

Marcio Martins Loureiro



**Mapeamento Físico e Genético do
Gluconacetobacter diazotrophicus PAL5**

Universidade Federal do Rio de Janeiro
Centro de Ciências da Saúde
Instituto de Bioquímica Médica
Núcleo de Estudos de Genomas Johanna Döbereiner

Maio de 2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Mapeamento Físico e Genético do *Gluconacetobacter diazotrophicus* PAL5

Marcio Martins Loureiro

Tese de Doutorado apresentada ao Programa de Pós-graduação em Química Biológica, Instituto de Bioquímica Médica, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro (UFRJ), como parte dos requisitos necessários à obtenção do título de Doutor em Ciências.

Orientadores: Prof^o Dr. Orlando Bonifácio Martins
Prof^o Dr. Paulo Cavalcanti Gomes Ferreira

Rio de Janeiro - RJ - Brasil

Maio de 2007

Ficha Catalográfica

Loureiro, Marcio Martins

Mapeamento Físico e Genético do *Gluconacetobacter diazotrophicus* PAL5 / Marcio Martins Loureiro - Rio de Janeiro: UFRJ / IBQM, 2007.

203 p.

Orientadores: Orlando Bonifácio Martins e Paulo Cavalcanti Gomes Ferreira.

Tese (Doutorado) - UFRJ/ Instituto de Bioquímica Médica / Programa de Pós-graduação em Química Biológica, 2007. Referências Bibliográficas: f.141-165.

1. *Gluconacetobacter diazotrophicus*. 2. Mapeamento físico. 3. PFGE. 4. Sequenciamento genômico. 5. Fechamento de Gaps. I. Martins, Orlando Bonifácio. Ferreira, Paulo Cavalcanti Gomes. II. Universidade Federal do Rio de Janeiro, Instituto de Bioquímica Médica, Programa de Pós-graduação em Química Biológica. III. Título.

Mapeamento Físico e Genético do *Gluconacetobacter diazotrophicus* PAL5

Marcio Martins Loureiro

Orientadores: Prof. Dr. Orlando Bonifácio Martins

Prof. Dr. Paulo Cavalcanti Gomes Ferreira

Tese de Doutorado submetida ao Programa de Pós-graduação em Química Biológica, Instituto de Bioquímica Médica, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro (UFRJ), como parte dos requisitos necessários à obtenção do título de Doutor em Ciências.

Aprovada por:

Prof. Dr. Orlando Bonifácio Martins
Instituto de Bioquímica Médica – UFRJ

Prof. Dr. Paulo Cavalcanti Gomes Ferreira
Instituto de Bioquímica Médica – UFRJ

Prof. Dr. Franklin David Rumjanek
Instituto de Bioquímica Médica – UFRJ

Prof. Dr. Rodolpho Mattos Albano
Departamento de Bioquímica – UERJ

Prof. Dra. Ana Coelho
Departamento de Genética – UFRJ

Prof. Dr. Marcius da Silva Almeida
Instituto de Bioquímica Médica – UFRJ

Prof. Dra. Vânia Lúcia Muniz de Pádua
Centro Universitário Estadual da Zona Oeste – UEZO

Rio de Janeiro
Maio de 2007

AGRADECIMENTOS

A todos os amigos e familiares por seus constantes apoio e incentivo no decurso da minha vida acadêmica, o que contribuiu diretamente para o sucesso da minha empreitada científica.

Ao corpo docente das disciplinas oferecidas ao programa de pós-graduação em Química Biológica do Instituto de Bioquímica Médica – CCS – UFRJ, em virtude dos conhecimentos transmitidos no decorrer das atividades didáticas e laboratoriais.

A todos os amigos do programa de pós-graduação em Química Biológica do Instituto de Bioquímica Médica – UFRJ, pelo companheirismo demonstrado em todos os momentos.

A todos os membros da Comissão de Pós-Graduação do Instituto de Bioquímica Médica e funcionários da Secretaria de Pós-Graduação do referido instituto, em especial a Tereza Cristina Lima, devido ao enorme empenho e atenção dispensados a todos os alunos do curso de Pós-Graduação em Química Biológica.

Aos membros integrantes do Projeto Riogene pelo empenho demonstrado durante todas as fases de execução do sequenciamento e anotação do genoma do *Gluconacetobacter diazotrophicus* PAL5. Especialmente os alunos e pesquisadores do Laboratório de Biologia Molecular – Instituto de Bioquímica Médica – CCS – UFRJ, Laboratório de Biologia Molecular de Plantas – Jardim Botânico – RJ e Laboratório de Genética e Bioquímica – Embrapa-Agrobiologia – RJ pelo companheirismo e colaborações realizadas no decurso da fase experimental desta tese.

Aos diretores do Instituto de Ciências Biológicas – UFRJ e Instituto de Bioquímica Médica – UFRJ pelo empenho, dedicação e incentivo à pesquisa em todos os seus aspectos dentro dos referidos institutos.

Ao Conselho Nacional de Apoio a Pesquisa (CNPq), à Fundação de Amparo à Pesquisa Científica do Estado do Rio de Janeiro (FAPERJ), e Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES) pelo apoio financeiro ao projeto.

A todos os autores do artigo que relata a construção do mapa físico e genético do *Gluconacetobacter diazotrophicus* PAL5 incluso nesta tese, devido as suas valiosas contribuições no desenvolvimento deste estudo.

Aos meus orientadores Prof. Dr. Orlando Bonifácio Martins e Prof. Dr. Paulo Cavalcanti Gomes Ferreira, pela seriedade que conduzem a pesquisa científica em seus laboratórios, e pela ampliação dos meus conhecimentos nos campos da Bioquímica, Microbiologia, Biologia Molecular e Genômica estrutural. E ainda pelo valioso empenho no processo de orientação de todos os seus alunos, sejam eles pertencentes ao curso técnico, graduação, mestrado ou doutorado; e por seus valiosos comentários sobre este manuscrito, durante as fases de experimentação e escrituração.

Ao Professor Rodolpho Mattos Albano – Departamento de Bioquímica – UERJ, pelo auxílio nas reações de sequenciamento de DNA, hibridização com sondas radioativas e demais comentários realizados acerca dos dados contidos neste manuscrito.

Aos professores Marcelo Rosado Fantappié e Marcius da Silva Almeida, ambos professores adjuntos do Instituto de Bioquímica Médica – UFRJ, pelas laboriosas revisões e valiosas sugestões proferidas acerca dos dados contidos neste manuscrito.

Aos Membros pertencentes à Banca Examinadora desta tese, composta pelo Prof. Dr. Franklin David Rumjanek (Instituto de Bioquímica Médica – UFRJ), Prof. Dra. Ana Coelho (Departamento de Genética – UFRJ), Prof. Dr. Rodolpho Mattos Albano (Departamento de Bioquímica – UERJ), Prof. Dr. Marcius da Silva Almeida (Instituto de Bioquímica Médica – UFRJ) e Prof. Dra. Vânia Lúcia Muniz de Pádua (Centro Universitário Estadual da Zona Oeste – UEZO), pelos valiosos comentários e sugestões proferidas sobre este manuscrito.

RESUMO

Foi construído um mapa físico e genético de macrorestrição do cromossomo do *Gluconacetobacter diazotrophicus* PAL5 através da metodologia de eletroforese em gel de campo pulsado (PFGE). Para execução de análises por esta metodologia, digestões simples e duplas do cromossomo foram realizadas com as enzimas de restrição *I-CeuI*, *SwaI* e *PacI*. Os fragmentos de DNA gerados foram posicionados no mapa físico, utilizando análises combinatórias dos resultados de hibridização, posição dos sítios de restrição de endonucleases e tamanho dos fragmentos DNA. *G. diazotrophicus* PAL5 possui 4 operons ribossomais em um cromossomo circular, com tamanho estimado de 4232 ± 47 Kb. Os resultados dos experimentos de hibridização levaram ao posicionamento de 42 marcadores genéticos neste cromossomo, e subsequente alinhamento de *contigs* do Projeto Riogene (97,6% da cobertura do genoma), demonstrando o valor da construção de mapas físicos, no processo de finalização de projetos de seqüenciamento genômico total executados pela metodologia de *shotgun*. No decurso do processo de anotação do genoma do *G. diazotrophicus* PAL5, foram observados 2 *contigs* com 38,8 Kb e 16,6 Kb, os quais foram classificados como plasmídeos A e B, respectivamente, com base nos resultados de anotação, hibridização e *nested* PCR. Este manuscrito também relata as características gerais do genoma deste microrganismo, tais como: composição do genoma, percentual de GC, número de ORFS, percentual de seqüências codantes e distribuição gênica por categorias e subcategorias funcionais. Além disso, também são descritas algumas particularidades encontradas neste genoma, tais como: ausência de algumas enzimas relacionadas com metabolismo de glicose, genes relacionados com produção de bacteriocina e exopolissacarídeos, organização e posicionamento de grupos gênicos relacionados com quimiotaxia, motilidade celular e fixação biológica de nitrogênio.

Palavras-Chave: *Gluconacetobacter diazotrophicus*; Mapeamento físico; PFGE; Sequenciamento genômico; Fechamento de Gaps.

ABSTRACT

A macrorestriction physical and genetic map of *Gluconacetobacter diazotrophicus* PAL5 chromosome was constructed through of pulsed field gel electrophoresis (PFGE) methodology. For analysis execution by this methodology, single and double digestions of the chromosome were carried out with *I-CeuI*, *SwaI* and *PacI* restriction enzymes. The generated DNA fragments were positioned into physical map, using combinatory analysis of hybridization results, restriction enzymes sites position and DNA fragments length. *G. diazotrophicus* has 4 ribosomal operons into a circular chromosome, with estimated length of 4232 ± 47 Kb. The hybridization results allow the positioning of 42 genetic markers on chromosome, and subsequently alignment of contigs from Riogene Project (97,6% of genome coverage), demonstrating the value of physical map construction, in finishing processes of whole genome shotgun sequencing projects. During annotation process of the *G. diazotrophicus* PAL5 genome, 2 contigs were observed with 38.8 Kb and 16.6 Kb, which were classified as plasmids A and B, respectively, on the basis of results from genomic annotation, hybridization and nested PCR. This manuscript also related the general characteristics of the genome from this endophyte, such as: genome composition, GC percentage, ORFs number, coding sequence percentage and gene distribution by functional categories and subcategories. Furthermore it is described some particularities encountered in this genome, such as: absence of some enzymes related with glucose metabolism, genes related with bacteriocin and exopolissacharide production, organization and positioning of gene clusters related to chemotaxis, cellular motility and biological nitrogen fixation.

Key-Words: *Gluconacetobacter diazotrophicus*; Physical mapping; PFGE; Genomic sequencing; Gap closure.

SUMÁRIO

Índice de Figuras.....	12
Índice de Tabelas.....	16
Lista de Abreviaturas.....	17
1 – Introdução.....	20
1.1 – A Fixação Biológica de Nitrogênio (FBN).....	20
1.2 – O <i>Gluconacetobacter diazotrophicus</i>	23
1.3 – Potencial biotecnológico do <i>Gluconacetobacter diazotrophicus</i>	29
1.4 – Mapeamento Físico de Genomas.....	33
1.5 – O seqüenciamento genômico e o seu atual <i>status</i> de desenvolvimento.....	36
1.6 – A metodologia de <i>shotgun</i>	38
1.7 – Montagem do genoma e processo de finalização.....	40
1.8 – Anotação de genomas.....	42
1.9 – Considerações gerais desta obra.....	44
2 – Objetivos.....	46
2.1 – Objetivos gerais.....	46
2.2 – Objetivos específicos.....	46
3 – Material e Métodos.....	47
3.1 – Cepas bacterianas e condições de crescimento.....	47
3.2 – Preparação de DNA genômico de alto peso molecular para PFGE	48
3.3 – Digestão dos blocos de PFGE contendo DNA genômico.....	49
3.4 – Eletroforese em Gel de Campo Pulsado (PFGE).....	49
3.5 – Transferência de DNA fracionado em gel de agarose para membranas de nylon através de capilaridade (<i>Southern Blotting</i>).....	51
3.6 – Seleção de clones contendo genes de cópia única e RNAs ribossomais.....	52

3.7 – Construção de Sondas Radioativas de DNA.....	52
3.8 – Experimentos de hibridização com sondas radioativas	53
3.9 – Construção da biblioteca genômica de Cromossomos Artificiais Bacterianos (BACs).....	54
3.10 – Crescimento dos clones da biblioteca de BACs, extração e quantificação do DNA.....	55
3.11 - Digestão, fracionamento e documentação dos perfis de restrição dos BACs.....	55
3.12 – Análise das imagens geradas a partir dos géis de digestão dupla dos BACs.....	56
3.13 – Seleção de BACs para reações de sequenciamento e hibridizações.....	56
3.14 – Anotação das ORFs de <i>G. diazotrophicus</i> PAL5 e mineração de dados obtidos na anotação.....	57
4 – Resultados.....	60
4.1 – Digestões do cromossomo do <i>G. diazotrophicus</i> PAL5 e perfis de PFGE.....	60
4.2 – Determinação do número e ordem dos genes de RNAs ribossomais.....	64
4.3 – Hibridizações com sondas de genes de cópia única.....	64
4.4 – Construção do mapa físico do cromossomo do <i>G. diazotrophicus</i> PAL5.....	83
4.5 – Alinhamento de <i>contigs</i> do Projeto Riogene no cromossomo do <i>G. diazotrophicus</i> PAL5.....	83
4.6 – Presença de plasmídeos no genoma do <i>G. diazotrophicus</i> PAL5.....	85
4.7 – Mapeamento físico do genoma do <i>G. diazotrophicus</i> PAL5 utilizando Cromossomos Artificiais Bacterianos (BACs).....	95
4.8 – Características gerais do genoma do <i>G. diazotrophicus</i> PAL5.....	103
4.9 – Algumas particularidades do genoma do <i>G. diazotrophicus</i> PAL5	113
5 – Discussão.....	123

5.1 – Mapeamento físico do cromossomo do <i>G. diazotrophicus</i> PAL5 através de fragmentos de macrorestrição.....	123
5.2 – Detecção de 2 plasmídeos no genoma de <i>G. diazotrophicus</i> PAL5.....	128
5.3 – Mapeamento físico do cromossomo do <i>G. diazotrophicus</i> PAL5 através de Cromossomos Artificiais Bacterianos (BACs).....	130
5.4 – Anotação do genoma do <i>G. diazotrophicus</i> PAL5.....	132
6 – Conclusões.....	137
7 – Referências bibliográficas.....	141
8 – Anexos.....	166

ÍNDICE DE FIGURAS

Figuras		Páginas
Figura 1	Fotomicrografias do <i>Gluconacetobacter diazotrophicus</i> PAL5.....	24
Figura 2	Perfis de PFGE do DNA genômico de <i>Gluconacetobacter diazotrophicus</i> PAL5 observado em 4 diferentes condições de eletroforese para detecção de todos os fragmentos de DNA.....	61
Figura 3	Hibridizações obtidas com sondas dos genes dos RNAs ribossomais, contra perfil de PFGE do cromossomo do <i>Gluconacetobacter diazotrophicus</i> PAL5 gerado com a enzima <i>I-CeuI</i>	68
Figura 4	Hibridizações obtidas com sondas dos genes de cópia única <i>hemE</i> , <i>hemH</i> , <i>kdpC</i> , <i>mfs</i> e <i>pyrB</i>	70
Figura 5	Hibridizações obtidas com sonda do gene de cópia única <i>dhbF</i>	71
Figura 6	Hibridizações obtidas com sondas dos genes de cópia única <i>lig3</i> , <i>mexI</i> e <i>mgtA</i>	72
Figura 7	Hibridizações obtidas com sondas dos genes de cópia única <i>lsdA</i> , <i>nifA</i> e <i>pilT</i>	73
Figura 8	Hibridizações obtidas com sondas dos genes de cópia única <i>fumC</i> , <i>hutH</i> , <i>kup</i> e <i>ubiE</i>	74
Figura 9	Hibridizações obtidas com sondas dos genes de cópia única <i>araE</i> , <i>araJ</i> , <i>clpP</i> , <i>csp</i> , <i>fcuA</i> , <i>hlyD</i> , <i>ilvD</i> , <i>lhr</i> e <i>xseA</i>	75
Figura 10	Hibridizações obtidas com sondas dos genes de cópia única <i>dnaE</i> , <i>gntP</i> , <i>mcpA</i> , <i>pepC</i> e <i>sucD</i>	76
Figura 11	Hibridizações obtidas com sondas dos genes de cópia única <i>acoC</i> , <i>pdhA</i> , <i>trpA</i> , <i>xdhA</i> e <i>zntA</i>	77
Figura 12	Hibridizações obtidas com sondas dos genes de cópia única <i>gltB</i> e	

	<i>nuoG</i>	78
Figura 13	Hibridizações obtidas com sondas dos genes de cópia única <i>metE</i> e <i>xylF</i>	79
Figura 14	Hibridizações obtidas com sondas dos genes ribossomais <i>rriI</i> e <i>rriS</i>	80
Figura 15	Hibridizações obtidas com sonda do gene ribossomal <i>rriF</i>	81
Figura 16	Mapa físico do cromossomo do <i>Gluconacetobacter diazotrophicus</i> PAL5, utilizando as enzimas de restrição <i>I-CeuI</i> , <i>PacI</i> e <i>SwaI</i>	84
Figura 17	Alinhamento de 16 <i>contigs</i> gerados no Projeto Riogene (maio de 2006), sobre o mapa físico do cromossomo do <i>Gluconacetobacter diazotrophicus</i> PAL5, com base nos resultados de hibridização.....	86
Figura 18	Sobreposição de alinhamentos dos <i>contigs</i> gerados no Projeto Riogene nas montagens de Maio e Dezembro de 2006, em relação ao mapa físico do cromossomo do <i>Gluconacetobacter diazotrophicus</i> PAL5.....	87
Figura 19	Hibridizações obtidas com sonda de região de cópia única do <i>contig</i> 12 (38,8 Kb), contra diferentes tipos de membranas contendo DNA genômico.....	89
Figura 20	Hibridizações obtidas com sonda de região de cópia única do <i>contig</i> 16 (16,6 Kb), contra diferentes tipos de membranas contendo DNA genômico.....	90
Figura 21	Gel de agarose ilustrativo dos 1315 perfis de restrição dos BACs, submetidos a digestões duplas com as enzimas de restrição <i>HindIII</i> and <i>BamHI</i>	96
Figura 22	Ilustração de 1 segmento de 1 mapa físico de BACs gerado no software FPC, após processamento de géis no <i>software Image 3.0i</i>	97
Figura 23	Representação esquemática dos 117 BACs selecionados a partir dos	

	mapas físicos gerados no <i>software</i> FPC, exibindo sua respectivas coberturas ao longo do genoma do <i>G. diazotrophicus</i> PAL5.....	98
Figura 24	Perfis de Restrição de 7 BACs que hibridizaram contra sondas radioativas dos genes de RNAr 16s e 23s do <i>G. diazotrophicus</i> PAL5.....	100
Figura 25	Ilustração de um segmento do mapa físico de BACs, exibindo o alinhamento de <i>contigs</i> da montagem de novembro de 2004, baseado nas sequências de DNA geradas na <i>Clemson University Genomic Institute</i>	101
Figura 26	Representação esquemática do alinhamento de 7 <i>contigs</i> (Blastn) da montagem de Novembro de 2004, baseado em 14 seqüências de DNA oriundas de 7 BACs, geradas na <i>Clemson University Genomic Institute</i>	102
Figura 27	Distribuição dos genes detectados no genoma do <i>Gluconacetobacter diazotrophicus</i> PAL5 de acordo com a sua classificação funcional.....	106
Figura 28	Distribuição dos genes detectados no genoma do <i>Gluconacetobacter diazotrophicus</i> PAL5 de acordo com sua classificação funcional, considerando os alinhamentos com <i>e-values</i> inferiores a 10^{-3} (Blastp) e presença de domínios protéicos conservados (PFAM)	112
Figura 29	Representação gráfica do metabolismo de glicose, evidenciando enzimas que estão ausentes no genoma do <i>Gluconacetobacter diazotrophicus</i> PAL5.....	114
Figura 30	Posicionamento de alguns genes e grupos gênicos analisados em maiores detalhes, em relação a 6 <i>contigs</i> gerados na montagem de Dezembro de 2006, alinhados no mapa físico do cromossomo do <i>Gluconacetobacter diazotrophicus</i> PAL5.....	116
Figura 31	Página do Sistema Bionotes exibindo propriedades da ORF codificante da bacteriocina do <i>Gluconacetobacter diazotrophicus</i> PAL5.....	117

Figura 32	Alinhamento da sequência de 274 aminoácidos da bacteriocina de <i>Gluconacetobacter diazotrophicus</i> PAL5, em relação a bacteriocina Linocina M18 de <i>Rhodospirillum rubrum</i>	118
Figura 33	Organização de 1 grupo gênico relacionado com quimiotaxia do <i>Gluconacetobacter diazotrophicus</i> PAL5, em relação a grupos gênicos de quimiotaxia do <i>Rhizobium etli</i> e <i>Caulobacter crescentus</i>	121
Figura 34	Organização do grupo de genes <i>nif-fix</i> do <i>Gluconacetobacter diazotrophicus</i> PAL5, em relação aos grupos de genes <i>nif-fix</i> do <i>Bradyrhizobium japonicum</i> USDA110 e <i>Rhizobium sp.</i> (cepa NGR 234)...	122

ÍNDICE DE TABELAS

Tabelas		Páginas
Tabela 1	Número e tamanho estimado dos fragmentos de DNA, gerados através de digestões simples e duplas do cromossomo de <i>Gluconacetobacter diazotrophicus</i> PAL5.....	62
Tabela 2	Descrição das sondas de DNA dos genes de cópia única e ribossomais, utilizadas para construção do mapa físico do cromossomo do <i>Gluconacetobacter diazotrophicus</i> PAL5.....	65
Tabela 3	Resultados obtidos nos experimentos de hibridização com sondas dos genes de cópia única e ribossomais nos 6 diferentes perfis de PFGE.....	82
Tabela 4	Resultados de anotação do <i>contig</i> 12 (38,8 Kb) realizado no sistema Bionotes, pelos membros integrantes do Projeto Riogene.....	91
Tabela 5	Resultados de anotação do <i>contig</i> 16 (16,6 Kb) realizado no sistema Bionotes, pelos membros integrantes do Projeto Riogene.....	94
Tabela 6	Características gerais do genoma do <i>Gluconacetobacter diazotrophicus</i> PAL5.....	104
Tabela 7	Classificação funcional dos genes detectados no genoma do <i>Gluconacetobacter diazotrophicus</i> PAL5.....	105
Tabela 8	Distribuição dos genes anotados pelos membros integrantes do Projeto Riogene no sistema Bionotes, subdivido nas categorias e subcategorias funcionais mais freqüentes.....	107
Tabela 9	Número e percentual de ORFs detectadas no genoma do <i>Gluconacetobacter diazotrophicus</i> PAL5, inicialmente classificadas como proteínas hipotéticas, e posteriormente classificadas funcionalmente.....	111

LISTA DE ABREVIATURAS

Alfresco	Programa de comparação de genomas
Artemis	Sistema de anotação de Biossêquências
ATP	Adenosina 5' trifosfato
BACs	Cromossomos artificiais bacterianos
BLAST	Ferramenta de alinhamento de biossêquências
BSA	Soroalbumina bovina
CAP3	Programa de montagem de seqüências de DNA
cDNA-AFLP	Polimorfismo de tamanhos de fragmentos amplificados a partir de cDNA
COG	Agrupamento de grupos ortólogos
dATP	Desoxiadenosina trifosfato
dCTP	Desoxicitosina trifosfato
DNA	Ácido desoxirribonucléico
DO _{600 nm}	Densidade ótica na faixa de luz visível a 600 nanômetros
EC	Categoria enzimática
EDTA	Ácido etilenodiaminotetracético
EMBL	Banco de dados de armazenamento de seqüências de nucleotídeos
ESP	EDTA – Sarkosil – Proteinase K
FBN	Fixação biológica de nitrogênio
FPC	Programa de montagem de <i>contigs</i> baseado em perfis de restrição de BACs
GDB	Banco de dados de armazenamento de seqüências de nucleotídeos
Glimmer	Programa de predição de genes
GSDB	Banco de dados de armazenamento de seqüências de nucleotídeos
HCl	Ácido clorídrico
ISs	Seqüências de inserção

ITS/RFLP	Polimorfismo de tamanhos de fragmentos de restrição de regiões intergênicas
Kb	Quilobases
K ₂ HPO ₄	Fosfato de potássio
LB	Caldo Luria-Bertani
LB-X-gal-IPTG	Meio de cultura Luria-Bertani acrescido de 5-bromo-4-cloro-3-indolil-β-D-galactopiranosida e isopropil-β-tiogalactopiranosida
Mb	Megabases
MgSO ₄	Sulfato de magnésio
MLEE	Eletroforese enzimática multilocus
mM	Milimolar
Modeller	Algoritmo de predição de estruturas proteicas
N ₂	Gás dinitrogênio
NaCl	Cloreto de sódio
NAD-GDH	Glicose desidrogenase ligada a nicotinamida adenosina dinucleotídeo
NaOH	Hidróxido de sódio
NH ₄ ⁺	Íon amônio
NO ₃ ⁻	Íon nitrato
nt	Nucleotídeo
ORFfinder	Ferramenta de predição de genes
ORFs	Fases de leitura aberta
pb	Pares de base
PCR	Reação de polimerase em cadeia
PDB	Banco de dados de estruturas terciárias de proteínas
PFAM	Banco de dados de seqüências de estruturas de domínios protéicos longos

PFGE	Eletoforese em gel de campo pulsado
pH	Potencial de hidrogênio
PHYLP	Pacote de análise filogenética
PIR	Banco de dados de seqüências de aminoácidos
PMSF	Fenilmetilsulfonil fluorido
PQQ-GDH	Pirroquinolina quinona ligada a uma glicose desidrogenase
PROSITE	Banco de dados de seqüências protéicas conservadas associadas com funções específicas em um organismo
RBSfinder	Ferramenta de localização de sítios de ligação ribossomais
RNA	Ácido ribonucléico
RNAr	Ácido ribonucleico ribossomal
RNase	Ribonuclease
rpm	Rotação por minuto
SDS	Dodecil sulfato de sódio
SSC	Solução salina-citrato
SUCEST	Projeto de sequenciamento de ESTs de cana de açúcar
SWISS-PROT	Banco de dados suíço de seqüências de aminoácidos
TBE	Tris – ácido bórico – EDTA
THREADER	Ferramenta de predição de estruturas protéicas
Tris	Tris hidroximetil amino metano
U	Unidades
µg	Micrograma
µl	Microlitro

1 – INTRODUÇÃO

1.1 – A Fixação Biológica de Nitrogênio (FBN)

O nitrogênio é considerado o principal macronutriente limitante, para a maioria das espécies vegetais, pois estas requerem nitrogênio do solo ou da atmosfera, para o seu crescimento e acumulação de biomassa (Gastal & Lemaire, 2002; Luciński, Polcyn & Ratajczak, 2002). O nitrogênio derivado do solo é absorvido pelas plantas principalmente como íon nitrato (NO_3^-) ou íon amônio (NH_4^+), e em menor quantidade, através de pequenas moléculas orgânicas, tais como aminoácidos (Gastal & Lemaire, 2002; Luciński, Polcyn & Ratajczak, 2002).

Os íons nitrato e amônio são extremamente instáveis no solo, podendo ser transformados em formas não absorvíveis ao vegetal ou serem lixiviadas, indicando que a maior parte do nitrogênio utilizado pela planta venha da atmosfera (Raven, Handley & Andrews, 2004). Em contraste, nas últimas cinco décadas, a utilização de fertilizantes nitrogenados nos cultivos, tem aumentado mais de 20 vezes no mundo, sendo considerados geralmente ineficientes, pois somente cerca de 50% é utilizado pelo vegetal (Mifflin & Habash, 2001).

O nitrogênio derivado da atmosfera (gás dinitrogênio - N_2), não é utilizado diretamente no metabolismo vegetal, portanto, ele deve ser fixado na forma de outra molécula que seja utilizável, através de um processo conhecido como Fixação Biológica de Nitrogênio (FBN), que é realizada por alguns procariontes conhecidos como microrganismos diazotróficos, capazes de converter N_2 a NH_4^+ . Nesta simbiose, o vegetal fornece fontes de carbono ao microrganismo e em troca recebe nitrogênio fixado na forma de NH_4^+ e aminoácidos (Barea *et al.*, 2005; Oldroyd, Harrison & Udvardi, 2005). Sob este aspecto a FBN é extremamente vantajosa, pois não causa impactos ambientais, reduz custos de cultivo, e apresenta alta eficiência de utilização de nitrogênio (Gastal & Lemaire, 2002; Jeuffroy, Ney & Ourry, 2002).

A FBN é um processo que demanda uma alta quantidade de energia, e é catalisada por uma metaloenzima de duplo componente chamada nitrogenase, que é composta por duas subunidades proteicas, a Fe-proteína (dinitrogenase redutase) e a MoFe-proteína (dinitrogenase), que apresentam alta homologia entre diferentes microrganismos diazotróficos (Ureta & Nordlund, 2002; Fischer & Newton, 2005; Kaiser *et al.*, 2005). A síntese e a atividade da nitrogenase requer a expressão de diversos genes *nif* específicos. No fim da década passada, o maior grupo de genes *nif* de *G. diazotrophicus* foi clonado e seqüenciado, sendo que dentre estes genes, os mais estudados foram os genes *nifA* (codificante do ativador transcricional NifA), e os genes *nifHDK* (codificantes das Fe- e MoFe-proteínas que compõe a nitrogenase) (Sevilla *et al.*, 1997; Teixeira *et al.* 1999; Perlova *et al.*, 2003).

A atividade da nitrogenase (redução do N_2 a NH_4^+) em microrganismos diazotróficos aeróbios, requer quantidades substanciais de ATP derivados da respiração aeróbia, entretanto a enzima nitrogenase é sensível ao O_2 . A proteção da nitrogenase pode ocorrer de 3 diferentes maneiras: 1- formação de nódulos nas raízes, que sabidamente correspondem a estruturas especializadas na regulação da quantidade de O_2 disponível ao microrganismo diazotrófico (Dong *et al.*, 2002); 2- proteção conformacional da enzima nitrogenase; 3- controle da taxa de respiração celular, com finalidade de produzir uma concentração de oxigênio, compatível com a atividade da nitrogenase (Ureta & Nordlung 2002).

Os estudos sobre FBN em Gramíneas no Brasil, iniciaram-se a aproximadamente 40 anos, sob a liderança da pesquisadora Johanna Döbereiner (Embrapa – Agrobiologia – RJ), período este em que ocorreu a descoberta de diversas bactérias fixadoras de nitrogênio atmosférico, em vários países de climas tropicais e subtropicais, especialmente no Brasil, incluindo espécies rizosféricas (*Beijerinckia fluminensis* e *Azotobacter paspali*), associativas (*Azospirillum lipoferum*, *Azospirillum brasilense*, *Azospirillum amazonense*) e endofíticas (*Herbaspirillum seropedicae*, *Herbaspirillum rubrisubalbicans*, *Gluconacetobacter*

diazotrophicus, *Burkholderia brasilensis* e *Burkholderia tropica*), os quais são considerados os principais microrganismos de interesse para grupos de pesquisa, que estudam a associação de microrganismos diazotróficos com vegetais (Baldani *et al.*, 1997; Baldani & Baldani, 2005).

Dentre as espécies supracitadas, devemos destacar as diazotróficas endofíticas, pois geralmente estabelecem associações com diversas espécies e genótipos vegetais, além de fixarem nitrogênio mais eficientemente que as diazotróficas rizosféricas ou associativas, pois estabelecem contato direto com os tecidos vegetais (James *et al.*, 2001; Munõz-Rojas, Fuentes-Ramírez & Caballero-Mellado, 2005; Muthukumarasamy *et al.*, 2006). Apesar destas espécies apresentarem alto potencial de implementação agrícola, a primeira espécie bacteriana fixadora de nitrogênio com características endofíticas (*H. seropedicae*), foi descoberta somente em 1984 por Baldani e colaboradores, a partir de raízes lavadas e com superfícies esterilizadas de milho, sorgo e arroz (Baldani, 1984; Baldani & Baldani, 2005).

A partir desta descoberta até o ano de 1990, ocorreu uma intensificação das pesquisas acerca da colonização de tecidos vegetais por bactérias diazotróficas, e iniciaram-se estudos direcionados para a elucidação de aspectos relacionados com interações planta-bactéria, além de ocorrer em 1988, o isolamento de outro importante endófito fixador de nitrogênio denominado *G. diazotrophicus*, que foi isolado por Cavalcante & Döbereiner, a partir de estudos em tecidos internos de colmos e raízes de cana de açúcar (Cavalcante & Döbereiner, 1988; Baldani & Baldani, 2005).

Posteriormente, novas bactérias endofíticas fixadoras de nitrogênio foram identificadas por diferentes grupos de pesquisa, cujas principais espécies foram *Herbaspirillum rubrisubalbicans* (Baldani *et al.*, 1996), *Herbaspirillum frisingense* (Kirchhof *et al.*, 2001), *Azospirillum dobereineriae* (Eckert *et al.*, 2001), *Gluconacetobacter johannae* e *Gluconacetobacter azotocaptans* (Fuentes-Ramírez *et al.*, 2001) e *Burkholderia tropica* (Reis

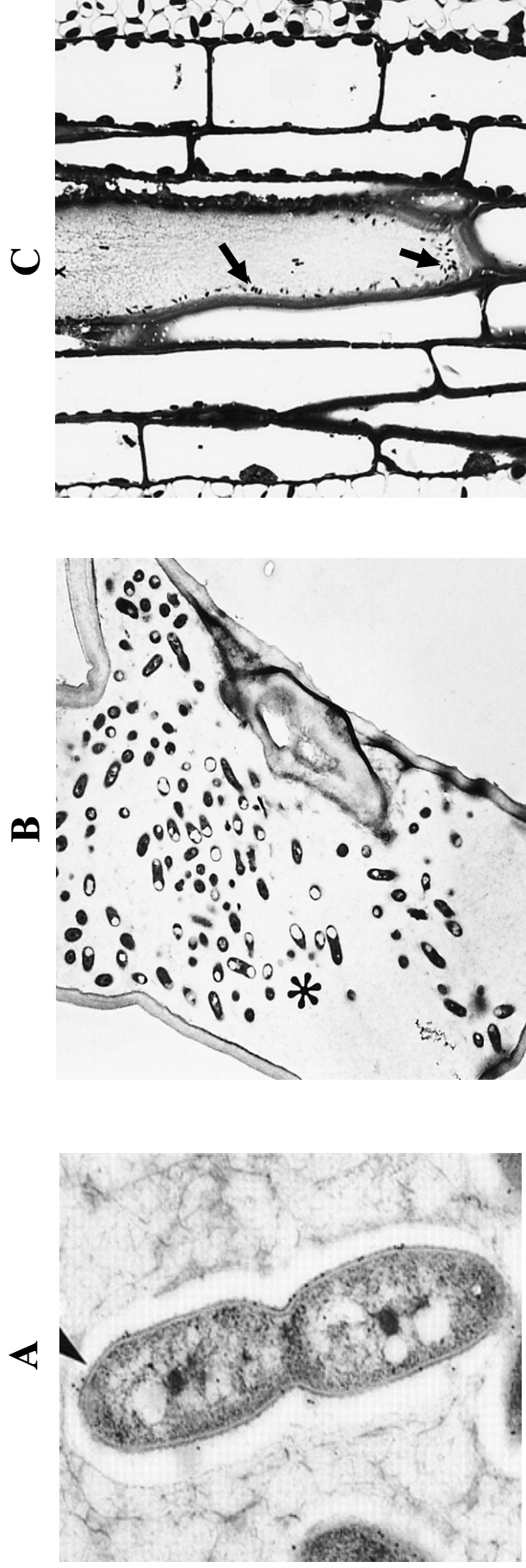
et al., 2004), as quais estão inclusas nos principais gêneros bacterianos de interesse para grupos de pesquisa em FBN, que visam demonstrar a sua importância econômica, assim como o seu importante papel no gerenciamento integrado de nutrientes, na produção agrícola sustentável e proteção ambiental (Loiret *et al.*, 2004; Baldani & Baldani, 2005; Muthukumarasamy *et al.*, 2006).

1.2 – O *Gluconacetobacter diazotrophicus*

Em 1988, Cavalcante & Döbereiner, isolaram a partir da cana de açúcar, uma bactéria capaz de fixar nitrogênio atmosférico com alta tolerância a acidez, que inicialmente foi denominada *Sacharobacter nitrocaptans*. Posteriormente, este microrganismo foi reclassificado como *Acetobacter diazotrophicus*, nova espécie dentro do gênero *Acetobacter sp.* com base em estudos fenotípicos, taxonômicos e de hibridização de DNA-rRNA e DNA-DNA (Gillis *et al.*, 1989). Atualmente, este microrganismo encontra-se incluso na classe das α -Proteobactérias, ordem *Rhodospirillales*, família *Acetobacteraceae* e gênero *Gluconacetobacter*, sendo denominado *Gluconacetobacter diazotrophicus*, com base em análise de seqüências parciais de rRNA 16S e características fenotípicas (Yamada, Hoshino & Ishikawa, 1997).

O *G. diazotrophicus* é um bacilo Gram negativo, reto e com extremidades arredondadas, que possui flagelos laterais ou peritríquios (Figura 1A), de metabolismo microaeróbico, que apresenta a capacidade de fixar nitrogênio atmosférico e produzir ácido acético a partir de etanol, além da capacidade de redução de acetileno. Suas colônias apresentam pigmentação marrom quando cultivadas em placas de cultura contendo Meio Batata (com alto teor de açúcar) e alaranjada em placas de cultura contendo Meio LGIP (com azul de bromotimol) (Döbereiner *et al.*, 1995). O *G. diazotrophicus* possui temperatura ótima de crescimento e Fixação de Biológica de Nitrogênio a 30 °C e o pH inicial de crescimento

Figura 1 - Fotomicrografias do *Gluconacetobacter diazotrophicus* PAL5



A – Microscopia eletrônica de transmissão exibindo a morfologia do *Gluconacetobacter diazotrophicus* PAL5; B – Microscopia eletrônica de transmissão exibindo o *Gluconacetobacter diazotrophicus* PAL5 no espaço intercelular (*) adjacente a uma cavidade sub-estomatal infectada C – Microscopia eletrônica de transmissão exibindo nas setas o *Gluconacetobacter diazotrophicus* PAL5 dentro dos vasos xilemáticos. Fonte: James *et al.*, 2001.

entre 5,5 e 6,0. Adicionalmente, cabe ressaltar que este microrganismo é incapaz de crescer a 37 °C (Gillis et al., 1989).

Este microrganismo é encontrado no espaço intercelular (Figura 1B), realizando associação endofítica em todas as estruturas da cana de açúcar, e com os mais variados genótipos deste vegetal, especialmente em áreas agrícolas onde a adição de fertilizantes nitrogenados nos cultivares é baixa (Gillis *et al.*, 1989; Reis, Olivares & Döbereiner, 1994; Fuentes-Ramírez *et al.*, 1999; Steenhouldt & Vanderleyden, 2000; Fischer & Newton, 2005).

Além da cana de açúcar, *G. diazotrophicus* tem sido detectado em associação com outras espécies vegetais, tais como: batata doce (*Ipomoea batatas*) (Döbereiner, Reis & Lazarini, 1988), café (*Coffea arabica*), nas gramíneas *Eleusine coracana* e *Pennisetum purpureum*, ananás (*Ananas comosus*) (Santos, Bustillos-Cristales & Caballero-Mellado, 2001), cenoura (*Daucus carota*), rabanete (*Raphanus sativus*) e beterraba (*Beta vulgaris*) (Munõz-Rojas, Fuentes-Ramírez & Caballero-Mellado, 2005).

Cabe ressaltar que esta habilidade de colonizar vegetais incluídos nas famílias *Poaceae*, *Convolvulaceae*, *Rubiaceae*, *Bromeliaceae*, *Apiaceae*, *Brassicaceae* e *Chenopodiaceae*, enfatiza o importante potencial benéfico do *G. diazotrophicus* para diferentes hospedeiros vegetais (Döbereiner, Reis & Lazarini, 1988; Reis, Olivares & Döbereiner, 1994; Santos, Bustillos-Cristales & Caballero-Mellado, 2001; Munõz-Rojas, Fuentes-Ramírez & Caballero-Mellado, 2005).

Apesar do *G. diazotrophicus* ser encontrado em associação com diversos tipos de vegetais, e em diferentes países de climas tropicais ou subtropicais, quando foram analisadas diferentes cepas bacterianas, oriundas de regiões geográficas distintas, através de experimentos de eletroforese enzimática multilocus (MLEE), ou ainda através da análise de cepas isoladas de variedades de cana de açúcar no Brasil, mantidas durante 30 anos em banco de germoplasma, e analisadas pela técnica de polimorfismo de tamanhos de fragmentos de

restrição de regiões intergênicas (ITS/RFLP), observou-se uma baixa variabilidade genética nesta espécie bacteriana (Caballero-Mellado & Martinez-Romero, 1994; Santos *et al.* 1999; Munõz-Rojas & Caballero-Mellado, 2003).

Adicionalmente, é comumente relatada a ocorrência de decréscimo populacional deste endófito, ao longo dos estágios de desenvolvimento do vegetal e através do aumento de concentrações de fontes de nitrogênio, entretanto não é bem conhecida a influência destes fatores na redução populacional. Além disso, alguns fatores abióticos, tais como desidratação e salinidade, também podem causar redução populacional, pois afetam a atividade de enzimas relacionadas com o catabolismo de glicose e ciclo dos ácidos tricarbóxicos, diminuindo a taxa de produção de ATP pelo microrganismo (James *et al.*, 2001; Munõz-Rojas & Caballero-Mellado, 2003; Tejera *et al.*, 2003; Baldani & Baldani, 2005).

Recentemente, Munõz-Rojas, Fuentes-Ramírez & Caballero-Mellado (2005), descreveram uma atividade antagonista em cepas de genótipos predominantes de *G. diazotrophicus*, contra cepas da mesma espécie e espécies estritamente relacionadas do gênero *Gluconacetobacter spp.*, incluindo *G. johannae*, *G. azotocaptans* e *G. liquefaciens*, onde foi detectado a produção de bacteriocinas, que consistem em peptídeos biologicamente ativos com atividade tóxica, produzidos e secretados por microrganismos contra cepas da mesma espécie ou espécies estritamente relacionadas, desempenhando um importante papel na colonização deste endófito em vegetais, pois favorece a colonização de cepas produtoras de bacteriocina em detrimento de cepas sensíveis, o que pode explicar o fato da variabilidade genética de *G. diazotrophicus* ser baixa, e ainda o fato de ocorrer redução populacional deste endófito, ao longo do desenvolvimento do vegetal colonizado.

Uma importante característica da interação entre vegetais e *G. diazotrophicus*, reside no fato deste endófito colonizar os espaços intercelulares da maioria dos órgãos vegetais, e promover o crescimento vegetal sem causar nenhum sintoma de doença (Baldani & Baldani,

2005). Estudos de expressão gênica de cana de açúcar durante a associação com *G. diazotrophicus*, utilizando diferentes abordagens tais como cDNA-AFLP *fingerprinting*, perfis transcricionais gerados pelo SUCEST (*Sugarcane EST Sequencing Project*) e microarranjos, sugeriram uma participação efetiva da cana de açúcar na associação, pois diversos genes envolvidos em diferentes processos fisiológicos vegetais, foram classificados como genes diferencialmente expressos durante a associação (Nogueira *et al.* 2001). Por isto há interesse na caracterização de vias sinalizadoras, pelas quais a cana de açúcar é capaz de decifrar os sinais bacterianos e respondê-los adequadamente para que ocorra o sucesso da associação (Vargas *et al.* 2003), além dos mecanismos moleculares que promovem o crescimento vegetal através da associação com endófitos (Baldani & Baldani, 2005).

O *G. diazotrophicus* é hábil para infectar e colonizar os tecidos radiculares e partes aéreas da cana de açúcar, entretanto, inicialmente este microrganismo coloniza a superfície das raízes, e posteriormente as infecta pelas junções das raízes laterais e/ou através de injúrias localizadas nas extremidades das raízes, geradas devido ao atrito com o solo durante o seu desenvolvimento. Após penetrar nas raízes, este microrganismo produz uma grande colonização do colmo da base do vegetal, e move-se para as partes aéreas utilizando os vasos xilemáticos (Figura 1C). Relativamente grandes populações deste endófito são encontradas nos tecidos de partes aéreas do vegetal, o que sugere estas partes serem os principais sítios para fixação de nitrogênio, devido ao baixo nível de oxigenação e energia disponível em forma de sacarose (James *et al.* 1994, 2001; Baldani & Baldani, 2005).

Este endófito é encontrado em número aproximado de 103 células/g de tecido fresco de cana de açúcar (Munõz-Rojas & Caballero-Mellado, 2003; Muthukumarasamy *et al.*, 2006), e aparentemente seu processo de colonização no vegetal é favorecido pela própria planta, pois recentemente Blanco e colaboradores (2005), detectaram através de preparações de glicoproteínas purificadas de cana de açúcar, marcadas com fluoresceína isotiocianato, que

estas glicoproteínas da superfície celular vegetal, interagem especificamente com receptores presentes na parede celular bacteriana, caracterizando a primeira etapa da discriminação biológica de um endófito simbiote compatível ao vegetal.

O *G. diazotrophicus* possui propriedades fisiológicas que o distingue de outros microrganismos diazotróficos, tais como a sua capacidade de crescer preservando a sua capacidade de fixar nitrogênio sob altas concentrações de sacarose (30%) e baixo pH (3,0). Contudo, a atividade da nitrogenase é parcialmente inibida por NH_4^+ em altas concentrações de sacarose, por via de um processo desconhecido que não envolve modificações covalentes (ADP-ribosilação) do seu componente Fe-proteína. Por outro lado, a nitrogenase, não é afetada quando altas concentrações de NO_3^- (25 mM) está presente no meio, pois não possui a enzima nitrato redutase (Baldani *et al.*, 1997; Reis & Döbereiner, 1998; Fischer & Newton, 2005). A manutenção da fixação de nitrogênio na presença de nitrato é um benefício adicional, devido ao fato do potencial de implementação *G. diazotrophicus* como agro-estimulante, em cultivares tratados previamente com fertilizantes à base de nitrato.

O *G. diazotrophicus* é capaz de crescer em altas concentrações de sacarose, que consiste na principal fonte de carbono em seu habitat natural, entretanto, este microrganismo não é capaz de captar diretamente este dissacarídeo para ser metabolizado. A capacidade de crescer em sacarose, deve-se a produção de uma enzima com atividade sacarolítica extracelular denominada levansucrase, que é secretada para o espaço periplasmático, com finalidade de disponibilizar glicose e frutose para o crescimento do endófito (Alvarez & Martinez-Drets, 1995; Tejera *et al.*, 2003; Arrieta *et al.* 2004; Baldani & Baldani, 2005).

A oxidação extracelular de glicose no espaço periplasmático, é considerada a principal rota de assimilação de glicose em *G. diazotrophicus*, onde ocorre a conversão de glicose para ácido glucônico, catalisada por uma enzima de membrana denominada glicose desidrogenase tipo II, que é composta pela associação da pirroloquinolina quinona com uma glicose

desidrogenase (PQQ-GDH). Neste organismo, o gluconato pode ser posteriormente metabolizado através do ramo oxidativo da via das pentoses-fosfato. Entretanto, é importante mencionar que o catabolismo de glicose também pode ocorrer em baixos níveis no ambiente intracelular, sendo neste caso catalisado por uma glicose desidrogenase ligada a nicotinamida adenosina dinucleotídeo (NAD-GDH) (Luna *et al.*, 2006). Adicionalmente, existem relatos de que o *G. diazotrophicus*, além da glicose e da frutose, é capaz de utilizar galactose, manose, inositol, arabinose, manitol, glicerol e etanol como fontes de carbono para o seu metabolismo (Tejera *et al.*, 2003; Loiret *et al.*, 2004).

Este endófito é reconhecido por possuir a via das pentoses-fosfato, constituindo a principal via operante no catabolismo de glicose 6-fosfato, em concordância com outros membros da família *Acetobacteraceae*, pois enzimas chave das vias Embden-Meyerhof-Parnas (6-fosfofrutoquinase e 1-fosfofrutoquinase) e Entner-Doudoroff (6-fosfogliconato desidratase e 2-ceto-3-deoxi-6-fosfogliconato aldolase) não são detectadas, através de análise de atividade enzimática, em extrato total de células de *G. diazotrophicus*. Adicionalmente, através desta metodologia, são detectadas todas as enzimas participantes do ciclo dos ácidos tricarbóxicos, assim como enzimas associadas com cadeia respiratória para oxidação total de glicose e cetogluconatos (Alvarez & Martinez-Drets, 1995; Luna *et al.*, 2002).

1.3 – Potencial biotecnológico do *Gluconacetobacter diazotrophicus*

O potencial biotecnológico do *G. diazotrophicus* é enfatizado pela demonstração que mais de 60% do nitrogênio, pode ser adquirido através da fixação biológica em variedades de cana de açúcar, reforçando sua importância para a agricultura sustentável no Brasil, onde baixas quantidades de fertilizantes nitrogenados são aplicadas nos cultivares (Lima, Boddey & Döbereiner, 1987; Urquiaga, Cruz & Boddey, 1992; Döbereiner, Baldani & Urquiaga, 1995; Baldani *et al.*, 1997; Baldani & Baldani, 2005). Devido a este fato, um entendimento

detalhado sobre os mecanismos pelos quais ocorre regulação de expressão dos genes *nif* em *G. diazotrophicus*, é de grande importância para exploração dos benefícios da FBN no cultivo da cana de açúcar, e no melhoramento desta interação, e possivelmente da interação deste endófito com outros vegetais economicamente importantes, assim como na exploração do potencial benéfico para ambas as partes envolvidas nesta associação planta-bactéria (Perlova, 2003).

Um exemplo da importância da FBN em nosso país, reside no fato de que o Brasil aplica as menores doses de fertilizantes nitrogenados nos cultivares (51 Kg ha^{-1}), quando comparado aos maiores usuários mundiais destes insumos (EUA – 105 Kg ha^{-1} e Europa – 335 Kg ha^{-1}), o que gera um grande impacto econômico, pois o nosso país é responsável por aproximadamente 30% do cultivo mundial de cana de açúcar. Além disso, o programa Pro-álcool produz em média 14 bilhões de litros de etanol por ano, o que permite reduzir o consumo de 260 mil barris de petróleo por dia. A nível ecológico, cabe ressaltar o fato deste combustível alternativo ser bem menos nocivo ao meio ambiente que os combustíveis fósseis (Dobereiner, Baldani & Urquiaga, 1995).

O Programa Pro-álcool foi desenvolvido com intuito de substituir em larga escala derivados do petróleo, evitando assim a dependência externa do produto. A partir de sua instauração (1975) até o ano de 2000, foram produzidos cerca de 5,6 milhões de veículos movidos a álcool hidratado, e também passou a haver um acréscimo de até 25% de álcool na gasolina consumida pelos veículos nacionais, o que ocasionou a redução na importação de aproximadamente 550 milhões de barris de petróleo, e ainda proporcionou uma economia na ordem de 11,5 bilhões de dólares aos cofres públicos (Biodieselbr Online, 2006).

Entretanto, no final da década de 80, ocorreu uma fase de estagnação do programa, motivado pelos baixos preços do petróleo no cenário mundial, crise de abastecimento de álcool, e redução de estímulo de utilização deste combustível alternativo, afetando assim a sua

credibilidade, ocasionando um significativo decréscimo da demanda, e conseqüentemente, das vendas de automóveis movidos a álcool. Atualmente, está ocorrendo no Brasil, uma nova expansão dos canaviais do interior paulista e do nordeste, além da ampliação e construção de 40 novas usinas, que pretendem gerar cerca de 360 mil novos empregos diretos, e 900 mil indiretos pela iniciativa privada, pois se acredita que o álcool terá um papel cada vez mais importante como combustível, no Brasil e no mundo, motivado principalmente pela tecnologia dos motores *flex fuel*, que hoje já é oferecido para quase todos os modelos de automóveis, e representaram cerca de 49,5% do total de vendas (Biodieselbr Online, 2006).

Adicionalmente, convém ressaltar que a mais de duas décadas é conhecida a eficácia da FBN realizada pelo *G. diazotrophicus*, onde foi verificado que a quantidade de nitrogênio fixada biologicamente (139×10^6 a 170×10^6 toneladas de nitrogênio por ano), é equivalente ao dobro da quantidade fixada com adubação, utilizando fertilizantes nitrogenados (Peoples & Craswell, 1992), sem levar em conta o fato deste tipo de adubação ser extremamente dispendiosa (cerca de 150 milhões de dólares ao ano para o governo brasileiro) e tóxica para o meio ambiente e manipuladores (Döbereiner, 1992).

