

## RESSALVA

Alertamos para ausência de Capa, Folha de rosto e Ficha catalográfica, não incluídas pelo(a) autor(a) no arquivo original.

## **DADOS CURRICULARES DO AUTOR**

**RODRIGO MATHEUS PEREIRA** – Nascido em 24 de janeiro de 1975, no município de Ribeirão Preto, Estado de São Paulo. Ingressou no curso de Agronomia na Faculdade de Ciências Agrárias e Veterinárias da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Campus de Jaboticabal, em Janeiro de 1995. No mesmo campus, em agosto de 1996, fundou, em conjunto com colaboradores, a Consultoria Agropecuária Júnior (CAP<sub>JR</sub>), empresa júnior da qual participou por 4 anos, ocupando diferentes cargos executivos. Em janeiro de 2001 obteve o título de Engenheiro Agrônomo. Iniciou o curso de mestrado em Microbiologia em agosto de 2001. Obteve o título de Especialista em Bioinformática, pelo Laboratório Nacional de Computação Científica (LNCC), Petrópolis-RJ, em novembro de 2002. Obteve o título de Mestre em Microbiologia em setembro de 2003. Iniciou o curso de Doutorado em Microbiologia em março de 2004. Iniciou estágio em bioinformática, no laboratório de biotecnologia do centro nacional de pesquisa da soja – Embrapa, em abril de 2007.

*“O único homem que está isento de erros é aquele que não arrisca acertar.”*  
**Albert Einstein**

Aos meus pais,

**Norival e Vera**

que me ensinaram a importância do estudo,  
por darem amor, apoio e compreensão,  
pelo esforço e trabalho que me possibilitaram o estudo,  
pela educação, lições de vida e moral.

As minhas irmãs,

**Ana Carolina e Juliana,**

Pelo amor, amizade e apoio,

A minha esposa,

**Thaís,**

pelo amor, amizade e apoio,

A minha filha,

**Giovanna,**

que transformou completamente minha vida,  
possibilitando viver a grande e maravilhosa experiência de ser pai.

**DEDICO**

## **AGRADECIMENTOS**

À minha família, pelo amor, apoio e confiança em todos os momentos;

À minha orientadora prof. Dra, Eliana G. de Macedo Lemos, a quem muito admiro pela competência profissional e genialidade, agradeço pela orientação, pelos conselhos, pela paciência, pela amizade, pelo carinho quase maternal que surgiu ao longo de todos esses anos e pela oportunidade de crescimento pessoal e profissional.;

À Dra. Lúcia M. Carareto Alves agradeço pela paciência, pela ajuda na tese, pela amizade, pelo apoio em momentos decisivos da minha vida e pela alegria de viver que contagia a todos em sua volta;

Ao Dr. João Carlos Campanharo agradeço a amizade e pelos constantes auxílios prestados no laboratório;

Ao amigo Érico Leandro da Silveira, agradeço a grande amizade, aos muitos anos de convívio na república, a parceria em meus primeiros artigos científicos, as muitas aventuras vividas ao longo desses anos que jamais serão esquecidas;

Ao amigo Maurício Egídeo Cantão agradeço pela imensa ajuda na análise dos dados da tese, pela grande amizade no mundo real e virtual por tantos anos, pelos muitos ensinamentos de bioinformática, pelo companheirismo nos diversos “turnos noturnos”, pela paciência para explicar e sempre ajudar;

Ao amigo Luciano Takeshi Kishi, agradeço aos muitos ensinamentos de bioinformática, a amizade ao longo dos anos de convívio no laboratório, o companheirismo dos “turnos noturnos”, as inúmeras vezes que me ajudou nas diversas análises realizadas;

Aos amigos Jackson A. Marcondes de Souza e Tiago José Lopes, agradeço pela imensa ajuda na bancada e nas primeiras análises respectivamente, no início de minha tese, agradeço as explicações, a paciência e amizade no laboratório;

Ao amigo Tehuni Orlando González agradeço os conselhos, a amizade, a alegria e aos muitos momentos divertidos que propiciou ao longo do período que pudemos conviver;

Agradeço também a amizade dos amigos e amigas: Douglas, Renato, Dimitrov, Denilson, Silvana, Eliamar, Viviane e todos integrantes do laboratório que auxiliaram direta ou indiretamente na execução desta tese e tornaram o convívio no laboratório sempre muito agradável e divertido;

A Deus por tudo que tenho e que sou.

Muito Obrigado!

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>1</b>
<b>2. REVISÃO BIBLIOGRÁFICA.....</b>	<b>3</b>
2.1 TAXONOMIA DE PROCARIOTOS.....	3
2.2 MECANISMOS RESPONSÁVEIS PELA EVOLUÇÃO DE GENOMAS.....	5
2.3 ABORDAGENS DE ALTA RESOLUÇÃO NO RELACIONAMENTO TAXONÔMICO DE MICRORGANISMOS .....	10
2.4 A IMPORTÂNCIA DA CLASSIFICAÇÃO SISTEMÁTICA DE BACTÉRIAS DIAZOTRÓFICAS.....	13
<b>3. MATERIAL E MÉTODOS.....</b>	<b>16</b>
3.1. FASES ANTERIORES À OBTENÇÃO DOS MICROARRANJOS DE DNA.....	16
3.1.1. CONSTRUÇÃO DA BIBLIOTECA DE DNA DE <i>B. ELKANII</i> SEMIA 587 POR <i>SHOTGUN</i> .....	16
3.1.2 OBTENÇÃO DOS CLONES DE DNA.....	18
3.1.2.1. Fragmentação por nebulização .....	18
3.1.2.2. Precipitação do DNA.....	18
3.1.2.3. Reparo dos terminais dos fragmentos .....	18
3.1.2.4. Reação com Kinase.....	19
3.1.2.5. Obtenção dos fragmentos a partir do gel.....	19
3.1.2.6. Eluição dos fragmentos reparados do gel.....	20
3.1.2.7. Cálculo da concentração do DNA.....	20
3.1.2.8. Reação de ligação.....	20
3.1.2.9. Transformação de células.....	20
3.1.3. SELEÇÃO DOS CLONES.....	21
3.1.4. SEQÜENCIAMENTO DOS CLONES.....	21
3.1.4.1. Extração do DNA plasmidial.....	21
3.1.4.2. Reação de PCR e seqüenciamento.....	23

3.1.4.3 Análise das sequências geradas .....	23
3.2 CONSTRUÇÃO DOS MICROARRANJOS DE DNA.....	24
3.2.1. CONSTRUÇÃO DA SONDA (CHIP) DE <i>Bradyrhizobium elkanii</i> SEMIA 587 .....	24
3.2.1.1. Amplificação do dna plasmidial.....	24
3.2.2. IMPRESSÃO ROBÓTICA DOS PRODUTOS DE PCR.....	24
3.2.3 FIXAÇÃO DOS DNAs IMPRESSOS NAS LÂMINAS.....	25
3.2.4. CULTIVO DAS ESTIRPES.....	25
3.2.5. EXTRAÇÃO E PURIFICAÇÃO DO DNA.....	26
3.2.6 BIOCHIP DE DNA DE BRADYRHIZOBIUM ELKANII SEMIA 587.....	27
3.2.8. SÍNTESE DE DNAs FLUORESCENTEMENTE MARCADOS PARA HIBRIDIZAÇÃO (BOWTELL E SAMBROOK 2003-MODIFICADO).....	28
3.2.9. HIBRIDIZAÇÕES.....	28
3.3. ANÁLISE DOS DADOS.....	29
3.4 AGRUPAMENTO FUNCIONAL DOS RESULTADOS OBTIDOS.....	31
3.5 ANÁLISE FILOGENÉTICA.....	32
<b>4. RESULTADOS E DISCUSSÃO.....</b>	<b>34</b>
4.1. CONSTRUÇÃO DAS BIBLIOTECAS PLASMIDEAIS BE587, SEQÜENCIAMENTO E SELEÇÃO DE GENES PARA COMPOSIÇÃO DO MICROARRANJO DE DNA.....	34
4.2 HIBRIDIZAÇÃO E ANÁLISE DAS IMAGENS DE MICROARRANJO DE DNA ....	35
4.3. COMPARAÇÃO DE GENOMAS POR HIBRIDIZAÇÃO.....	37
4.4. ANÁLISE FILOGENÉTICA.....	49
<b>5. CONCLUSÕES .....</b>	<b>55</b>
<b>6. REFERÊNCIAS.....</b>	<b>56</b>
<b>7. APÊNDICE .....</b>	<b>72</b>

## COMPARAÇÃO FILOGENÉTICA DE GENOMAS DE RIZÓBIOS POR HIBRIDIZAÇÃO ATRAVÉS DE MICROARRAY

**RESUMO** - Diversos estudos com bactérias fixadoras de nitrogênio buscam identificar genes responsáveis pela fixação de nitrogênio, nodulação e simbiose, assim como conhecer seu genoma e compreender melhor o metabolismo delas. No entanto ainda há poucos trabalhos sobre o genoma de *Bradyrhizobium elkanii*, bactéria importante no contexto da produção de alimentos e de utilização comercial. Esse é o primeiro estudo a comparar 2654 genes de *Bradyrhizobium elkanii* SEMIA 587, com as bactérias *Bradyrhizobium japonicum* LGM 6138, *Azorhizobium caulinodans* LGM 6465, *Mesorhizobium huakuii* LGM 14107, *Rhizobium leguminosarum* bv. *Trifolii* LGM 8820 e *Sinorhizobium meliloti* LGM 6133 através da comparação de genomas por hibridização (CGH) usando Microarray. Portanto o objetivo deste trabalho foi realizar uma taxonomia molecular através da comparação de genomas por hibridização DNA-DNA, utilizando a tecnologia de microarranjos de DNA (CGH Microarray), para buscar novos genes que possam ser úteis na classificação filogenética de rizóbios. A técnica de CGH Microarray permitiu encontrar sessenta e cinco genes comuns entre todas as bactérias analisadas, após realizar comparação e classificação de todos genes encontrados em comum entre elas. Os resultados obtidos ao se comparar três árvores filogenéticas baseadas em diferentes genes, confirmaram que alterar a quantidade e as sequências nas análises, causa variação nas árvores filogenéticas como já observado em outros microrganismos. Os resultados sugerem que quanto maior for o número de genes ortólogos usados na filogenia, mais confiáveis serão as árvores filogenéticas obtidas. As bactérias que sofreram menos alterações de agrupamento nas diferentes árvores foram *B. elkanii* e *B. japonicum*.

**Palavras-chaves:** *Bradyrhizobium elkanii*, genes ortólogos, filogenia

## PHYLOGENETIC COMPARATION OF GENOME OF RHIZOBIA BY HIBRIDIZATION THROUGH MICROARRAY

**ABSTRACT** - Diverse studies with fixing nitrogen bacteria search to identify responsible genes for the nitrogen fix, nodulation and symbiosis, as well as knowing its genome and understanding the metabolism of them better. However still has few works about genome of *Bradyrhizobium elkanii*, important bacterium in the context of the production of foods and commercial use. This is the first study to compare 2654 genes of *Bradyrhizobium elkanii* SEMIA 587, with the bacteria *Bradyrhizobium japonicum* LGM 6138, *Azorhizobium caulinodans* LGM 6465, *Mesorhizobium huakuii* LGM 14107, *Rhizobium leguminosarum* bv. *Trifolii* LGM 8820 and *Sinorhizobium meliloti* LGM 6133 through the CGH Microarray. Therefore aim of this work was to carry a molecular taxonomy through the comparison of genomes for hybridization DNA-DNA, being used the technology of microarrangements of DNA (CGH Microarray), to search new genes that can be useful in the phylogenetic classification of rhizobia. The technique of CGH Microarray allowed to find sixty and five orthologs genes between all the analyzed bacteria, after to carry through comparison and classification of all genes in common between them. The results gotten to if comparing three established phylogenetic trees in different genes, had confirmed that to modify the amount and the sequences in the analyses, cause variation in the phylogenetic trees as already observed in other microorganisms. The results suggest that how much bigger will be the number of used ortholog genes in the phylogeny, more trustworthy will be the gotten phylogenetic trees. The results had confirmed that when modifying the amount of sequences in the phylogenetic analyses, had variation in the phylogenetic trees as already observed in other microorganism. The bacteria that had suffered little alterations from grouping in the different trees had been *B. elkanii* and *B. japonicum*.

**Keywords:** *Bradyrhizobium elkanii*, ortholog genes, phylogeny

## I. INTRODUÇÃO

Bactérias fixadoras de nitrogênio realizam importante papel na agricultura, dado ao potencial que algumas apresentam em efetuarem associações simbióticas com leguminosas. Nestas interações, as rizobactérias podem fixar nitrogênio diretamente da atmosfera reduzindo-o a amônia, que será assimilada em aminoácidos e proteínas utilizados pelas plantas hospedeiras.

As associações melhor compreendidas incluem algumas bactérias Gram-negativas, da ordem *Rhizobiales*, dos gêneros *Rhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Mesorhizobium*, *Sinorhizobium* e *Alorhizobium*. Estas bactérias são coletivamente denominadas de rizóbios. Inicialmente o termo rizóbio, referia-se a membros do gênero *Rhizobium*. No entanto com o passar dos anos, o termo tornou-se usado por todas as bactérias que são capazes de nodular e fixar nitrogênio em associação com plantas leguminosas e que no início da taxonomia de procariotos, pertenciam todas ao mesmo gênero *Rhizobium*, ou eram muito proximamente relacionadas a ele (WILLENS, 2006).

A inoculação de culturas agronomicamente importantes com bactérias fixadoras de nitrogênio traz grandes benefícios. Ela dispensa a aplicação de adubos nitrogenados no solo, reduzindo significativamente os custos da produção, além de evitar desequilíbrios ecológicos pela poluição do ambiente. Atualmente, no Brasil, são recomendadas quatro estirpes de rizóbios para o uso em inoculantes comerciais: as estirpes de *Bradyrhizobium* SEMIA 5019, SEMIA 587, SEMIA 5079 e SEMIA 5080 (VARGAS *et al.*, 1992; MENDES *et al.*, 2001).

A indicação de uma estirpe de rizóbio para utilização em inoculantes só ocorre após detalhados estudos que permitem avaliar seu potencial como simbiote fixador do nitrogênio, além da sua capacidade em competir com outras bactérias nativas do solo e que atuam na nodulação de determinada leguminosa. O Brasil é um dos países que possui maior biodiversidade, e à medida que a fixação biológica do nitrogênio assume importância nos sistemas de produção agrícola, aumenta o

número de coleções de rizobactérias isoladas. Contudo, o número de estirpes classificadas sistematicamente continua reduzido.

Trabalhos dedicam-se ao estudo de bactérias fixadoras de nitrogênio buscando identificar genes responsáveis pela fixação de nitrogênio, nodulação e simbiose, assim como compreender melhor o metabolismo e o genoma das mesmas. Como exemplos é possível citar pesquisas realizadas com *Bradyrhizobium japonicum* que teve seu genoma sequenciado (KANECO *et al.*, 2002) e analisaram a expressão por microarray (HAUSER *et al.* 2006). Do mesmo modo *Sinorhizobium meliloti* (GALIBERT *et al.*, 2001), *Mesorhizobium Loti* (KANECO *et al.*, 2000), *Rhizobium etli* (GONZALEZ *et al.*, 2006) e *Rhizobium leguminosarum* (YUNG *et al.*, 2006) tiveram seus genomas sequenciados e analisados. No entanto ainda há poucos trabalhos envolvendo o genoma de *Bradyrhizobium elkanii* (DALL ACQUA, 2004; SOUZA, 2006) descrito por RUMJANEK e colaboradores em 1993, essa bactéria é igualmente importante no contexto da produção de alimentos e de utilização comercial.

O presente trabalho comparou 2654 sequências de *Bradyrhizobium elkanii* SEMIA 587, com as bactérias *B. japonicum* LGM 6138, *Azorhizobium caulinodans* LGM 6465, *Mesorhizobium huakuii* LGM 14107, *Rhizobium leguminosarum* bv. *Trifolii* LGM 8820 e *Sinorhizobium meliloti* LGM 6133 através de CGH microarray.

Esse é o primeiro estudo a comparar o genoma parcial do *B. elkanii* ao de outras cinco bactérias fixadoras de nitrogênio, com o objetivo de realizar uma taxonomia molecular através da comparação de genomas por hibridização DNA-DNA, utilizando a tecnologia de microarranjos (CGH microarray), para buscar novos genes que possam ser úteis na classificação filogenética de rizóbios.

## II. REVISÃO BIBLIOGRÁFICA

### 2.1. TAXONOMIA DE PROCARIOTOS

Em procariotos a taxonomia pode ser descrita como a ciência que determina a classificação (criação de novos taxa), identificação (alocação de linhagens dentro de espécies conhecidas) e nomenclatura (VANDAMME *et al.*, 1996) desses organismos. Esta ciência produziu um sistema estável, previsível e altamente informativo que tem colaborado para o avanço de vários ramos da ciência, incluindo não somente a microbiologia, mas também a genômica, ciências médicas, ecologia de microrganismos, biotecnologia, evolução e epidemiologia (ROSELLÓ-MORA, 2005).

Sistemas baseados somente em propriedades morfológicas e comportamentais, levaram a sérios erros de classificação microbiana, nos mais variados grupos de bactérias. Tais métodos microbiológicos tradicionais baseados em características fenotípicas, como propriedades morfológicas, fisiológicas e bioquímicas, governaram por décadas a taxonomia microbiana e forneceram informação descritiva para a estruturação de diversos taxa bacterianos (BOONE & CASTENHOLZ, 2001).

Após o surgimento da taxonomia numérica (SNEATH & SOKAL, 1962) e o auxílio da computação, dados fenotípicos começaram a ser analisados por coeficientes numéricos que expressam similaridade entre linhagens com o auxílio da computação. A taxonomia numérica veio proporcionar maior objetividade aos esquemas de classificação microbiana, nesse sentido essa abordagem pressupõe a utilização de um grande número de testes bioquímicos (100 a 200) e uma amostragem grande e diversificada de linhagens, sendo os resultados expressos em porcentagens (VANDAMME *et al.*, 1996). A aplicação de taxonomia numérica levou à avanços significativos na classificação dos microrganismos, mais especificamente das bactérias (GOODFELLOW, 2000).

Uma variedade de características ecológicas e fenotípicas associadas ao uso da homologia DNA-DNA é freqüentemente usada na classificação de microrganismos e foi denominada de taxonomia polifásica por COLWELL (1970 ab). Colwell propôs a integração da informação do nível molecular ao ecológico para se obterem identificações e classificações mais precisas e confiáveis. A taxonomia polifásica tem como objetivo classificar microrganismos através da integração de diferentes tipos de dados e informações. Isto inclui dados fenotípicos (testes bioquímicos, composição de ácidos graxos, etc), dados genotípicos (métodos baseados em padrões de bandas diferentes) e dados filogenéticos (seqüências de genes ribossomais). (VANDAMME. *et al.*, 1996).

Desde 1970, métodos físico-químicos de hibridação DNA-DNA são usados para delimitar espécies procarióticas. Nesta abordagem, a similaridade genética total entre isolados é medida pelo grau de hibridização dos genomas dentro de determinadas condições padronizadas. Isolados que exibem valores de hibridização DNA-DNA superiores a 70% e menores que 5% de diferença na temperatura de fusão adequada são considerados como sendo da mesma espécie. Embora possa haver ajustes nesses valores de hibridização, pois de acordo com o relacionamento dos grupos taxonômicos analisados, isolados que exibem menos do que 50% hibridização DNA-DNA pertencem a espécies diferentes. (GEVERS *et al.*, 2005).

Geralmente a identificação de linhagens em nível de espécie por meio de testes fenotípicos e de seqüências 16S rRNA é um dos principais problemas na taxonomia de procariotos pois as características fenotípicas e de seqüências das espécies existentes são muito semelhantes. Neste caso, o critério decisivo para identificação de espécies ainda é a similaridade obtida em ensaios de hibridização de DNA entre genomas de dois organismos (STACKEBRANDT *et al.*, 2002).

Apesar de ainda ser adotada como um padrão a hibridização de DNA-DNA, esta é a parte mais laboriosa e demorada na taxonomia de procariotos, pois necessita de equipamentos especiais e técnicos treinados, o que faz com que essa metodologia seja realizada somente em poucos laboratórios especializados ao redor do mundo. Além disso, a hibridização DNA-DNA é uma técnica mal adaptada

à rápida identificação de procariotos, assim como não é aplicável à classificação de procariotos não cultiváveis, que são a maioria na biosfera. Os resultados obtidos com esta técnica não são cumulativos em bases de dados, sendo que cada experimento novo deve incluir as linhagens de referência (GEVERS *et al.*, 2005).

## 2.2 MECANISMOS RESPONSÁVEIS PELA EVOLUÇÃO DE GENOMAS

Com o aumento de genomas seqüenciados e depositados em bancos de dados públicos o uso desses dados para acessar o relacionamento filogenético entre espécies de procariotos também tem aumentado. Embora as seqüências de genes 16S rRNA e a hibridização DNA-DNA continuem a serem usadas como critérios de classificação de espécies, um número muito maior de informações adicionais taxonômicas pode ser extraído de seqüências de genomas completos. Além disso, a resolução de seqüências do gene 16S rRNA na análise de espécies próximas é suficiente para separar espécies diferentes, mas não é capaz de diferenciar estirpes de uma mesma espécie (COENYE *et al.*, 2005).

A disponibilidade de seqüências de genomas completos, de organismos proximamente relacionados, apresenta a oportunidade de reconstruir eventos de evolução de genomas, permitindo conhecer outras forças responsáveis por sua variação além da substituição de nucleotídeos. Através dela é possível identificar eventos de rearranjos cromossômicos específicos como aquisição e perda de gene, transferência horizontal e eventos de duplicação de genes (SNEL *et al.*, 2002; KUNING *et al.*, 2003; HONG *et al.*, 2004). Todos esses mecanismos podem afetar a classificação filogenética.

