela 3).

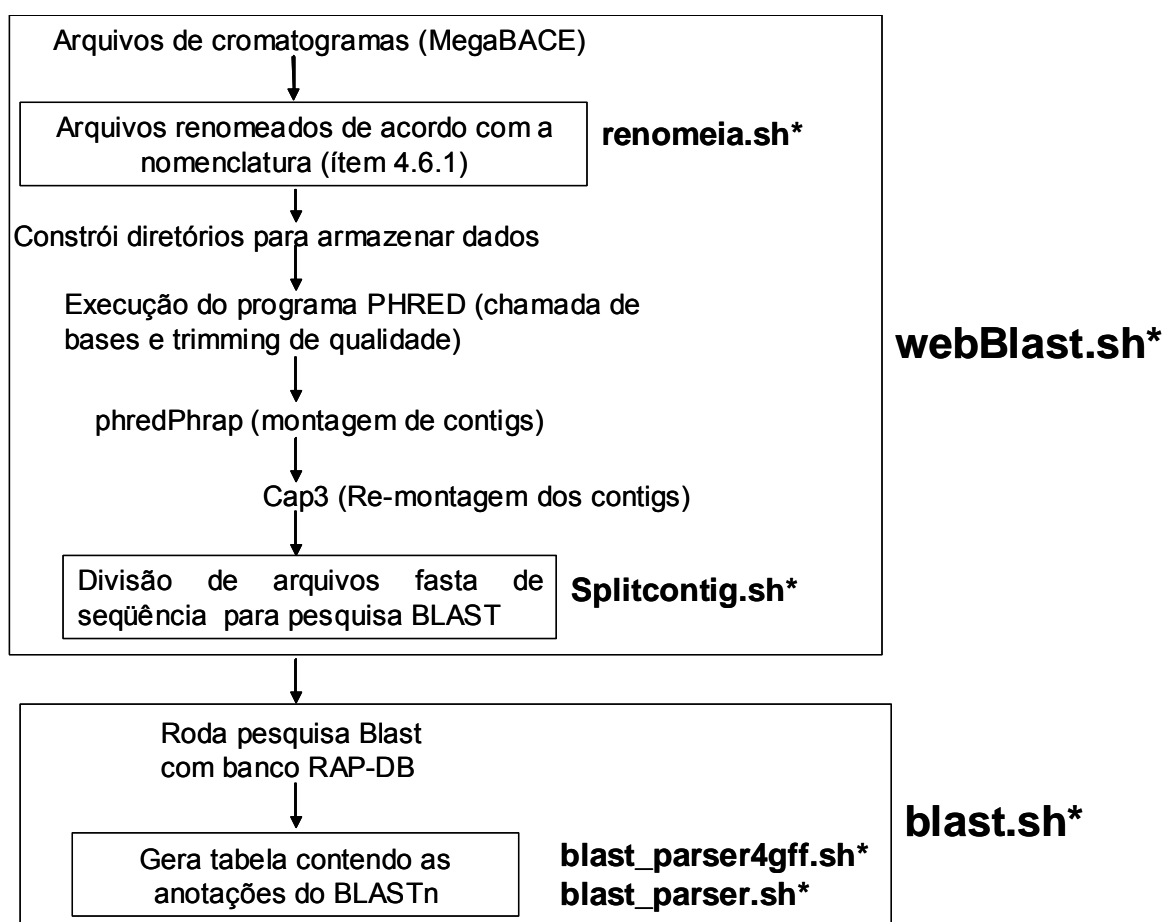
FIGURA 3 – FLUXOGRAMA GERAL DA ANÁLISE COMPUTACIONAL DE EST



#### 4.6.1 Nomenclatura utilizada para nomear as amostras dos cromatogramas

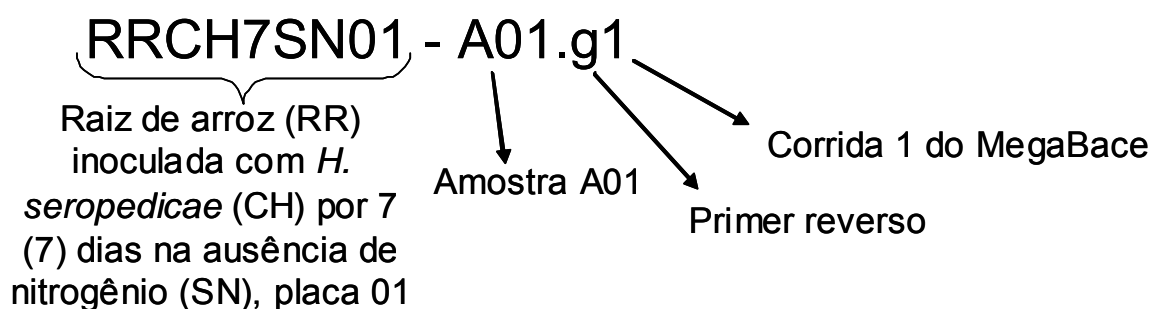
Como mostra a Figura 5 para nomear os cromatogramas produzidos pelo MegaBace, utilizou-se a nomenclatura adotada contendo o nome da biblioteca (RRCH7SN), o número da placa, o nome da amostra, o primer utilizado para o seqüenciamento (g: reverso; b: universal) e o número da corrida no Megabace.

FIGURA 4 – FLUXOGRAMA DE ANÁLISE DE SEQÜÊNCIAS EST DE ARROZ.



\* “script” utilizado para executar a tarefa da caixa ao lado correspondente. O código fonte dos *scripts* está apresentado em anexo.

FIGURA 5 – NOMENCLATURA UTILIZADA PARA NOMEAR CROMATOGRAMAS



#### 4.6.2 Scripts para edição e alinhamento das seqüências

O *script* WebBlast.sh (ANEXO 3) realizou várias etapas desde a criação de diretórios a análise de comparação das seqüências. As etapas executadas pelos *scripts* estão descritas na Tabela 2. Dessa forma, primeiro foram criados os diretórios para armazenamento de dados gerados pelos programas. Em seguida o arquivo da seqüência do vetor foi copiado para o diretório raiz do projeto criado. As seqüências então passaram por um filtro de vetor e "*triming*" de qualidade, onde o comando: **phred -id ../chromat\_dir/ -trim\_alt ../chromat\_dir -trim\_phd -pd ../phd\_dir**, realizado no *software* Phred, seleciona seqüências de boa qualidade e mascara seqüências de vetor.

HOUDE et al (2006) ao analisar seus resultados de EST de trigo observou que os resultados do programa "Contig Assembly Program" (CAP3) produziram resultados mais consistentes que o Phrap e sugere CAP3 como o *software* mais apropriado para montagem de ESTs. Assim utilizou-se para a *clusterização* o CAP3 e esse constitui o próximo passo do *script*. O programa foi executado com as opções **-o 30 -p 90**, que requer um mínimo de 30pb para efetuar o alinhamento entre duas seqüências, com uma identidade de pelo menos 90%.

TABELA 2 – LISTA DOS "SCRIPTS" CRIADOS E UTILIZADOS

Script	Anexo	Atribuições
renomeia.sh	1	Renomear os cromatogramas gerado pelo seqüenciador megaBase
Splitfastafila.sh	2	Dividir os arquivos fasta dos <i>contigs</i> e <i>singlets</i> em arquivos fasta separados
WebBlast.sh	3	Integrar a execução de todos os programas e <i>scripts</i> desde renomear diretórios a execução do Cap3
blast.sh	4	Executar a pesquisa Blast e obter informações do resultado do Blast para gerar tabela
blast_parser4gff.sh	5	Obter resultado de anotações da seqüência com maior e-value do banco de dados RAP-DB
Blast_parser.sh	6	Obter resultados sobre o alinhamento do blast como valor e-value e identidade

#### 4.6.3 Banco de dados público utilizado para a comparação

Um banco de dados com informações sobre anotação chamado "Banco de dados RAP" (RAP-DB; *Rice Annotation Project - Data Base*) foi desenvolvido pelo *International Rice Genome Sequencing* (IRGSP). Este banco integra a seqüência do genoma de *Oryza sativa* subespécie *japonica* de IRGSP com as anotações do projeto de anotação (RAP-DB) (OHYANAGI et al, 2006).

O RAP-DB além de conter a seqüência do genoma apresenta o *loci* RAP com o correspondente lócus IDs representando os genes anotados. Cada lócus possui um ou mais transcritos variantes com seqüência anotada RAP. Regiões previstas como codificadora de proteína foram também empregadas como loci previsto RAP. Os transcritos do *Institute for Genomic Research* (TIGR) derivado de anotações da montagem TIGR foi adicionado ao RAP-DB através de mapeamento destes no genoma IRGSP. Cada transcrito apresenta *links* para o *Gene ontology*, informações sobre o domínio ou motivo estrutural, informações sobre a expressão do transcrito (OHYANAGI et al, 2006).



#### 4.6.4 *Scripts* para a comparação das seqüências com o banco de dados público

Para fazer a comparação das seqüências obtidas os arquivos *fasta.contigs* e *fasta.singlets* gerados pelo CAP3 precisam ser divididos em arquivos separados, um para cada *contig*. Para esta etapa o *script* *WebBlast.sh* (ANEXO 3) executa o *script* *Splitfastafile.sh* (ANEXO 2). Os arquivos criados são então utilizados na pesquisa de comparação com o banco de dados RAP-DB de arroz do IRGSP (<http://rapdb.lab.nig.ac.jp/index.html>), utilizando a ferramenta BLAST. As análises foram feitas utilizando apenas comparação com nucleotídeos (BLASTn).

Para a obtenção dos resultados do BLAST foi criado o *script* *Blast\_parser.sh* (ANEXO 6) que pesquisa o arquivo de resultado gerado pelo BLAST e obtém informações como “e-value”, número de acesso do EST do banco de dados, tamanho da seqüência submetida a pesquisa e do banco de dados, valor de identidade e outros. Para obter resultados sobre as informações da anotação foi criado o *script* *blast\_parser4gff.sh* (ANEXO 5). O *script* *Blast.sh* (ANEXO 4) executa os dois últimos *scripts* descritos gerando um arquivo com as informações do alinhamento da pesquisa BLAST. A classificação de EST quanto a sua função foi realizada através de comparação com o bando de dados Swiss-Prot utilizando a terminologia do *Gene Ontolgy* (<http://www.geneontology.org/>).