Recentemente Fischer & Newton (2005), relataram que na associação com a cana de açúcar, o *G. diazotrophicus* pode fixar mais de 150 Kg de nitrogênio por hectare durante o desenvolvimento do vegetal, o que representa cerca de 50% do total de nitrogênio requerido para a colheita, corroborando com os dados de Muthukumarasamy e colaboradores (2006), que relataram a estabilidade da biomassa na colheita, mesmo ocorrendo redução de 50% (140 Kg/ha) ou total privação de fertilizantes nitrogenados nos cultivares, sinalizando a possibilidade de redução de custos no cultivo de cana de açúcar através da inoculação do *G. diazotrophicus* nas plantações.

A habilidade do *G. diazotrophicus* aumentar a taxa de crescimento da cana de açúcar tem sido largamente documentada, e este benefício tem sido associado em parte com a

transferência do nitrogênio bacteriológicamente fixado ao vegetal, e ainda com a produção de fatores estimulantes do crescimento vegetal, destacando-se a produção dos fitohormônios auxina e giberilina, que tem sido amplamente detectado em culturas deste endófito (Lee *et al.*, 2000; Muthukumarasamy *et al.*, 2002; Munõz-Rojas & Caballero-Mellado, 2003; Lee *et al.*, 2004; Muthukumarasamy *et al.*, 2006). Entretanto, o primeiro indício da produção de fitohormônios por *G. diazotrophicus* foi demonstrado por Sevilla e colaboradores (2001), através da utilização de mutantes *nif*⁻ (deficientes no processo de FBN), que foram capazes de estimular o crescimento vegetal em experimentos onde o nitrogênio não foi o fator limitante, levantando-se a hipótese da produção destes fatores de crescimento vegetal.

Adicionalmente, cabe ressaltar que a biossíntese de fitohormônios não é uma característica exclusiva de vegetais superiores, pois algumas espécies bacterianas, fungos e algas podem produzi-los, e induzir efeitos marcantes no desenvolvimento e crescimento vegetal (Lee *et al.*, 2004). Em experimentos de inoculação de diferentes genótipos de cana de açúcar micropropagadas, com cepas selecionadas de *G. diazotrophicus*, ocorreram diversas alterações anatômicas e fisiológicas na planta hospedeira. Entre as modificações anatômicas ocorridas, podemos destacar o aumento de sítios mitóticos nas raízes laterais, a emersão de raízes laterais com alteração na geometria das raízes, através do aumento da porção de raízes finas e do sistema radicular total, e entre as alterações fisiológicas, destaca-se o aumento da acumulação de nutrientes, da biomassa vegetal, e da atividade das H⁺-ATPases da fração microsomal das células radiculares, assim como os seus respectivos conteúdos proteicos (Olivares *et al.*, 2002; Baldani & Baldani, 2005).

O *G. diazotrophicus* também apresenta importância biotecnológica relacionada a processos de controle biológico, devido a sua atividade antagonista contra *Xanthomonas albilineans*, como resultado da produção de uma bacteriocina similar a lisozima que é capaz de inibir o crescimento deste fitopatógeno de cana de açúcar, responsável pela doença da

folha escaldada, possivelmente através da destruição da parede celular da cepa patogênica (Piñon *et al.*, 2002; Blanco *et al.*, 2005).

Além disso, o *G. diazotrophicus* fermenta açúcares, diminuindo o pH do meio para abaixo de 3.0, o que conseqüentemente gera uma atividade antagonista contra um importante fitopatógeno da cana de açúcar, denominado *Colletotrichum falcatum* (Muthukumarasamy, Revathi & Vadivelu, 2000). E devido as suas características endofíticas, o *G. diazotrophicus*, tem sido utilizado como vetor de expressão de genes heterólogos de interesse, como por exemplo, os genes *cry3A* e *cry1Ab* isolados de *Bacillus thuringiensis*, que são responsáveis pela atividade entomopatogênica contra insetos coleópteros e lepdópteros, causadores de danos para a cana de açúcar (Baldani & Baldani, 2005).

1.4 – Mapeamento Físico de Genomas

Mapas físicos são considerados poderosas ferramentas para localização e isolamento de genes, estudos de organização de genomas, e como fase preliminar para um seqüenciamento genômico eficiente (Heber, Hoheisel & Vingron, 2000). Em especial, os mapas físicos gerados por eletroforese em gel de campo pulsado (PFGE) de grandes fragmentos de DNA, associado com experimentos de hibridização (Majumder *et al.*, 1996; Marcone & Seemüller, 2001), e mapas físicos gerados a partir de digestões de cromossomos artificiais bacterianos (BACs) com endonucleases, associado com técnicas de seqüenciamento e hibridização (Tomkins *et al.*, 2001; Berent & Messick, 2003; Goto *et al.*, 2003; Battistoni *et al.*, 2005), os quais correspondem a alternativas para acessar a organização genômica de organismos, cujo suas seqüências genômicas completas, não estão disponíveis em bancos de dados.

Mapeamento de fragmentos de macrorestrição tem sido aplicado para examinar a estrutura e a topologia de genomas bacterianos por muitos anos, fornecendo valiosas

informações para elucidação de questões relacionadas à regulação global do metabolismo, rearranjos genômicos, e ainda dinâmica de aspectos evolucionários da organização genômica (Hesekorn, 1989; Kuepel, Piletier & Hill, 1990; Riley & Sanderson, 1990). Além de serem eficazes na estimação do tamanho de genomas, e na detecção de aberrações grosseiras na estrutura do cromossomo (Majumder *et al.*, 1996; Weinel *et al.*, 2001).

Mapeamento físico utilizando BACs digeridos com endonucleases, associado com técnicas de sequenciamento e hibridização, tem sido empregado no decurso de projetos de sequenciamento genômico com finalidade de determinar o posicionamento de marcadores moleculares na estrutura do genoma de diversos organismos, assim como auxiliar na montagem, ordenação e orientação de *contigs* durante o processo de finalização de sequenciamentos genômicos (Tao *et al.*, 2001; Battistoni *et al.*, 2005).

Mapeamentos físicos proporcionam medições diretas do conteúdo genômico de vários microorganismos, onde os menores genomas foram encontrados em *Mycoplasma sp.* (600 – 1200 Kb), seguidos de organismos incluídos nos gêneros *Archae*, *Neisseria*, *Helicobacter*, *Haemophilus* e *Streptococcus* (2 Mb). Os gêneros *Rhodobacter*, *Salmonella*, *Escherichia*, *Bacillus* e *Pseudomonas* apresentam genomas entre 4 e 5 Mb. E os maiores genomas foram detectados em *Streptomyces sp.* (8 Mb), *Bradyrhizobium sp.* (8,7 Mb), *Myxococcus sp.* (9,5 Mb) e *Anabaema sp.* (6,5 Mb) (Fonstein & Haselkorn, 1995).

O agrupamento dos genomas em progressão de 1, 2, 4 e 8 Mb é consistente com antigas idéias de evolução via duplicação ou fusão de diferentes genomas (Fonstein & Haselkorn, 1995). Por outro lado, o tamanho do genoma pode variar drasticamente dentro do mesmo grupo taxonômico (1 a 5 Mb para *Spirochaeta sp.*; 2,7 a 6,5 Mb para *Cyanobacteria sp.*). Além disso, eventos significativos de alteração de tamanho genômico, tais como deleções de 2 Mb em *Streptomyces ambofaciens* (Lebond *et al.*, 1991) foram reportadas, onde

o tamanho do genoma de diferentes cepas podem variar de 6 a 8 Mb (Fonstein and Haselkorn, 1995).

Os dois principais mecanismos envolvidos em alterações de tamanho de genomas consistem na acumulação de elementos de transposição, os quais tem contribuído na evolução de muitos genomas, e ainda a acumulação de seqüências repetitivas (Hancock, 2002). Por outro lado, a redução do tamanho dos genomas, ocorre através da conservação de genes pertencentes a vias metabólicas essenciais a sobrevivência do microrganismo, enquanto que os demais genes têm sido perdidos no decurso da evolução, refletindo a racionalização do equilíbrio entre estabilidade dos principais processos fisiológicos, e requerimentos para a adaptabilidade ambiental (Koonin, Mushegian & Rudd, 1996).

A nível ecológico, as bactérias que apresentam os menores genomas geralmente colonizam nichos estáveis, e freqüentemente são capazes de colonizar apenas um limitado número de organismos hospedeiros, enquanto que bactérias com genomas maiores, tendem a ocupar ambientes altamente complexos e variáveis como o solo, pois codificam um número maior proteínas, inclusas em diferentes famílias e associadas com variados processos fisiológicos (Bentley & Parkhill, 2004).

Através de mapeamento físico, o antigo dogma de que os microrganismos de uma forma geral possuem genomas circulares, caiu ao se identificar em *Borrelia burgdorferi* (Ferdows & Barbour, 1989), um replicon linear. E também houve a demonstração de novos tipos de organização genômica, onde temos como exemplo o genoma de *Agrobacterium tumefaciens* (Allardet-Servent *et al.*, 1993), que apresenta um cromossomo linear (2,2 Mb) e um outro circular (3,2 Mb). E ainda, *Sinorhizobium meliloti* que apresenta 3 replicons circulares, composto por um cromossomo (3,7 Mb) e dois megaplasmídios denominados pSyma (1,4 Mb) e pSymb (1,7 Mb), os quais podem ser transferidos para outros gêneros bacterianos, em particular *Agrobacterium sp.* (Barloy-Hubler *et al.*, 2000a; Capela *et al.*,

2001). Convém mencionar que esta variedade de replicons encontrados em células bacterianas, atualmente torna cada vez mais difícil a definição de genoma (Bentley & Parkhill, 2004).

Atualmente, mapeamentos físicos têm sido desenvolvidos como estratégias de verificação e aceleração da montagem de genomas completos, durante a fase de fechamento de buracos em genomas bacterianos (fase de finalização), em projetos de seqüenciamento genômico que utilizam a metodologia de *shotgun*. A fase de finalização é resistente à automação por ferramentas de bioinformática, e os problemas de montagem do genoma devem ser resolvidos caso a caso pelo grupo de pesquisa. Nestas circunstâncias, os mapas físicos fornecem valiosas informações, para o controle de qualidade das montagens de *contigs* e *scaffolds*, além de auxiliarem na determinação do posicionamento e orientação destes *contigs* no genoma (Chan & Goodwin, 1999; Shimizu, *et al.*, 2001; Weinel *et al.*, 2001).

1.5 – O seqüenciamento genômico e o seu atual *status* de desenvolvimento.

O seqüenciamento genômico surgiu a partir do desenvolvimento das técnicas de clonagem utilizando enzimas de restrição (Sanger, Nicklen & Coulson, 1977), seqüenciamento automático de DNA (Smith *et al.* 1986), e uso de corantes químicos nas reações de seqüenciamento (Prober *et al.* 1987). Entretanto, nos últimos 15 anos, as técnicas de seqüenciamento evoluíram muito, como por exemplo a utilização de capilares ao invés de placas de eletroforese, que aumenta a quantidade de bases lidas por reação em um menor tempo de execução. Um outro fator muito importante refere-se à leitura das bases do fragmento de DNA que antes era feito manualmente pelo pesquisador, e atualmente é feito automaticamente por um laser acoplado ao seqüenciador (Santos, 2005).

O início da era genômica foi precisamente datada em 28 de julho de 1995, que consiste na data de publicação de um artigo descrevendo uma seqüência genômica completa de 1,8

Mb, oriunda de *Haemophilus influenzae*, seguido pela divulgação de seqüência completa de um genoma de 0,58 Mb, pertencente ao *Mycoplasma genitalium*, que ocorreu em menos de três meses após (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995). Adicionalmente, cabe ressaltar que estes projetos genomas iniciaram após a implementação dos projetos referentes a *Escherichia coli* e *Bacillus subtilis*, que utilizaram a metodologia de *shotgun* hierárquico, a qual consiste no sequenciamento de insertos sub-clonados, de fragmentos de DNA com localização conhecida no cromossomo (Daniels *et al.*, 1992; Kunst, Vassarotti & Danchin, 1995).

Em contraste, os genomas de *H. influenzae* e *M. genitalium*, foram seqüenciados pela metodologia de *shotgun* randômico, que consiste no sequenciamento de milhares de clones com insertos pequenos de DNA diversas vezes, seguido pela montagem de seqüências contínuas de DNA (*contigs*), e fechamento de buracos no genoma utilizando reação de polimerase em cadeia (PCR), e subsequente sequenciamento dos fragmentos de DNA obtidos, utilizando uma variedade de padrões e programas de computador desenhados para executarem tal tarefa. (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995; Koonin, Mushegian & Rudd, 1996; Fraser *et al.*, 2002).

De 1977 até os dias atuais, rápidos avanços na metodologia associada com o sequenciamento de DNA, utilizada para obter a seqüência genômica completa de um organismo, tem alterado as estratégias de estudo de sistemas genéticos de vários organismos, em especial bactérias, tornando o processo relativamente simples, pelo menos para alguns genomas (Kaneko *et al.*, 2002; Bentley & Parkhill, 2004). Dentre estes avanços, destaca-se o desenvolvimento de diversos vetores para construção de bibliotecas genômicas de *shotgun*, o desenvolvimento e aprimoramento de estratégias para fechamento de buracos genômicos e métodos mais sofisticados e robustos para localização e anotação de genes, reduzindo drasticamente o tempo e custos requeridos em cada estágio de um projeto genoma (Fraser *et*

al., 2002). Conseqüentemente, o aumento da disponibilização de novas seqüências genômicas completas em bancos de dados, contribui para um melhor entendimento de processos relacionados à replicação, metabolismo, expressão gênica e evolução bacteriana (Moszer, 1998; Kaneko *et al.*, 2002).

Atualmente existem relatos na literatura, da existência de 2037 projetos de seqüenciamento genômico, sendo que 1.650 projetos estão em andamento (996 projetos de organismos procariotos, 608 de eucariotos e 46 metagenomas), e apenas 387 genomas foram seqüenciados completamente e publicados até o momento (319 de bactérias, 27 de arqueas e 41 de eucariotos). Convém ressaltar, que os projetos genomas bacterianos em andamento, estão divididos em três principais grupos de interesse: bactérias de interesse médico (46,7%), bactérias de interesse biotecnológico (29,8%) e bactérias de interesse ambiental e/ou agrícola (22,2%). Apenas 19% de todos os projetos de seqüenciamento genômico, possuem suas informações disponíveis e publicadas até o momento, sinalizando para uma tendência de crescimento significativo de informações a serem disponibilizadas em bancos de dados nos próximos anos (Kyrpides, 1999; Benson *et al.* 2005; Santos, 2005; GOLD, 2006).

1.6 – A metodologia de *shotgun*

Apesar dos avanços tecnológicos na manipulação de moléculas de DNA, ainda existe um limitante no número de bases que podem ser lidas pelos procedimentos laboratoriais de sequenciamento do DNA. Nestes experimentos, mesmo havendo a possibilidade de se trabalhar com segmentos longos de DNA, somente as 800 primeiras bases podem ser lidas ou seqüenciadas diretamente. Uma alternativa freqüentemente utilizada para lidar com essa limitação em projetos genomas bacterianos, é a metodologia de *shotgun* (Sanger, Nicklen & Coulson, 1977; Edwards *et al.*, 1990; Waterson, Lander & Sulston, 2002).

A metodologia de *shotgun* consiste na fragmentação da molécula de DNA em milhares de fragmentos, com tamanhos variando entre 1 e 2 Kb, através de digestões com enzimas de restrição ou fragmentações mecânicas (sonicação ou nebulização). Em seguida os fragmentos gerados são clonados em plasmídios linearizados, gerando moléculas de DNA recombinante, que posteriormente são inseridos em *Escherichia coli*, para construção das chamadas bibliotecas genômicas. A partir do cultivo dos clones destas bibliotecas genômicas, realiza-se extração de plasmídeos, seguido de seqüenciamento de seus insertos, para obtermos as seqüências que serão utilizadas na reconstrução da seqüência genômica completa, através de um processo conhecido como montagem de DNA (Sanger *et al.*, 1980; Deininger, 1983; Waterson, Lander & Sulston, 2002).

Teoricamente a metodologia de *shotgun* é bastante simples, entretanto, alguns fatores intrínsecos dos procedimentos laboratoriais, podem complicar o processo de montagem do DNA, entre eles podemos citar o fato de que alguns insertos contidos nos plasmídeos inseridos em *E. coli*, podem codificar proteínas tóxicas para a bactéria hospedeira, impossibilitando o seqüenciamento de alguns fragmentos de DNA que apresentam estes genes tóxicos, além de possíveis falhas do manipulador e da reação de seqüenciamento, tais como sinais fracos ou até mesmo ausentes (Santos, 2005).

Uma outra particularidade da metodologia de *shotgun* consiste na utilização de softwares como o *Phred*, que utiliza métodos probabilísticos, para identificar o valor de qualidade de uma amostra, o que equivale à probabilidade de uma base ter sido lida erroneamente, durante a interpretação dos dados experimentais gerados pelas reações de seqüenciamento. Uma base é considerada de boa qualidade, quando é atribuído valor acima de 20, cujo este valor significa uma probabilidade de erro de 0,01 ou 1%, portanto, valores acima de 20 refletem uma taxa de erro ainda menor (Ewing & Green, 1998).

1.7- Montagem do genoma e processo de finalização

Quando o seqüenciamento pela metodologia de *shotgun* termina, as seqüências dos clones (*reads*) são utilizadas para montar diversas seqüências genômicas maiores (*contigs*). Os *contigs* são separados um dos outros por buracos de seqüenciamento, os quais suas seqüências são desconhecidas, além disso, cada *contig* possui áreas com baixa qualidade (principalmente nas extremidades), o que interfere na montagem do genoma. Outros fatores de interferência nas montagens genômicas, consistem na presença de grandes quantidades de seqüências repetitivas, e na impossibilidade de clonagem de seqüências de genes tóxicos (Santos *et al.*, 2004; Santos, 2005).

A eliminação de buracos entre os *contigs* é conhecida como processo de finalização, que geralmente inicia-se após o fim do seqüenciamento pela metodologia de *shotgun*. Entretanto, até hoje não existe um indício evidente que determine o fim do seqüenciamento por esta metodologia, e devido a este fato, os grupos de pesquisa geralmente seqüenciam uma cobertura de aproximadamente 6 a 8 vezes do genoma, antes de iniciarem o processo de finalização (Kaneko *et al.*, 2000; Capela *et al.*, 2001; Kaneko *et al.*, 2002; Prust *et al.*, 2005).

O primeiro passo para realização da finalização de um genoma, é tentar determinar a ordem e orientação dos *contigs*. Esta tarefa pode ser realizada das seguintes formas: através do uso da metodologia de *mate-pair reads*, que torna possível ligar *contigs* separados por um ou mais *reads* formando os chamados *scaffolds*, que consistem num agrupamento de *contigs*, exibindo regiões que não foram seqüenciadas (Santos *et al.*, 2004; Santos, 2005); através de mapeamentos físicos de perfis de digestão de BACs (cromossomos artificiais bacterianos) associados com técnicas de hibridização e seqüenciamento (Capela *et al.* 1999; Barloy-Hubler *et al.*, 2000a, 2000b; Capela *et al.* 2001; Tomkins *et al.* 2001; Goto *et al.*, 2003; Loureiro *et al.*, 2005); ou através de mapas físicos de macrorestrição, gerados com a

metodologia de PFGE associado com técnicas de hibridização (Lai *et al.*, 1999; Blank & Stemke, 2000; Taketani, 2001; Weinel *et al.*, 2001; Loureiro *et al.*, 2003).

Para seqüenciar essas regiões de buracos genômicos existem duas estratégias: o resseqüenciamento e a reação da polimerase em cadeia (PCR). O resseqüenciamento utiliza clones que já foram seqüenciados pela metodologia de *shotgun*, onde são refeitas as suas reações de seqüenciamento, visando melhorar a qualidade das seqüências em regiões flanqueadoras de buracos no genoma, assim como conseguir obter a seqüência completa dos clones selecionados pela metodologia de *mate-pair reads*, que foi utilizada no processo de geração de *scaffolds* (Santos *et al.*, 2004; Santos, 2005).

Por outro lado, a PCR é altamente eficiente para cobertura de regiões que apresentam seqüências repetitivas ou de genes tóxicos, pois esta técnica baseia-se na produção de oligonucleotídeos, que anelam nas bordas dos *contigs* que estão ligados em *scaffolds*, com o objetivo de amplificar as regiões ainda não seqüenciadas. Esta é considerada a estratégia mais eficiente de todas, pois amplifica fragmentos de DNA que até o momento não foram seqüenciados por outras metodologias, além de ser considerada a metodologia mais indicada, para fechamento de buracos existentes entre *contigs* ordenados e orientados, através dos dados obtidos a partir de mapeamento físico (Taketani, 2001; Weinel *et al.*, 2001; Nelson *et al.*, 2002; Santos *et al.*, 2004; Matsunaga *et al.*, 2005; Loureiro, 2005; Santos, 2005).

Tendo como premissa básica que o objetivo da fase de finalização de um genoma, é criar uma única seqüência de DNA contínua e com uma mínima taxa de erro (Gordon, Desmarais & Green, 2001), podemos intuitivamente concluir que esta fase é a mais cara e mais lenta, dentro de um projeto genoma que utiliza a metodologia de *shotgun* (Branscomb & Predki, 2002). A maior parte dos procedimentos e análises realizadas na fase de finalização de um genoma, tais como: ordem e orientação dos *contigs*, desenho de oligonucleotídeos para execução de PCRs combinatórios, e técnicas especiais de seqüenciamento como o *primer*

walking são executadas manualmente, o que torna esta fase cerca de quatro vezes mais cara e mais lenta que toda a fase de seqüenciamento (Gordon, Desmarais & Green, 2001; Santos *et al.*, 2004; Loureiro, 2005; Santos, 2005)

1.8 – Anotação de genomas.

A análise das seqüências biológicas de um projeto de sequenciamento genômico, através da utilização das mais variadas ferramentas de bioinformática e a comparação com os mais variados bancos de dados, compreende uma etapa final conhecida como anotação, que consiste na conversão dos dados de uma seqüência biológica, em informações biologicamente relevantes. A anotação consiste em uma meta-informação, ou seja, uma descrição de características em mais alto nível da seqüência biológica (Lemos, 2004).

Dentre as principais ferramentas de bioinformática, utilizadas para execução do processo de anotação, destacam-se: o BLAST (*Basic Local Alignment Sequence Tool*), que consiste numa ferramenta utilizada para verificação de homologia de uma determinada seqüência, com seqüências estudadas previamente por outros pesquisadores (Altschul, 1990; Altschul, 1998; WU-BLAST, 2006); o RBSFinder, que localiza sítios de ligação ribossomal (TIGR, 2006); programas de predição de genes como o GLIMMER (Delcher *et al.*, 1999) e o ORFFinder (NCBI, 2006a); algoritmos para predição de estruturas de proteínas como o Modeller (Sali, 2006) e o Threader (Jones, 2006); pacotes de análise filogenética como o PHYLIP (*PHYLogeny Inference Package*) (Felsenstein, 2006); e programas projetados para comparação de vários genomas como o Alfreco (Sanger, 2006).

Com relação aos bancos de dados, podemos destacar os que armazenam seqüências de nucleotídeos, tais como o Genbank-NT (NCBI, 2006b), GSDB (Harger *et al.*, 1998), GDB (Letovsky *et al.*, 1998) e o EMBL (Cochrane *et al.*, 2006); os que armazenam seqüências de aminoácidos como o SWISS-PROT (Boeckmann, 2003), PIR (Wu *et al.*, 2002) e o Genbank-

NR (NCBI, 2006b); bancos de dados de estruturas terciárias de proteínas como o PDB - *Protein Data Bank* (Westbrook *et al.* 2002); padrões de seqüências proteicas conservadas associadas com funções específicas num organismo, tais como o PROSITE (Falquet *et al.*, 2002); e bibliotecas que oferecem padrões de seqüências de estruturas de domínios proteicos mais longos, como o PFAM (Bateman *et al.*, 2004).

O processo de anotação apresenta 4 etapas distintas, onde a primeira consiste na anotação à nível de nucleotídeos, cujo o principal objetivo é a localização física de ORFs (fases de leitura aberta) nas seqüências de DNA, assim como elementos repetitivos, promotores e sítos de ligação ribossomais. A segunda etapa da anotação ocorre em nível de proteínas, cujos objetivos residem na determinação de identidades e prováveis funções dos genes, visando identificar genes presentes e eventualmente ausentes num determinado genoma. A terceira etapa ocorre à nível de processo, onde os grupos de pesquisa buscam realizar a identificação de diversas vias metabólicas. A quarta etapa consiste na análise das seqüências reguladoras presentes no genoma, visando implementar uma anotação funcional eficiente, e aprimorar a compreensão das diversas vias metabólicas presentes no organismo, em especial, aquelas que apresentam importância econômica, científica, tecnológica e acadêmica (Koonin, Mushegian & Rudd, 1996; Downs & Escalante-Semerena, 2000; Schawartz, 2000; Sistema Bionotes 2006b).

Em linhas gerais, podemos afirmar que a anotação de um genoma consiste num processo iterativo, que requer a utilização de diversas ferramentas de bioinformática, com finalidade de refinar os dados brutos gerados a partir de um projeto de sequenciamento genômico (Sistema Bionotes, 2006b), através da utilização de ferramentas que possibilitem a localização de regiões codificantes numa seqüência de DNA (Delcher *et al.*, 1999; NCBI, 2006a), a predição de função e estrutura da proteína codificada por um determinado gene (Jones, 2006; Sali, 2006), assim como a determinação de relações filogenéticas entre

diferentes organismos (Felsenstein, 2006; Sanger, 2006). Desta forma, são elucidadas questões relacionadas com organização genômica, aspectos bioquímicos, fisiológicos e evolutivos do organismo em estudo, de espécies relacionadas e/ou espécies que ocorrem em associação (Koonin, Mushegian & Rudd, 1996; Moszer, 1998).

Atualmente existem diversos sistemas de anotação, onde devemos destacar o Bionotes (Lemos, Seibel & Casanova, 2003) e o Artemis (Rutherford *et al.*, 2000), que consistem em sistemas de anotações biológicas que possuem por objetivo facilitar a interpretação de biosseqüências, a partir de análises realizadas em diversas fontes de dados externas, como por exemplo o Genbank (NCBI, 2006b), PIR (Wu *et al.*, 2002), SWISSPROT (Boeckmann, 2003) e o PROSITE (Falquet *et al.*, 2002), além de possuir outros aplicativos externos como o BLAST (Altschul, 1990; Altschul, 1998), CAP3 (Huang & Madan, 1999), Phred (Ewing *et al.*, 1998; Ewing & Green, 1998), Phrap (Gordon, Abajian & Green, 1998), Crossmatch (UWGC, 2006), Glimmer (Delcher *et al.*, 1999), tRNAScan (Lowe & Eddy, 1997) e RBSfinder (TIGR, 2006), o que permite desde a consulta até o cadastramento, atualização e remoções de anotações.

1.9 – Considerações gerais desta obra

Neste trabalho, foi descrita uma análise detalhada do cromossomo do *G. diazotrophicus* PAL5, através da metodologia de PFGE associada com experimentos de hibridização, com subsequente construção de um mapa físico e genético, reportando o tamanho do genoma e o número e orientação dos operons ribossomais. E um mapeamento físico, utilizando cromossomos artificiais bacterianos (BACs), associados com sequenciamento de clones selecionados e hibridizações com sondas radioativas.

Além de enfatizar a importância da utilização dos dados de mapeamento físico e genético do genoma deste endófito, no alinhamento de *contigs* e *scaffolds* durante a fase de

finalização do seqüenciamento genômico desta bactéria. E descrever uma análise geral dos resultados de anotação deste genoma, com finalidade de relatar as principais particularidades do genoma deste microrganismo, que consiste no único membro do subgrupo das α -proteobactérias com capacidade de fixação de nitrogênio atmosférico, e a única bactéria com características endofíticas relatada em diversos hospedeiros vegetais, que apresenta o seu genoma completamente seqüenciado.

2 – OBJETIVOS

2.1 – Objetivos gerais

Realizar um mapeamento físico e genético do cromossomo do *G. diazotrophicus* PAL5. Para isso, será utilizada a metodologia de PFGE, associada com experimentos de hibridização, e digestões duplas de cromossomos artificiais bacterianos (BACs) associados com técnicas de sequenciamento e hibridização, visando a determinação do posicionamento de marcadores genéticos neste cromossomo, além de descrever uma análise geral dos resultados de anotação deste genoma.

2.2 – Objetivos específicos

- 2.2.1 – Determinar o tamanho, morfologia e composição do genoma de *G. diazotrophicus* PAL5;
- 2.2.2 – Relatar o número e a orientação dos genes de RNA ribossomais na estrutura do cromossomo deste endófito;
- 2.2.3 – Posicionar marcadores moleculares na estrutura do cromossomo deste microrganismo, para auxiliar o alinhamento de *contigs* e *scaffolds* durante a fase de finalização do Projeto Riogene;
- 2.2.4 – Descrever uma análise geral dos resultados de anotação do genoma desta bactéria, realizada pelos membros integrantes do Projeto Riogene;
- 2.2.5 – Realizar uma mineração de dados acerca de algumas particularidades genômicas deste endófito, observadas durante o processo de anotação.

3 – MATERIAL E MÉTODOS

3.1 – Cepas bacterianas e condições de crescimento

A linhagem de células utilizadas em todos os experimentos do Projeto Riogene (Projeto Riogene, 2006) foi o *Gluconacetobacter diazotrophicus* PAL5 (ATCC49037; BR11281), que consiste numa linhagem celular isolada de raízes de cana de açúcar cultivadas em uma plantação do estado de Alagoas, utilizada em diversos estudos sobre fixação biológica de nitrogênio, interação planta-bactéria e produção de bacteriocinas e fitohormônios.

O *G. diazotrophicus* PAL5 foi cultivado em caldo Dygs (2,0 g l⁻¹ de glicose; 1,5 g l⁻¹ de peptona; 2,0 g l⁻¹ de extrato de levedura; 0,5 g l⁻¹ de K₂HPO₄; 0,5 g l⁻¹ de MgSO₄.7H₂O; 1,5 g l⁻¹ de ácido glutâmico) pH 6,0, em agitador giratório a 30 °C e 200 rpm, durante aproximadamente 40 h, para obter uma cultura com densidade ótica (DO_{600 nm}) de 1,3.

Adicionalmente, foram realizadas culturas controle em placas de Ágar Dygs (Caldo Dygs suplementado com 1,5% de Ágar bacteriológico) a 30 °C durante 48 h, para cada cultura utilizada nas extrações de DNA, com finalidade de verificar a pureza das culturas utilizadas. A cepa estoque é mantida congelada a -70 °C em meio de cultura Dygs líquido suplementado com glicerol a 20%.

Os clones de *Escherichia coli* oriundos das bibliotecas genômicas de *shotgun* do Projeto Riogene, foram cultivadas em caldo Luria-Bertani (LB) (5 g l⁻¹ de extrato de levedura; 10 g l⁻¹ de triptona; 10 g l⁻¹ de NaCl) pH 7,0, suplementado com 100 µg/mL de ampicilina, em agitador giratório a 37 °C e 200 rpm durante 24 h. As bibliotecas de *shotgun* são mantidas congeladas a -70 °C, em placas de 96 poços com 150 µl de caldo LB suplementado com 100 µg/mL de ampicilina e glicerol a 20% por poço.

3.2 – Preparação de DNA genômico de alto peso molecular para PFGE

O DNA genômico utilizado na metodologia de PFGE, foi preparado a partir de células contidas em 1 mL de cultura de *G. diazotrophicus* PAL5, em fase logarítmica de crescimento ($DO_{600\text{ nm}} = 1,3$), que foram sedimentadas e lavadas com 500 μL de solução salina a 0,85% estéril. Estas células foram ressuspensas em 25 μL de tampão EDTA 50 mM (pH 8,0), 10 μL de lisozima (10 mg/mL), 190 μL de solução de lise (6 mM de Tris-HCl - pH 7,5; 100 mM de EDTA pH 7,5; 1 M de NaCl; 0,5% de Brij58; 0,2% de Deoxicolato; 0,5% de Sarkosil) e 225 μL de Agarose de baixo ponto de fusão a 2% aquecida a 50 °C. Esta mistura foi cuidadosamente homogeneizada e moldada em forma de blocos de agarose.

Os blocos de agarose foram transferidos para tubos falcon de 15 mL e suspensos em 5 mL de solução de lise suplementada com 20 $\mu\text{g/mL}$ de Rnase A, e incubados a 37 °C por 16 h sob leve agitação. Após o período de incubação, a solução de lise foi retirada, e os blocos foram lavados duas vezes com 5 mL de tampão CHEF-TE (100 mM de Tris-HCl, pH 7,5; 100 mM de EDTA, pH 7,5), durante 30 minutos cada lavagem. Após as duas lavagens, o tampão CHEF-TE foi substituído por 5 mL de tampão ESP (0,4 M de EDTA, pH 9,3; 1% de sarkosil; 1 mg/mL de proteinase K), e incubou-se os blocos por 24 h a 50 °C em banho-maria.

Após o tratamento dos blocos com proteinase K, realizou-se inativação desta protease através de duas lavagens dos blocos, com 5 mL de tampão CHEF-TE acrescido de 1 mM de PMSF (fenilmetilsulfonyl fluorido), em cada lavagem de 30 minutos a 50 °C. Após inativação das proteases, os blocos foram lavados 4 vezes em 5 mL de tampão CHEF-TE, sob leve agitação a 37 °C durante 30 minutos cada lavagem. Após a última lavagem com CHEF-TE, este tampão foi substituído por 5 mL de etanol 70%, que permite a estocagem dos blocos a 4 °C durante vários meses.

3.3 – Digestão dos blocos de PFGE contendo DNA genômico.

As enzimas de restrição de raro corte utilizados neste estudo foram *I-CeuI* (5'...TAAC TATAACGGTCCTAA▼GGTAGCGA...3'; 3'...ATTGATATTGCCAG▲GATTCCATCG CT...5'), *PacI* (5'...TTAAT▼TAA...3'; 3'...AAT▲TAATT...5') e *SwaI* (5'...ATTT▼AA AT...3'; 3'...TAAA▲TTTA...5') (New England Biolabs®). Os blocos de agarose contendo o DNA genômico foram cortados em 4 pedaços de tamanho iguais, colocados em tubo eppendorf de 1,5 mL imersos em 100 µl de tampão 1× apropriado para a enzima de restrição que se deseja digerir o DNA, e incubados a 4 °C durante 1 h.

Após o período de incubação, este tampão foi substituído por 100 µl de novo tampão 1×, contendo 10 µg de BSA (soro albumina bovina) e 10U da enzima, e incubado em temperaturas apropriadas durante 6 h, de acordo com as recomendações do fabricante da enzima de restrição. O DNA genômico foi digerido através de digestões simples e duplas com as enzimas de restrição acima citadas, o que gerou 6 diferentes perfis de restrição do DNA genômico do *G. diazotrophicus* PAL5. Nas digestões duplas, após o período de incubação da primeira enzima, as soluções de restrição foram descartadas, e os blocos de agarose foram lavados uma vez com CHEF-TE durante 1 h a 4 °C, antes da execução das segundas digestões com tampões, enzimas e temperaturas apropriadas.

3.4 – Eletroforese em Gel de Campo Pulsado (PFGE)

Todos os géis de PFGE foram submetidos à eletroforese no sistema de PFGE Gene Navigator (GE-Amersham Biosciences®). Os parâmetros de eletroforese, tais como: concentração da agarose, tampão de corrida, tempo de pulso, voltagem e temperatura do tampão de corrida, variaram de acordo com as condições experimentais requeridas para detecção dos fragmentos de DNA, em diferentes faixas de tamanho.

Para separação de fragmentos de DNA compreendidos entre 70 e 1640 Kb, utilizou-se os seguintes parâmetros: agarose 1%, TBE 0,5× (6,05 g l⁻¹ de Tris base; 3,09 g l⁻¹ de ácido bórico; 0,18 g l⁻¹ de Na₂EDTA) a 13 °C, 6 V/cm e um programa com 3 blocos apresentando diferentes tempos de pulso (Bloco A - 80 a 100 s durante 16 h; Bloco B - 100 a 120 s durante 8 h; e Bloco C - 120 a 140 s durante 16 h).

Na análise de fragmentos de DNA com tamanhos entre 2 e 650 Kb, foram empregados 2 diferentes programas de eletroforese, com parâmetros distintos a saber: Para fragmentos de DNA, com tamanho compreendidos entre 50 e 650 Kb, utilizou-se gel de agarose a 1%, TBE 0,5× a 6 V/cm e 13 °C, com pulso inicial de 1 s e final de 50 s durante 23 h. Para fragmentos com tamanhos entre 2 e 50 Kb empregou-se gel de agarose a 1%, TAE 1× (4,84 g l⁻¹ de Tris base; 1,14 mL l⁻¹ de ácido acético glacial; 400 µl l⁻¹ de EDTA 0,5 M, pH 8,0) a 4 V/cm e 13 °C, com pulso inicial de 0,1 s e final de 10 s durante 16 h.

Para fracionamento de fragmentos de DNA com tamanhos acima de 1 Mb, foi utilizado gel de agarose 0,8%, TBE 0,5× a 12 °C, e um programa de eletroforese dividido em dois blocos. O primeiro bloco consistiu de um pulso inicial de 1000 s durante 24 h, seguido de um bloco com tempo de pulso inicial de 1000 e final 4000 s durante 192 h.

Na produção dos géis de PFGE, foram utilizados um dos três diferentes marcadores de peso molecular adquiridos comercialmente (New England Biolabs®), de acordo com as faixas de tamanhos que foram analisadas. Para fragmentos de DNA entre 225 e 1900 Kb, utilizou-se os cromossomos de levedura; para fragmentos variando de 50 a 1000 Kb utilizou-se o *lambda ladder*; e para fragmentos entre 0,13 e 194 Kb utilizou-se o marcador *lambda low range*.

Após o procedimento de eletroforese, todos os géis foram corados com brometo de etídio (100 µg/mL) durante 15 min, e fotografados sobre luz ultravioleta para documentação dos dados obtidos. Posteriormente, os tamanhos dos fragmentos de DNA observados nos géis,

foram determinados através da média dos valores, atribuídos aos fragmentos de DNA em 10 diferentes géis, analisados no software Fragment Analysis (Molecular Dynamics®).

3.5 – Transferência de DNA fracionado em gel de agarose para membranas de nylon através de capilaridade (*Southern Blotting*).

Para realização deste procedimento, cortou-se o gel próximo aos *lanes* desejados, e realizou-se depurinação através do seu tratamento com 200 mL de solução depurinante (HCl 0,25 N), durante 15 min a temperatura ambiente sob leve agitação, seguida de desnaturação com 200 mL de solução desnaturante (NaOH 0,5 M; NaCl 1,5 M) por 45 min, e neutralização com 200 mL de solução neutralizante (Tris-HCl 1 M, pH 7,5; NaCl 1,5 M) durante 1 h, ambas a temperatura ambiente e sob leve agitação.

Durante o procedimento de neutralização, foi realizada a montagem do suporte de transferência de DNA para membrana de nylon (Hybond+; GE-Amershan Biosciences®), utilizando-se placas de vidro, potes plásticos e papel de filtro 3 mm. Após neutralização do gel, colocou-se o gel invertido sobre o suporte, e sobre este uma membrana de nylon, previamente cortada em tamanho adequado e tratada com água deionizada por 5 min.

Sobre a membrana colocou-se 3 papéis de filtro 3 mm, cortados de acordo com o tamanho do gel, e sobre estes uma pilha de papel toalha de aproximadamente 15 cm de altura, uma placa de vidro e um peso de aproximadamente 500 g. Adicionou-se 500 mL de tampão SSC 10× (NaCl 1,5 M; Citrato de Sódio 15 mM) pH 7,0, em pote plástico localizado na base do sistema de transferência, e procedeu-se a transferência durante 48 h. Após este período, as membranas foram incubadas em estufa a 80 °C durante 2 h, para ocorrer fixação do DNA.

3.6 – Seleção de clones contendo genes de cópia única e RNAs ribossomais

Clones com insertos contendo genes de cópia única, oriundos das bibliotecas de *shotgun* do Projeto Riogene, que apresentavam os três seguintes critérios, foram selecionados para produção de sondas radioativas nos experimentos de hibridização: 1) clones com insertos presentes em *contigs* maiores que 10 Kb; 2) clones cujo suas seqüências não apresentam identidade, com seqüências conservadas em outros membros da mesma família proteica; e 3) quando a seqüência do clone estivesse presente em grandes *contigs*, elas deveriam estar distante entre 100 e 200 Kb de outro marcador molecular selecionado.

Para seleção de clones contendo insertos com genes de RNAs ribossomais, foram alinhadas através da utilização do Blastn (Altschul *et al.*, 1990; Altschul, 1998), seqüências dos RNAs ribossomais 16S, 23S e 5S de *G. diazotrophicus*, contra o banco de dados de *reads* do Projeto Riogene (Sistema Bionotes, 2006a).

3.7 – Construção de Sondas Radioativas de DNA

Os clones selecionados na biblioteca de *shotgun* (item 3.6) foram cultivados em caldo LB, conforme descrito no item 3.1, e submetidos ao método rápido de lise alcalina descrito por Sambrook & Russell (2001), para proceder extração de plasmídeos. Estes plasmídeos foram submetidos à amplificação de seus insertos, através da metodologia da reação de polimerase em cadeia (PCR), utilizando pares de primers de acordo com os vetores utilizados para produção das bibliotecas de *shotgun* (Projeto Riogene, 2006), e demais condições estabelecidas na rotina laboratorial.

Os produtos de PCR foram submetidos a eletroforese em gel de agarose, de onde se purificou as bandas correspondentes ao tamanho de cada um produto de PCR, e seqüenciados com o kit de sequenciamento Big-Dye terminator, nos seqüenciadores automáticos de DNA ABI377 ou Megabace (Applied Biosystems®), para confirmar a identidade dos insertos. Para

marcação das sondas radioativas, os produtos de PCR previamente purificados e ressequenciados, foram submetidos a uma marcação com primers randômicos (pd (N)₆) e [α -³²P] dCTP (GE-Amersham Biosciences®), com exceção para a sonda do RNAr 5S, a qual foi marcada através de PCR radioativo, com par de primers desenhados para tal finalidade e [γ -³²P]dATP (GE-Amersham Biosciences®). Ambos os protocolos de marcação de sondas foram realizados segundo Sambrook & Russell (2001). Antes da utilização das sondas, elas foram previamente desnaturadas por aquecimento a 99 °C durante 5 min e rapidamente submetidas a choque térmico a 0 °C durante 2 min.

3.8 – Experimentos de hibridização com sondas radioativas

As membranas de nylon produzidas de acordo com o item 3.5, foram pré-hibridizadas a 42 °C em solução de pré-hibridização (formamida 50%; SDS 0,5%; SSC 6×; dendhardt's 5×; 100 µg de DNA de timo de bezerro desnaturado) durante 2 h. Após este período, foram adicionados 50 µl da sonda radioativa marcada de acordo com o item 3.7, e a incubação continuou durante 24 h à 42 °C.

Após a incubação das membranas com as sondas, o tampão de hibridização contendo o excesso de sonda, foi estocado em tubo falcon de 50 mL dentro de caixa de acrílico para o decaimento do isótopo radioativo, e as membranas foram lavadas com diferentes tampões, sob condições variáveis de estringência para retirar o excesso de sonda e minimizar os ruídos de hibridização, de acordo com a seguinte seqüência de tampões: 1^a lavagem – SSC 2× e SDS 0,1% durante 5 min a temperatura ambiente; 2^a lavagem – SSC 2× e SDS 0,5% durante 15 min a temperatura ambiente sob agitação; 3^a lavagem – SSC 0,1× e SDS 0,5% durante 1 h a 37 °C sob agitação; 4^a lavagem – SSC 0,1× e SDS 0,5%, durante 1 h a 68 °C sob agitação; e 5^a lavagem - SSC 6× durante 5 min a temperatura ambiente.

Após as lavagens, as membranas foram secas, embaladas em filme de polipropileno, e colocadas em exposição direta contra tela de estoque de emissão de fósforo, dentro de um cassete de exposição do sistema *phosphor image* (Molecular Dynamics®), durante 48 h a temperatura ambiente. Posteriormente à exposição, a tela de estoque de emissão de fósforo foi escaneada com o *software Storm* (Molecular Dynamics®), onde foram geradas imagens com extensão tiff, para documentação dos experimentos de hibridização.

Em seguida, as imagens geradas foram comparadas com as fotos dos géis de PFGE, utilizadas para produzir as membranas de cada experimento de hibridização, e os resultados foram anotados. Todos os experimentos de hibridização foram realizados pelo menos duas vezes para verificar a reprodutibilidade dos dados.

3.9 – Construção da biblioteca genômica de Cromossomos Artificiais Bacterianos (BACs)

A biblioteca genômica de BACs foi produzida na *Clemson University Genomic Institute* – Carolina do Norte – EUA, a partir de fragmentos de DNA genômico do *G. diazotrophicus* PAL5, com tamanhos compreendidos entre 100 e 140 Kb, os quais foram gerados a partir de digestões parciais com a enzima de restrição *HindIII* (New England Biolabs®), e eletroeluídos a partir de géis de PFGE. Estes fragmentos de DNA foram clonados no plasmídeo vetor pCUGI_BAC1 (Luo *et al.*, 2001), e transformados em *E. coli* DH10B.

Os clones contendo plasmídeos com insertos de DNA genômico do *G. diazotrophicus* PAL5, foram selecionados em meio de cultura LB-X-gal-IPTG (LB: Meio Luria bertani; X-gal: 5-bromo-4-cloro-3-indolil- β -D-galactopiranosida; IPTG: isopropil- β -tiogalactopiranosida), suplementado com 12,5 μ g/mL de cloranfenicol, pois o plasmídeo vetor apresenta o gene Cm^r que confere resistência a este antimicrobiano. Os clones selecionados foram inoculados em microplacas de 384 poços, contendo 50 μ l de Caldo LB

suplementado com glicerol a 20% e 12,5 µg/mL de cloranfenicol por poço, com auxílio de robô Qbot, o que gerou um total de 5760 clones (15 placas).

3.10 – Crescimento dos clones da biblioteca de BACs, extração e quantificação do DNA

As placas da biblioteca de BACs foram replicadas no Núcleo de Genomas Johanna Döbereiner – Instituto de Bioquímica Médica – CCS – UFRJ, e 1 µl das culturas destes clones foram cultivados em tubos falcon contendo 10 mL de Caldo LB, suplementado com 12,5 µg/mL de cloranfenicol, em agitador giratório a 37 °C e 200 rpm, durante 24 h. As culturas produzidas foram centrifugadas a 7000 rpm durante 5 min, e as células obtidas foram transferidas para tubos de eppendorf de 1,5 mL, para proceder-se o método rápido de lise alcalina descrito por Sambrook & Russell (2001) e extrair os BACs. Após extração, os BACs foram quantificados em gel de agarose, utilizando-se como padrão diluições de DNA do bacteriófago Lambda (New England Biolabs®).

3.11 - Digestão, fracionamento e documentação dos perfis de restrição dos BACs.

Aproximadamente de 1 µg de BACs dos clones analisados foram transferidos para novos tubos eppendorf, e submetidos a digestões duplas com 10U das enzimas de restrição *HindIII* e *BamHI* (New England Biolabs®) em 20 µl de tampão E 1× a 37 °C, durante 24 h. Após digestão, adicionou-se 3 µl de tampão de amostra (0,15% de Azul de Bromotimol; 0,15% Xilenocianol e 50% Glicerol), e procedeu-se concentração dos fragmentos de DNA em *speed vac* para um volume de 10 µl de amostra, os quais foram fracionados em gel de agarose a 1% e TAE 1×, em cuba refrigerada a 4 °C durante 16 h a 60V. Como padrão de peso molecular, utilizou-se 500 ng do marcador *Hilo marker* (DNA de bacteriófago lambda, submetido a digestões simples com as enzimas *HindIII* e *SalI* - New England Biolabs®). Após eletroforese, os géis foram corados com *Syber green* (Sigma®) 1000× durante 45 min, e

escaneados no *software Storm* (Molecular Dynamics®), a partir do qual foram geradas imagens com extensão tiff.

3.12 – Análise das imagens geradas a partir dos géis de digestão dupla dos BACs

As imagens dos géis geradas no *software storm* foram transferidas para o *software Image 3.0i* (Sanger Institute – Cambridge – Inglaterra), onde foi realizado o processamento dos géis, que consiste na determinação de fragmentos de DNA representativos de cada perfil de restrição dos BACs, assim como dos fragmentos dos marcadores moleculares. Ao término do processamento dos géis, são gerados arquivos de saída reconhecidos pelo *software FPC* (*Fingerprint Contig Software*; Sanger Institute – Cambridge – Inglaterra), que foi utilizado para comparação entre os perfis de restrição dos BACs depositados no banco de dados. A partir desta comparação, o *software FPC* realiza a montagem de mapas físicos, com base nas sobreposições entre os BACs analisados, deduzidas a partir de homologias observadas entre os perfis de restrição.

3.13 – Seleção de BACs para reações de sequenciamento e hibridizações

A partir de 2 mapas físicos de BACs gerados no *software FPC*, utilizando 2 diferentes níveis de estrigência na análise: *cutoff* 1e-12 (mais estrigente) e *cutoff* 1e-04 (menos estrigente), ambos com nível de tolerância 7, foram selecionados 117 clones, os quais seus insertos de DNA realizam 2 coberturas completas do maior *contig* gerado sob a condição mais estrigente, e 2 coberturas completas do genoma do *G. diazotrophicus* PAL5, sob a condições menos estrigente. Os BACs destes clones foram submetidos a reações de sequenciamento na *Clemson University Genomic Institute* – Carolina do Norte – EUA, e os perfis de restrição dos BACs foram submetidos a hibridizações com sondas radioativas dos genes de RNAr 16s e 23s, no Núcleo de Estudos de Genomas Johanna Döbereiner – Instituto de Bioquímica

Médica – CCS – UFRJ, de acordo com os procedimentos descritos nos itens 3.5, 3.6, 3.7 e 3.8.

3.14 – Anotação das ORFs do *G. diazotrophicus* PAL5 e mineração de dados obtidos na anotação.

Todo o processo de anotação do genoma do *Gluconacetobacter diazotrophicus* PAL5, foi executado no sistema Bionotes (Lemos, Seibel & Casanova, 2003; Lemos, 2004), pelos membros integrantes do Projeto Riogene. Foi realizado um processo de anotação manual, onde o primeiro passo consistiu na validação das ORFs, pois o programa Glimmer (Delcher *et al.*, 1999) utilizado para predição de ORFs, é considerado um programa que gera um grande número de falsos positivos, onde diversas ORFs podem apresentar sobreposições completas. Nestas situações, a seleção das ORFs corretas foram baseadas nos resultados de Blast (Altschul *et al.*, 1990; Altschul, 1998), onde dentre as ORFs sobrepostas, as que apresentaram melhores alinhamentos em relação aos bancos de dados, foram validadas.

Dentre as variantes de Blast (WU-BLAST, 2006) utilizadas no processo de anotação, o Blastp foi utilizado como ferramenta inicial, pois compara seqüências de aminoácidos, contra seqüências de proteínas depositadas nos bancos de dados públicos, realizando uma análise menos estridente que outras ferramentas, tais como o Blastn (utilizado para comparação de seqüências de nucleotídeos), devido ao fato de que os mais variados organismos podem utilizar códons diferenciados, para codificar o mesmo aminoácido.

A inferência de prováveis funções das proteínas codificadas pelo genoma de *G. diazotrophicus* PAL5, com base nos resultados de Blastp obedeceram os seguintes critérios:

1- o alinhamento deve ter *e-value*¹ inferior a 10^{-5} ; 2- Porcentagens de identidade superiores a 30%; e 3- Alinhamento de 60% ou mais da seqüência da ORF, foram considerados suficientes

¹ Valor de expectativa. Este valor equivale à probabilidade do alinhamento ocorrer por acaso. Quanto mais próximo de zero, maior será a chance do alinhamento ter significado biológico.

para caracterizar homologia, com seqüências de proteínas depositadas em banco de dados. Quando não houve detecção de sobreposição de seqüências de ORFs, e mesmo assim o resultado de Blast indicou uma similaridade muito baixa ou nenhuma em relação aos bancos de dados, estas seqüências foram classificadas como seqüências de proteínas hipotéticas (Sistema Bionotes, 2006b).