### Duplicação Gênica

Duplicação gênica é considerada como um importante mecanismo que facilita a adaptação a mudanças ambientais e na exploração de novos nichos (HOOPER *et al.*, 2003). Analisando a organização genômica de parálogos em

genomas bacterianos foi verificado que a maioria dos genes duplicados, em *Bacteria*, aparece em pequenos eventos de duplicação gênica. Genes parálogos compreendem uma significativa fração (até 44%) da capacidade de produção de proteínas de genomas bacterianos, uma fração do que é observado tem forte correlação com o tamanho do genoma. A respeito da organização de genes parálogos nos genomas, foi observado que 15% da coleção desses genes em 106 genomas bacterianos investigados consistiam de duplicatas em *tandem*, 9,5% localizada no bloco de segmentos duplicados e a maioria 75,5% é perdida em sua organização como consequência da dinâmica do genoma. (GEVERS *et al.*, 2004). A maioria de duplicações de blocos parecem ser um típico *operon* de três a quatro genes relacionados indicando uma possível história de duplicação do *operon* e sua retenção.

Estudando a expansão de genes de linhagens específicas JORDAN e companheiros (2001) analisaram a contribuição das duplicações gênicas nas diferenças genômicas entre genomas de procariotos. Verificaram que grupos de genes parálogos foram gerados exibindo a divergência específica de linhagens procarióticas, da mesma forma que havia sido inferida na comparação de sequências de genes 16S rRNA. Também foi observado que estas famílias de genes parálogos provavelmente estão contribuindo substancialmente nas diferenças fenotípicas entre linhagens bacterianas.

#### Transferência Horizontal de Genes (HTG) ou Transferência Lateral de Genes (LGT)

Alguns autores usam a expressão transferência horizontal de genes para indicar que houve transferência de genes entre gêneros e/ou espécies de bactérias diferentes. Outros usam para esse mesmo fenômeno a expressão transferência lateral de genes, portanto HTG e LGT são expressões sinônimas para o mesmo fenômeno.

Além da duplicação gênica com posterior divergência funcional, procariotos tem um mecanismo alternativo de adaptação gênica a seu ambiente. A introdução

de novos genes ou alelos por transferência horizontal de genes (HTG) permite a adaptação para nichos específicos, os quais podem iniciar uma diversificação e especiação bacteriana (OCHMAN *et al.*, 2000; COHAN, 2001).

Embora HTG seja conhecido por estar envolvido na transmissão de genes de resistência a antibióticos, por exemplo, somente na análise comparativa de múltiplos genomas procarióticos a escala de HTG tornou-se aparente (COENYE *et al.*, 2005). A evolução como é conhecida parece ser impossível sem um inter-relacionamento entre a variação verticalmente gerada e horizontalmente adquirida (WOESE, 2000).

A consciência do impacto potencial de HTG tem sido traduzida dentro de um conceito teórico de três categorias de genes: (i) um “núcleo compacto” de genes resistentes para HTG, (ii) um “núcleo leve” com genes que são raramente transferidos e (iii) a “concha” de genes os quais são suscetíveis a transferência horizontal de genes. Contudo HTG é um tema controverso na genômica, pois qualquer evidência filogenética de HTG pode também ser explicada através da combinação de duplicação gênica e eventos de perda de genes de uma linhagem específica (PHILIPPE *et al.*, 2003). Apesar disso HTG tem sido observada e quantificada diretamente através de técnicas de fluorescência a laser e o uso de gene de expressão (SORENSEN *et al.*, 2003).

### Perda de genes

Embora os genomas bacterianos possam adquirir genes, pelos processos de duplicação e transferência horizontal de genes, eles não crescem infinitamente, e a causa disso são eventos de balanceamento através da perda de genes. Isso ocorre porque genomas bacterianos podem proteger apenas uma quantidade finita de informações contra mutações e perdas. Deleções de cromossomos servem para eliminar genes que falham em fornecer uma função significativa, isto é, o conjunto de DNA adquirido, assim como sequências ancestrais supérfluas. (MIRA *et al.*, 2001).

A perda de genes, assim como o ganho deles, estão relacionados à adaptação da bactéria a diferentes ambientes. A taxa de deleção varia grandemente entre espécies diferentes (MARRI *et al.*, 2006).

A redução de genomas tem sido documentada em grupos bacterianos que apresentam estilo de vida associado ao hospedeiro. O tecido do hospedeiro fornece constantemente muitos metabólitos intermediários que eliminam a pressão para manutenção de muitos genes responsáveis pela biosíntese desses metabólitos. A adaptação de microrganismos intracelulares experimenta, portanto uma indução deletional. Isso ocorre porque bactérias associadas a hospedeiros possuem pequena população genética quando comparadas aos parentes de vida livre, como resultado dos altos níveis de fixação de leves mutações deletérias. A redução do genoma é possivelmente uma das principais forças por trás da evolução de parasitas e/ou bactérias intracelulares e é chamada de evolução por redução (MORAN, 2003; KLASSON *et al.*, 2004).

#### Rearranjos cromossômicos

O rearranjo genômico é um dos principais elementos responsáveis pela evolução dos genomas, entre todos os fenômenos que influenciam a variação do mesmo. A organização e evolução do genoma bacteriano não podem ser compreendidas antes do completo seqüenciamento de dois organismos próximos com posterior comparação deles. O nível de conservação pode ser alto quando organismos são filogeneticamente próximos, principalmente porque os rearranjos ainda não tiveram tempo de ocorrer. A avaliação da conservação da ordem dos genes no genoma em relação à escala evolucionária demonstrou que o número de substituições de aminoácidos, assim como o grau de rearranjo do genoma, constantemente aumentam ao longo do tempo de divergência (SUYAMA *et al.*, 2001). A conservação da ordem dos genes em um genoma pode ser usada como medida filogenética para estudar o relacionamento entre espécies (KORBEL *et al.*, 2002). O grau de plasticidade do genoma é dependente da quantidade de seqüências repetidas e móveis como transposons, plasmídeos e bacteriófagos.

Cromossomos contendo alta densidade de repetições têm altas taxas de rearranjos liderando acelerada perda de ordem gênica. (ROCHA, 2003). Portanto esse é um dos fatores que atuam fortemente nos estudos filogenéticos, agrupando ou separando os grupos de procariotos analisados.

#### Efeitos dos mecanismos de evolução de genomas na taxonomia de procariotos

Os mecanismos de duplicação gênica, perda gênica, transferência horizontal de genes e rearranjos cromossômicos possibilitam refazer a história evolucionária, explicando o relacionamento filogenético de microrganismos.

Os métodos genotípicos usados na taxonomia bacteriana, como determinação de conteúdo G+C, hibridização DNA-DNA e seqüenciamento do gene 16S rRNA, não são fortemente influenciados pelas forças que alteram os genomas. O conteúdo G+C de um genoma não é significativamente afetado por rearranjos cromossômicos, perda ou duplicação gênica. Por outro lado a transferência horizontal de genes pode afetar o conteúdo G+C dependendo da quantidade de genes transferidos e composição dos mesmos. Os métodos baseados em tipagem de organismos onde um padrão de bandas é gerado incluem PFGE (pulsed-field gel electrophoresis) e vários métodos baseados em PCR, entre eles BOX, RAPD e AFLP. O padrão obtido por esses métodos pode ser alterado pelas forças que alteram os genomas. Rearranjos cromossômicos podem exercer forte efeito sobre os padrões de bandeamento obtidos por “fingreprint”.

Características fenotípicas e quimiotaxonômicas são influenciadas indiretamente pelas mudanças no genoma. O efeito de duplicação gênica irá depender da taxa evolucionária de duplicação dos genes no genoma. O gene duplicado pode envolver uma nova função benéfica, podendo conduzir a aquisição de novas características bioquímicas, ou cada cópia pode adotar partes das tarefas originais anteriores a sua duplicação, o que pouco afetará as características bioquímicas. Do mesmo modo a aquisição de genes por transferência horizontal pode resultar em novas aptidões, no caso dos genes adquiridos codificarem

funções que não estavam presentes no genoma que o recebeu, enquanto a inativação do gene ou perda irá normalmente ser acompanhada pela perda de aptidões específicas (CONYE *et al.*, 2005).

Com o avanço da ciência novos métodos de análise de microrganismos estão sendo introduzidos, possibilitando estudar os mecanismos de evolução de genomas e fornecendo novas oportunidades para a sistemática de procariotos.

### **2.3. ABORDAGENS DE ALTA RESOLUÇÃO NO RELACIONAMENTO TAXONÔMICO DE MICRORGANISMOS**

Para representar a biodiversidade nos seus diferentes níveis, a taxonomia polifásica integra dados fenotípicos, quimiotaxômicos, moleculares e genômicos. Além disso, a taxonomia polifásica é um consenso entre sistematas (VANDAMME *et al.*, 1996). Neste contexto, diversas técnicas genômicas baseadas em padrões de banda (*fingerprints*), por exemplo, *AFLP*, *rep-PCR* e ribotipagem, foram aplicadas na década passada (DIJKSHOORN *et al.*, 2001). Vários estudos independentes mostraram uma alta correlação entre a similaridade de padrões de *AFLP* e de hibridização DNA-DNA para diversos grupos taxonômicos modelos. Por este motivo, o *AFLP* foi sugerido como uma alternativa para as hibridizações de DNA (STACKEBRANDT *et al.*, 2002; THOMPSON *et al.*, 2004). Apesar da técnica de *AFLP* ser rápida, altamente discriminatória e dos resultados poderem ser acumulados em bases de dados locais, a comparação de padrões de *AFLP* gerados em diferentes laboratórios é muito difícil, comprometendo a criação de bancos de dados públicos para a identificação de procariotos.

A disponibilidade de sequências de genomas completos possibilitou o desenvolvimento de abordagens como filogenia baseada em conteúdo de genes ortólogos (DAUBIN *et al.*, 2002; DUTILH *et al.*, 2007), seqüências assinatura (COENYE *et al.*, 2004) e alinhamento de sequências concatenadas (DAUBIN *et al.*, 2003). Embora a filogenia possa ser feita através de seqüenciamento e alinhamento

de genomas completos (EDWARDS-INGRAM *et al.*, 2004) o tempo e o custo tornam esse método impraticável se for realizado para todos os grupos de bactérias que se deseja analisar.

Técnicas como MLST, MLSA e técnicas baseadas no conteúdo genômico como CGH Microarray atualmente são as mais precisas na abordagem taxonômica de procariotos.

A técnica de tipagem de seqüências baseadas em multilocus – MLST (multi locus sequence typing - MAIDEN *et al.*, 1998) tem sido usada com freqüência para distinguir estirpes de uma mesma espécie bacteriana ou espécies proximamente relacionadas (BALDWIN *et al.*, 2005; HANAGE *et al.*, 2005). Nessa técnica, cada seqüência diferente é assinalada com um diferente número de alelos e cada estirpe é definida sem ambigüidade para sete genes, os quais correspondem ao número de alelos nos sete *loci*. Em outras palavras a abordagem básica dessa técnica é concatenar seqüências de cinco a sete genes constitutivos (housekeeping), que estejam presentes nas bactérias a serem analisadas. O relacionamento entre estirpes de espécies é exibido por um dendrograma usando matriz de diferença de pares nos perfis alélicos (HANAGE *et al.*, 2006).

Já a técnica MLSA (multilocus sequence analysis) é uma variação da MLST, trata-se de um método de caracterização genotípica que se baseia no uso de múltiplos genes que codificam proteínas, sendo usada para estudar grupos mais diversificados de procariotos (por exemplo: todos componentes de um gênero) (GEVERS *et al.*, 2005).

Trabalhos sugerem que um grande número de genes “core”, ou genes núcleos, poderiam ser usados para uma análise filogenética mais acurada (DAUBIN *et al.*, 2002; KURTZMAN *et al.*, 2003). Genes “core” podem ser definidos como o conjunto de todos os genes exibidos como ortólogos entre todos os membros de um grupo evolucionário coerente como o filo *Proteobacteria*, por exemplo (HARRIS *et al.*, 2003). Constatou-se que um conjunto de genes “core” de vinte ou mais seqüências resultam em uma análise filogenética acurada (ROKAS *et al.*, 2003). Entretanto o seqüenciamento desse número de genes para cada análise

filogenética tornaria a técnica pouco rotineira, pois embora seqüenciar genes esteja se tornando uma prática comum, nem todos os laboratórios possuem recursos suficientes para realização constante de sequenciamentos. Por outro lado as análises baseadas no alinhamento de seqüências de um ou poucos *loci* podem gerar inferências filogenéticas diferentes dependendo de quais genes são selecionados para o estudo (VAN BERKUM *et al.*, 2003; EARDLY. *et al.*, 2005; DEWHIRST *et al.*, 2005).

Seqüências de genes permitem a construção de árvores filogenéticas que refletem a história evolucionária das espécies em um dado clade e deste modo fornecem uma taxonomia natural. O problema de filogenias que se baseiam apenas na seqüência de um gene (um locus), é a variação das árvores filogenéticas montadas quando se usam genes diferentes para o mesmo grupo de organismos comparados. (SICHRITZ-PONTEN *et al.*, 2001; KURTZMAN *et al.*, 2003, KWON *et al.*, 2005)

Um método alternativo as técnicas baseadas em um ou poucos genes, é a comparação de genomas por hibridização através de Microarray (CGH microarray). Uma maneira comum de realizar essa técnica é usar genomas completos, ou grandes conjuntos de genes e “open reads frames” (ORFs), de uma estirpe seqüenciada (PORWOLLIK *et al.*, 2002). A hibridização de amostras de estirpes dentro do “array” identifica genes que estão ausentes, presentes ou que são altamente divergentes da estirpe de referência. O relacionamento genético é determinado baseado nos padrões de presença e ausência entre estirpes. Inferir relacionamentos filogenéticos por CGH microarray é muito mais vantajoso do que métodos baseados em somente um locus ou múltiplos loci, porque as lâminas contendo microarranjos podem conter muitas centenas de DNAs. Desse modo uma simples lâmina contendo microarranjos tem potencial para uma ampla identificação de um grande número de microrganismos ou genes de uma só vez, assim muito mais informação é incorporada na análise (EDWARDS-INGRAN *et al.*, 2004).

Estudos recentes têm empregado essa técnica para comparar genomas e realizar inferências filogenéticas (PORWOLLIK *et al.*, 2002; EDWARDS-INGRAN *et*

*al.*, 2004; GIUNTINI *et al.*, 2005; HAUSER *et al.* 2006; JING *et al.* 2006; NASH *et al.*, 2006, AAKRA. *et al.*, 2007). Estirpes bacterianas proximamente relacionadas de *Helicobacter pylori* tiveram seus genomas comparados (SALAMA *et al.* 2000), assim como diferentes estirpes de *Pseudomonas* (CHO E TIEDJE, 2001), estirpes de *Salmonella* (PORWOLLIK *et al.*, 2002), isolados de *Pseudomonas aeruginosa* (WOLFGANG *et al.*, 2003), espécies diferentes de *Salmonella* (CHAN *et al.*, 2003), estirpes diferentes de *Rickettsia prowazekii* (GE *et al.*, 2004), *Escherichia coli* K-12 e *Shigella* (FUKIYA *et al.*, 2004). Também foram realizados estudos utilizando CGH microarray entre bactérias muito diferentes como *Klebsiella pneumoniae* 342 e *Escherichia coli* K-12 (DONG *et al.*, 2001) e na análise do plasmídeo pSymA de *Sinorhizobium meliloti* (GIUNTINI *et al.*, 2005).

Porém outros microrganismos como fungos também estão sendo estudados pela técnica de CGH microarray. WATANABE *et al.*, (2004) obtiveram sucesso ao comparar estirpes de leveduras usando um “chip” com 6000 genes, levantando informações sobre a determinação de espécies e alterações no número de cópias de genes entre elas. Em outro estudo EDWARDS-INGRAM *et al.*, (2004) realizaram um estudo filogenético de *Saccharomyces cerevisiae* e *S. Boulardii* que possibilitou a obtenção de explicações sobre alguns eventos evolucionários relativos a filogenia dessas espécies. A técnica também vem sendo usada para estudos sobre o câncer (LAGE *et al.*, 2003) e até mesmo em organismos mais complexos como o genoma do rato (CHUNG *et al.*, 2004).

#### **2.4. A importância da classificação sistemática de bactérias fixadoras de nitrogênio.**

O interesse agrônômico pelas rizobactérias origina-se no fato de que algumas delas possuem habilidade em fixar o nitrogênio atmosférico, quando em simbiose mutualista com plantas da família *Fabaceae*, de grande interesse na alimentação humana.

A nomenclatura das espécies de *Rhizobium* foi originalmente concebida na crença de que uma classificação natural poderia ser baseada na especificidade da variação das espécies e estirpes bacterianas em relação às suas plantas hospedeiras. Há muitos problemas nesta metodologia de nomenclatura, incluindo o fato de existir aproximadamente 18000 espécies de leguminosas diferentes. Além disso, estirpes consideradas “específicas” para algumas plantas, podem variar devido a possíveis combinações de material genético. Também é necessário observar que algumas estirpes podem nodular muitas plantas diferentes. (YOUNG *et al.*, 2001; BROUGHTON 2003; WILLEMS 2006).

Os rizóbios podem ser isolados dos nódulos, mas não são facilmente identificados quando isolados do solo. Contudo, selecionar estirpes de rizóbio mais eficientes no processo de fixação biológica de nitrogênio para ser utilizado como inoculante comercial, requer estudos detalhados. Não há um critério formal para a definição de um gênero bacteriano, mas padrões mínimos sugeridos para a descrição de novos gêneros e espécies de rizóbios que nodulam raízes e caules podem ser considerados como referência.

O conceito de espécie para procariotos é assunto de grande discussão entre microbiologistas (COHAN 2002). Isso porque de um lado há a idéia de que todos os organismos, incluindo os procariotos, estão dentro de grupos de fenótipos similares, apresentam elevada identidade de sequências, conteúdo genômico parecido e mesmo ambiente (ROSELLÓ-MORA *et al.*, 2001).

Do outro lado há aqueles que defendem o ponto de vista onde, espécies procarióticas são um grupo sujeito a forças dinâmicas específicas, e adotam um modelo de espécie baseado nas forças evolucionárias sugeridas por padrões empíricos de variação genética. Um desses modelos é motivado principalmente pela extensiva evidência de transferência horizontal de genes (HTG) e recombinação homóloga observada entre parentes próximos (RODRIGUEZ VALERA 2004; COHAN *et al.*, 2007). Esta visão sugere que espécies procarióticas são grupos de bactérias ecologicamente heterogêneas cuja diversidade é forçada

por HTG e recombinação homóloga (LAWRENCE, 2002; GOGARTEN *et al.*, 2002), que podem auxiliar na adaptação às mudanças ambientais.

Entretanto, a definição de espécie bacteriana mais usada é: “uma categoria que circunscreve um grupo genômico coerente de indivíduos isolados exibindo um alto grau de similaridade em muitas características independentes, comparativamente testadas dentro de condições altamente padronizadas” (STACKEBRANDT *et al.*, 2002). Em outras palavras, espécies bacterianas são consideradas como sendo grupos de estirpes que são caracterizadas por certo grau de consistência fenotípica, por um significativo grau de (50-70%) hibridização do DNA e mais do que 97% de identidade de seqüência do gene 16S rRNA.(VANDAMME *et al.*, 1996, URSING *et al.*,1995).

A presença de numerosas seqüências de genes e de árvores filogenéticas depositadas em bancos de dados “on-line”, derivadas tanto dos genes 16S rRNA e *nifH*, somados aos resultados obtidos na análise de outras características (a presença e natureza de polissacarídeos, a utilização e metabolismo de carboidratos, a composição protéica, etc.) ampliaram e facilitaram a identificação de rizobactérias, mas o número de espécies classificadas sistematicamente continua reduzido. Pesquisadores buscam solucionar problemas taxonômicos relativos aos rizóbios baseando-se na técnica de análise do 16S rRNA.

Analizando 20 seqüências de rizóbios, YOUNG *et al.*, (2001) observaram que algumas seqüências podem esconder uma ou mais possíveis novas espécies. Em outro trabalho YOUNG *et al.*, (2004), através da análise desse mesmo gene enfatiza a necessidade de unir os gêneros *Rhizobium* e *Agrobacterium* em um só, pois tomando como referência somente esta técnica, as diferenças não dão suporte suficiente para a divisão dos mesmos. Por isso houve a mudança da classificação taxonomica do *Agrobacterium larrymoorei*, para *Rhizobium larrymoorei* baseada em YOUNG 2004 e aceita no manual Bergeys por GARRIT *et al.*, 2005.

Nesse contexto a utilização da classificação taxonômica baseada em vários genes torna-se evidente, pois seria mais imprecisa a classificação das bactérias pelas técnicas que utilizam apenas um gene .

### III. MATERIAL E MÉTODOS

#### 3.1. Fases anteriores à obtenção dos Microarranjos de DNA

Antes da obtenção final do microarranjo de DNA de *Bradyrhizobium elkanii* utilizado neste estudo, várias etapas tiveram que ser concluídas (PRADO, 2004; DALL ACQUA, 2004; SOUZA, 2006).

##### 3.1.1. Construção da Biblioteca de DNA de *B. elkanii* SEMIA 587 por “shotgun”

As células bacterianas de *Bradyrhizobium elkanii* SEMIA 587 foram cultivadas em meio TY (Tryptona 5g/L, Extrato de Levedura 3g/L e CaCl<sub>2</sub>.2H<sub>2</sub>O 0,87g/L) por um período de 48 horas, a 28°C, sob agitação de 140 rpm. A suspensão contendo bactérias foi então submetida à centrifugação (12.000xg, 30 minutos a 4°C). O sedimento foi lavado com solução salina (NaCl 0,85%) para remoção de polissacarídeos extracelulares e possíveis resíduos do meio de cultivo. Após a centrifugação e lavagem as células foram mantidas a -20°C (estoque) para posterior utilização.

Para a extração de DNA (SAMBROOK *et al.*, 1989), as células foram retiradas do estoque e ressuspensas em 1mL de solução salina (NaCl 0,85%). As suspensões foram transferidas para microtubos e centrifugadas a 15.600xg por 5 minutos, sendo este passo repetido por mais uma vez para a completa remoção de resíduos do meio de cultivo. Depois da dupla lavagem os sedimentos foram ressuspensos em 1mL de solução NaCl-EDTA (0,15M - 0,01M) e incubados a 37°C durante 10 minutos, em tubos Corex. Após esse período, adicionaram-se 500µL de lisozima (5mg/mL) preparada em tampão Tris-EDTA-dextrose (10mM;

24mM; 50mM), procedendo-se então a uma nova incubação a 37°C, por 30 minutos. Posteriormente, adicionaram-se 500µL de solução de SDS (duodecil sulfato de sódio) a 20%, e novamente realizou-se uma incubação a 56°C, durante 20 minutos.