TABELA 3 - PROGRAMAS DE BIOINFORMÁTICA UTILIZADOS

Programa de bioinformática	Atribuições	Referência
Phred	Recebe os arquivos de cromatogramas, faz a chamada de bases, atribui valor de qualidade para cada base e faz <i>trimming</i> de qualidade.	EWING et al, 1998 EWING, GREEN, 1998
Phrap	Realiza a montagem das seqüências contíguas a partir de regiões de sobreposição	<a href="http://www.phrap.org/phredphrap/phrap.html">http://www.phrap.org/phredphrap/phrap.html</a>
Consed	Edita as seqüências contíguas e permite visualização das montagens	GORDON et al, 1998
Cap3	Realiza montagem das seqüências	HUANG, MADAN, 1999
Blast	Compara seqüências contra um banco de dados	ALTSCHUL et al, 1997

## 5. RESULTADOS E DISCUSSÃO

### 5.1 CULTIVO DE *Oryza sativa*

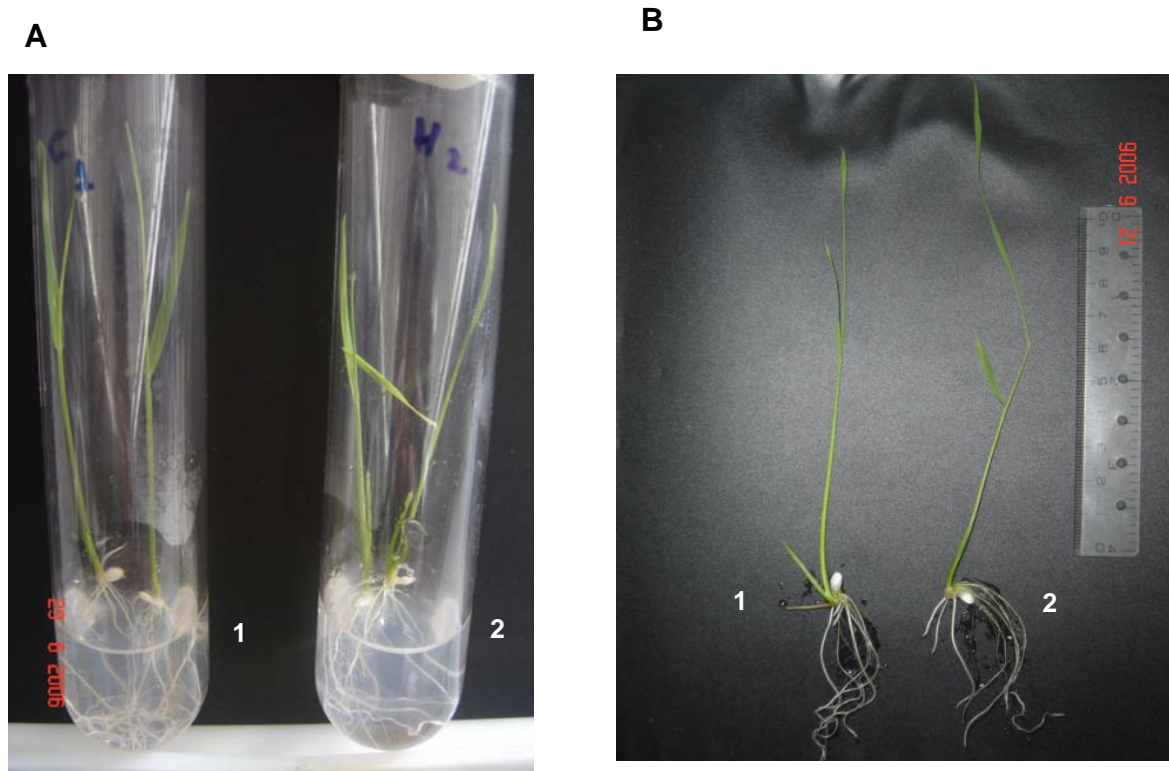
Com o objetivo de identificar genes diferencialmente expressos em plantas de arroz inoculadas com *Herbaspirillum seropedicae* SMR1, sementes de arroz foram germinadas e cultivadas como descrito na metodologia (item 4). O sistema e condições de cultivo de *Oryza sativa* foram otimizados para que a manipulação, inoculação e coleta das raízes induzissem um mínimo de estresse para a planta. O sistema de cultivo *in vitro* foi o mesmo utilizado por BALDANI V.L., BALDANI J.I., DÖBEREINER (2000), com exceção da concentração de ágar do meio de cultivo, que foi de 0,2 %, (originalmente 0,6 %). A menor concentração de ágar facilitou a remoção das plantas no momento da extração de RNA. As sementes desinfectadas foram colocadas para germinar no escuro diretamente sobre a superfície do meio, os tubos foram selados e mantidos em câmara de BOD com temperatura e luminosidade controlada. Como pode ser observadas na Figura 6, as raízes desenvolveram-se no interior do meio semi-sólido, o que permitiu a sustentação das plantas. No 14º dia de crescimento das plantas, o número de bactérias colonizadoras foi determinado através de contagem de unidades formadoras de colônia (UFC). Uma média de  $10,34 \cdot 10^8$  UFC/g de tecido fresco de raiz foi obtida, confirmando a eficiente colonização de arroz por *H. seropedicae*. Não foi recuperada nenhuma bactéria nas plantas não inoculadas.

As plantas inoculadas com *H. seropedicae* não apresentaram diferenças de crescimento evidentes quando comparadas com plantas não inoculadas. Para analisar quantitativamente o tamanho das plantas, 12 plantas controle e 12 inoculadas tiveram seu peso seco determinado. As plantas inoculadas tiveram uma média de peso seco de  $0,017 \text{ g} \pm 0,004$  enquanto as controles tiveram  $0,02 \text{ g} \pm 0,003$ . As semelhanças observadas podem ser decorrentes do curto período do experimento, pois as sementes ainda contavam com sua reserva nutritiva, que sustentou um bom estado nutricional (Figura 6). Além disso, o tempo de inoculação de 7 dias provavelmente não é suficiente para promover diferenças de crescimento das plantas. RONCATO-MACARI e colaboradores (2003) observaram diferença

significativa de crescimento entre plantas de arroz inoculadas e não inoculadas com *H. seropedicae* após 30 dias em laboratório.

Entretanto, a ausência de diferença entre as plantas inoculadas e o controle negativo não interfere nas análises posteriores, uma vez que o objetivo é identificar genes envolvidos na interação planta-bactéria e estes devem ser induzidos já em etapas iniciais após a inoculação. A alteração na expressão gênica induzida pela interação com microrganismos é bem estabelecida em rizóbios, onde mudanças significativas iniciam algumas horas após a inoculação (CÁRDENAS, 2006).

FIGURA 6 – SISTEMA DE CULTIVO E PLANTAS DE *Oryza sativa* COM 14 DIAS DE CRESCIMENTO



Legenda:

A

- 1 – planta controle com 14 dias, não inoculado e sem nitrogênio adicionado ao meio de cultura.  
2 – planta com 14 dias, inoculada com *H. seropedicae* SMR1 no 7º dia ( $10^8$  células.mL<sup>-1</sup>).

B

- 1– planta controle com 14 dias, não inoculado e sem nitrogênio adicionado ao meio de cultura.  
2– planta com 14 dias, inoculada com *H. seropedicae* SMR1 no 7º dia ( $10^8$  células.mL<sup>-1</sup>).

## 5.2 EXTRAÇÃO DE RNA TOTAL E CONSTRUÇÃO DE BIBLIOTECAS DE cDNA

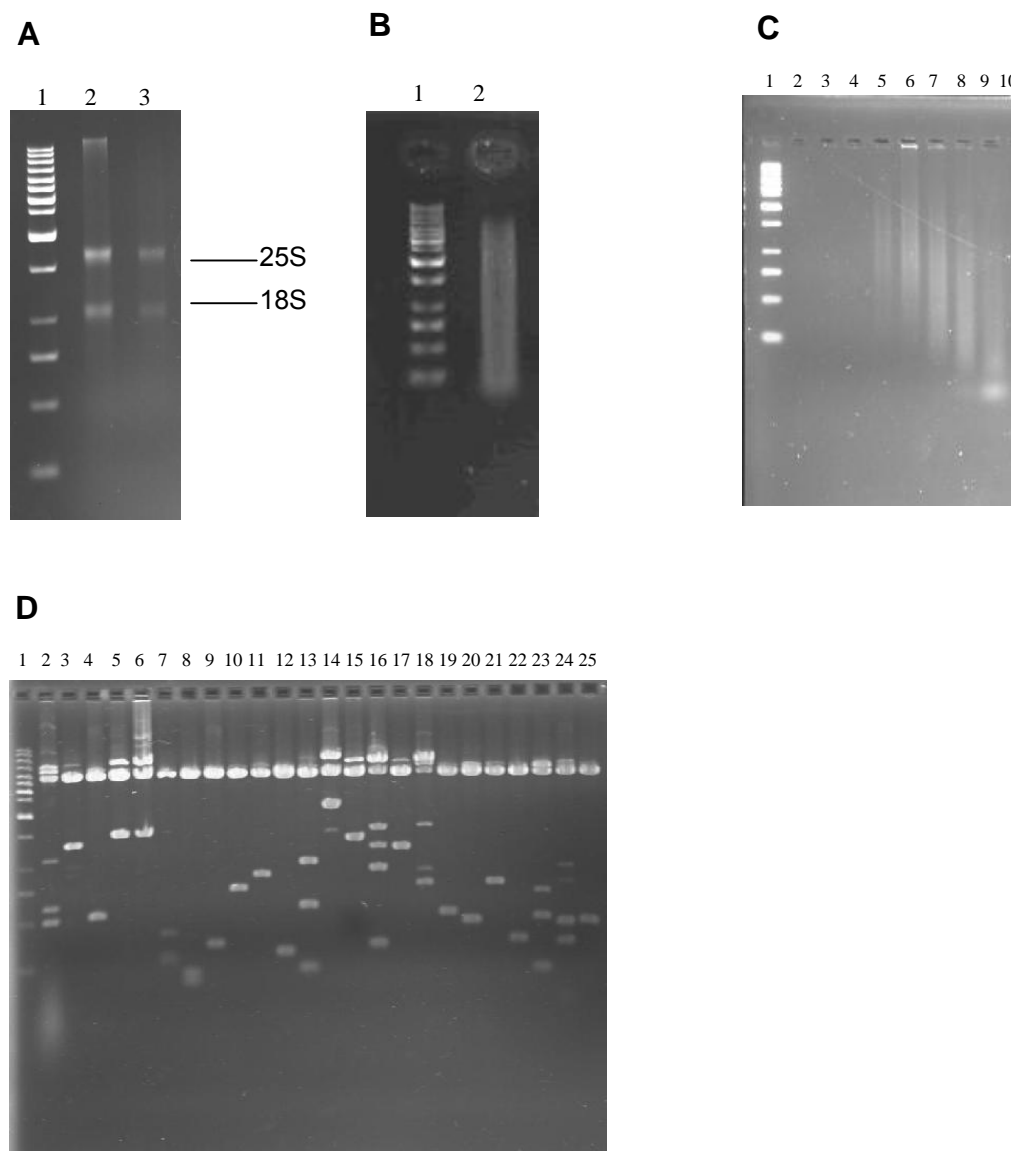
As raízes de arroz submetidas a diferentes tratamentos tiveram seu RNA total extraído e este foi utilizado como molde para a síntese de cDNA (Figura 7 A). O cDNA foi amplificado com oligonucleotídeos contendo sítios *Sfi*I A (5') e *Sfi*I B (3') (Figura 7 B). Uma vez que a síntese de cDNA foi realizada utilizando iniciador oligo-dT, o produto obtido deve conter apenas cópias de mRNA. O cDNA obtido foi clivado com a enzima de restrição *Sfi*I, os fragmentos com mais de 500pb selecionados (Figura 7 C) e ligados ao vetor pDNR-LIB. Após ligação e transformação foram obtidas diversas colônias (Tabela 4) das quais uma amostra de 24 foram submetidas à extração plasmidial. Os plasmídeos isolados foram digeridos com as enzimas de restrição *Hind*III e *Eco*RI, liberando o inserto de cDNA (Figura 7 D). O tamanho médio dos insertos obtidos está descrito na Tabela 4, sendo tais valores semelhantes aos obtidos anteriormente em outros projetos (RAUYARE et al 2001; REDDY, RAMAKRISHNA, SEKHAR, 2002; VANTELON-DEBOUT et al 2003).

Uma vez obtidas as quatro bibliotecas, deu-se início a etapa de seqüenciamento. Apenas a biblioteca RRCH7SN foi analisada neste trabalho. O seqüenciamento dos clones das outras bibliotecas será concluído posteriormente e permitirá uma análise comparativa entre as bibliotecas.