Posteriormente, realizou-se uma análise na base de dados PFAM (Bateman, 2004), que apresenta uma coleção de famílias e domínios protéicos. O PFAM possui seqüências conservadas de cerca de 6200 famílias proteicas, das quais aproximadamente 75% das proteínas encontradas no SWISS-PROT (Boeckmann *et al.*, 2003), possuem pelo menos um domínio do PFAM, e cada resultado relatado no PFAM acompanha um comentário sobre o domínio encontrado, facilitando a identificação da possível função da ORF. Após a predição da possível função das ORFs, estas foram validadas e anotadas às suas respectivas categorias e subcategorias funcionais, assim como o nome do gene, número de EC (*Enzyme category*), número de COG (*Clusters of Orthologous Groups*), organismos relacionados e comentários adicionais dos pesquisadores acerca da ORF.

Adicionalmente, na tentativa de determinar prováveis funções para ORFs de algumas proteínas hipotéticas, foram analisados os alinhamentos de seqüências do *G. diazotrophicus* PAL5 (Blastp), que apresentaram alinhamentos com *e-values* inferiores a 10^{-3} , e alinhavam com domínios protéicos conservados depositados no PFAM. As proteínas hipotéticas com valores de *e-value* superiores a 10^{-3} nos seus alinhamentos, continuaram a receber esta classificação.

A mineração dos dados de anotação de todas as ORFs validadas, foi realizada através do acesso direto no banco de dados do sistema Bionotes de anotação, onde todas as informações pertinentes as ORFs validadas pelos membros integrantes do Projeto Riogene,

foram acessadas para desenvolvimento de análises estatísticas e observação das principais particularidades do genoma deste importante endófito.

4 – RESULTADOS

4.1 – Digestões do cromossomo do *G. diazotrophicus* PAL5 e perfis de PFGE

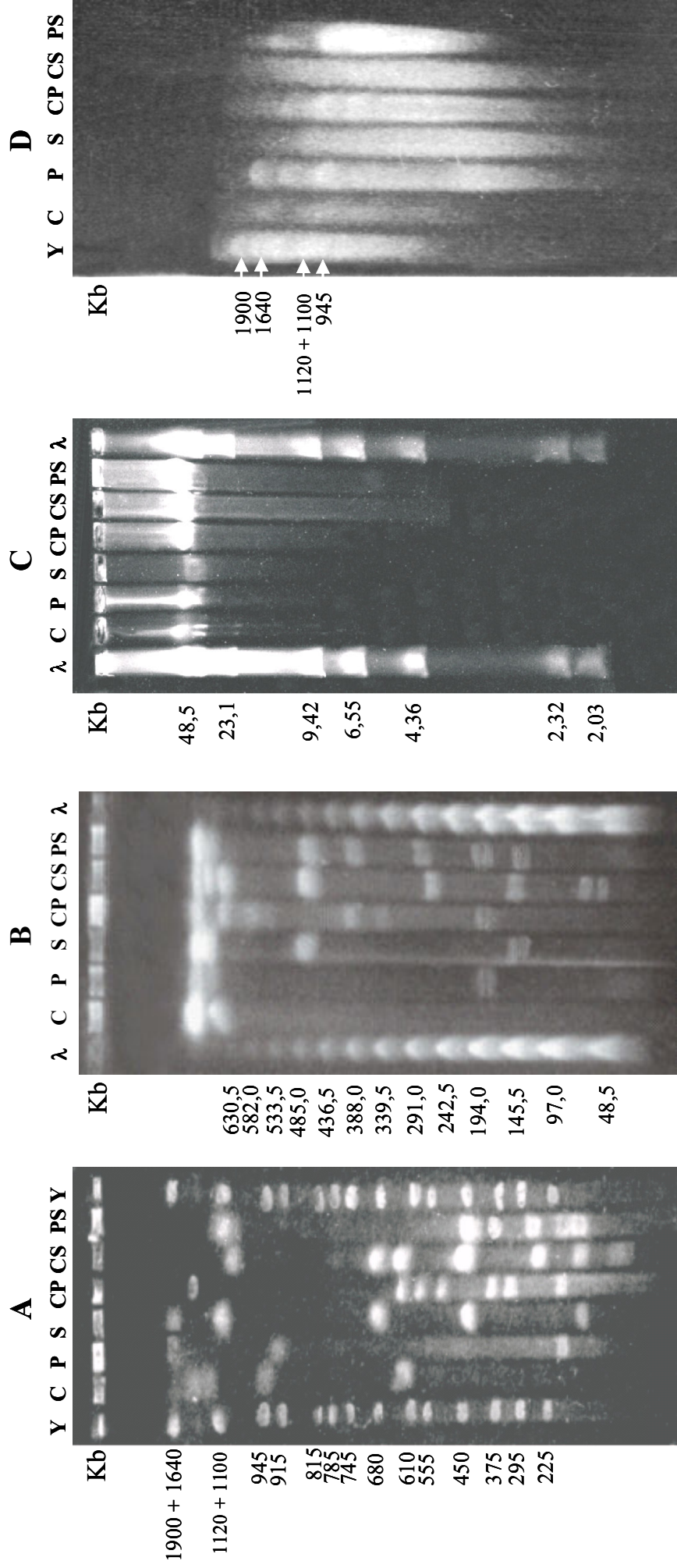
O cromossomo do *G. diazotrophicus* PAL5 foi digerido através de digestões simples e duplas, as quais geraram 6 diferentes perfis de PFGE. As digestões simples consistiram em digestões com as enzimas de restrição de raro corte *I-CeuI*, *PacI* e *SwaI*, que geraram perfis de PFGE identificados como C, P e S, respectivamente. As digestões duplas consistiram em digestões com as seguintes combinações de enzimas de restrição: *I-CeuI* + *PacI*, *I-CeuI* + *SwaI* e *PacI* + *SwaI*, que geraram perfis de PFGE identificados como CP, CS e PS, respectivamente (Figura 2 e Tabela 1).

Os géis de PFGE foram submetidos a 4 diferentes programas de eletroforese (Figura 2A-D), para separação de todos os fragmentos de DNA detectáveis por esta metodologia. Contudo, os géis produzidos sob as condições de eletroforese da figura 2A, foram considerados mais representativos, devido ao fato de que os perfis de PFGE exibiam todos os fragmentos de DNA detectados neste estudo.

Nas outras 3 condições de eletroforese utilizadas (Figura 2B-D), não houve detecção de fragmento de DNA com tamanho acima de 1634 Kb (Figura 2D), ou fragmento de DNA com tamanho compreendido entre 2 e 70 Kb (Figura 2B-C). Adicionalmente, sob as condições de PFGE empregadas na figura 2B, alguns fragmentos de DNA classificados neste trabalho como bandas duplas puderam ser separados, ex: bandas P4 e P5, bandas S5 e S6 (Tabela 1), o que confirmou a existência destes fragmentos. Os perfis de PFGE gerados por digestões simples revelaram 4 a 6 diferentes fragmentos de DNA, com tamanhos variando de 130 a 1634 Kb, enquanto que as digestões duplas, exibiram de 9 a 11 fragmentos de DNA com tamanhos variando de 20 a 1357 Kb (Figura 2 e Tabela 1).

Os fragmentos de DNA gerados em cada perfil de PFGE foram nomeados de acordo com as enzimas de restrição que os criaram (C, P, S, CP, CS ou PS), seguido por um número

Figura 2 – Perfis de PFGE do DNA genômico de *Gluconacetobacter diazotrophicus* PAL5 observado em 4 diferentes condições de eletroforese para detecção de todos os fragmentos de DNA



Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; λ – marcador de PFGE *Lambda ladder* (50 – 1000 Kb) ou marcador de PFGE *Lambda Low Range* (2,03 – 194 Kb); **A**- Eletroforese realizada em gel de agarose a 1% e TBE 0,5× a 6 V/cm e 13 °C, numa seqüência de 3 blocos (bloco A consistiu de um tempo de pulso inicial de 80 s e um final de 100 s durante 16 h; bloco B consistiu de um tempo de pulso inicial de 100 s, e um final de 120 s durante 8 hs; e o bloco C consistiu de um tempo de pulso inicial de 120, e um final de 140 s durante 16 h); **B** – Eletroforese realizada em gel de agarose 1% e TBE 0,5× a 6 V/cm e 13 °C com tempo de pulso inicial de 1 s e final de 50 s durante 23 hs. **C** – Eletroforese realizada em gel de agarose a 1% e TAE 1× a 4 V/cm e 13 °C com tempo de pulso inicial de 0,1 e final de 10 s durante 16 h; **D**- Eletroforese realizada em gel de agarose a 0,8% e TBE 0,5× a 2 V/cm e 12 °C, numa seqüência de 2 blocos (bloco A consistiu de um tempo de pulso de 1.000 s durante 24 h; seguido do bloco B que consistiu de um tempo de pulso inicial de 1.000 e um final de 4.000 s durante 192 h.

Tabela 1 – Número e tamanhos estimados dos fragmentos de DNA, gerados através de digestões simples e duplas do cromossomo do *Gluconacetobacter diazotrophicus* PAL5

Fragmento	Digestões Simples						Digestões Duplas					
	<i>I-CeuI</i>		<i>PacI</i>		<i>SwaI</i>		<i>I-CeuI + PacI</i>		<i>I-CeuI + SwaI</i>		<i>PacI + SwaI</i>	
	(Kb) ^e	Fragmento	(Kb)	Fragmento	(Kb)	Fragmento	(Kb)	Fragmento	(Kb)	Fragmento	(Kb)	Fragmento
C1	1400	P1	1577	S1	1634	CP1	1357	CS1	1052	PS1	1094	
C2	1235	P2	≅ 1400 ^a	S2	1152	C4	625	S3	680	PS2	1066 ^a	
C3	965	P3	927	S3	680	CP2	576	CS2	≅ 680 ^a	S4	469	
C4	625	P4	200	S4	467	CP3	508	C4	625	PS3	390	
-	-	P5	≅ 200 ^a	S5	142	CP4	393	S4	468	PS4	285	
-	-	-	-	S6	130	CP5	352	CS3	260	PS5	≅ 285 ^a	
-	-	-	-	-	-	P4	200	CS4	152	P4	200	
-	-	-	-	-	-	P5	≅ 200 ^a	S6	130	P5	≅ 200 ^a	
-	-	-	-	-	-	CP6	<20 ^b	CS5	80	S5	142	
-	-	-	-	-	-	-	-	CS6	70	S6	130	
-	-	-	-	-	-	-	-	-	-	PS6	<20 ^b	
TOTAL^c	4225		≅ 4300		4205		≅ 4211^d		≅ 4197		≅ 4261^d	

a – Banda dupla; b – Fragmento estimado pela digestão do cromossomo, porém não observado em géis de PFGE; c – A média das somas dos fragmentos de restrição gerados nas digestões simples e duplas (C, P, S e CS) do cromossomo do *G. diazotrophicus* PAL5 foi de 4232 Kb com desvio padrão de 47 Kb, desconsiderando os perfis que tiveram fragmentos não observados em géis de PFGE (CP e CS); d – Este total desconsiderou o tamanho estimado dos fragmentos não observados em géis de PFGE; e – Kb, Quilobases.

arábico seqüencial, de acordo com o decréscimo de tamanho dos fragmentos analisados em cada perfil, como exemplo temos os fragmentos C1, C2, C3 e C4 (gerados na digestão simples com a endonuclease *I-CeuI*), que apresentam tamanhos estimados em 1400, 1235, 965 e 625 Kb, respectivamente (Tabela 1). Nas digestões duplas, quando um determinado fragmento de DNA não foi digerido pela segunda enzima de restrição, este fragmento recebeu a mesma nomenclatura da digestão simples, como exemplo temos o perfil de PFGE CP (gerado na digestão dupla com as endonucleases *I-CeuI* e *PacI*), que apresentam os fragmentos C4, P4 e P5 (Tabela 1). Adicionalmente, sob diferentes condições de PFGE, o DNA genômico de *G. diazotrophicus* PAL5 não digerido, não entrou na malha do gel de agarose, sugerindo que este endófito apresenta um cromossomo circular (Dados não exibidos).

A Tabela 1 exibe o número e o tamanho estimado de 45 fragmentos de DNA, gerados através de digestões simples e duplas do cromossomo do *G. diazotrophicus* PAL5. Dentre estes, 2 fragmentos de DNA denominados CP6 e PS6 (obtidos nas digestões duplas CP e PS, respectivamente), não foram observados nos experimentos conduzidos neste estudo. Por outro lado, se considerarmos o número de fragmentos de DNA observados nas digestões simples, o número de fragmentos de DNA esperado nas digestões duplas, a soma dos tamanhos dos demais fragmentos de DNA detectados nos perfis de PFGE CP e PS, e o tamanho estimado do genoma, podemos sugerir a existência destes fragmentos de DNA, e estimar que apresentam tamanhos inferiores a 20 Kb. O fato dos fragmentos CP6 e PS6 não serem observados nos géis de PFGE, sugerem a ocorrência de baixos sinais fluorescentes nestes fragmentos de DNA, quando analisados em géis corados com brometo de etídio ou *syber green*.

A estimativa de tamanho do cromossomo do *G. diazotrophicus* PAL5 em cada perfil de PFGE, foi obtido a partir do somatório dos tamanhos de todos os fragmentos de DNA observados nos perfis de PFGE, gerados a partir de digestões simples e duplas do

cromossomo, onde estes valores variaram de 4197 Kb a 4300 Kb. O tamanho do cromossomo foi estimado em 4232 Kb, com desvio padrão de ± 47 Kb, através da média dos valores de tamanhos de cromossomo, estimados para cada perfil de PFGE. (Tabela 1).

4.2 – Determinação do número e orientação dos operons ribossomais

Através da digestão do cromossomo do *G. diazotrophicus* PAL5 com a enzima de restrição *I-CeuI*, a qual reconhece uma seqüência de 26 pb altamente conservada na subunidade 23S rRNA (gene *rrl*) (Ramos-Díaz & Ramos, 1998; Zé-Zé, Tenreiro & Paveia, 2000; El-Osta *et al.*, 2002), foi verificado a presença de 4 fragmentos de DNA em gel de agarose, sugerindo a existência de 4 operons ribossomais neste genoma (Perfil de PFGE C da figura 2A e Tabela 1).

Para confirmar esta hipótese, foram realizadas hibridizações com as sondas de DNA dos genes ribossomais 16S (*rrs*), 23S (*rrl*) e 5S (*rrf*) (Tabela 2). Nestes experimentos, esperava-se 4 fragmentos de DNA hibridizados no perfil de PFGE C para cada sonda utilizada, entretanto, foram observados apenas 3 fragmentos de DNA hibridizados neste perfil de PFGE, com cada sonda supracitada. As sondas dos genes *rrs* e *rrl* hibridizaram com os fragmentos C2, C3 e C4, enquanto que a sonda do gene *rrf* hibridizou com os fragmentos C1, C2 e C3 (Figura 3).

4.3- Hibridizações com sondas de genes de cópia única

Baseado nos resultados de hibridização (Figuras 4-15 e Tabela 3), gerados com sondas radioativas de genes de cópia única (Tabela 2), foram verificadas sobreposições entre fragmentos individuais de DNA, nos 6 diferentes perfis de PFGE do DNA cromossomal de *G. diazotrophicus* PAL5, uma vez que estas sondas demonstraram hibridizações com apenas 1

Tabela 2 – Descrição das sondas de DNA dos genes de cópia única e ribossomais, utilizadas para construção do mapa físico do cromossomo do *Gluconacetobacter diazotrophicus* PAL5

Gene ^a	Função da seqüência ^b	Fonte de seqüência similar ^c	Probabilidade ^d
<i>acoC</i>	Dihidrolipoamido acetiltransferase	<i>Pseudomonas fluorescens</i> Pf-5	5e-86
<i>araE</i>	Transportador simporte de açúcar-próton	<i>Streptomyces ambofaciens</i> ATCC 23877	9e-20
<i>araJ</i>	Permease de efluxo de arabinose	<i>Magnetospirillum magnetotacticum</i> MS-1	7e-14
<i>clpP</i>	Subunidade proteolítica da protease Clp dependente de ATP	<i>Gluconobacter oxydans</i> 621H	1e-88
<i>csp</i>	Proteína de choque ao frio	<i>Gluconobacter oxydans</i> 621H	5e-24
<i>dhbF</i>	Peptídeo sintase não ribossomal DhbF	<i>Erwinia carotovora</i> subsp. atroseptica SCRI1043	3e-68
<i>dnaE</i>	Cadeia alfa da DNA polimerase III	<i>Gluconobacter oxydans</i> 621H	5e-131
<i>fcuA</i>	Receptor ferricromo FcuA	<i>Gluconobacter oxydans</i> 621H	5e-89
<i>fumC</i>	Fumarato hidratase classe II	<i>Rhodospseudomonas palustris</i> CGA009	1e-55
<i>glhB</i>	Cadeia pesada da glutamato sintase (NADPH)	<i>Acidiphilium cryptum</i> JF-5	5e-120
<i>gntP</i>	Permease transportadora de Gluconato	<i>Gluconobacter oxydans</i> 621H	8e-111
<i>hemE</i>	Uroporfirinogênio III descarboxilase	<i>Gluconobacter oxydans</i> 621H	3e-109
<i>hemH</i>	Ferroquelatase	<i>Acidiphilium cryptum</i> JF-5	6e-51
<i>hlyD</i>	Proteína de secreção da família HlyD	<i>Ralstonia eutropha</i> JMP134	1e-36
<i>hutH</i>	Proteína histidina amonia-liase (Histidase)	<i>Gluconobacter oxydans</i> 621H	6e-64
<i>ilvD</i>	Dihidroxiácido desidratase	<i>Gluconobacter oxydans</i> 621H	2e-92
<i>kdpC</i>	Cadeia C da ATPase transportadora de potássio	<i>Granulibacter bethesdensis</i> CGDNIH1	3e-57
<i>kup</i>	Proteína de aquisição de potássio	<i>Bradyrhizobium</i> sp. BTAi1	6e-69
<i>lhr</i>	Superfamília II de helicases e derivados inativados	<i>Burkholderia dolosa</i> AUO158	6e-22

Tabela 2 – Continuação.

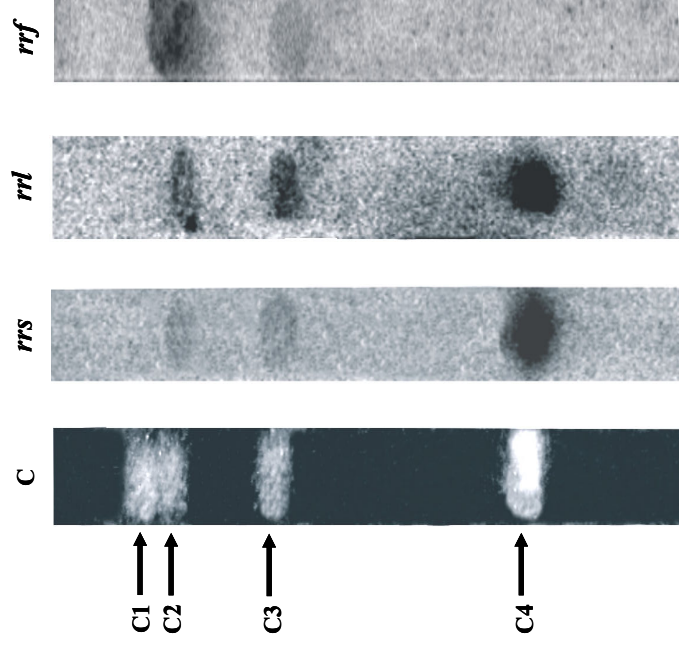
Gene^a	Função da seqüência^b	Fonte de seqüência similar^c	Probabilidade^d
<i>lig3</i>	ATP-dependente DNA ligase	<i>Rhizobium</i> sp. NGR234	2e-107
<i>lsdA</i>	Precursor da levansucrase	<i>Gluconacetobacter diazotrophicus</i>	3e-24
<i>mcpA</i>	Proteína de quimiotaxia aceptora de metil	<i>Rhodospirillum centenum</i>	9e-60
<i>metE</i>	5-metiltetrahidropterilglutamato-homocisteína metiltransferase	<i>Caulobacter crescentus</i> CB15	1e-111
<i>mexI</i>	Transportador RND de efluxo de múltiplas drogas	<i>Pseudomonas entomophila</i> L48	2e-98
<i>mfs</i>	Proteína de fusão de membrana MFS de efluxo de múltiplas drogas	<i>Acinetobacter baumannii</i> ATCC 17978	7e-42
<i>mgtA</i>	ATPase importadora de Mg ²⁺	<i>Bradyrhizobium japonicum</i> USDA 110 DNA	2e-59
<i>nifA</i>	Regulador transicional específico da Mo/Fe nitrogenase	<i>Gluconacetobacter diazotrophicus</i>	1e-133
<i>nuoG</i>	Subunidade Gama da NADH desidrogenase	<i>Nitrosococcus oceanii</i> ATCC 19707	5e-41
<i>pdhA</i>	Componente alfa da piruvato desidrogenase E1	<i>Gluconobacter oxydans</i> 621H	1e-73
<i>pepC</i>	Fosfoenolpiruvato carboxilase	<i>Granulibacter bethesdensis</i> CGDNIHI	3e-132
<i>pilT</i>	Receptor de membrana externa similar a proteína PilT	<i>Serratia proteamaculans</i> 568	6e-11
<i>pyrB</i>	Aspartato carbamoiltransferase	<i>Gluconobacter oxydans</i> 621H	1e-98
<i>rrf</i>	RNA ribossomal 5S	<i>Bartonella henselae</i> strain Houston-1	2e-36
<i>rriI</i>	RNA ribossomal 23S	<i>Acetobacter intermedius</i>	0.0
<i>rriS</i>	RNA ribossomal 16S	<i>Gluconacetobacter diazotrophicus</i>	0.0
<i>sucD</i>	Sub-unidade alfa da succinil-CoA sintetase	<i>Ralstonia metallidurans</i> CH34	1e-56
<i>trpA</i>	Proteína de transportadora de membrana induzida por tolueno	<i>Serratia proteamaculans</i> 568	4e-16
<i>ubiE</i>	Metiltransferase dependente de SAM	<i>Mesorhizobium</i> sp. BNC1	4e-50
<i>xdhA</i>	Xantina desidrogenase	<i>Gluconobacter oxydans</i> 621H	5e-150

Tabela 2 – Continuação.

Gene^a	Função da seqüência^b	Fonte de seqüência similar^c	Probabilidade^d
<i>xseA</i>	Subunidade maior da exodeoxirribonuclease VII	<i>Gluconobacter oxydans</i> 621H	8e-42
<i>xyfF</i>	Transportador ABC de D-xilose, proteína de ligação periplasmática de D-xilose	<i>Pseudomonas syringae</i> pv. tomato cepa DC3000	7e-71
<i>zntA</i>	ATPase transportadora de cátions tipo P	<i>Sinorhizobium meliloti</i>	6e-58

a – nomes dos genes que foram utilizados para construção de sondas; b – Função das seqüências utilizadas para construção de sondas; c- microrganismos que apresentam melhores alinhamentos em relação a esta seqüência no banco de dados NCBI, utilizando-se a ferramenta Blastn; d- valor de *e-value* observado no alinhamento da seqüência da sonda contra o banco de dados NCBI utilizando a ferramenta Blastn.

Figura 3 – Híbridizações obtidas com sondas dos genes dos RNAs ribossomais, contra perfil de PFGE do cromossomo do *Gluconacetobacter diazotrophicus* PAL5 gerado com a enzima *I-CeuI*.



C1, C2, C3 e C4 – Fragmentos de DNA gerados a partir da digestão do cromossomo de *G. diazotrophicus* PAL5 com a enzima *I-CeuI*; *rrs* – sonda para o gene do RNA ribossomal 16S; *rrl* – sonda para o gene do RNA ribossomal 23S; *rrf* – sonda para o gene do RNA ribossomal 5S.

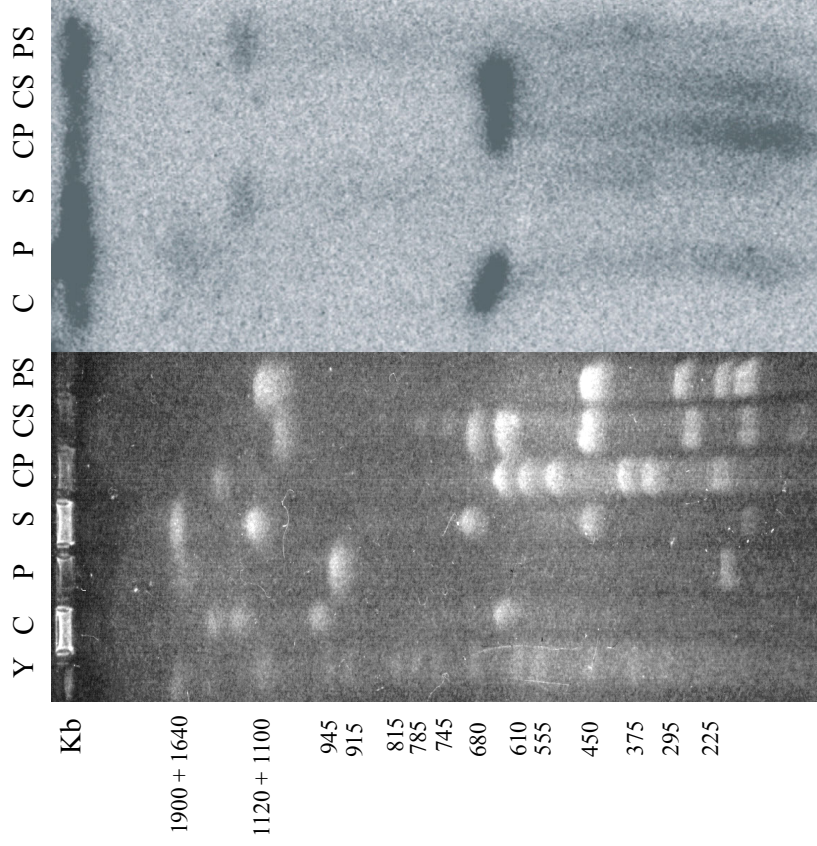
fragmento de DNA em cada perfil de PFGE. Um exemplo desta abordagem é demonstrado na Figura 4, onde as sondas radioativas dos genes *hemE*, *hemH*, *kdpC*, *mfs* e *pyrB*, demonstram sobreposição entre os fragmentos de DNA C4, P2, S2 e PS1 contidos nestes perfis de PFGE (Tabela 3).

Desta mesma forma, todos os resultados de hibridização foram analisados e anotados (Figuras 5 a 15), permitindo a detecção de 12 grupos de genes, que demonstraram diferentes perfis de hibridização (Tabela 3), indicando proximidade de genes alocados no mesmo grupo. Adicionalmente, estes resultados foram utilizados para determinar o posicionamento de marcadores moleculares no cromossomo bacteriano.

Todos os fragmentos de DNA hibridizaram com pelo menos uma das 42 sondas radioativas utilizadas neste estudo, exceto os fragmentos P5, S6, CP6 e PS6. A ausência de hibridizações nos fragmentos P5 e S6, pode ser explicado devido ao fato de que o mapa físico do cromossomo do *G. diazotrophicus* PAL5, foi construído com a utilização de número reduzido de sondas, portanto, não foram desenhadas sondas para estes fragmentos, porém os mesmos foram detectados no gel de PFGE da figura 2B. Com relação aos fragmentos CP6 e PS6, cabe ressaltar que estes fragmentos não foram detectados em nenhum gel de PFGE, e sugere-se que possuam tamanhos inferiores a 20 Kb. Os 4 fragmentos de DNA que não apresentaram sinais de hibridização, foram montados no mapa físico do cromossomo do *G. diazotrophicus* PAL5 (Figura 16), através de análise combinatória de tamanho dos fragmentos de DNA, e posicionamento dos sítios de restrição da enzima que gerou o fragmento.

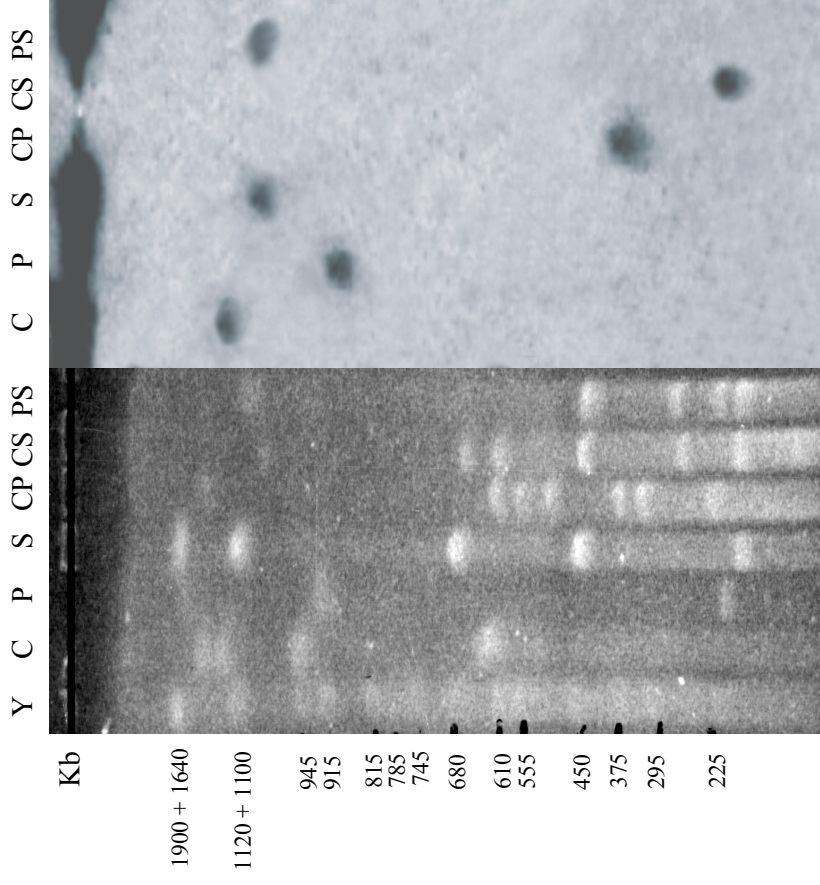
Os fragmentos de DNA P1 e P2 (Perfil de PFGE P da Figura 2A; Tabela 1), CS2 e S3 (Perfil de PFGE CS da Figura 2A; Tabela 1), PS1 e PS2, PS4 e PS5 (Perfil de PFGE PS da Figura 2A; Tabela 1), são indistinguíveis sob as condições de eletroforese empregadas nos géis de PFGE. Contudo, os resultados de hibridização com sondas de genes de cópia única,

Figura 4 – Híbridizações obtidas com sondas dos genes de cópia única *hemE*, *hemH*, *kdpC*, *mfs* e *pyrB*^a



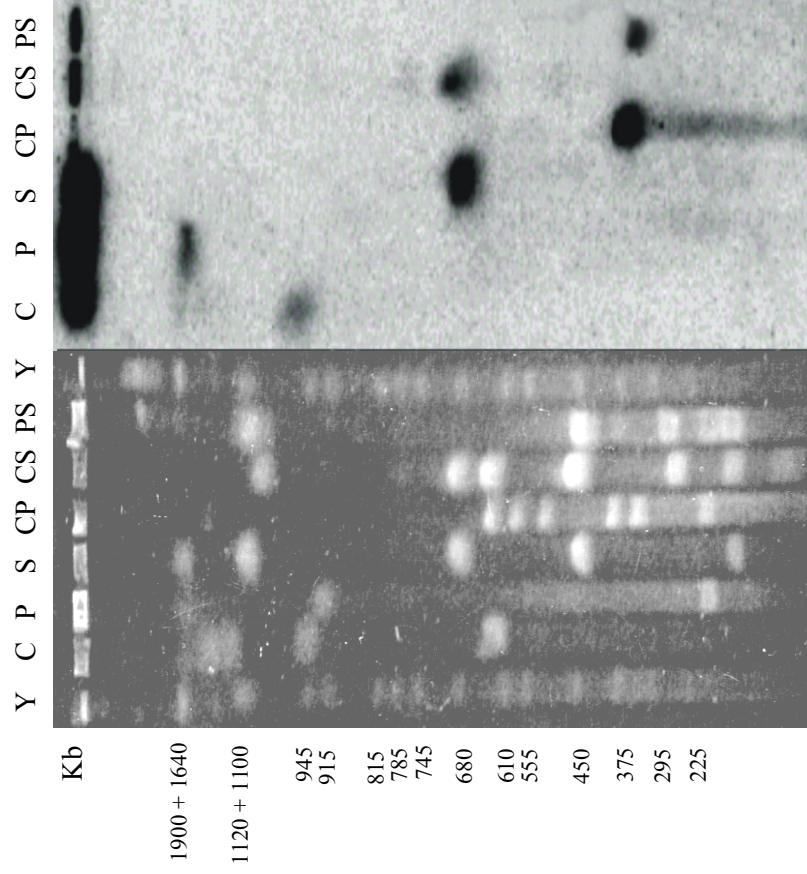
a – As sondas destes cinco genes deram os mesmos resultados de hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y- cromossomos de levedura; *hemE* – gene da enzima uroporfirinogênio III descarboxilase; *hemH* – gene da enzima ferroquelatase; *kdpC* – gene da cadeia C da ATPase transportadora de potássio; *mfs* – gene da proteína fusionada a membrana de efluxo de múltiplas drogas MFS; *pyrB* – gene da enzima aspartato carbamoiltransferase.

Figura 5 – Híbridizações obtidas com sonda do gene de cópia única *dhbF*



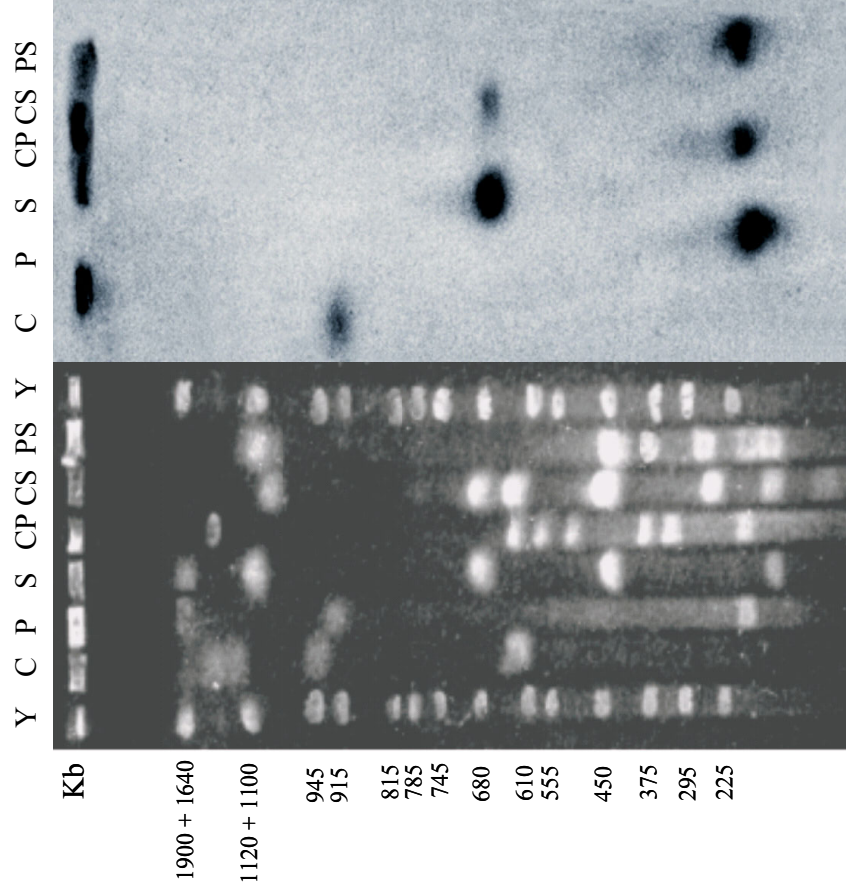
Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y - cromossomos de levedura; *dhbF* - Peptídeo sintetase não-ribossomal DhbF.

Figura 6 – Hibridizações obtidas com sondas dos genes de cópia única *lig3*, *mexI* e *mgtA*^a



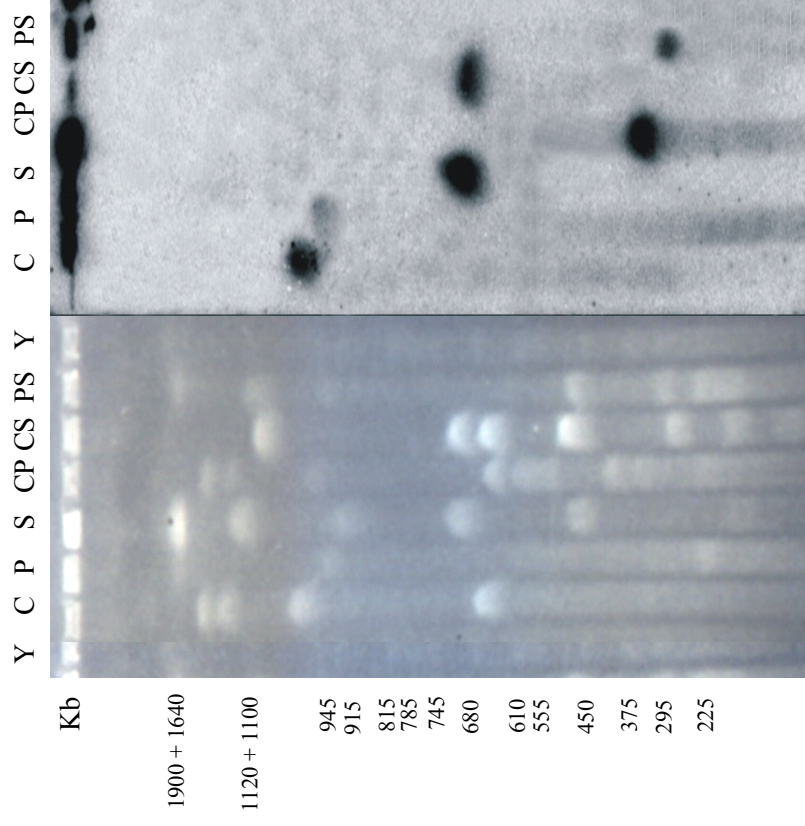
a – As sondas destes três genes deram os mesmos resultados de hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – cromossomos de levedura; *lig3* – gene da enzima DNA ligase dependente de ATP; *mexI* – gene do transportador de efluxo de múltiplas drogas RND; *mgtA* – gene da ATPase importadora de Mg²⁺.

Figura 7 – Híbridizações obtidas com sondas dos genes de cópia única *lsdA*, *nifA* e *pilT*^a



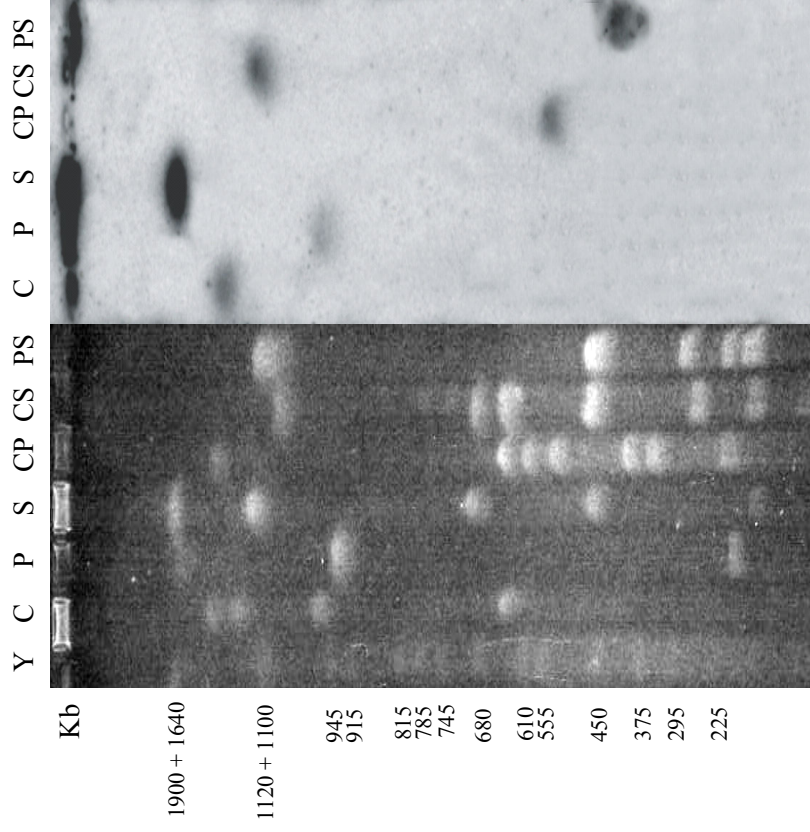
a – As sondas destes três genes deram os mesmos resultados de hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; *lsdA* – gene do precursor da levansucrase; *nifA* – gene do regulador transcricional específico da nitrogenase Mo/Fe; *pilT* – gene do receptor de membrana externa, similar a proteína PilT.

Figura 8 – Híbridos obtidos com sondas dos genes de cópia única *fumC*, *hutH*, *kup* e *ubiE*^a



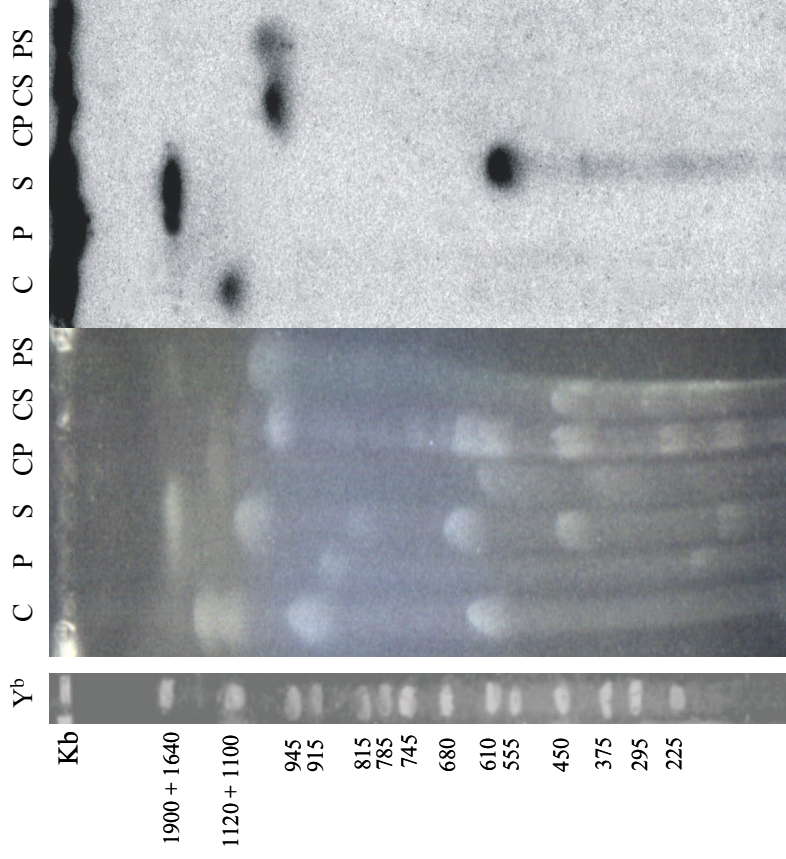
a – As sondas destes quatro genes deram os mesmos resultados de hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; *fumC* – gene da enzima fumarato hidratase classe II; *hutH* – gene da proteína histidina amônia-liase (Histidase); *kup* – gene da proteína de assimilação de potássio; *ubiE* – gene da enzima metiltransferase dependente de SAM.

Figura 9 – Hibridizações obtidas com sondas dos genes de cópia única *araE*, *araJ*, *clpP*, *csp*, *fcuA*, *hlyD*, *ilvD*, *lhr* e *xseA*^a



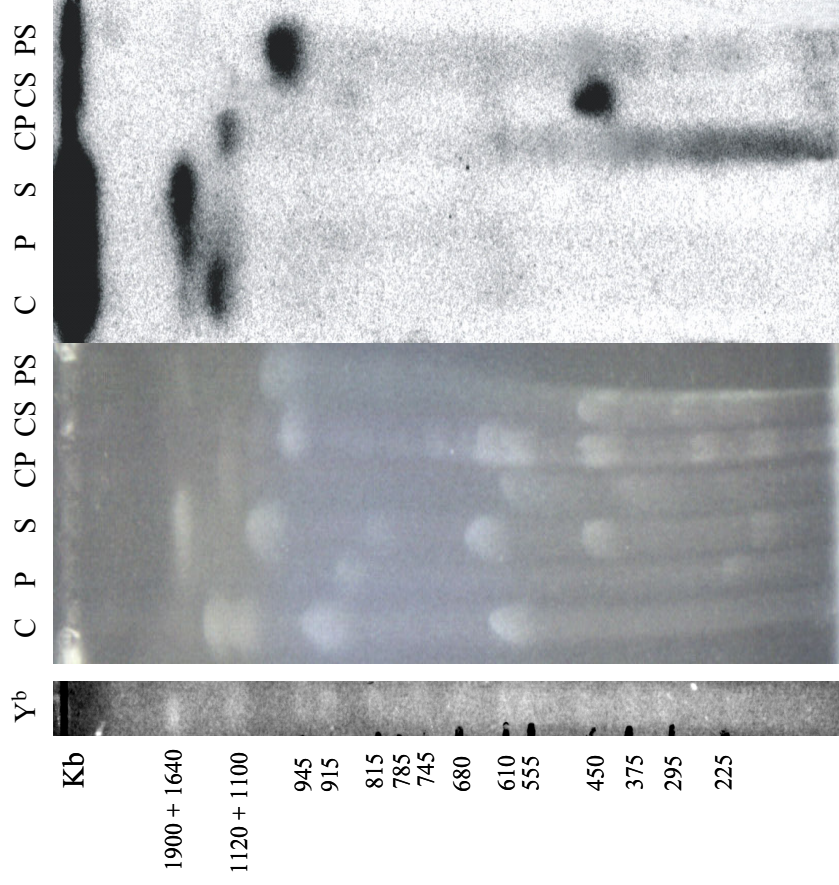
a – As sondas destes nove genes deram os mesmos resultados de hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; *araE* – gene do transportador simporte de açúcar-próton; *araJ* – gene da permease de efluxo de arabinose; *clpP* – gene da sub-unidade proteolítica da protease Clp dependente de ATP; *csp* – gene da proteína de choque ao frio; *fcuA* – gene do receptor ferricromo FcuA; *hlyD* – Proteína de secreção da família hlyD; *ilvD* – gene da enzima dihidroxiácido desidratase; *lhr* – gene da helicase da superfamília II e derivados inativados; *xseA* – gene da grande sub-unidade da exodeoxirribonuclease VII.

Figura 10 – Híbridizações obtidas com sondas dos genes de cópia única *dnaE*, *gntP*, *mcpA*, *pepC* e *sucD*^a



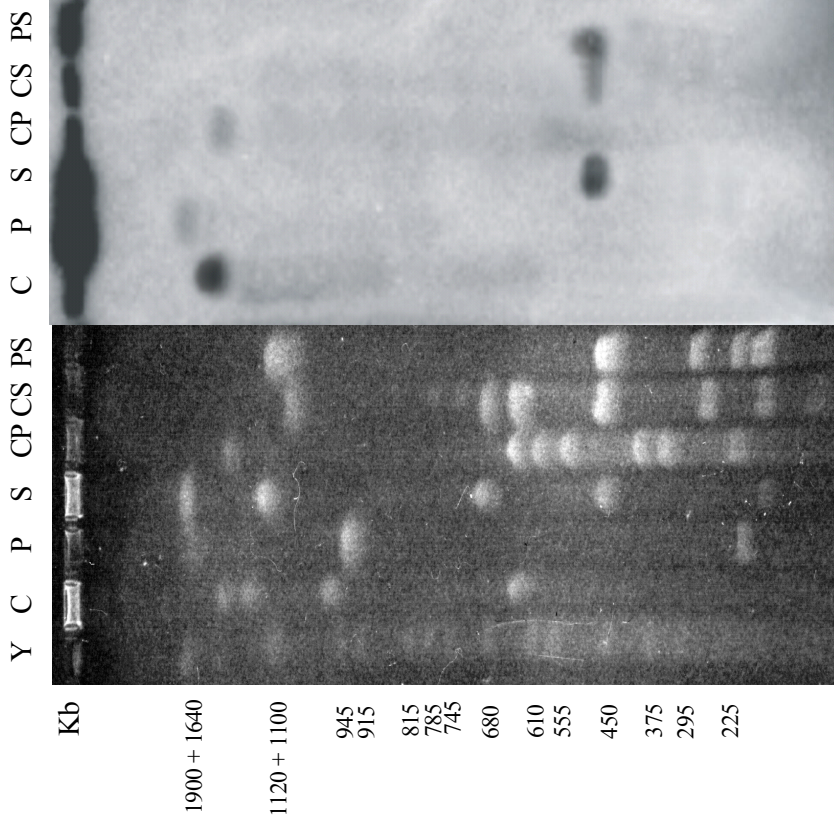
a – As sondas destes cinco genes deram os mesmos resultados de hibridização. b – perfil de eletroforese extraído do mesmo gel que foi utilizado para hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*J-CeuI*), P (*PacI*), S (*SwaI*), CP (*J-CeuI* + *PacI*), CS (*J-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; *dnaE* – gene da cadeia alfa da DNA polimerase III; *gntP* – gene da permease transportadora de gluconato; *mcpA* – gene da proteína de quimiotaxia aceptora de metil; *pepC* – gene da enzima fosfoenolpiruvato carboxilase; *sucD* – gene da subunidade alfa da succinil-CoA sintetase

Figura 11 – Híbridizações obtidas com sondas dos genes de cópia única *acoC*, *pdhA*, *trpA*, *xdhA* e *zntA*^a



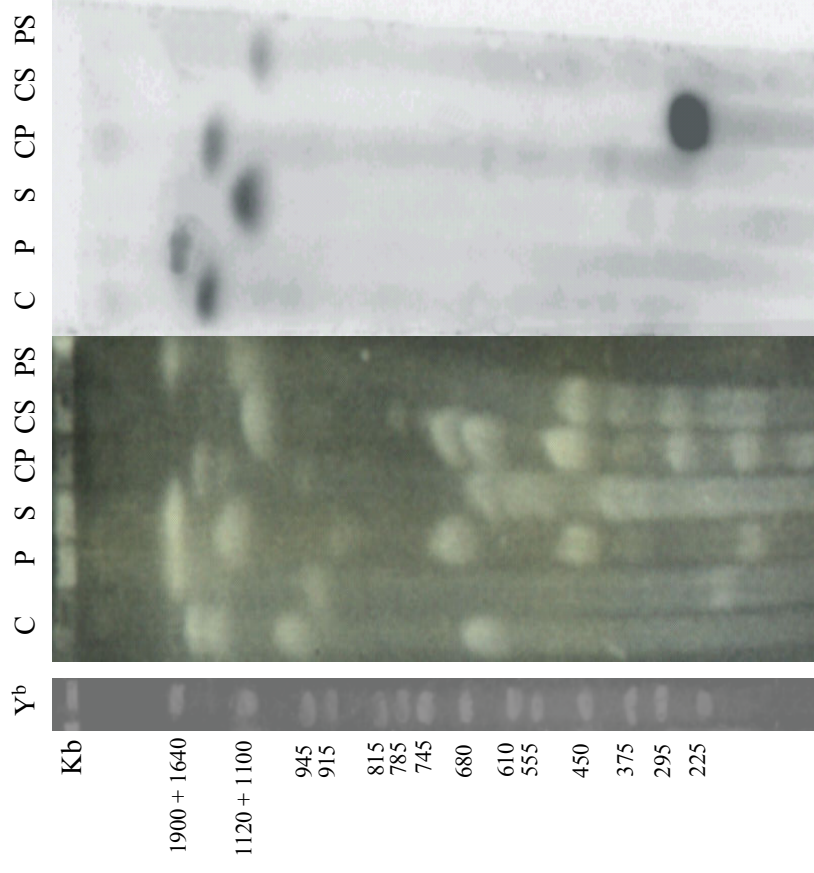
a – As sondas destes cinco genes deram os mesmos resultados de hibridização. b – perfil de eletroforese extraído do mesmo gel que foi utilizado para hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; *acoC* – gene da enzima dihidrolipoamido acetiltransferase; *pdhA* – gene do componente alfa da piruvato desidrogenase E1; *trpA* – gene da proteína de transporte interno de membrana induzida por bomba de exclusão de tolueno; *xdhA* – gene da proteína xantina desidrogenase XdhA; *zntA* – gene da ATPase transportadora de cátion tipo P.