Decorrido o tempo de incubação foram adicionados à suspensão, 500µL de solução de perclorato de sódio 5M, gota a gota com agitação manual suave. Em seguida adicionaram-se 2mL de solução de fenol-clorofórmio-álcool isoamílico (25:24:1), deixando-se por 1 hora em agitação orbital a 220 rpm, em banho de gelo e no interior de uma câmara fria.

Após esse período o material foi centrifugado a 12.000xg, por 20 minutos, a 4°C. Coletou-se então, com auxílio de uma pipeta Pasteur a fase superior transferindo-a para tubos de ensaio limpos. À fase recuperada foi adicionada 2mL de solução clorofórmio-álcool isoamílico (24:1), procedendo-se novamente uma agitação orbital a 220 rpm em banho de gelo e no interior de uma câmara fria 20 minutos. O material foi centrifugado a 5.860xg por 20 minutos, a 4°C, em seguida coletou-se a fase superior, a qual foi armazenada novamente em novos tubos Corex. O processo foi novamente repetido adicionando-se 2mL de clorofórmio-álcool isoamílico até a recuperação e transferência da fase superior para tubos limpos, aos quais foram adicionados 2 volumes de etanol 95% (gelado), agitando-se manualmente e incubando-se a -20°C, por 12 horas para a precipitação do DNA.

Depois desta fase de incubação as amostradas foram centrifugas a 5.860xg por 20 minutos a 4°C, em seguida descartando-se o sobrenadante. O precipitado foi lavado com 1mL de etanol 70% (gelado), rapidamente e de maneira cuidadosa, por duas vezes para evitar a perda do DNA. Os tubos foram então vertidos sobre papel absorvente e depois mantidos à temperatura ambiente por 35- 40 minutos para eliminar o excesso de etanol. O sedimento seco foi ressuspendido em 1mL de tampão TE a 10°C (Tris-HCl 10mM; EDTA 1mM) e o DNA foi transferido para tubos tipo *ependorf* e estocados a -20°C.

O DNA foi quantificado em espectrofotômetro com uma absorbância de 260nm.

### **3.1.2. Obtenção dos clones de DNA**

#### **3.1.2.1. Fragmentação por nebulização**

Foram adicionados 50µg de DNA em um nebulizador contendo tampão Tris-HCl (1M pH 8,0), MgCl<sub>2</sub> (1M), glicerol esterilizado (50%), água filtrada e esterilizada (q.s.p. 2mL). A nebulização deu-se por 25 segundos a uma pressão de 3kgf/cm<sup>2</sup> com nitrogênio. As amostras foram centrifugadas a 4°C, por 2 minutos, a 12.000xg para concentrar o material no nebulizador.

#### **3.1.2.2. Precipitação do DNA**

O DNA fragmentado foi transferido para tubos de 1,5 mL (máximo de 400µL de DNA). Foi acrescentado 0,1 volume de acetato de sódio 3M pH 5,2, e depois levemente homogeneizado, em seguida foram acrescentados 2,5 volumes de etanol 95% (temperatura ambiente).

Os tubos foram incubados em gelo seco por 20 minutos, procedendo-se a uma centrifugação a 15.294xg, por 30 minutos a 4°C. Descartou-se o sobrenadante, lavando-se em seguida as amostras com 1 mL de etanol 70%. Uma nova centrifugação foi realizada a 15.294xg por 15 minutos a 4°C. O sobrenadante foi descartado com pipetador e o material sedimentado foi seco a vácuo. A amostra foi ressuspendida em tampão TE, originando no final 1 microtubo com 44 µL de DNA.

#### **3.1.2.3. Reparo dos terminais dos fragmentos**

Para cada amostra contendo 44 µL de DNA acrescentaram-se 50 mM MgCl<sub>2</sub>, 0,5 mM de dNTP e 3U/µL de DNA polimerase bacteriófago T4, esse material foi incubado em temperatura ambiente, por 15 minutos. Posteriormente adicionou-se 2µL (10U) de fragmento Klenow da DNA polimerase I de *E. coli* e foi realizada uma incubação em temperatura ambiente, por 15 minutos. Aos fragmentos em reparo foi adicionado 40 µL de água e 100 µL de uma solução 1:1 de fenol/clorofórmio à temperatura ambiente (fenol saturado com Tris-HCl pH 8). A mistura foi agitada em *vortex* por 1 minuto, seguida de centrifugação a 15.294xg, por 2 minutos, a 4°C. A

fase aquosa foi transferida para um outro tubo de 1,5 mL e adicionou-se 100  $\mu$ L de éter etílico saturado com água (5:1 V:V). Centrifugou-se o material a 15.294xg, por 5 minutos, em temperatura ambiente. A fase orgânica superior foi retirada, deixando-se o tubo aberto para a evaporação completa do éter. No tubo foi adicionado 0,1 volume de NaCl 1M e 2,5 volumes de etanol absoluto, misturando-se por inversão, em seguida a amostra foi deixada em gelo seco, por 30 min. O material foi novamente centrifugado a 12.000xg, por 30 minutos, a 4°C, sendo o sobrenadante descartado e o sedimento (“*pellet*”) lavado com 1mL de etanol 70% (temperatura ambiente). O material foi centrifugado a 12.000xg, por 15 minutos, em temperatura ambiente, descartando-se o sobrenadante, sendo então seco a vácuo e ressuspendido em 46  $\mu$ L de água filtrada e esterilizada.

#### **3.1.2.4. Reação com Kinase**

Aos fragmentos de DNA foram adicionados 20mM riboATP, 10U de polinucleotídeo Kinase do bacteriófago T4, tampão polinucleotídeo Kinase do bacteriófago T4 (10x) e 100mM de DTT (Dithiothreitol). As amostras foram incubadas a 37°C por 30 minutos. A reação foi inativada por tratamento com calor a 65°C por 5 minutos, adicionando-se em seguida 2 $\mu$ L de tampão de corrida (40% de glicerol e 0,02% de azul de bromofenol) para cada 20  $\mu$ L de DNA. Essas amostras foram submetidas à eletroforese em gel de agarose com baixo ponto de fusão (“*Low Melt*”) 0,8% com tampão TBE 1x (Tris 89mM, Ácido Bórico 89mM e EDTA 2,5mM; pH 8,3) sem brometo de etídio. Um marcador foi aplicado juntamente com uma pequena amostra do material para corar e documentar. A eletroforese foi realizada a 90V por aproximadamente 2 horas.

#### **3.1.2.5. Obtenção dos fragmentos a partir do gel**

Procedeu-se literalmente o corte do gel entre os poços que continham o marcador e uma pequena amostra do material, resultando dois fragmentos de gel, uma com o material nebulizado e outra com o marcador e amostra para corar. A porção do gel que continha o marcador foi corada com brometo de etídio, e sob um

transiluminador com UV, marcou-se com pequenos cortes a região do tamanho de fragmento desejado (de 600 a 1000pb). O gel corado foi colocado ao lado do gel com o material reparado e com uma lamínula foi cortada a região que continha o material reparado, colocando-se a região de gel cortada em tubos *ependorf* (aproximadamente 400mg de gel/tubo).

#### **3.1.2.6. Eluição dos fragmentos reparados do gel**

Os fragmentos selecionados de DNA foram eluídos do gel através do kit *Concert™ Rapid PCR Purification System* (Gibco BRL - Cat No. 11458-015).

#### **3.2.1.7. Cálculo da concentração do DNA**

Para a quantificação do DNA fragmentado pipetou-se 5 µL do material em 40 µL de água (diluição de 10x). Realizou-se leitura em espectrofotômetro com um comprimento de onda de 260 nm. Para a confirmação da presença do DNA e do tamanho dos fragmentos foi realizada uma eletroforese em gel de agarose 0,8% com tampão TBE (1x), submetida a 90 V, por 1,5 horas.

#### **3.2.1.8. Reação de ligação**

Foram colocados, em um tubo de 1,5 mL, 400 ng de DNA num volume máximo de 20 µL juntamente com o vetor (100 ng de vetor pUC18-*Sma*I) agitando-se gentilmente com o pipetador e adicionando-se 20U de T4 DNA Ligase. A mistura foi incubada pelo período de 16 horas, à 16°C.

#### **3.2.1.9. Transformação de células**

Em tubos do tipo "Falcon", de 15 mL, foram adicionados 5 µL de DNA ligado com o vetor. Nesse momento, as células competentes (*E. coli* DH5α) foram removidas da temperatura -85°C e foram colocadas em banho frio por ±5 minutos, misturando-se gentilmente. As células, então, foram adicionadas (50 µL) aos tubos com DNA sendo misturados gentilmente e incubados em banho de gelo por 20

minutos. Os tubos foram então levados ao banho-maria a 42°C sem agitação, por 1 minuto. Após esse período, os tubos foram retornados ao gelo, por 2 minutos, adicionando-se, após isso, 1 mL de meio SOC [Tryptona (20g/L), Extrato de Levedura (5g/L), NaCl (0,5g/L), KCl 1M (2,5mL/L), MgCl<sub>2</sub> 1M (10mL/L) e Glicose 1M (200μL/L)] no tubo. Os tubos foram incubados a 37°C, com agitação de 150xg, durante 2 horas. A cultura de células do tubo foi então cultivada em 10 placas contendo 100 μL (por placa) de meio 2TY (Tryptona 16g/L, Extrato de Levedura 10g/L e NaCl 5g/L) contendo ampicilina (50mg/mL de meio). Após a solidificação do meio, aplicou-se, com alça de Drigalski, uma solução de 300 μL de X-gal (0,0175g de *Blue Gal*; 0,5mL de Dimetil Formamida e 2,8mL de água filtrada) por placas. As placas foram incubadas a 37°C por um período de 24 horas.

### 3.1.3 Seleção dos clones

Após a incubação foram selecionadas as colônias de coloração branca. A inserção do fragmento de DNA dá-se na região do gene *lacZ* responsável pela síntese de β-galactosidase que quebra o substrato X-gal, originando coloração azul. Portanto, as células que receberam o inserto formam colônias brancas uma vez que não são mais capazes de sintetizarem a β-galactosidase e as que não receberam, formam colônias azuis. Os clones transformados foram repicados com palitos de madeira autoclavados para placas estéreis de poliestireno para cultivo de bactérias (*96-well assay plate* de 250μL cada poço) contendo 150μL de meio *Circle Grow* (BIO 101, Inc; Cat. No. 3000-142) com ampicilina (50mg/mL). As placas foram armazenadas em estufa BOD (Biological Oxygen Demand) por 24 horas. Após esse período, foram adicionados 100 μL de solução de glicerol 40%. Essas placas constituíram-se no estoque da biblioteca de clones e foram estocadas a -85°C para posterior seqüenciamento dos clones. No final foram obtidas 94 placas totalizando 9.024 clones.

### 3.1.4 Seqüenciamento dos Clones

#### 3.1.4.1. Extração do DNA plasmidial

Foram transferidos 5  $\mu$ L de suspensão de cada clone estocado para placas de cultivo de bactérias (Mega Titer, "96 wells", 2,5mL cada poço), contendo 1mL de meio Circle Grow. As placas foram incubadas com agitação (230 rpm), a 37°C, por 24 horas.

Após a incubação, as placas foram centrifugadas por 5 minutos, os sedimentos foram ressuspensos em 240  $\mu$ L de solução GTE [Glicose 50mM; Tris-HCl 23mM (pH 8,0); EDTA 10mM], e agitados em *vortex* até a dissolução do sedimento. As placas foram novamente centrifugadas, e os sedimentos foram adicionados 80 $\mu$ L de solução de GTE acrescidos de 5 $\mu$ g de ribonuclease A. As amostras foram agitadas em *vortex* e 60  $\mu$ L, de cada uma foram transferidos para microplacas de polipropileno. A essas suspensões adicionou-se 60 $\mu$ L de solução de lise (NaOH 0,2N/ SDS 1%) e estas foram misturadas gentilmente por inversão (5 vezes) e incubadas por 10 minutos em temperatura ambiente. Após isso, adicionou-se 60  $\mu$ L de Acetato de potássio 3M, agitou-se suavemente por inversão e as placas foram incubadas por 10 minutos em temperatura ambiente. As amostras foram novamente incubadas a 90°C, por 30 minutos, e então resfriadas em gelo, por 5 minutos. As placas foram centrifugadas por 4 minutos, a 20°C, a 3220xg e todo o material foi removido para placas Millipore providas de filtro e acopladas em placas de polipropileno novas. Esse conjunto de placas foi então centrifugado por 5 minutos, a 3220xg, 20°C. As placas com filtro foram removidas e descartadas e às placas de polipropileno, contendo o material filtrado, foram adicionados 110  $\mu$ L de isopropanol absoluto. O material foi misturado por inversão e as placas foram centrifugadas por 45 minutos, 3220xg, à 20°C, com o objetivo de precipitar o DNA plasmidial. O sobrenadante foi então descartado e o precipitado foi lavado com 200  $\mu$ L de etanol 70% gelado. As placas foram centrifugadas por 5 minutos nas mesmas condições de rotação e temperatura descritas anteriormente. Após esse

procedimento, o sobrenadante foi descartado e o precipitado seco em temperatura ambiente, por 60 minutos. O DNA plasmidial precipitado foi então ressuscitado em 75  $\mu$ L de água filtrada autoclavada e quantificado por eletroforese em um gel de agarose 0,8%, a 100 V por 60 minutos, usando como padrões de peso molecular 50, 100 e 200 ng do plasmídeo pGEM.

#### **3.1.4.2 Reação de PCR e seqüenciamento**

O DNA plasmidial extraído e quantificado foi usado numa concentração de 100 ng, com 2  $\mu$ L de *primer* (10U) (M13-pUC #1201-R ou #1211-F), 2 $\mu$ L de *Big Dye* [*Big Dye<sup>TM</sup> Terminator Cycle Sequencing Ready Reactions* (Applied Biosystems – Cat No. 4314419)], 2  $\mu$ L de tampão de reação e água filtrada e esterilizada q.s.p. 10 $\mu$ L. A reação foi conduzida em termociclador com os seguintes ciclos: 2 minutos a 96°C para uma desnaturação inicial e 40 ciclos de 10 segundos a 96°C (desnaturação), 20 segundos a 52°C (pareamento) e 4 minutos a 60°C para extensão final. Após a amplificação, o material foi precipitado, desnaturado com Hi-Di-Formamida a 95°C, por 5 minutos, sendo então seqüenciado no 3700 DNA Analyzer ABI Prism (Applied Biosystems).

Para o sequenciamento, todo o procedimento descrito anteriormente foi realizado com 96 clones escolhidos ao acaso a fim de se verificar a validade das bibliotecas (avaliar a porcentagem de vetores presentes nas seqüências, verificar se o fragmento escolhido está distribuído por todo o genoma da bactéria e confirmar o tamanho do inserto) através do software Phred/phrap/Consed.

#### **3.1.4.3 Análise das seqüências geradas**

Após o sequenciamento, os eletroferogramas de seqüências de DNA foram analisados pelo programa PhredPhrap (EWING & GREEN, 1998) e Consed (GORDON *et al.*, 1998) a fim de gerar seqüências no formato FASTA, avaliar a qualidade e realizar alinhamento das mesmas. Seqüências apresentando valores

phred igual ou maior que 20 foram consideradas satisfatórias e usadas no estudo. Posteriormente as seqüências FASTA foram submetidas ao “GenBank” do “National Center for Biotechnology Information” (NCBI), através do programa BLASTN, para comparação com seqüências similares de nucleotídeos depositadas no banco de dados.

A biblioteca Be587 foi analisada em busca de possíveis ORFs, procurando as melhores seqüências para seleção de clones que foram utilizados na montagem do microarranjo Be587. Durante este processo foi utilizado o programa “OC Identifier tool” (CANTÃO *et al.*, 2007). O consenso das seqüências foi comparado também com outros genomas de rizóbios já seqüenciados e depositados no rhizobase (<http://bacteria.kazusa.or.jp/rhizobase/>).

## **3.2 Construção dos Microarranjos de DNA**

### **3.2.1. Construção da Sonda (*Chip*) de *Bradyrhizobium elkanii* SEMIA 587**

#### **3.2.1.1. Amplificação do DNA plasmidial**

A reação de PCR foi realizada em um volume total de 20µL contendo tampão (1x), MgCl<sub>2</sub> (1,5mM), dNTPs (200µM), Taq DNA polimerase (1U), oligonucleotídeo iniciador (*primer* universal) M13-pUC #1201-R e #1211-F (5 pmoles de cada), DNA plasmidial (25ng) e água esterilizada para completar o volume. As reações foram realizadas em termociclador sendo utilizados os seguintes passos: 1 ciclo de desnaturação a 96°C por 1 minuto, 40 ciclos (20 segundos a 96°C, 30 segundos a 50°C e 4 minutos a 72°C) e 1 ciclo final a 72°C por 5 minutos. As amostras amplificadas foram mantidas a 4°C.

### 3.2.2. Impressão robótica dos produtos de PCR

Os produtos purificados (15 $\mu$ L) foram adicionados em placas (96 *wells*) contendo 15 $\mu$ L de DMSO (50%). A impressão dos microarranjos foi feita em lâminas de vidro tratadas com aminosilanos por um equipamento robotizado de alta precisão (GMS 417 *Arrayer*, Affymetrix/Genetic MicroSystems). Os pontos nas lâminas ou *spots* foram impressos a uma distância determinada (250 $\mu$ m) um do outro na forma de *array*, em duplicata.

### 3.2.3 Fixação dos DNAs Impressos nas lâminas

Os DNAs impressos foram reidratados com vapor d'água a 75°C por cerca de 5 segundos (as lâminas foram mantidas suspensas, com a face contendo DNA impresso em contato com o vapor) e incubadas por 1 minuto a 70°C. Em seguida, o DNA foi fixado em uma câmara de UV *cross-linker* regulada para 1300x100 $\mu$ J/cm<sup>2</sup>. As lâminas foram novamente incubadas a 70°C por 2 horas e mantidas em dessecador até o momento da hibridização.

### 3.2.4. Cultivo das bactérias

As bactérias *Azorhizobium caulidonans* LGM 6465, *Bradyrhizobium elkanii* Semia 587, *Bradyrhizobium japonicum* LGM 6138, *Mesorhizobium huakuii* LGM 14107, *Rhizobium leguminosarum* *bv.* *trifolii* LGM 8820, *Sinorhizobium meliloti* LGM 6133 utilizadas no trabalho foram cultivadas em meio sólido de extrato de levedura-manitol (YMA, pH 6,8) (Vincent, 1970). Posteriormente foram cultivadas em erlenmeyers contendo meio TY (triptona 5g/L, extrato de levedura 3g/L e CaCl<sub>2</sub>.2H<sub>2</sub>O 0,87g/L), (Beringer, 1974). As células bacterianas foram mantidas por um período de 72 horas, a 28°C, sob agitação (148 rpm), após este período foram submetidas à centrifugação (12000xg, 30 minutos a 4°C). O sedimentado foi lavado com solução salina (NaCl 0,85%) para remoção de polissacarídeos extracelulares e

de possíveis resíduos do meio de cultura. Após a centrifugação e lavagem as células bacterianas foram usadas para extração de DNA.

### **3.2.5. Extração e purificação do DNA genômico**

As células cultivadas como descrito no item anterior foram ressuspensas em 1mL de solução salina. As suspensões foram transferidas para tubos *ependorf* e centrifugadas a 15.600 x g por 5 minutos. Esta operação foi repetida mais uma vez para a total remoção de resíduos do meio de cultivo.

Os sedimentos foram ressuspensos em 1mL de solução NaCl-EDTA (0,15M; 0,01M) e incubados a 37°C por 10 minutos, em tubos Corex. Após esse período, adicionaram-se 500µL de lisozima (5mg/mL) preparada em tampão Tris-EDTA-dextrose (10mM; 24mM; 50mM) e levados a uma nova incubação a 37°C por 30 minutos. Posteriormente, adicionaram-se 500µL de solução de SDS (duodecil sulfato de sódio) a 20%, e novamente realizou-se uma outra incubação a 56°C durante 20 minutos.

Decorrido o tempo de incubação, foram adicionados à suspensão 500µL de solução de perclorato de sódio 5M, gota a gota com agitação suave. Em seguida, adicionaram-se 2mL de solução de fenol-clorofórmio-álcool isoamílico (25:24:1), deixando-se por 1 hora em agitação orbital a 220 rpm em banho de gelo. Após esse período de incubação, o material foi centrifugado a 12.000xg, por 20 minutos, a 4°C. Coletou-se então a fase superior com auxílio de uma pipeta *Pasteur*, transferindo-a para outros tubos de ensaio limpos. À fase recuperada adicionaram-se 2mL de solução clorofórmio-álcool isoamílico (24:1) procedendo-se uma agitação orbital por 20 minutos a 220 rpm, novamente em banho de gelo.

O material foi centrifugado a 5.860xg por 20 minutos a 4°C, e após esse passo, realizou-se a coleta da fase superior, a qual foi novamente transferida para tubos limpos. Mais uma vez, 2mL de clorofórmio-álcool isoamílico foram

adicionados, repetindo o procedimento anterior, até a recuperação e transferência da fase superior para tubos limpos onde foram adicionados 2 volumes de etanol 95% (gelado) seguido de uma agitação manual suave e incubação a -20°C por 12 horas para a precipitação do DNA. Após as 12 horas, centrifugou-se o material a 5.860xg por 20 minutos a 4°C, descartando-se então o sobrenadante. Após centrifugação o precipitado foi lavado por duas vezes com 1mL de etanol 70% (gelado) de maneira rápida e cuidadosa para evitar a perda do DNA. Os tubos foram então vertidos sobre papel absorvente, e depois uma espera de 30 a 40 minutos em temperatura ambiente para eliminar o excesso de etanol. O DNA foi transferido para um microtubo plástico através da adição de 1 mL de tampão TE (Tris-HCl 10mM; EDTA 1mM) e estocado a -20 °C, posteriormente foi quantificado em espectrofotômetro a uma absorbância de 260 nm.

### **3.2.6 Biochip de DNA de *Bradyrhizobium elkanii* SEMIA 587**

Os fragmentos de DNAs obtidos por PCR das sequências contendo 2654 genes selecionados foram dispostos em 33 placas estéreis de poliestireno (96-*well assay plate*, 250µL) constituindo o estoque de DNA dos clones, mantidos a -20°C até o momento da amplificação. Este material foi submetido à eletroforese e esse estoque de DNA foi utilizado na PCR para a obtenção e análise dos fragmentos. O material amplificado pela PCR foi impresso (*spotted*) nas lâminas para a fabricação do microarranjo de DNA de *B. elkanii*.