TABELA 4 – DADOS GERAIS DAS BIBLIOTECAS DE cDNA PRODUZIDAS

<b>Biblioteca</b>	<b>Concentração total de RNA (ng.µL<sup>-1</sup>)</b>	<b>Número de placas estocadas</b>	<b>Média de tamanho de inserto (pb)</b>	<b>Número de amostras seqüenciadas</b>
RRCH7SN	7,84	24	1290	912
RRCH5SN	9,02	31	863	192
RRSHSN	27	29	800	384
RRSHCN	9,88	5	958	96
Total	-	8544 clones	-	1584

FIGURA 7 – PERFIL ELETROFORÉTICO DE RNA TOTAL, cDNA E RESTRIÇÃO DE CLONES

**A - Perfil eletroforético de RNA total extraído de raízes de arroz**

- 1 - marcador de massa molecular 1Kb ladder;
- 2 – RNA total extraído de raízes de arroz não inoculadas com *H. seropedicae*;
- 3 - RNA total extraído de raízes de arroz inoculadas com *H. seropedicae* por 5 dias

**B - Perfil eletroforético de cDNA após reação de PCR**

- 1 – Marcador de massa molecular 1Kb ladder;
- 2 – cDNA amplificado com 20 ciclos a partir de RNA total da coluna 3 da figura 7A;

**C- Perfil eletroforético de fragmentos de cDNA da biblioteca RRCH7SN após coluna de fracionamento**

- 1 – marcador de massa molecular 1Kb ladder
- 2 – vazio
- 3 a 10 – Frações eluídas e unidas duas a duas

**D - Perfil eletroforético após restrição dos clones obtidos**

- 1 – marcador de massa molecular 1Kb ladder
- 2 a 25 - Perfil de restrição de clones aleatórios da biblioteca RRCH7SN.

### 5.3 ANÁLISE DAS SEQÜÊNCIAS DA BIBLIOTECA RRCH7SN

O seqüenciamento da biblioteca RRCH7SN foi realizado utilizando-se o *primer* M13-reverso, que permite seqüenciar a extremidade do cDNA correspondente a região 3' do mRNA (3' EST). Apenas uma placa contendo 96 amostras foi seqüenciada utilizando o *primer* universal. Tradicionalmente, os projetos de seqüenciamento de EST favorecem a leitura da ponta 5' EST por conter uma freqüência maior de região codificadora quando comparada com a ponta 3', que geralmente possui significativa região não traduzida. Todavia, o final 3' EST também é informativo porque comumente oferece seqüências únicas que podem ser utilizadas para distinguir genes parálogos (RUDD, 2003).

Um total de 912 clones foi utilizado no seqüenciamento, porém nem todos produziram seqüência. As regiões de baixa qualidade das extremidades 5' e 3' das seqüências obtidas foram removidas (*trimming* de qualidade) utilizando parâmetros padrão do programa Phred. A figura 8 mostra a distribuição das seqüências de acordo com o seu comprimento e o índice de qualidade, após o processo de *trimming* de qualidade. A maioria das seqüências apresentou uma boa qualidade (índice Phred>15) numa extensão maior que 100 bp, sendo que apenas 14% continha menos que 100 pb com índice de qualidade *Phred* maior ou igual a 15.

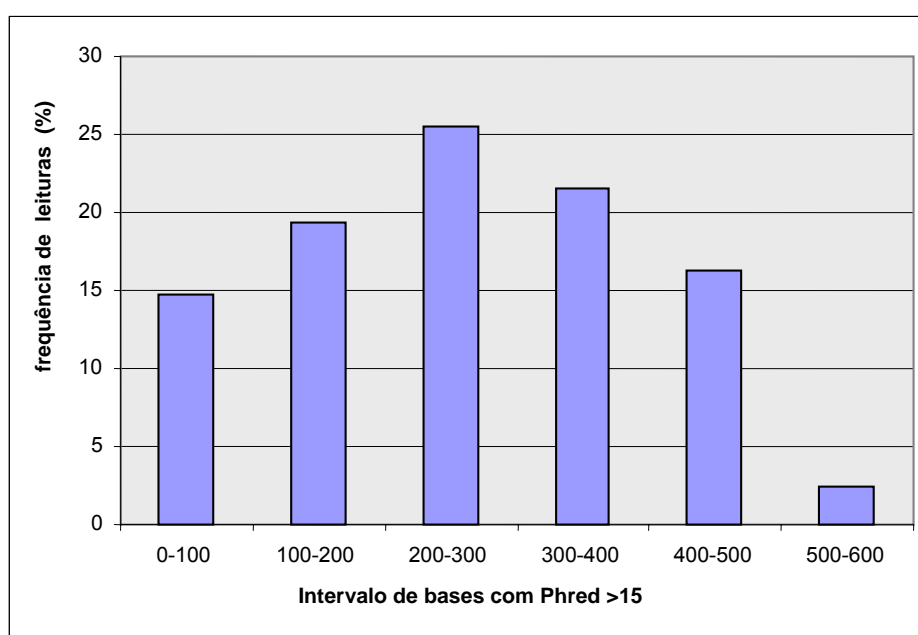
Todas as seqüências obtidas foram utilizadas na montagem, com o objetivo de identificar regiões contíguas e clones idênticos, que podem representar genes mais abundantemente expressos. Como resultado, obteve-se 446 seqüências únicas (*singlets*) e 36 seqüências contíguas (*contigs*), ou seja, 482 ESTs diferentes. Ventelon-Debout e colaboradores (2003) obtiveram aproximadamente 70 % de seqüências únicas quando analisaram mais de 5.000 clones de diferentes bibliotecas de EST de arroz. A maior freqüência observada no presente trabalho (85%) provavelmente deve-se ao menor número de clones seqüenciados. Após a montagem, as seqüências foram comparadas com o banco de dados público RAP-DB (Rice Annotation Project - Data Base) utilizando a ferramenta Blast.

O RAP-DB é resultado de um trabalho cooperativo realizado por várias instituições envolvidas com o seqüenciamento do genoma de arroz. Apesar do enorme número de ESTs de arroz depositadas (mais de um milhão) em todos os

bancos de seqüências, o RAP-DB contém aproximadamente 75.000 seqüências curadas e anotadas, refletindo o número de genes de arroz, estimado em 38.000.

As Tabelas 5 e 6 mostram um resumo dos resultados de comparação com *e-value* menor que  $10^{-6}$  para busca de similaridade no RAP-DB com o programa BLAST.

FIGURA 8 – DISTRIBUIÇÃO DAS SEQÜÊNCIAS OBTIDAS DE ACORDO COM COMPRIMENTO E QUALIDADE



\* Cada coluna representa a porcentagem de leituras contendo o número de bases indicado com Phred maior ou igual a 15. As regiões de baixa qualidade das extremidades das leituras utilizadas foram removidas anteriormente a análise (*trimmig*), sendo que o número total de leituras restantes foi de 499.

Como se observa na Figura 9, 46 % das ESTs analisadas não são similares a nenhum EST do banco de dados (*e-value* >  $10^{-6}$ ). Uma vez que uma dada seqüência não se encontra depositada em bancos de dados de EST, ela pode ser considerada como nova, ou seja, um gene ainda não detectado em nenhuma das condições estudadas. Novos genes são definidos por Reddy et. al. (2002) como ESTs que não mostram alinhamento com qualquer outra seqüência de nucleotídeos no banco de dados utilizando um valor de corte de *e-value* de  $10^{-6}$ . Portanto, 46% das seqüências obtidas neste trabalho correspondem a essa categoria, sendo

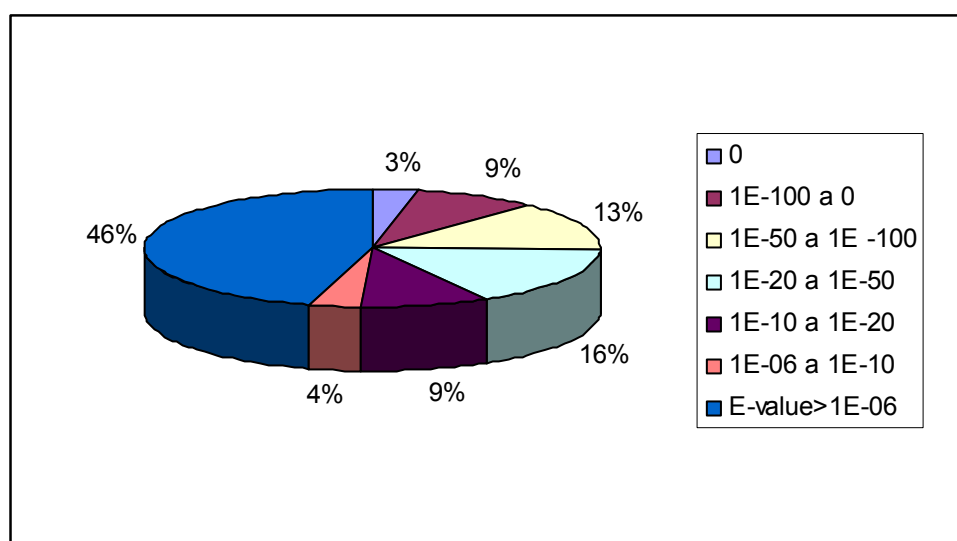


potenciais genes novos. Ventelon-Debout e colaboradores (2003) obtiveram uma frequência média de 42 % de seqüências deste tipo nas diferentes bibliotecas de cDNA de arroz analisadas. Estes autores compararam as seqüências obtidas e traduzidas em aminoácidos com o banco de dados NCBI (*National Center for Biotechnology Information*), o que em princípio poderia diminuir o número de genes novos identificados.

Todas as ESTs obtidas neste trabalho apresentaram alta identidade (97-100%) com o genoma de arroz, confirmando sua origem.

A análise da frequência com que um dado gene é obtido na biblioteca de cDNA reflete sua abundância no tecido ou condição em estudo. A comparação de frequências entre bibliotecas/condições permite identificar genes importantes para o processo. Ao analisarmos o número de clones presentes em cada *contig* obtido, notamos que o número máximo foi observado nos *contigs* 11, 18 e 16, que possuem 7, 6 e 4 leituras, respectivamente. Quando comparado com o banco de dados, estas seqüências possuem maior similaridade com genes envolvidos com síntese de ATP, dobramento de proteínas e peptidases, que correspondem a genes geralmente expressos em altos níveis.

FIGURA 9 – DISTRIBUIÇÃO DO *E-value* DAS SEQÜÊNCIAS ANALISADAS



Legenda: Frequência de *e-value* após alinhamento de nucleotídeos das seqüências obtidas contra o banco RAP-DB.