Figura 12 – Híbridizações obtidas com sondas dos genes de cópia única *gltB* e *nuoG*^a



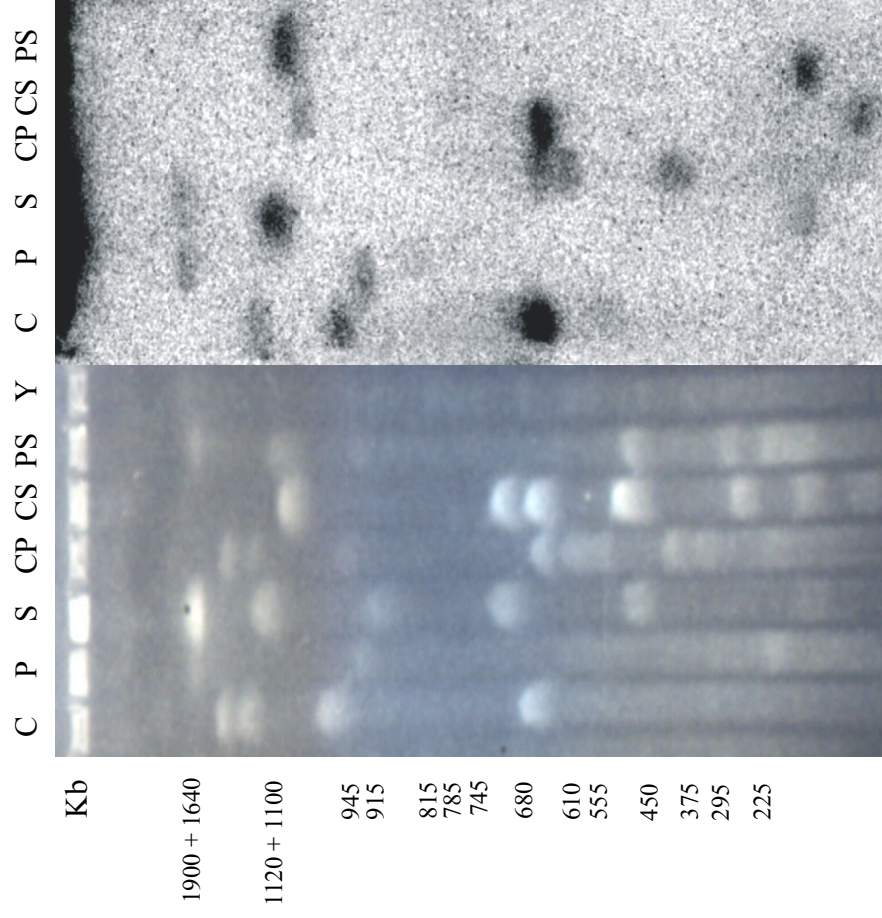
a – As sondas destes dois genes deram os mesmos resultados de hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; *gltB* – gene da grande cadeia da glutamato sintetase (NADPH); *nuoG* – gene da cadeia G da NADH-quinona oxidoreductase.

Figura 13 – Híbridizações obtidas com sondas dos genes de cópia única *metE* e *xyIF*^a



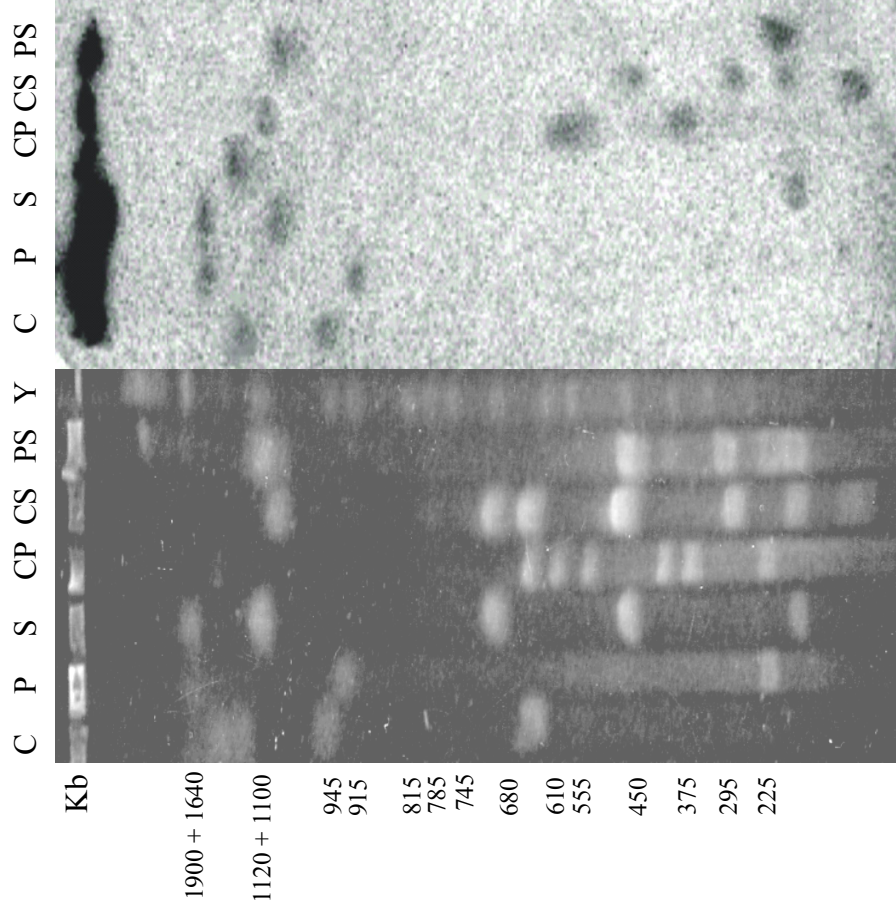
a – As sondas destes dois genes deram os mesmos resultados de hibridização. b – Perfil de eletroforese extraído do mesmo gel que foi utilizado no experimento de hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; *metE* – gene da enzima 5-metiltetrahidropteroil triglutamato-homocisteína metiltransferase; *xyIF* – gene do transportador ABC de D-xilose, proteína de ligação de D-xilose periplasmática.

Figura 14 – Híbridizações obtidas com sondas dos genes ribossomais *rrl* e *rrs* ^a



a – As sondas destes dois genes deram os mesmos resultados de hibridização. Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; *rrl* – gene do RNA ribossomal 23S; *rrs* – gene do RNA ribossomal 16S.

Figura 15 – Híbridizações obtidas com sonda do gene ribossomal *rrf*



Perfis de PFGE gerados a partir de digestões simples ou duplas: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* + *PacI*), CS (*I-CeuI* + *SwaI*), PS (*PacI* + *SwaI*); Y – Cromossomos de levedura; *rrf* – gene do RNA ribossomal 5S.

Tabela 3 – Resultados obtidos nos experimentos de hibridização com sondas dos genes de cópia única e ribossomais nos 6 diferentes perfis de PFGE.

Grupo ^a	Genes ^b	<i>I-CeuI</i> ^c	<i>PacI</i> ^c	<i>SwaI</i> ^c	<i>I-CeuI + PacI</i> ^c	<i>I-CeuI + SwaI</i> ^c	<i>PacI + SwaI</i> ^c
1	<i>hemE, hemH, kdpC, mfs</i> e <i>pyrB</i>	C4	P2	S2	C4	C4	PS1
2	<i>dhbF</i>	C3	P2	S2	CP4	CS4	PS1
3	<i>lig3, mexI</i> e <i>mgtA</i>	C3	P2	S3	CP4	S3	PS5
4	<i>lsdA, nifA</i> e <i>pilT</i>	C3	P4	S3	P4	S3	P4
5	<i>fumC, hutH, kup</i> e <i>ubiE</i>	C3	P3	S3	CP5	S3	PS4
6	<i>araE, araJ, clpP, csp, fcuA, hlyD, ihvD, lhr</i> e <i>xseA</i>	C2	P3	S1	CP3	CS1	PS3
7	<i>dnaE, gntP, mcpA, pepC</i> e <i>sucD</i>	C2	P1	S1	CP2	CS1	PS2
8	<i>acoC, pdhA, trpA, xdhA</i> e <i>zntA</i>	C1	P1	S1	CP1	CS2	PS2
9	<i>gltB</i> e <i>nuoG</i>	C1	P1	S4	CP1	S4	S4
10	<i>metE</i> e <i>xylF</i>	C1	P1	S2	CP1	CS3	-
11	<i>rrl</i> e <i>rrs</i>	C2/C3/C4	P1/P2/P3	S1/S2/S5	C4/CP2/CP5	CS1/C4/CS5	PS1/PS2/S5
12	<i>rrf</i>	C1/C2/C3	P1/P2/P3	S1/S2/S5	CP1/CP3/CP4	CS2/CS3/CS4/CS6	PS1/PS2/S5

a – Grupo de genes que apresentam o mesmo perfil de hibridização; b – Nome dos genes utilizados para construção de sondas (vide tabela 2); c – Digestões simples e duplas geradas a partir do cromossomo do *G. diazotrophicus* PAL5 (vide tabela 1).

demonstraram que os fragmentos P2, CS2, PS2 e PS5 são bandas duplas, que exibem sobreposições com diferentes fragmentos de DNA, incluídos em outros perfis de PFGE, e conseqüentemente exibiram sinais de hibridizações com sondas de genes alocados em diferentes grupos (Tabela 3).

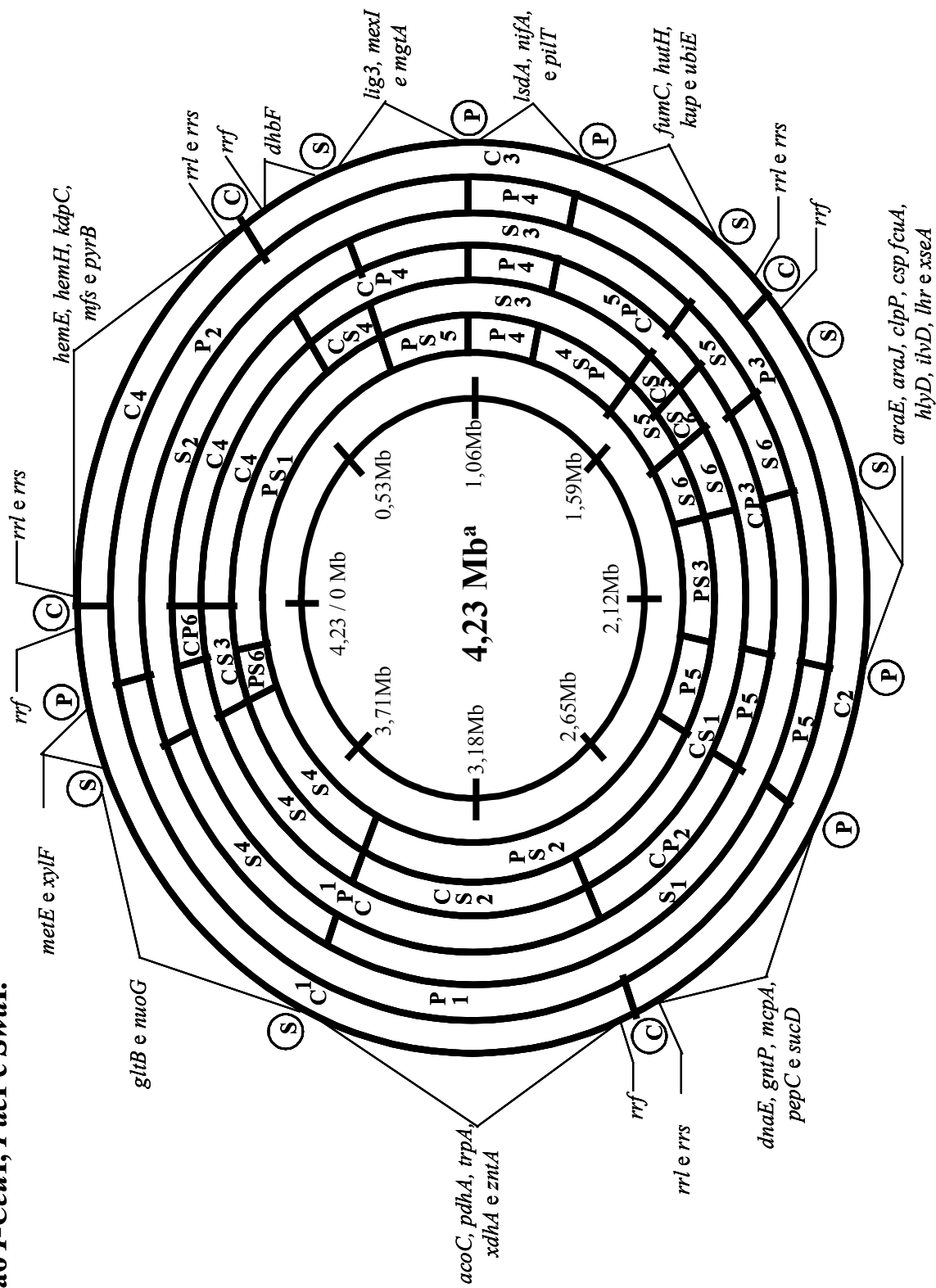
4.4 – Construção do mapa físico do cromossomo do *G. diazotrophicus* PAL5

Considerando sobreposições entre a maioria dos fragmentos de DNA, oriundos dos 6 diferentes perfis de PFGE analisados, com base nos resultados de hibridização utilizando sondas radioativas de genes de cópia única e genes ribossomais (Tabela 3), tamanho dos fragmentos de DNA, existência de bandas duplas nos perfis de PFGE (Tabela 1) e posicionamento dos sítios de restrição das endonucleases utilizadas neste estudo, 45 fragmentos de DNA de 6 diferentes perfis de PFGE, foram ordenados no mapa físico do cromossomo circular de *G. diazotrophicus* PAL5 (Figura 16), utilizando um reduzido número de sondas radioativas, e dispensando o uso de digestões parciais do cromossomo, PFGE bidimensional e clones de ligação. No total, os resultados de hibridização permitiram a alocação de 42 marcadores moleculares no mapa físico do cromossomo do *G. diazotrophicus* PAL5 (Figura 16).

4.5 – Alinhamento de *contigs* do Projeto Riogene no cromossomo do *G. diazotrophicus* PAL5

Quando este mapeamento físico do cromossomo do *G. diazotrophicus* PAL5 foi concluído (Maio de 2006), o seqüenciamento genômico total deste microrganismo atingiu 99,99% das seqüências genômicas, montadas em 18 *contigs* com tamanhos variando de 12,8 a 1251,3 Kb (Dados não exibidos). Adicionalmente, com o posicionamento de 42 marcadores moleculares no mapa físico (Figura 16; Tabela 2), cujas seqüências de DNA estavam alocadas em diferentes *contigs*, foi possível realizar o alinhamento de 16 *contigs* neste cromossomo

Figura 16 – Mapa Físico do Cromossomo do *Gluconacetobacter diazotrophicus* PAL5, utilizando as enzimas de restrição *I-CeuI*, *PacI* e *SwaI*.



C, P e S – sítios de restrição das enzimas *I-CeuI*, *PacI* e *SwaI*, respectivamente. Os fragmentos de restrição foram nomeados de acordo com os critérios estabelecidos nos materiais e métodos (vide Tabela 1). Os fragmentos de DNA CP6 e PS6 foram preditos a partir da digestão do cromossomo e alinhados no mapa físico, porém não foram observados nos géis de PFGE. Os marcadores genéticos foram descritos na Tabela 2. A posição exata dos marcadores moleculares nos fragmentos de DNA não foram determinadas neste mapa físico. O círculo interno representa a escala em megabases do cromossomo do *G. diazotrophicus* PAL5. a – este valor representa a média de tamanho estimado do cromossomo do *G. diazotrophicus* PAL5.

(Figura 17), totalizando 4137,5 Kb à nível de seqüências ordenadas e orientadas, o que consiste em 97,6% de cobertura do cromossomo deste endófito, quando comparado ao tamanho estimado pela metodologia de PFGE. Entretanto, no alinhamento dos *contigs* desta montagem, os *contigs* 5, 8, 9, 10, 11, 14, 15, 17 e 18, não puderam ser precisamente alinhados e ordenados, apesar de todos estarem alocados na mesma região do fragmento C2.

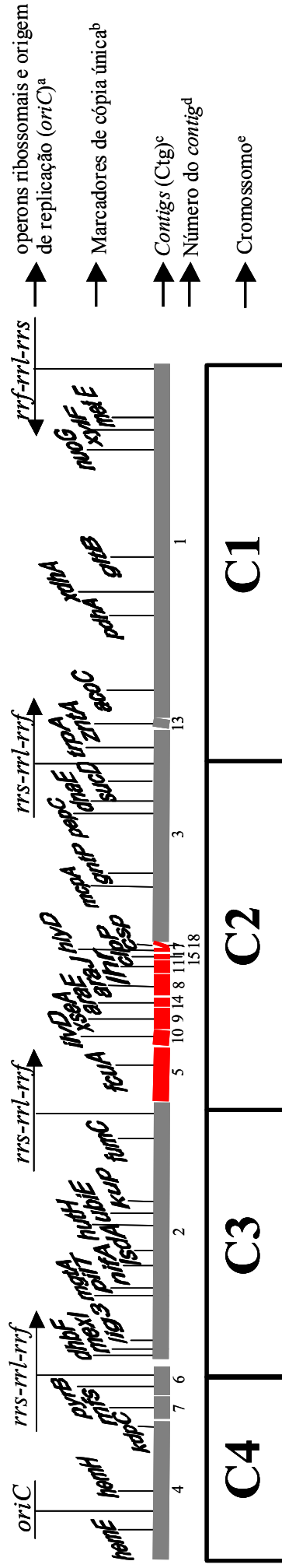
Para construção deste mapa físico, foi utilizado um reduzido número de experimentos de hibridização, especialmente em regiões onde foram alinhados grandes *contigs* gerados pelo Projeto Riogene, sobre o mapa físico do cromossomo de *G. diazotrophicus* PAL5 (Figura 16). Duas regiões onde foram alinhados no mapa físico, os fragmentos de DNA P5 e S6 (Figura 16), não apresentam marcadores genéticos. Entretanto, estes fragmentos foram observados em géis de PFGE (Figura 2B), e foi demonstrado na Figura 17, que a região cromossomal ocupada por estes fragmentos são cobertas pelos *contigs* 3 e 5, que apresentam tamanhos de 749,2 e 191,5 Kb, respectivamente.

A Figura 18 exhibe a sobreposição do alinhamento dos 16 *contigs*, gerados no Projeto Riogene em maio de 2006, em relação aos 6 *contigs* gerados na montagem de dezembro de 2006, onde confirmou-se o alinhamento dos *contigs* 1, 2, 3, 4, 6, 7 e 13 (montagem de Maio de 2006), que tiveram os buracos genômicos fechados através da metodologia de PCR combinatório. Adicionalmente, quando foram analisados os *contigs* 5 e 2 (montagem de dezembro de 2006), foi possível realinhar e reordenar precisamente os *contigs* 5, 8, 9, 10, 11, 14, 15, 17 e 18 (montagem de maio de 2006), no cromossomo do *G. diazotrophicus* PAL5.

4.6 – Presença de plasmídeos no genoma do *G. diazotrophicus* PAL5

Neste estudo foram alinhados no mapa físico do cromossomo do *G. diazotrophicus* PAL5, apenas 16 dos 18 *contigs* gerados no projeto de sequenciamento genômico deste microrganismo (Figura 17). Contudo, os dados de anotação destes outros dois *contigs*

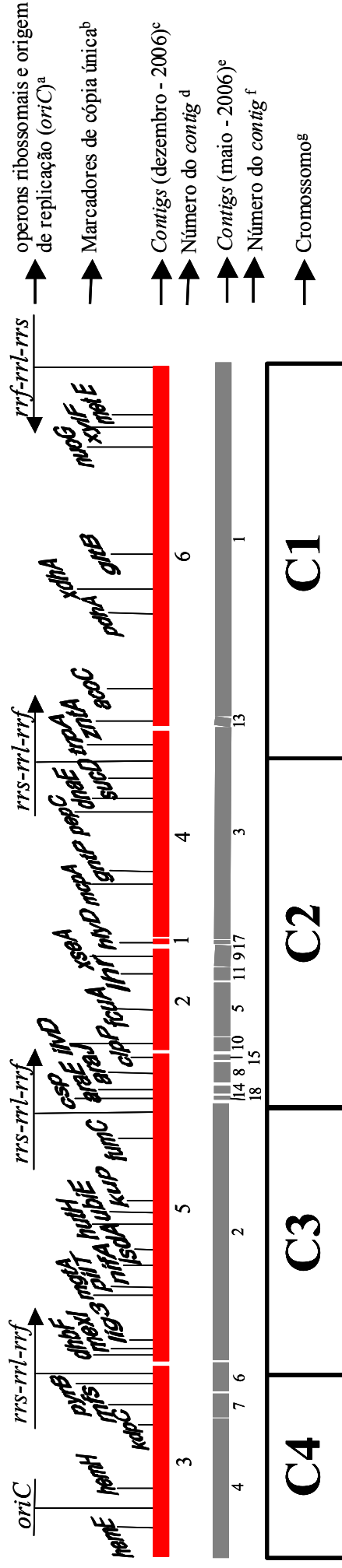
Figura 17 – Alinhamento de 16 contigs gerados no Projeto Riogene (maio de 2006), sobre o mapa físico do cromossomo do *Gluconacetobacter diazotrophicus* PAL5, com base nos resultados de hibridização*



MÉDIA DO TAMANHO DO CROMOSSOMO CIRCULAR 4232 Kb ± 47 Kb

As seqüências das sondas dos genes ribossomais e de cópia única, foram utilizadas no alinhamento de contigs do Projeto Riogene. * Os 16 contigs alinhados totalizam 97,6% de cobertura do cromossomo. Os contigs em vermelho não puderam ser ordenados e orientados precisamente através da metodologia empregada, porém todos estão alocados nesta mesma região do fragmento C2. O tamanho dos fragmentos de DNA gerados com a enzima *I-CeuI* e os contigs do Projeto Riogene, assim como a posição dos marcadores moleculares estão em escala. a – Localização e orientação dos operons ribossomais e localização da origem de replicação. b – Nomenclatura dos genes de cópia única utilizados para construção de sondas de DNA (vide tabela 2). c - Alinhamento dos contigs montados no Projeto Riogene. d – Os tamanhos dos contigs alinhados foram: contig 1 (1251,3 Kb), 2 (910,1 Kb), 3 (749,2 Kb), 4 (492,4 Kb), 5 (191,5 Kb), 6 (101,0 Kb), 7 (82,5 Kb), 8 (75,5 Kb), 9 (75,1 Kb), 10 (50,2 Kb), 11 (44,2 Kb), 13 (33,7 Kb), 14 (32,9 Kb), 15 (20,9 Kb), 17 (14,1 Kb), 18 (12,8 Kb). e – alinhamento dos fragmentos de DNA gerados com a enzima *I-CeuI* no mapa físico do cromossomo do *Gluconacetobacter diazotrophicus* PAL5.

Figura 18 – Sobreposição de alinhamentos dos contigs gerados no Projeto Riogene nas montagens de maio e dezembro de 2006, em relação ao mapa físico do cromossomo do *Gluconacetobacter diazotrophicus* PAL5.



MÉDIA DE TAMANHO DO CROMOSSOMO CIRCULAR

4232 Kb ± 47 Kb

As seqüências das sondas dos genes ribossomais e de cópia única foram utilizadas no alinhamento de contigs do Projeto Riogene. O tamanho dos fragmentos de DNA gerados com a enzima *I-CeuI* e os contigs do Projeto Riogene, assim como a posição dos marcadores moleculares estão em escala. a – Localização e orientação dos operons ribossomais e localização da origem de replicação. b – Nomenclatura dos genes de cópia única utilizados para construção de sondas de DNA (vide Tabela 2). c – Alinhamento dos contigs montados no Projeto Riogene (dezembro de 2006). d – Os tamanhos dos contigs alinhados foram: contig 1 (14,1 Kb), 2 (361,0 Kb), 3 (675,9), 4 (749,2 Kb), 5 (1052,2 Kb), 6 (1285,0 Kb); e – Alinhamento dos contigs montados no Projeto Riogene (maio de 2006). f – Os tamanhos dos contigs alinhados foram: contig 1 (1.251,3 Kb), 2 (910,1 Kb), 3 (749,2 Kb), 4 (492,4 Kb), 5 (191,5 Kb), 6 (101,0 Kb), 7 (82,5 Kb), 8 (75,5 Kb), 9 (75,1 Kb), 10 (50,2 Kb), 11 (44,2 Kb), 13 (33,7 Kb), 14 (32,9 Kb), 15 (20,9 Kb), 17 (14,1 Kb), 18 (12,8 Kb). g – Alinhamento dos fragmentos de DNA gerados com a enzima *I-CeuI* no mapa físico do cromossomo do *Gluconacetobacter diazotrophicus* PAL5.

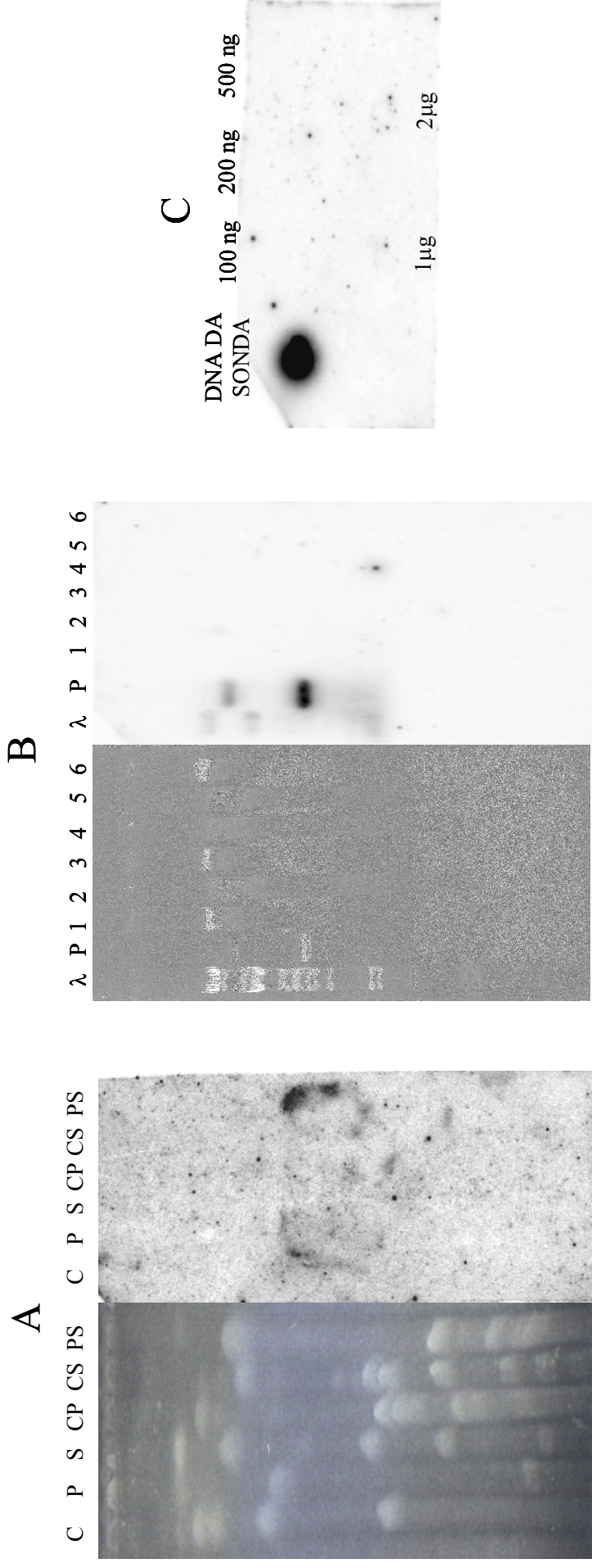
restantes (Tabelas 4 e 5), denominados *contigs* 12 e 16, sugerem que estes sejam plasmídeos. Convém ressaltar, que estes 2 plasmídeos não foram detectados através de purificação utilizando gradiente de cloreto de cério, método de lise alcalina (Sambrook e Russel, 2001), ou maxi-prep utilizando o kit comercial Quiagem Tip (Quiagem®) em células de *G. diazotrophicus* PAL5, sedimentadas a partir de 3 L de cultura a 1,3 DO_{600nm} (dados não exibidos).

Adicionalmente, quando sondas radioativas produzidas a partir de seqüências de cópia única destes plasmídeos, foram hibridizadas contra membranas produzidas a partir de géis de PFGE (Figura 19A; Figura 20A), assim como DNA genômico digerido com diferentes enzimas de restrição imobilizados em membrana de nylon (Figura 19B; Figura 20B), e *Dot blot* de DNA genômico em concentrações variando de 100 ng até 2 µg (Figura 19C; Figura 20C), não houve detecção de sinais de hibridização, sugerindo uma origem extra-cromossomal para estas seqüências. A possibilidade destas seqüências serem contaminantes foi considerada, porém esta hipótese foi afastada, pois os clones cujo suas seqüências geraram estes *contigs*, foram obtidos a partir de diferentes clonagens, pertencem a diferentes bibliotecas genômicas, e a montagem foi realizada portanto, com alto grau de confiabilidade (Dados não exibidos).

Os *contigs* 12 e 16 apresentam tamanhos de 38,8 e 16,6 Kb, respectivamente, e a circularidade destas seqüências foram verificadas a partir de *nested* PCRs (Sambrook & Russell, 2001), utilizando-se 4 pares de *primers* desenhados para anelarem em regiões de cópia única, próximas as extremidades destes *contigs*, os quais geraram fragmentos esperados de 2.8 (*contig* 12) e 1.6 Kb (*contig* 16), que posteriormente foram submetidos a reações de sequenciamento (Dados não exibidos).

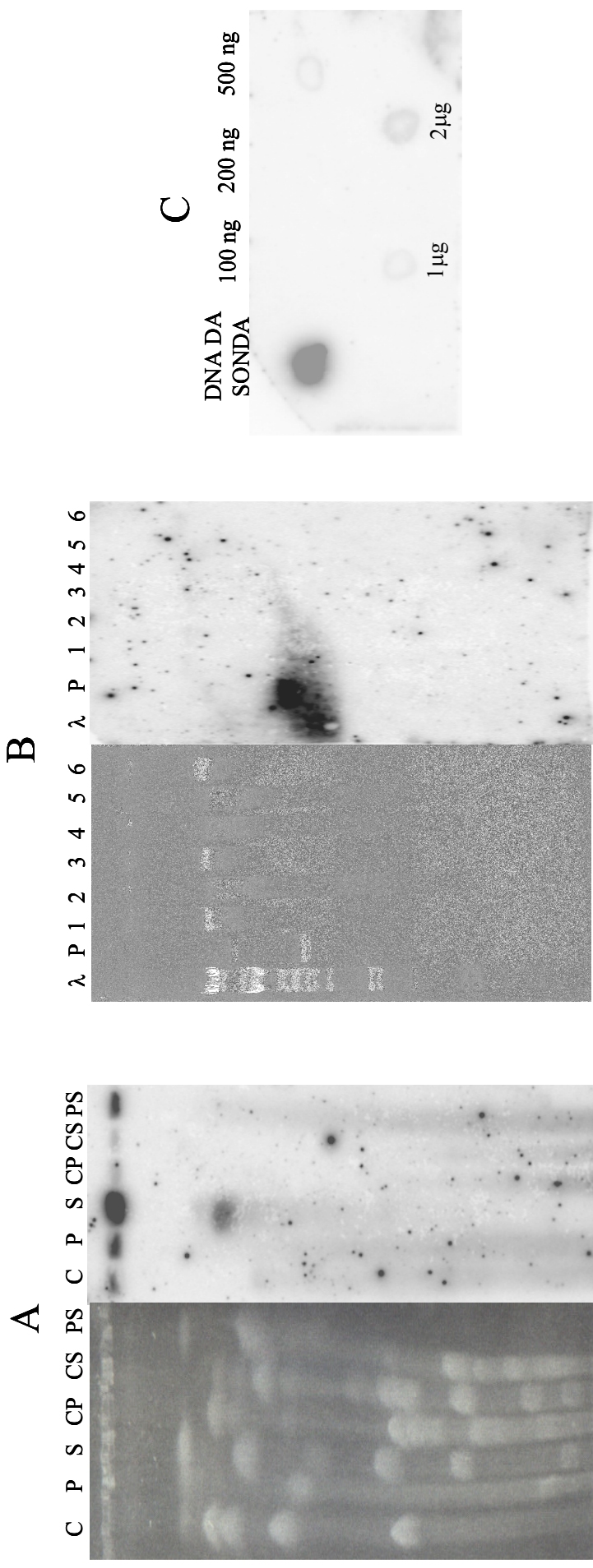
Analisando-se detalhadamente a anotação destes *contigs* (Tabelas 4 e 5), observa-se a presença genes tipicamente plasmidiais, destacando-se os genes que codificam a produção de

Figura 19 – Híbridizações obtidas com sonda de região de cópia única do *contig 12* (38,8 Kb), contra diferentes tipos de membrana contendo DNA genômico.



A – Resultados de hibridização contra membrana de PFGE; C, P, S, CP, CS e PS – Perfis de PFGE gerados a partir de digestões simples ou duplas do DNA cromossomal de *G. diazotrophicus* PAL5. **B** – Resultados de hibridização contra membrana contendo diferentes digestões simples de DNA genômico; 1 a 6 – DNA genômico de *G. diazotrophicus* PAL5 digerido com diferentes enzimas de restrição (1- *Bam*HI, 2- *Eco*RV, 3- *Hind*III, 4- *Pst*I, 5- *Sal*I, 6- *Xba*I); λ - Lambda DNA digerido com *Pst*I; P - plasmídeo pGEM. **C** – *Dot blot* de DNA genômico de *G. diazotrophicus* PAL5 espotados em diferentes concentrações (100 ng a 2 μ g).

Figura 20 – Híbridizações obtidas com sonda de região de cópia única do *contig 16* (16,6 Kb), contra diferentes tipos de membrana contendo DNA genômico.



A – Resultados de hibridização contra membrana de PFGE; C, P, S, CP, CS e PS – Perfis de PFGE gerados a partir de digestões simples ou duplas do DNA cromossomal de *G. diazotrophicus* PAL5. **B** – Resultados de hibridização contra membrana contendo diferentes digestões simples de DNA genômico; 1 a 6 – DNA genômico de *G. diazotrophicus* PAL5 digerido com diferentes enzimas de restrição (1- *Bam*HI, 2- *Eco*RV, 3- *Hind*III, 4- *Pst*I, 5- *Sal*I, 6- *Xba*I); λ - Lambda DNA digerido com *Pst*I; P - plasmídeo pGEM. **C** – *Dot blot* de DNA genômico de *G. diazotrophicus* PAL5 espotados em diferentes concentrações (100 ng a 2 μg).

Tabela 4 – Resultados de anotação do contig 12 (38,8 Kb) realizado no Sistema Bionotes, pelos membros integrantes do Projeto Riogene.

Número da ORF	CDS ^a	Função da seqüência similar ^b	Fonte da seqüência similar ^c	Probabilidade ^d
1	223..750	Proteína hipotética	-	-
2	1120..1965	Proteína hipotética conservada	<i>Burkholderia vietnamiensis</i> G4	2e-55
3	2327..2863	Proteína hipotética	-	-
4	3105..3542	Proteína hipotética	-	-
5	4488..4949	Proteína hipotética	-	-
6	5060..5356	Proteína hipotética	-	-
7	5356..5547	Proteína hipotética	-	-
8	5675..6229	Proteína hipotética	-	-
9	6244..6525	Proteína hipotética	-	-
10	6509..6757	Proteína hipotética	-	-
11	6981..7565	DNA recombinase	<i>Aeromonas punctata</i>	2e-28
12	7553..8464	Proteína hipotética conservada	<i>Mesorhizobium</i> sp. BNC1.	3e-52
13	8528..9355	Endonuclease	<i>Gluconobacter oxydans</i> 621H	5e-59
14	9422..9697	Proteína hipotética	-	-
15	10092..10766	Proteína TrbI de transferência conjugal	<i>Mesorhizobium loti</i>	2e-11
16	11096..11725	Proteína hipotética	-	-
17	11730..12722	Proteína hipotética	-	-
18	12727..13596	Proteína hipotética	-	-
19	13547..14557	Proteína hipotética	-	-
20	14557..15003	Proteína hipotética	-	-
21	15014..15823	Proteína hipotética	-	-
22	15823..18366	Proteína B4 do sistema de secreção tipo IV	<i>Erythrobacter litoralis</i> HTCC2594	3e-21
23	18379..18732	Proteína hipotética	-	-
24	18739..19077	Proteína hipotética	-	-
25	19061..19267	Proteína hipotética	-	-

Tabela 4 – Continuação.

Número da ORF	CDS	Função da seqüência similar	Fonte da seqüência similar	Probabilidade
26	19642..20040	Proteína hipotética	-	-
27	20183..21184	Proteína E do sistema de secreção tipo II	<i>Solibacter usitatus</i> Ellin6076	7e-36
28	21237..22037	Proteína hipotética conservada	<i>Burkholderia cenocepacia</i> HI2424	1e-21
29	22146..22433	Proteína hipotética	-	-
30	22728..23108	Proteína hipotética	-	-
31	23201..23746	Proteína hipotética	-	-
32	24033..24959	Proteína hipotética	-	-
33	25331..26737	Proteína de mobilização	<i>Erwinia carotovora</i> subsp. <i>atroseptica</i>	1e-35
34	27181..27531	Proteína hipotética	-	-
35	27534..28175	Proteína de partição cromossomal	<i>Nitrosospira multififormis</i> ATCC25196	1e-14
36	28169..28369	Proteína hipotética	-	-
37	28393..28545	Proteína hipotética	-	-
38	28735..29637	Proteína TraL de transferência conjugal	<i>Nitrosomonas eutropha</i> C71	5e-36
39	29726..30034	Proteína hipotética	-	-
40	30067..31386	Iniciador de replicação repC	<i>Gluconobacter oxydans</i> 621H	2e-16
41	31485..31925	Proteína hipotética	-	-
42	32135..32533	Proteína BfpH de biogênese de pili tipo IV	<i>Gluconobacter oxydans</i> 621H	1e-24
43	32533..33003	Proteína hipotética	-	-
44	33086..33499	Proteína hipotética	-	-
45	33529..33957	Proteína hipotética	-	-
46	33989..34540	Proteína hipotética	-	-
47	34618..34815	Proteína hipotética I	-	-
48	35005..35439	Proteína hipotética	-	-
49	35524..35868	Proteína SSB ligadora de DNA de fita simples	<i>Shewanella</i> sp. ANA-3.	1e-11
50	36247..36687	Proteína plasmidial	<i>Pseudomonas putida</i>	2e-11
51	36702..36878	Proteína hipotética	-	-

Tabela 4 – Continuação.

Número da ORF	CDS	Função da seqüência similar	Fonte da seqüência similar	Probabilidade
52	37002..37484	Proteína hipotética conservada	<i>Mesorhizobium sp.</i> BNC1	2e-07
53	37477..37827	Proteína hipotética	-	-
54	37891..38445	Proteína hipotética	-	-

ORF – fase de leitura aberta (*open reading frames*); CDS - seqüência codificante (*coding sequence*). a- Posição da seqüência codificante; b – Funções preditas para as ORFs, com base nos resultados de análises de Blastp no banco de dados do NCBI, assim como no UniProt e PFAM; c- microrganismos que apresentam melhores alinhamentos em relação a esta seqüência no banco de dados NCBI, utilizando-se a ferramenta Blastp; d – *e-value* observado no alinhamento da seqüência da sonda contra o banco de dados NCBI utilizando a ferramenta Blastp.

Tabela 5 – Resultados de anotação do *contig 16* (16,6 Kb) realizado no Sistema Bionotes, pelos membros integrantes do Projeto Riogene.

Número da ORF	CDS ^a	Função da seqüência similar ^b	Fonte da seqüência similar ^c	Probabilidade ^d
1	1..1589	Componente VirD2 do sistema de secreção tipo IV	<i>Rhodospirillum rubrum</i>	e-152
2	1710..1923	Proteína hipotética	-	-
3	2230..3291	Transposase e derivados inativados	<i>Mesorhizobium loti</i> MAFF303099	e-138
4	2751...3510	Transposase	<i>Mesorhizobium loti</i> MAFF303099	1e-76
5	2877..3762	Proteína hipotética	-	-
6	3763..4462	Transglicosilase lítica de mureína solúvel	<i>Bradyrhizobium japonicum</i> USDA 110	8e-43
7	4537..5077	Proteína de transferência conjugal TraF	<i>Bradyrhizobium japonicum</i> USDA 110	5e-50
8	5076..5571	Proteína hipotética conservada	<i>Mesorhizobium sp.</i> BNC1	4e-54
9	5570..5810	Proteína hipotética conservada	<i>Bradyrhizobium japonicum</i> USDA 110	6e-19
10	5809..5998	Metilase de modificação de DNA	<i>Mesorhizobium sp.</i> BNC1	5e-13
11	5997..6657	ATPases envolvidas em partição cromossomal	<i>Mesorhizobium sp.</i> BNC1	8e-76
12	6632..7685	Proteína A de replicação	<i>Mesorhizobium loti</i> MAFF303099	e-102
13	7687..8017	Proteína hipotética conservada	<i>Mesorhizobium loti</i> MAFF303099	7e-21
14	8197..9547	Transportador NhaA de Na ⁺ /H ⁺	<i>Gluconobacter oxydans</i> 621H	9e-89
15	9567..10263	Proteína hipotética conservada	<i>Mesorhizobium sp.</i> BNC1	2e-13
16	10238..10814	Proteína hipotética conservada	<i>Caulobacter crescentus</i> CB15	4e-04
17	10756..11038	Proteína hipotética conservada	<i>Mesorhizobium sp.</i> BNC1	2e-20
18	11258..11663	Proteína hipotética conservada	<i>Bradyrhizobium japonicum</i> USDA 110	5e-41
19	12134..12740	Proteína hipotética conservada	<i>Bradyrhizobium japonicum</i> USDA 110	4e-17
20	13768..14554	Glutationina S-transferase	<i>Pseudomonas aeruginosa</i> PAO1	3e-52
21	14622..15690	ATPase de segregação cromossômica	<i>Mesorhizobium sp.</i> BNC1	e-106
22	15606..16611	Transposase	<i>Gluconobacter oxydans</i> 621H	e-123

ORF – fase de leitura aberta (*open reading frames*); CDS – Seqüência codificante (*coding sequence*). a – Posição da seqüência codificante; b – Funções preditas para as ORFs, com base nos resultados de análises de Blastp no banco de dados do NCBI, assim como no UniProt e PFAM; c – Microorganismos que apresentam melhores alinhamentos em relação a esta seqüência no banco de dados NCBI, utilizando-se a ferramenta Blastp; d – *e-value* observado no alinhamento da seqüência da sonda contra o banco de dados NCBI utilizando a ferramenta Blastp.

proteínas pertencentes ao sistema de transferência horizontal de genes por conjugação (Proteína TrbI, Proteína TraF; proteína TraL; proteína BfpH; e proteínas de via de secreção Tipo IV: componentes VirD2 e B4), e alguns genes que não necessariamente apresentam origem cromossomal, tais como os genes que codificam a produção de proteínas de mobilização, proteína A de particionamento, ATPase de segregação cromossômica, proteínas envolvidas em replicação de DNA, Fator C de replicação, transposase e resolvase.

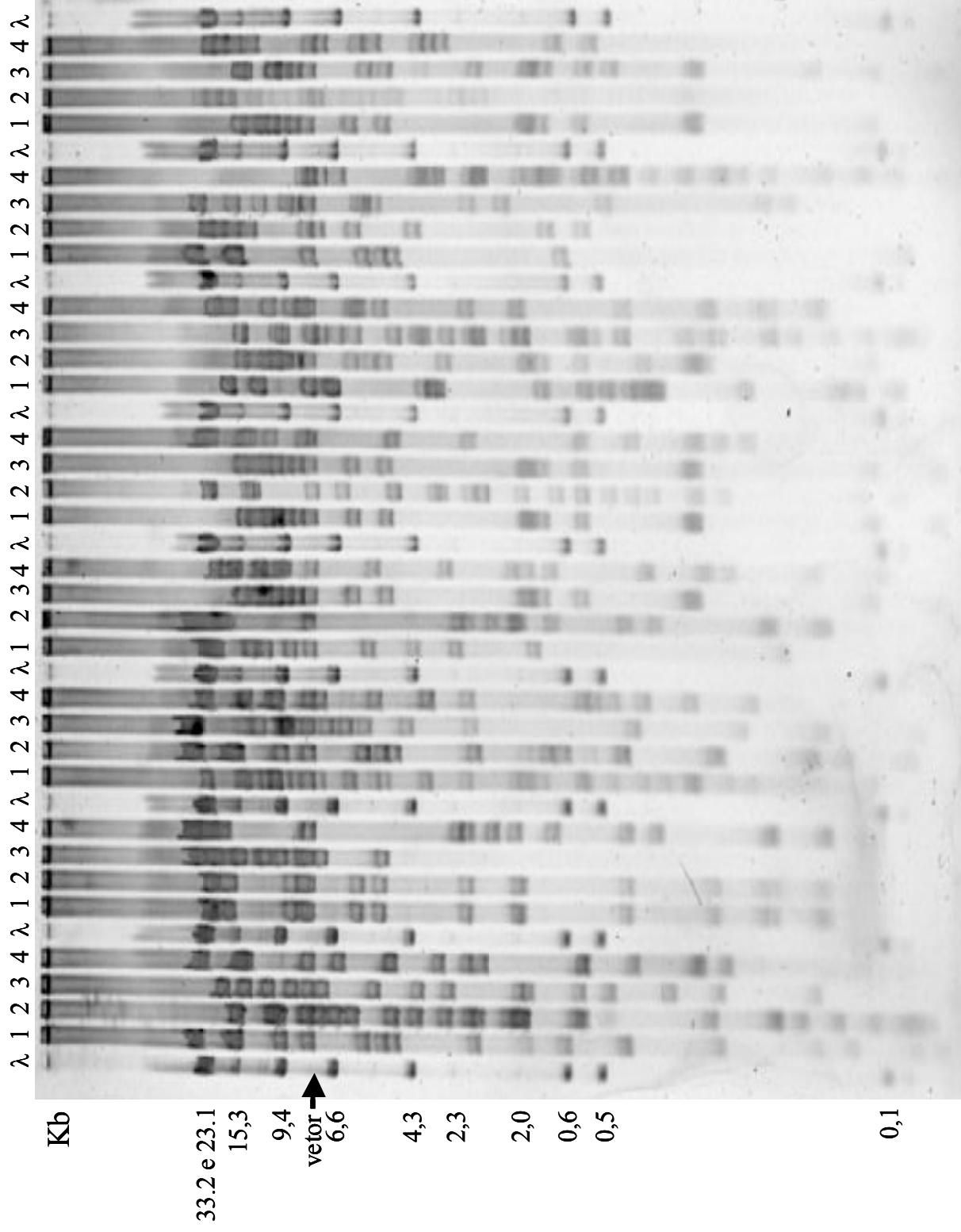
4.7 – Mapeamento físico do genoma do *G. diazotrophicus* PAL5 utilizando Cromossomos Artificiais Bacterianos (BACs)

Neste estudo foram analisados 1315 perfis de restrição de BACs, submetidos a digestões duplas com as enzimas *HindIII* e *BamHI* (Figura 21), o que equivale a 37,3 coberturas do genoma do *G. diazotrophicus* PAL5, considerando o tamanho médio dos BACs (120 Kb), da biblioteca genômica construída na *Clemson University Genomic Institute* – Carolina do Norte – EUA, e o tamanho do genoma estimado através do mapeamento físico por fragmentos de macrorestrição (4232 Kb; Figura 16).

Os 1315 perfis de restrição de BACs foram processados no *software image 3.0i*, para posterior utilização de dados na montagem de mapas físicos, utilizando o *software* FPC (Figura 22). No *software* FPC, através da utilização de uma condição mais estridente de análise (*cutoff* 1e-12) foram gerados 21 *contigs*, onde o maior *contig* apresenta em sua estrutura 916 BACs e aproximadamente 1600 Kb de tamanho, por outro lado, numa montagem menos estridente (*cutoff* 1e-04), foi montado apenas 1 *contig* que apresenta os 1315 BACs analisados em sua estrutura (dados não exibidos).

A partir das montagens supracitadas, foram selecionados 117 clones cujo seus BACs realizavam 2 coberturas do maior *contig* da montagem mais estridente, e 2 coberturas do genoma na montagem menos estridente (Figura 23). Estes 117 clones foram submetidos a

Figura 21 – Gel de agarose ilustrativo dos 1315 perfis de restrição dos BACs, submetidos a digestões duplas com as enzimas de restrição *Hind*III and *Bam*H1.



1, 2, 3 e 4 – Perfis de digestão dos Cromossomos Artificiais Bacterianos (BACs); λ - *Hilo marker*.

Figura 22 – Ilustração de 1 segmento de 1 mapa físico de BACs gerado no software FPC, após processamento de géis no *software Image 3.0i*.