Para a construção do microarranjo em lâminas todo o material foi impresso com uma replica e distância de 250µm entre cada *spot* (espaçamento entre o centro do primeiro *spot* e o centro do quarto *spot*) e a geometria usada foi desenhada utilizando-se um esquema de placas 4x4. Os microarranjos foram impressos em lâminas cobertas com aminosilano comercializadas pela *Corning* (Cat. Nº 40004).

Para a obtenção do DNA marcado utilizou-se a metodologia proposta por BOWTELL & SAMBROOK (2003) modificada por DALL ACQUA 2004, onde se

realizou a quebra do DNA genômico a ser marcado resultando em fragmentos de 500 a 1000pb.

### **3.2.8. Síntese de DNAs fluorescentemente marcados para hibridização (Bowtell e Sambrook 2003-modificado)**

O DNA extraído das estirpes bacterianas foi fragmentado por sonicação (*Ultrassom Branson Sonifier 250*) com pulsos de 10 segundos, 60W, por 1 minuto, (em banho de gelo) a fim de obter-se fragmentos de 500 a 1000pb. O DNA fragmentado de cada estirpe (1,5µg para cada amostra) foi incubado a 95°C por 5 minutos, sendo então resfriado imediatamente no gelo. Adicionou-se a solução do kit *Random Primers DNA Labeling System* (Invitrogen®), contendo cada um dos nucleotídeos [2µL de cada dATP, dGTP, dCTP; 1µL de TTP e 15µL de *tampão random primers*] juntamente com 4µL de Cy3-dUTP (Amersham Pharmacia) ou Cy5-dUTP (Amersham Pharmacia), 1U de *Taq* DNA polimerase e água destilada q.s.p. 25µL. Após a adição de cada um dos reagentes da reação as amostras foram incubadas por 2 horas, a 28°C para que a reação de incorporação dos fluoróforos ocorresse. Os fragmentos de DNA marcados foram coletados e lavados por centrifugação em filtros Microcon YM-30 (Amicon Millipor Bioseparations) (DALL ACQUA 2004). Neste experimento todas as amostras de *B. elkanii* foram marcadas com o fluoróforo Cy-3, todas demais estirpes foram marcadas com o fluoróforo Cy-5.

### **3.2.9. Hibridizações.**

A hibridização se deu no equipamento GeneTac Hybridization (Genetic Microsystems), na qual as lâminas de vidro contendo os microarranjos com DNA do *B. elkanii* foram acopladas.

Os DNAs marcados das estirpes de rizóbios em estudo foram misturados a 37,5 µl de solução de hibridização, contendo 8 µl de líquido bloqueador RPN3601 (Amershan), 19 µl de SSC [20x] e 5,5 µl de 2% SDS, e pré-desnaturados a 95°C por 2 min.

Cada mistura de DNA foi distribuída sobre a lâmina e hibridizada a 42°C durante 12 horas. Após hibridização, a lâmina foi lavada automaticamente e sequencialmente em 2x SSC/0,5% SDS, 0,5x SSC e 0,05x SSC, a 25°C. Cada lavagem correspondeu a um período de 15 min, com 10 seg de fluxo e 20 seg de incubação, durante 10 ciclos. O período de secagem da lâmina foi de 15 min..Todos procedimentos foram realizados no equipamento GeneTac Hybridization (Genetic Microsystems).

Foram utilizadas três lâminas para cada experimento, ou seja, foram realizados em triplicata as extrações de DNA, síntese de DNAs marcados e hibridizações Onde cada hibridização comparou sempre o DNA de *B. elkanii* (Cy3) contra todas estirpes de rizóbios em estudo (Cy5).

### **3.3. Análise dos dados.**

Os sinais fluorescentes foram obtidos pelo GMS-418 Arrayer Scanner (Affymetrix). As leituras foram feitas sob diferentes comprimentos de onda para captar a excitação dos marcadores fluorescentes contidos nos DNAs: 532 nm (Cy3) e 635 nm (Cy5).

As imagens TIFF resultantes para cada fluoróforo, Cy3 ou Cy5, foram reunidas em uma sobreposição e analisadas, determinando-se a densidade do pixel

(intensidade) para cada spot no microarranjo, utilizando-se o software ImaGene 5.5 (Biodiscovery). Uma grade de círculos independentes, correspondentes a cada spot de DNA nos arranjos, foi desenhada sobre a imagem para designar cada spot a ser quantificado. A quantificação foi calculada pela mediana da intensidade de todos os pixels referentes ao sinal de hibridização de cada spot. Os pixels classificados como background foram automaticamente subtraídos pelo software. A localização e identificação de cada gene no arranjo foram definidas em um arquivo texto criado com auxílio do programa CloneTracker 2 (Biodiscovery).

Para ajustar as diferenças sistemáticas na intensidade relativa de cada canal, os dados quantificados foram exportados e transformados pelo software GeneSight 5.5 (Biodiscovery). A utilização de métodos de suavização de diagramas de dispersão tornou-se a abordagem mais eficaz na normalização e correção das intensidades pré-análises estatísticas (QUACKENBUSH, 2002). Este processo, definido como normalização, foi aplicado seguindo os parâmetros de correção lowess (locally weighted linear regression), com um método de normalização local (o algoritmo é aplicado em subsets físicos dos dados, isto é um subgrid) (Quackenbush, 2002).

Os DNAs de *Azorhizobium caulidonans*, *Bradyrhizobium japonicum*, *Mesorhizobium huakuii*, *Rhizobium leguminosarum* *bv. Trifolii* e *Sinorhizobium meliloti*, foram comparados ao genoma parcialmente seqüenciado de *Bradyrhizobium elkanii* contido no microarranjo de DNA. Essa comparação foi feita através da determinação da razão ( $r$ ) da mediana das intensidades de cada ORF nos spots, nas triplicatas de microarranjos. As razões representam o nível relativo de DNAs hibridizados para cada gene de *B. elkanii*, em cada comparação de genoma. As razões (Cy5/Cy3) foram calculadas de tal forma que  $\text{Log } r$  (base 2), do valor absoluto das hibridizações, apresentou-se negativos para intensidades que foram superiores em *B. elkanii*, ou nulos para genes que estavam em ambos genomas com igual intensidade. Em contrapartida, os valores mostraram-se positivos para intensidades superiores nas estirpes comparadas, ou seja, genes com menor número de cópias em *B. elkanii*. Baseado nas observações de

TABOADA *et al.*, (2004) na qual genes foram considerados conservados se o sinal diferenciou por menos do que um fator de dois ou seja, Log da razão variando entre -1.0 e +1.0.

### **3.4 Agrupamento funcional dos resultados obtidos.**

Os genes foram distribuídos através de diferentes grupos funcionais, de acordo com suas funções celulares. Para conhecer as funções de cada gene nas bactérias estudadas foi realizada uma anotação funcional automática para cada comparação de genoma realizada. Um conjunto de “scripts perl” (desenvolvidos pelo bioinformata Maurício E. Cantão) e o programa BlastX foi utilizado na anotação funcional dos genes.

As seqüências de nucleotídeos de *B. elkanii*, correspondentes aos genes que hibridizaram nas lâminas, foram comparadas ao *B. japonicum* (KANECO *et al.*, 2002) pelo programa BlastX, para verificação de similaridade em todas as estirpes estudadas. Posteriormente o relatório gerado pelo BlastX foi filtrado por um script perl. As seqüências que apresentaram alta similaridade foram categorizadas através de outro script, baseando-se na categorização do genoma de *B. japonicum* (KANECO *et al.*, 2002) que foi anotado de acordo com as categorias funcionais de Mônica Riley (1993). Os dados categorizados no presente estudo foram os mesmos relativos ao *B. japonicum* (KANECO *et al.*, 2002) utilizados no banco de dados Rhizobase ([www.kazusa.org.jp/rhizobase](http://www.kazusa.org.jp/rhizobase)). Seqüências categorizadas como hipotéticas passaram por nova comparação de similaridade, usando o programa BlastX, contra o banco NCBI, para avaliar possível mudança na anotação realizada.

### 3.5. Análise Filogenética

Três árvores filogenéticas foram construídas para comparar os diferentes níveis de resolução filogenética quando baseadas em diferentes números de genes, todas utilizaram o método da máxima parcimônia.

A primeira árvore filogenética na figura 3, utilizou apenas seqüências do gene 16S rRNA para cada bactéria analisada. As seqüências de nucleotídeos foram recuperadas do NCBI, alinhadas no programa ClustalX (THOMPSON *et al.*, 1997) e posteriormente utilizadas no programa Mega2 (KUMAR *et al.*, 2001), adotando o método da máxima parcimônia.

A segunda árvore filogenética, (figura 4), foi construída após a verificação de quais genes estavam presentes em todas as bactérias analisadas com as restrições citadas a seguir. Os genes, após os cálculos para retirada de background, realização da normalização e conferência das medianas, tiveram o valor de log da razão transferidos para planilhas, esses valores foram colocados em uma tabela no programa Excel. Os resultados das hibridizações, na forma de log da razão, foram colocados em colunas lado a lado, para todas as bactérias analisadas. As linhas da tabela que não continham valores, ou seja, que não hibridizaram em uma ou mais bactérias, ou apresentavam valores muito negativos, foram excluídas desta análise. Um intervalo de Log da razão entre -1 e 1 foi adotado baseando-se em TABOADA *et al.*, (2004). As seqüências desses genes para cada bactéria analisada foram buscadas no Genbank, concatenadas e finalmente alinhadas. Só então foi construída uma árvore filogenética usando o programa MEGA 2 (KUMAR, *et al.*, 2001) através do método da máxima parcimônia.

Na terceira árvore filogenética, representada na figura 5, 1972 genes que hibridizaram nas três lâminas, após serem analisados nos softwares ImaGene e GeneSight, tiveram o valor de log da razão verificados e transformados em 0 e 1. Valores do log da razão maiores ou iguais a 0,5 foram considerados como 1 indicando presença em ambas bactérias hibridizadas, quando menores que 0,5 foram considerados como ausentes ou 0 (WAN *et al.*, 2007). Uma matriz binária foi

construída com esses valores e analisada no programa Phylip (FELSENSTEIN, 2004), para a construção de uma árvore filogenética utilizando máxima parcimônia.

## IV. RESULTADOS E DISCUSSÃO

### 4.1. Construção das bibliotecas plasmideais Be587, seqüenciamento e seleção de genes para composição do microarranjo de DNA.

O seqüenciamento genômico de *B. elkanii* SEMIA 587 teve início no Laboratório de Bioquímica de Microrganismos e Plantas (LBMP) a partir de 2002 (PRADO, 2004), quando foram contabilizados 9024 clones de bibliotecas de “shotgun”. Posteriormente SOUZA (2006) ampliou as bibliotecas de “shotgun” Be587 em 9471 clones, atingindo um montante de 18495 clones. A partir do seqüenciamento de DNA destes clones foram obtidas 16016 seqüências com tamanho de fragmentos entre 0,6 a 1,6 Kb.

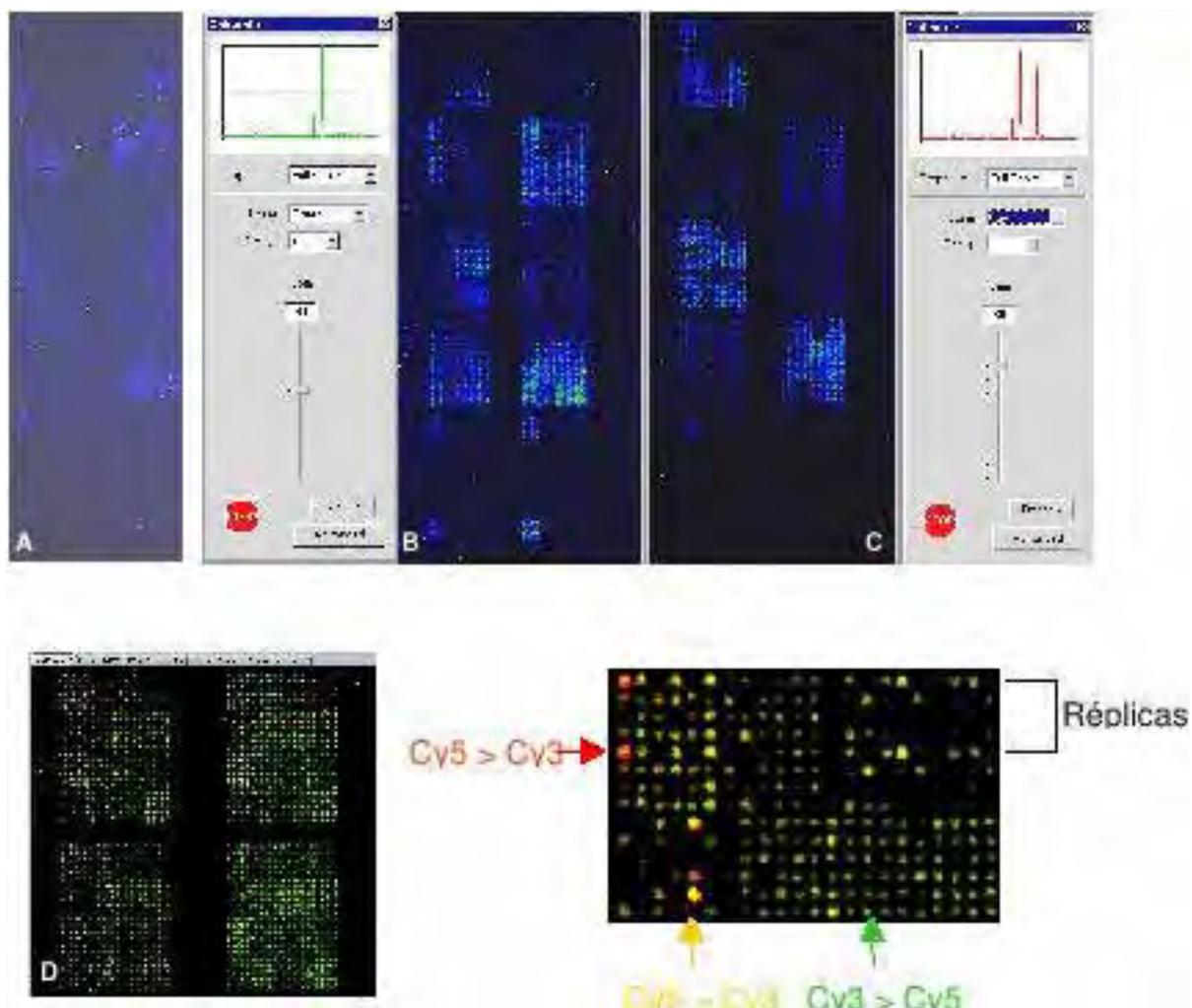
Após análises computacionais 4265 seqüências foram excluídas por apresentarem problemas e 11751 foram utilizadas para montagem do genoma Be587. Dentre as seqüências selecionadas, 10176 apresentaram mais de 400 bases com qualidade phred igual ou superior a 20, equivalente a um percentual maior que 86% do total de seqüências. A montagem genômica produziu 2522 “contigs” compostos por 8043 seqüências. Quando comparado aos 8317 genes de *B. japonicum* (KANEKO *et al.*, 2002), observou-se uma média de 3250 genes identificados em *B. elkanii* SEMIA 587, o que corresponde a 39% de genes identificados em *B. japonicum* USDA 110.

A ferramenta “OC Identifier” selecionou clones contendo ORF única para cada seqüência. Esses clones contendo ORFs únicas foram selecionados para compor o microarranjo de DNA Be587. Esta abordagem de seleção de clones possibilitou a composição do microarranjo de DNA de *B. elkanii* 587, contendo 2654 genes (CANTÃO *et al.*, 2007). A presença de genes relacionados às categorias de diferentes grupos funcionais, incluindo genes para nodulação e fixação do nitrogênio, permite explorar este microarranjo para ensaios diversificados.

## 4.2 Hibridização e análise das imagens de microarranjo de DNA

As imagens observadas na Figura 1 apresentam uma lâmina do microarranjo de DNA Be587, obtidas pelo GMS-418 Arrayer Scanner (Affymetrix), contendo DNA sonda dos genes selecionados que a compõem, e os resultados da hibridização com os alvos de DNA marcados fluorescentemente. O aspecto geral destas imagens ilustra o procedimento para os diferentes experimentos realizados, uma vez que as imagens adquiridas na forma de “false color” são muito semelhantes. A Figura 1A corresponde à lâmina do microarranjo anterior ao processo de hibridização, contendo apenas os produtos amplificados por PCR e os controles impressos. As Figuras 1B e 1C apresentam imagens do microarranjo hibridizado, obtidas nos comprimentos de onda dos canais verde (Cy3) e vermelho (Cy5), respectivamente. A intensidade dos “spots” observados indica a eficiência da hibridização da sonda contida no microarranjo com os alvos de DNA marcados fluorescentemente.

A imagem sobreposta (Figura 1D) foi obtida através da composição realizada pelo programa ImaGene (Biodiscovery). Ressalta-se que a coloração das imagens são falsas, uma vez que o sinal emitido pela fluorescência é decodificado pelo programa de análise e interpretado de forma matemática. Contudo, a sobreposição dos canais produz uma imagem que permite uma rápida visualização qualitativa de um experimento de microarranjo de DNA, revelando informações sobre o balanço de cores, uniformidade da hibridização, uniformidade dos “spots”, “background” e artefatos que possam ocorrer no experimento. Também é possível se ter uma impressão grosseira e superficial, do número de genes que são hibridizados entre duas bactérias, uma vez que a coloração dos “spots” (verde, vermelho ou amarelo) indica em qual bactéria, ou em ambas, um gene está presente (Figura 1).



**Figura 1:** Imagens “false color” dos “spots” que compõem o microarranjo de DNA De *B. elakanii* 587, impresso em lâmina de vidro. (A) DNA sonda sem hibridização; (B) DNA sonda hibridizado escaneado no canal verde (532 nm/Cy3); (C) DNA sonda hibridizado escaneado no canal vermelho (635 nm/Cy5); (D) Imagem de uma composição obtida através da sobreposição das leituras verde/vermelho, realizada pelo programa ImaGene. O detalhe mostra as relações entre os sinais para Cy3 e Cy5.

Após a quantificação dos sinais hibridizados, procede-se a normalização dos dados. Esta técnica matemática minimiza as discrepâncias dos dados de diferentes experimentos devido a variáveis não biológicas tais como preparação de amostras, condições de hibridização, marcação, quantidade de sonda, entre outros. As intensidades da fluorescência vermelha (Cy5) são freqüentemente inferiores às intensidades verdes (Cy3), devido a uma incorporação relativamente menor dos nucleotídeos Cy5. Assim, dado as diferenças observadas entre as intensidades vermelha e verde, causada por fatores não inerentes à resposta biológica, a normalização pode ser definida como um nivelamento matemático dos dados (YANG *et al.*, 2002).

#### **4.3. Comparação de genomas por hibridização**

Para a obtenção dos genes ortólogos entre todas as bactérias no estudo, inicialmente foram realizadas comparações por hibridização dos DNAs das espécies com o genoma parcial *B. elkanii* nas lâminas de vidro em triplicatas.

Os genes foram distribuídos através de diferentes grupos funcionais, de acordo com suas funções celulares, baseado nas categorias funcionais de Mônica Riley (1993), como pode ser visto no anexo 1 do apêndice.

Inicialmente foram realizadas comparações de categorização entre os bancos de dados Rhizobase, GO (ASHBURNER, *et al.*, 2000) e COG (TATUSOV *et al.*, 2004). Embora muito utilizado na atualidade o GO (Gene Ontology) inicialmente foi desenvolvido para auxiliar no processo de anotação de ORFs oriundas de eucariotos, talvez por conta disso ainda não apresente um amplo volume de dados de procariotos, o que pode ter causado o baixo desempenho na categorização automática dos resultados do presente estudo. Embora o COG (Cluster Ortolog Genes) também armazene dados de procariotos, os resultados apresentados foram inferiores ao apresentado pelo Rhizobase. Por esse motivo os dados categorizados no presente estudo foram obtidos através de comparação com o *B. japonicum* (KANECO *et al.*, 2002) depositado no banco de dados Rhizobase

([www.kazusa.org.jp/rhizobase](http://www.kazusa.org.jp/rhizobase)). Esse genoma foi anotado baseado nas categorias funcionais de Mônica Riley (1993).

**Tabela 1 – Número de genes em comum entre o genoma parcial de *B. elkanii* e *B. japonicum* distribuídos por grupos funcionais.**

Categorias	Únicos	Porcentagem correspondente no <i>B. elkanii</i>
1-Aminoácido e biossíntese	62	98
2-Biossíntese de cofatores, grupos prostéticos e carreadores	69	97
3-Envelope celular	38	95
4-Processos celulares	88	96
5-Metabolismo intermediário central	80	96
6-Metabolismo energético	122	96
7-Metabolismo de ácidos graxos, fosfolipídeos e esterol	91	97
8-Purinas, pirimidinas, nucleosídeos e nucleotídeos	33	97
9-Funções regulatórias	121	96
10- Recombinação, reparo e replicação do DNA	50	96
11-Transcrição	19	92
12-Tradução	79	96
13-Proteínas de transporte e ligação	235	94
14-Outras categorias	343	95
15-Hipotéticas	530	96
<b>Total</b>	<b>1960</b>	<b>97</b>

A hibridização realizada entre *B. elkanii* e *B. japonicum* foi a que apresentou maior número de genes em comum, como pode ser observado na tabela 1 acima, com destaque para a categoria hipotética.

Nas lâminas de *B. elkanii* haviam mais de uma cópia do mesmo gene, no entanto, somente os genes únicos, ou seja, aqueles que não apresentaram repetições foram incluídos na tabela 1. A segunda coluna mostra em porcentagem quanto equivale os genes únicos, coluna da esquerda, em relação ao total de genes únicos do *B. elkanii*.

Embora a categoria hipotética tenha apresentado um elevado número de genes em comum entre as bactérias comparadas na tabela 1, ao observar o valor de porcentagem, é possível verificar que a categoria de aminoácidos e biossíntese, foi a que apresentou a maior porcentagem de genes em comum. A distribuição dos genes em subcategorias funcionais pode ser observada no anexo 3 do apêndice.