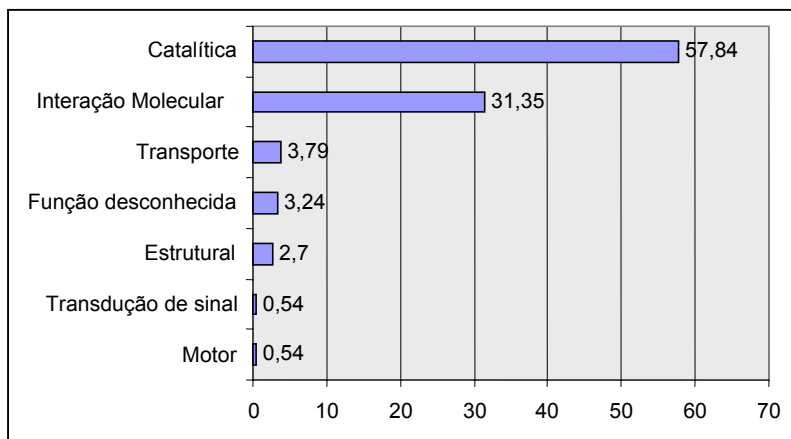
Para determinar as funções das proteínas codificadas pelas ESTs seqüenciadas, estas foram comparadas com o banco de dados (Figura 10) e classificadas usando a terminologia do *Gene Ontology* (GO). Inicialmente estas foram agrupadas de acordo com o processo biológico no qual estão envolvidas, com sua função molecular e quanto ao componente celular onde exercem sua função. Quanto à função molecular, a subclasse “atividade catalítica” foi a mais dominante (57,84 %) seguida pela subclasse “interação molecular” (31,35 %). Quanto à classe “compartimento celular” a maior parte dos genes pertence a subclasse “intracelular” (68,83 %), onde a maioria das proteínas está presente nas organelas celulares e no citoplasma. Em relação à classe “processo biológico” a subclasse mais abundante foi “processo celular” (74,53 %) seguida de “processo metabólico” (14,15 %) e “resposta a um estímulo” (10,38 %). Esta última subclasse se refere a estímulos bióticos e abióticos como estresse, disponibilidade de nutrientes, luz e defesa a patógenos. Os genes agrupados nesta classe podem estar envolvidos também na interação arroz – *H. seropedicae*, mostrando que a biblioteca inoculada pode conter genes expressos em resposta a presença de *H. seropedicae* e poderá ser utilizada em análises comparativas com as bibliotecas não inoculadas. Estes genes são fortes candidatos para futuras análises de expressão individual, através de PCR quantitativa, hibridização, fusão com genes repórteres, etc. Dentre os genes identificados nas bibliotecas com possível envolvimento na interação *H. seropedicae*-arroz encontra-se o do clone denominado RRCH7SN11-G03.g. A seqüência obtida apresenta 95 % de identidade com o gene *Zm-hir3* de milho (*Zea mays*). O gene *Zm-hir3* expressa uma proteína que faz parte de uma família envolvida na resposta hipersensitiva e morte celular programada em plantas (NADIMPALLI et al., 2000). Aparentemente, a atividade da proteína está diretamente envolvida com a regulação de canais iônicos que controlam a permeabilidade celular e induzem a morte celular. Apesar de *H. seropedicae* não induzir nenhum sintoma de doença em arroz, tais proteínas podem estar envolvidas no controle de interação ou nível de colonização.

Apesar de diferentes bibliotecas de cDNA terem sido construídas, utilizando raízes de arroz não inoculadas e inoculadas com *H. seropedicae* e com diferente disponibilidade de nitrogênio (Tabela 4), apenas a biblioteca de raízes inoculadas com *H. seropedicae* e incubadas por 7 dias foi parcialmente seqüenciada e

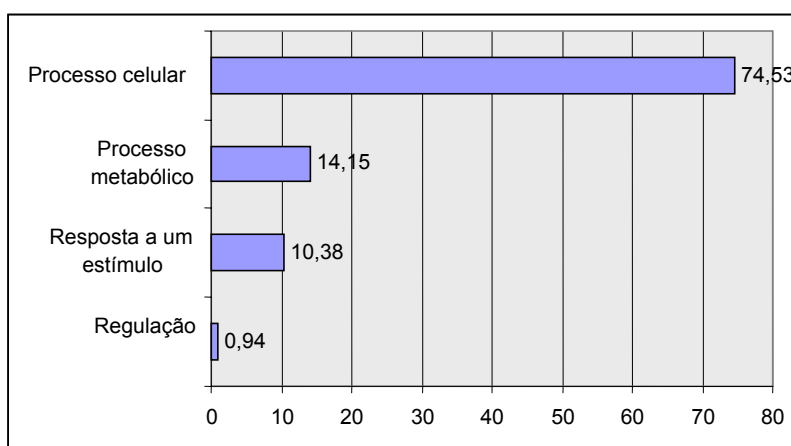
analisada. As análises revelam que uma grande proporção dos genes não tem semelhança com ESTs depositadas nos bancos de dados de arroz. Esta grande proporção pode ser resultado do tamanho da região 3'-UTR observada em mRNAs de eucariotos. Todas as seqüências obtidas foram também comparadas com os dados de seqüenciamento do genoma de arroz e possuem alta similaridade com tais seqüências, o que confirma sua origem. O seqüenciamento de um maior número de clones desta e de outras bibliotecas deverá permitir uma análise comparativa de expressão e a identificação de novos genes candidatos para análise futura, utilizando novas abordagens (CAVALCANTE et al, 2007).

FIGURA 10 – ONTOLOGIA USANDO SWISS-PROT

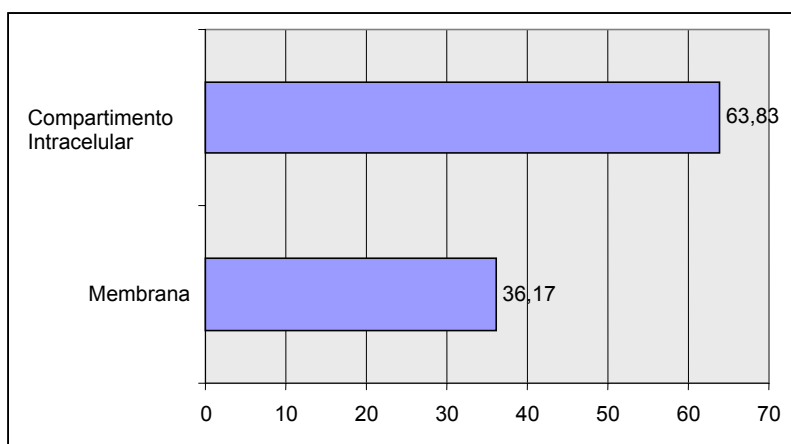
A



B



C



Legenda: A – Ontologia para Função Biológica

B – Ontologia para Processo Biológico

C – Ontologia para Componente Celular

TABELA 5 – RESULTADO DA PESQUISA BLAST PARA OS “SINGLETs”

EST	Número de Acesso RAP-DB	E-value	Provável proteína
RRCH7SN00-A02.g0	Os06g0672600 Os06g0672600	1,00E-31	desconhecida
RRCH7SN00-A08.g0	AK121113 Os02g0252600	2,00E-42	Aminotransferase
RRCH7SN02-D09.g3	AK109426 Os09g0133200	1,00E-87	proteína da família Glicose/ribitol dehidrogenase
RRCH7SN02-E03.g3	AK121942 Os07g0662500	2,00E-60	fator de Elongação 1-beta' (EF-1-beta')
RRCH7SN02-E09.g2	AK067965 Os02g0834700	4,00E-88	hipotética conservada YfCH
RRCH7SN00-B02.g0	AK099023 Os12g0550800	2,00E-14	hipotética
RRCH7SN02-F06.g2	AK061400 Os02g0200900	1,00E-74	rica em repetições de Leucine, subtipo contendo cisteína
RRCH7SN02-F10.g2	AK100508 Os08g0265500	3,00E-45	proteína da família ATPase tipo AFG1
RRCH7SN02-G05.g2	AK067466 Os01g0613900	1,00E-10	proteína de função desconhecida UPF0052 e da família CofD
RRCH7SN00-B03.g0	AK068820 Os11g0220300	5,00E-95	hipotética
RRCH7SN02-G07.g3	AK072680 Os01g0862000	8,00E-52	proteína da família Alfa/beta hidrolase
RRCH7SN02-G08.g3	AK065134 Os02g0818000	1,00E-78	proteína contendo domínio CBS
RRCH7SN02-G10.g3	AK119892 Os06g0557100	6,00E-78	proteína contendo domínio kinase
RRCH7SN00-B04.g0	Os01g0332200 Os01g0332200	2,00E-35	desconhecida
RRCH7SN02-G11.g1	AK067055 Os02g0644600	0.0	proteína contendo domínio preniltransferase
RRCH7SN02-H11.g1	AK110754 Os02g0546700	3,00E-08	hipotética
RRCH7SN02-H12.g3	AK102517 Os01g0243700	1,00E-166	Beta1, tipo glucanase 3
RRCH7SN03-B08.b1	AK066885 Os04g0517500	3,00E-40	Fosfoenolpiruvato carboxilase kinase
RRCH7SN05-A04.g2	Os06g0243300 Os06g0243300	6,00E-48	desconhecida
RRCH7SN05-A06.g2	AK073929 Os04g0687100	4,00E-37	Solúvel inorgânico pirofosfatase
RRCH7SN05-A07.g1	AK064594 Os01g0355400	4,00E-31	hipotética conservada
RRCH7SN00-B08.g0	Os12g0289400 Os12g0289400	1,00E-129	desconhecida
RRCH7SN05-A07.g2	AK120192 Os10g0427100	3,00E-19	hipotética conservada
RRCH7SN05-B08.g1	AK109799 Os11g0435400	3,00E-68	hipotética
RRCH7SN05-C05.g2	AK121903	5,00E-52	desconhecida
RRCH7SN05-C06.g2	AK073793 Os06g0203800	3,00E-68	kinase 1 tipo ERECTA
RRCH7SN05-C11.g1	AK108972	1,00E-175	desconhecida
RRCH7SN05-D02.g2	AK105552 Os01g0311600	7,00E-11	proteína da família Sulfotransferase
RRCH7SN05-E02.g1	AK100647 Os04g0450100	2,00E-47	proteína de função desconhecida da família DUF341
RRCH7SN05-E06.g2	Os09g0457100 Os09g0457100	3,00E-56	proteína da família Citochromo P450
RRCH7SN05-E07.g2	AK062356 Os06g0318800	4,00E-99	hipotética
RRCH7SN00-C02.g0	AK102090 Os02g0216300	3,00E-12	hipotética conservada
RRCH7SN05-F04.g2	AK060895 Os01g0754500	0.0	proteína de função desconhecida da família DUF1421
RRCH7SN00-C04.g0	AK106259 Os10g0551100	3,00E-91	hipotética conservada
RRCH7SN05-G08.g1	AK106155 Os04g0418000	1,00E-166	hipotética conservada
RRCH7SN05-H02.g2	AK112099 Os04g0569100	1,00E-135	Proteína contendo motivo hélice-volta-hélice, semelhante ao repressor lambda
RRCH7SN00-C05.g0	AK119907 Os01g0738900	4,00E-12	hipotética
RRCH7SN05-H07.g2	Os07g0525000 Os07g0525000	1,00E-55	Peptidase C48, da família SUMO/Sentrin/Ubl1
RRCH7SN05-H08.g2	Os10g0165400 Os10g0165400	1,00E-06	hipotética conservada