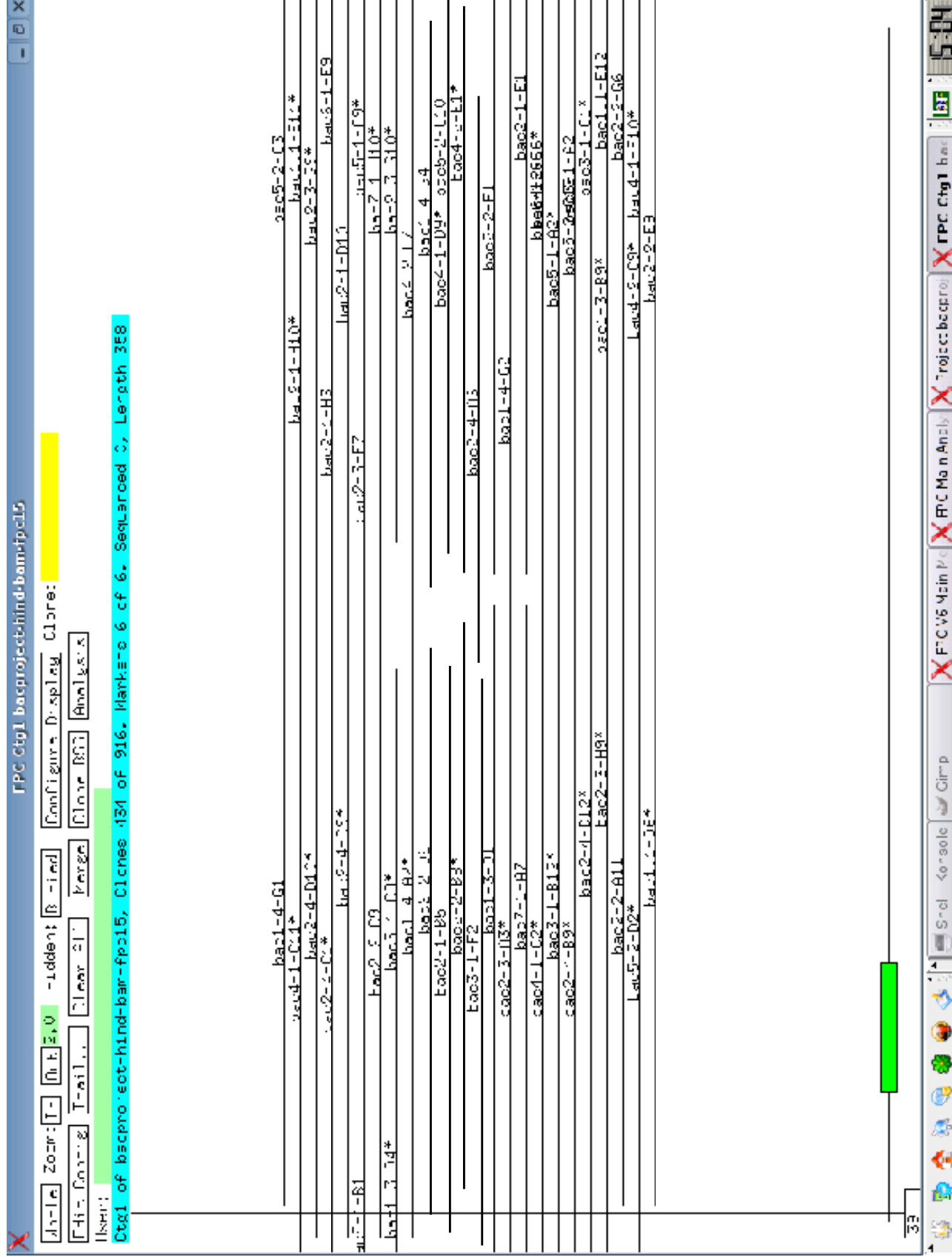
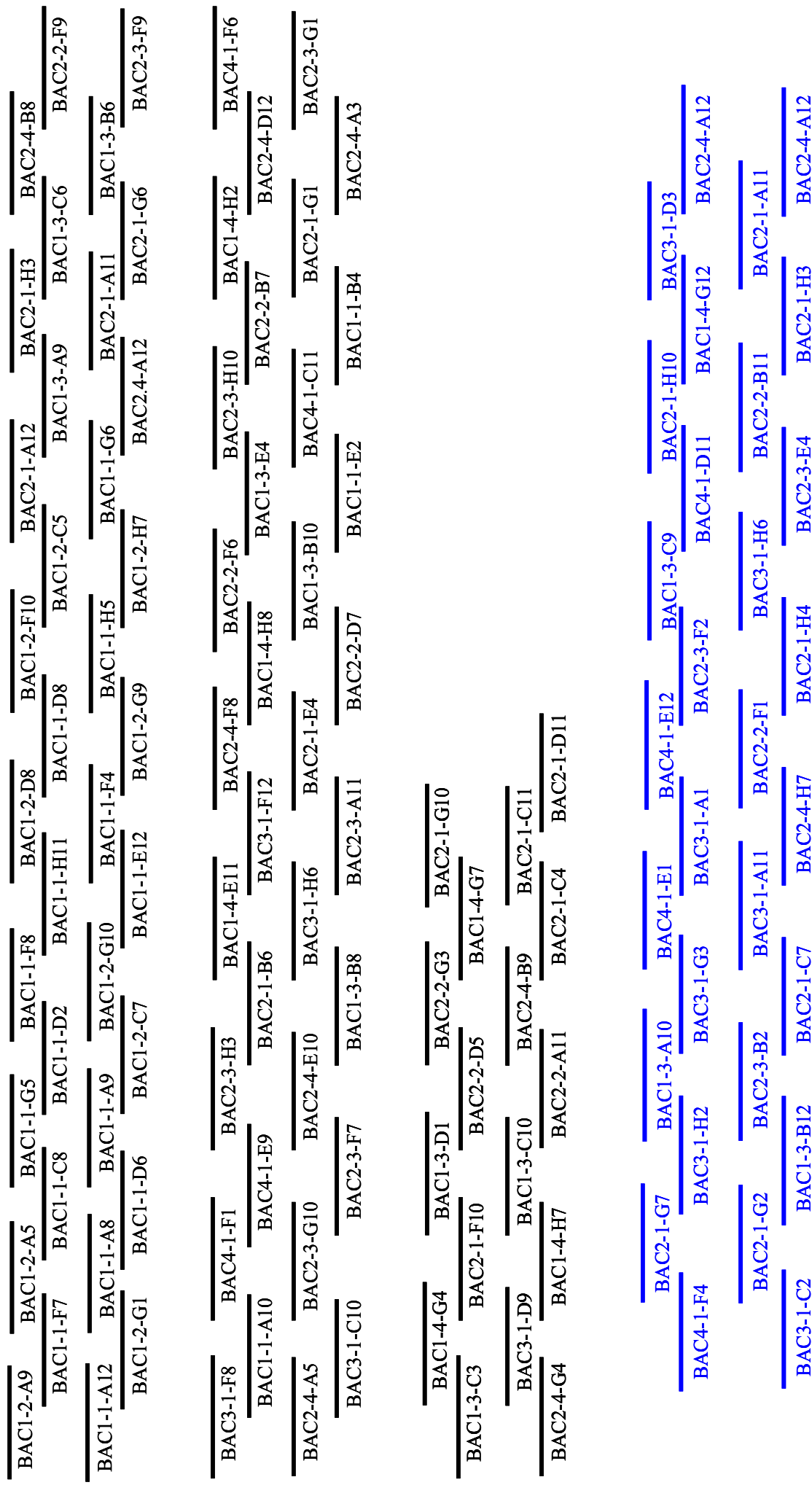


Figura 23 – Representação esquemática dos 117 BACs selecionados a partir dos mapas físicos gerados no software FPC, exibindo sua respectivas coberturas ao longo do genoma do *G. diazotrophicus* PAL5.



Em preto estão representados 87 Cromossomos Artificiais Bacterianos (BACs) selecionados a partir do mapa físico gerado no software FPC com uma stringência menor (*Cutoff* 1e-04), os quais realizam 2 coberturas de todo o genoma. Em azul estão representados 30 BACs selecionados a partir do mapa físico gerado no software FPC com uma stringência maior (*Cutoff* 1e-12), os quais realizam 2 coberturas do maior *contig* gerado neste mapa.

hibridizações com sondas radioativas dos genes de RNAr 16s e 23s, com finalidade de determinar o posicionamento do grupo de genes ribossomais nestes mapas físicos, onde verificou-se sinais de hibridização para 7 BACs distintos (Figura 24), os quais estão alocados em 4 pontos diferentes do mapa físico de BACs.

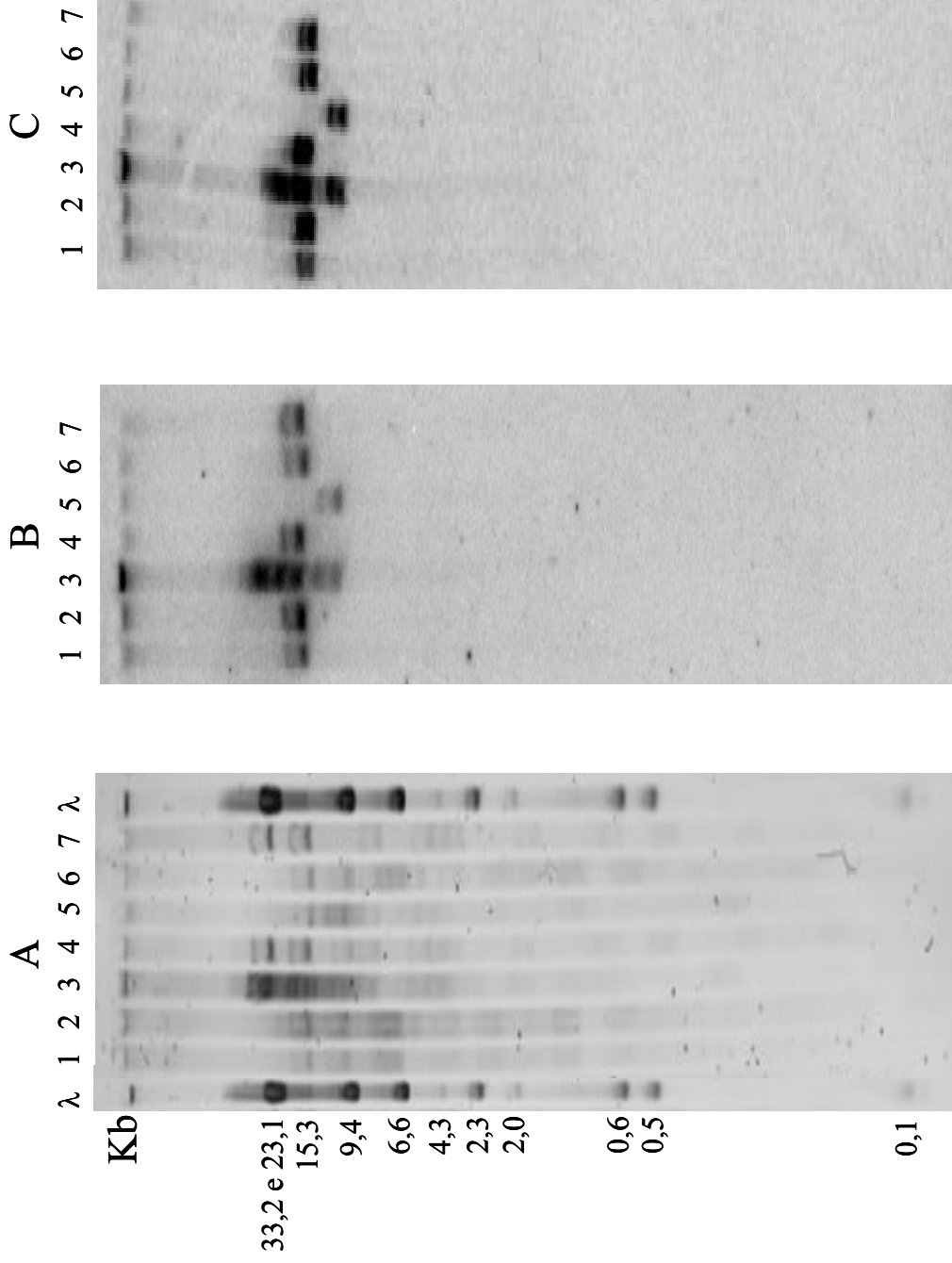
Os 117 clones selecionados (Figura 23) foram submetidos à reação de sequenciamento na *Clemson University Genomic Institute* – Carolina do Norte – EUA, devido ao fato de que o protocolo utilizado para tal finalidade, não foi padronizado às condições de trabalho dos laboratórios integrantes do Projeto Riogene, após inúmeras tentativas de padronização sem sucesso. Por outro lado, das 668 reações de sequenciamento realizadas (2 extremidades dos insertos de DNA, contidos nos 117 BACs, em duplicata), apenas 43 reações resultaram em seqüências de boa qualidade (6,4% de aproveitamento).

Adicionalmente, quando estas 43 seqüências de boa qualidade foram alinhadas (Blastn) contra o banco de dados de *contigs* do Projeto Riogene (<http://www.riogene.lncc.br>), 22 seqüências alinharam com seqüências repetitivas no genoma deste endófito, ou seja, apenas 21 seqüências puderam ser aproveitadas para o alinhamento de *contigs* neste genoma (dados não exibidos).

Das 21 seqüências de DNA aproveitadas do sequenciamento dos BACs, foi possível alinhar 14 *contigs* (Figura 25) no mapa físico de BACs do *G. diazotrophicus* PAL5, entretanto cabe ressaltar, que no período em que foi realizado este mapeamento (novembro de 2004), existiam 561 *contigs* depositados no banco de dados do Projeto Riogene, e a maior parte deles apresentavam tamanhos inferiores a 30 Kb.

Por outro lado, 14 seqüências oriundas de 7 BACs foram extremamente importantes para o alinhamento de 7 *contigs* da montagem de novembro de 2004 (Figura 26), os quais totalizam 1.176,9 Kb a nível de seqüências genômicas do *G. diazotrophicus* PAL5, o que representa 27,8% de cobertura do genoma, quando comparado ao tamanho genômico

Figura 24 - Perfis de Restrição de 7 BACs que hibridaram contra sondas radioativas dos genes de RNAr 16s e 23s do *G. diazotrophicus* PAL5*



* Estes resultados de hibridização foram detectados entre os 117 clones, que foram selecionados para seqüenciamento de DNA na *Clemson University Genomic Institute* - Carolina do Norte - EUA. λ - Hilo marker. A - Perfis de restrição dos BACs que hibridaram com sondas dos genes do RNAr 16S e 23S. B - hibridizações obtidas com sonda do gene do RNAr 16s; C - hibridizações obtidas com sonda do gene do RNAr 23s.

Figura 25 – Ilustração de um segmento do mapa físico de BACs, exibindo o alinhamento de contigs da montagem de novembro de 2004, baseado nas sequências de DNA geradas na *Clemson University Genomic Institute*.

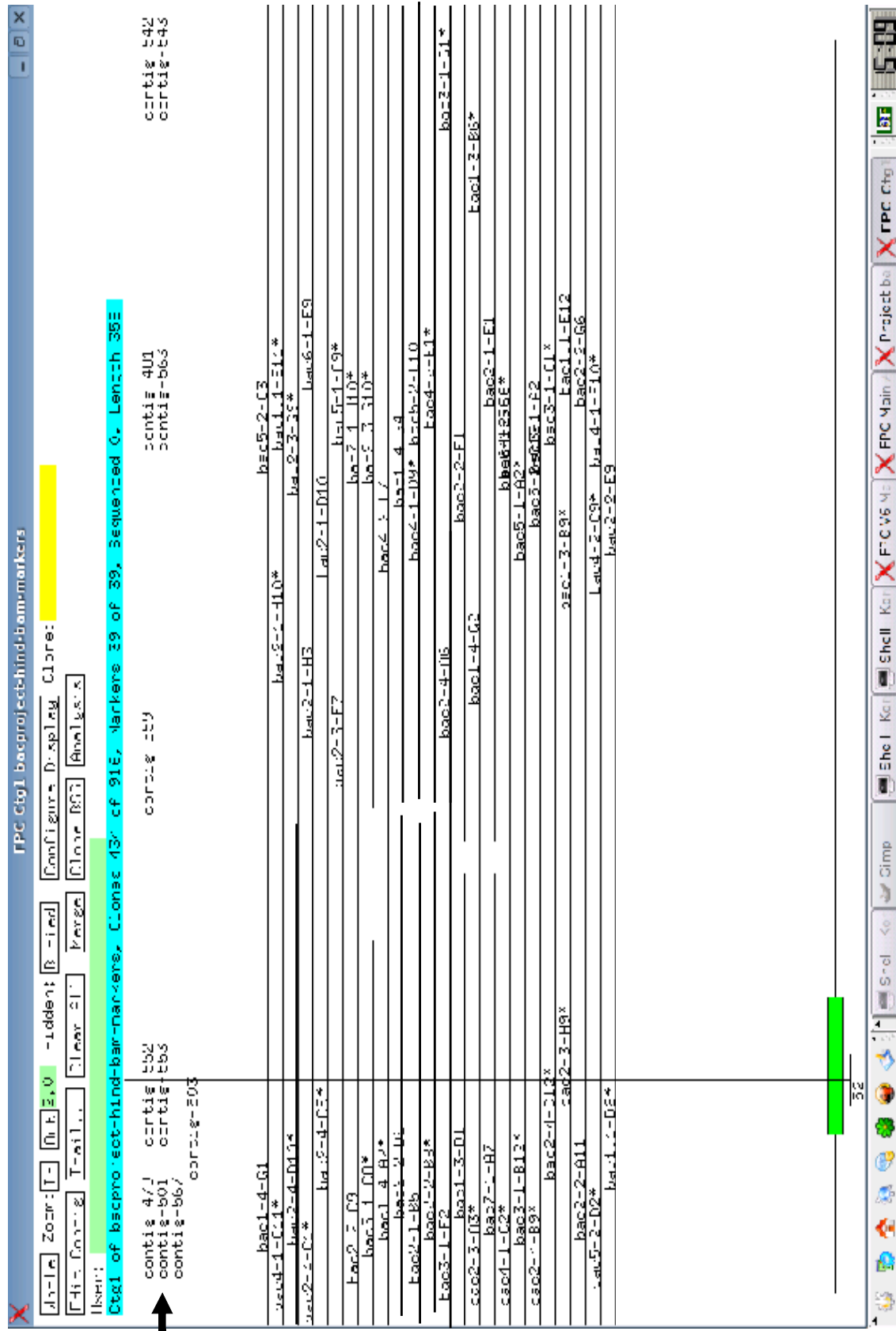
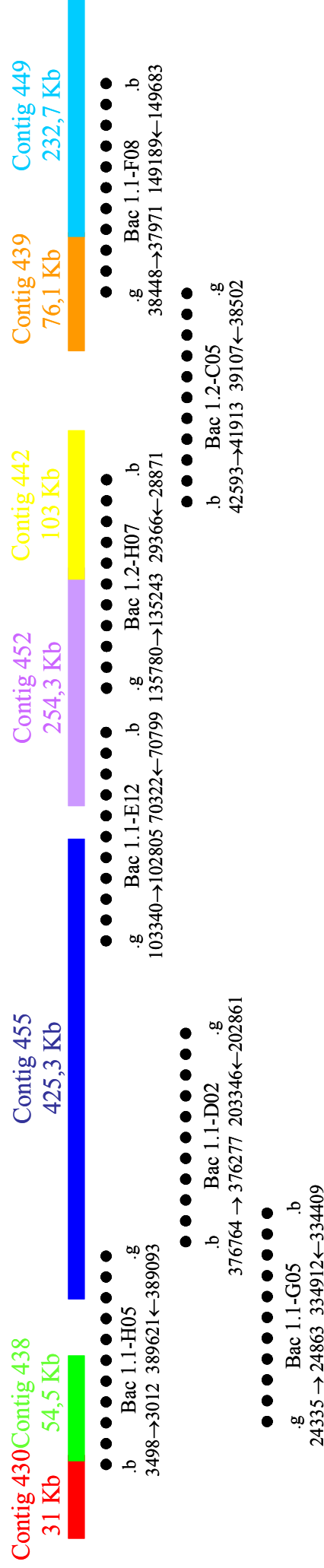


Figura 26 – Representação esquemática do alinhamento de 7 contigs (Blastn) da montagem de Novembro de 2004, baseado em 14 seqüências de DNA oriundas de 7 BACs, geradas na *Clemson University Genomic Institute.**



* Estes dados foram utilizados para montagem do maior contig deste genoma (denominado como contig 1 na montagem de Maio de 2006, e contig 6 na montagem de dezembro de 2006). Em cores estão representados os contigs ordenados e alinhados, assim como seus respectivos tamanhos. Em linhas pontilhadas pretas estão representados os BACs e seus respectivos alinhamentos contra os contigs do Projeto Riogene. Nesta representação esquemática de alinhamento foram ordenados e orientados 7 contigs, os quais totalizam 1.176,9 Kb a nível de seqüências genômicas do *G. diazotrophicus* PAL5, o que representa 27,8% de cobertura do genoma, quando comparado ao tamanho genômico estimado pelo mapa físico de macrorestrição (PFGE). O tamanho dos BACs representados nesta figura não estão em escala.

estimado pelo mapa físico de macrorestrição (4232 Kb; Figura 16). Adicionalmente, os dados de alinhamento destes 7 BACs, serviram de subsídio para a montagem do maior *contig* deste genoma, denominado como *contig* 1 na montagem de Maio de 2006, e *contig* 6 na montagem de dezembro de 2006 (Figura 18).

4.8 – Características gerais do genoma do *G. diazotrophicus* PAL5

A Tabela 6 exibe as características gerais do genoma do *G. diazotrophicus* PAL5, que consiste na presença de um cromossomo circular com tamanho de 4137,5 Kb, além de 2 plasmídeos circulares designados como A e B, com tamanhos de 38,8 e 16,6 Kb, respectivamente. Este genoma apresenta 3657 genes, sendo que 3581 localizados no cromossomo, 54 no plasmídeo A e 22 no plasmídeo B. A média de tamanho gênico é de 978 bases, o que confere produção de proteínas com tamanho médio de 326 aminoácidos, além de uma densidade gênica de 1,1 gene por Kb e um percentual total de 90,5% de seqüências codificantes de proteínas e RNAs. Entre outras características, destaca-se a presença de 4 operons ribossomais e 71 tRNAs, assim como um alto percentual de GC neste genoma (66,4%), e uma distribuição equitativa dos genes ao longo da fita *sense* e *anti-sense* (53,3% e 46,7%, respectivamente).

A Tabela 7 e a Figura 27 exibem a distribuição e a classificação funcional, dos 3657 genes detectados no genoma do *G. diazotrophicus* PAL5, onde se destaca um maior percentual de genes relacionados com a classe funcional metabolismo central e intermediário (18,7%; 684 genes), que apresenta genes relacionados com biossíntese de aminoácidos, lipídios, nucleotídeos e co-fatores, metabolismo de polissacarídeos e oligossacarídeos, fixação biológica de nitrogênio, degradação de proteínas, entre outros (Tabela 8). A segunda maior população de genes se enquadram na categoria funcional superfície celular (9,2%; 337 genes),

Tabela 6 – Características gerais do genoma do *Gluconacetobacter diazotrophicus* PAL5.

Característica	Total
Tamanho do Cromossomo	4137,5 Kb
Tamanho do Plasmídio circular A	38,8 Kb
Tamanho do Plasmídio circular B	16,6 Kb
Proporção total de GC no genoma	66,4%
Número e percentual de bases codificadoras de proteínas e RNAs	3794574 (90,5%)
Número de genes	3657
Proporção total de GC nos genes	67,1%
Média de tamanho gênico	978 bases
Média de aminoácidos por proteína codificada	326
Densidade gênica	1,1 gene por Kb
Número e percentual de genes codificados na fita sense	1951 (53,3%)
Número e percentual de genes codificados na fita antisense	1706 (46,7%)
Número de tRNAs	71
Número de operons ribossomais	4

Tabela 7 – Classificação funcional dos genes detectados no genoma do *Gluconacetobacter diazotrophicus* PAL5

Classe funcional	Número de ORFs	Percentual de ORFs
Metabolismo central e intermediário	684	18,7%
Superfície celular (membranas internas e externas, proteínas secretadas, estruturas de superfície, etc)	337	9,2%
Transferência de Informações (transcrição, tradução, modificação de DNA e RNA)	227	6,2%
Metabolismo energético (glicólise, transporte de elétrons, etc.)	164	4,5%
Patogenicidade/Adaptação/Chaperoninas	143	3,9%
Reguladores	120	3,3%
Elementos oriundos de Bacteriófagos e seqüências de inserção (ISs)	63	1,7%
Degradação de grandes moléculas	60	1,6%
Degradação de pequenas moléculas	13	0,4%
Pseudogenes e genes parciais	4	0,1%
Proteínas hipotéticas conservadas	44	1,2%
Proteínas hipotéticas	1802	49,2%
Total	3657	100%

ORF – Fase de leitura aberta (*Open Reading Frames*)

Figura 27 – Distribuição dos genes detectados no genoma do *Gluconacetobacter diazotrophicus* PAL5 de acordo com a sua classificação funcional.

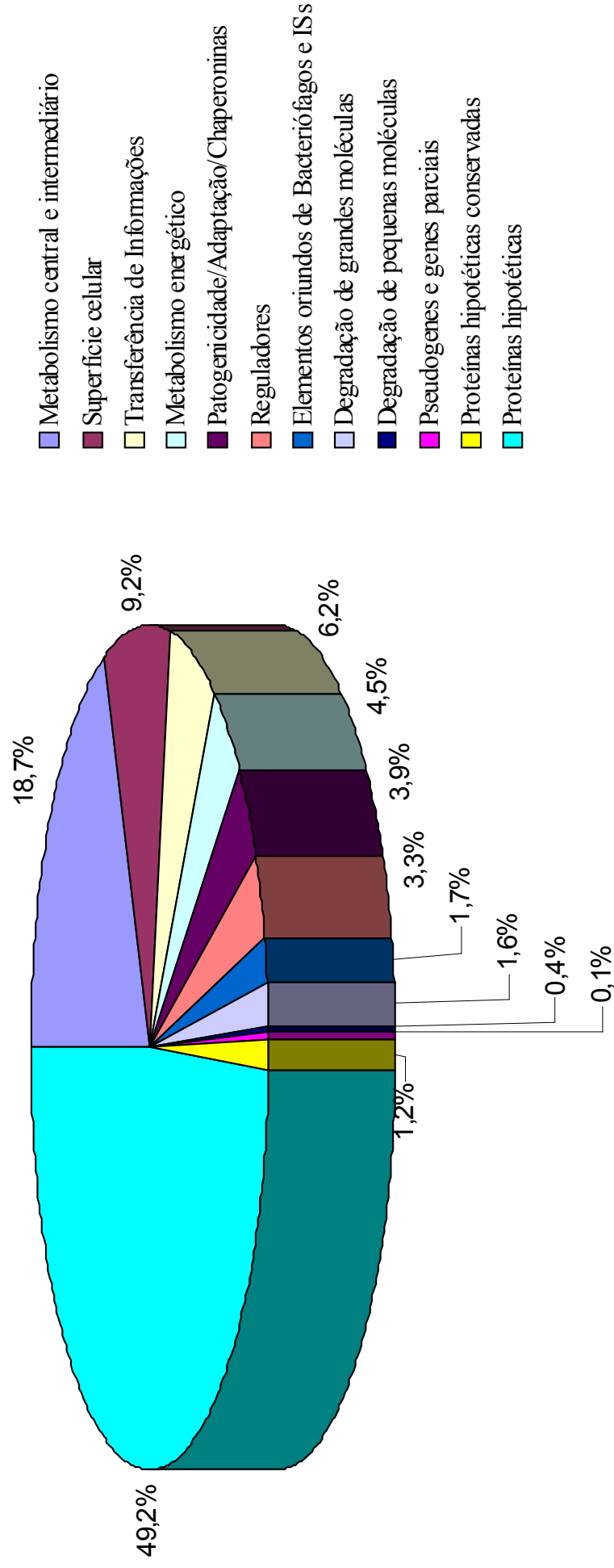


Tabela 8 - Distribuição dos genes anotados pelos membros integrantes do Projeto Riogene no sistema Bionotes, subdivido nas categorias e subcategorias funcionais mais freqüentes.

Categorias funcionais	Subcategorias funcionais	N	% categoria (% genoma)
Metabolismo central e intermediário	Biossíntese de aminoácidos	127	18,5%
	Biossíntese de lipídios	75	10,9%
	Metabolismo de polissacarídeos e oligossacarídeos	50	7,3%
	Biossíntese de nucleotídeos	45	6,6%
	Fixação biológica de nitrogênio	20	2,9%
	Biossíntese de cofatores	17	2,5%
	Tradução e modificação de proteínas	11	1,6%
	Degradação de proteínas	10	1,5%
	Replicação e reparo de DNA	10	1,5%
	Divisão celular	8	1,2%
	Outros	311	45,5%
	Total da categoria funcional	684	100% (18,7%)
Superfície celular	Proteínas ligadoras e transportadoras	202	59,9
	Proteínas de membrana externa, interna e parede celular	125	37,1
	Outros	10	3,0%
	Total da categoria funcional	337	100% (9,2%)
Transferência de Informações	Replicação e reparo de DNA	81	35,7%
	Transcrição e tradução	76	33,5%
	Proteínas ribossomais	51	22,5%
	Modificação de DNA e RNA	8	3,5%
	Outros	11	4,8%
	Total da categoria funcional	227	100% (6,2%)

Tabela 8 – Continuação.

Categorias funcionais	Subcategorias funcionais	N	% categoria (% genoma)
Metabolismo energético	Cadeia transportadora de elétrons	92	56,1
	Ramo oxidativo da via das pentose-fosfato	28	17,1
	Piruvato desidrogenase	12	7,3
	Glicólise	10	6,1
	Ciclo dos ácidos tricarboxílicos	9	5,5
	Outros	13	7,9
	Total da categoria funcional	164	100% (4,5%)
Patogenicidade, adaptação e chaperoninas	Resistência a drogas	32	22,4%
	Motilidade celular	25	17,5%
	Adaptação a condições ambientais atípicas	21	14,7%
	Quimiotaxia	20	14,0%
	Destoxificação celular	12	8,4%
	Divisão celular	11	7,7%
	Adaptação osmótica	9	6,3%
	Chaperoninas	8	5,6
	Outros	5	3,4%
	Total da categoria funcional	143	100% (3,9%)
Reguladores	lysR	20	16,7%
	Proteínas quinases	16	13,2%
	Sistemas de duplo componente	11	9,2%
	Assimilação de nitrogênio	8	6,7%
	tetR	7	5,8%

Tabela 8 – Continuação.

Categorias funcionais	Subcategorias funcionais	N	% categoria (% genoma)
	gntR	5	4,2%
	asnC	3	2,5%
	araC	2	1,7%
	arsR	2	1,7%
	lacI	2	1,7%
	marR	2	1,7%
	luxR	2	1,7%
	merR	1	0,8%
	PadR	1	0,8%
	Outros	38	31,6%
	Total da categoria funcional	120	100% (3,3%)
Elementos oriundos de Bacteriófagos e ISs	Elementos com funções relacionadas a transposons	50	79,3%
	Elementos lateralmente adquiridos	9	14,3%
	Elementos com funções relacionadas a bacteriófagos	3	4,8%
	Elementos com funções relacionadas a plasmídios	1	1,6%
	Total da categoria funcional	63	100% (1,7%)

onde destacam-se os genes relacionados com produção de proteínas ligadoras e transportadoras, e proteínas de membrana externa, interna e parede celular (Tabela 8).

Com relação a outras categorias funcionais descritas na Tabela 7 e Figura 27, destaca-se a categoria transferência de informações, que apresenta 227 genes (6,2% do total do genoma) relacionados principalmente com replicação e reparo de DNA, transcrição e tradução gênica, e proteínas ribossomais, seguido pelas categorias funcionais metabolismo energético (4,5%; 164 genes), que apresenta genes relacionados com cadeia transportadora de elétrons, ramo oxidativo da via das pentoses, complexo piruvato desidrogenase, glicólise e ciclo dos ácidos tricarboxílicos (Tabela 8); e a categoria funcional patogenicidade/adaptação/chaperoninas (3,9%; 143 genes), que apresenta genes relacionados com quimiotaxia, mobilidade celular, resistência a drogas, adaptações a condições ambientais atípicas, destoxificação celular e adaptação osmótica, ente outros (Tabela 8).

Outras duas categorias funcionais que merecem destaque (Tabela 7; Figura 27) consistem na categoria reguladores (3,3%; 120 genes), que engloba reguladores inclusos em diversas sub-categorias, tais como: lysR, proteínas quinases, sistemas duplo componentes, reguladores de assimilação de nitrogênio, tetR e gntR; e na categoria elementos oriundos de bacteriófagos e seqüências de inserção (ISs) (1,7%; 63 genes), onde predominam genes de elementos com funções relacionadas a transposons (Tabela 8). E ainda com relação à classificação funcional, podemos observar um grande número de genes (49,2%; 1802 genes), que codificam a síntese de proteínas hipotéticas, pois suas seqüências não apresentam similaridade, com seqüências de outros microrganismos depositadas em bancos de dados.

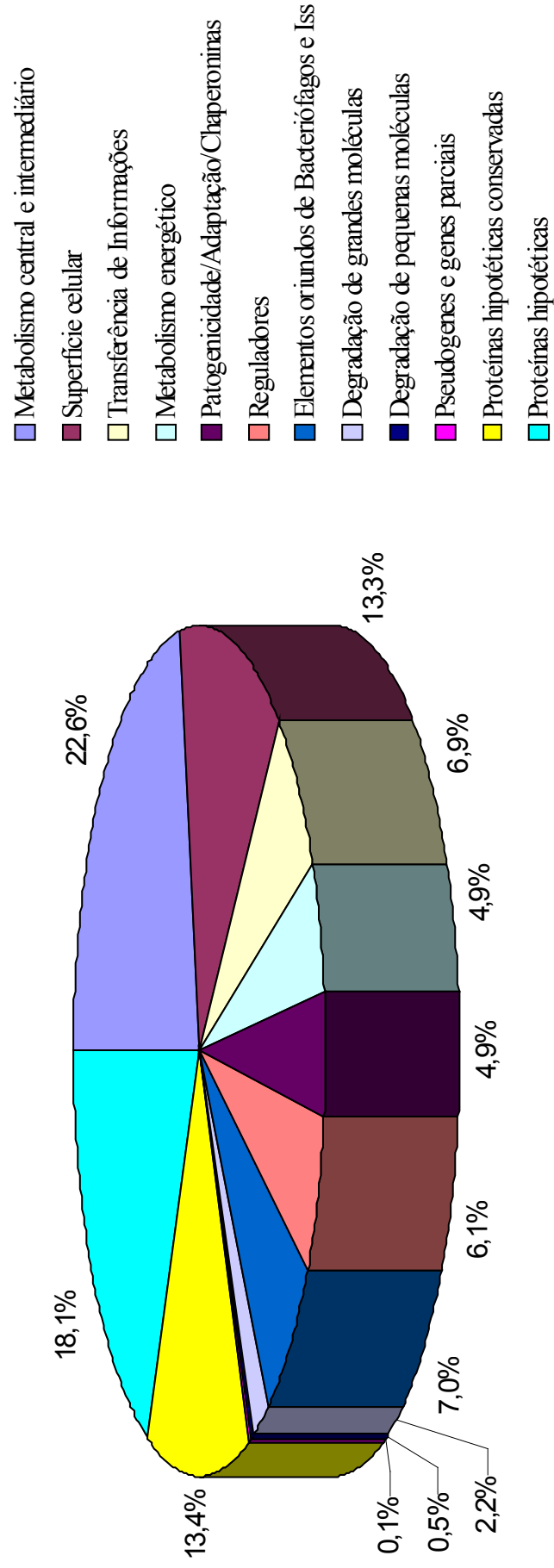
Através da análise das 1802 proteínas (49,2% do genoma) classificadas como proteínas hipotéticas (Tabela 9; Figura 28), considerando-se a presença de domínios protéicos conservados (Pfam) e alinhamentos com *e-value* inferior a 10^{-3} (Blastp), foram verificadas similaridades com domínios protéicos conservados (Pfam), nas seqüências de 687 proteínas

Tabela 9 – Número e percentual de ORFs detectadas no genoma do *Gluconacetobacter diazotrophicus* PAL5, inicialmente classificadas como proteínas hipotéticas e posteriormente classificadas funcionalmente*.

Classe funcional	Número de ORFs	Percentual de ORFs no genoma
Metabolismo central e intermediário	138	3,9%
Superfície celular (membranas internas e externas, proteínas secretadas, estruturas de superfície, etc)	149	4,1%
Transferência de Informações (transcrição, tradução, modificação de DNA e RNA)	24	0,7%
Metabolismo energético (glicólise, transporte de elétrons, etc.)	16	0,4%
Patogenicidade/Adaptação/Chaperoninas	38	1,0%
Reguladores	103	2,8%
Elementos oriundos de Bacteriófagos e seqüências de inserção (ISs)	193	5,3%
Degradação de grandes moléculas	20	0,5%
Degradação de pequenas moléculas	6	0,1%
Pseudogenes e genes parciais	0	0%
Proteínas hipotéticas conservadas	447	12,2%
Proteínas hipotéticas	668	18,2%
Total	1802	49,2%

* Predição de função baseada na presença de domínios protéicos conservados (PFAM) e alinhamentos com *e-values* inferiores a 10^{-3} (Blastp).

Figura 28 – Distribuição dos genes detectados no genoma do *Gluconacetobacter diazotrophicus* PAL5 de acordo com sua classificação funcional, considerando os alinhamentos com *e-values* inferiores a 10^{-3} (Blastp) e presença de domínios protéticos conservados (PFAM).

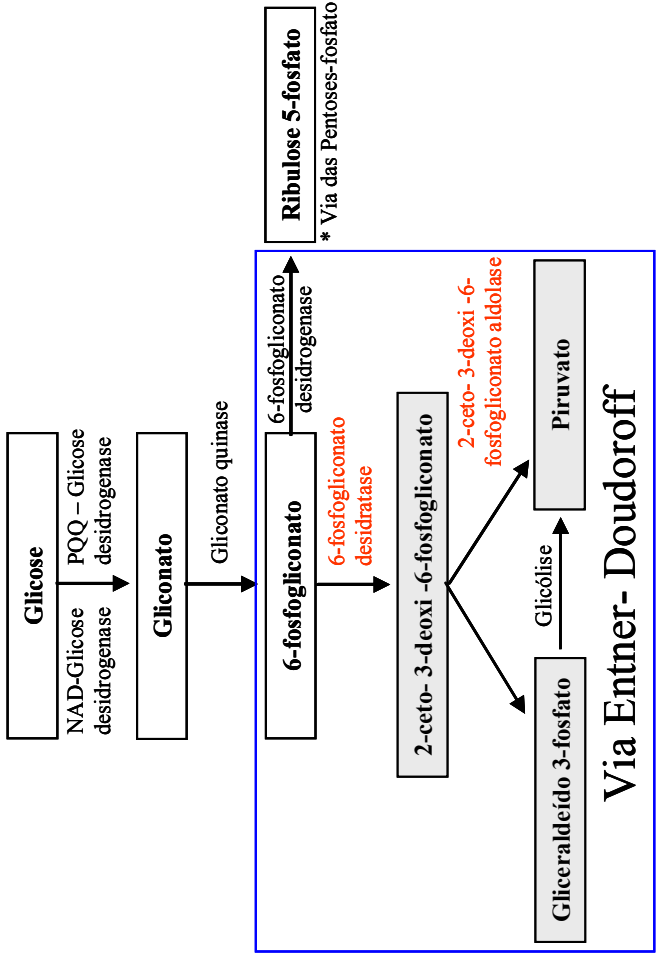
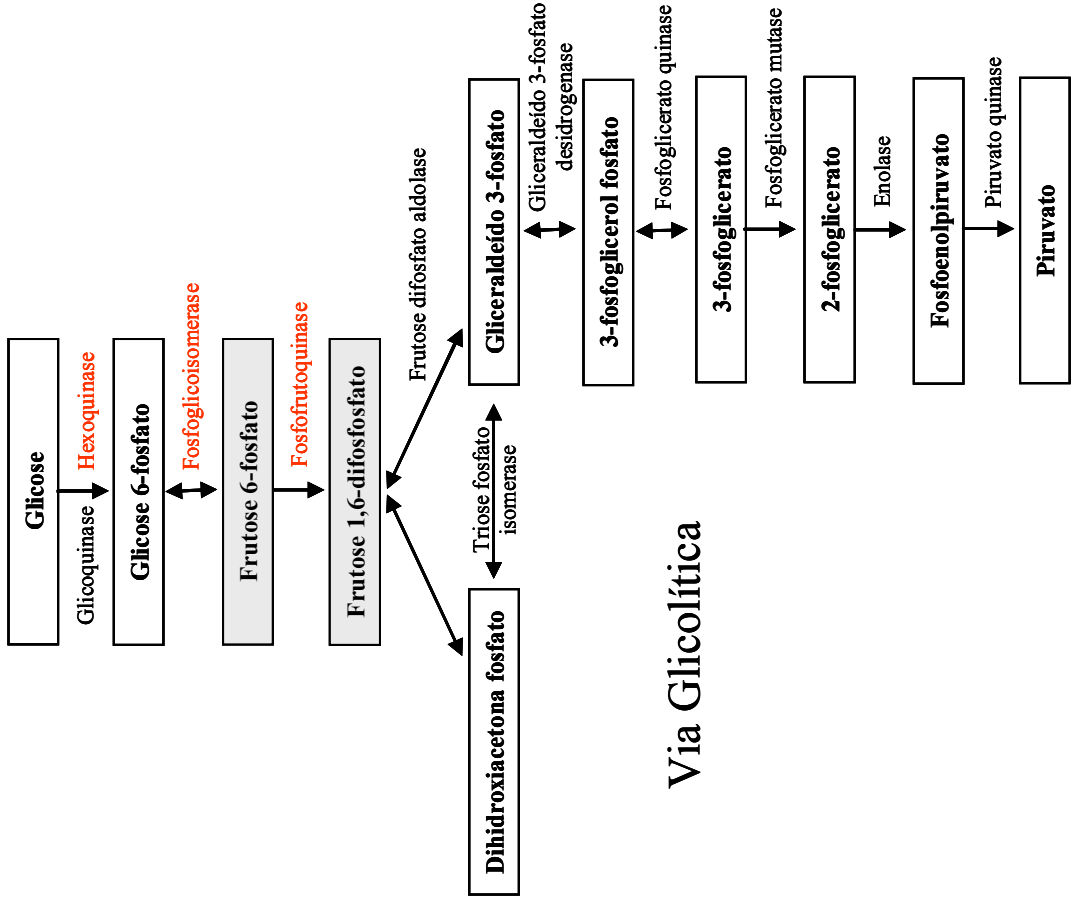


hipotéticas (18,8% do genoma) (Tabela 9; Figura 28), onde verificou-se um maior percentual de predição de funções, para genes inclusos na categorias funcionais: Elementos oriundos de bacteriófagos e seqüências de inserção (ISs) (5,3% do genoma; 193 genes); Superfície celular (4,1% do genoma; 149 genes); Metabolismo central e intermediário (3,9% do genoma; 138 genes); Reguladores (2,8% do genoma; 103 genes); e Patogenicidade/adaptação/chaperoninas (1,0% do genoma; 38 genes). Adicionalmente, as seqüências de 447 proteínas hipotéticas (12,2% do genoma), foram classificadas como proteínas hipotéticas conservadas, pois apresentaram similaridade com seqüências de proteínas hipotéticas, encontradas nos genomas de outros organismos seqüenciados. Outras 668 seqüências de proteínas hipotéticas (18,2% do genoma), continuaram recebendo esta designação, pois não apresentaram homologia com seqüências depositadas nos bancos de dados.

4.9 – Algumas Particularidades do genoma do *G. diazotrophicus* PAL5

No decurso da anotação do genoma do *G. diazotrophicus* PAL5, não foram detectadas a presença de 3 enzimas da fase preparatória da glicólise (hexoquinase, fosfoglicoisomerase e fosfofrutoquinase). Entretanto, ao invés de possuir hexoquinase, este endófito possui duas ORFs codificantes da enzima glicoquinase, que acredita-se ser a responsável pela produção de glicose 6-fosfato nesta célula. Além disso, as enzimas 6-fosfogliconato desidratase e 2-ceto-3-desoxi-6-fosfogliconato aldolase, que consistem em enzimas chaves da via metabólica de fermentação Entner-Doudoroff, também não foram detectadas (Figura 29). Adicionalmente, foi observada uma grande quantidade de genes relacionados com: biossíntese de todas as enzimas do ramo oxidativo da via das pentoses-fosfato (28 genes), complexo piruvato desidrogenase (12 genes), ciclo dos ácidos tricarboxílicos (9 genes) e cadeia transportadora de elétrons (92 genes) (Tabela 8). A

Figura 29 – Representação gráfica do metabolismo de glicose, evidenciando enzimas que estão ausentes no genoma do *Gluconacetobacter diazotrophicus* PAL5



As enzimas representadas em vermelho não foram detectadas no genoma do *G. diazotrophicus* PAL5. Nas caixas cinza estão representados os intermediários metabólicos ausentes em *G. diazotrophicus* PAL5. PQQ – Pirroloquinolina quinona, NAD – nicotinamida adenosina dinucleotídeo.

ausência das enzimas supracitadas, sugere que o ramo oxidativo da via das pentoses-fosfato, consista na principal via metabólica operante no catabolismo de glicose 6-fosfato deste endófito.

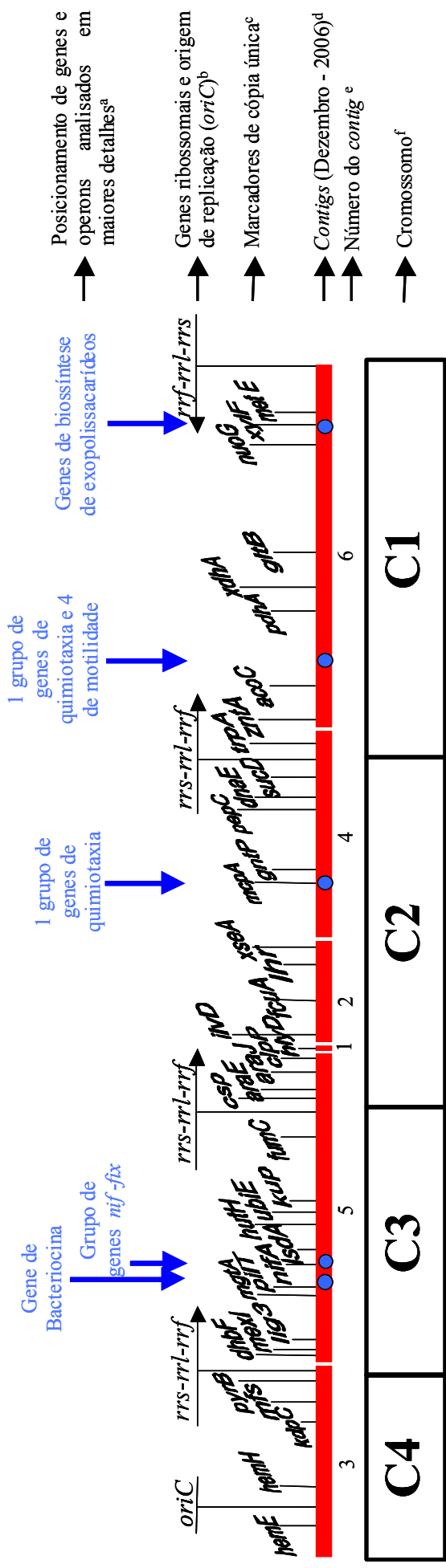
No que tange à capacidade de produção de bacteriocinas² pelo *G. diazotrophicus* PAL5, foi observado a presença de um gene de cópia única, localizado próximo ao grupo de genes *nif-fix*, no *contig* 5 da montagem de dezembro de 2006 (Figura 30). Este gene consiste numa seqüência de 822 nucleotídeos, codificante de uma proteína de 274 aminoácidos, que apresenta maior similaridade com a bacteriocina Linocina M18 de *Rhodospirillum rubrum* (Figuras 31 e 32). Além disso, o *G. diazotrophicus* PAL5 exibe na categoria funcional patogenicidade, adaptação e chaperoninas, um grande número de genes relacionados com resistência a drogas (32), mobilidade (25), adaptação a condições ambientais atípicas (21), quimiotaxia (20) e destoxificação celular (12) (Tabela 8).

Dentre os 32 genes relacionados com resistência a drogas, destacam-se 16 genes *mdt* que estão associados com resistência a múltiplas drogas. Estes se encontram dispersos por todo o cromossomo, entretanto, não foram detectados em seqüências plasmidiais. Adicionalmente, os genes relacionados com adaptação a condições atípicas (principalmente proteínas relacionadas a choque térmico) e destoxificação celular (glutathionina S-transferase, proteínas de resistência a cobre, proteínas de resistência a ácido fusárico e acriflavina, alquil hidroperóxido redutase e superóxido dismutase, principalmente), também foram encontrados dispersos por todo o cromossomo e ausentes nos plasmídios, exceto glutathionina S-transferase que foi detectado no plasmídio B (Tabela 5).

Dos 20 genes relacionados com quimiotaxia em *G. diazotrophicus* PAL5 (Tabela 8), 8 codificam a produção de receptores de quimiotaxia McpA (genes *mcpA*) e 12 codificam proteínas

² Bacteriocinas consistem em peptídeos biologicamente ativos com atividade tóxica, produzidos e secretados por algumas bactérias contra cepas bacterianas da mesma espécie ou de espécies diferentes.

Figura 30 – Posicionamento de alguns genes e grupos gênicos analisados em maiores detalhes, em relação a 6 contigs gerados na montagem de dezembro de 2006, alinhados no mapa físico do cromossomo do *Gluconacetobacter diazotrophicus* PAL5.



MÉDIA DO TAMANHO DO CROMOSSOMO CIRCULAR

4232 Kb ± 47 Kb

As seqüências das sondas dos genes ribossomais e de cópia única foram utilizadas no alinhamento de contigs do Projeto Riogene. O tamanho dos fragmentos de DNA gerados com a enzima *I-CeuI* e os contigs do Projeto Riogene, assim como a posição dos marcadores moleculares estão em escala. a – Posicionamento da seqüências codificantes nos contigs: bacteriocina (seqüência codante entre os nucleotídeos 417033 e 417854), grupo de genes *nif-fix* (432227 e 455262), grupo de genes relacionados com quimiotáxia (203909 e 211404), 1 grupo de genes relacionados com quimiotáxia e 4 com motilidade celular (1153795 e 1212200) e genes de biossíntese de exopolissacarídeos (227978 e 247925; 293964 e 297808); b – Localização e orientação dos operons ribossomais e localização da origem de replicação. c – Nomenclatura dos genes de cópia única utilizados para construção de sondas de DNA (vide Tabela 2). d – Alinhamento dos contigs montados no Projeto Riogene em dezembro de 2006. e – Os tamanhos dos contigs alinhados foram: contig 1 (14,1 Kb), 2 (361,0 Kb), 3 (675,9), 4 (749,2 Kb), 5 (1052,2 Kb), 6 (1285,0 Kb); f – Alinhamento dos fragmentos de DNA gerados com a enzima *I-CeuI* no mapa físico do cromossomo do *Gluconacetobacter diazotrophicus* PAL5.

Figura 31 – Página do Sistema Bionotes exibindo propriedades da ORF codificante da bacteriocina de *Gluconacetobacter diazotrophicus* PAL5.

BioNotes *Gluconacetobacter diazotrophicus*

Version Index Search Run Blast Download Contigs > 10000 Overview

Glimmer ORF Search Result - Detail

[Contig Search Result](#) Annotations

ORF Properties								
ORF	Contig	Reading Frame	Sequence Length		Position		Codon	
			Nucleotide	Aminoacid	Start	End	Start	Stop
0691	contig5	+3	822	274	417033	417854	atg	tga

Community Annotation	
Status	Gene
Valid	Putative bacteriocin protein

Nucleotide Sequence

```

ATGAACAATCTGCATCGGGAACCTCGCGCCGATTTCTGAAGCCGCATGGGCGCAGATCGAG
GAAGAGGCTCCCACGCTCAAGCGCTACCTGGCCGCCCGCGCGTGGTCGATGTGCCG
GAGGCGAAGGATTCGGGTTTTTCGGCTGTTCGGCACCAGCCATGTCGAGCGCATCGACGCA
CCCGGAAGCGATATTCGCGCCCGTCCGGCGCAACGTGCTGCCGCTGGTGGAACTGCGCGTG
CCCTTCACGCTGGCGCGGGACGCGATCGACGATGTGGAGCGTGGCGCGGGTGATTTCGGAC
TGGCAACCCCTGAAGGACGCCGCGAAGAAAATCGCGTTTCGCCGAAGACCGGGCGGTATTC
GACGGTTACGCCGCTGCCGGAATCCTGGGCCTTCGTGAGGGCACCTCGAACCCGAAACTG
GCCCTGCCGCTTCCGCCAGCGACTATCCGGCGGCCATCGCGGCGGCGCTGAACCAGCTT
CGCCTTGCCGCGGTCAACGGACCCTACGCCGTGGTGCCTCGGCGCCGCGCTCTATACCGCG
CTGAGCGGCGGCGACGATGAGGGATATCCCGTCTTCGGCATATCGAGAGCCTGATCGAC
GGCAAGATCGTCTGGGCCCCGGCGATCGAGGGGGGCTTCGTCTGTCCACCCGTGGCGGC
GATTCGAACTGGATATCGGCCAGGACTTTTCCATCGGGTACAGCAGCCATTCCGCGGAT
TCCGTGGAGCTTTACCTTCAGGAGAGCTTTACGTTCCAGTTGCTGACGACCGAGGCATCC
GTGGCATTGTCAGCTGCGAAGAACGAAGCGCCCTCGGCCCG
    
```

Local Online NCBI Blast

Parte inferior do formulário

Aminoacid Sequence

```

MNNLHRELAPISEAAWAQIEEEASRTLKRYLAARRVVDVPEAKGFVFSAVGTGHVERIDA
PGSDIRAVRRNVLPVELRVPFTLARDAIDDVERGAGDSWQPLKDAAKKIAFAEDRAVF
DGYAAAGILGLREGTSNPKLALPSSASDYPAIAAALNQLRLAGVNGPYAVVLGAGVYTA
LSGGDDEGYPVFRHIESLIDGKIVWAPAIEGGFVLSTRGGDFELDIGQDFSIGYSSHSAD
SVELYLQESFTFQLLTTEASVALSAKNEAPSAP
    
```

Local Online NCBI Blast

Parte inferior do formulário

[Local Offline NCBI BlastP ORF x NR](#)

Accession	Comment	Score	E Value
YP_426062.1	Linocin_M18 bacteriocin protein [Rhodospirillum rubrum ATCC 11170]	229 bits (583)	1e-59

* A seqüência de aminoácidos desta ORF, apresenta um domínio conservado, com melhor *score* de alinhamento de PFAM com a bacteriocina linocina M18 (*Score* = 368.4 ; *e-value* = 9e-108).