Dos 2654 genes, incluindo os genes repetidos de *B. elkanii*, 2087 eram diferentes ou únicos. Desses 2087 genes únicos, ou sem repetições do *B. elkanii* presentes na lâmina, 1960 apresentaram-se em comum entre as duas bactérias da tabela 1. No entanto esse elevado número de genes em comum já era esperado, pois ambas pertencem à mesma família *Bradyrhizobiaceae* (GARRITY *et al.*, 2006) confirmando a classificação taxonômica atual como adequada para estas bactérias.

Classificado inicialmente por SKERMAN *et al.*, 1980, como *Rhizobium japonicum*, teve seu nome alterado para *Bradyrhizobium japonicum* por JORDAN (1982). Atualmente é classificado como filo: *Proteobacteria*, classe: *Alpha Proteobacteria*, ordem: *Rhizobiales*, família: *Bradyrhizobiaceae*, ordem: *Bradyrhizobium*, espécie: *Bradyrhizobium japonicum*.

A espécie *Azorhizobium caulinodans* foi determinada por DREYFUS *et al.*, (1988) que observaram características fenotípicas e genotípicas singulares na bactéria que nodula plantas de *Sesbania rostrata*.

Na tabela 2 é possível observar os genes em comum entre *B. elkanii* e *A. caulinodans*, que baseados na técnica de CGH microarray, apresentaram elevado número de genes em comum na maioria das categorias funcionais. Entre as bactérias do estudo, *A. caulinodans* apresentou um elevado número de genes em comum na maioria das categorias funcionais, no entanto o valor total foi o menor de todas as bactérias comparadas.

Uma das possíveis explicações para esse resultado é que esta espécie é a única entre as bactérias comparadas a nodular o caule de diferentes espécies de *Sesbania*. No anexo 4 do apêndice é possível observar a distribuição dos genes em subcategorias funcionais.

**Tabela 2 - Número de genes em comum entre o genoma parcial de *B. elkanii* e *Azorhizobium caulinodans* distribuídos por grupos funcionais.**

Categorias	Únicas	Porcentagem correspondente no <i>B. elkanii</i>
1-Aminoácido e biossíntese	68	99
2-Biossíntese de cofatores, grupos prostéticos e carreadores	74	99
3-Envelope celular	40	98
4-Processos celulares	91	99
5-Metabolismo intermediário central	85	92
6-Metabolismo energético	124	98
7-Metabolismo de ácidos graxos, fosfolipídeos e esterol	92	98
8-Purinas, pirimidinas, nucleosídeos e nucleotídeos	33	92
9-Funções regulatórias	126	99
10- Recombinação, reparo e replicação do DNA	53	98
11-Transcrição	21	47
12-Tradução	86	99
13-Proteínas de transporte e ligação	245	99
14-Outras categorias	212	60
15-Hipotéticas	417	76
<b>Total</b>	<b>1767</b>	<b>85</b>

Atualmente a bactéria *Azorhizobium caulinodans* encontra-se classificada sistematicamente, de acordo com GARRITY (2005), no filo: *Proteobacteria*, Classe: *Alpha Proteobacteria*, Ordem *Rhizobiales*, Família: *Hyphomicrobiaceae*, Gênero: *Azorhizobium*, espécie: *Azorhizobium caulinodans*.

Batizado em homenagem a um professor chinês de microbiologia do solo, pioneiro na investigação de microssimbiontes isolados da planta *Astragalus sinicus*, a bactéria *Mesorhizobium huakuii* foi inicialmente classificada como *Rhizobium huakuii* (CHEN *et al.*, 1991). Posteriormente JARVIS *et al* (1997) a classificaram como *Mesorhizobium huakuii*. Atualmente de acordo com GARRITY (2005) essa bactéria encontra-se classificada como filo: *Proteobacteria*, classe: *Alpha*

*Proteobacteria*, ordem: *Rhizobiales*, família: *Phyllobacteriaceae*, gênero: *Mesorhizobium*, espécie: *Mesorhizobium huakuii*.

Uma característica interessante dessa bactéria é que ela apresenta pouca capacidade, ou na maioria das vezes é incapaz, de nodular outra leguminosa, que não seja a *Astragalus sinicus*, planta utilizada para adubação verde na rotação com a cultura do arroz em campos da China, Coréia e Japão.

Na tabela 3 é possível observar os genes em comum entre *B. elkanii* e *M. huakuii* quando comparados pela técnica de CGH microarray, o anexo 5 do apêndice revela a distribuição dos genes em subcategorias funcionais.

**Tabela 3 - Número de genes em comum entre o genoma parcial de *B. elkanii* e *Mesorhizobium huakuii* distribuídos por grupos funcionais.**

Categorias	Únicos	Porcentagem correspondente no <i>B. elkanii</i>
1-Aminoácido e biossíntese	67	97
2-Biossíntese de cofatores, grupos prostéticos e carreadores	73	97
3-Envelope celular	40	98
4-Processos celulares	89	97
5-Metabolismo intermediário central	85	92
6-Metabolismo energético	119	94
7-Metabolismo de ácidos graxos, fosfolipídeos e esterois	89	95
8-Purinas, pirimidinas, nucleosídeos e nucleotídeos	33	92
9-Funções regulatórias	126	99
10- Recombinação, reparo e replicação do DNA	51	94
11-Transcrição	22	49
12-Tradução	83	95
13-Proteínas de transporte e ligação	240	97
14-Outras categorias	268	76
15-Hipotéticas	461	84
Total	1846	88

A tabela 3 apresenta um número intermediário de genes em comum entre *B. elkanii* e *M. huakuii*, sendo maiores do que os apresentados nas tabelas 2 e 5, porém menores do que nas tabelas 1 e 4.

Após a comparação de *B. japonicum*, a comparação que apresentou maior número de genes em comum com *B. elkanii* foi *Rhizobium leguminosarum*, como pode ser visto ao comparar os valores das tabelas 4 e 1. A distribuição de genes em subcategorias funcionais pode ser observada no anexo 6 do apêndice. Atualmente é classificado como filo: *Proteobacteria*, classe: *Alphaproteobacteria*, ordem: *Rhizobiales*, família: *Rhizobiaceae* gênero: *Rhizobium*, espécie: *Rhizobium leguminosarum* (GARRITY 2005).

**Tabela 4 - Número de genes em comum entre o genoma parcial de *B. elkanii* e *Rhizobium leguminosarum* bv. *Trifolii* distribuídos por grupos funcionais.**

Categorias	Únicos	Porcentagem correspondente no <i>B. elkanii</i>
1-Aminoácido e biossíntese	65	94
2-Biossíntese de cofatores, grupos prostéticos e carreadores	72	96
3-Envelope celular	39	95
4-Processos celulares	85	92
5-Metabolismo intermediário central	91	99
6-Metabolismo energético	107	85
7-Metabolismo de ácidos graxos, fosfolipídeos e esterois	88	94
8-Purinas, pirimidinas, nucleosídeos e nucleotídeos	35	97
9- Funções de Regulação	123	97
10- Recombinação, reparo e replicação do DNA	50	93
11-Transcrição	44	98
12-Tradução	54	62
13-Proteínas de transporte e ligação	231	93
14-Outras categorias	329	93
15-Hipotéticas	510	93
<b>Total</b>	<b>1918</b>	<b>92</b>

Esta espécie é subdividida ainda em três biovars : *phaseoli*, *trifolii* e *viciae* (GARRITY 2005) baseados na sua especificidade de nodulação em plantas como feijão, trevo e ervilhaca respectivamente.

A espécie *Sinorhizobium meliloti*, originalmente classificada como *Rhizobium meliloti* teve a classificação atual proposta por DE LAJUDIE *et al.*, (1994). A classificação atual desta bactéria é filo: *Proteobacteria*, classe: *Alphaproteobacteria*, ordem: *Rhizobiales*, família: *Rhizobiaceae*, gênero: *Sinorhizobium*, espécie: *Sinorhizobium meliloti* (GARRITY 2005). Entre todas as bactérias comparadas a *B. elkanii* foi a que apresentou menor número de genes em comum como pode ser observado na tabela 5. No anexo 7 do apêndice também é possível verificar a distribuição dos genes em comum distribuídos em subcategorias.

**Tabela 5 - Número de genes em comum entre o genoma parcial de *B. elkanii* e *Sinorhizobium meliloti* distribuídos por grupos funcionais**

Categorias	Únicos	Porcentagem correspondente no <i>B. elkanii</i>
1-Aminoácido e biossíntese	61	88
2-Biossíntese de cofatores, grupos prostéticos e carreadores	59	79
3-Envelope celular	36	88
4-Processos celulares	82	89
5-Metabolismo intermediário central	80	87
6-Metabolismo energético	112	89
7-Metabolismo de ácidos graxos, fosfolipídeos e esterois	85	90
8-Purinas, pirimidinas, nucleosídeos e nucleotídeos	29	81
9-Funções regulatórias	109	86
10- Recombinação, reparo e replicação do DNA	51	94
11-Transcrição	18	40
12-Tradução	77	89
13-Proteínas de transporte e ligação	228	92
14-Outras categorias	319	90
15-Hipotéticas	487	89
<b>Total</b>	<b>1833</b>	<b>88</b>

Esta bactéria que teve seu genoma completamente seqüenciado por GALIBERT et al 2001, é capaz de nodular alfafa (*Medicago sativa*), soja, além de plantas dos gêneros *Melilotus* e *Trigonella* (GARRITY 2005).

A espécie *Bradyrhizobium elkanii* foi determinada por KUYKENDAL et al., (1992). Esta espécie de bactéria, assim como *B. japonicum*, normalmente induz a formação de nódulos em espécies de plantas de soja (*Glycine*) e siratro (*Macroptilium atropurpureum*). Contudo, *B. elkanii* também induz nódulos em plantas do fenótipo não nodulante de soja *rjlrj1*, “black-eyed peas” (*Vigna*) e “mungbean”. RUMJANEK et al., (1993) classificou as estirpes brasileiras 29W (SEMIA 5019) e 587 (SEMIA 587) como pertencentes à espécie *Bradyrhizobium elkanii* baseado na similaridade de sequências 16S rDNA e reatividade de anticorpos contra sorogrupos de estirpes internacionais, USDA76 e USDA 31.

Estas linhagens de *B. elkanii* brasileiras atualmente levam o prefixo da Seção de Micobiologia Agrícola (SEMIA) e fazem parte da coleção de rizóbios da FEPAGRO. Assim, a classificação sistemática atual da estirpe SEMIA 587, utilizada neste estudo, com a inclusão da nova família *Bradyrhizobiaceae* (GARRITY et al., 2006), pode ser observada abaixo: Filo: *Proteobacteria*; Classe: *Alpha Proteobacteria*; Ordem: *Rhizobiales*; Família: *Bradyrhizobiaceae*; Gênero: *Bradyrhizobium* Espécie: *Bradyrhizobium elkanii*.

**Tabela 6 - Número de genes únicos de *B. elkanii* na lâmina e as porcentagens relativas a todas bactérias comparadas distribuídos por grupos funcionais.**

Categories	B. <i>elkanii</i> únicos	B. <i>japonicum</i> %	A. <i>caulinodans</i> %	M. <i>huakuii</i> %	R. <i>leguminosarum</i> %	S. <i>melioloti</i> %
1-Aminoácido e biossíntese	69	90	99	97	94	88
2-Biossíntese de cofatores, grupos prostéticos e carreadores	75	92	99	97	96	79
3-Envelope celular	41	93	98	98	95	88
4-Processos celulares	92	96	99	97	92	89
5-Metabolismo intermediário central	92	87	92	92	99	87
6-Metabolismo energético	126	97	98	94	85	89
7-Metabolismo de ácidos graxos, fosfolipídeos e esterol	94	97	98	95	94	90
8-Purinas, pirimidinas, nucleosídeos e nucleotídeos	36	92	92	92	97	81
9-Funções regulatórias	127	95	99	99	97	86
10- Recombinação, reparo e replicação do DNA	54	93	98	94	93	94
11-Transcrição	45	42	47	49	98	40
12-Tradução	87	91	99	95	62	89
13-Proteínas de transporte e ligação	248	95	99	97	93	92
14-Outras categorias	354	97	60	76	93	90
15-Hipotéticas	547	97	76	84	93	89
Total	2087	94	85	88	92	88

Na tabela 6 é possível observar que o *B. japonicum* é a bactéria que apresenta a maior porcentagem total de genes em comum com *B. elkanii*, especialmente devido as categorias: Outras categorias e a Hipotéticas, que foram maiores do que todas as bactérias comparadas. Também apresentou uma distribuição razoavelmente uniforme de genes em comum com *B. elkanii* entre todas as categorias, com exceção feita à categoria transcrição.

Outro dado interessante é que *A. caulinodans* apresentou um grande número de genes em comum na maioria das categorias, superando até mesmo o *B. japonicum*. Entretanto as categorias: Outras categorias, Hipotéticas e transcrição, foram suficientemente diferentes para fazer com que esta bactéria apresentasse a menor porcentagem total de genes em comum com *B. elkanii*.

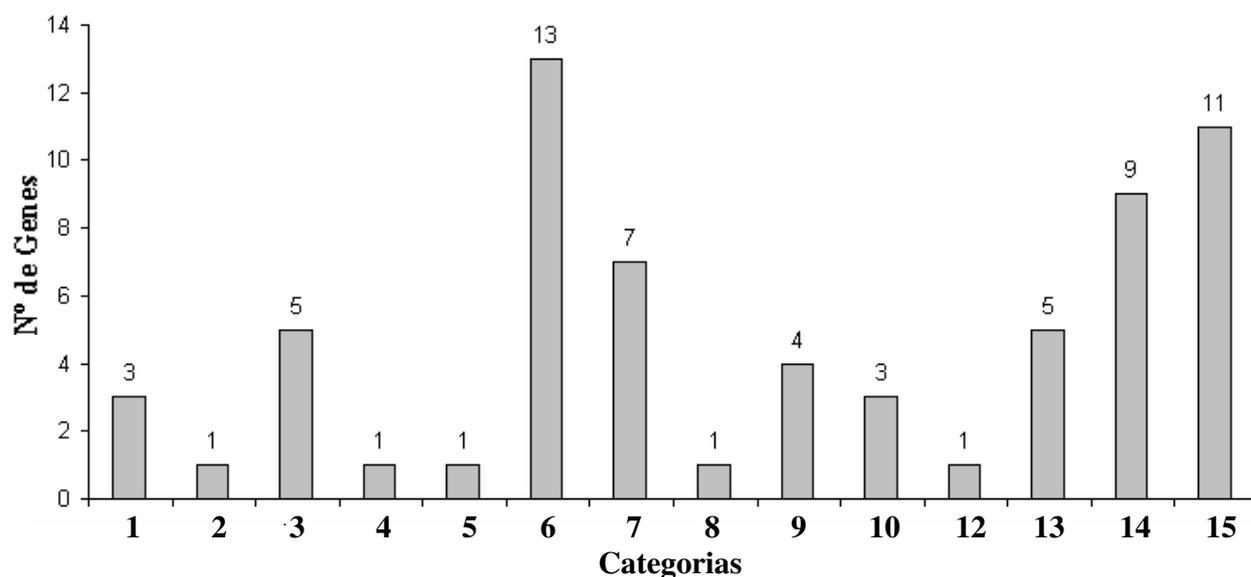
Após *B. japonicum* a bactéria que apresentou maior porcentagem total de genes em comum com o *B. elkanii* foi o *R. leguminosarum bv Trifolii*. Isso poderia justificar o porque das bactérias pertencentes a família *Bradyrhizobiaceae* terem sido classificadas como rizóbios no início da taxonomia bacteriana.

Contudo é importante ressaltar que o genoma do *B. elkanii* utilizado no estudo ainda encontra-se incompleto, por isso os valores dessa comparação poderão sofrer alterações quando essa análise puder ser feita com o genoma completo.

Todas as comparações citadas anteriormente tiveram como objetivo fornecer dados para a realização da busca dos genes que estavam em comum entre todas as bactérias.

Através dessas análises foi possível observar sessenta e cinco seqüências de genes em comum entre todas as bactérias analisadas, (figura 2), a distribuição desses genes em subcategorias funcionais, assim como a identificação de cada gene pode ser observada no anexo 2 do apêndice.

Posteriormente as seqüências de *B. elkanii* foram identificadas, após serem comparadas as seqüências de proteínas de *B. japonicum* no banco de dados Rhizobase e posteriormente as armazenadas no banco de dados de proteínas do NCBI.



**Figura 2.** Número de genes em comum entre o genoma parcial de *B. elkanii* e *Sinorhizobium meliloti*, *M. huakuii*, *R. leguminosarum* bv. *Trifolii*, *B. japonicum*, *A. caulinodans* distribuídos por categorias funcionais dentro do intervalo do log da razão variando entre -1 e 1.

1-Aminoácido e biossíntese	9-Funções regulatórias
2-Biossíntese de cofatores, grupos prostéticos e carreadores	10- Recombinação, reparo e replicação do DNA
3-Envelope celular	12-Tradução
4-Processos celulares	13-Proteínas de transporte e ligação
5-Metabolismo intermediário central	14-Outras categorias
6-Metabolismo energético	15-Hipotéticas
7-Metabolismo de ácidos graxos, fosfolipídeos e esteroi	
8-Purinas, pirimidinas, nucleosídeos e nucleotídeos	

Inicialmente foi observado através de anotação automática que a maioria dos genes em comum pertencia à categoria “Hipotéticos”. Essa categoria sempre apresentou o maior número de genes categorizados como pode ser observado nas tabelas 1, 2, 3, 4 e 5. Nos genomas seqüenciados de *B. japonicum* (KANECO *et al.*, 2002), *S. meliloti* ( GALIBERT *et al.*, 2001) e *R. leguminosrum* (YOUNG *et al.*, 2006) também foi observado um elevado número de genes hipotéticos ou desconhecidos.

Para reduzir o número de genes hipotéticos, foi realizada uma anotação manual dos genes pertencentes a esta categoria. Como resultado a categoria “Metabolismo Energético” passou a apresentar maior número de genes em comum entre todas as categorias analisadas, ganhando dois genes, embora o número de genes hipotéticos não tenha sofrido grande redução, como pode ser observado na figura 2.

Em trabalho recente SOUZA (2006) observou genes pertencentes a esta categoria sendo mais expressos em *B. elkanii* quando cultivados em meio definido para rizóbio (RDM). Ainda segundo o autor isto ocorreu porque no período inicial de adaptação ao meio de cultivo (fase lag) as bactérias no meio RDM, com qualidade nutricional mais limitada do que os demais meios, necessitavam de maior número de genes para sua adaptação.

Foram encontrados genes para todas as categorias funcionais, exceto transcrição. Contudo cabe ressaltar que o genoma do *B. elkanii* no presente estudo encontra-se parcialmente seqüenciado, constando na lâmina de microarray, apenas 2654 genes. Portanto as proporções de genes observados em comum entre todas as bactérias analisadas poderão ser diferentes em uma futura análise que disponha do genoma completo.

#### 4.4. Análise Filogenética

Três árvores filogenéticas com todas as bactérias presentes no estudo foram construídas para comparar diferentes níveis de resolução filogenética quando baseadas em diferentes números de genes. Para todas as árvores foi utilizado o método da máxima parcimônia, pois com exceção da árvore baseada somente na seqüência do gene 16S rRNA, o número de sítios analisados nas demais árvores foi muito grande.

Na primeira árvore filogenética, representada pela Figura 3 foi utilizado apenas seqüências do gene 16S rRNA recuperadas do Genbank (NCBI) para cada bactéria presente no estudo. O seqüenciamento do gene 16S rRNA tem grande participação na maneira como as bactérias são classificadas, já que a taxonomia polifásica, método mais aceito na atualidade, utiliza-se dele, juntamente com a hibridização de DNA e análise dos dados fenotípicos.

A seqüência do gene 16S rRNA foi suficiente para separar as seis bactérias do presente estudo, onde dois grupos foram formados. No primeiro grupo *Rhizobium leguminosarum* e *Sinorhizobium meliloti* estão próximos, seguido pelo *Mesorhizobium huakuii* e mais distantemente o *Azorhizobium caulinodans*. A proximidade observada entre *S. meliloti*, *R. leguminosarum* e *M. huakuii* também foi observada por van BERKUM *et al.*, (2003), mas em árvores construídas com base em seqüências do gene 23S rRNA.

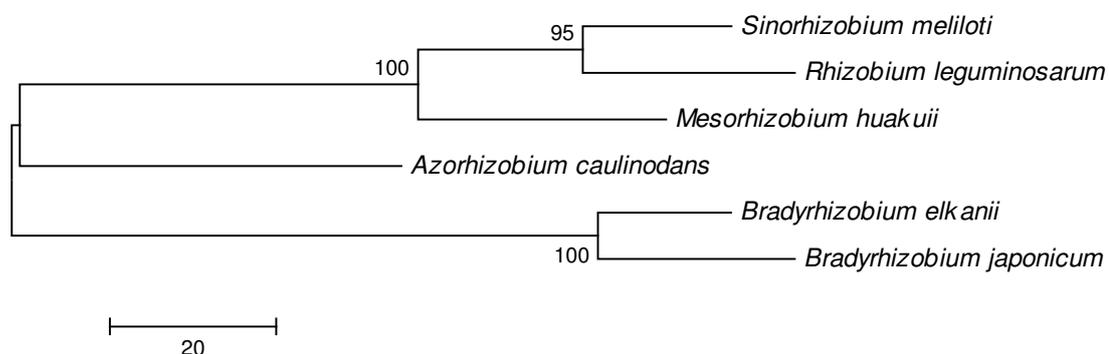


Figura 3 - Árvore filogenética construída usando genes 16S rRNA de cada bactéria. Foi utilizado o programa MEGA2 e o método da máxima parcimônia.

Para investigar o motivo desse agrupamento e das incongruências observadas ao se comparar a árvore construída com base no gene 16S rRNA, uma extensa análise estatística de genes rRNA nas bactérias estudadas foi realizada fornecendo evidências que sugerem uma história de recombinação do gene 16S rRNA (BROUGHTON, 2003) entre espécies de *Mesorhizobium* e *Sinorhizobium* van BERKUM *et al.*, (2003). No segundo grupo as bactérias *B. elkanii* e *B. japonicum* aparecem agrupadas como esperado por pertencerem ao mesmo gênero.