RRCH7SN05-H12.g2	AK103010 Os03g0776000	1,00E-37	Glicose-6-fosfate isomerase
RRCH7SN11-A05.g1	Os07g0138200 Os07g0138200	3,00E-29	desconhecida
RRCH7SN11-A06.g1	Os04g0242200 Os04g0242200	1,00E-114	desconhecida
RRCH7SN11-A07.g1	AK061241 Os09g0449700	1,00E-177	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN11-A09.g1	AK106587 Os03g0643100	3,00E-74	hipotética conservada
RRCH7SN11-A11.g1	AK070780 Os09g0516600	0.0	Glioxalase II
RRCH7SN11-B01.g1	AK068557 Os09g0419100	2,00E-17	hipotética
RRCH7SN11-B02.g1	Os10g0211200 Os10g0211200	0.0	desconhecida
RRCH7SN11-B05.g1	Os01g0522400 Os01g0522400	2,00E-11	proteína contendo domínio duplo repetido ARM
RRCH7SN00-C09.g0	AK103318 Os12g0210000	4,00E-36	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN11-B06.g1	Os09g0328700 Os09g0328700	8,00E-29	hipotética conservada
RRCH7SN11-B08.g1	AK059098 Os10g0539500	1,00E-164	Histona H4
RRCH7SN11-B09.g1	AK073891 Os05g0591500	0.0	hipotética conservada
RRCH7SN11-B10.g1	AK065107 Os01g0603000	2,00E-75	proteína da família de ligação Hialuronan/mRNA
RRCH7SN11-C01.g1	AK059215 Os01g0590700	1,00E-131	Proteína de inativação de ribossomo
RRCH7SN11-C02.g1	AK100614 Os11g0433200	1,00E-06	Peptidase A22B
RRCH7SN11-C04.g1	AK099802 Os03g0365200	1,00E-80	Peptidase C48, da família SUMO/Sentrin/Ubl1
RRCH7SN11-C05.g1	AK105283 Os03g0221300	0.0	hipotética conservada
RRCH7SN11-D02.g1	AK068475 Os12g0166500	1,00E-74	hipotética
RRCH7SN11-D03.g1	AK106600 Os09g0111300	1,00E-64	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN11-D05.g1	Os09g0506200 Os09g0506200	5,00E-18	desconhecida
RRCH7SN11-D06.g1	AK107916 Os08g0399700	2,00E-42	hipotética
RRCH7SN00-C12.g0	AK121635 Os07g0150200	8,00E-31	ribossomal 40S S12
RRCH7SN11-D08.g1	Os11g0432100 Os11g0432100	2,00E-14	desconhecida
RRCH7SN11-D11.g1	AY452768 Os12g0626400	1,00E-99	proteína da família Squaleno/fitoeno sintase
RRCH7SN11-E01.g1	Os06g0246900 Os06g0246900	1,00E-61	desconhecida
RRCH7SN11-E06.g1	Os06g0269100 Os06g0269100	2,00E-14	desconhecida
RRCH7SN11-E10.g1	AK110657	3,00E-89	desconhecida
RRCH7SN11-E11.g1	AK062152 Os01g0717000	0.0	GmCK1p
RRCH7SN11-E12.g1	AK069351 Os02g0435900	1,00E-108	hipotética conservada
RRCH7SN11-F03.g1	AK105034 Os01g0348900	1,00E-153	proteína contendo domínio lectina relacionada a Jacalin
RRCH7SN11-F04.g1	Os04g0271500 Os04g0271500	3,00E-28	desconhecida
RRCH7SN11-F05.g1	AK067239 Os01g0137800	4,00E-21	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN11-F06.g1	AK066198 Os07g0475900	6,00E-95	proteína contendo domínio ACT de ligação a aminoácido
RRCH7SN11-F08.g1	AK103228 Os11g0550300	2,00E-17	hipotética conservada
RRCH7SN11-F09.g1	AK102885 Os01g0531200	1,00E-119	Ubiquitina ligase SINAT4
RRCH7SN11-F10.g1	Os05g0279700 Os05g0279700	2,00E-60	desconhecida
RRCH7SN11-G01.g1	AK100944 Os01g0211800	7,00E-17	proteína contendo domínio de fator de transcrição zipper Basic-leucina (bZIP)
RRCH7SN11-G03.g1	AK111545 Os05g0591900	0.0	proteína de resposta hipersensitiva induzida
RRCH7SN11-G05.g1	AK073388 Os01g0813500	1,00E-33	Vacuolar classificação 36
RRCH7SN11-G08.g1	AK065369 Os11g0547000	2,00E-57	FKF1 (Adagio 3)

RRCH7SN11-G09.g1	AK120787 Os08g0387200	1,00E-128	proteína de função desconhecida contendo domínio DUF81
RRCH7SN11-G11.g1	AK106928 Os02g0437300	6,00E-89	hipotética
RRCH7SN11-H03.g1	AK074000 Os07g0445600	3,00E-44	hipotética conservada
RRCH7SN11-H04.g1	AK102995 Os07g0184900	1,00E-176	DNA helicase dependente de ATP
RRCH7SN11-H05.g1	AK107917 Os01g0274300	2,00E-17	hipotética
RRCH7SN13-A06.g1	AK067461 Os07g0492100	2,00E-22	proteína da família UvrD/REP helicase
RRCH7SN13-A08.g1	AK065811 Os05g0267100	6,00E-16	Cellular de ligação a retinaldehide, contendo domínio C-terminal
RRCH7SN13-A10.g1	AK103076 Os04g0378200	2,00E-36	Proteína contendo motivo alfa estéril (SAM)
RRCH7SN13-B01.g1	AK103944 Os05g0156300	1,00E-150	Protein disulfide-isomerase precursor
RRCH7SN13-B02.g1	AK068820 Os11g0220300	7,00E-73	hipotética
RRCH7SN13-B10.g1	AK066902 Os11g0460800	3,00E-29	Peptidase S10, da família serine carboxipeptidase
RRCH7SN13-C03.g1	Os11g0534600 Os11g0534600	2,00E-07	desconhecida
RRCH7SN13-C07.g1	AK061826 Os01g0358400	0.0	ribossomal 40S S4
RRCH7SN13-C08.g1	AK101718 Os03g0746800	2,00E-29	Peptidase S9A, prolyl oligopeptidase, contendo domínio N-terminal beta-propeller
RRCH7SN00-D08.g0	Os04g0111700 Os04g0111700	1,00E-23	desconhecida
RRCH7SN13-C10.g1	AK122138 Os02g0221900	4,00E-77	proteína da família Citochromo P450
RRCH7SN13-D01.g1	AK072582 Os11g0194600	8,00E-17	hipotética
RRCH7SN13-D10.g1	AK069124 Os07g0526700	1,00E-84	Terpenol ciclase/prenil transferase
RRCH7SN13-F10.g1	AK100434 Os04g0672700	3,00E-08	proteína contendo domínio preniltransferase
RRCH7SN13-F11.g1	Os11g0246600 Os11g0246600	2,00E-71	Peptidase S1C, da família hrtA/degQ
RRCH7SN13-F12.g1	Os10g0399000 Os10g0399000	5,00E-08	desconhecida
RRCH7SN13-G03.g1	AK102017 Os01g0727900	8,00E-11	hipotética conservada
RRCH7SN13-G05.g1	AK068786 Os01g0866300	1,00E-111	Proteína tipo VAMP (Proteína de membrana associada a vesículas)
RRCH7SN13-G09.g1	AK066224 Os10g0420900	1,00E-11	hipotética
RRCH7SN13-H02.g1	AK108461 Os07g0623100	1,00E-125	ER66 (Fragmento)
RRCH7SN13-H03.g1	AK067979 Os06g0135000	2,00E-21	hipotética conservada
RRCH7SN16-A01.g1	AK064780 Os09g0482100	7,00E-41	82 de choque térmico
RRCH7SN16-A08.g1	AK101053 Os05g0452500	3,00E-08	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN16-A10.g1	AK065527 Os11g0235300	1,00E-162	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN16-B08.g1	Os06g0555300 Os06g0555300	1,00E-19	desconhecida
RRCH7SN16-B10.g1	AK111629 Os02g0730300	2,00E-77	transportador 8 de potássio (AtPOT8) (AtHAK8)
RRCH7SN16-B11.g1	AK066050 Os02g0648000	8,00E-18	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN16-C02.g1	AK064893 Os10g0189100	3,00E-54	proteína da família fofoglicomutase/fofomannomutase
RRCH7SN16-C05.g1	AK110485	5,00E-32	desconhecida
RRCH7SN16-C06.g1	AK099539 Os02g0601800	1E-154	proteína contendo domínio Bromo
RRCH7SN16-C07.g1	AK110009	4,00E-15	desconhecida
RRCH7SN16-D05.g1	Os11g0228900 Os11g0228900	3,00E-23	desconhecida
RRCH7SN16-D12.g1	AK107927 Os02g0438500	3,00E-10	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN01-A09.g1	AK105855 Os06g0531200	4,00E-94	proteína da família DJ-1

RRCH7SN16-E02.g1	AK106395 Os11g0136800	1,00E-58	proteína de especificidade dupla phosphatase 19
RRCH7SN16-E06.g1	AK064860 Os03g0839300	5,00E-36	hipotética conservada
RRCH7SN16-F02.g1	AK068012 Os04g0346100	2,00E-75	40S ribossomal S27."
RRCH7SN01-B01.g2			desconhecida
RRCH7SN16-G02.g1	AK067239 Os01g0137800	5,00E-36	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN16-G06.g1	AK065192 Os08g0229600	5,00E-78	hipotética
RRCH7SN16-G07.g1	AK065611 Os06g0701300	2,00E-49	proteína contendo domínio ABC-1
RRCH7SN16-G09.g1	Os06g0321500 Os06g0321500	2,00E-37	desconhecida
RRCH7SN16-G10.g1	Os05g0218800 Os05g0218800	2,00E-27	desconhecida
RRCH7SN16-H04.g1	AK119907 Os01g0738900	3,00E-15	hipotética
RRCH7SN16-H11.g1	AK059890 Os03g0778100	1,00E-150	subunidade III do centro de reação do fotosistema I
RRCH7SN17-A05.g2	Os03g0697300 Os03g0697300	1,00E-26	desconhecida
RRCH7SN17-A11.g1	Os09g0290500 Os09g0290500	5,00E-11	desconhecida
RRCH7SN17-C11.g2	Os11g0310900 Os11g0310900	1,00E-102	desconhecida
RRCH7SN17-D10.g1	AK068914 Os02g0187400	1,00E-35	proteína da família Glicerofosforil diester fosfodiesterase
RRCH7SN01-B12.g1	Os01g0288900 Os01g0288900	3,00E-16	desconhecida
RRCH7SN17-E01.g2	AK108482 Os02g0747500	2,00E-40	AtHVA22e (tipo induzida por Ácido abscísico)
RRCH7SN17-E10.g2	AK120701 Os12g0575300	9,00E-85	hipotética
RRCH7SN17-F10.g2	AB111810 Os09g0482100	1,00E-109	82 de choque térmico
RRCH7SN17-G02.g2	AK120884 Os06g0319700	6,00E-08	60S ribossomal L31
RRCH7SN17-G04.g2	AK069976 Os02g0593600	7,00E-13	rica em repetições de Leucina
RRCH7SN23-A02.g1	AK109584 Os04g0397800	5,00E-11	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN23-A11.g1	Os05g0370400 Os05g0370400	1,00E-83	desconhecida
RRCH7SN23-B02.g1	AK120160 Os07g0569100	1,00E-106	Remorin, contendo domínio C-terminal
RRCH7SN23-D02.g1	Os10g0450700 Os10g0450700	2,00E-12	desconhecida
RRCH7SN23-D09.g1	Os09g0475300 Os09g0475300	1,00E-111	desconhecida
RRCH7SN01-C03.g2	AK106630 Os07g0629500	8,00E-52	Transcrito não codificador de proteína, não caracterizado.
RRCH7SN23-E11.g1	AK072896 Os06g0137500	1,00E-139	proteína contendo domínio Brix
RRCH7SN23-F01.g1	AK063438 Os02g0123100	1,00E-51	proteína 2 de controle da divisão celular
RRCH7SN23-G06.g1	AK073072 Os09g0295200	1,00E-08	hipotética
RRCH7SN23-G08.g1	AK071296 Os05g0122900	1,00E-74	BLE1
RRCH7SN23-G10.g1	Os02g0122100 Os02g0122100	9,00E-37	desconhecida
RRCH7SN23-G11.g1	AK071032 Os11g0206000	1,00E-10	hipotética
RRCH7SN23-H03.g1	AK111553 Os02g0672200	6,00E-29	Argonauta e Dicer protein, contendo domínio PAZ
RRCH7SN01-C04.g2	Os02g0694600 Os02g0694600	3,00E-49	Proteína rica em cisteínas, contendo domínio dedos de Zn.
RRCH7SN23-H09.g1	AK065967 Os12g0255200	3,00E-34	Ca(2+)- nuclease dependente
RRCH7SN23-H11.g1	Os12g0208600 Os12g0208600	7,00E-38	desconhecida
RRCH7SN01-C12.g2	AK058365 Os08g0127600	8,00E-08	DnaJ de choque térmico contendo domínio N-terminal
RRCH7SN01-D07.g2	AK066224 Os10g0420900	2,00E-42	hipotética
RRCH7SN01-D11.g2	Os05g0118300 Os05g0118300	1,00E-106	desconhecida
RRCH7SN01-E08.g2	AB007193 Os01g0866400	1,00E-102	Fructose-1,6-bisfosfatase
RRCH7SN01-F04.g2	Os04g0227600 Os04g0227600	4,00E-38	desconhecida