Figura 32 – Alinhamento da sequência de 274 aminoácidos da bacteriocina de *Gluconacetobacter diazotrophicus* PAL5, em relação a bacteriocina Linocina M18 de *Rhodospirillum rubrum* *.

```

Query: 1 MNLLHRELAPISEAAWAQIEEEASRTLKRYLAARRVVDVPEAKGFGFSAVGTGHVERI-- 58
      MN+L R+LAPIS AWA+IE EA TL LAAR+VVD G+ S+V G E +
Sbjct: 1 MNDLMRDLAPISAKAWAEIETEARGTLTTLAARKVVDFKGPLGWDASSVSLGRTEALAE 60

Query: 59 --DAPGSD--IRAVRRNVLPVLELRVFFTLARDAIDDVERGAGDSDWQPLKDAAKKIAFA 114
      A GS + +R V PL+EL VPFTL R ++ + RGA D+D P+ +AA+ IA A
Sbjct: 61 EPKAAGSAAVTVRKRAVQPLIELCVFTLKRAELEAIARGASDADLLDPVIEAARAIAIA 120

Query: 115 EDRAVFDGYAAAGILGREGTSNPKLALPXXXXXXXXXXLNQLRLAGVNGPYAVVLG 174
      EDRAVF G+AA GI G+ E ++ L LP L IR GV+GPYA+VIG
Sbjct: 121 EDRAVFHGFAGGITGIGEASAEHALDLPADLADFFGVLVRALAVLRDRGVDGPPYALVLG 180

Query: 175 AGVYTAALSGDDEG-YPVFRHIESLIDGKIVWAPAIEGGFVLSTRGGDFELDIDIGQDFSIG 233
      VY L G YPV +H+ L +G ++WAP ++G ++S RGGDFEL +G+DFSIG
Sbjct: 181 RTVYQQLMETTTTPGGYPVLQHVRRIFEGPLIWAAPGVDGAMLISQRGGDFELTVGRDFSIG 240

Query: 234 YSSHADSVELYLQESFTFQLLTTEASVALSAAKNEAPSA 273
      Y H A SV LYLQES TF+ L EA+V L A A
Sbjct: 241 YHDHDAQSVHLYLQESMTFRCLGPEAAVPLRGLSQAATKA 280

```

* Os Resultados de Blastp exibiram score de 229 bits (583) e e-value de 1e-59.

de quimiotaxia Che (genes *che*). Estes genes formam 2 grupos gênicos distintos com 5 (*cheW*, *cheR*, *mcpA*, *cheA* e *cheB*) e 8 genes (*mcpA*, *cheY*, *cheA*, *cheW*, *cheR*, *cheB*, *cheY* e *cheD*), respectivamente. Entretanto, outros 6 genes *mcpA* e 1 gene *cheL* encontram-se dispersos ao longo do cromossomo. Cabe ressaltar que estes genes apresentam maior similaridade, com seqüências oriundas do *Rhizobium etli*, e ambos microrganismos apresentam uma organização gênica muito similar ao *Caulobacter crescentus* (Figura 33).

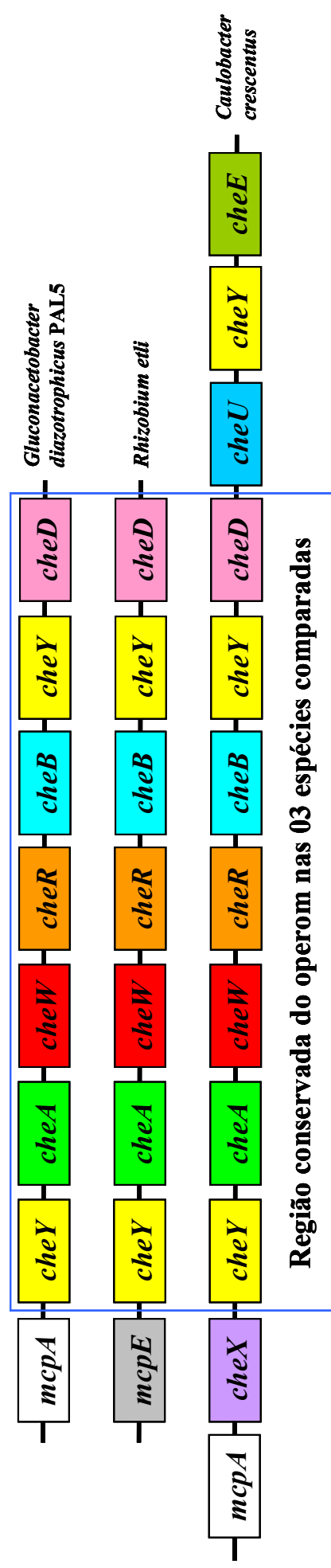
Os genes relacionados com motilidade celular (Tabela 8) formam 4 grupos gênicos distintos, com 5, 8, 8 e 4 genes, respectivamente, e codificam a síntese de todas as proteínas necessárias à produção de flagelos (proteínas de corpúsculo basal, proteínas motoras, proteínas de ganchos e anéis flagelares, e flagelinas). Estes genes apresentaram maiores homologies em relação aos genes de *Caulobacter crescentus* e de microrganismos inclusos na ordem *Rhizobiales* (Dados não exibidos). Uma outra característica interessante é que os 4 grupos de genes relacionados com motilidade, e 1 relacionado com quimiotaxia (grupo gênico da Figura 33), estão alocados um ao lado do outro no cromossomo do *G. diazotrophicus* PAL5, ao longo de uma seqüência de aproximadamente 58400 nucleotídeos, presente na extremidade terminal do *contig* 6 (seqüências codantes entre os nucleotídeos 1153795 e 1212200), enquanto que o outro grupo de genes relacionados com quimiotaxia, encontra-se alocado na porção inicial do *contig* 4 (seqüências codantes entre os nucleotídeos 203909 e 211404) (Figura 30).

Especial destaque deve ser atribuído a 11 genes relacionados à produção de exopolissacarídeos, que foram inclusos na categoria funcional superfície celular. Quatro destes genes foram classificados como genes *exo* (*exoF*, *exoO*, *exoP* e *exoY*), 7 como genes *gum* (*gumB*, *gumD*, *gumE*, *gumH*, *gumJ*, *gumK* e *gumM*), e 38 como genes responsáveis pela produção de glicosiltransferases, inclusos na categoria funcional de metabolismo central e intermediário

(Tabela 8). Estes genes são requeridos para biossíntese de polissacarídeos de superfície celular (principalmente succinoglicanos), que apresentam importante papel no mecanismo de interação planta-bactéria (Kaneko *et al.*, 2002; Thieme *et al.*, 2005; Krause *et al.*, 2006). Cabe ressaltar que os genes *exo* e *gum*, apresentaram maior homologia com genes de microrganismos incluídos na ordem *Rhizobiales*, e estão alocados na porção inicial do *contig* 6 (Figura 30), enquanto os genes de glicosiltransferases são encontrados distribuídos ao longo de todo o cromossomo bacteriano.

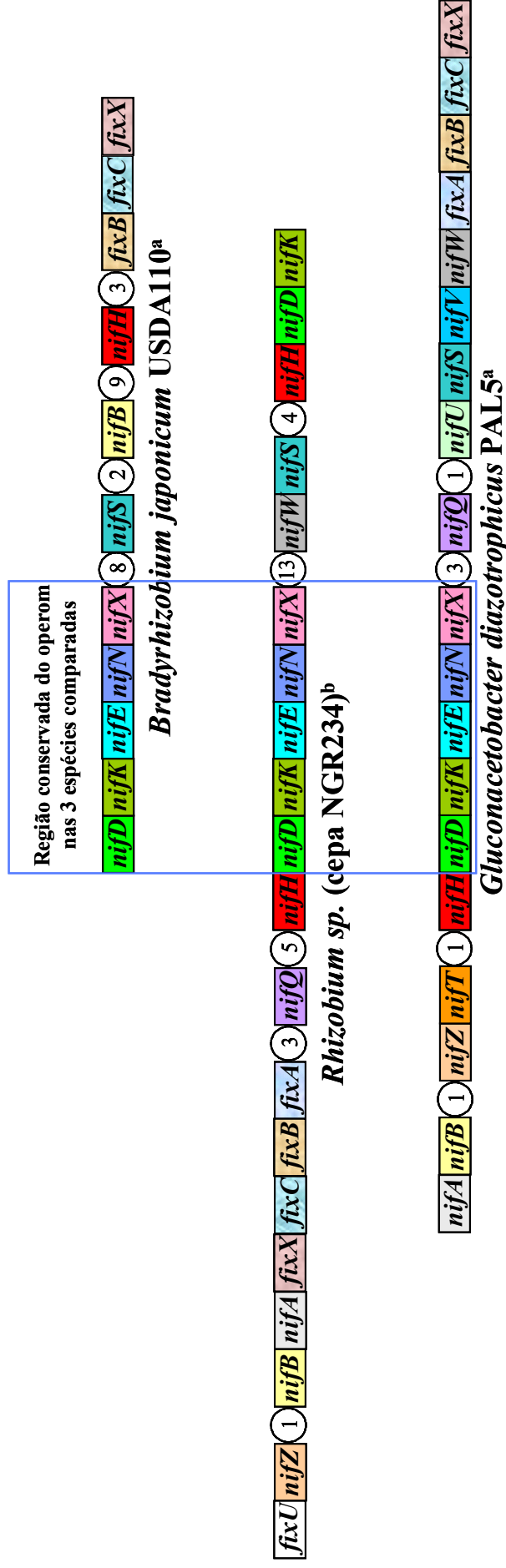
No cromossomo do *G. diazotrophicus* PAL5, foi observado a presença de um grupo gênico *nif-fix* composto por 19 genes (15 genes *nif* e 4 *fix*) (Figura 34), localizado nas seqüências codantes entre os nucleotídeos 432227 e 455262 do *contig* 5 (montagem de dezembro de 2006) (Figura 30). Este grupo gênico teve seu posicionamento determinado no cromossomo de *G. diazotrophicus* PAL5, através de experimentos de hibridização, com sonda radioativa do gene *nifA* (regulador transcricional da nitrogenase) contra membranas de PFGE (Figura 7). A montante do gene *nifA*, encontra-se os genes *nifB*, *nifZ*, *nifT*, *nifH*, *nifD*, *nifK*, *nifE*, *nifN*, *nifX*, *nifQ*, *nifU*, *nifS*, *nifV*, *nifW*, *fixA*, *fixB*, *fixC* e *fixX* (Figura 34). Cabe ressaltar que as seqüências destes genes apresentaram maiores similaridades, com seqüências do plasmídeo de simbiose do *Rhizobium sp.* (cepa NGR234), e seqüências cromossômicas do *Bradyrhizobium japonicum* USDA110. Quando comparada a organização destes genes nas 3 espécies, observou-se grande similaridade entre os genes *nifHDKENX*, de *G. diazotrophicus* PAL5 e *Rhizobium sp.* (cepa NGR234). Adicionalmente, foi observado 3 genes *mod* (*modA*, *modB* e *modC*) a jusante do grupo de genes *nif-fix*, que são responsáveis pela produção de proteínas relacionadas com aquisição de molibdênio, e ao longo dos *contigs* 5 e 6 (Figura 30), foram observados um total de 8 reguladores transcpcionais relacionados com assimilação de nitrogênio, onde destaca-se os genes *ntr*, e uma cópia adicional do gene *nifU* alocada no *contig* 4 (Dados não exibidos).

Figura 33 – Organização de um grupo gênico relacionado com quimiotaxia do *Gluconacetobacter diazotrophicus* PAL5, em relação a grupos gênicos de quimiotaxia do *Rhizobium etli* e *Caulobacter crescentus*



mcp – Genes que codificam receptores de quimiotaxia; *che* – Genes que codificam proteínas de quimiotaxia

Figura 34 – Organização do grupo de genes *nif-fix* do *Gluconacetobacter diazotrophicus* PAL5, em relação ao grupo de genes *nif-fix* de *Bradyrhizobium japonicum* USDA110 e *Rhizobium* sp. (cepa NGR 234).



O número arábico representado dentro dos círculos, representa o número de genes de proteínas hipotéticas encontradas entre genes *nif* ou *nif* e *fix*, ao longo das seqüências genômicas dos microrganismos analisados. a – Operon localizado no cromossomo do organismo; b – Operon localizado em plasmídeo de simbiose.

5 – DISCUSSÃO

5.1 – Mapeamento físico do cromossomo do *G. diazotrophicus* PAL5 através de fragmentos de macrorestrição

Para execução do mapeamento físico do cromossomo do *Gluconacetobacter diazotrophicus* PAL5, foram escolhidas as enzimas de restrição de raro corte *PacI* e *SwaI* (Tabela 1), devido ao grande percentual de GC em seu genoma (66,4%). O objetivo desta estratégia foi gerar um pequeno número de fragmentos de DNA (5 e 6, respectivamente) facilmente manipuláveis pela metodologia de PFGE. Adicionalmente, estas enzimas foram relatadas por outros autores que mapearam cromossomos ricos em GC (Ramos-Díaz & Ramos, 1998; Widjaja, Suwanto & Tjahjono, 1999; Llamas *et al.*, 2002), como enzimas adequadas à construção de mapas físicos de macrorestrição, pois clivam estes cromossomos em um pequeno número de grandes fragmentos de DNA.

Com o objetivo de determinar o número e a posição dos operons ribossomais (operons *rrn*), o cromossomo do *G. diazotrophicus* PAL5 foi submetido a digestões com a enzima de restrição *I-CeuI* (Tabela 1), que reconhece uma seqüência de 26 pb altamente conservada na subunidade ribossomal 23S (gene *rrl*). Convém mencionar que a utilização desta enzima foi relatada na literatura por diversos autores (Ramos-Díaz & Ramos, 1998; Widjaja, Suwanto & Tjahjono, 1999; Melkerson-Watson *et al.*, 2000; Zé-Zé, Tenreiro & Paveia, 2000; Dudez *et al.*, 2002, El-Osta *et al.*, 2002), como uma estratégia precisa de localização de genes *rrl* em cromossomos bacterianos.

Na Tabela 1 é exibido o número e os tamanhos estimados dos fragmentos de DNA, gerados a partir de digestões simples e duplas do cromossomo de *G. diazotrophicus* PAL5. Entretanto, cabe ressaltar que os fragmentos CP6 e PS6 não foram detectados sob nenhuma

condição de eletroforese empregada neste estudo (Figura 2A–D). Os resultados obtidos sugerem que estes fragmentos de DNA possuem tamanhos inferiores a 20 Kb³. Diversos autores (Römling *et al.*, 1989; Chan & Goodwin, 1999; Melkerson-Watson *et al.*, 2000), relatam baixos sinais fluorescentes em fragmentos de DNA inferiores a 20 Kb, em géis corados com brometo de etídio ou syber green. Os resultados demonstrados neste trabalho sugerem a ocorrência de baixos sinais fluorescentes, nos fragmentos de DNA preditos com tamanhos inferiores a 20 Kb nos géis de PFGE.

O tamanho do cromossomo do *G. diazotrophicus* PAL5 foi estimado em 4.232 Kb, com desvio padrão de 47 Kb, através da utilização da metodologia de PFGE (Tabela 1). Adicionalmente, os dados do Projeto Riogene revelam que o cromossomo do *G. diazotrophicus* PAL5 apresenta 4137,5 Kb (Tabela 6), à nível de seqüências de DNA montadas em 6 *contigs* (Figura 18). Estes dados sugerem que mapeamentos físicos são ferramentas eficientes na estimacão de conteúdos genômicos, em concordância com diversos autores que utilizaram esta metodologia para aferição do tamanho e morfologia de cromossomos bacterianos (Ramos-Díaz & Ramos, 1998; Blank & Stemke, 2000; Zé-Zé, Tenreiro & Paveia, 2000; Marcone & Seemüller, 2001; Oana *et al.*, 2002). Adicionalmente, é importante salientar que a estimativa correta do conteúdo genômico de um microrganismo, corresponde a uma valiosa informação em projetos de seqüenciamento genômico, pois através deste dado pode-se estimar o seu *status*, ou seja, quantas vezes o genoma foi seqüenciado, e qual o total de cobertura do genoma que um determinado projeto apresenta, auxiliando assim na tomada de decisões em relação à abordagem no seqüenciamento, principalmente em projetos que utilizam a metodologia de *shotgun*.

³ Predição baseada no tamanho estimado do cromossomo circular do *G. diazotrophicus* PAL5 e número de cortes causados pelas enzimas *I-CeuI*, *PacI* e *SwaI* quando utilizadas individualmente.

Para confirmar a hipótese de que o cromossomo do *G. diazotrophicus* PAL5 possui 4 operons ribossomais, foram realizadas hibridizações com sondas de DNA correspondentes aos genes 16S (*rrs*), 23S (*rrl*) e 5S (*rrf*) (Figura 3; Figura 14; Figura 15; Tabela 2), que geralmente estão organizados nesta ordem em genomas bacterianos (Ramos-Díaz & Ramos, 1998; Bricker, 2000). Os resultados de hibridização sugerem uma estrutura circular de cromossomo, contendo 4 operons ribossomais, sendo que um dos operons possui uma orientação invertida em relação aos demais. O operon ribossomal invertido encontra-se alocado entre os fragmentos C1 e C4 (Figura 17). Os dados que suportam esta afirmação, se baseiam no forte sinal obtido no fragmento C4, e ausência de sinal para o fragmento C1, nas hibridizações com as sondas *rrs* e *rrl* (Figura 3; Figura 14). Por outro lado, nos resultados obtidos com a sonda *rrf* ocorreu o oposto (Figura 3; Figura 15). Estes resultados estão em concordância com diversos autores (Ramos-Díaz & Ramos, 1998; Blank & Stemke, 2000; Padovan *et al.*, 2000; Zê-Zê, Tenreiro & Paveia, 2000; El-Osta *et al.*, 2002; Oana *et al.*, 2002), que utilizaram esta estratégia para confirmar o número, localização e inversões de operons ribossomais.

Através dos resultados de hibridização com 42 sondas diferentes (Figuras 4-15; Tabela 2), foram verificadas sobreposições entre fragmentos individuais de DNA, incluídos nos 6 diferentes perfis de PFGE do DNA cromossomal de *G. diazotrophicus* PAL5, permitindo a detecção de 12 grupos de genes, que demonstraram diferentes perfis de hibridização (Tabela 3). Estes dados foram utilizados para determinar o posicionamento destes marcadores moleculares, assim como executar o alinhamento de 45 fragmentos de DNA e 15 sítios de restrição neste cromossomo bacteriano (Figura 16), dispensando o uso de digestões parciais do cromossomo, PFGE bidimensional e clones de ligação. Estratégia similar foi utilizada na construção do mapa físico do cromossomo da rickettsia *Cowdria ruminantium*, realizado por Villiers *et al.* (2000).

No mapa físico do cromossomo do *G. diazotrophicus* PAL5, foram alocados 42 marcadores genéticos (Figura 16), e dentre estes marcadores, alguns genes como o *nifA* e *lsdA*, estão presentes em diversos genomas integrados a operons, sugerindo a presença de genes relacionados em sua periferia. Adicionalmente, foram detectados os genes *hemE* e *hemH*, que são considerados marcadores estratégicos, devido ao fato de que em muitos genomas bacterianos foram descritos como genes flanqueadores da origem de replicação cromossômica (*oriC*) (Capela *et al.*, 2001; Kaneko *et al.*, 2002). Por esta razão, no mapa físico do cromossomo do *G. diazotrophicus* PAL5, o fragmento C4 foi alocado na posição zero, pois este é o menor fragmento de DNA que hibridizou com os genes *hemE* e *hemH*. Adicionalmente, os resultados de anotação do genoma deste endófito, também revelaram uma grande inversão no conteúdo de GC, assim como a presença dos genes *parAB* (envolvidos em particionamento cromossomal), *gidAB* (genes de divisão inibidos por glicose), *rho* (codante do fator de terminação transcricional Rho) e *dnaJQ* (responsável pela biossíntese das chaperoninas DnaJ e DnaQ), próximos ao gene *hemE* (Dados não exibidos), que correspondem a critérios utilizados para estabelecer o posicionamento correto da origem de replicação nos mais variados genomas (Capela *et al.*, 2001; Nierman *et al.*, 2001; Kaneko *et al.*, 2002).

A utilização de sondas de DNA de genes de cópia única (Tabela 2), demonstrou uma grande importância no Projeto Riogene, pois diversos genes selecionados estavam alocados em diferentes *contigs*, com tamanhos superiores a 10 Kb, permitindo o alinhamento destes *contigs* no cromossomo bacteriano, assim como o procedimento de geração de *scaffolds* (Figura 17), facilitando o processo de finalização deste projeto de sequenciamento genômico, realizado pela metodologia de *shotgun*. O processo de finalização vem sendo executado através da metodologia de PCRs combinatórios, utilizando oligonucleotídeos desenhados para anelarem em regiões de

cópia única, localizadas o mais próximo possível das extremidades dos *contigs*, seguido de reações de seqüenciamento dos produtos de PCR obtidos. Esta estratégia reduziu o número de experimentos de hibridização, e preveniu um problema comumente observado em outros mapas físicos, que consiste no agrupamento de marcadores genéticos em um pequeno segmento do cromossomo (Villiers *et al.*, 2000; Dudez *et al.*, 2002).

Quando este mapeamento físico foi concluído, o seqüenciamento genômico total do *G. diazotrophicus* PAL5, tinha alcançado 99,99% da seqüência genômica montada em 18 *contigs*, com tamanhos girando de 12,8 a 1251,3 Kb (legendas das Figuras 17 e 18; Tabelas 4 e 5). Os 42 marcadores genéticos alocados neste mapa físico, permitiram realizar o alinhamento de 16 *contigs* (Figura 17), totalizando 4137,5 Kb, o que consiste em 97,6% da cobertura do genoma. Além do mais, os resultados de alinhamento dos *contigs* de *G. diazotrophicus* PAL5 no mapa físico, sugerem um alto grau de integridade dos *contigs* montados no projeto de seqüenciamento genômico, e auxiliaram a ordenação e orientação destes *contigs*, facilitando o processo de finalização através da metodologia de PCRs combinatórios.

A integridade dos *contigs* montados no Projeto Riogene foi confirmada, quando realizou-se a sobreposição do alinhamento dos 16 *contigs*, gerados no Projeto Riogene em maio de 2006, em relação 6 *contigs* gerados na montagem de dezembro de 2006 (Figura 18), pois verificou-se o fechamento de buracos genômicos entre os *contigs*: 4, 7 e 6; 2, 18, 14, 8 e 15; 10, 5, 11 e 9; e 13 e 1 (montagem de maio de 2006), o que gerou os *contigs* 3, 5, 2 e 6 da montagem de dezembro de 2006, respectivamente, além de confirmar o posicionamento dos *contigs* 3 e 17, que na montagem de dezembro de 2006 receberam a nomenclatura 4 e 1, respectivamente. Cabe ressaltar que estes dados enfatizam a importância do alinhamento de *contigs* e geração de *scaffolds*, no decurso do processo de finalização do genoma deste endófito.

Os dados supracitados estão em concordância com estudos prévios (Weinel *et al.*, 2001; Blank & Stemke, 2000; Lai *et al.*, 1999), que consideram mapas físicos de macrorestrição como ferramentas eficientes, para o controle de qualidade das seqüências montadas, a partir dos dados de seqüenciamento gerados pela metodologia de *shotgun*, uma vez que esta abordagem permite a detecção de possíveis erros de montagem dentro da seqüência de um *contig*, ou erros de ordenação de *contigs* em *scaffolds*, e ainda auxilia na re-ordenação de uma montagem, especialmente em regiões flanqueadoras de genes ribossomais.

5.2 – Detecção de dois plasmídeos no genoma de *G. diazotrophicus* PAL5

Os *contigs* 12 e 16, obtidos na montagem de maio de 2006, foram classificados neste estudo como plasmídeos, baseado em quatro observações: 1 – Estas seqüências não apresentaram sinais de hibridização, ao utilizar sondas específicas para regiões de cópia única destes plasmídeos, contra os mais variados tipos de membranas contendo DNA genômico (Figuras 19 e 20), o que sugere uma origem extracromossomal para estas seqüências; 2 – Os genes anotados nestas seqüências, são comumente encontrados na estrutura de plasmídeos (Tabela 4; Tabela 5); 3 – Estes *contigs* correspondem a longas seqüências de DNA (38,8 e 16,6 Kb), que apresentam alta qualidade de montagem; 4 – Estes *contigs* foram montados a partir de seqüências de clones oriundos de diferentes bibliotecas genômicas.

Adicionalmente, a anotação dos *contigs* 12 e 16 (Tabela 4; Tabela 5) revelou em suas seqüências, a existência de genes relacionados com produção de proteínas pertencentes ao sistema de transferência horizontal de genes por conjugação (Proteína TrbI, Proteína TraF; proteína TraL; proteína BfpH; e proteínas de via de secreção tipo IV: componentes VirD2 e B4), assim como genes que codificam a produção de proteínas de mobilização, proteína A de partição

cromossomal, ATPase de segregação cromossômica, proteínas envolvidas em replicação de DNA, Fator C de replicação, transposase e resolvase. Cabe ressaltar que todos estes genes são compatíveis com a descrição de diversos autores, que relatam a presença destes em seqüências plasmídicas de diferentes microrganismos, tais como: *Agrobacterium tumefaciens* (Suzuki *et al.*, 2000), *Caulobacter crescentus* (Nierman *et al.*, 2001), *Xylella fastidiosa* (Bhattacharyya *et al.*, 2002), *Pseudomonas syringae* pv. *Maculicola* (Stavrínides & Guttman, 2004), *Azoarcus spp.* (Rabus *et al.*, 2005) e *Xanthomonas campestris* pv. *vesicatoria* (Thieme *et al.*, 2005).

O fato dos *contigs* 12 e 16 (versão de maio de 2006) serem classificados como plasmídeos, tendo como base os resultados de anotação é bastante aceitável, pois apesar do *G. diazotrophicus* PAL5, ser descrito na literatura como uma cepa que não apresenta plasmídeos (Munõz-Rojas, Fuentes-Ramírez & Caballero-Mellado, 2005), e destes plasmídeos não serem detectados por técnicas convencionais de purificação (Sambrook & Russell, 2001), acredita-se que eles estejam presentes em cópia única na célula de *G. diazotrophicus* PAL5, o que dificulta a sua detecção em gel de agarose. Entretanto, estes plasmídeos foram detectados no seqüenciamento genômico, pois a partir da clonagem de suas seqüências, foi possível amplificá-las e seqüenciá-las.

Para reforçar esta hipótese, existem relatos da descoberta de dois plasmídeos adicionais em *Pseudomonas syringae* pv. *tomato* DC3000, com tamanhos de 67,5 e 73,7 Kb, que passou a ser descrita como possuidora de quatro plasmídeos (Stavrínides & Guttman, 2004), e também foi descoberto um plasmídeo adicional de 19,1 Kb em *Xanthomonas campestris* pv. *vesicatoria* (Thieme *et al.*, 2005), que também passou a ser descrita como possuidora de quatro plasmídeos. Cabe ressaltar que somente após a implementação de projetos de seqüenciamento genômico destes microrganismos, foi possível detectar a existência de tais plasmídeos.

5.3 – Mapeamento físico do cromossomo do *G. diazotrophicus* PAL5 através de Cromossomos Artificiais Bacterianos (BACs).

Na análise dos 1315 perfis de restrição de BACs, submetidos a digestões duplas com as enzimas *HindIII* e *BamHI* (Figura 21), verificou-se uma grande redundância dos perfis de processados no *software Image 3.0i*, o que pode explicar o fato de serem obtidos 21 *contigs* sob uma condição mais estrigente de análise (*cutoff* 1e-12), mesmo após ter sido analisado o equivalente a 37,3 coberturas do genoma do *G. diazotrophicus* PAL5.

Cabe ressaltar que este valor de estringência é utilizado nos mais variados estudos de mapeamento físico de BACs, sob justificativa destes parâmetros minimizarem os erros de montagem e maximizarem o tamanho dos *contigs* gerados (Tao *et al.*, 2001; Tomkins *et al.*, 2001; González *et al.*, 2005), e ainda o fato de que outros grupos de pesquisa (Battistoni *et al.*, 2005; Goto, *et al.*, 2003) utilizaram parâmetros similares de análise, e obtiveram coberturas completas de genomas, com um número inferior de BACs analisados.

Adicionalmente, a partir dos 117 clones selecionados (Figura 23), foram realizadas 668 reações de sequenciamento na *Clemson University Genomic Institute* – Carolina do Norte – EUA, entretanto, apenas 43 reações resultaram em seqüências de boa qualidade (6,4% de aproveitamento), e dentre estas, 22 seqüências alinharam com seqüências repetitivas no genoma deste endófito, sendo aproveitadas apenas 21 seqüências para o alinhamento de *contigs* neste genoma (Figuras 25 e 26).

Em virtude dos fatos supracitados, associado com a difícil padronização do protocolo de sequenciamento dos BACs e alto custo das reações (8 vezes mais dispendiosa do que um sequenciamento comum), não foram realizadas novas rodadas de sequenciamento. Adicionalmente, é importante mencionar que foi observado neste estudo e por Tomkins e

colaboradores (2001), um baixo aproveitamento à nível de número de bases de qualidade, em seqüências oriundas de BACs, após remoção das seqüências do vetor (aproximadamente 100 bases).

Os melhores resultados obtidos com o mapeamento físico de BACs, foram as hibridizações com sondas radioativas dos genes de RNAr 16s e 23s (Figura 24), onde verificou-se que 7 BACs distintos estavam alocados em 4 pontos diferentes do mapa físico de BACs, corroborando com os dados do mapeamento físico de macrorestrição (Loureiro *et al.*, 2003; Santos *et al.*, 2004; Loureiro *et al.*, 2005; Figuras 16 e 17) e resultados de anotação obtidos pelo Projeto Riogene (Sistema Bionotes, 2006a; Tabela 6), que sugerem a presença de 4 operons ribossomais no cromossomo deste endófito. Adicionalmente cabe ressaltar que outros autores (Berent & Messick, 2003; Battistoni *et al.*, 2005), utilizaram estratégia similar para determinação do número de operons ribossomais.

Além disso, 14 seqüências oriundas de 7 BACs foram extremamente importantes para o alinhamento de 7 *contigs* da montagem de Novembro de 2004 (Figura 26), os quais representam 27,8% de cobertura do genoma (1176,9 Kb), quando comparado ao tamanho genômico estimado pelo mapa físico de macrorestrição (4232 Kb; Figura 16), pois estas seqüências forneceram subsídios para a montagem do maior *contig* deste genoma, denominado como *contig 1* na montagem de Maio de 2006, e *contig 6* na montagem de dezembro de 2006 (Figura 26). Estes resultados corroboram com outros autores (Battistoni *et al.*, 2005; Tomkins *et al.*, 2001) que relatam a eficácia da utilização de mapeamentos físicos de BACs para verificação de montagens genômicas geradas a partir de seqüências oriundas de bibliotecas de *shotgun*.

5.4 – Anotação do genoma do *G. diazotrophicus* PAL5

Os resultados de anotação do genoma do *G. diazotrophicus* PAL5, revelaram que este microrganismo não possui as enzimas fosfoglicoisomerase e fosfofrutoquinase, que fazem parte da fase preparatória da glicólise, e as enzimas 6-fosfogliconato desidratase e 2-ceto-3-desoxi-6-fosfogliconato aldolase que fazem parte da via metabólica de fermentação Entner-Doudoroff (Figura 29). Isto sugere que o ramo oxidativo da via das pentoses-fosfato seja a principal via metabólica operante para o catabolismo de glicose 6-fosfato. Estes resultados corroboram com dados prévios (Alvarez & Martinez, 1995; Luna *et al.*, 2002), que relataram o fato desta característica ser comum ao *G. diazotrophicus* PAL5 e aos demais membros da família *Acetobacteriaceae*, entretanto, estes estudos realizaram apenas análise de atividade enzimática em extrato total de células. Portanto, esta tese representa o primeiro relato a nível genômico do *G. diazotrophicus* PAL5, descrevendo a ausência das enzimas supracitadas, partindo-se do princípio que estas enzimas são bastante conservadas nos mais variados organismos.

Foram observados em *G. diazotrophicus* PAL5, 11 genes relacionados à produção de exopolissacarídeos (4 genes *exo* e 7 genes *gum*) e 38 responsáveis pela produção de glicosiltransferases, que são genes requeridos para biossíntese de polissacarídeos de superfície celular (principalmente succinoglicanos). Acredita-se que estes apresentem importante papel de reconhecimento celular, no decurso de mecanismos de interação planta-bactéria (Krause *et al.*, 2006), pois os genes *exo* e genes de glicosiltransferases, foram descritos no genoma de *Bradyrhizobium japonicum* USDA110 (Kaneko *et al.* 2002), como genes envolvidos na produção destes polissacarídeos, os quais correspondem a requisitos necessários ao reconhecimento planta-bactéria, e conseqüente ocorrência do processo de FBN. Adicionalmente, os genes *gum* foram descritos em *Xanthomonas campestris* pv. *Vesicatoria* (Thieme *et al.*, 2005), como importantes

genes relacionados ao processo de interação deste fitopatógeno com seus hospedeiros vegetais. Com base nestes dados, é possível especular que a eficiência das interações benéficas entre *G. diazotrophicus* PAL5 e seus hospedeiros vegetais, esteja relacionada em parte com estes genes responsáveis pela biossíntese de exopolissacarídeos.

Na categoria funcional patogenicidade/adaptação/chaperoninas, foram observados 45 genes (1,2% do genoma) relacionados com quimiotaxia e motilidade (Tabela 8) com maiores similaridades em relação a seqüências oriundas do *Caulobacter crescentus*. Cabe mencionar que *C. crescentus* é considerado como modelo de estudo de genes inclusos nesta categoria funcional, pois cerca de 2,5% do seu genoma (94 genes) são devotados a motilidade celular e quimiotaxia, além de possuir genes relacionados com degradação de compostos contendo anéis aromáticos, quebra de fontes de compostos orgânicos derivados de plantas, e ainda muitos fatores sigma extracitoplasmáticos, que conferem ao organismo a habilidade de responder a uma enorme variedade de flutuações ambientais (Nierman *et al.*, 2001). A partir desta observação, é possível inferir que o *G. diazotrophicus* PAL5, seria capaz de realizar associações benéficas com hospedeiros vegetais através da utilização destes genes.

Os dados que suportam a inferência supracitada, foi a observação de dois grupos de genes relacionados com quimiotaxia no genoma de *G. diazotrophicus* PAL5, contendo genes codantes de receptores de quimiotaxia McpA (genes *mcpA*) e proteínas de quimiotaxia Che (genes *che*), com maiores similaridades em relação a seqüências oriundas do *Rhizobium etli* (Gonzalez *et al.*, 2006). Adicionalmente, os genes relacionados com quimiotaxia destes dois microrganismos, apresentaram uma organização gênica muito similar ao *C. crescentus* (Figura 33), que é reconhecido como microrganismo modelo de estudos em quimiotaxia (Nierman *et al.*, 2001). Além disso, o *G. diazotrophicus* PAL5 apresenta 4 grupos de genes distintos, contendo genes

relacionados à biossíntese de flagelos, alocados entre os nucleotídeos 1153795 e 1212200 do *contig* 6 (montagem de dezembro de 2007; Figura 30), que apresentam maiores similaridades em relação aos genes de *C. crescentus* e microrganismos inclusos na ordem *Rhizobiales*.

O *G. diazotrophicus* PAL5 é descrito na literatura como uma cepa que apresenta atividade antagonista contra cepas de espécies diferentes, espécies estritamente relacionadas, ou da mesma espécie (Munõz-Rojas, Fuentes-Ramírez & Caballero-Mellado, 2005), além de ser um microrganismo capaz de adaptar-se a diferentes flutuações ambientais, especialmente, no decurso de competições intra e inter-específicas com microrganismos fitopatogênicos, rizosféricos e endofíticos, o que favorece o sucesso do processo de colonização deste endófito (Capela *et al.*, 2001; Pinõn *et al.*, 2002; Blanco *et al.*, 2005). Adicionalmente, os dados de anotação do genoma deste microrganismo fornecem subsídios para sustentar tais afirmações, pois foi detectada a presença de um gene de cópia única, localizado próximo ao grupo de genes *nif-fix* (Figura 30), codificante de uma proteína similar a bacteriocina Linocina M18 de *Rhodospirillum rubrum* (Figuras 31 e 32), e um grande número de genes relacionados com resistência a drogas (principalmente genes *mdt*), adaptação a condições ambientais atípicas (principalmente proteínas relacionadas a choque térmico) e destoxificação celular (como glutationina S-transferase, proteínas de resistência ao cobre, proteínas de resistência a ácido fusárico e acriflavina, alquil hidroperóxido redutase e superóxido dismutase) (Tabela 8).

Detectou-se no cromossomo do *G. diazotrophicus* PAL5, a presença de um grupo de genes *nif-fix* composto por 19 genes (15 genes *nif* e 4 *fix*) (Figura 34), localizado nas seqüências codantes entre os nucleotídeos 432227 e 455262 do *contig* 5 (montagem de dezembro de 2006; Figura 30). Este operom apresenta os genes *nifABZTHDKENXQUSVW* e *fixABCX* (Figura 34), e as respectivas seqüências destes genes apresentam maiores similaridades com seqüências do

plasmídeo de simbiose do *Rhizobium sp.* (cepa NGR234) (Freiberg *et al.*, 1997) e seqüências cromossômicas do *Bradyrhizobium japonicum* USDA110 (Kaneko *et al.*, 2002). Quando comparado a organização gênica deste grupo de genes nas três espécies, observou-se grande similaridade entre os genes *nifHDKENX* (responsáveis pela biossíntese da nitrogenase) de *G. diazotrophicus* PAL5 e *Rhizobium sp.* (cepa NGR234), o que sugere a possibilidade do *G. diazotrophicus* PAL5 ter adquirido estes genes relacionados ao processo de FBN, através de mecanismos de transferência horizontal de genes entre estas α -proteobactérias.

Neste genoma foi observado um grande número de genes com funções desconhecidas (1802 genes; 49,2% do genoma), sendo classificados como genes de proteínas hipotéticas, pois não apresentam homologia com seqüências depositadas nos bancos de dados. Este percentual corrobora com a literatura, pois nos mais variados genomas, 35 a 55% das seqüências geralmente recebem esta classificação (Schawartz, 2000; Capela *et al.*, 2001; Thieme *et al.*, 2005), e especula-se que quando as funções deste grande grupo de genes começarem a serem exploradas, novas vias metabólicas serão descobertas, e provavelmente causarão impactos importantes para a mais variadas áreas, incluindo a medicina e a biotecnologia (Fraser, Eisen & Salzberg, 2000).

Através da análise das 1802 proteínas (49,2% do genoma) classificadas como proteínas hipotéticas (Tabela 9; Figura 22), considerando-se a presença de domínios protéicos conservados (PFAM) e *e-values* inferiores a 10^{-3} (Blastp), verificou-se que 687 (18,8% do genoma) foram incluídas em outras categorias funcionais, 447 (12,2% do genoma) foram classificadas como proteínas hipotéticas conservadas, e 668 proteínas (18,2% do genoma) continuaram a serem classificadas como proteínas hipotéticas do *G. diazotrophicus* PAL5. Entretanto, convém enfatizar que a comparação de seqüências é a ferramenta mais prevalente para predição de funções de uma proteína, entretanto deve ser considerada apenas como uma primeira fase para elucidação da

função de uma proteína, tendo em mente que o trabalho experimental é necessário para validar estas predições de funções (Downs & Escalante-Semerena, 2000).

Em vista dos fatos expostos acima, os projetos de seqüenciamento genômico, de uma forma geral, abrem inúmeras possibilidades para novos desafios científicos. Entre eles, destacam-se os projetos genoma funcional e genoma estrutural (Lemos, 2004). No caso de projetos genomas funcionais bacterianos, seus principais objetivos residem no entendimento sobre a interação de uma bactéria com seus hospedeiros (Feezor *et al.*, 2005; Shelburne *et al.*, 2005; Virtaneva *et al.*, 2005), assim como na realização de estudos de genômica estrutural de ORFs que apresentam funções desconhecidas, e baixa ou nenhuma similaridade de seqüência com outros microrganismos, pois estas ORFs podem revelar a presença de vias metabólicas específicas do organismo estudado (Shelburne *et al.*, 2005). Além disso pode-se realizar genômica comparativa entre diferentes estirpes de um determinado microrganismo, visando entender sua predileção por determinados ambientes ou hospedeiros, sua biodiversidade genética, organização gênica e mecanismos de interação com o hospedeiro e outros microrganismos (Kao *et al.*, 2004; Feezor *et al.*, 2005; Guimil *et al.*, 2005; Hickman, Tifrea & Harwood, 2005; Shelburne *et al.*, 2005; Virtaneva *et al.*, 2005).

Em relação aos projetos proteomas bacterianos, seus principais objetivos consistem em entender qual é o conjunto de proteínas expressas a partir de um determinado genoma, em condições normais, de patogênese, ou quando submetidos a diferentes condições ambientais e/ou de estresse (Cordwell, Nouwens & Walsh, 2001; Huang, Triche & Jong, 2002; Rosen & Ron, 2002; Nouwens, Walsh & Cordwell, 2003).

6 – CONCLUSÕES

1 – A utilização das endonucleases de restrição de raro corte *I-CeuI*, *PacI* e *SwaI*, para realização de digestões simples e duplas do cromossomo do *G. diazotrophicus* PAL5 e subsequente realização da metodologia de PFGE, demonstrou-se eficaz na determinação do tamanho e morfologia do cromossomo deste endófito.

2 – A utilização de experimentos de hibridização utilizando sondas radioativas de DNA, para detecção de genes de cópia única e genes ribossomais, nos fragmentos de DNA gerados pela metodologia de PFGE, propiciou o alinhamento de 45 fragmentos de DNA oriundos de 6 diferentes perfis de PFGE, no mapa físico do cromossomo do *G. diazotrophicus* PAL5, assim como na determinação do número e ordem dos genes ribossomais, e posicionamento de marcadores moleculares neste cromossomo.

3 – A estratégia de seleção de genes de cópia única, presentes em grandes *contigs* gerados pelo Projeto Riogene, e subsequente produção de sondas radioativas de DNA destes genes, que foram hibridizadas contra membranas de PFGE, geradas a partir de digestões simples e duplas do cromossomo do *G. diazotrophicus* PAL5, auxiliou na determinação do posicionamento de marcadores moleculares neste cromossomo, e conseqüentemente, favoreceu o alinhamento de 16 *contigs* no mapa físico do cromossomo deste endófito (Montagem de maio de 2006), o que possibilitou a geração de grandes *scaffolds*, durante a fase de finalização deste projeto de seqüenciamento genômico.

4 – A utilização dos dados de mapeamento físico e genético do cromossomo do *G. diazotrophicus* PAL5, foi de suma importância na fase de finalização do seqüenciamento genômico deste microrganismo, executado pela metodologia de *shotgun*, pois através do alinhamento de 16 *contigs* (97,6% da cobertura do genoma; Montagem de maio de 2006), houve uma redução do número de experimentos de PCR combinatórios, utilizados para obtenção da seqüência genômica completa deste endófito.

5 – Através do alinhamento dos 6 *contigs* gerados na montagem de dezembro de 2006, em relação aos 16 *contigs* gerados na montagem de maio de 2006, os quais foram alinhados no mapa físico do cromossomo do *G. diazotrophicus* PAL5, foi possível confirmar o alinhamento e/ou realinhar todos os *contigs* pertencentes ao cromossomo deste endófito, pois estes *contigs* tiveram os seus buracos genômicos fechados pela metodologia de PCR combinatório.

6 – Os dados de anotação do genoma do *G. diazotrophicus* PAL5, realizado no sistema Bionotes pelos membros integrantes do Projeto Riogene, reforçam a hipótese de que dois *contigs* obtidos neste projeto de seqüenciamento genômico sejam plasmídeos, com tamanhos estimados em 38,8 e 16,6 Kb, respectivamente, pois foi observada uma alta qualidade de montagem destas seqüências, além do fato de terem sido obtidas a partir da montagem de seqüências de clones oriundos de diferentes bibliotecas genômicas e apresentarem genes compatíveis com diversos plasmídeos descritos na literatura.

7 – A partir do mapeamento físico do cromossomo do *G. diazotrophicus* PAL5 utilizando cromossomos artificiais bacterianos (BACs), foi possível confirmar o número de operons

ribossomais presentes neste genoma, e realizar a ordenação e orientação de 7 *contigs* gerados na montagem de novembro de 2004, o que forneceu subsídios para a montagem do maior *contig* gerado neste projeto de sequenciamento genômico.

8 – Na categoria funcional metabolismo energético, foi observado que o *G. diazotrophicus* PAL5, apresenta a via das pentoses-fosfato como principal via metabólica operante no catabolismo de glicose 6-fosfato, pois este endófito não possui genes relacionados com produção das enzimas fosfoglicoisomerase e fosfofrutoquinase (fase preparatória da via glicolítica), e 6-fosfogliconato desidratase e 2-ceto-3-deoxi-6-fosfogliconato aldolase (via de fermentação Entner-Doudoroff).

9 – A presença de grande número de genes relacionados à categoria funcional patogenicidade/adaptação/chaperoninas (3,9% do genoma, 143 genes), em especial genes relacionados com quimiotaxia, motilidade celular, resistência a drogas, adaptação a condições ambientais atípicas, destoxificação celular e produção de bacteriocina, ajuda a explicar a capacidade do *G. diazotrophicus* PAL5, em realizar associações benéficas com diversos hospedeiros vegetais e sugerem que este endófito seja capaz de responder a uma enorme variedade de flutuações ambientais, especialmente em competições intra e inter-específicas com microrganismos fitopatogênicos, rizosféricos e endofíticos

10 – O fato dos 19 genes inclusos no grupo de genes *nif-fix* do *G. diazotrophicus* PAL5 apresentarem maiores similaridades em relação a genes presentes no plasmídeo de simbiose do *Rhizobium sp.* (cepa NGR234) e cromossomo do *Bradyrhizobium japonicum* USDA110, além do

fato destes três microrganismos apresentarem uma organização similar dos genes *nifDKENX*, sugerem a possibilidade do *G. diazotrophicus* PAL5 ter adquirido estes genes relacionados ao processo de FBN, por meio de mecanismos de transferência horizontal de genes.

11 – No genoma do *G. diazotrophicus* PAL5 foi observado um grande número de genes (49,2% do genoma; 1802 genes), que não puderam ter as suas funções preditas e foram classificados como genes de proteínas hipotéticas (tiveram alinhamentos com valores de *e-value* superiores a 10^{-3}). A exploração dos dados relativos a estes genes, pode levar a descoberta de novas vias metabólicas intrínsecas deste endófito.

7 – REFERÊNCIAS BIBLIOGRÁFICAS

- Allardet-Servent, A.; Michaux-Charachon, S.; Jumas-Bilak, E.; Karayan, L. & Ramuz, M. Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. **J. Bacteriol.**, 175: 7869-7874, 1993.
- Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W. & Lipman, D.J. Basic local alignment search tool. **J. Mol. Biol.**, 215(3): 403-410, 1990.
- Altschul, S.F. Fundamentals of database searching. *In: Trends guide to bioinformatics*. Elsevier Science, 1988. p. 7-9.
- Alvarez, B. & Martinez-Drets. Metabolic characterization of *Acetobacter diazotrophicus*. **Can. J. Microbiol.**, 41: 918-924, 1995.
- Arrieta, J.G.; Sotolongo, M.; Menéndez, C.; Alfonso, D.; Trujillo, L.E.; Soto, M.; Ramírez, R. & Hernández, L. A Type II Protein Secretory Pathway Required for Levansucrase Secretion by *Gluconacetobacter diazotrophicus*. **J. Bacteriol.**, 186(15): 5031-5039, 2004.
- Baldani, J.I. **Ocorrência e caracterização de *Azospirillum amazonense* em comparação com outras espécies deste gênero em raízes de milho, sorgo e arroz**. 1984. Dissertação de Mestrado. Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro.
- Baldani, J.I.; Pot, B.; Kirchhof, G.; Falsen, E.; Baldani, V.L.; Olivares, F.L.; Hoste, B.; Kersters, K.; Hartmann, A.; Gillis, M. & Dobereiner, J. Emended description of *Herbaspirillum*; inclusion of [*Pseudomonas*] *rubrisubalbicans*, a milk plant pathogen, as *Herbaspirillum rubrisubalbicans* comb. nov.; and classification of a group of clinical isolates (EF group 1) as *Herbaspirillum* species 3. **Int. J. Syst. Bacteriol.** 46(3): 802-810, 1996.

- Baldani, J.I. & Baldani, V.L.D. History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. **An. Acad. Bras. Ciênc.**, 77(3): 549-579, 2005.
- Baldani, J.I.; Caruso, L.; Baldani, V.L.D.; Goi, S.R. & Döbereiner, J. Recent advances in BNF with non-legume plants. **Soil Biol. Biochem.**, 29: 911-922, 1997.
- Barea, J.M.; Pozo, M.J.; Azcón, R. & Azcón-Aguilar, C. Microbial co-operation in the rhizosphere. **J. Exp. Bot.** 56(417): 1761-1778, 2005.
- Barloy-Hubler, F.; Capela, D.; Batut, J. & Galibert, F. High-resolution physical map of the pSymb megaplasmid and comparison of the three replicons of *Sinorhizobium meliloti* strain 1021. **Cur. Microbiol.**, 41: 109-113, 2000a.
- Barloy-Hubler, F.; Capela, D.; Barnett, M.J.; Kalman, S.; Federspiel, N. A.; Long, S. R. & Galibert, F. High-resolution physical map of the *Sinorhizobium meliloti* 1021 pSyma megaplasmid. **J. Bacteriol.**, 182(4): 1185-1189, 2000b.
- Bateman, A.; Coin, L.; Durbin, R.; Finn, R.D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E.L.; Studholme, D.J.; Yeats, C. & Eddy, S.R. The Pfam protein families database. **Nucleic Acids Res.**, 32(1): D138-D141, 2004.
- Battistoni, F.; Bartels, D.; Kaiser, O.; Reamon-Buettner, S.M.; Hurek, T. & Reinhold-Hurek, B. Physical map of the *Azoarcus sp.* Strain BH72 genome based on a bacterial artificial chromosome library as a platform for genome sequencing and functional analysis. **FEMS Microbiol. Lett.**, 249: 233-240, 2005.
- Bhattacharyya, A.; Stilwagen, S.; Reznik, G.; Feil, H.; Feil, W.S.; Anderson, I.; Bernal, A.; D'Souza, M.; Ivanova, N.; Kapatral, V.; Larsen, N.; Los, T.; Lykidis, A.; Selkov J.R.E.; Walunas, T.L.; Purcell, A.; Edwards, R.A.; Hawkins, T.; Haselkorn, R.; Overbeek, R.;

- Kyripides, N.C. & Predki, P.F. Draft Sequencing and Comparative Genomics of *Xylella fastidiosa* Strains Reveal Novel Biological Insights. **Genome Res.**, 12: 1556-1563, 2002.
- Benson, D.A.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J. & Wheeler, D.L. GenBank. **Nucleic Acids Res.**, 31(1): 23-7, 2005.
- Bentley, S.D. & Parkhill, J. Comparative Genomic Structure of Prokaryotes. **Annu. Rev. Genet.**, 38: 771-791, 2004.
- Berent, L.M. & Messick, J.B. Physical map and genome sequencing survey of *Mycoplasma haemofelis* (*Haemobartonella felis*). **Infection and Immunity**, 71(6): 3657-3662, 2003.
- Biodieselbr Online. **PróAlcool – Programa Brasileiro de Álcool**. Disponível em: <http://www.biodieselbr.com/proalcool/pro-alcool.htm>. Acesso em 08 de novembro de 2006.
- Blanco, Y.; Arroyo, M.; Legaz, M.E. & Vicente, C. Isolation from *Gluconacetobacter diazotrophicus* cell walls of specific receptors for sugarcane glycoproteins, which act as recognition factors. **J. Chromatog.**, 1093: 204-211, 2005.
- Blank, W.A. & Stemke, G.W. A physical and genetic map of the *Mycoplasma hyopneumoniae* strain J genome. **Can. J. Microbiol.** 46: 832-840, 2000.
- Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.C.; Estreicher, A.; Gasteiger, E.; Martin, M.J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S. & Schneider, M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. **Nucleic Acids Res.** 31(1): 365-370, 2003.
- Branscomb, E. & Predki, P. On the high value of low standards. **J. Bacteriol.**, 184(23): 6406-6409, 2002.
- Bricker, B.J. Characterization of the three ribosomal RNA operons *rrnA*, *rrnB* and *rrnC*, from *Brucella melitensis*. **Gene.**, 255: 117-126, 2000.