Em outro estudo envolvendo 12 espécies de rizóbios foi observada inconsistência na posição do *Rhizobium galegae* entre árvores filogenéticas obtidas a partir de sequências do gene 16S rRNA e do gene *dnaK*. (EARDLY. *et al.*, 2005). Algumas discrepâncias também foram observadas para membros do gênero *Rhizobium* nos relacionamentos filogenéticos entre isolados coreanos e estirpes de referência dos gêneros *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium*, e *Sinorhizobium* quando analisados através de sequências do gene 16S rRNA e região ITS (KWON *et al.*, 2005). DEWHIRST *et al.*, (2005) também observaram discrepâncias na comparação de árvores filogenéticas obtidas a partir de sequências de 16S rRNA e 23S rRNA ao analisar diferentes espécies de *Heliobacter*.

A segunda árvore filogenética foi construída baseada nas seqüências de dez proteínas que estavam presentes em todas as bactérias analisadas.

Foi buscado no NCBI as seqüências das proteínas correspondentes às 65 seqüências de nucleotídeos que hibridizaram nas lâminas para todas bactérias do estudo. Entretanto a bactéria *Mesorhizobium huakuii*, quando pesquisada no banco de proteínas do NCBI (15/10/2007), revelou somente trinta e três proteínas armazenadas, das quais somente 10 correspondiam ao conjunto das 65 buscadas inicialmente. Esta bactéria assim como o *B. elkanii* ainda não teve seu genoma completamente seqüenciado. Esse baixo número de proteínas armazenadas impediu que uma árvore filogenética contendo as sessenta e cinco seqüências

observadas em todas bactérias fosse construída, reduzindo o número total a apenas dez.

As seqüências das proteínas utilizadas foram: NodA, NodB, N-acetylglucosaminyltransferase, citrate synthase, ATP synthase beta, transposase, glycosyl transferase, glutamine synthetase, enoyl CoA hydratase e Gyrase B.

Já havia sido observado em estudo anterior que o gene *citrate synthase*, apresentava-se comum entre diferentes gêneros de rizóbios, além disso, este gene também poderia ser usado para a construção de árvores filogenéticas (HERNÁNDEZ-LUCAS *et al.*, 2004). De forma similar genes como *NodA*, *NodB* e N-acetylglucosaminyltransferase (*NodC*) já tiveram suas presenças relatadas em diferentes gêneros de rizóbios (KANECO *et al.*, 2002; GALIBERT *et al.*, 2001; YOUNG *et al.*, 2006).

Como na maioria das vezes os genes hipotéticos ou pouco conhecidos apresentam-se em grande concentração nos organismos seqüenciados, COMAS *et al.*, (2007) sugerem que eles apresentam grande importância nos mecanismos de evolução do genoma.

Como é possível observar na figura 4 dois grandes grupos distintos foram identificados, um contendo quatro bactérias e outro apenas duas. No maior grupo *B. japonicum*, *B. elkanii*, com razoável grau de identidade agruparam-se mais proximamente. Esse resultado já era esperado, pois ambas foram classificadas como sendo do mesmo gênero.

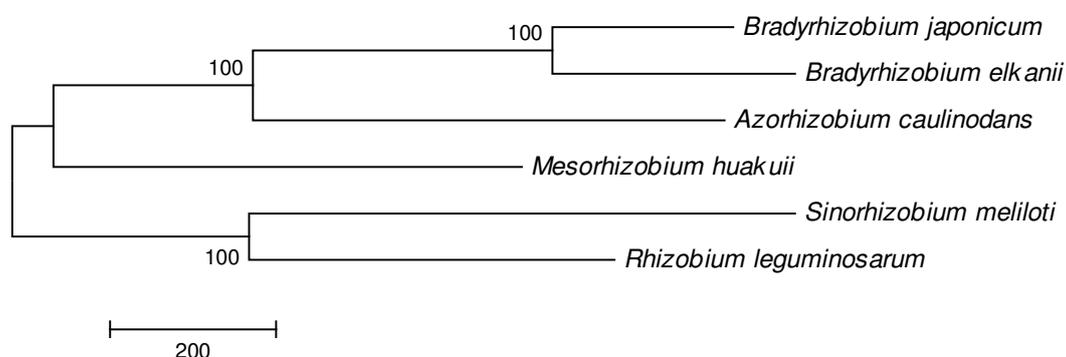


Figura 4 - Árvore filogenética construída utilizando seqüências das proteínas: NodA, NodB, N-acetylglucosaminyltransferase, citrate synthase, ATP synthase beta, transposase, glycosyl transferase, glutamine synthetase, enoyl CoA hydratase, Gyrase B. Foi utilizado o programa MEGA2 e o método da máxima parcimônia. na escala evolutiva.

A seqüência do gene 16S rRNA é muito utilizada na taxonomia de procaríotos, porque esse gene evolui lentamente, além do produto gênico ser essencial à sobrevivência da bactéria e funcionalmente conservado. Entretanto basear a filogenia bacteriana apenas na seqüência desse gene não apresenta resultados adequados. Observou-se que a topologia de árvores obtidas das seqüências do gene 16S rRNA foram significativamente diferentes das árvores construídas com base nas seqüências do gene 23S rRNA e região do espaço transcrito interno (ITS) (van BERKUM *et al.*, 2003).

Já no segundo grupo é possível observar apenas duas bactérias *S. meliloti* e *R. leguminosarum* que embora estejam em um mesmo grupo, apresentam baixo grau de identidade. Mesmo utilizando um conjunto de apenas dez seqüências concatenadas alguns agrupamentos das árvores representadas na figura 5 puderam se repetir na figura 4 embora ROKAS *et al.*, (2003) sugeriram um número de genes maior ou igual a vinte para uma análise genética acurada (DAUBIN *et al.*, 2002; KURTZMAN *et al.*, 2003).

A terceira árvore filogenética, representada na figura 5, baseou-se no total de genes hibridizados na lâmina, exceto aquelas que não hibridizaram nas três lâminas.

A figura 5 exibe uma árvore filogenética com topologia muito diferente das árvores filogenéticas anteriores. Nela é possível observar que embora apresente um número de substituições menor do que na árvore anterior, *B. elkanii* e *B. japonicum* continuam no mesmo grupo, porém dessa vez *A. caulinodans* aparece no mesmo grupo do gênero *Bradyrhizobium*, diferente do que ocorreu na figura 3, mas similar ao relacionamento da figura 4.

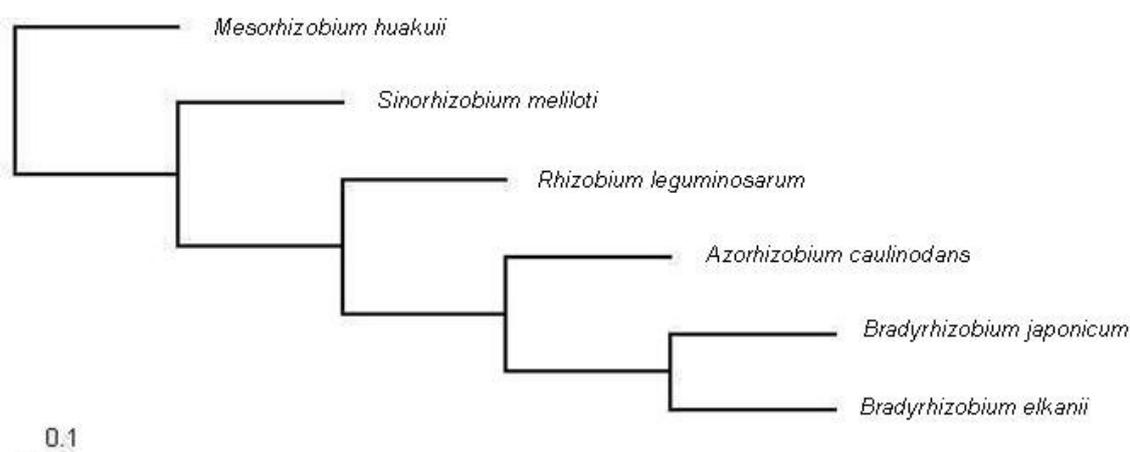


Figura 5 - Árvore filogenética baseada nas informações provenientes de 1972 Orfs das lâminas de microarray. Foram utilizados o programa Phylip e o método da máxima parcimônia.

As espécies *R. leguminosarum*, *S. meliloti* e *M. huakuii* se apresentaram com uma topologia completamente diferente da observada nas figuras 3 e 4. Discrepâncias em árvores filogenéticas baseada em apenas um gene foram

observadas anteriormente (van BERKUM *et al.*, 2003; EARDLY. *et al.*, 2005; KWON *et al.*, 2005; DEWHIRST *et al.*, 2005).

O resultado observado na figura 5 confirma diversas observações anteriores, que indicam que filogenias baseadas em apenas um gene podem ser muito suscetíveis à variação nas árvores filogenéticas construídas. Os resultados reforçam a sugestão de utilizar muitas seqüências na construção de árvores filogenéticas para obter resultados confiáveis (DAUBIN *et al.*, 2002; KURTZMAN *et al.*, 2003, EDWARDS-INGRAN *et al.*, 2004).

## V. CONCLUSÕES

1. O presente trabalho revelou sessenta e cinco genes ortólogos entre os genomas das bactérias *Azorhizobium caulinodans* LGM 6465, *Bradyrhizobium elkanii* SEMIA 587, *Bradyrhizobium japonicum* LGM 6138, *Mesorhizobium huakuii* LGM 14107, *Rhizobium leguminosarum* bv. *Trifolii* LGM 8820 e *Sinorhizobium meliloti* LGM 6133.
2. Os resultados observados no presente trabalho sugerem que quanto maior for o número de genes ortólogos usados na filogenia, mais confiáveis serão as árvores filogenéticas obtidas.
3. A técnica de CGH microarray foi adequada para a busca de genes ortólogos entre os diferentes genomas analisados.

## VI. REFERÊNCIAS

AAKRA, A.; NYQUIST, O. L.; SNIPEN, L.; REIERSEN, T. S.; NES, I. F. Survey of Genomic Diversity among *Enterococcus faecalis* Strains by Microarray-Based Comparative Genomic Hybridization **Applied Environmental Microbiology**, v.73, p.2207-2217, 2007.

ASHBURNER, M.; BALL, C. A.; BLAKE, J.A.; BOTSTEIN, D.; BUTLER, H.; CHERRY, J.M.; DAVIS, A.P.; DOLINSKI, K; DWIGHT, S.S.; EPPIG, J.T.; HARRIS, M.A.; HILL, D.P.; ISSEL-TARVER, L.; KASARSKIS, A.; LEWIS, S.; MATESE, J.C.; RICHARDSON, J E.; RINGWALD, M.; RUBIN, G. M.; SHERLOCK, G. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v.25, p.25 – 29, 2000.

BALDWIN, A.; MAHENTHIRALINGAM, E.; THICKETT, K. M.; HONEYBOURNE, D.; MAIDEN, M. C. J.; GOVAN, J. R.; SPEERT, D. P.; LIPUMA, J. J., VANDAMME, P.; DOWSON, C. G. Multilocus sequence typing scheme that provides both species and strain differentiation for the *Burkholderia cepacia* complex. **Journal Clinic Microbiology**, v.43, p.4665–4673, 2005.

BOONE, D.R.; CASTENHOLZ, R.W. 2001. *Bergey's Manual of Systematic Bacteriology* 2<sup>nd</sup> ed., Volume One, Springer-Verlag, USA.

BOWTELL, D.; SAMBROOK, J. *DNA microarray – A molecular cloning manual*. 1st ed. Cold Spring Harbor Laboratory Press, New York, 712 pp. 2003.

BROUGHTON, W. J. Roses by Other Names: Taxonomy of the *Rhizobiaceae*. **Journal of Bacteriology**, v.185, p.2975-2979, 2003.

CANTÃO M.E., FERREIRA J.E., LEMOS E.G.M. Optimal Clone Identifier for Genomic Shotgun Libraries: "OC Identifier tool". **Genetic Molecular Research**, v.6, p.1, 2007.

CHAN, K.; BAKER, S.; KIM, S.S.; DETWEILER, C.S. *et al.*, Genomic Comparison of *Salmonella enterica* Serovars and *Salmonella bongori* by Use of an *S. enterica* Serovar Typhimurium DNA Microarray. **Journal of Bacteriology**, v. 185, p. 553–563, 2003.

CHEN, W.X., LI, G.S., QI, Y.L., WANG, E.T., YUAN, H.L., AND LI, J.L. "Rhizobium huakuii sp. nov. isolated from the root nodules of *Astragalus sinicus*." **International Journal of Systematic Bacteriology** v. 41, p.275-280, 1991.

CHO, J.C.; TIEDJE, J. M. Bacterial Species Determination from DNA-DNA Hybridization by Using Genome Fragments and DNA Microarrays. **Applied and Environmental Microbiology**, v.67, n<sup>o</sup>8, p.3677-3682, 2001.

CHUNG, Y.J.; JONKERS, J. KITSON, H.; FIEGLER, H. *ET AL.*, A Whole-Genome Mouse BAC Microarray With 1-Mb Resolution for Analysis of DNA Copy Number Changes by Array Comparative Genomic Hybridization. **Genome Research**, v.14, p.188-196, 2004.

COENYE, T.; GEVERS, D.; VAN DE PEER, Y.; VANDAMME, P.; SWINGS, J. Towards a prokaryotic genomic taxonomy. **FEMS Microbiology Review**, v.29, p.147-167, 2005.

COENYE, T.; VANDAMME, P. Use of the genomic signature in bacterial classification and identification. **Systematic Applied Microbiology**, v.27, p.175–185, 2004.

COHAN, F.M. Bacterial species and speciation. **Systematic Biology** v.50, p.513 – 524, 2001.

COHAN, F.M. What are bacterial species? **Annual Review Microbiology** v.56, p.457–487, 2002.

COHAN, F. M. ; PERRY, E. B. A Systematics for Discovering the Fundamental Units of Bacterial Diversity. **Current Biology** v.17, p.373-386, 2007.

COLWELL, R. R. Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and related *Vibrio* species. **Journal Bacteriology**, v.104, p.410-433, 1970a.

COLWELL, R.R. 1970b. Polyphasic taxonomy of bacteria. *In* Culture Collections of Microorganisms, pp. 421-436. H. Iizuka & T. Hasegawa. (eds.) Tokyo, University of Tokyo Press.

COMAS, I.; MOYA, A.; GONZÁLEZ-CANDELAS, F. Phylogenetic signal and functional categories in Proteobacteria genomes. **BMC Evolutionary Biology** 2007, 7(Suppl 1):S7. <http://www.biomedcentral.com/content/pdf/1471-2148-7-S1-S1.pdf>

DAUBIN, V.; GOUY, M.; PERRIERE, G. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. **Genome Research**, v.12, p.1080–1090, 2002.

DAUBIN, V.; MORAN, N.A.; OCHMAN, H. Phylogenetics and the cohesion of bacterial genomes. **Science**, v301, p.829–832, 2003.

DALL' ACQUA, W. R. (2004) **DNA microarrays e hibridização DNA/DNA na caracterização de estirpes de rhizobium**. 58f Dissertação (Mestrado em

Produção Vegetal) Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal.

DE LAJUDIE, P.; WILLEMS, A.; POT, B.; DEWETTINCK, D.; MAESTROJUAN, G.; NEYRA, M.; COLLINS, M.D.; DREYFUS, B.; KERSTERS, K.; GILLIS, M. "Polyphasic taxonomy of rhizobia: emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov." *Int. J. Syst. Bacteriol.* v.44, p.715-733, 1994.

DEWHIRST, F. E.; ZELI, S.; SCIMECA, M. S.; STOKES, L. N.; BOUMENNA, T.; CHEN, T.; PASTER, B. J.; FOX, J.G.; Discordant 16S and 23S rRNA Gene Phylogenies for the Genus *Helicobacter*: Implications for Phylogenetic Inference and Systematics **Journal Bacteriology**, v.187, p. 6106-6118, 2005.

DIJKSHOORN, L.; TOWNER, K. J.; STRUELENS, M.; 2001. New approaches for the generation and analysis of microbial typing data. First edition, Elsevier. Amsterdam.

DONG, YUEMEI, GLASNER, JEREMY D., BLATTNER, FREDERICK R., TRIPLETT, ERIC W. Genomic Interspecies Microarray Hybridization: Rapid Discovery of Three Thousand Genes in the Maize Endophyte, *Klebsiella pneumoniae* 342, by Microarray Hybridization with *Escherichia coli* K-12 Open Reading Frames. **Applied and Environmental Microbiology**, Washington, v. 67, p. 1911–1921, 2001.

DREYFUS, B., GARCIA, J.L., GILLIS, M. "Characterization of *Azorhizobium caulinodans* gen. nov., sp. nov., a stem-nodulating nitrogen-fixing bacterium isolated from *Sesbania rostrata*." **International Journal Systematic Bacteriology** v.38, p.89-98, 1998.

DUTILH, B.E.; VAN NOORT, V.; VAN DER HEIJDEN, R. T. J. M.; BOEKHOUT, T.; SNEL, B.; HUYNEN, M.A. Assessment of phylogenomic and orthology approaches for phylogenetic inference. **Bioinformatics**, v.23, p.815-824, 2007.

<http://bioinformatics.oxfordjournals.org/cgi/content/abstract/btm015v1>

Acessado em 28/10/2007.

EARDLY, B. D.; NOUR, S. M., VAN BERKUM, P., SELANDER, R. K. Rhizobial 16S rRNA and dnaK Genes: Mosaicism and the Uncertain Phylogenetic Placement of *Rhizobium galegae*. **Applied Environmental Microbiology**, v.71, p.1328-1335, 2005.

EDWARDS-INGRAM, L.C.; GENT, M. E.; HOYLE, D.C.; HAYES, A.; STATEVA, L. I.; OLIVER, S.G. Comparative Genomic Hybridization provides new insights into the molecular taxonomy of the *Saccharomyces sensu strictu* complex. **Genome Research**, v.14, p.1043-1051, 2004.

EWING, B.; HILIER, L.D.; WENDL, M.C.; GREEN, P. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. **Genome Research**, v.8, p.175-185, 1998.

FELSENSTEIN, J. PHYLIP (Phylogeny Inference Package) version 3.6.

Department of Genome Sciences, University of Washington, Seattle, 2004.

FUKIYA, SATORU, MIZOGUCHI, HIROSHI, TOBE, TORU, MORI, HIDEO Extensive Genomic Diversity in Pathogenic *Escherichia coli* and *Shigella* Strains Revealed by Comparative Genomic Hybridization Microarray. **Journal Bacteriology**, v.186, p. 3911-3921, 2004.

GALIBERT F; FINAN TM; LONG SR; PUHLER A; ABOLA P; AMPE F; BARLOY-HUBLER F; BARNETT MJ; BECKER A; BOISTARD P; BOTHE G; BOUTRY M; BOWSER L; BUHRMESTER J; CADIEU E; CAPELA D; CHAIN P; COWIE A; DAVIS RW; DREANO S; FEDERSPIEL NA; FISHER RF; GLOUX S; GODRIE T; GOFFEAU A; GOLDING B; GOUZY J; GURJAL M; HERNANDEZ-LUCAS I; HONG A; HUIZAR L; HYMAN RW; JONES T; KAHN D; KAHN ML; KALMAN S; KEATING DH; KISS E; KOMP C; LELAURE V; MASUY D; PALM C; PECK MC; POHL TM; PORTETELLE D; PURNELLE B; RAMSPERGER U; SURZYCKI R; THEBAULT P; VANDENBOL M; VORHOLTER FJ; WEIDNER S; WELLS DH; WONG K; YEY KC; BATUT J. The composite genome of the legume symbiont *Sinorhizobium meliloti*. **Science**. V.293(5530), p.668-72, 2001. <http://www.sciencemag.org/cgi/content/full/293/5530/668>

GARRITY, G.M. (editor) **Bergey's Manual of Systematic Bacteriology**, second edition, vol. 2 (The *Proteobacteria*), part C (The *Alpha*-, *Beta*-, *Delta*-, and *Epsilonproteobacteria*), Springer, New York, 2005.

GARRITY, G.M.; BELL, J.A.; LILBURN, T. *Bradyrhizobiaceae* fam. nov. **International Journal of Systematic and Evolutionary Microbiology**, v. 56, p.1-6, 2006.

GE, HONG, CHUANG, YAO-YU ERIC, ZHAO, SHUPING, TONG, MIN, TSAI, MONG-HSUN, TEMENAK, JOSEPH J., RICHARDS, ALLEN L., CHING, WEI-MEI. Comparative Genomics of *Rickettsia prowazekii* Madrid E and Breinl Strains. **Journal of Bacteriology**, v.186, p.556–565, 2004.

GEVERS, D.; VANDEPOELE, K.; SIMILLION, C.; VAN DE PEER, Y. Gene duplication and biased functional retention of paralogs in bacterial genomes. **Trends Microbiology**, v.12, p.148 –155, 2004.

GEVERS, D. COHAN, F. M.; LAWRENCE, J. G.; SPRATT, B.G.; COENYE, T.; FEIL, E.J.; STACKEBRANDT, E.; VAN DE PEER, Y.; VANDAMME, P.; THOMPSON, F. L.; SWINGS, J. Re-evaluating bacterial species. **Nature Microbiology Reviews** v.3, p.733–739., 2005.

GIUNTINI, E.; MENGONI, A.; FILIPPO, C.; CAVALIERI, D.; AUBIN-HORTH, N.; LANDRY, C.R.; BECKER, A.; BAZZICALUPO, M. Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of *Sinorhizobium meliloti* natural strains. **BMC Genomics** v.6 p.158, 2005. <http://www.biomedcentral.com/1471-2164/6/158>

GOGARTEN J.P.; DOOLITTLE, W.F.; LAWRENCE J.G.; Prokaryotic evolution in light of gene transfer. **Molecular Biology Evolution**, v.19, p.2226– 2238, 2002.

GOODFELLOW, M. 2000. Microbial systematics: background and uses. In: Applied Microbial Systematics Priest, F.G. & Goodfellow, M. (eds), Kluwer Academic Publishers, Dordrecht.

HANAGE, W. P.; KAIJALAINEN, T.; HERVA, E.; SAUKKORIIPI, A.; SYRJANEN,R.; SPRATT,B.G. Using multilocus sequence data to define the pneumococcus. **Journal Bacteriology**, v.187, p.6223–6230, 2005.

HANAGE W.P.; FRASER C.; SPRATT G. Sequences, sequence clusters and bacterial species. **Philosophical Transactions of Royal Society London B** v.361, p.1917–1927, 2006.