RRCH7SN01-F07.g2	AK070228 Os03g0191200	1,00E-26	proteína contendo domínio WW/Rsp5/WWP
RRCH7SN01-F08.g2	AY177699 Os01g0726400	1,00E-35	fator de transcrição, contendo domínio MADS-box
RRCH7SN01-G06.g2	Os10g0557700 Os10g0557700	1,00E-44	hipotética conservada
RRCH7SN01-G11.g2	AK071004 Os06g0492000	8,00E-35	proteína da família CDP-álcool fosfatidiltransferase
RRCH7SN01-H08.g2	AK068039 Os04g0489600	2,00E-89	Proteína contendo domínio de dimerização Hélice-laço-hélice
RRCH7SN02-A01.g3	AK102700 Os03g0441500	5,00E-96	proteína contendo domínio relacionado ao transportador ABC
RRCH7SN02-A02.g2	AK068810 Os09g0426100	0.0	proteína de função desconhecida contendo domínio DUF6
RRCH7SN02-B01.g3	AY626826	1,00E-128	desconhecida
RRCH7SN02-B05.g2	AK121894 Os12g0510000	6,00E-69	hipotética
RRCH7SN02-B06.g2	AK109438 Os05g0562200	1,00E-43	proteína da família 19 induzida pela seca
RRCH7SN02-B09.g2	Os07g0583900 Os07g0583900	1,00E-54	desconhecida
RRCH7SN02-C07.g2	AK069978 Os02g0529500	1,00E-64	proteína C tipo SNAP25 (Fragmento)
RRCH7SN02-D01.g2	AK119685 Os12g0111600	1,00E-133	Proteína induzida por hipóxia
RRCH7SN02-D02.g2	Os08g0339500 Os08g0339500	2,00E-48	desconhecida
RRCH7SN02-D06.g3	AK071906 Os07g0474700	2,00E-23	hipotética
RRCH7SN02-D07.g3	AK119884	5,00E-45	desconhecida
RRCH7SN02-D08.g2	AK120281 Os03g0348800	6,00E-47	hipotética

\*Os resultados mostrados nesta tabela são para os alinhamentos com *e-value* menor que  $10^{-06}$

TABELA 6 – RESULTADO DA PESQUISA BLAST PARA OS “CONTIGS”

mRNA	Erro/10kb	e-value	Número de Acesso RAP-DB	Provável proteína	Número de clones
Contig2	26,38	2,00E-19	<a href="#">Os09g0244100 Os09g0244100</a>	proteína da família En/Spm como transposon	2
Contig3	0,01	2,00E-11	<a href="#">Os01g0326900 Os01g0326900</a>	desconhecida	2
Contig4	8,6	2,00E-11	<a href="#">Os11g0681800 Os11g0681800</a>	desconhecida	2
Contig8	66,53	1,00E-39	<a href="#">AK108357 Os02g0216400</a>	hipotética	2
Contig9	6,27	2,00E-58	<a href="#">Os01g0816800 Os01g0816800</a>	desconhecida	2
Contig10	72,13	8,00E-34	<a href="#">AK064150 Os05g0283000</a>	hipotética conservada	2
Contig11	36,51	0,00E+00	<a href="#">AK102263 Os03g0648400</a>	DnaJ homóloga a ANJ1.	8
Contig13	10,33	5,00E-39	<a href="#">AK067239 Os01g0137800</a>	Transcrito não codificador de proteína, não caracterizado.	2
Contig14	15,2	1,00E-168	<a href="#">AK073685 Os02g0592300</a>	desconhecida	2
Contig16	203	1,00E-126	<a href="#">AK121120 Os10g0527100</a>	Chloroplast ATP synthase a chain precursor	4
Contig17	165,8	0,00E+00	<a href="#">Os07g0104300 Os07g0104300</a>	desconhecida	3
Contig18	32,2	0,00E+00	<a href="#">AK108972 </a>	proteína da família peptidase M24	6
Contig20	132,45	1,00E-11	<a href="#">AK110649 Os01g0797700</a>	Transcrito não codificador de proteína, não caracterizado.	2
Contig21	88,31	7,00E-11	<a href="#">Os07g0178200 Os07g0178200</a>	desconhecida	2
Contig26	173,07	9,00E-40	<a href="#">Os12g0468100 Os12g0468100</a>	desconhecida	2
Contig28	23,55	0,00E+00	<a href="#">AK059208 </a>	desconhecida	3

\*Os resultados mostrados nesta tabela são para os alinhamentos com *e-value* menor que  $10^{-06}$

## 6 – CONCLUSÃO

Nossa estratégia de trabalho permitiu a obtenção de quatro bibliotecas de cDNA representativas de genes expressos de raízes de arroz inoculadas e não inoculadas com *H. seropedicae* e na presença ou ausência de fonte de nitrogênio. As bibliotecas obtidas constituem uma importante fonte de material para futuros projetos de análise de interação planta-bactéria. A biblioteca de raízes inoculadas com *H. seropedicae* por 7 dias foi parcialmente seqüenciada e permitiu a obtenção de uma coleção de ESTs que podem conter genes diferencialmente expressos. As 482 ESTs seqüenciadas foram analisadas quanto a sua provável função e revelou genes que podem estar envolvidos na interação planta-bactéria. Uma caracterização aprofundada e análises funcionais desses genes devem permitir a confirmação de genes relacionados a tal processo e contribuir para o entendimento da interação de *H. seropedicae* com arroz.

## 7 – REFERÊNCIAS BIBLIOGRÁFICAS

ALTSCHUL, S.F.; MADDEN, T.L.; SCHÄFFER, A.A.; et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, v.25, n.17, p.3389-3402, 1997.

ARANGARASAN, V.; PALANIAPPAN, S.P.; CHELLIAH, S., Inoculation effects of diazotrops and phosphobacteria on rice. **Indian Journal of Microbiology**, v. 38, p. 111–112, 1998.

BALDANI, J.I.; BALDANI, V.L.D.; SELDIN, L.; et al. Characterization of *Herbaspirillum seropedicae* gen. nov., sp. nov., a root-associated nitrogen-fixing bacterium. **International Journal Systematic Bacteriology**, v. 36, p. 86–93, 1986.

BALDANI, V.L.D. **Efeito da inoculação de *Herbaspirillum* spp. no processo de colonização e infecção de plantas de arroz e ocorrência e caracterização parcial de uma nova bactéria diazotrófica seropédica.** Itaguaí, RJ, 1996. 238 f. Tese (Doutorado em Agronomia) – Instituto de Agronomia, Universidade Federal Rural do Rio de Janeiro.

BALDANI, V.L.D.; BALDANI, J.I.; DÖBEREINER, J. Inoculation of rice plants with the endophytic diazotrophs *Herbaspirillum seropedicae* and *Burkholderia* spp. **Biology and Fertility of Soils**, v. 30, p. 485–491, 2000.

BARRAQUIO, W.L.; REVILLA, L.; LADHA, J.K. Isolation of endophytic bacteria from wetland rice. **Plant and Soil**, v. 194, p. 15-24, 1997.

BARRY, G. F. The Use of the Monsanto Draft Rice Genome Sequence in Research. **Plant Physiology**, v. 125, p. 1164–1165, 2001.

BODDEY, R.M.; OLIVEIRA, O.C. de; URQUIAGA, S.; et al. Biological nitrogen fixation associated with sugar cane and rice: contributions and prospects for improvement. **Plant and Soil**, v. 174, p. 195-209, 1995.

CANUTO, E.L.; OLIVEIRA, A.L.M.; REIS, V.M.; et al. Evaluation of the biological nitrogen fixation contribution in sugarcane plants originated from seeds and inoculated with nitrogen-fixing endophytes. **Brazilian Journal of Microbiology**, v. 34, p. 62-64, 2003.

CÁRDENAS, L.; ALEMÁN, E.; NAVA, N. et al. Early responses to Nod factors and mycorrhizal colonization in a non-nodulating *Phaseolus vulgaris* mutant. **Planta**, v. 223, p. 746–754, 2006.

CAREL, N.; BERNARDI, G. Two classes of genes in plants. **Genetics**, v. 154, p. 1819-1825, 2000.

CAVALCANTE, J. J. V.; VARGAS, C.; NOGUEIRA, E. M.; et al. Members of the ethylene signalling pathway are regulated in sugarcane during the association with nitrogen-fixing endophytic bacteria. **Journal of Experimental Botany**, v. 58, n. 3, p. 673–686, 2007.

CRUZ, L.M.; SOUZA, E.M.; WEBER, O.B. et al. 16S ribosomal DNA characterization of nitrogen-fixing bacteria isolated from banana (*Musa* spp.) and pineapple (*Ananas comosus* (L.) Merril). **Applied and Environmental Microbiology**, v. 67, n.5, p. 2375-2379, 2001.

DONG, Z.; MCCULLY, M.E.; CANNY, M.J. Does *Gluconacetobacter diazotrophicus* live and move in the xylem of sugarcane stems? Anatomical and physiological data. **Annals of Botany**, v. 80, p. 147-158, 1997.

ELBELTAGY A.; NISHIOKA K.; SATO T.; et al. Endophytic Colonization and In Planta Nitrogen Fixation by a *Herbaspirillum* sp. Isolated from Wild Rice Species. **Applied and Environmental Microbiology**, v. 67, p. 5285-5293, 2001.

ENGELHARD, M.; HUREK, T.; REINHOLD-HUREK, B. Preferential occurrence of diazotrophic endophyte, *Azoarcus* sp. in wild rice species and land races of *Oryza sativa* in comparison with modern races. **Environmental Microbiology**, v. 2, p. 131-141, 2000.

EWING, B.; GREEN, P. Base-Calling of Automated Sequencer Traces Using *Phred*. II. Error Probabilities. **Genome Research**, v. 8, n. 3, p. 186-194, 1998.

EWING, B.; HILLIER, L.; WENDL, M.C. et al. Base-calling of automated sequencer traces using *Phred*. I. Accuracy Assessment. **Genome Research**, v. 8, n. 3, p. 175-185, 1998.

FISCHER, K.S. Frontier project on nitrogen fixation in rice: looking ahead. In: The Quest for Nitrogen Fixation in rice. Eds. LADHA, J.K. & REDDY, pag. 25-31

Proceedings of a workshop, 9-12 August 1999, Los Baños Laguna, Philippines IRRI (International Rice Research Institute). ISBN 971-22-0112-0

GOFF, S. A.; RICKE, D.; LAN, T.; et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. japonica). **Science**, v. 296, p. 92-100, 2002.

GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: a graphical tool for sequence finishing. **Genome Research**, v. 8, p. 195-202, 1998.

GYANESHWAR, P.; JAMES, E.K.; MATHAN, N.; et al. Endophytic Colonization of Rice by a Diazotrophic Strain of *Serratia marcescens*. **Journal of Bacteriology**, v. 183, p. 2634-2645, 2001.

GYANESHWAR, P.; JAMES, E.K.; REDDY P.M.; et al. *Herbaspirillum* colonization increases growth and nitrogen accumulation in aluminium - tolerant rice varieties. **New Phytologist**, v. 154, p. 131-145, 2002.

HOUDE, M.; BELCAID, M.; OUELLET, F. et al. Wheat EST resource for functional genomics of abiotic stress. **BMC Genomics**, v.7, p.149, 2006.