- Caballero-Mellado, J. & Martinez-Romero, E. Limited genetic diversity in the endophytic sugarcane bacterium *Acetobacter diazotrophicus*. **Appl. Environ. Microbiol.**, 60: 1532-1537, 1994.
- Capela, D.; Barloy-Hubler, F.; Gatiús, M. T.; Gouzy, J. & Galibert, F. A high-density physical map of *Sinorhizobium meliloti* 1021 chromosome derived from bacterial artificial chromosome library. **Proc. Natl. Acad. Sci. USA.**, 96(16): 9357-9362, 1999.
- Capela, D.; Barloy-Hubler, F.; Gouzy, J.; Bothe, G.; Ampe, F.; Batut, J.; Boistard, P.; Becker, A.; Boutry, M.; Cadieu, E.; Dreano, S.; Gloux, S.; Godrie, T.; Goffeau, A.; Kahn, D.; Kiss, E.; Lelaure, V.; Masuy, D.; Pohl, T.; Portetelle, D.; Puhler, A.; Purnelle, B.; Ramsperger, U.; Renard, C.; Thebault, P.; Vandenbol, M.; Weidner, S. & Galibert, F. Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. **Proc. Natl. Acad. Sci. USA.**, 98(17): 9877-9882, 2001.
- Cavalcante, V.A. & Döbereiner, J. A new acid-tolerant nitrogen-fixing bacterium associated with sugarcane. **Plant Soil.**, 108: 23-31, 1988.
- Chan, J.W.Y.F. & Goodwin, P.H. A physical map of the chromosome of *Xanthomonas campestris* pv. *Phaseoli* var. *fuscans* BXP65. **FEMS Microbiol. Lett.**, 180: 85-90, 1999.
- Cochrane, G.; Aldebert, P.; Althorpe, N.; Andersson, M.; Baker, W.; Baldwin, A.; Bates, K.; Bhattacharyya, S.; Browne, P.; Van den Broek, A.; Castro, M.; Duggan, K.; Eberhardt, R.; Faruque, N.; Gamble, J.; Kanz, C.; Kulikova, T.; Lee, C.; Leinonen, R.; Lin, Q.; Lombard, V.; Lopez, R.; McHale, M.; McWilliam, H.; Mukherjee, G.; Nardone, F.; Pastor, M. P.; Sobhany, S.; Stoehr, P.; Tzouvara, K.; Vaughan, R.; Wu, D.; Zhu, W. & Apweiler, R.

- EMBL Nucleotide Sequence Database: developments in 2005. **Nucleic Acids Res.**, 34: D10-D15, 2006.
- Cordwell, S. J.; Nouwens, A. S. & Walsh, B. J. Comparative proteomics of bacterial pathogens. **Proteomics**, 1(4): 461-472, 2001.
- Corpet, F. Multiple sequence alignment with hierarchical clustering. **Nucleic Acids Res.** 16(22): 10881-10890, 1988.
- Daniels, D.; Plunkett, G.; Burland, V. & Blattner, F.R. Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. **Science**, 257: 771-778, 1992.
- Deininger, P.L. Approaches to rapid DNA sequence analysis. **Anal Biochem.**, 135(2): 247-263, 1983.
- Delcher, A.L.; Harmon, D.; Kasif, S.; White, O. & Salzberg, S.L. Improved microbial gene identification with GLIMMER. **Nucleic Acids Res.**, 27(23): 4636-4641, 1999.
- Döbereiner, J. Recent changes in concepts of plant bacteria interactions: Endophytic N₂ fixing bacteria. **Braz. Soc. Genet.**, 44: 310-313, 1992.
- Döbereiner, J.; Baldani, V.L.D. & Baldani, J.I. **Como isolar e identificar bactérias diazotróficas de plantas não leguminosas**. Brasília: EMBRAPA-SPI: Itaguaí, RJ: EMBRAPA-CNPAB, 1995, 60p.
- Döbereiner, J.; Baldani, V.L.D. & Urquiaga, S. Importance of biological dinitrogen fixation for biofuel programmes. *In*: INTERNATIONAL SYMPOSIUM ON SUSTAINABLE AGRICULTURE FOR THE TROPICS: THE ROLE OF BIOLOGICAL NITROGEN FIXATION. Angra dos Reis – Rio de Janeiro – Brazil, 1995, p. 108.

- Döbereiner, J.; Reis, V. & Lazarini, A.C. New N₂ fixing bacteria in association with cereals and sugarcane. *In*: Bothe, H., de Briyn, F.J., Newton, W.E. **Nitrogen Fixation**, Stuttgart: Gustav Fisher, 1988, p. 717-722.
- Dong, Z.; Zelmer, C.D.; Canny, M.J.; McCully, M.E.; Luit, B.; Pan, B.; Faustino, R.S.; Pierce, G.N. & Vessey, J.K. Evidence for protection of nitrogenase from O₂ by colony structure in the aerobic diazotroph *Gluconacetobacter diazotrophicus*. **Microbiology**, 148: 2293-2298, 2002.
- Downs, D.M. & Escalante-Semerena, J.C. Impact of Genomics and Genetics on the Elucidation of Bacterial Metabolism. **Methods**, 20: 47-54, 2000.
- Dudez, A.M.; Chaillou, S.; Hissler, L.; Stentz, R.; Champomier-Vergès, M.C.; Alpert, C.A. & Zagorec, M. Physical and genetic map of the *Lactobacillus sakei* 23K chromosome. **Microbiology**, 148: 421-431, 2002.
- Eckert, B.; Weber, O.B.; Kirchhof, G.; Halbritter, A.; Stoffels, M. & Hartmann, A. *Azospirillum doebereineriae* sp. nov., a new nitrogen-fixing bacterium associated with the C4-grass *Miscanthus*. **Int. J. Syst. Evolut. Microbiol.**, 51: 17-26, 2001.
- Edwards, A.; Voss, H.; Rice, P.; Civitello, A.; Stegemann, J.; Schwager, C.; Zimmermann, J.; Erfle, H.; Caskey, C.T. & Ansorge, W. Automated DNA sequencing of the human HPRT locus. **Genomics**, 6: 593-608, 1990.
- El-Osta, Y.G.A.; Hillier, A.J.; Davidson, B.E. & Dobos, M. Pulsed field gel electrophoretic analysis of the genome of *Lactobacillus gasseri* ATCC33323, and construction of a physical map. **Electrophoresis**, 23: 3321-3331, 2002.
- Ewing, B.; Hillier, L.; Wendl, M. & Green, P. Base-calling of automated sequencer traces using Phred I accuracy assessment. **Genome Res.**, 8(3): 175-185, 1998.

- Ewing, B. & Green, P. Base-calling of automated sequencer traces using Phred II error probabilities. **Genome Res.**, 8(3): 186-194, 1998.
- Falquet, L.; Pagni, M.; Bucher, P.; Hulo, N.; Sigrist, C.J.; Hofmann, K. & Bairoch, A. The PROSITE database, its status in 2002. **Nucleic Acids Res.**, 30: 235-238, 2002.
- Feezor, R.J.; Cheng, A.; Paddock, H.N.; Baker, H.V. & Moldawer, L.L. Functional genomics and gene expression profiling in sepsis: beyond class prediction. **Clin Infect Dis.**, 41(S7): S 427-435, 2005.
- Ferdows, M.S. & Barbour, A.G. Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the lyme disease agent. **Proc. Natl. Acad. Sci. USA.** 86: 5969-5973, 1989.
- Felsenstein, J. **Department of Genetics, University of Washington, PHYLIP: the PHYLogeny Inference Package.** Disponível em: <http://evolution.genetics.washington.edu/phylip.html>. Acesso em: 10 de maio de 2006.
- Fisher, K. & Newton, W.E. Nitrogenase proteins from *Gluconacetobacter diazotrophicus*, a sugarcane-colonizing bacterium. **Biochim. Biophys. Acta**, 1750: 154-165, 2005.
- Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A. & Merrick, J.M. Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd. **Science**. 269: 496-512, 1995.
- Fonstein, M. & Haselkorn, R. Physical mapping of bacterial genomes. **J. Bacteriol.**, 177(12): 3361-3369, 1995.
- Fraser, C.M.; Eisen, J.A. & Salzberg, S.L. Microbial genome sequencing. **Nature**, 406: 799-803, 2000.

- Fraser, C.M.; Eisen, J.A.; Nelson, K.E.; Paulsen, I.T. & Salzberg, S.L. The Value of Complete Microbial Genome Sequencing (You Get What You Pay For). **J. Bacteriol.**, 184(23): 6403-6405, 2002.
- Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischman, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G. & Kelley, J.M. The minimal gene complement of *Mycoplasma genitalium*. **Science**, 270: 397-403, 1995.
- Freiberg, C.; Fellay, R.; Bairoch, A.; Broughton, W.J.; Rosenthal, A. & Perret, X. Molecular basis of symbiosis between Rhizobium and legumes. **Nature.**, 387(6631): 394-401, 1997.
- Fuentes-Ramírez, L.E.; Caballero-Mellado, J.; Sepúlveda, J. & Martínez-Romero, E. Colonization of sugarcane by *Acetobacter diazotrophicus* is inhibited by high N-fertilization. **FEMS Microbiol. Ecol.**, 29: 117-128, 1999.
- Fuentes-Ramírez, L.E.; Bustillos-Cristales, R.; Tapia-Hernandez, A.; Jimenez-Salgado, T.; Wang, E.T.; Martinez-Romero, E. & Caballero-Mellado, J. Novel nitrogen-fixing acetic acid bacteria, *Gluconacetobacter johannae* sp. nov. and *Gluconacetobacter azotocaptans* sp. nov. associated with coffee plants. **Int. J. Syst. Evolut. Microbiol.** 51: 1305-1314, 2001.
- Gastal, F. & Lemaire, G. N uptake and distribution in crops: an agronomical and ecophysiological perspective. **J. Exp. Bot.**, 53(370): 789-799, 2002.
- Gillis, M.; Kersters, K.; Hoste, B.; Janssens, D.; Kroppenstedt, R.M.; Stephan, M.P.; Teixeira, K.R.S.; Döbereiner, J. & Deley, J. *Acetobacter diazotrophicus* sp. Nov., a nitrogen-fixing acetic acid bacterium associated with sugarcane. **Int. J. Syst. Bacteriol.** 39: 361-364, 1989.
- GOLD. **Genomes OnLine Database v2.0**. Disponível em: <http://www.genomesonline.org>. Acesso em 10 de maio de 2006.

- González, J.; Nefedov, M.; Bosdet, I.; Casals, F.; Calvete, O.; Delprat, A.; Shin, H.; Chiu, R.; Mathewson, C.; Wye, N.; Hoskins, R.A.; Schein, J.E.; Jong, P. & Ruiz, A. **A BAC-based physical map of the *Drosophila buzzatii* genome.** *Genome Research*, 15: 885-892, 2005.
- Gonzalez, V.; Santamaria, R. I.; Bustos, P.; Hernandez-Gonzalez, I.; Medrano-Soto, A.; Moreno-Hagelsieb, G.; Janga, S. C.; Ramirez, M. A.; Jimenez-Jacinto, V.; Collado-Vides, J. & Davila, G. The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons. **Proc. Natl. Acad. Sci. U.S.A. In press**
- Gordon, D.; Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. **Genome Res.**, 8(3): 195-202, 1998.
- Gordon, D.; Desmarais, C. & Green, P. Automated finishing with autofinish. **Genome Res.** 11(4): 614-625, 2001.
- Goto, T.; Todo, K.; Miyamoto, K. & Akimoto, S. Bacterial artificial chromosome library of *Fingoldia magna* ATCC 29328 for genetic mapping and comparative genomics. **Microbiol Immunol.**, 47(12): 1005-1016, 2003.
- Guimil, S.; Chang, H. S.; Zhu, T.; Sesma, A.; Osbourn, A.; Roux, C.; Ioannidis, V.; Oakeley, E.J.; Docquier, M.; Descombes, P.; Briggs, S.P. & Paszkowski, U. Comparative transcriptomics of rice reveals an ancient pattern of response to microbial colonization. **Proc. Natl. Acad. Sci. USA.**, 102(22): 8066-8070, 2005.
- Harger, C.; Skupski, M.; Bingham, J.; Farmer, A.; Hoisie, S.; Hraber, P.; Kiphart, D.; Krakowski, L.; McLeod, M.; Schwertfeger, J.; Seluja, G.; Siepel, A.; Singh, G.; Stamper, D.; Steadman, P.; Thayer, N.; Thompson, R.; Wargo, P.; Waugh, M.; Zhuang, J.J. & Schad, P.A. The Genome Sequence DataBase (GSDB): improving data quality and data access. **Nucleic Acids Res.**, 26: 21-26, 1998.

- Hancock, J.M. Genome size and the accumulation of simple sequence repeats: implications of new data from genome sequencing projects. **Genetica**, 115: 93-103, 2002.
- Heber, S.; Hoheisel, J. & Vingron, M. Application of bootstrap techniques to physical mapping. **Genomics**, 69: 235-241, 2000.
- Hesekorn, R. Excision elements interrupt nitrogen fixation operons in cyanobacteria. *In*: Berg, D.E. & Howe, M.M. **Mobile DNA**. American Society for Microbiology. Washington, D.C., 1989. p. 735-742.
- Hickman, J. W.; Tifrea, D. F. & Harwood, C. S. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. **Proc. Natl. Acad. Sci. USA.**, 102(40):14422-14427, 2005.
- Huang, S. H.; Triche, T. & Jong, A. Y. Infectomics: genomics and proteomics of microbial infections. **Funct. Integr. Genomics**. 1(6): 331-344, 2002.
- Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. **Genome Res.**, 9(9): 868-877, 1999.
- James, E.K.; Reis, V.M.; Olivares, F.L.; Baldani, J.I. & Döbereiner, J. Infection of sugarcane by the nitrogen-fixing bacterium *Acetobacter diazotrophicus*. **J. Exp. Bot.**, 45: 757-766, 1994.
- James, E.K.; Olivares, F.L.; Oliveira, A.L.M.; Reis, J.R.F.B.; Silva, L.G. & Reis, V.M. Further observations on the interaction between sugar cane and *Gluconacetobacter diazotrophicus* under laboratory and greenhouse conditions. **J. Exp. Bot.**, 52: 747-760, 2001.
- Jeuffroy, M.H.; Ney, B. & Ourry, A. Integrated physiological and agronomic modelling of N capture and use within the plant. **J. Exp. Bot.**, 53(370): 809-823, 2002.
- Jones, D.T. **Threader – Protein fold recognition by optimal protein sequence threading**. Disponível em <http://bioinf.cs.ucl.ac.uk/threader/>. Acesso em 10 de maio de 2006.

- Kaiser, B.N.; Gridley, K.L.; Brady, J.N.; Phillips, T. & Tyerman, S.D. The Role of Molybdenum in Agricultural Plant Production. **Ann. Botany.**, 96: 745-754, 2005..
- Kaneko, T.; Nakamura, Y.; Sato, S.; Asamizu, E.; Kato, T.; Sasamoto, S.; Watanabe, A.; Idesawa, K.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kiyokawa, C.; Kohara, M.; Matsumoto, M.; Matsuno, A.; Mochizuki, Y.; Nakayama, S.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takeuchi, C.; Yamada, M. & Tabata, S. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. **DNA Res.**, 7(6): 331-338, 2000.
- 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 Res.**, 9(6):189-197, 2002.
- Kao, K.C.; Yang, Y.L.; Boscolo, R.; Sabatti, C.; Roychowdhury, V. & Liao, J. C. Transcriptome-based determination of multiple transcription regulator activities in *Escherichia coli* by using network component analysis. **Proc. Natl. Acad. Sci. USA.** 101(2): 641-646, 2004.
- Kirchhof, G.; Eckert, B.; Stoffels, M.; Baldani, J.I.; Reis, V.M. & Hartmann, A. *Herbaspirillum frisingense* sp. nov., a new nitrogen-fixing bacterial species that occurs in C4-fiber plants. **Int. J. Syst. Evolut. Microbiol.**, 51: 157-168, 2001.
- Koonin, E.V.; Mushegian, A. R. & Rudd, K.E. Sequencing and analysis of bacterial genomes. **Curr. Biology.**, 6(4): 404-416, 1996.
- Krause, A.; Ramakumar, A.; Bartels, D; Battistoni, F.; Bekel, T.; Boch, J.; Böhm, M.; Friedrich, F.; Hurek, T.; Krause, L.; Linke, B.; McHardy, A.C.; Sarkar, A.; Schneiker, S.; Syed, A.A.; Thauer, R.; Vorhölter, F.J.; Weidner, S.; Puhler, A.; Reinhold-Hurek, B.; Kaiser, O. &

- Goesmann, A. Complete genome of the mutualistic N₂-fixing grass endophyte *Azoarcus sp.* strain BH72. **Nat. Biotech.**, 24(11): 1385-1391, 2006.
- Kuepel, P.L.; Piletier, A.J. & Hill, T.M. *Tus* and the terminators: inhibition of replication in *Escherichia coli*. in: Drlica, K. & Riley, M. **The Bacterial Chromosome**. American Society for Microbiology. Washington, D.C., 1990, p. 297-306.
- Kunst, F.; Vassarotti, A. & Danchin, A. Organization of the European *Bacillus subtilis* genome sequencing project. **Microbiology.**, 141: 249-255, 1995.
- Kyrpides, N. Genomes OnLine Database (GOLD): a monitor of complete and ongoing genome projects world wide. **Bioinformatics.**, 15: 773-774, 1999.
- Lai, Z.; Jing, J.; Aston, C.; Clarke, V.; Apodaca, J.; Dimalanta, E.T.; Carucci, D.J.; Gardner, M.J.; Mishra, B.; Anantharaman, T.S.; Paxia, S.; Hoffman S.L.; Venter, C.; Huff, E.J. & Schwartz, D.C. A shotgun optical map of the entire *Plasmodium falciparum* genome. **Nat. Genet.**, 23(3): 309- 313, 1999.
- Leblond, P.; Demuyter, P.; Simomnet, J.M. & Decaris, B. Genetic instability and associated genome plasticity in *Streptomyces ambofaciens*: pulsed field gel electrophoresis evidence for large DNA alterations limited genomic region. **J. Bacteriol.**, 173: 3422-3429, 1991.
- Lee, S.; Reth, A.; Meletzus, D.; Sevilla, M. & Kennedy, C. Characterization of a Major Cluster of *nif*, *fix*, and Associated Genes in a Sugarcane Endophyte, *Acetobacter diazotrophicus*. **J. Bacteriol.** 182(24): 7088-7091, 2000.
- Lee, S.; Flores-Encarnación, M.; Contreras-Zentella, M.; Garcia-Flores, L.; Escamilla, J.E. & Kennedy, C. Indole-3-Acetic Acid Biosynthesis Is Deficient in *Gluconacetobacter diazotrophicus* Strains with Mutations in Cytochrome *c* Biogenesis Genes. **J. Bacteriol.**, 186(16): 5384-5391, 2004.

- Lemos, M. **Workflow para bioinformática**. 2004. 224p. Tese de Doutorado em Ciências da Computação. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- Lemos, M.; Seibel, L.F.B. & Casanova, M.A. 2003. **Sistema de Anotações em Biossequências**. Monografia da ciência da computação número 04/03, Departamento de Informática. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- Letovsky, S. I.; Cottingham, R. W.; Porter, C. J. & Li, P. W. GDB: the Human Genome Database. **Nucleic Acids Res.**, 26(1): 94-99, 1998.
- Lima, E.; Boddey, R.M. & Döbereiner, J. Quantification of biological nitrogen fixation associated with sugarcane using a ¹⁵N aided nitrogen balance. **Soil Biol. Biochem.**, 19: 165-170, 1987.
- Llamas, I.; Sánchez, M.J.; Argandoña, M.; Béjar, V.; Quesada, E. & Moral, A.D. Analysis of the genome of the moderate halophile *Halomonas eurihalina*. **Curr. Microbiol.**, 45: 233-239, 2002.
- Loiret, F.G.; Ortega, E.; Kleiner, D.; Ortega-Rodés, P.; Rodés, R. & Dong, Z. A putative new endophytic nitrogen-fixing bacterium *Pantoea sp.* from sugarcane. **J. App. Microbiol.**, 97: 504-511, 2004.
- Loureiro, M. M.; Gomes, U. A.; Taketani, R. G.; Martins, O. B. & Ferreira, P. C. G. Physical Mapping of *Gluconacetobacter diazotrophicus* PAL5 genome. In. XXXII REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE BIOQUÍMICA E BIOLOGIA MOLECULAR – SBBQ. 2003. Caxambú – MG. Anais da XXXII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular – SBBq.
- Loureiro, M.M.; Turque, A.S.; Santos, M.B.Q.; França, L.M.; Silveira, C.B.; Pádua, V.L.M.; Martins, O.B. & Ferreira, P.C.G. Physical mapping of *Gluconacetobacter diazotrophicus*

- PAL5 genome using macrorestriction analysis and bacterial artificial cromosomes *In*. XXXIV REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE BIOQUÍMICA E BIOLOGIA MOLECULAR – SBBQ. 2005. Águas de Lindóia – SP. Anais da XXXIV Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular – Sbbq.
- Lowe, T.M. & Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. **Nucleic Acids Res.** 25(5): 955-964, 1997.
- Luciński, R.; Polcyn, W. & Ratajczak, L. Nitrate reduction and nitrogen fixation in symbiotic association *Rhizobium* - legumes. **Acta Bioch. Pol.**, 49(2): 537-546, 2002.
- Luna, M.F.; Bernardelli, C.E.; Mignone, C.F. & Boiardi, J.L. Energy Generation by Extracellular Aldose Oxidation in N₂-Fixing *Gluconacetobacter diazotrophicus*. **Appl. Environ. Microbiol.** 68(4): 2054-2056, 2002.
- Luna, M.F.; Bernardelli, C.E.; Galar, M.L. & Boiardi, J.L. Glucose Metabolism in Batch and Continuous Cultures of *Gluconacetobacter diazotrophicus* PAL 3. **Curr. Microbiol.** 52: 163-168, 2006.
- Luo, M.; Wang, Y.H.; Frisch, D.; Joobeur, T.; Wing, R.A. & Dean, R.A. Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon Fusarium wilt (*Fom-2*). **Genome.** 44: 154-162, 2001.
- Majumder, R.; Sengupta, S.; Khetawat, G.; Bhadra, R.K; Roychoudhury, S. & Das, J. Physical map of the genome of *Vibrio cholerae* 569B and localization of genetic markers. **J. Bacteriol.**, 178: 1105-1112, 1996.
- Marcone, C. & Seemüller, E. A chromosome map of the European stone fruit yellows phytoplasma. **Microbiology.**, 147: 1213-1221, 2001.

- Matsunaga, T.; Okamura, Y.; Fukuda, Y.; Wahyudi, A.T.; Murase, Y. & Takeyama, H. Complete Genome Sequence of the Facultative Anaerobic Magnetotactic Bacterium *Magnetospirillum* sp. strain AMB-1. **DNA Res.**, 12: 157-166, 2005.
- Melkerson-Watson, L.J.; Rode, C.K.; Zhang, L.; Foxman, B. & Bloch, C.A. Integrated genomic map from Uropathogenic *Escherichia coli* J96. **Infect. Immun.**, 68: 5933-5942, 2000.
- Miflin, B.J. & Habash, D.Z. The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. **J. Exp. Bot.**, 53(370): 979-987, 2002.
- Moszer, I. The complete genome of *Bacillus subtilis*: from sequence annotation to data management and analysis. **FEBS Lett.**, 430: 28-36, 1998.
- Munõz-Rojas, J. & Caballero-Mellado, J. Population Dynamics of *Gluconacetobacter diazotrophicus* in Sugarcane Cultivars and Its Effect on Plant Growth. **Microb. Ecol.**, 46: 454-464, 2003.
- Munõz-Rojas, J.; Fuentes-Ramírez, L.E. & Caballero-Mellado, J. Antagonism among *Gluconacetobacter diazotrophicus* strains in culture media and in endophytic association. **FEMS Microbiol. Ecol.**, 54: 57-66, 2005.
- Muthukumarasamy, R.; Revathi, G. & Vadivelu, M. 2000. Antagonistic potential of N₂-fixing *Acetobacter diazotrophicus* against *Colletotrichum falcatum* Went: a causal organism of red-rot of sugarcane. **Curr. Science.**, 78: 1063-1065, 2005.
- Muthukumarasamy, R.; Rebatí, G.; Seshadri, S. & Lakshminarasimhan, C. *Gluconacetobacter diazotrophicus* (syn. *Acetobacter diazotrophicus*), a promising diazotrophic endophyte in tropics. **Curr. Science.**, 83(2): 137-145, 2002.

Muthukumarasamy, R.; Govindarajan, M.; Vadivelu, M. & Revathi, G. N-fertilizer saving by the inoculation of *Gluconacetobacter diazotrophicus* and *Herbaspirillum sp.* in micropropagated sugarcane plants. **Microbiol. Res.**, 161: 238-245, 2006.

NCBI – National Center for Biotechnology Information. **Open Reading Frame Finder**. Disponível em: <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>. Acesso em 10 de maio de 2006a.

NCBI – National Center for Biotechnology Information. **The Blast Databases**. Disponível em: <ftp://ftp.ncbi.nih.gov/blast/db/>. Acesso em 10 de maio de 2006b.

Nelson, K.E.; Weinel, C.; Paulsen, I.T.; Dodson, R.J.; Hilbert, H.; Santos, V.A.P.M.; Fouts, D.E.; Gill, S.R.; Pop, M.; Holmes, M.; Brinkac, L.; Beanan, M.; DeBoy, R.T.; Daugherty, S.; Kolonay, J.; Madupu, R.; Nelson, W.; White, O.; Peterson, J.; Khouri, H.; Hance, I.; Lee, P.C.; Holtzapple, E.; Scanlan, D.; Tran, K.; Moazzez, A.; Utterback, T.; Rizzo, M.; Lee, K.; Kosack, D.; Moestl, D.; Wedler, H.; Lauber, J.; Stjepandic, D.; Hoheisel, J.; Straetz, M.; Heim, S.; Kiewitz, C.; Eisen, J.; Timmis, K.N.; Dusterhöft, A.; Tümmeler, B. & Fraser, C.M. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. **Environm. Microbiol.**, 4(12): 799-808, 2002.

Nierman, W.C.; Feldblyum, T.V.; Laub, M.T.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Alley, M.R.K.; Ohta, N.; Maddocki, J.R.; Potocka, I.; Nelson, W.C.; Newton, A.; Stephens, C.; Phadkei, N.D.; Ely, B.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.F.; Smit, J.; Craven, M.B.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf, A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Venter, C.; Shapiro, L. & Fraser, C.N. Complete genome sequence of *Caulobacter crescentus*. **Proc. Natl. Acad. Sci. USA**. 98(7): 4136-4141, 2001.

- Nogueira, E.M.; Vinagre, F.; Masuda, H.P.; Vargas, C.; Pádua, V.L.M.; Silva, F.R.; Santos, R.V.; Baldani, J.I.; Ferreira, P.C.G. & Hemerly, A.S. Expression of sugar cane genes induced by inoculation with *Gluconacetobacter diazotrophicus* and *Herbaspirillum rubrisubalbicans*. **Genet. Mol. Biol.**, 24: 199-206, 2001.
- Nouwens, A.S.; Walsh, B.J. & Cordwell, S.J. Application of proteomics to *Pseudomonas aeruginosa*. **Adv. Biochem. Eng. Biotechnol.**, 83: 117-140, 2003.
- Oana, K.; Okimura, Y.; Kawakami, Y.; Hayashida, N.; Shimosaka, M.; Okazaki, M.; Hayashi, T. & Ohnishi, M. Physical and genetic map of *Enterococcus faecium* ATCC19434 and demonstration of intra and interspecific genomic diversity in enterococci. **FEMS Microbiol. Lett.**, 207: 133-139, 2002.
- Oldroyd, G.E.D.; Harrison, M.J. & Udvardi, M. Peace Talks and Trade Deals. Keys to Long-Term Harmony in Legume-Microbe Symbioses. **Plant Physiol.**, 137: 1205-1210, 2005.
- Olivares, F.L.; Reis, V.M. & Façanha, A.R. The role of endophytic diazotrops in sugarcane root morphogenesis and development. *In*: Finan, T. M.; O'brian, M.R.; Layzell, D.B.; Vessey, J.K. & Newton, W. **Nitrogen fixation: Global perspectives**. Oxon: CAB International. 2002, p. 476-477.
- Padovan, A.C.; Firrao, G.; Schneider, B. & Gibb, K.S. Chromosome mapping of the sweet potato little leaf phytoplasma reveals genome heterogeneity within the phytoplasmas. **Microbiology.**, 146: 893-902, 2000.
- Peoples, M.B. & Craswell, E.T. Biological nitrogen fixation: investments, expectations and actual contributions to agriculture. **Plant Soil**, 141: 13-39, 1992.

- Perlova, O.; Ureta, A.; Nordlund, S. & Meletzus, D. Identification of Three Genes Encoding PII-Like Proteins in *Gluconacetobacter diazotrophicus*: Studies of Their Role(s) in the Control of Nitrogen Fixation. **J. Bacteriol.**, 185(19): 5854-5861, 2003.
- Piñon, D.; Casas, M.; Blanch, M.; Fontaniella, B.; Blanco, Y.; Vicente, C.; Solas, M.T. & Legaz, M.E. *Gluconacetobacter diazotrophicus*, a sugar cane endosymbiont, produces a bacteriocin against *Xanthomonas albilineans*, a sugar cane pathogen. **Res. Microbiol.** 53: 345-351, 2002.
- Prober, J.M.; Trainor, G.L.; Dam, R.J.; Hobbs, F.W.; Robertson, C.W.; Zagursky, R.J.; Cocuzza, A.J.; Jensen, M.A. & Baumeister, K. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. **Science**, 238: 336-341, 1987.
- Projeto Riogene. **Riogene - Virtual Institute of Genomic Research**. Disponível em <http://www.riogene.lncc.br>. Acesso em: 10 de maio de 2006.
- Prust, C.; Hoffmeister, M.; Liesegang, H.; Wiezer, A.; Fricke, W.F.; Ehrenreich, A.; Gottschalk, G. & Deppenmeier, U. Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*. **Nat Biotechnol.**, 23(2): 195-200, 2005.
- Rabus, R.; Kube, M.; Heider, J.; Beck, A.; Heitmann, K.; Widdel, F. & Reinhardt, R. The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. **Arch. Microbiol.**, 183: 27-36, 2005.
- Ramos-Díaz, M.A. & Ramos, J.L. Combined Physical and genetic map of the *Pseudomonas putida* KT2440 chromosome. **J. Bacteriol.**, 180: 6352-6363, (1998).
- Raven, J.A.; Handley, L.L. & Andrews, M. Global aspects of C/N interactions determining plant-environment interactions. **J. Exp. Bot.**, 55(394): 11-25, 2004.

- Reis, V.M. & Döbereiner, J. Effect of high sugar concentration on nitrogenase activity of *Acetobacter diazotrophicus*. **Arch. Microbiol.**, 171: 13-18, 1998.
- Reis, V.M.; Olivares, F.L. & Döbereiner, J. Improved methodology for isolation and identification of *Acetobacter diazotrophicus* and confirmation of its endophytic habitat. **World J. Microbiol. Biotechnol.**, 10: 401-405, 1994.
- Reis, V.M.; Estrada-de los Santos, P.; Tenorio-Salgado, S.; Vogel, J.; Stoffels, M.; Guyon, S.; Mavingui, P.; Baldani, V.L.; Schmid, M.; Baldani, J.I.; Balandreau, J.; Hartmann, A. & Caballero-Mellado, J. *Burkholderia tropica* sp. nov., a novel nitrogen-fixing, plant-associated bacterium. **Int. J. Syst. Evol. Microbiol.**, 54(6): 2155-2162, 2004.
- Riley, M. & Sanderson, E. Comparative genetics of *Escherichia coli* and *Salmonella typhimurium*. In: Drlica, K. & Riley, M. **The bacterial chromosome**. American Society for Microbiology. Washington, D.C. 1990, p. 85-95.
- Römling, U.; Grothues, D.; Bautsch, W. & Tümmler, B. A physical genome map of *Pseudomonas aeruginosa* PAO. **EMBO J.**, 8: 4081-4089, 1989.
- Rosen, R. & Ron, E. Z. Proteome analysis in the study of the bacterial heat-shock response. **Mass. Spectrom. Rev.** 21(4): 244-265, 2002.
- Rutherford, K.; Parkhill, J.; Crook, J.; Horsnell, T.; Rice, P.; Rajandream, M. A. & Barrell, B. Artemis: sequence visualization and annotation. **Bioinformatics**, 16(10): 944-945, 2000.
- Sali, A. **MODELLER – A Program for protein structure modeling**. Disponível em <http://salilab.org/modeller/manual/manual.html>. Acesso em 10 de maio de 2006.
- Sambrook, J. and Russell, D.W. **Molecular cloning: a laboratory manual**. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 2001.

- Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. **Proc. Natl. Acad. Sci. USA.**, 74: 5463-5467, 1977.
- Sanger, F.; Coulson, A.R.; Barrell, B.G.; Smith, A.J. & Roe, B.A. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. **J. Mol. Biol.**, 143(2): 161-178, 1980.
- Sanger Institute. **Informatics Analysis Software: Alfresco, November 2003**. Disponível em <http://www.sanger.ac.uk/Software/Alfresco>. Acesso em 10 de maio de 2006.
- Santos, S.T.; Santos, A.M.; Reis, V.M.; Teixeira, K.R.S. & Baldani, J.I. PCR/RFLP analysis of *Acetobacter diazotrophicus* strains isolated from sugarcane genotypes originated from different countries. *In*: INTERNATIONAL CONGRESS ON NITROGEN FIXATION, 12. Foz de Iguaçu, PR, Brazil, 1999, p.611.
- Santos, M.B.Q. **Montagem e finalização do genoma de *Gluconacetobacter diazotrophicus* PAL5**. 2005, 105 p. Dissertação de Mestrado em Química Biológica. Rio de Janeiro. Universidade Federal do Rio de Janeiro-CCS-Instituto de Bioquímica Médica. Rio de Janeiro.
- Santos, M.B.Q.; Loureiro, M.M.; Turque, A.S.; França, L.M.; Pádua, V.L.M; Martins, O.B.; Ferreira, P.C.G. & Riogene Genomic Sequencing Consortium. *Gluconacetobacter diazotrophicus* PAL5: a progress report. *In*. XXXIII REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE BIOQUÍMICA E BIOLOGIA MOLECULAR – SBBQ. Caxambú - MG. 2004. Anais da XXXIII Reunião da Sociedade Brasileira de Bioquímica e Biologia Molecular – SBBq.
- Santos, P.E.L.; Bustillos-Cristales, R. & Caballero-Mellado, J. *Burkholderia*, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. **Appl. Environ. Microbiol.**, 67: 2790-2798, 2001.

- Schwartz, I. Microbial Genomics: From Sequence to Function. **Genomics**. 6(5): 493-495, 2000.
- Sevilla, M.; Meletzus, D.; Teixeira, K.R.S.; Lee, S.; Nutakki, A.; Baldani, J.I & Kennedy, C. Analysis of *nif* and regulatory genes in *Acetobacter diazotrophicus*. **Soil Biol. Biochem.** 29: 871-874, 1997.
- Sevilla, M.; Burris, R.H.; Guanapala, N. & Kennedy, C. Comparison of benefit to sugarcane plant growth and $^{15}\text{N}_2$ incorporation following inoculation of sterile plants with *Acetobacter diazotrophicus* wild type and *nif*-mutant strains. **Mol. Plant-Microbe Interact.** 14: 358-366, 2001.
- Shelburne, S.A.; Sumbly, P.; Sitkiewicz, I.; Granville, C.; DeLeo, F.R. & Musser, J.M. Central role of a bacterial two-component gene regulatory system of previously unknown function in pathogen persistence in human saliva. **Proc. Natl. Acad. Sci. USA.**, 102(44): 16037-16042, 2005.
- Shimizu, T.; Ohshima, S.; Ohtani, K; Shimizu, T. & Hayashi, H. Genomic map of *Clostridium perfringens* Strain 13. **Microbiol. Immunol.**, 45: 179-189, 2001.
- Sistema Bionotes. ***Gluconacetobacter diazotrophicus* community**. Disponível em <http://www.bionotes.com.br>. Acessado em 10 de maio de 2006a.
- Sistema Bionotes. **Riogene: Anotação do genoma do *Gluconacetobacter diazotrophicus***. Disponível em: <http://www.bionotes.com.br>. Acessado em 10 de maio de 2006b.
- Smith, L.M.; Sanders, J.Z.; Kaiser, R.J.; Hughes, P.; Dodd, C.; Connell, C.R.; Heiner, C.; Kent, S.B.H. & Hood, L.E. Fluorescence detection in automated DNA sequence analysis. **Nature**, 321: 674-679, 1986.

- Stavrínides, J. & Guttman, D.S. Nucleotide Sequence and Evolution of the Five-Plasmid Complement of the Phytopathogen *Pseudomonas syringae* pv. *maculicola* ES4326. **J. Bacteriol.**, 186(15): 5101-5115, 2004.
- Steenhoudt, O. & Vanderleyden, J. *Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. **FEMS Microbiol.**, 24: 487-506, 2000.
- Suzuki, K.; Hattori, Y.; Uraji, M.; Ohta, N.; Iwata, K.; Murata, K.; Kato, A. & Yoshida, K. Complete nucleotide sequence of a plant tumor-inducing Ti plasmid. **Gene**, 242: 331-336, 2000.
- Taketani, R.G. **Mapeamento físico do genoma de *Gluconacetobacter diazotrophicus* PAL5.** 2001, 61 p. Monografia de Bacharelado em Microbiologia e Imunologia. Rio de Janeiro. Universidade Federal do Rio de Janeiro – CCS – Instituto Microbiologia Professor Paulo de Góes. Rio de Janeiro.
- Tao, Q.; Chang, Y.L.; Wang, J.; Chen, H.; Islam-Faridi, M.N.; Scheuring, C.; Wang, B.; Stelly, D.M. & Zhang, H.B. Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. **Genetics**. 158: 1711-1724, 2001.
- Teixeira, K.R.S.; Wulling, M.; Morgan, T.; Galler, R.; Zellermann, E.M.; Baldani, J.I.; Kennedy, C. & Meletzus, D. Molecular analysis of the chromosomal region encoding the *nifA* and *nifB* genes of *Acetobacter diazotrophicus*. **FEMS Microbiol. Lett.**, 176(2): 301-309, 1999.
- Tejera, N.A.; Ortega, E.; Gonzalez-Lopez, J. & Lluch, C. Effect of some abiotic factors on the biological activity of *Gluconacetobacter diazotrophicus*. **J. App. Microbiol.** 95: 528-535, 2003.

Thieme, F.; Koebnik, R.; Bekel, T.; Berger, C.; Boch, J.; Büttner, D.; Caldana, C.; Gaigalat, L.; Goesmann, A.; Kay, S.; Kirchner, O.; Lanz, C.; Linke, B.; McHardy, A.C.; Meyer, F.; Mittenhuber, G.; Nies, D.H.; Niesbach-Klösgen, U.; Patschkowski, T.; Rückert, C.; Rupp, O.; Schneiker, S.; Schuster, S.C.; Vorhölter, F.J.; Weber, E.; Pühler, A.; Bonas, U.; Bartels, D. & Kaiser, O. Insights into Genome Plasticity and Pathogenicity of the Plant Pathogenic Bacterium *Xanthomonas campestris* pv. vesicatoria Revealed by the Complete Genome Sequence. **J. Bacteriol.**, 187(21): 7254-7266, 2005.

TIGR – The Institute for Genomic Research. **RBSFinder**. Disponível em <http://www.tigr.org/software/genefinding.shtml>. Acesso em 10 de maio de 2006.

Tomkins, J.P.; Wood, T. C.; Stacey, M.G.; Loh, J.T.; Judd, A.; Goicoechea, J.L.; Stacey, G.; Sadowsky, M.J. & Wing, R.A. A marker-dense physical map of the *Bradyrhizobium japonicum* genome. **Genome Res.** 11(8): 1434-1440, 2001.

Ureta, A. & Nordlund, S. Evidence for Conformational Protection of Nitrogenase against Oxygen in *Gluconacetobacter diazotrophicus* by a Putative FeSII Protein. **J. Bacteriol.**, 184(20): 5805-5809, 2002.

Urquiaga, S.; Cruz, H.S. & Boddey, R.M. Contribution of nitrogen fixation to sugarcane: nitrogen-15 and nitrogen-balance estimates. **Soil Sc. Soc. AM. J.**, 56: 105-114, 1992.

UWGC – University of Washington Genome Center. **Crossmatch – Sequence alignment tool**. Disponível em <http://www.genome.washington.edu/UWGC/analysistools/Swat.cfm>. Acesso em 10 de maio de 2006.

Vargas, C.; Pádua V.L.M.; Nogueira, E.M; Vinagre, F.; Masuda, H.P.; Silva, F.R.; Baldani, J.I.; Ferreira, P.C.G. & Hemerly, A.S. Signaling pathways mediating the association between

- sugarcane and endophytic diazotrophic bacteria: a genomic approach. **Symbiosis**, 35: 159-180, 2003.
- Villiers, E.P.; Brayton, K.A.; Zweygarth, E. & Allsopp, B.A. Genome size and genetic map of *Cowdria ruminantium*. **Microbiology**, 146: 2627-2634, 2000.
- Virtaneva, K.; Porcella, S. F.; Graham, M. R.; Ireland, R. M.; Johnson, C. A.; Ricklefs, S. M.; Babar, I.; Parkins, L.D.; Romero, R.A.; Corn, G.J.; Gardner, D.J.; Bailey, J.R.; Parnell, M.J. & Musser, J.M. Longitudinal analysis of the group A Streptococcus transcriptome in experimental pharyngitis in cynomolgus macaques. **Proc. Natl. Acad. Sci. USA**, 102(25): 9014-9019, 2005.
- Waterson, R.H.; Lander, E.S. & Sulston, J.E. On the sequencing of the human genome. **Proc. Natl. Acad. Sci. USA**. 99(6): 3712-3716, 2002.
- Weinel, C.; Tümmler, B.; Hilbert, H.; Nelson, K.E. & Kiewitz, C. General method of rapid Smith/Birnstiel mapping adds for gap closure in shotgun microbial genome sequencing projects: application to *Pseudomonas putida* KT2440. **Nucleic Acids Res.** 29: e110, 2001.
- Westbrook, J.; Feng, Z.; Jain, S.; Bhat, T. N.; Thanki, N.; Ravichandran, V.; Gilliland, G. L.; Bluhm, W.; Weissig, H.; Greer, D. S.; Bourne, P. E. & Berman, H. M. The protein data bank: unifying the archive. **Nucleic Acids Res.**, 30(1): 245-248, 2002.
- Widjaja, R.; Suwanto, A. & Tjahjono, B. Genome size and macrorestriction map of *Xanthomonas campestris* pv. *Glycines* YR32 chromosome. **FEMS Microbiol. Lett.**, 175: 59-68, 1999.
- Wu, C.H.; Huang, H.; Arminski, L.; Castro-Alvear, J.; Chen, Y.; Hu, Z.Z.; Ledley, R.S.; Lewis, K.C.; Mewes, H.W.; Orcutt, B.C.; Suzek, B.E.; Tsugita, A., Vinayaka, C.R.; Yeh, L.S.; Zhang, J. & Barker, W.C. The protein information resource: an integrated public resource of functional annotation of proteins. **Nucleic Acids Res.**, 30(1): 35-37, 2002.

WU-BLAST – **Washington University BLAST Archives-BLAST Manual**. Disponível em <http://blast.wustl.edu/doc/blast1.pdf>. Acesso em 10 de maio de 2006.

Yamada, Y.; Hoshino, K. & Ishikawa, T. The phylogeny of acetic acid bacteria based on the partial sequences of 16S ribosomal RNA: the elevation of the subgenus *Gluconacetobacter* to the generic level. **Bioscience Biotechnology Biochemistry**, 61: 1244-1251, 1997.

Zé-Zé, L.; Tenreiro, R. & Paveia, H. The *Oenococcus oeni* genome: physical and genetic mapping of strain GM and comparison with the genome of a divergent strain, PSU-1. **Microbiology**, 146: 3195-3204, 2000.

8 – ANEXOS



Paulo Cavalcanti Gomes Ferreira
Instituto de Bioquímica Médica
Universidade Federal do Rio de Janeiro
Prédio do CCS / Bl. D -subsolo – sala 05
Cid. Universitária – Ilha do Fundão
Rio de Janeiro - RJ - 21.941-590 - Brasil
Tel. +55 21 2562-6751
Tel. Secretaria. +55 21 2562-6789
Fax. +55-21-2270-8647
e-mail: paulof@bioqmed.ufrj.br

April 09 th, 2007.

Dear Sir

Please find enclosed the manuscript entitled “Physical and Genetic Map of the *Gluconacetobacter diazotrophicus* PAL5 chromosome”, by Marcio Martins Loureiro, Marcelo Bertalan, Aline da Silva Turque, Leonardo Mega França, Vânia Lúcia Muniz de Pádua, José Ivo Baldani, Orlando Bonifácio Martins and Paulo Cavalcanti Gomes Ferreira. We would like to submit this work to be published as *article* in *Microbial genomics and bioinformatics* section from Canadian Journal of Microbiology.

In this work, we describe a PFGE-based analysis of the *Gluconacetobacter diazotrophicus* PAL5 chromosome, and subsequent construction of a physical and genetic map, reporting the chromosome size, number and orientation of the rRNA (*rrn*) operons. In addition, hybridization experiments allowed the positioning of 42 molecular markers on this chromosome, representing a first demonstration of the genomic organization of this important endophyte.

This manuscript represents an original work that is not being considered for publication, in whole or in part, in another journal, book, conference proceedings, or government publication. This manuscript is part of a doctoral thesis by Marcio Martins Loureiro, from which other manuscripts may be generated. Additionally, is convenient to relate that all authors contributed substantially to the manuscript production and approved the final submission.

The complete contact information of each author is listed below:

M. M. Loureiro (loureiro@bioqmed.ufrj.br), M. Bertalan (bertalan@bioqmed.ufrj.br), A. S. Turque (turque@bioqmed.ufrj.br), L. M. França (leomegaf@bioqmed.ufrj.br), V. L. M. Pádua

(vanpadua@bioqmed.ufrj.br), O. B. Martins (omartins@bioqmed.ufrj.br), P. C. G. Ferreira (paulof@bioqmed.ufrj.br)

Núcleo de Estudo de Genomas Johanna Döbereiner, Instituto de Bioquímica Médica, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro, RJ, Brasil.

J. I. Baldani (ibaldani@cnpab.embrapa.br)

Laboratório de Genética e Bioquímica, Embrapa-Agrobiologia, Seropédica, RJ, Brasil.

Additionally, we would like recommend the following scientist to act as reviewers of our manuscript:

1 – **Dr. E. P de Villiers** – Division of Parasitology and Tropical Veterinary Medicine, Faculty of Veterinary Medicine, Utrecht University, PO Box 80.165, 3508 TD Utrecht, The Netherlands. Tel.: +33 30 253 6923. Fax: + 31 30 254 0784. e-mail: e.devilliers@vet.uu.nl

2 – **Dr. Marian Dobos** – Division of Laboratory Medicine, School of Medical Sciences, RMIT University, Melbourne, Victoria, Australia. Fax: +61 3 9925 7063. e-mail: dobos@rmit.edu.au

3 – **Dr. Rogério Tenreiro** – *Address 1* - Departamento de Biologia Vegetal, FCUL, Campo Grande, 1749-016 Lisboa, Portugal – *Address 2*: Centro de Genética e Biologia Molecular, UL, 1749-016 Lisboa, Portugal. Tel: +351 21 750 0000. Fax: +351 21 750 0048. e-mail: rpat@fc.ul.pt

4 – **Dr. Monique Zagorec** – Unité Flore Lactique et Environnement Carné, INRA, Domaine de Vilvert, F-78350 Jouy-en-Josas, France. Tel.: + 33 1 3465 2289. Fax: +33 1 3465 2105. e-mail: zagorec@diamant.jouy.inra.fr

5 – **Dr. L. Elizabeth Bertani** – Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. Tel: + 1 626 395 4917. Fax: + 1 626 405 9452. e-mail: lebert@cco.caltech.edu

6- **Dr. Paul H. Goodwin** - Department of environmental Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1. Tel: +1 519 824 4120. Fax: +1 519 837 0442. e-mail: pgoodwin@uoguelph.ca

Sincerely yours,

Paulo Cavalcanti Gomes Ferreira - PhD

e-mail: paulof@bioqmed.ufrj.br ; ferreira@cshl.edu

Loureiro et al.

Physical and genetic map of *G. diazotrophicus* chromosome

Physical and Genetic Map of the *Gluconacetobacter diazotrophicus* PAL5 chromosome

Marcio Martins Loureiro, Marcelo Bertalan, Aline da Silva Turque, Leonardo Mega França, Vânia Lúcia Muniz de Pádua, José Ivo Baldani, Orlando Bonifácio Martins and Paulo Cavalcanti Gomes Ferreira.

M. M. Loureiro (loureiro@bioqmed.ufrj.br), M. Bertalan (bertalan@bioqmed.ufrj.br), A. S. Turque (turque@bioqmed.ufrj.br), L. M. França (leomegaf@bioqmed.ufrj.br), V. L. M. Pádua (vanpadua@bioqmed.ufrj.br), O. B. Martins (omartins@bioqmed.ufrj.br), P. C. G. Ferreira¹ (paulof@bioqmed.ufrj.br)

Núcleo de Estudo de Genomas Johanna Döbereiner, Instituto de Bioquímica Médica, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro, RJ, Brasil.

J. I. Baldani (ibaldani@cnpab.embrapa.br)

Laboratório de Genética e Bioquímica, Embrapa-Agrobiologia, Seropédica, RJ, Brasil.

¹Corresponding author

Paulo Cavalcanti Gomes Ferreira

Av. Brigadeiro Trompovsky, S/N – Núcleo de Estudo de Genomas Johanna Döbereiner, Instituto de Bioquímica Médica, Centro de Ciências da Saúde, Sala D5 – Subsolo, Universidade Federal do Rio de Janeiro, 21491-590, RJ, Brasil.

Phone +55 21 2562-6751 Fax +55 21 2270-8647

e-mail: paulof@bioqmed.ufrj.br; ferreira@cshl.edu

Abstract: A macrorestriction map of the *Gluconacetobacter diazotrophicus* PAL5, an important nitrogen-fixing endophyte of several plants such as sugarcane, coffee, sweet potato and pineapple, was constructed using pulsed-field gel electrophoresis (PFGE) and DNA hybridization. For PFGE analysis, single and double digestions of the chromosome were carried out with *I-CeuI*, *SwaI* and *PacI* restriction enzymes. The resulting DNA fragments were positioned into physical map, using combinatory analysis of hybridization results, restriction enzymes sites position and DNA fragments length. *G. diazotrophicus* has 4 *rrn* operons into a circular chromosome, with estimated length of 4,240 Kb \pm 41.6 Kb. The hybridization results allow the positioning of 42 genetic markers on chromosome (39 single copy and 3 repeated elements), as well as the detection of 1 *rrn* operon with inverted orientation.

Key words: *Gluconacetobacter diazotrophicus*; PFGE; physical map; genetic map; *rrn* operons.

Introduction

Gluconacetobacter diazotrophicus is a nitrogen-fixing acetic acid α -proteobacterium, found in endophytic association with several sugarcane cultivars, especially in agricultural areas where nitrogen fertilizer input is low (Gillis et al. 1989; Reis et al. 1994; Fuentes-Ramírez et al. 1999; Fischer and Newton 2005). Besides sugarcane, it has been detected into other plant species such as Cameroon grass (*Pennisetum purpureum*), sweet potato (*Ipomoea batatas*) (Döbereiner et al. 1988), coffee (*Coffea arabica*), the grass *Eleusine coracana* and pineapple (*Ananas comosus*) (Santos et al. 2001; Muñoz-Rojas et al. 2005). The ability to colonize a number of families such as *Poaceae*, *Convolvulaceae*, *Rubiaceae* and *Bromeliaceae* suggests that *G. diazotrophicus* possibly has the potential to benefit different plant hosts (Döbereiner et al. 1988; Reis et al. 1994; Santos et al. 2001; Muñoz-Rojas et al. 2005).

G. diazotrophicus biotechnological potential is emphasized by the demonstration that up to 60% of the nitrogen can be acquired through biological nitrogen fixation in some sugarcane varieties (Urquiaga et al. 1992; Baldani et al. 1997; Baldani and Baldani 2005), and the ability to promote plant growth by means of phyto-hormone production, such as auxins and gibberellin (Muthukumarasamy et al. 2002; Muñoz-Rojas and Caballero-Mellado 2003; Lee et al. 2004; Muthukumarasamy et al. 2006). Furthermore, a potential biological control role has been attributed to *G. diazotrophicus*, due to its antagonistic activity against *Xanthomonas albilineans*, as a result of the production of a bacteriocin (Piñón et al. 2002; Blanco et al. 2005), and against *Colletotrichum falcatum*, possibly because of its ability to ferment sugars and reduce medium at pH below 3.0 (Muthukumarasamy et al. 2000).

G. diazotrophicus is able to fix nitrogen and grow at high sucrose concentration (30%) and low pH (3.0). However, the nitrogenase activity is partially inhibited by NH_4^+ at high sucrose

concentrations, though is not affected by high concentration of NO_3^- (25mM) (Baldani et al. 1997; Reis and Döbereiner 1998; Fischer and Newton 2005). The persistence of nitrogen fixation in the presence of nitrate is an additional benefit, because *G. diazotrophicus* could be used as agro-stimulant even in cultivars previously treated with nitrate fertilizers.