HARRIS,J.K.; KELLEY, S.T.; SPIEGELMAN, G.B.; PACE, N.R. The genetic core of the universal ancestor. **Genome Research** 13:407-412, 2003.

HAUSER, F.; LINDEMANN, A.; VUILLEUMIER, S.; PATRIGNANI, A.; SCHLAPBACH, R.; FISCHER, H.M. HENNECKE, H. Design and validation of a partial-genome microarray for transcriptional profiling of the *Bradyrhizobium japonicum* symbiotic gene region. **Molecular Genetics and Genomics** 275: 55–67, 2006.

HERNÁNDEZ-LUCAS, I.; ROGEL-HERNÁNDEZA, M.A.; SEGOVIAB, L.; ROJAS-JIMÉNEZ, K.; MARTÍNEZ-ROMEROA, E. Phylogenetic Relationships of Rhizobia Based on Citrate Synthase Gene Sequences. **Systematic and Applied Microbiology** , v.27, p.703-706, 2004.

HONG,S.H.; KIM,T.Y.; LEE,S.Y. Phylogenetic analysis based on genome-scale metabolic pathway reaction content. **Applied Microbiology Biotechnology**, v.65, p.203 –210, 2004.

HOOPER, S.D.; BERG, O.G. On the nature of gene innovation:duplication patterns in microbial genomes. **Molecular Biology Evolution** v.20, p.945 –954, 2003.

JARVIS, B.D.W.; VAN BERKUM, P.; CHEN, W.X.; NOUR, S.M.; FERNANDEZ, M.P.; CLEYET-MAREL, J.C.; GILLIS, M. "Transfer of *Rhizobium loti*, *Rhizobium huakuii*, *Rhizobium ciceri*, *Rhizobium mediterraneum*, and *Rhizobium tianshanense* to *Mesorhizobium* gen. nov." *Int. J. Syst. Bacteriol.* v.47, p.895-898,1997.

JING, W.; XYAOBING, Z.; JUNPING, P. YANG, E.; WEN, B.; JIAN, Y.; JIE, D.; LILIAN, S.; XYNGIE, X.; QI, J. Genomic compositions and phylogenetic analysis of *Shigella boydii* subgroup. **Science in China: Series C Life Sciences** v.49 -1, p. 46-52, 2006.

JORDAN, D.C. "Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from

leguminous plants." **International Journal of Systematic Bacteriology**, v.32, p.136-139., 1982.

JORDAN, I.K.; MAKAROVA, K.S.; SPOUGE, J.L.; WOLF, Y.I.; KOONIN, E.V. Lineage-specific gene expansions in Bacterial and Archaeal genomes. **Genome Research**, v.11, p. 555 –565, 2001.

KANEKO, T.; NAKAMURA, Y.; SATO, S.; MINAMISAWA, K.; UCHIUMI, T.; SASAMOTO, S.; WATANABE, A.; IDESAWA, K.; IRIGUCHI, M.; KAWASHIMA, K.; KOHARA, M.; MATSUMOTO, M.; SHIMPO, S.; TSURUOKA, H.; WADA, T.; YAMADA, M.; TABATA, S. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. **DNA Reserach** v.9, p.189-197,2002.

KLASSON, L.; ANDERSSON, S.G.E. Evolution of minimal- gene-sets in host-dependent bacteria. **Trends in Microbiology**, v.12, p.37 – 43, 2004.

KORBEL, J.O.; SNEL, B.; HUYNEN, M.A.; BORK, P. SHOT: a web server for the construction of genome phylogenies, **Trends in Genetics**, v18, p.158 –162, 2002.

KUMAR, S.; TAMURA, K.; JAKOBSEN, I.B. NEI, M.; MEGA2: molecular evolutionary genetics analysis software **Bioinformatics**, v.17, p.1244-1245, 2001.

KUNING, V.; OUZOUNIS, C.A. The balance of driving forces during genome evolution in prokaryotes. **Genome Research**, v.13, p.1589 –1594, 2003.

KURTZMAN, C.P.; ROBNETT, C.J. Phylogenetic relationships among yeasts of the "*Sacharomyces comples*" determined from multigene sequence analyses. **FEMS Yeast Research**, v.3, p.417-432, 2003.

KUYKENDALL, L.D., Saxena, B., Devine, T.E., and Udell, S.E. "Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov." **Canadian Journal Microbiology**, v.38, p.501-505, 1992.

KWON, S.W.; PARK, J.Y.; KIM, J.S.; KANG, J.W.; CHO, Y.H.; LIM, C.K. PARKER, M.A.; LEE, G.B. Phylogenetic analysis of the genera *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium* and *Sinorhizobium* on the basis of 16S rRNA gene and internally transcribed spacer region sequences. **International Journal of Systematic Evolution Microbiology**, v.55, p.263–270, 2005.

LAGE, J.M.; Leamon, J.H.; Pejovic, T.; Hamann, S.; Lacey, M. Whole Genome Analysis of Genetic Alterations in Small DNA Samples Using Hyperbranched Strand Displacement Amplification and Array–CGH. **Genome Research**, v.13, p.294-307, 2003.

LAWRENCE, J. G. Gene transfer in bacteria: speciation without species. **Theoretical Population Biology**, v.61, p.449–460, 2002.

MAIDEN, M. C. J.; BYGRAVES, J. A. B.; FEIL, E; MORELLI, G.; RUSSELL, J.E.; URWIN, R.; ZHANG, Q.; ZHOU, J.; ZURTH, K.; CAUGANT, D.A.; FEAVERS, I. M.; ACHTMAN, M.; SPRATT, B.G. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. **Proceedings of the National Academy of Sciences** v.95, p.140–3145, 1998.

MARRI, P.R.; HAO, W.; GOLDING, G.B. Gene Gain and Gene Loss in *Streptococcus*: Is It Driven by Habitat?. **Molecular Biology Evolution**, v.23, p.2379-2391, 2006.

MIRA, A.; OCHMAN, H.; MORAN, N.A. Deletional bias and the evolution of bacterial genomes. **Trends in Genetics**, v.17, p.589–596, 2001.

MORAN, N.A. Tracing the evolution of gene loss in obligate bacterial symbionts. **Current Opinion in Microbiology**, v.6, p.512 –518, 2003.

NASH, J.H.E.; FINDLAY, W.A.; LUEBBERT, C.C.; MYKYTCZUK, O.L.; FOOTE, S.J.; TABOADA, N.E.; CARRILLO, C.D.; BOYD, J.M.; COLQUHOUN, D.J.; REITH, M.E.; BROWN, L.L. Comparative genomics profiling of clinical isolates of *Aeromonas salmonicida* using DNA microarrays. **BMC Genomics**, 2006.  
<http://www.biomedcentral.com/1471-2164/7/43>.

OCHMAN, H., LAWRENCE, J.G. AND GROISMAN, E.A. Lateral gene transfer and the nature of bacterial innovation. **Nature**, v.405, p.299 –304, 2000.

PHILIPPE, H.; DOUADY, C.J. Horizontal gene transfer and phylogenetics. **Current Opinion in Microbiology**, v.6, p.1 –8, 2003.

PORWOLLIK, S. *et al.*, Evolutionary genomics of Salmonella: Gene acquisitions revealed by microarray analysis. **Proceedings of the National Academy of Sciences**, V.99, p.8956-8961, 2002.

PRADO, A.L. (2004). **Hibridização DNA-DNA e expressão de genes simbióticos de *Bradyrhizobium japonicum* e *B. elkanii* através de “DNA Microarrays”**. 90 f. Tese (Doutorado em Produção Vegetal) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal.

QUACKENBUSH, J. Microarray data normalization and transformation. **Nature Genetics Supplement**, v.32, p.496-501, 2002.

RILEY, M. Functions of the gene products of *Escherichia coli*, **Microbiology Review**, v.57, p.862–952., 1993.

ROCHA, E.P.C. DNA repeats lead to the accelerated loss of gene order in bacteria. **Trends in Genetics**, v.19, p.600–603, 2003.

RODRIGUEZ-VALERA, F. Environmental genomics, the big picture? **FEMS Microbiology Letters**, v.231, p.153–158, 2004.

ROKAS, A.; WILLIAMS, B.L.; KING, N.; CARROL, S.B. Genome-scale approaches to resolving incongruence in molecular phylogenies. **Nature**, v.425, p.798-804, 2003.

ROSSELLO-MORA, R. Updating Prokaryotic Taxonomy. **Journal Bacteriology**, v.187, p. 6255-6257, 2005.

ROSSELLO-MORA, R.; AMANN, R. The species concept for prokaryotes. **FEMS Microbiology Review**, v.25, p.39–67, 2001.

RUMJANEK, N.G.; DOBERT, R.C.; BERKUM, P.; TRIPLETT, E.W. Common soybean inoculant strains in Brazil are members of *Brayrhizobium elkanii*. **Applied and Environmental Microbiology**, v.59, p.4371-4373, 1993.

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

SALAMA, N.; GUILLEMIN, K.; MCDANIEL, T.K.; SHERLOCK, G.; TOMPKINS, L.; FALKOW, S. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains **Proceedings of the National Academy of Sciences**, v 97, p.14668-14673, 2000.

SKERMAN, V.B.D., MCGOWAN, V., AND SNEATH, P.H.A. (editors): "Approved lists of bacterial names." *International Journal Systematics Bacteriology* v.30, p.225-420, 1980.

STACKEBRANDT, E.; FREDERIKSEN, W.; GARRITY, G.M.; GRIMONT, P.A.; KAMPFER, P.; MAIDEN, M.C.; NESME, X.; ROSSELLO-MORA, R.; SWINGS, J.; TRUPER, H.G.; VAUTERIN, L.; WARD, A.C.; WHITMAN, W.B. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. **International Journal Systematic Evolution Microbiology** v.52, p.1043 –1047, 2002.

SICHRITZ-PONTEN, T; ANDERSSON, S.G. A phylogenomic approach to microbial evolution. **Nucleic Acids Research**, v.29, p.545-552, 2001.

SNEL, B.; BORK, P.; HUYNEN, M.A. Genomes in flux: the evolution of archaeal and proteobacterial gene content. **Genome Research** v.12, p.17 –25, 2002.

SNEATH, P.H.; R. R. SOKAL. Numerical taxonomy. **Nature**, v.193, p.855-860, 1962.

SORENSEN, S. J.; SORENSEN, A. H.; HANSEN, L. H.; OREGAARD, G.; VEAL, D. DIRECT detection and quantification of horizontal gene transfer by using flow cytometry and gfp as a reporter gene. **Current Microbiology**, v. 47, p.129 –133, 2003.

SOUZA, J.A.M. (2006) **Perfil transcricional de *bradyrhizobium elkanii* semia 587 "in vitro" e em simbiose com soja (*Glycine max* L. Merrill) através de microarranjo de dna**. 152f. Tese (Doutorado em Microbiologia Agropecuária) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal.

SUYAMA, M.; BORK, P.; Evolution of prokaryotic gene order: genome rearrangements in closely related species. **Trends in Genetics**, v.17, p.10 –13, 2001.

TABOADA, E.N.; ACEDILLO, R.R.; LUEBBERT, C.C.; FINDLAY, W.A.; NASH, J.H.; A new approach for the analysis of bacterial microarray-based Comparative Genomic Hybridization: insights from an empirical study. **BMC Genomics**, v.6:78, 2005. Acessado em 2/10/2007

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=15918914>

TATUSOV, R. L.; FEDOROVA, N.D.; JACKSON, J.D.; JACOBS, A.R.; KIRYUTIN, B.; KOONIN, E.V.; KRYLOV, D.M.; MAZUMDER, R.; MEKHEDOV, S.L.; NIKOLSKAYA, A.N.; RAO, B.S.; SMIRNOV, S.; SVERDLOV, A.V.; The COG database: an updated version includes eukaryotes. **BMC Bioinformatics** 2003. <http://www.biomedcentral.com/1471-2105/4/41>

THOMPSON J.D.; GIBSON T.J. ; PLEWNIAK F.; JEANMOUGIN F. ; HIGGINS D.G. ; The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Research** v.25, p.4876-4882, 1997.

THOMPSON, F.L.; LIDA, T.; SWINGS, J. Biodiversity of vibrios. **Microbiology Molecular Biology Review**, v.68, p.403-431, 2004.

URSING, J.B.; ROSSELLO-MORA, R. A.; GARCIAVALDES, E.; LALUCAT, J. Taxonomic note –a pragmatic approach to the nomenclature of phenotypically similar genomic groups. **International Journal Systematic Bacteriology** v.45, p.604, 1995.

VANDAMME, P.; POT, B.; GILLIS, M.; VOS, P.; KERSTERS, K.; SWINGS, J.; Polyphasic taxonomy, a consensus approach to bacterial systematics. **Microbiology Reviews**, v.60, p.407-438, 1996.

VAN BERKUM, P.; TEREFEWOR, Z.; PAULIN, L.; SUOMALAINEN, S.; LINDSTROM, K.; EARDLY, B.D. Discordant phylogenies with the *rrn* loci of *Rhizobia*. **Journal. Bacteriology**, v.185, p.2988–2998, 2003.

VASUDEVAN, S.; WOLF, Y.I.; YIN, J.I.; NATALE, D.A. The COG database: an updated version includes eukaryotes. **BMC Bioinformatics** v.4, n.41, 2003.

WAN, Y.; BROCHAT, S.L.; CALL, D.R.; Validation of Mixed-Genome Microarrays as a Method for Genetic Discrimination. **Applied Environmental Microbiology** V.73, p.1425-1432, 2007.

WARD, D.M.; COHAN, F.M. BHAYA, D.; HEIDELBERG, J.F. KÜHL, M.; GROSSMAN, A. Genomics, environmental genomics and the issue of microbial species, **Heredity**, 1–13, 2007.

WATANABE, T.; MURATA, Y.; OKA, S.; IWAHASHI, H. A new approach to species determination for yeast strains: DNA microarray-based comparative genomic hybridization using a yeast DNA microarray with 6000 genes. **Yeast**, v.21, p.351-365, 2004.

WILLEMS, A. The taxonomy of rhizobia: an overview. **Plant and Soil** v.287, p.3–14, 2006.

WOESE, C.R. Interpreting the universal phylogenetic tree. **Proceedings of the National Academy of Sciences USA**, v97, p.8392 –8396, 2000.

WOLFGANG, M. C. *et al.*, Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. **Proceedings of the National Academy of Sciences USA**, v.100, p.8484-8489, 2003.

YOUNG J. M. Renaming of *Agrobacterium larrymoorei* Bouzar and Jones 2001 as *Rhizobium larrymoorei* (Bouzar and Jones 2001) comb. nov. **International Journal Systematic and Evolutionary Microbiology**. 54, 149, 2004.

YANG, Y.H.; DUDOIT, S.; LUU, P.; LIN, D.M.; PENG, V.; NGAI, J.; SPEED, T.P. Normalization for cDNA microarray data: A robust composite method addressing 126 single and multiple slide systematic variation. **Nucleic Acids Research**, v.30(4): e15, 2002.

YOUNG, J.P.W.; CROSSMAN, L.C.; JOHNSTON, A.W.B.; THOMSON, N.R.; GHAZOU, Z.F.; HULL, K.H.; WEXLER, M.; CURSON, A.R.J.; TODD, J.D.; POOLE, P.S.; MAUCLINE, T.H.; EAST, A.K.; QUAIL, M.A.; CHURCHER, C.; ARROWSMITH, C.; CHEREVACH, I.; CHILLINGWORTH, T.; CLARKE, K.; CRONIN, A.; DAVIS, P.; FRASER, A.; HANCE, Z.; HAUSER, H.; JAGELS, K.; MOULE, H.; MUNGALL, K.; NORBERTCZAK, H.; RABBINOWITSCH, E.; SANDERS, M.; SIMMONDS, M.; WHITEHEAD, S.; PARKHILL, J. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components **Genome Biology** 7:R34 2006.

<http://genomebiology.com/2006/7/4/R34>

YOUNG J. M.; KUYKENDALL, L. D.; MARTINEZ-ROMERO E.; KERR, A.; SAWADA, H. A Revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all. *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. **International Journal Systematic Bacteriology**, v. 51, p.89–103, 2001.

## VII. APÊNDICE

**Anexo 1 – Lista das quinze categorias e subcategorias funcionais baseadas em Mônica Riley (1993) e utilizadas no estudo.**

<b>1</b>	<b>Biossíntese de aminoácidos</b>
	1.aminoácidos aromáticos; 2.aspartato; 3.aminoácidos de cadeia ramificada; 4.glutamato (assimilação de nitrogênio); 5. serina (assimilação de enxofre)
<b>2</b>	<b>Biossíntese de Cofatores, Grupos Prostéticos e Carreadores</b>
	1.biotina; 2.carotenóide; 3.cobalamina, 4.grupo heme, 5.ficobilina e porfirina; 6.ácido fólico; 7.lipoato ; 8.menaquinona e ubiquinona; 9.molibdopterina; 10.nicotinato e nicotinamida; 11.pantotenato; 12.piridoxina; 13. quinolinato; 14.riboflavina; 15.tiamina; 16.tiredoxina, glutaredoxina e glutatona
<b>3</b>	<b>Envelope Celular</b>
	1.membranas, lipoproteínas e porinas; 2.mureína e peptidoglicano; 3.polissacarídeos de superfície, lipolissacarídeos e antígenos; 4.estruturas de superfície
<b>4</b>	<b>Processos Celulares</b>
	1.divisão celular; 2.morte celular; 3.chaperoninas; 4.quimiotaxia; 5.destoxificação; 6.secreção de proteínas e peptídeos; 7.transformação; 8.transferências de plasmídeos)
<b>5</b>	<b>Metabolismo Intermediário Central</b>
	1.açúcares aminados; 2.compostos fosfóricos; 3.fixação de nitrogênio; 4.metabolismo do nitrogênio; 5.polissacarídeos e glicoproteínas
<b>6</b>	<b>Metabolismo Energético</b>
	1.Aminoácidos e aminas; 2.via do glicolato; 3.glicólise; 4.vias da pentose-fosfato; 5.metabolismo do piruvato e acetil-CoA; 6.açúcares, ciclo do ácido tricarboxílico; 7.respiração
<b>7</b>	<b>Metabolismo de Ácidos Graxos, Fosfolipídeos e Esterol</b>
<b>8</b>	<b>Purinas, Pirimidinas, Nucleosídeos e Nucleotídeos</b>
	1.interconversões e “salvage” de nucleosídeos e nucleotídeos; 2.biossíntese de purinas e pirimidinas ribonucleotídicas
<b>9</b>	<b>Funções de Regulação</b>
<b>10</b>	<b>Replicação, Recombinação e Reparo de DNA</b>
<b>11</b>	<b>Transcrição</b>
	1.degradação de RNA; 2.síntese de RNA, modificação e transcrição de DNA
<b>12</b>	<b>Tradução</b>
	1.tRNA aminoacil sintetases e modificação de tRNA; 2.proteínas ribossomais, degradação de proteínas, peptídeos e glicoproteínas; 3.nucleoproteínas; modificação protéica e fatores de tradução
<b>13</b>	<b>Transporte e Proteínas de Ligação</b>
<b>14</b>	<b>Outras Categorias</b>
	1.adaptações e condições atípicas; 2.sensibilidade à drogas e análogos; 3.sensibilidade à radiação; 4.proteínas de “WD-repeat” (repetição de triptofano e aspartato); 5.hidrogenase; 6.transposons; 7.simbiose; 8.Outros
<b>15</b>	<b>Hipotéticas</b>
	1.proteínas hipotéticas conservadas; 2.sem similaridade

**Anexo 2 – Distribuição nas 15 categorias funcionais dos 65 genes observados em comum entre todas bactérias do estudo.**

<b>Grupos funcionais</b>	<b>ID</b>	<b>Produto gênico (Gene)</b>	<b>e-value</b>	<b>score</b>
<b>1. Biossíntese de aminoácidos</b>				
Família do aspartato	blI5902	threonine dehydratase ( <i>tlvA</i> )	5e-66	245(626)
Família da Serina / Assimilação de Enxofre	blr1481	phosphoadenosine phosphosulfate reductase EC 1.8.4.8 ( <i>cysH</i> )	9e-89	321(822)
Glutamato	blI1069	probable glutamine synthetase (EC 6.3.1.2)	e-157	552(1422)
<b>2. Biossíntese de Cofatores, Grupos Prostéticos e Carreadores</b>				
Outros	blI4939	methyltransferase	3e-72	(266) 680
<b>3-Envelope celular</b>				
Família da mureína e peptidoglicano	blr5156	putative D-alanyl-D-alanine carboxypeptidase	1e-81	297(791)
Polissacarídeos de superfície, lipopolissacarídeos e antígenos	blI5980	glucose-1-phosphate cytidyltransferase	8e-18	85.5 (210)
Polissacarídeos de superfície, lipopolissacarídeos e antígenos	NP774803	probable glycosyl transferase	1e-77	373 (958)
Estrutura de superfície	blr2079	rtxA homolog	2e-65	243 (621)
Membranas, lipoproteínas e porinas	YP_779224.1	lipoprotein A	4e-49	198 (503)
<b>4. Processos Celulares</b>				
Quimiotaxia	blI5814	probable flagellar basal-body rod protein ( <i>figB</i> )	4e-41	162 (411)
		Score = 162 bits		

(411), Expect = 4e-41

**5. Metabolismo intermediário central**

Fixação do Nitrogênio	blr4342	amino transferase ( <i>nifS</i> )	3e-56	213 (542)
-----------------------	---------	-----------------------------------	-------	-----------

**6. Metabolismo energético**

Aminoácidos e aminas	bl13177	probable arginine/lysine/ornithine decarboxylases	1e-134	472(1214)
Aminoácidos e aminas	blr0807	succinate-semialdehyde	5e-45	174 (442)
Piruvato desidrogenase	bl17292	probable pyruvate	6e-09	402(1032)
Acúcares	blr6758	probable transaldolase	1e-104	373 (957)
Ciclo dos ácidos tricarbóxicos	blr4839	citrate synthase ( <i>glTA</i> )	4e-86	312 (799)
Respiração	blr1174	putative cytochrome C ( <i>coxG</i> )	1e-72	268 (684)
Respiração	blr6340	multicopper oxidase ( <i>copA</i> )	1e-100	358 (918)
Respiração	YP_001373144.1	Gene info ETC complex I subunit conserved region	2e-15	87 (212)
Respiração	bl10440	ATP synthase beta	9e-88	318 (816)
Respiração	blr6340	probable hydrogenlyase component( <i>hycC</i> )	6e-88	318 (815)
Outros	blr2585	ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit( <i>cbbL</i> )	1e-106	379 (972)
Outros	YP_533890.1	GCN5-related N-acetyltransferase	1e-53	213 (543)
Outros	bl11244	flavin dependant oxidoreductase	3e-80	293 (749)