HUANG, X.; MADAN, A. CAP3: A DNA Sequence Assembly Program. **Genome Research**, v. 9, p. 868-877, 1999.

IRGSP. International Rice Genome Sequencing project  
([http://rgp.dna.affrc.go.jp/IRGSP/nature436\\_793-800/nature05.html](http://rgp.dna.affrc.go.jp/IRGSP/nature436_793-800/nature05.html)) Acesso em 19/08/2005)

IRRI. **Rice Today**, v. 5, n. 3, Metro Manila, Filipinas, 2006.

JAMES, E.K.; GYANESHWAR, P.; MATHAN, N. et al. Infection and Colonization of Rice Seedlings by the Plant Growth-Promotion Bacterium *Herbaspirillum seropedicae* Z67. **Molecular Plant-Microbe Interaction**, v. 15, n. 9, p. 894-906, 2002.

JANTASURIYARAT, C.; GOWDA, M.; HALLER, K.; et al. Large-Scale Identification of Expressed Sequence Tags Involved in Rice and Rice Blast Fungus Interaction. **Plant Physiology**, v. 138, p. 105-115, 2005.

KENNEDY, I. R.; CHOUDHURY, A.T.M.A.; KECSKÉS, M. L. Non-symbiotic bacterial diazotrophs in crop-farming systems: can their potential for plant growth promotion be better exploited?. **Soil Biology & Biochemistry**, v. 36, p. 1229–1244, 2004.

LADHA, J. K.; KIRK, G. J. D.; BENNETT J.; et al. Opportunities for increased nitrogen use efficiency from improved lowland rice germplasm. *Field Crops Research*, v. 56, p. 41–71, 1998.

LADHA, J.K.; REDDY, P.M. Extension of nitrogen fixation to rice – necessity and possibilities. **GeoJournal**, v. 35, p. 363-372, 1995.

LADHA, J. K.; REDDY, P. M. Nitrogen fixation in rice systems: state of knowledge and future prospects. *Plant and Soil*, v. 252, p. 151–167, 2003.

LAMBAIS, M.R. *In silico* differential display of defense-related expressed sequence tags from sugarcane tissues infected with diazotrophic endophytes. **Genetics and Molecular Biology**, v. 24, p. 103-111, 2001.

MACDOWELL, J.M.; DANGL, J.L. Signal transduction in the plant immune response. **Trends in Biochemical Science**, v. 25, n. 2, p. 79-82, 2000.

MACHADO, H.B.; FUNAYAMA, S.; RIGO, L.U.; et al. Excretion of ammonium by *Azospirillum brasilense* mutants resistant to ethylenediamine. **Canadian Journal of Microbiology**, v. 37, p. 549-553, 1991.

MALARVIZHI, P.; LADHA, J.K. Influence of available nitrogen and rice genotype on associative nitrogen fixation. **Soil Science Society of America Journal**, v. 63 p. 93-99, 1999.

MATSUMOTO, T.; WU, J.; KANAMORI, H. et al (International Rice Genome Sequencing Project). The map-based sequence of the rice genome. **Nature**, v. 436, p. 793-800, 2005.

MEHNAZ, S.; MIRZA, M.S.; HAURAT, J.; et al. Isolation and 16S rRNA sequence analysis of the beneficial bacteria from the rhizosphere of rice. **Canadian Journal of Microbiology**, v. 47, p. 110-117, 2001.

MICHALEK, W.; WESCHKE, W.; PLEISSNER, K.P.; et al. EST analysis in barley defines a unigene set comprising 4,000 genes. **Theoretical Applied Genetic**, v. 104, n. 1, p. 97–103, 2002.

MILLER, J. H. **Experiments in molecular genetics**. New York. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y, 1990.

MUTHUKUMARASAMY R.; CLEENWERCK I.; REVATHI G.; et al. Natural association of *Gluconacetobacter diazotrophicus* and diazotrophic *Acetobacter peroxydans* with wetland rice. **Systematic and Applied Microbiology**, v. 28, p. 277–286, 2005.

MUTHUKUMARASAMY, R.; REVATHI, G.; LAKSHMINARASIMHAN, C. Diazotrophic associations in sugar cane cultivation in South India. **Tropical Agriculture**, v. 76, p. 171–178, 1999.

NADIMPALLI, R; YALPANI, N.; JOHAL, G.S. et al. Prohibitins, Stomatins, and Plant Disease Response Genes Compose a Protein Superfamily That Controls Cell Proliferation, Ion Channel Regulation, and Death. **The Journal of Biological Chemistry**, v. 275, n. 38, p. 29579–29586, 2000.

NIEUWENHOVE, C. V.; MERCKX, R.; HOLM, L. V.; et al. Dinitrogen fixation activity of *Azorhizobium caulinodans* in the rice (*Oryza sativa* L.) rhizosphere assessed by nitrogen balance and nitrogen-15 dilution methods. **Biology and Fertility of Soils**, v. 33, p.25 –32. 2001.

NOGUEIRA, E.M.; VINAGRE, F.; MASUDA, H.P.; et al. Expression of sugarcane genes induced by inoculation with *Gluconacetobacter diazotrophicus* and *Herbaspirillum rubrisubalbicans*. **Genetics and Molecular Biology**, v. 24, p. 199-206, 2001.

OGIHARA, Y.; MOCHIDA, K.; NEMOTO, Y.; et al. Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. **Plant Journal**, v. 33, p. 1001–1011, 2003.

OLIVARES, F.L.; JAMES, E.K.; BALDANI, J.I.; et al. Infection of mottled stripe disease susceptible and resistant varieties of sugarcane by the endophytic diazotroph *Herbaspirillum*. **New Phytologist**, v. 135, p. 723-737, 1997.

OLIVEIRA, A.L.M.; URQUIAGA, S.; DÖBEREINER J.; et al. The effect of inoculating endophytic N<sub>2</sub>-fixing bacteria on micropropagated sugarcane plants. **Plant and Soil**, v. 242, p. 205–215, 2002.

PEREIRA, J.A.R.; CAVALCANTE, V.A.; BALDANI, J.I.; et al. Field inoculation of sorghum and rice with *Azospirillum* spp. and *Herbaspirillum seropedicae*. **Plant and Soil**, v.110, p. 269–274, 1988.

RAUYAREE, P.; CHOI, W.; FANG, E.; et al. Genes expressed during early stages of rice infection with the rice blot fungus *Magnaporthe grisea*. **Molecular Plant Pathology**, v. 2, n. 6, p. 347-354, 2001.

REDDY, A.R.; RAMAKRISHNA, W.; SEKHAR, A.C. et al. Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22). **Genome**, v. 45, p. 204-211, 2002.

REINHOLD-HUREK, B.; HUREK, A. T. Interactions of gramineous plants with *Azoarcus* spp. and other diazotrophs: identification, localization and perspectives to study their function. **Critical Reviews Plant Science**, v. 17, p. 29-54, 1998a.

REINHOLD-HUREK, B.; HUREK, A. T. Life in grasses: diazotrophic endophytes. **Trends in Microbiology**, v. 6, p. 139-144, 1998b.

REIS, V.M., BALDANI, J.I., BALDANI, V.L.D., DOÖBEREINER, J. Biological dinitrogen fixation in gramineae and palm trees. **Critical Reviews in Plant Sciences** 19, 227–247. 2000.

ROGER, P.A.; LADHA, J.K. Biological N<sub>2</sub> fixation in wetland rice fiels: estimations and contribution to nitrogen balance. **Plant and Soil**, v. 141, p. 41-55, 1992.

RONCATO-MACCARI, L. D.B.; RAMOS, H. J.O.; PEDROSA, F. O. Endophytic *Herbaspirillum seropedicae* expresses nif genes in gramineous plants. **FEMS Microbiology Ecology**, v. 45, p. 39-47, 2003.

RONCATO-MACCARI, L. D. B.; RAMOS, H. J. O.; PEDROSA F. O. et al. Root colonization, systemic spreading and contribution of *Herbaspirillum seropedicae* to growth of rice seedling. **Symbiosis**, v. 35, p. 261-270, 2003.



RONNING, C.M.; STEGALKINA, S.S.; ASCENZI, R.A.; et al. Comparative analyses of potato expressed sequence tag libraries. **Plant Physiology**, v. 131, p. 419–429, 2003.

RUDD, S. Expressed sequence tags: alternative or complement to whole genome sequences? **TRENDS in Plant Science**, vol. 8, n. 7, p. 321-329, 2003.

RUIZ-LOZANO, J.M.; ROUSSEL, H.; GIANINAZZI, S. et al. Defense genes are differentially induced by a mycorrhizal fungus and *Rhizobium sp* in wild-type and symbiosis-defective pea genotypes. **Molecular Plant-Microbe Interaction**, v. 12, p. 976-984, 1999.

SHRESTHA, R. K.; J. K. LADHA. Genotypic variation in promotion of rice nitrogen fixation as determined by <sup>15</sup>N dilution. **Soil Science Society America Journal**, v. 60, p. 1815-1821, 1996.

SALSE, J.; PIÉGU, B.; COOKE, R.; et al. New *in silico* insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome. **The Plant Journal**, v. 38, p. 396-409, 2004.

SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Molecular cloning; a laboratory manual**. 2. ed. New York: Cold Spring Harbor Laboratory Press, 1989.

SASAKI, T.; BURR, B. International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. **Current Opinion in Plant Biology**, v. 3, p. 138—141, 2000.

SOUZA, E. M.; PEDROSA, F. O.; MACHADO, H. B.; DRUMMOND, M.; YATES, M. G. The N-terminus of the NifA protein of *Herbaspirillum seropedicae* is probably involved in sensing of ammonia. **In I. A. Tikhonovich, N. A. Provorov, V. I. Romanov, and W. E. Newton (ed.), Nitrogen fixation: fundamentals and applications. Kluwer Academic Publishers, Dordrecht, The Netherlands.** p. 260. 1995.

STOLTZFUS, J.R.; SO, R.; MALARVITHI, P.P.; et al. Isolation of endophytic bacteria from rice and assessment of their potential for supplying rice with biologically fixed nitrogen. **Plant and Soil**, v. 194, p. 25–36, 1997.

UEDA, T.; SUGA, Y.; YAHIRO N.; et al. Genetic diversity of N<sub>2</sub>-fixing bacteria associated with rice roots by molecular evolutionary analysis of a nifD library. **Canadian Journal of Microbiology**, v. 41, p. 235-240, 1995a.

UEDA, T.; SUGA, Y.; YAHIRO, N.; et al. Remarkable N<sub>2</sub>-fixing bacterial diversity detected in rice roots by molecular evolutionary analysis of *nifH* gene sequences. **Journal of Bacteriology**, v. 177, p. 1414-1417, 1995b.

VENTELON-DEBOUT, M.; NGUYEN, T.T.H.; WISSOCQ, A.; et al. Analysis of the transcriptional response to Rice Yellow Mottle Virus infection in *Oryza sativa indica* and *japonica cultivars*. **Molecular Genetics and Genomics**, v. 270, p. 253-262, 2003.

YAMAMOTO, K.; SASAKI, T. Large-scale EST sequencing in rice. **Plant Molecular Biology**, v. 35, p. 135–144, 1997.

YU, J.; HU, S.; WANG, J.; et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. indica). **Science**, v. 296, p. 79-92, 2002.

YU, J.; WANG, J.; LIN, W.; et al. The Genomes of *Oryza sativa*: A History of Duplications. **Plos Biology**, v. 3, p. 266-281, 2005.

WU, J.; MAEHARA, T.; SHIMOKAWA, T.; et al. A Comprehensive Rice Transcript Map Containing 6591 Expressed Sequence Tag Sites. **The Plant Cell**, v. 14, p. 525–535, 2002.