Physical maps are considered powerful tools for localization and isolation of genes, studying the organization of genomes, and as a preliminary step for efficient sequencing (Heber et al. 2000). Additionally, physical maps generated by pulsed field gel electrophoresis (PFGE) of large DNA fragments, followed by hybridization and fingerprinting methods, is an alternative to assess the genome organization in organisms for which the complete genome sequence is not available (Majunder et al. 1996; Marcone and Seemüller 2001). Macrorestriction mapping has been applied to the examination of bacterial genome structures and topology for many years (Weinel et al. 2001). Currently, physical mapping strategies have been further developed to verify and accelerate the assembly, in the gap closure phase of microbial genome shotgun-sequencing projects, acting as quality control of sequence and contig assemblies (Chan and Goodwin 1999; Shimizu et al. 2001; Weinel et al. 2001).

Here, we describe the PFGE-based analysis of the *G. diazotrophicus* PAL5 chromosome, and subsequent construction of a physical and genetic map, reporting the genome size, number and orientation of the rRNA (*rrn*) operons. Moreover, hybridization results of DNA probes against macrorestriction DNA fragments enabled the location of 42 genetic markers on this chromosome, representing a first demonstration of the genomic organization of this important endophyte.

Material and methods

Bacterial strains and growth conditions

G. diazotrophicus PAL5 strain (ATCC49037; BR11281) was cultured in Dygs broth (glucose 2.0 g, peptone 1.5 g, yeast extract 2.0 g, K₂HPO₄ 0.5 g, MgSO₄·7H₂O 0.5 g, glutamic acid 1.5 g per liter) pH 6.0, in a gyratory shaker at 30 °C and 200 rpm during approximately 40 h, to obtain a culture with 1.3 optical density (OD₆₀₀). Additionally, we performed control cultures in Agar Dygs plates (Dygs broth supplemented with 1.5% bacteriologic agar) at 30 °C during 48 h, from each culture used for DNA extraction, to confirm the culture purity. The stock strain is maintained frozen at –70 °C in Dygs broth supplemented with 20% glycerol.

Escherichia coli clones from shotgun library were grown in Luria-Bertani (LB) broth (yeast extract 5 g, tryptone 10 g, NaCl 10 g per liter) pH 7.0, supplemented with 100 µg/mL of ampicillin, in a gyratory shaker at 37 °C and 200 rpm during 24 h. The shotgun library are maintained frozen at –70 °C, in 96 well plates with 150 µl of LB broth supplemented with 100 µg/mL of ampicillin and 20% glycerol per well.

Preparation of high molecular weight genomic DNA

Genomic DNA was prepared from 1 mL of *G. diazotrophicus* PAL5 cultures, in logarithmic phase of growth (OD₆₀₀ 1.3). Briefly, cells were suspended in 25 µl of EDTA 50 mM (pH 8.0) buffer, and 10 µl of lysozyme (10 mg/mL), 190 µl of lysis solution (6 mM Tris-HCl – pH 7.5, 100 mM EDTA – pH 7.5, 1 M NaCl, 0.5% Brij58, 0.2% Deoxycholate, 0.5% Sarkosyl), and 225 µl of 2% low-melting-point agarose at 50 °C were added. This mix was gently homogenized and molten in blocks form. The blocks were suspended in 5 mL of lysis solution

supplemented with 20 µg/mL of Rnase A, and incubated at 37 °C for 16 h with gentle shaking. After the incubation period, the lysis solution was replaced by 5 mL of a solution containing 0.4 M EDTA (pH 9.3), 1% sarkosyl, and 1 mg/mL of proteinase K, and incubation continued for 24 h at 50 °C. The agarose blocks were treated with 1 mM phenylmethylsulfonyl fluoride (PMSF), and washed four times for 30 min each in 5 mL of CHEF-TE (100 mM Tris-HCl pH 7.5; 100 mM EDTA pH 7.5). The blocks were stored in 5 mL of 70% ethanol at 4 °C for several months.

Restriction digestions

The rare cut restriction enzymes used in this study were *I-CeuI*, *PacI* and *SwaI* (New England Biolabs, Ontario, Canada). The agarose blocks containing the genomic DNA were sliced in 4 pieces, and immersed into 1× restriction buffer supplied by the manufacturer and incubated at 4 °C. After 1 hour, the buffer was removed and 100 µl of fresh 1× buffer, containing 10 µg of bovine serum albumin (BSA) and 10 U of the enzyme were added, and incubated in the appropriate temperatures, according to the manufacturer's recommendations during 6 h. In double digestions, after the incubation period, the restriction buffers were discarded, and the sliced blocks were washed once with CHEF-TE during 1 h at 4 °C, before of the second digestions in the proper buffers.

Pulsed field gel electrophoresis (PFGE) and Southern blotting

Electrophoresis was carried out in the Gene Navigator PFGE system (GE-Amersham Biosciences, Baie d'Urfe, Canada) with contour-clamped homogeneous field (CHEF). The electrophoresis parameters such as agarose concentration, buffers, pulse times, running voltage and buffer temperature varied, depending on the required experimental conditions, for detection of

the DNA fragments in different length ranges (see legends of the figure 1A-C). Yeast chromosome PFGE marker (225 to 1900 Kb), Lambda ladder PFGE marker (50 to 1000 Kb) and Lambda low range marker (0.13 to 194 Kb) from New England Biolabs, were used as size standards. The gels were stained with ethidium bromide and photographed. The DNA fragments length was estimated by the mean of length values, attributed to the fragments in 10 different gels, using the results generated by the Fragment Analysis Software (GE-Amershan Biosciences). For southern blot, the gels were depurined, denatured, neutralized and blotted to nylon membranes (Hybond+ from GE-Amershan Biosciences) (Sambrook and Russell 2001).

DNA library and sequencing

G. diazotrophicus PAL5 was cultivated as described above, and genomic DNA was isolated according to standard protocol (Sambrook and Russell 2001). One DNA library with average insert sizes of 1 Kb were generated from sonified genomic DNA. These DNA fragments were gel excised, concentrated, end-repaired, cloned into pUC18 vector at the *Sma*I site and electroporated into *Escherichia coli* DH10B (Invitrogen, California, USA). Out of this DNA library, a total of 3,456 clones were double-end sequenced twice in MegaBACE 4000 automated DNA sequencer (GE-Amershan Biosciences).

Selection of clones containing rRNA and single copy genes

The double-end sequencing reactions generated 6,501 DNA sequences which were aligned against non-redundant NCBI database, using the basic local alignment search tool (BLAST) for selection of clones used for DNA probes construction. Clones with inserts containing potential single copy genes of *G. diazotrophicus* PAL5 were chosen for hybridization: 1) their sequences

should be unique in the library 2) their sequences should not show high identity with sequences from others proteins of the same family; 3) their sequences should show homology to published single copy genes. Clones with inserts containing rRNA genes were selected. The sequences of the used clones for DNA probes construction are available under GenBank accession numbers EF508271 - EF508312.

DNA probe construction

Selected clones from the shotgun library were cultured in LB broth as described above, and plasmid DNA was prepared by the rapid alkaline lysis method (Sambrook and Russell 2001). These plasmids were submitted to PCR amplification, in 50 µl reaction volume containing 50 to 100 ng of DNA, 10 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM MgCl₂, 50 pmol of each primer (universal M13 forward and reverse primers) and 2.5 U *Taq* polymerase (Promega, Madison, USA), using the following cycling conditions: an initial denaturation at 95°C for 5 min, followed by 35 cycles with a profile of 1 min at 95 °C (denaturation), 30 s at 50 °C (annealing) and 2 min at 72 °C (extension), followed by a final extension for 5 min at 72 °C. The PCR products were purified by gel excision, and resequenced to confirm the identity. PCR products were labeled by random prime reactions with [α -³²P]dCTP, except the 5S rRNA probe that was labeled with [γ -³²P]dATP through radioactive PCR (Sambrook and Russell 2001).

Southern hybridization

Hybridizations were carried out at 42 °C in the following solution: 50% formamide, 0.5% SDS, 6× SSC, 5× Dendhardt's solution and 100 µg of DNA calf thymus during 24 h. The membranes were then washed under variable conditions, depending on the desired stringency,

dried and exposed into a cassette with a storage phosphor screen during 48 h. After exposure, the screen was scanned with Storm software (GE-Amersham Biosciences) for documentation of the hybridization results. Following, the processed images were compared with the pictures of the gels used in each experiment. All hybridization experiments were realized at least twice to verify the reproducibility of the results.

Results

The DNA fragments generated in each restriction profile were named according to the restriction enzymes that created them (C, P, S, CP, CS or PS), followed by a sequential number according to the decreasing size. In the double digestions, when a given fragment was not digested by a second restriction enzyme, this fragment received the same nomenclature of the single digestion (table 1). Under different electrophoresis conditions, the undigested genomic DNA of *G. diazotrophicus* PAL5 did not entry into the gels, suggesting that is a circular structure (data not shown).

Genomic DNA was digested by single and double digestions, which generated 6 different fingerprinting profiles. The single digestions consisted of digestions with the restriction enzymes *I-CeuI*, *PacI* and *SwaI*, generating profiles identified as C, P and S, respectively. The double digestions consisted of digestions with the following restriction enzymes combinations: *I-CeuI* plus *PacI*, *I-CeuI* plus *SwaI* and *PacI* plus *SwaI*, generating profiles identified as CP, CS and PS, respectively (table 1).

The PFGE profiles generated by single digestions revealed 4 to 6 different DNA fragments, with length ranging from 130 to 1634 Kb, while those of the double digestions have shown 9 to 11 DNA fragments, with length varying from 20 to 1357 Kb. The *G. diazotrophicus*

PAL5 chromosome size was estimated in $4,240 \text{ Kb} \pm 41.6 \text{ Kb}$, through the mean from the estimated genome sizes of all PFGE profiles, generated with single and double digestions, and these values ranged from 4,197 Kb to 4,300 Kb (table 1).

For separation of all detectable DNA fragments, the PFGE gels were run at 3 different electrophoresis programs (figure 1A-C). However, the gels produced under electrophoresis condition of the figure 1A, were considered the most representative, because the PFGE profiles exhibit all detectable DNA fragments. In the others 2 electrophoresis conditions used (figure 1B-C), DNA fragments with length between 2 and 70 Kb were not detected. In addition, under the PFGE conditions employed in figure 1B, some bands classified in this work as doublets, i.e. bands P4 and P5, bands S5 and S6 were separated, confirming their existence. Electrophoresis procedures performed for detection of larger DNA fragments did not detect any additional bands (data not shown).

Table 1 shows the number and estimated length of 45 DNA fragments, generated by single and double digestions of the *G. diazotrophicus* PAL5 chromosome. Among these, 2 fragments designated CP6 and PS6 (from CP and PS double digestions, respectively), were not detected in any experiment conducted in this study. By other hand, considering the number of DNA fragments observed in single digestions, the expected number of DNA fragments in double digestions, the length sum of the detected DNA fragments in PFGE profiles CP and PS, and the estimated length of *G. diazotrophicus* PAL5 chromosome, our results suggest the existence of these DNA fragments and estimate that its present length smaller than 20 Kb. The fact of the CP6 and PS6 fragments were not be observed on PFGE gels, suggest the diffusion of these fragments from blocks during preparations.

The intron-encoded restriction endonuclease *I-CeuI*, which recognizes a highly conserved 26 bp sequence into the large 23S rRNA subunit, cleaved *G. diazotrophicus* chromosome in four fragments, suggesting the existence of 4 *rrn* operons (PFGE profile C in figure 1A; table 1). To confirm this hypothesis, we made hybridizations with DNA probes of the 16S (*rrs*), 23S (*rrl*) and 5S (*rrf*) rRNA genes (table 2), that generally are arranged in this sequential orientation on bacterial genomes. Therefore, we expected 4 DNA fragments hybridized in the PFGE profile C, for each radioactive probe used. However, our results showed that only 3 DNA fragments hybridized in this profile for these probes (figure 2A; table 3).

The *rrs* and *rrl* probes hybridized with the fragments C2, C3 and C4, while the *rrf* probe hybridized with the fragments C1, C2 and C3 (figure 2A; table 3), confirming that the *G. diazotrophicus* PAL5 chromosome is circular and has an inverted orientation of 1 *rrn* operon, located between the fragments C4 and C1, based on the strong signal of the C4 fragment obtained in hybridization with *rrs* and *rrl* probes, and absence of signal for the C1 fragment, while with the *rrf* probe occurs the opposite.

Based on the hybridization results (table 3) generated with single copy gene probes (table 2), we verified overlaps between individual DNA fragments from the six different PFGE profiles of the chromosomal DNA, because these probes showed unequivocal hybridizations with one fragment into each profile. An example of this approach is shown in figure 2B, where the *nifA* probe was hybridized and shows overlap between the DNA fragments C3, P4 and S3 contained in these profiles (table 3). Likewise, all hybridization results were analyzed and annotated, making possible the detection of 12 gene clusters that showed different hybridization profiles (table 3), indicating proximity of genes located on the same cluster. Furthermore, these data were used to determine the molecular markers positioning on this bacterial chromosome.

All restriction fragments conferred hybridization signals, with at least one of the 42 DNA probes, except fragments CP6 and PS6 that probably diffused from blocks during PFGE preparations, and fragments P5 and S6 because we could not select from our DNA library probes that gave hybridization signals with these fragments. However, those 4 DNA fragments that were not detected by hybridization were assembled into the physical map by combinatorial analysis of DNA fragment length and restriction site position of the restriction enzyme which generated them.

The DNA fragments P1 and P2 (PFGE profile P from figure 1A; table 1), CS2 and S3 (PFGE profile CS from figure 1A; table 1), PS1 and PS2, PS4 and PS5 (PFGE profile PS from figure 1A; table 1) were undistinguishable one from other on PFGE gels. However, the hybridization results with single copy genes probes showed that the bands P2, CS2, PS2 and PS5 were doublets, that exhibited overlaps with different individual fragments included in other PFGE profiles, and consequently gave hybridization signals with genes probes located in different gene clusters (table 3)

Considering the overlaps between the majority of DNA fragments from different PFGE profiles, *rrn* genes hybridization results (table 3), DNA fragments length, existence of doublets in PFGE profiles (table 1) and location of restriction enzymes sites, 45 DNA fragments from six different PFGE profiles, were ordered into physical map of the *G. diazotrophicus* PAL5 chromosome (figure 3), using a reduced number of radioactive probes, and avoiding the use of partial digestions of the chromosome, two-dimensional PFGE and linking clones. In total, the hybridizations allowed 42 genetic markers to be allocated into the *G. diazotrophicus* PAL5 physical map (figure 3).

Discussion

Because the *G. diazotrophicus* PAL5 genome has a high GC content (66.4%), the restriction enzymes *PacI* and *SwaI* were chosen to obtain the DNA fragments used to construct the physical map described here. The aim of this strategy was generate a low number of DNA fragments, easily manageable by PFGE methodology. Additionally, these rare-cutting restriction enzymes were related by other authors that mapped GC rich chromosomes as suitable enzymes to construct macrorestriction physical maps of bacterial chromosomes because they cleave these chromosomes into a small number of large DNA fragments (Ramos-Díaz and Ramos 1998; Widjaja et al. 1999; Llamas et al. 2002).

In addition, with the purpose of determinate the number, position and orientation of *rrn* operons, following the strategy previously suggested for precise location of the *rrl* genes on chromosomes (Ramos-Díaz and Ramos 1998; Widjaja et al. 1999; Melkerson-Watson et al. 2000; Zé-Zé et al. 2000; Dudez et al. 2002; El-Osta et al. 2002), we used an intron-encoded restriction endonuclease *I-CeuI*, that recognizes a highly conserved 26 bp restriction site into *rrl* gene.

Among the 45 DNA fragments generated in single and double digestions, only the fragments named as CP6 and PS6 (table 1) were not detected in any experiment, but the results suggest that these predicted DNA fragments probably diffused of the blocks during preparations. This argument is supported by related results from other studies that observed diffusion of smaller DNA fragments from agarose blocks during preparations, or low fluorescence signals in staining procedures (Römling et al. 1989; Chan and Goodwin 1999; Melkerson-Watson et al. 2000).

Physical methods provided direct measurements of genome content for various microorganisms (Ramos-Díaz and Ramos 1998; Zé-Zé et al. 2000; Blank and Stemke 2000; Marcone and Seemüller 2001; Oana et al. 2002). In this study, the *G. diazotrophicus* PAL5

genome size was estimated in 4,240 Kb \pm 41.6 Kb. In addition, results based on hybridization using the *rrn* genes demonstrated that *G. diazotrophicus* has four *rrn* operons in a circular chromosome, and one *rrn* operon is an inverted orientation in relation to others. These results are in agreement with several authors that used this strategy to confirm the number, location and inversions of *rrn* operons (Ramos-Díaz and Ramos 1998; Blank and Stemke 2000; Padovan et al. 2000; Zé-Zé et al. 2000; El-Osta et al. 2002; Oana et al. 2002).

The use of single copy gene probes (table 2), permitted the detection of 12 gene clusters, that showed different hybridization profiles and revealed hybridization signals with the majority of DNA fragments (except fragments P5, S6, CP6 and PS6)(table 3), and together with the *rrn* genes hybridization data, DNA fragments length and restriction sites position, allowed the assembly of the *G. diazotrophicus* PAL5 chromosome physical map, as well as the positioning of molecular markers on its structure. Similar strategy was used for construction of a physical and genetic map of the *Cowdria ruminantium* chromosome (Villiers et al. 2000).

Among 42 genetic markers allocated into the *G. diazotrophicus* PAL5 physical map (Figure 3), some genes such as *mcpA*, *nifA* and *lsdA*, are present in several genomes as part of operons, suggesting the presence of related genes in their neighborhood. Additionally, the *hemE* and *hemH* genes, which are considered strategic markers, because in some bacterial genomes are described as flanking genes of the replication origin (*oriC*) (Capela et al. 2001; Kaneko et al. 2002), were mapped. Consequently, in the *G. diazotrophicus* PAL5 physical map (Figure 3), the fragment C4 was allocated in position zero, because this is the smaller DNA fragment that hybridized with the *hemE* and *hemH* probes.

The hybridization results described here in this paper are being used by the Riogene Genomic Sequencing Consortium (<http://www.riogene.lncc.br>) to help obtain the genomic

sequence of *G. diazotrophicus* PAL5. This approach is in agreement with previous studies (Lai et al. 1999; Blank and Stemke 2000; Weinel et al. 2001), which considered PFGE-based physical maps as an efficient tool, for quality control of sequence assembly generated from shotgun sequencing data, once this strategy allows the detection of possible misassemblies within sequence contigs or contig scaffolds, and helps the re-ordering of an assembly, especially in the flanking regions of *rrn* genes.

The main contribution of this work resides on the fact that little was known on genetic organization of *G. diazotrophicus* PAL5, an important endophytic nitrogen-fixing microorganism, isolated from crops of tropical and subtropical regions, in association with different plant hosts. The data presented here is a first step to understand its genomic organization and will contribute for elucidation of the complete genome sequence of this microorganism.

Acknowledgements

We thank the Conselho Nacional de Apoio a Pesquisa Científica (CNPq) and Fundação de Amparo a Pesquisa Científica do Estado do Rio de Janeiro (Faperj), by financial support to the project (Grants numbers 680040/01-3 and 171.474/2004, respectively).

References

- Baldani, J.I., and Baldani, V.L.D. 2005. History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *An. Acad. Bras. Ciênc.* **77(3)**: 549–579.
- Baldani, J.I., Caruso, L., Baldani, V.L.D., Goi, S.R., and Döbereiner, J. 1997. Recent advances in BNF with non-legume plants. *Soil Biol. Biochem.* **29**: 911–922.
- Blanco, Y., Arroyo, M., Legaz, M.E., and Vicente, C. 2005. Isolation from *Gluconacetobacter diazotrophicus* cell walls of specific receptors for sugarcane glycoproteins, which act as recognition factors. *J. Chromatogr.* **1093**: 204–211.
- Blank, W.A., and Stemke, G.W. 2000. A physical and genetic map of the *Mycoplasma hyopneumoniae* strain J genome. *Can. J. Microbiol.* **46**: 832–840.
- Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J., Boistard, P., Becker, A., Boutry, M., Cadieu, E., Dréano, S., Gloux, S., Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Masuy, D., Pohl, T., Portetelle, D., Pühler, A., Purnelle, B., Ramsperger, U., Renard, C., Thébault, P., Vandebol, M., Weidner, S., and Galibert, F. 2001. Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc. Natl. Acad. Sci. USA.* **98**: 9877–9882.
- Chan, J.W.Y.F., and Goodwin, P.H. 1999. A physical map of the chromosome of *Xanthomonas campestris* pv. *Phaseoli* var. *fuscans* BXP65. *FEMS Microbiol. Lett.* **180**: 85–90.
- Döbereiner, J., Reis, V., and Lazarini, A.C. 1988. New N₂ fixing bacteria in association with cereals and sugarcane. *In Nitrogen Fixation. Edited by H. Bothe, F.J. Briyn, and W.E. Newton.* Gustav Fisher, Stuttgart, Cologne. pp. 717–722.

- Dudez, A.M., Chaillou, S., Hissler, L., Stentz, R., Champomier-Vergès, M.C., Alpert, C.A., and Zagorec, M. 2002. Physical and genetic map of the *Lactobacillus sakei* 23K chromosome. *Microbiology*. **148**: 421–431.
- El-Osta, Y.G.A., Hillier, A.J., Davidson, B.E., and Dobos, M. 2002. Pulsed field gel electrophoretic analysis of the genome of *Lactobacillus gasseri* ATCC33323, and construction of a physical map. *Electrophoresis*. **23**: 3321–3331.
- Fischer, K., and Newton, W.E. 2005. Nitrogenase proteins from *Gluconacetobacter diazotrophicus*, a sugarcane-colonizing bacterium. *Biochim. Biophys. ACTA*. **1750**: 154–165.
- Fuentes-Ramírez, L.E., Caballero-Mellado, J., Sepúlveda, J., and Martínez-Romero, E. 1999. Colonization of sugarcane by *Acetobacter diazotrophicus* is inhibited by high N-fertilization. *FEMS Microbiol. Ecol.* **29**: 117–128.
- Gillis, M., Kersters, K., Hoste, B., Janssens, D., Kroppenstedt, R.M., Stephan, M.P., Teixeira, K.R.S., Döbereiner, J., and Deley, J. 1989. *Acetobacter diazotrophicus* sp. Nov., a nitrogen-fixing acetic acid bacterium associated with sugarcane. *Int. J. Syst. Bacteriol.* **39**: 361–364.
- Heber, S., Hoheisel, J., and Vingron, M. 2000. Application of bootstrap techniques to physical mapping. *Genomics*. **69**: 235–241.
- Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Irigushi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M., and Tabata, S. 2002. Complete genomic sequence of Nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* **9**: 189–197.

- Lai, Z., Jing, J., Aston, C., Clarke, V., Apodaca, J., Dimalanta, E.T., Carucci, D.J., Gardner, M.J., Mishra, B., Anantharaman, T.S., Paxia, S., Hoffman S.L., Venter, C., Huff, E.J., and Schwartz, D.C. 1999. A shotgun optical map of the entire *Plasmodium falciparum* genome. *Nature*. **23(3)**: 309–13.
- Lee, S., Flores-Encarnación, M., Contreras-Zentella, M., Garcia-Flores, L., Escamilla, J.E., and Kennedy, C. 2004. Indole-3-acetic acid biosynthesis is deficient in *Gluconacetobacter diazotrophicus* strains with mutations in cytochrome c biogenesis genes. *J. Bacteriol.* **186(16)**: 5384–5391.
- Llamas, I., Sánchez, M.J, Argandoña, M., Béjar, V., Quesada, E., and Moral, A.D. 2002. Analysis of the genome of the moderate halophile *Halomonas eurihalina*. *Curr. Microbiol.* **45**: 233–239.
- Majumder, R., Sengupta, S., Khetawat, G., Bhadra, R.K, Roychoudhury, S., and Das, J. 1996. Physical map of the genome of *Vibrio cholerae* 569B and localization of genetic markers. *J. Bacteriol.* **178**: 1105–1112.
- Marcone, C., and Seemüller, E. 2001. A chromosome map of the European stone fruit yellows phytoplasma. *Microbiology.* **147**: 1213–1221.
- Melkerson-Watson, L.J., Rode, C.K., Zhang, L., Foxman, B., and Bloch, C.A. 2000. Integrated genomic map from Uropathogenic *Escherichia coli* J96. *Infect. Immun.* **68**: 5933–5942.
- Munõz-Rojas, J., and Caballero-Mellado, J. 2003. Population dynamics of *Gluconacetobacter diazotrophicus* in sugarcane cultivars and its effect on plant growth. *Microb. Ecol.* **46**: 454–464.

- Munõz-Rojas, J., Fuentes-Ramírez, L.E., and Caballero-Mellado, J. 2005. Antagonism among *Gluconacetobacter diazotrophicus* strains in culture media and in endophytic association. FEMS Microbiol. Ecol. **54**: 57–66.
- Muthukumarasamy, R., Govindarajan, M., Vadivelu, M., and Revathi, G. 2006. N-fertilizer saving by the inoculation of *Gluconacetobacter diazotrophicus* and *Herbaspirillum sp.* in micropropagated sugarcane plants. Microbiol. Res. **161**: 238–245.
- Muthukumarasamy, R., Rebatí, G., Seshadri, S., and Lakshminarasimhan, C. 2002. *Gluconacetobacter diazotrophicus* (syn. *Acetobacter diazotrophicus*), a promising diazotrophic endophyte in tropics. Curr. Sci. **83**(2): 137–145.
- Muthukumarasamy, R., Revathi, G., and Vadivelu, M. 2000. Antagonistic potential of N₂-fixing *Acetobacter diazotrophicus* against *Colletotrichum falcatum* Went.: a causal organism of red-rot of sugarcane. Curr. Sci. **78**: 1063–1065.
- Oana, K., Okimura, Y., Kawakami, Y., Hayashida, N., Shimosaka, M., Okazaki, M., Hayashi, T., and Ohnishi, M. 2002. Physical and genetic map of *Enterococcus faecium* ATCC19434 and demonstration of intra and interspecific genomic diversity in enterococci. FEMS Microbiol. Lett. **207**: 133–139.
- Padovan, A.C., Firrao, G., Schneider, B., and Gibb, K.S. 2000. Chromosome mapping of the sweet potato little leaf phytoplasma reveals genome heterogeneity within the phytoplasmas. Microbiology. **146**: 893–902.
- Piñon, D., Casas, M., Blanch, M., Fontaniella, B., Blanco, Y., Vicente, C., Solas, M.T. and Legaz, M.E. 2002. *Gluconacetobacter diazotrophicus*, a sugar cane endosymbiont, produces a bacteriocin against *Xanthomonas albilineans*, a sugar cane pathogen. Microbiol. Res. **53**: 345–351.

- Ramos-Díaz, M.A., and Ramos, J.L. 1998. Combined Physical and genetic map of the *Pseudomonas putida* KT2440 chromosome. *J. Bacteriol.* **180**: 6352–6363.
- Reis, V.M., and Döbereiner, J. 1998. Effect of high sugar concentration on nitrogenase activity of *Acetobacter diazotrophicus*. *Arch. Microbiol.* **171**: 13–18.
- Reis, V.M., Olivares, F.L., and Döbereiner, J. 1994. Improved methodology for isolation and identification of *Acetobacter diazotrophicus* and confirmation of its endophytic habitat. *World J. Microb. Biot.* **10**: 401–405.
- Römling, U., Grothues, D., Bautsch, W., and Tümmler, B. 1989. A physical genome map of *Pseudomonas aeruginosa* PAO. *EMBO J.* **8**: 4081–4089.
- Sambrook, J., and Russell, D.W. 2001. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Santos, P.E.L., Bustillos-Cristales, R., and Caballero-Mellado, J. 2001. *Burkholderia*, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. *Appl. Environ. Microbiol.* **67**: 2790–2798.
- Shimizu, T., Ohshima, S., Ohtani, K., Shimizu, T., and Hayashi, H. 2001. Genomic map of *Clostridium perfringens* Strain 13. *Microbiol. Immunol.* **45**: 179–189.
- Urquiaga, S., Cruz, H.S., and Boddey, R.M. 1992. Contribution of nitrogen fixation to sugarcane: nitrogen-15 and nitrogen-balance estimates. *Soil Sci. Soc. Am. J.* **56**: 105–114.
- Villiers, E.P., Brayton, K.A., Zweygarth, E., and Allsopp, B.A. 2000. Genome size and genetic map of *Cowdria ruminantium*. *Microbiology.* **146**: 2627–2634.
- Weinel, C., Tümmler, B., Hilbert, H., Nelson, K.E., and Kiewitz, C. 2001. General method of rapid Smith/Birnstiel mapping adds for gap closure in shotgun microbial genome

sequencing projects: application to *Pseudomonas putida* KT2440. *Nucleic Acids Res.* **29**: e110.

Widjaja, R., Suwanto, A., and Tjahjono, B. 1999. Genome size and macrorestriction map of *Xanthomonas campestris* pv. *Glycines* YR32 chromosome. *FEMS Microbiol. Lett.* **175**: 59–68.

Zé-Zé, L., Tenreiro, R., and Paveia, H. 2000. The *Oenococcus oeni* genome: physical and genetic mapping of strain GM and comparison with the genome of a divergent strain, PSU-1. *Microbiology.* **146**: 3195–3204.

Table 1. Number and estimated length of the DNA fragments generated by single and double digestions of the *Gluconacetobacter diazotrophicus* PAL5 chromosome.

Single Digestions						Double Digestions					
<i>I-CeuI</i>		<i>PacI</i>		<i>SwaI</i>		<i>I-CeuI / PacI</i>		<i>I-CeuI / SwaI</i>		<i>PacI / SwaI</i>	
Fragment	MW [§] (Kb)	Fragment	MW [§] (Kb)	Fragment	MW [§] (Kb)	Fragment	MW [§] (Kb)	Fragment	MW [§] (Kb)	Fragment	MW [§] (Kb)
C1	1400	P1	1577	S1	1634	CP1	1357	CS1	1052	PS1	1094
C2	1235	P2	≅ 1400*	S2	1152	C4	625	S3	680	PS2	1066*
C3	965	P3	927	S3	680	CP2	576	CS2	≅ 680*	S4	469
C4	625	P4	200	S4	467	CP3	508	C4	625	PS3	390
-	-	P5	≅ 200*	S5	142	CP4	393	S4	468	PS4	285
-	-	-	-	S6	130	CP5	352	CS3	260	PS5	≅ 285*
-	-	-	-	-	-	P4	200	CS4	152	P4	200
-	-	-	-	-	-	P5	≅ 200*	S6	130	P5	≅ 200*
-	-	-	-	-	-	CP6	<20†	CS5	80	S5	142
-	-	-	-	-	-	-	-	CS6	70	S6	130
-	-	-	-	-	-	-	-	-	-	PS6	<20†
TOTAL‡	4225		≅ 4300		4205		≅ 4231		≅ 4197		≅ 4281

* These bands are doublets.

† Predicted fragment by chromosome digestion and alignment of larger DNA fragments, but not observed in PFGE gels.

‡ The mean of the sums of the restriction fragments generated from single and double digestions of the *G. diazotrophicus* PAL5 chromosome was 4,240 Kb with standard deviation of 41.6 Kb.

§ MW, molecular weight.

Table 2. Description of the *rrn* and single copy gene probes used for construction of the *Gluconacetobacter diazotrophicus* PAL5 chromosome physical map.

Gene*	Function of the sequence [†]	Source of similar sequence [‡]	Probability [§]
<i>acoC</i>	Dihydro-lipoamide acetyltransferase	<i>Pseudomonas fluorescens</i> Pf-5	5e-86
<i>araE</i>	Sugar-proton symporter	<i>Streptomyces ambofaciens</i> ATCC 23877	9e-20
<i>araJ</i>	Arabinose efflux permease	<i>Magnetospirillum magnetotacticum</i> MS-1	7e-14
<i>clpP</i>	ATP dependent Clp protease proteolytic subunit 2	<i>Gluconobacter oxydans</i> 621H	1e-88
<i>csp</i>	Cold shock protein	<i>Gluconobacter oxydans</i> 621H	5e-24
<i>dhbF</i>	Nonribosomal peptide synthetase Dhbf	<i>Erwinia carotovora</i> subsp. atroseptica SCRI1043	3e-68
<i>dnaE</i>	DNA polymerase III, alpha chain	<i>Gluconobacter oxydans</i> 621H	5e-131
<i>fcuA</i>	Ferrichrome receptor FcuA	<i>Gluconobacter oxydans</i> 621H	5e-89
<i>fumC</i>	Fumarate hydratase class II	<i>Rhodospseudomonas palustris</i> CGA009	1e-55
<i>gltB</i>	Glutamate synthase (NADPH) large chain	<i>Acidiphilium cryptum</i> JF-5	5e-120
<i>gntP</i>	Gluconate transporter permease	<i>Gluconobacter oxydans</i> 621H	8e-111
<i>hemE</i>	Uroporphyrinogen III decarboxylase	<i>Gluconobacter oxydans</i> 621H	3e-109
<i>hemH</i>	Ferrochelataase	<i>Acidiphilium cryptum</i> JF-5	6e-51
<i>hlyD</i>	Secretion protein HlyD	<i>Ralstonia eutropha</i> JMP134	1e-36
<i>hutH</i>	Histidine ammonia-lyase (Histidase) protein	<i>Gluconobacter oxydans</i> 621H	6e-64
<i>ilvD</i>	Dihydroxy-acid dehydratase	<i>Gluconobacter oxydans</i> 621H	2e-92
<i>kdpC</i>	Potassium-transporting ATPase C chain	<i>Granulibacter bethesdensis</i> CGDNIH1	3e-57
<i>kup</i>	Potassium uptake protein	<i>Bradyrhizobium</i> sp. BTAi1	6e-69
<i>lhr</i>	Superfamily II helicase and inactivated derivatives	<i>Burkholderia dolosa</i> AUO158	6e-22
<i>lig3</i>	ATP-dependent DNA ligase	<i>Rhizobium</i> sp. NGR234	2e-107
<i>lsdA</i>	Levansucrase precursor	<i>Gluconacetobacter diazotrophicus</i>	3e-24
<i>mcpA</i>	Methyl-accepting chemotaxis protein	<i>Rhodospirillum centenum</i>	9e-60
<i>metE</i>	5-methyltetrahydropteroyl-triglutamate-homocysteine methyltransferase	<i>Caulobacter crescentus</i> CB15	1e-111
<i>mexI</i>	Multidrug efflux RND transporter	<i>Pseudomonas entomophila</i> L48	2e-98
<i>mfs</i>	Multidrug efflux MFS membrane fusion protein	<i>Acinetobacter baumannii</i> ATCC 17978	7e-42
<i>mgtA</i>	Mg ²⁺ -importing ATPase	<i>Bradyrhizobium japonicum</i> USDA 110 DNA	2e-59
<i>nifA</i>	Mo/Fe nitrogenase specific transcriptional regulator	<i>Gluconacetobacter diazotrophicus</i>	1e-133
<i>nuoG</i>	NADH dehydrogenase gamma subunit	<i>Nitrosococcus oceani</i> ATCC 19707	5e-41
<i>pdhA</i>	Pyruvate dehydrogenase E1 component alpha	<i>Gluconobacter oxydans</i> 621H	1e-73
<i>pepC</i>	Phosphoenolpyruvate carboxylase	<i>Granulibacter bethesdensis</i> CGDNIH1	3e-132

Table 2. (continued)

Gene*	Function of the sequence[†]	Source of similar sequence[‡]	Probability[§]
<i>pilT</i>	PilT protein-like, outer membrane heme receptor	<i>Serratia proteamaculans</i> 568	6e-11
<i>pyrB</i>	Aspartate carbamoyltransferase	<i>Gluconobacter oxydans</i> 621H	1e-98
<i>rrf</i>	5S ribosomal RNA gene	<i>Bartonella henselae</i> strain Houston-1	2e-36
<i>rrl</i>	23S ribosomal RNA gene	<i>Acetobacter intermedius</i>	0.0
<i>rrs</i>	16S ribosomal RNA gene	<i>Gluconacetobacter diazotrophicus</i>	0.0
<i>sucD</i>	Succinyl-CoA synthetase, alpha subunit	<i>Ralstonia metallidurans</i> CH34	1e-56
<i>trpA</i>	Inner membrane transport protein of toluene-induced toluene exclusion pump	<i>Serratia proteamaculans</i> 568	4e-16
<i>ubiE</i>	SAM-dependent methyltransferase	<i>Mesorhizobium</i> sp. BNC1	4e-50
<i>xdhA</i>	Xanthine dehydrogenase XdhA protein	<i>Gluconobacter oxydans</i> 621H	5e-150
<i>xseA</i>	Exodeoxyribonuclease VII large subunit	<i>Gluconobacter oxydans</i> 621H	8e-42
<i>xyIF</i>	D-xylose ABC transporter, periplasmic-D xylose binding protein	<i>Pseudomonas syringae</i> pv. tomato str.DC3000	7e-71
<i>zntA</i>	Cation transport P-type ATPase	<i>Sinorhizobium meliloti</i>	6e-58

* Names of the genes that were used for probes construction.

† Function of the sequences used for probes construction.

‡ Microorganisms that present best alignment score of its sequence against NCBI data bank using blastn tool.

§ e-value score observed in alignment of the probe sequence against NCBI data bank using blastn tool.

Table 3. Hybridization experiments results with *rrn* and single copy gene probes

Cluster*	Genes [†]	<i>I-CeuI</i> [‡]	<i>PacI</i> [‡]	<i>Swal</i> [‡]	<i>I-CeuI / PacI</i> [‡]	<i>I-CeuI / Swal</i> [‡]	<i>PacI / Swal</i> [‡]
1	<i>hemE, hemH, kdpC, mfs</i> and <i>pyrB</i>	C4	P2	S2	C4	C4	PS1
2	<i>dhbF</i>	C3	P2	S2	CP4	CS4	PS1
3	<i>lig3, mexI</i> and <i>mgtA</i>	C3	P2	S3	CP4	S3	PS5
4	<i>nifA, lsdA</i> and <i>pilT</i>	C3	P4	S3	P4	S3	P4
5	<i>fumC, hutH, kup</i> , and <i>ubiE</i>	C3	P3	S3	CP5	S3	PS4
6	<i>araE, araJ, clpP, csp, fcuA, hlyD, lhr, ilvD</i> , and <i>xseA</i>	C2	P3	S1	CP3	CS1	PS3
7	<i>dnaE, gntP, mcpA, pepC</i> and <i>sucD</i>	C2	P1	S1	CP2	CS1	PS2
8	<i>acoC, pdhA, trpA, xdhA</i> and <i>zntA</i>	C1	P1	S1	CP1	CS2	PS2
9	<i>gliB</i> and <i>nuoG</i>	C1	P1	S4	CP1	S4	S4
10	<i>xyf</i> and <i>metE</i>	C1	P1	S2	CP1	CS3	-
11	<i>rrs</i> and <i>rrl</i>	C2/C3/C4	P1/P2/P3	S1/S2/S5	C4/CP2/CP5	CS1/C4/CS5	PS1/PS2/S5
12	<i>rrf</i>	C1/C2/C3	P1/P2/P3	S1/S2/S5	CP1/CP3/CP4	CS2/CS3/CS4/CS6	PS1/PS2/S5

* Gene clusters that present the same hybridization profile.

† Name of the genes used for probes construction (see table 2)

‡ Single and double digestions generated from *G. diazotrophicus* PAL5 chromosome (see table 1)

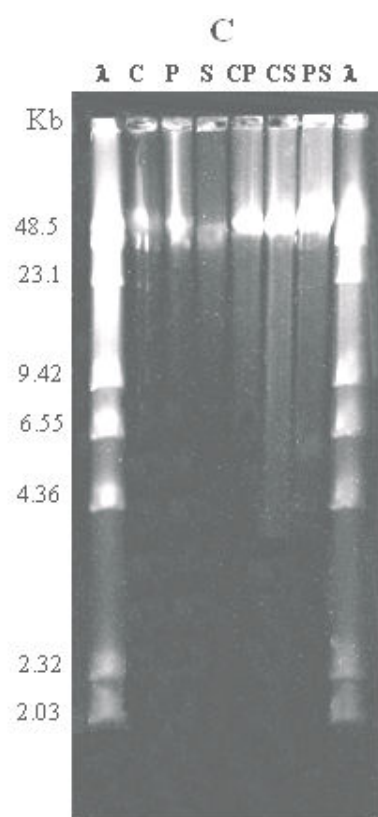
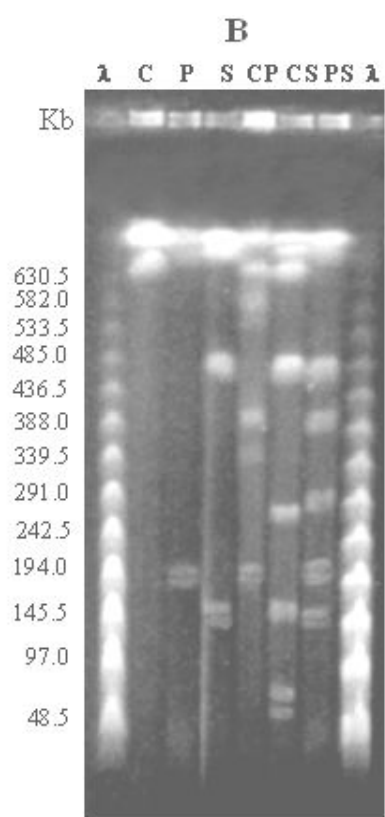
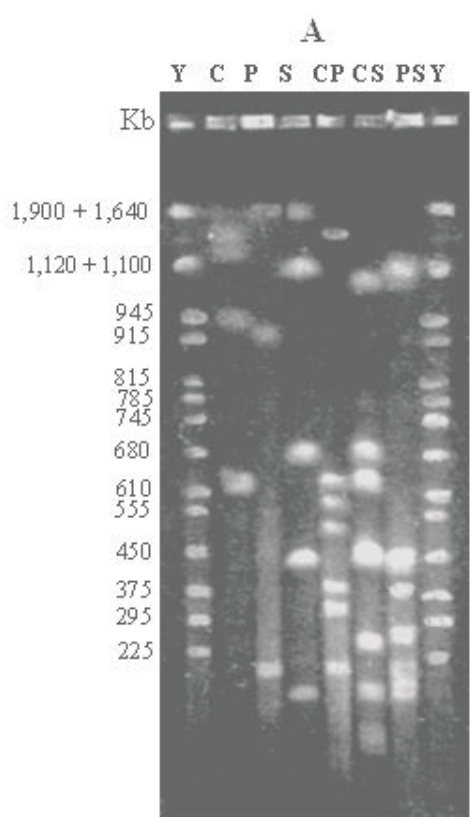


Figure 1. PFGE profiles of *Gluconacetobacter diazotrophicus* PAL5 genomic DNA observed in 3 different electrophoresis conditions. PFGE profiles generated from single or double digestions: C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* plus *PacI*), CS (*I-CeuI* plus *SwaI*), PS (*PacI* plus *SwaI*); Y- yeast chromosome PFGE marker; λ - Lambda ladder PFGE marker (50–1000 Kb) or Lambda Low Range PFGE marker (2.03–194 Kb); **A** - Electrophoresis procedure carried out in 1% agarose gel and running buffer containing 0.5 \times TBE at 6V/cm at 13 °C in a sequence of three ramps (ramp A consisted of an initial switch time of 80 s, a final switch time of 100 s during 16 h; ramp B consisted of an initial switch time of 100 s, a final switch time of 120 s during 8 h; and ramp C consisted of an initial switch time of 120 s, a final switch time of 140 s during 16 h; **B** - Electrophoresis procedure carried out in 1% agarose gel and running buffer containing 0.5 \times TBE at 6v/cm at 13°C with initial switch time of 1s and final switch time of 50s during 23h. **C** - Electrophoresis procedure carried out in 1% agarose gel and running buffer containing 1 \times TAE at 4 V/cm at 13 °C with initial switch time of 0.1 s and final switch time of 10 s during 16 h.

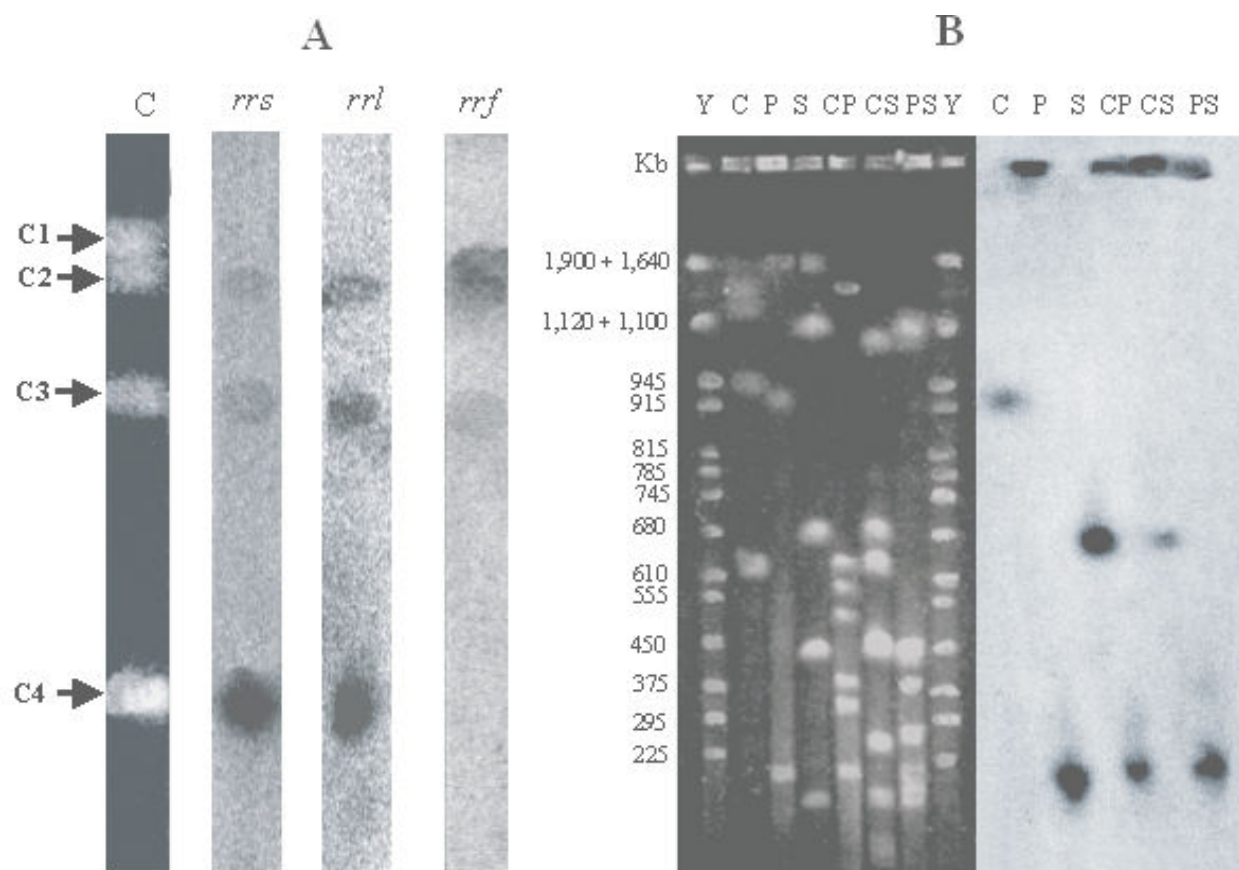


Figure 2. Hybridization with *rrn* probes against PFGE profile generated with *I-CeuI* digestion of *Gluconacetobacter diazotrophicus* PAL5 chromosome, and an example of hybridization results generated with single copy genes against PFGE profiles. **A** - hybridization results with *rrn* probes; **B**- hybridization results with *nifA* probe. The single and double digestions of the *G. diazotrophicus* PAL5 chromosome were C (*I-CeuI*), P (*PacI*), S (*SwaI*), CP (*I-CeuI* plus *PacI*), CS (*I-CeuI* plus *SwaI*), PS (*PacI* plus *SwaI*). Y - yeast chromosome PFGE marker. *rrs* - 16S ribosomal RNA gene; *rrl* - 23S ribosomal RNA gene ; *rrf* - 5S ribosomal RNA gene. C1, C2, C3 and C4 are the DNA fragments generated from chromosome digestion with *I-CeuI* restriction enzyme (see table 1).

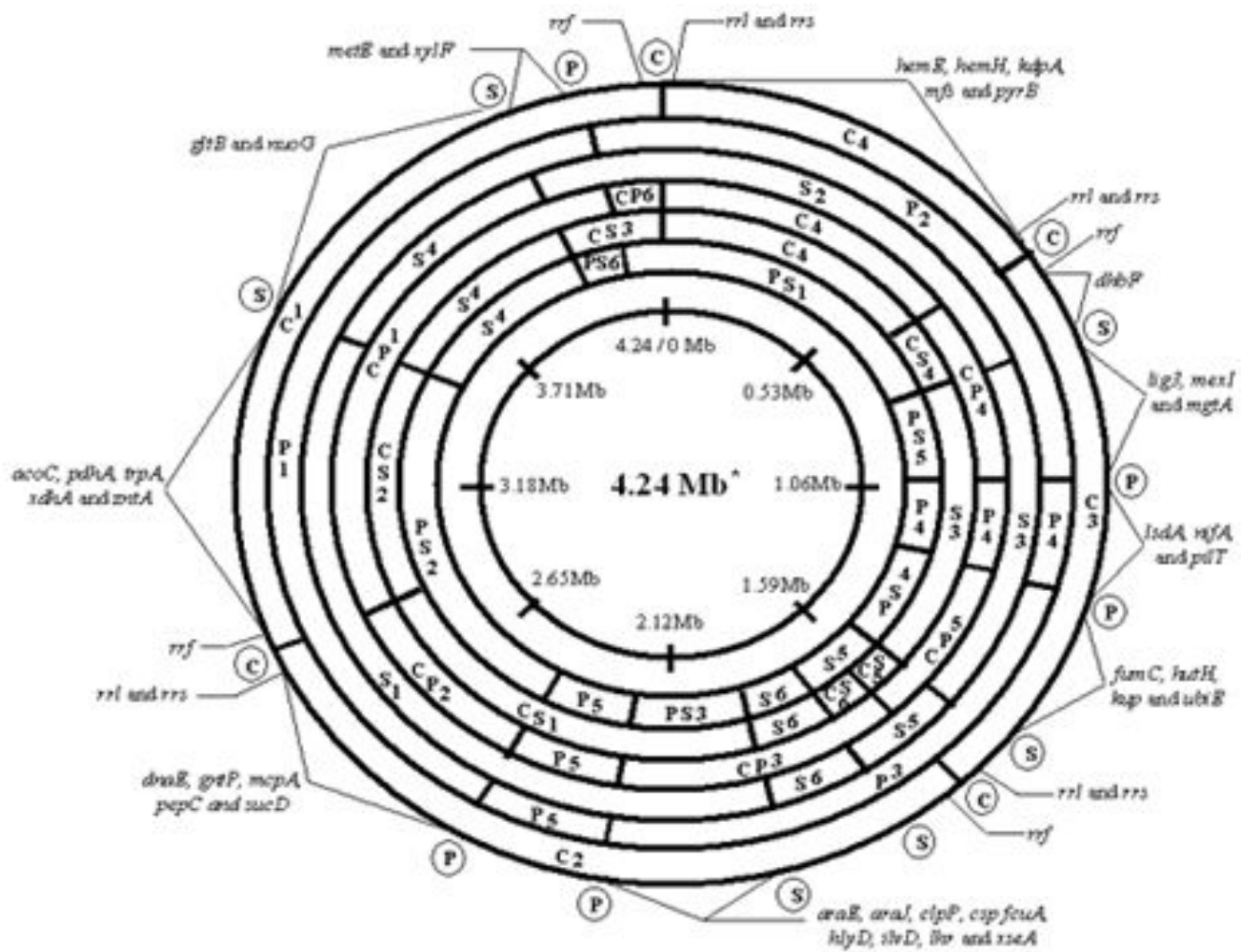


Figure 3. Physical map of *Gluconacetobacter diazotrophicus* PAL5 chromosome using the *I-ceuI*, *PacI* and *SwaI* restriction enzymes. C, P and S – restriction sites of the *I-CeuI*, *PacI* and *SwaI* restriction enzymes, respectively. Restriction fragments were named according to established criteria in materials and methods (see table 1). The DNA fragments CP6 and PS6 were predicted from chromosome digestion results and aligned on this physical map, but were not observed in PFGE gels. The genetic markers are described in table 2. The exact positioning of the molecular markers in DNA fragments, were not determined in this physical map. The innermost circle represents the scale in megabases of the *G. diazotrophicus* chromosome. * this value represent the length mean of the *G. diazotrophicus* chromosome.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)