**7-Metabolismo de ácidos graxos, fosfolípidos e esteroi**

	bl16363	putative acyl-CoA dehydrogenase EC 1.3.99.3	1e-113	401(1030)
--	---------	---	--------	-----------

blr0139	putative acyl-CoA dehydrogenase (EC 1.3.99.-)	1e-121	429 (1103)
blr7807	probable acyl-CoA ligase	1e-119	421 (1082)
YP_485722.1	Gene info thiolase	1e-57	226 (577)
blI2946	enoyl CoA hydratase ( <i>fadB</i> )	3e-43	124 (310)
YP_567411.1	Gene info acyl-CoA dehydrogenase-like	2e-71	273 (697)
ZP_00944826.1	C-5 sterol desaturase	2e-35	152 (385)

### 8. Purinas, Pirimidinas, Nucleosídeos e Nucleotídeos

Biossíntese de purinas e pirimidinas ribonucleotídicas	blI0007	probable ribonucleoside-diphosphate reductase (EC 1.17.4.1)	1e-122	431 (1109)
--	---------	---	--------	------------

### 9. Funções de Regulação

	blI0790	probable transcriptional regulator	9e-61	228 (580)
	blr3121	two-component response	5e-78	285 (729)
	blr7678	transcriptional regulatory protein AraC family	9e-74	271 (693)
	blr0154	two-component sensor histidine kinase	9e-76	278 (710)

### 10. Replicação, Recombinação e Reparo de DNA

	NP_948032.1	Gene info possible uracil-DNA glycosylase	2e-50	203 (516)
	blI0823	DNA Gyrase subunit B	1e-166	582 (1501)
	blI0051	replication protein A ( <i>repA</i> )	4e-31	129 (325)

<b>12. Tradução</b>			
Modificação protéica e fatores de tradução	blr0961	putative translation initiation factor	1e-119 421 (1082)
<b>13. Transporte e Proteínas de Ligação</b>			
	blI0251	putative efflux protein	5e-50 192 (488)
	blr3224	ABC transporter ATP-binding	1e-54 207(527)
	blr6248	ABC transporter ATP-binding	2e-28 120 (301)
	blI6293	HlyB/MsbA family ABC	1e-128 452(1162)
	blI6234	ABC transporter permease protein	5e-24 106(264)
<b>14. Outras Categorias</b>			
Sensibilidade à drogas e análogos	blr1515	RND multidrug efflux membrane permease( <i>acrA</i> )	9e-99 354 (908)
Sensibilidade a Radiação	blI4062	DNA repair protein ( <i>radA</i> )	1e-10 52 (122)
Simbiose	blr2025	acyltransferase ( <i>nodA</i> )	2e-73 270 (689)
Simbiose	gb AAZ43128.1	N-acetylglucosaminyltransferase ( <i>NodC</i> )	4e-135 483 (1243)
Simbiose	blr2026	nodulation chito-oligosaccharide deacetylase ( <i>nodB</i> )	6e-24 112 (281)
Transposons	blI8200	putative transposase	7e-99 334 (857)
Outros	blr6213	methanol dehydrogenase large subunit ( <i>mxoF</i> )	
Outros	blI3054	methylmalonyl-CoA mutase ( <i>mutA</i> )	1e-90 327 (839)
Outros	blr3805	putative aminotransferase (EC 2.6.1.-)	1e-100 359 (922)

## 15. Hipotéticas

Proteínas hipotéticas conservadas	blr6188	hypothetical protein	1e-101	362 (929)
Proteínas hipotéticas conservadas	blr0533	hypothetical protein	1e-99	357 (915)
Proteínas hipotéticas conservadas	blr6071	hypothetical protein	5e-43	169 (427)
Proteínas hipotéticas conservadas	blr3062	hypothetical protein	5e-87	315 (806)
Proteínas hipotéticas conservadas	bsr1458	hypothetical protein	4e-43	169 (428)
Proteínas hipotéticas conservadas	blr7676	hypothetical protein	5e-91	328 (841)
Proteínas hipotéticas conservadas	blr0721	hypothetical protein	1e-72	267 (683)
Proteínas hipotéticas conservadas	bl11403	hypothetical protein	1e-55	211 (536)
Proteínas hipotéticas conservadas	bl11430	hypothetical protein	9e-76	278 (710)
Proteínas hipotéticas conservadas	blr2978	hypothetical protein	4e-86	312 (799)
Sem similaridade	blr3498	unknown protein	5e-10	60 (142)

**Anexo 3 - Número de genes que estão em comum entre *B. elkanii* e *B japonicum*, distribuídos em grupos de subcategorias funcionais.**

Categoria Funcional	Subcategoria	Subcategoria Funcional	Diferentes
1	1	Família de aminoácidos aromáticos	8
1	2	Família do aspartato	14
1	3	Branched chain family	17
1	4	Família do glutamato / Assimilação do Nitrogênio	14
1	5	Família da Serina / Assimilação de Enxofre	9
2	1	Biotina	5
2	2	Carotenoide	2
2	3	Cobalamina, heme, phycobilin and porphyrin	23
2	4	Ácido Fólico	4
2	5	Lipoato	1
2	6	Menaquinone and ubiquinone	4
2	7	Molybdopterin	2
2	8	Nicotinato e nicotinamida	2
2	9	Patotenato	1
2	10	Piridoxina	1
2	11	Quinolinato	1
2	12	Riboflavina	2
2	13	Tiamina	4
2	14	Thioredoxin, glutaredoxin, and glutathione	15
2	15	Outros	2
3	1	Membranas, lipoproteínas e porinas	9
3	2	Murein sacculus e peptídeoglicano	16
3	3	Polissacarídeos de superfície, lipopolissacarídeos e antígenos	11
3	4	EStruturas de superfície	2
4	1	Divisão Celular	11
4	2	Morte celular	0
4	3	Chaperonas	14
4	4	Quimiotaxia	20
4	5	Desentoxicação	20
4	6	Secreção de peptídeos e proteínas	16
4	7	Transformação	0
4	8	Transferência de plasmídeos	7
5	1	Açúcares aminoácidos	0
5	2	Componentes fósforo	6
5	3	Fixação do Nitrogênio	41
5	4	Metabolismo do nitrogênio	7
5	5	Glicoproteínas e polissacarídeos	8
5	6	Outros	18
6	1	Aminoácidos e aminas	10
6	2	Via do Glicolato	2

6		3 Glicólise	5
6		4 Via da pentose fosfato	2
6		5 Metabolismo do Piruvato e do acetil-CoA	17
6		6 Piruvato desidrogenase	5
6		7 Açúcares	25
6		8 Ciclo dos ácidos tricarboxílicos	17
6		9 Respiração	30
6		10 Outros	9
8		1 Interconversions and salvage of nucleosides and nucleotides	0
8		2 Biossíntese de ribonucleotídeo Purina	25
8		3 Biossíntese de ribonucleotídeo pirimidina	8
11		1 Degradação de RNA	5
11		2 Síntese de RNA, modificação e transcrição do DNA	14
12		1 Modificação de tRNA e aminoacil sintetase tRNA	28
12		2 Proteínas Ribosomais	11
12		3 Degradação de proteínas, peptídeos e glycoproteínas	25
12		4 Nucleoproteínas	0
12		5 Modificação de proteína e fatores de tradução	15
14		1 Adaptações e condições atípicas	2
14		2 Sensitividade análoga e drogas	26
14		3 Sensitividade Radiação	2
14		4 Proteínas de repetição WD	0
14		5 Hidrogenase	3
14		6 Funções relacionadas a transposon	50
14		7 Simbiose	23
14		8 Outros	237
15		1 Proteínas hipotéticas conservadas	500
15		2 Sem similaridade	30
7	Não contém	Metabolismo de ácidos graxos, fosfolipídeos e esterois	91
9	Não contém	Funções regulatórias	121
10	Não contém	Recombinação, reparo e replicação do DNA	50
13	Não contém	Proteínas de transporte e ligação	235
<b>Total</b>			<b>1960</b>

**Anexo 4 - Número de genes que estão em comum entre *B. elkanii* e *Azorhizobium caulinodans* distribuídos em grupos de subcategorias funcionais.**

Categoria Funcional	Subcategoria	Subcategoria Funcional	Diferentes
1	1	Família de aminoácidos aromáticos	9
1	2	Família do aspartato	18
1	3	Aminoácidos de cadeia ramificada	16
1	4	Família do glutamato / Assimilação do Nitrogênio	15
1	5	Família da Serina / Assimilação de Enxofre	10
2	1	Biotina	5
2	2	Carotenoide	3
2	3	Cobalamina, heme, ficobilina e porfirina	22
2	4	Ácido Fólico	5
2	5	Lipoato	1
2	6	Menaquinona e ubiquinona	4
2	7	Molibdoproteína	2
2	8	Nicotinato e nicotinamida	2
2	9	Patotenato	1
2	10	Piridoxina	2
2	11	Quinolinato	1
2	12	Riboflavina	3
2	13	Tiamina	5
2	14	tiredoxina, glutaredoxina e glutathione	16
2	15	Outros	2
3	1	Membranas, lipoproteínas e porinas	10
3	2	Mureína e peptídeoglicano	15
3	3	Polissacarídeos de superfície, lipopolissacarídeos e antígenos	13
3	4	Estruturas de superfície	2
4	1	Divisão Celular	11
4	2	Morte celular	0
4	3	Chaperoninas	17
4	4	Quimiotaxia	19
4	5	Destoxicação	20
4	6	Secreção de peptídeos e proteínas	17
4	7	Transformação	0
4	8	Transferência de plasmídeos	7
5	1	Açúcares aminoácidos	0
5	2	Componentes fósforo	6
5	3	Fixação do Nitrogênio	42
5	4	Metabolismo do nitrogênio	9
5	5	Glicoproteínas e polissacarídeos	8
5	6	Outros	20
6	1	Aminoácidos e aminas	10
6	2	Via do Glicolato	2
6	3	Glicólise	5
6	4	Via da pentose fosfato	2

6	5	Metabolismo do Piruvato e do acetil-CoA	19
6	6	Piruvato desidrogenase	5
6	7	Acúcares	25
6	8	Ciclo dos ácidos tricarboxílicos	18
6	9	Respiração	29
6	10	Outros	9
8	1	Interconvenções e “salvage” nucleosídeos e nucleotídeos	0
8	2	Biossíntese de ribonucleotídeo Purina	25
8	3	Biossíntese de ribonucleotídeo pirimidina	8
11	1	Degradação de RNA	6
11	2	Síntese de RNA, modificação e transcrição do DNA	15
12	1	Modificação de tRNA e aminoacil sintetase tRNA	30
12	2	Proteínas Ribosomais	11
12	3	Degradação de proteínas, peptídeos e glycoproteínas	28
12	4	Nucleoproteínas	1
12	5	Modificação de proteína e fatores de tradução	16
14	1	Adaptações e condições atípicas	2
14	2	Sensitividade análoga e drogas	25
14	3	Sensitividade Radiação	2
14	4	Proteínas de repetição WD	0
14	5	Hidrogenase	3
14	6	Funções relacionadas a transposon	50
14	7	Simbiose	23
14	8	Outros	107
15	1	Proteínas hipotéticas conservadas	385
15	2	Sem similaridade	32
7	Não contém	Metabolismo de ácidos graxos, fosfolipídeos e esterol	92
9	Não contém	Funções regulatórias	126
10	Não contém	Recombinação, reparo e replicação do DNA	53
13	Não contém	Proteínas de transporte e ligação	245
<hr/>			
Total			1767
<hr/>			

**Anexo 5 - Número de genes que estão em comum entre *B. elkanii* e *Mesorhizobium huakuii* distribuídos por grupos funcionais divididos em subcategorias**

Categoria Funcional	Subcategoria	Subcategoria Funcional	Diferentes
1	1	Família de aminoácidos aromáticos	9
1	2	Família do aspartato	17
1	3	Aminoácidos de cadeia ramificada	16
1	4	Família do glutamato / Assimilação do Nitrogênio	15
1	5	Família da Serina / Assimilação de Enxofre	10
2	1	Biotina	5
2	2	Carotenoide	3
2	3	Cobalamina, heme, ficobilina e porfirina	23
2	4	Ácido Fólico	5
2	5	Lipoato	1
2	6	Menaquinona e ubiquinona	4
2	7	Molibdoproteína	2
2	8	Nicotinato e nicotinamida	2
2	9	Patotenato	1
2	10	Piridoxina	2
2	11	Quinolinato	1
2	12	Riboflavina	3
2	13	Tiamina	4
2	14	tiredoxina, glutaredoxina e glutathione	15
2	15	Outros	2
3	1	Membranas, lipoproteínas e porinas	10
3	2	Mureína e peptídeoglicano	16
3	3	Polissacarídeos de superfície, lipopolissacarídeos e antígenos	12
3	4	Estruturas de superfície	2
4	1	Divisão Celular	12
4	2	Morte celular	0
4	3	Chaperoninas	15
4	4	Quimiotaxia	18
4	5	Destoxicação	20
4	6	Secreção de peptídeos e proteínas	17
4	7	Transformação	0
4	8	Transferência de plasmídeos	7
5	1	Açúcares aminoácidos	0
5	2	Componentes fósforo	6
5	3	Fixação do Nitrogênio	42
5	4	Metabolismo do nitrogênio	10
5	5	Glicoproteínas e polissacarídeos	8
5	6	Outros	19
6	1	Aminoácidos e aminas	9
6	2	Via do Glicolato	2
6	3	Glicólise	5
6	4	Via da pentose fosfato	1

6	5	Metabolismo do Piruvato e do acetil-CoA	17
6	6	Piruvato desidrogenase	5
6	7	Acúcares	23
6	8	Ciclo dos ácidos tricarboxílicos	18
6	9	Respiração	30
6	10	Outros	9
8	1	Interconvenções e “salvage” nucleosídeos e nucleotídeos	1
8	2	Biossíntese de ribonucleotídeo Purina	24
8	3	Biossíntese de ribonucleotídeo pirimidina	8
11	1	Degradação de RNA	6
11	2	Síntese de RNA, modificação e transcrição do DNA	16
12	1	Modificação de tRNA e aminoacil sintetase tRNA	29
12	2	Proteínas Ribosomais	10
12	3	Degradação de proteínas, peptídeos e glycoproteínas	27
12	4	Nucleoproteínas	1
12	5	Modificação de proteína e fatores de tradução	16
14	1	Adaptações e condições atípicas	1
14	2	Sensitividade análoga e drogas	25
14	3	Sensitividade Radiação	2
14	4	Proteínas de repetição WD	0
14	5	Hidrogenase	3
14	6	Funções relacionadas a transposon	50
14	7	Simbiose	24
14	8	Outros	163
15	1	Proteínas hipotéticas conservadas	430
15	2	Sem similaridade	31
7	Não contém	Metabolismo de ácidos graxos, fosfolipídeos e esterol	89
9	Não contém	Funções regulatórias	126
10	Não contém	Recombinação, reparo e replicação do DNA	51
13	Não contém	Proteínas de transporte e ligação	240
<b>Total</b>			<b>1846</b>

**Anexo 6 - Número de genes em comum entre o genoma parcial de *B. elkanii* e *Rhizobium leguminosarum* bv. *Trifolii* distribuídos por grupos funcionais divididos em subcategorias**

Categoria Funcional	Subcategoria	Subcategoria Funcional	Diferentes
1	1	Família de aminoácidos aromáticos	8
1	2	Família do aspartato	17
1	3	Aminoácidos de cadeia ramificada	16
1	4	Família do glutamato / Assimilação do Nitrogênio	14
1	5	Família da Serina / Assimilação de Enxofre	10
2	1	Biotina	5
2	2	Carotenoide	3
2	3	Cobalamina, heme, ficobilina e porfirina	22
2	4	Ácido Fólico	5
2	5	Lipoato	1
2	6	Menaquinona e ubiquinona	4
2	7	Molibdoproteína	2
2	8	Nicotinato e nicotinamida	2
2	9	Patotenato	1
2	10	Piridoxina	2
2	11	Quinolinato	1
2	12	Riboflavina	3
2	13	Tiamina	4
2	14	tiredoxina, glutaredoxina e glutationa	15
2	15	Outros	2
3	1	Membranas, lipoproteínas e porinas	10
3	2	Mureína e peptídeoglicano	15
3	3	Polissacarídeos de superfície, lipopolissacarídeos e antígenos	12
3	4	Estruturas de superfície	2
4	1	Divisão Celular	11
4	2	Morte celular	0
4	3	Chaperoninas	15
4	4	Quimiotaxia	17
4	5	Destoxicação	18
4	6	Secreção de peptídeos e proteínas	17
4	7	Transformação	0
4	8	Transferência de plasmídeos	7
5	1	Açúcares aminoácidos	0
5	2	Componentes fósforo	6
5	3	Fixação do Nitrogênio	41
5	4	Metabolismo do nitrogênio	10
5	5	Glicoproteínas e polissacarídeos	8
5	6	Outros	18
6	1	Aminoácidos e aminas	8
6	2	Via do Glicolato	2

6	3	Glicólise	5
6	4	Via da pentose fosfato	1
6	5	Metabolismo do Piruvato e do acetil-CoA	17
6	6	Piruvato desidrogenase	5
6	7	Acúcares	21
6	8	Ciclo dos ácidos tricarboxílicos	15
6	9	Respiração	30
6	10	Outros	9
8	1	Interconvenções e “salvage” nucleosídeos e nucleotídeos	2
8	2	Biossíntese de ribonucleotídeo Purina	22
8	3	Biossíntese de ribonucleotídeo pirimidina	8
11	1	Degradação de RNA	5
11	2	Síntese de RNA, modificação e transcrição do DNA	15
12	1	Modificação de tRNA e aminoacil sintetase tRNA	29
12	2	Proteínas Ribosomais	11
12	3	Degradação de proteínas, peptídeos e glycoproteínas	25
12	4	Nucleoproteínas	1
12	5	Modificação de proteína e fatores de tradução	16
14	1	Adaptações e condições atípicas	1
14	2	Sensitividade análoga e drogas	24
14	3	Sensitividade Radiação	2
14	4	Proteínas de repetição WD	0
14	5	Hidrogenase	1
14	6	Funções relacionadas a transposon	49
14	7	Simbiose	23
14	8	Outros	229
15	1	Proteínas hipotéticas conservadas	482
15	2	Sem similaridade	24
7	Não contém	Metabolismo de ácidos graxos, fosfolipídeos e esterois	88
9	Não contém	Funções regulatórias	123
10	Não contém	Recombinação, reparo e replicação do DNA	50
13	Não contém	Proteínas de transporte e ligação	231
Total			1918

**Anexo 7- Número de genes em comum entre o genoma parcial de *B. elkanii* e *Sinorhizobium meliloti* distribuídos por grupos funcionais divididos em subcategorias**

Categoria Funcional	Subcategoria	Subcategoria Funcional	Diferentes
1	1	Família de aminoácidos aromáticos	9
1	2	Família do aspartato	15
1	3	Aminoácidos de cadeia ramificada	15
1	4	Família do glutamato / Assimilação do Nitrogênio	13
1	5	Família da Serina / Assimilação de Enxofre	9
2	1	Biotina	5
2	2	Carotenoide	2
2	3	Cobalamina, heme, ficobilina e porfirina	19
2	4	Ácido Fólico	3
2	5	Lipoato	1
2	6	Menaquinona e ubiquinona	2
2	7	Molibdoproteína	2
2	8	Nicotinato e nicotinamida	1
2	9	Patotenato	0
2	10	Piridoxina	1
2	11	Quinolinato	1
2	12	Riboflavina	2
2	13	Tiamina	3
2	14	tiredoxina, glutaredoxina e glutathione	15
2	15	Outros	2
3	1	Membranas, lipoproteínas e porinas	8
3	2	Mureína e peptídeoglicano	13
3	3	Polissacarídeos de superfície, lipopolissacarídeos e antígenos	13
3	4	Estruturas de superfície	2
4	1	Divisão Celular	12
4	2	Morte celular	0
4	3	Chaperoninas	14
4	4	Quimiotaxia	17
4	5	Destoxicação	16
4	6	Secreção de peptídeos e proteínas	16
4	7	Transformação	0
4	8	Transferência de plasmídeos	7
5	1	Açúcares aminoácidos	0
5	2	Componentes fósforo	6
5	3	Fixação do Nitrogênio	40
5	4	Metabolismo do nitrogênio	9
5	5	Glicoproteínas e polissacarídeos	7
5	6	Outros	18
6	1	Aminoácidos e aminas	9
6	2	Via do Glicolato	2
6	3	Glicólise	5
6	4	Via da pentose fosfato	2

6	5	Metabolismo do Piruvato e do acetil-CoA	15
6	6	Piruvato desidrogenase	4
6	7	Acúcares	21
6	8	Ciclo dos ácidos tricarboxílicos	15
6	9	Respiração	31
6	10	Outros	8
8	1	Interconvenções e “salvage” nucleosídeos e nucleotídeos	1
8	2	Biossíntese de ribonucleotídeo Purina	21
8	3	Biossíntese de ribonucleotídeo pirimidina	7
11	1	Degradação de RNA	5
11	2	Síntese de RNA, modificação e transcrição do DNA	13
12	1	Modificação de tRNA e aminoacil sintetase tRNA	26
12	2	Proteínas Ribosomais	10
12	3	Degradação de proteínas, peptídeos e glycoproteínas	26
12	4	Nucleoproteínas	1
12	5	Modificação de proteína e fatores de tradução	14
14	1	Adaptações e condições atípicas	1
14	2	Sensitividade análoga e drogas	25
14	3	Sensitividade Radiação	2
14	4	Proteínas de repetição WD	0
14	5	Hidrogenase	3
14	6	Funções relacionadas a transposon	47
14	7	Simbiose	22
14	8	Outros	219
15	1	Proteínas hipotéticas conservadas	460
15	2	Sem similaridade	27
7	Não contém	Metabolismo de ácidos graxos, fosfolipídeos e esterol	85
9	Não contém	Funções regulatórias	109
10	Não contém	Recombinação, reparo e replicação do DNA	51
13	Não contém	Proteínas de transporte e ligação	228
<hr/> Total			<hr/> 1833 <hr/>