WU, P.; ZHANG, G.; LADHA J. K.; et al. Molecular – marker- facilitated investigation on the ability to stimulate N<sub>2</sub> fixation in the rhizosphere by irrigated rice plants. **Theoretical Applied Genetic**, v. 91, p. 1171-1183, 1995.

## **ANEXOS**

ANEXO 1 – “*SCRIPT*” RENOMEIA.SH PARA RENOMEAR OS CROMATOGRAMAS  
GERADOS PELO MEGABACE

```

#!/bin/bash

#Renomeando as amostras com o nome do diretório fornecido

PROJNAME="$1"
BASEDIR="$2"

for LISTA in `ls $BASEDIR/seq/`
do
if [[ $LISTA == RR[CS]H[057][CS]N[0-9][0-9].[bg][0-9] ]]
then
echo Nome das pastas OK
PASTA=$(echo $LISTA | cut -d"." -f1)
DIRECTION=$(echo $LISTA | cut -d"." -f2)
else
echo Erro!Pasta \"$LISTA\" não está com nome padrão!!
continue
fi

for CHROMAT in `ls $BASEDIR/seq/$LISTA/`
do
POCO=$(echo $CHROMAT | cut -d"." -f1)
if cp $BASEDIR/seq/$LISTA/$CHROMAT
$BASEDIR/projetos/$PROJNAME/chromat_dir/$PASTA-$POCO.$DIRECTION
then
echo $PASTA-$POCO.$DIRECTION OK
else
echo Não foi possível alterar $BASEDIR/seq/
fi
done
done

exit 0

```

ANEXO 2 – “SCRIPT” SPLITFASTAFILE.SH PARA DIVIDIR O ARQUIVO FASTA  
DAS SEQUENCIAS EM ARQUIVOS SEPARADOS

```
#!/bin/bash

# Individualiza sequencias contidas em arquivos FASTA
#

FASTAFILE="$1"

# Individualizando as sequencias FASTA...
NUMCONTIG=$(echo | grep -c \> $FASTAFILE) # Conta as sequencias
NUM=$((($NUMCONTIG - 1)) # No. de sequencias menos 1 para uso no comando "csplit"
PREFIX=$(basename $FASTAFILE) # Pega o nome do arquivo FASTA
csplit -f $PREFIX $FASTAFILE "/>" {$NUM} # Coloca as sequencias em arquivos separados
rm -f ${PREFIX}00 # Remove o arquivo "00" vazio

exit 0
```

ANEXO 3 – “SCRIPT” WEBBLAST.SH PARA EXECUTAR TODOS OS PASSOS  
DESDE RENOMEAR CROMATOGRAS A EXECUÇÃO DO SOFTWARE CAP3



```

#!/bin/bash

# Criando diretório do projeto
BASEDIR="/home/liziane/ricedb"
PROJDIR="$BASEDIR/projetos"
PROJNAME="$1"

echo Criando diretórios para o projeto...
cd $PROJDIR
mkdir $PROJNAME
cd $PROJNAME
mkdir edit_dir phd_dir chromat_dir # Diretorios de montagem
mkdir contigs_dir singlets_dir # Diretorios de sequencia
mkdir blastn_dir tblastx_dir web_tblastx_dir # Diretorios para resultados BLAST

cp $BASEDIR/lib/vector.seq $PROJDIR/$PROJNAME #Copiando o arquivo vector.seq para a pasta
projeto

#Renomeando as amostras
bash $BASEDIR/bin/renomeia.sh $PROJNAME $BASEDIR

#Triming com PHRED
cd edit_dir
echo Rodando o "PHRED" com "trimming"...
phred -id ../chromat_dir -trim_alt ../chromat_dir -trim_phd -pd ../phd_dir &> /dev/null

#Rodando phredPhrap
echo Rodando "phredPhrap"...
phredPhrap > phredPhrap.screen 2> phredPhrap.error

#Rodando Cap
echo Rodando "CAP3"...
cap3 $PROJNAME.fasta.screen -o 30 -p 90 > cap3.screen 2> cap3.error
cd ../

#Primeiro os contigs são divididos
cd contigs_dir
echo Criando arquivos de contig em `pwd`
bash $BASEDIR/bin/splitfastafile.sh
$BASEDIR/projetos/$PROJNAME/edit_dir/$PROJNAME.fasta.screen.cap.contigs &> /dev/null
cd ../singlets_dir
echo Criando arquivos de singlets em `pwd`
bash $BASEDIR/bin/splitfastafile.sh
$BASEDIR/projetos/$PROJNAME/edit_dir/$PROJNAME.fasta.screen.cap.singlets &> /dev/null
cd ../
exit 0

```

ANEXO 4 – “*SCRIPT*” BLAST.SH PARA PESQUISA BLAST

```
#!/bin/bash
```

```
# Executa o BLAST para os contigs e singlets
```

```
BASEDIR="/home/liziane/ricedb"
```

```
PROJNAME="RRCH7SN"
```

```
BLASTDBP="$BASEDIR/lib/rapdb_rap1_all_orf_amino $BASEDIR/lib/rapdb_prediction_orf_amino  
$BASEDIR/lib/rapdb_rap1_rep_orf_amino"
```

```
BLASTDBN="$BASEDIR/lib/rapdb_rap1_all_orf_nuc $BASEDIR/lib/rapdb_rap1_all_transcripts  
$BASEDIR/lib/rapdb_prediction_orf_nuc $BASEDIR/lib/rapdb_rap1_rep_orf_nuc"
```

```
# Roda o BLAST
```

```
for ARQCONTIG in `ls $BASEDIR/projetos/$PROJNAME/contigs_dir`
```

```
do
```

```
blastall -p blastx -d "$BLASTDBP" -i $BASEDIR/projetos/$PROJNAME/contigs_dir/$ARQCONTIG  
-b5 -v5 -o blastx_dir/${ARQCONTIG}_blastx.out
```

```
blastall -p blastn -d "$BLASTDBN" -i $BASEDIR/projetos/$PROJNAME/contigs_dir/$ARQCONTIG  
-b5 -v5 -o blastn_dir/${ARQCONTIG}_blastn.out
```

```
done
```

```
for ARQCONTIG in `ls $BASEDIR/projetos/$PROJNAME/singlets_dir`
```

```
do
```

```
blastall -p blastx -d "$BLASTDBP" -i $BASEDIR/projetos/$PROJNAME/singlets_dir/$ARQCONTIG  
-b5 -v5 -o blastx_dir/${ARQCONTIG}_blastx.out
```

```
blastall -p blastn -d "$BLASTDBN" -i $BASEDIR/projetos/$PROJNAME/singlets_dir/$ARQCONTIG  
-b5 -v5 -o blastn_dir/${ARQCONTIG}_blastn.out
```

```
done
```

```
# Analisa os resultados do BLAST
```

```
for BLASTRESULT in `ls $BASEDIR/projetos/$PROJNAME/teste/blastx_dir`
```

```
do
```

```
echo ARQUIVO ANALISADO: $BLASTRESULT
```

```
bash blast_parser.sh $BASEDIR/projetos/$PROJNAME/teste/blastx_dir/$BLASTRESULT
```

```
bash blast_parser4gff.sh $BASEDIR/projetos/$PROJNAME/teste/blastx_dir/$BLASTRESULT
```

```
echo
```

```
echo "-----"
```

```
done
```

```
for BLASTRESULT in `ls $BASEDIR/projetos/$PROJNAME/teste/blastn_dir`
```

```
do
```

```
echo ARQUIVO ANALISADO: $BLASTRESULT
```

```
bash blast_parser.sh $BASEDIR/projetos/$PROJNAME/teste/blastn_dir/$BLASTRESULT
```

```
bash blast_parser4gff.sh $BASEDIR/projetos/$PROJNAME/teste/blastn_dir/$BLASTRESULT
```

```
echo
```

```
echo "-----"
```

```
done
```

```
exit 0
```

ANEXO 5 – “*SCRIPT*” BLAST\_PARSER4GFF.SH PARA CRIAR TABELA COM RESULTADOS DA PESQUISA BLAST COM INFORMAÇÕES DA ANOTAÇÃO

```

#!/bin/bash

# Executa o BLAST para os contigs e singlets
#

BLASTRESULT="$1"
GFFDIR="/home/liziane/ricedb/lib/rap-db/GFF_forDL"
GFFFILE=(all_loci.gff predicted_loci.gff predicted_transcript.gff)

CONTIGNAME=$(echo $BLASTRESULT | cut -d"." -f5)
echo
echo $CONTIGNAME

if grep "***** No hits found *****" $BLASTRESULT > /dev/null
then
echo Nenhum Hit encontrado para $BLASTRESULT
else
ORFID2=$(grep -m1 ^\> $BLASTRESULT | cut -d"|" -f2)
echo CONTIG: $CONTIGNAME
if [ $ORFID2 ]
then
echo ORF: $ORFID2

for ((i=0; i<${#GFFFILE[*]}; i++))
do
RESULT=$(grep $ORFID2 $GFFDIR/${GFFFILE[$i]})

if [ "$RESULT" ]
then
echo FOUND IN: ${GFFFILE[$i]}
echo $RESULT | tr ";" "\n"
else
echo NOT FOUND IN: ${GFFFILE[$i]}!
fi

done

else
echo ORF: Não encontrada!
fi
fi

exit 0

```

ANEXO 6 – "*SCRIPT*" BLAST\_PARSER.SH PARA CRIAR TABELA COM RESULTADOS DO BLAST COM INFORMAÇÕES SOBRE O ALINHAMENTO

```
#!/bin/bash

# Analisa um arquivo com resultado de busca com o BLAST
# Informações do primeiro hit são extraídas
#

BLASTRESULT="$1" # Arquivo de resultado BLAST

FLAG=

while read LINE
do

    if [ $FLAG ] && echo $LINE | grep ^Length > /dev/null
    then
        SUBJLENGTH=$(echo $LINE | cut -d"=" -f2)
        FLAG=
    elif [ $FLAG ]
    then
        DESC=$DESC$LINE
    elif echo $LINE | grep ^\> > /dev/null
    then
        DESC=$(echo $LINE | tr -d ">")
        FLAG=1
    elif echo $LINE | grep ^Query= > /dev/null
    then
        QUERY=$(echo $LINE | tr -s " " | cut -d" " -f2)
    elif echo $LINE | grep ^Length= > /dev/null
    then
        QUERYLENGTH=$(echo $LINE | cut -d"=" -f2)
    elif echo $LINE | grep \(. *letters\) > /dev/null
    then
        QUERYLENGTH=$(echo $LINE | cut -d"(" -f2 | cut -d" " -f1)
    elif echo $LINE | grep "Score =" > /dev/null
    then
        SCORE=$(echo $LINE | tr -s " " | cut -d" " -f3)
        EXPECT=$(echo $LINE | tr -s " " | cut -d" " -f8)
    elif echo $LINE | grep Identities > /dev/null
    then
        IDENT=$(echo $LINE | tr -s " " | cut -d" " -f3-4 | cut -d "," -f1)
        POSIT=$(echo $LINE | tr -s " " | cut -d" " -f7-8 | cut -d "," -f1)
        GAPS=$(echo $LINE | tr -s " " | cut -d" " -f11-12)
        break
    fi

done < "$BLASTRESULT"

# Imprime o resultado
echo QUERY: $QUERY
echo QUERYLENGTH: $QUERYLENGTH
echo DESCRIPTION: $DESC
echo SUBJECTLENGTH: $SUBJLENGTH
echo SCORE: $SCORE
echo E VALUE: $EXPECT
echo IDENTITY: $IDENT
echo POITION: $POSIT
echo GAPS: $GAPS

exit 0
```