



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE BIOCÊNCIAS  
DEPARTAMENTO DE BIOLOGIA CELULAR E GENÉTICA  
PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

**Análise comparativa dos genes de reparo do DNA em  
*Gluconacetobacter diazotrophicus* e *Gluconobacter oxydans***

CAROLINA CORADO DA SILVA OLIVEIRA

NATAL-RN  
2009



**UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE BIOCÊNCIAS  
DEPARTAMENTO DE BIOLOGIA CELULAR E GENÉTICA  
PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR**

**Análise comparativa dos genes de reparo do DNA em  
*Gluconacetobacter diazotrophicus* e *Gluconobacter oxydans***

**CAROLINA CORADO DA SILVA OLIVEIRA**

Dissertação apresentada ao Programa de Pós-graduação em Genética e Biologia Molecular do Centro de Bociências da Universidade Federal do Rio Grande do Norte, como parte dos requisitos para a obtenção do título de Mestre em Genética e Biologia Molecular.

**Orientador: Prof<sup>ª</sup>. Dra. Silvia Regina Batistuzzo de Medeiros**

**NATAL-RN  
2009**

# SUMÁRIO

Lista de figuras	iv
Lista de tabelas	vii
Resumo	viii
1. Introdução	1
1.1. Alfaproteobactérias	5
1.1.1. <i>Gluconacetobacter diazotrophicus</i>	6
1.1.2. <i>Gluconobacter oxydans</i>	7
1.1.3. <i>Caulobacter crescentus</i>	8
1.2. Sistemas de Reparo	10
1.2.1. Reparo por excisão	12
a. Reparo por excisão de bases (BER)	12
b. Reparo por excisão de nucleotídeos (NER)	15
c. Reparo de bases mal emparelhadas (MMR)	17
i. Reparo <i>Very Short Patch</i> (VSR)	18
1.2.2. Reparo direto	20
a. Fotorreativação	20
b. Alquiltransferência	22
c. Reparo oxidativo direto	22
1.2.3. Reparo recombinacional	23
1.2.4. Outras vias de reparo	25
a. Resposta SOS	25
b. Ligases	27
c. Helicases	27
1.3. Comparação gênica e filogenia	28
1.3.1. Grupos	29
1.3.2. Árvores filogenéticas	30
1.3.3. Análise comparativa	31
a. Métodos baseados em distância	31
b. Métodos baseados no caráter	32
1.3.4. Bootstrap – Análise de confiabilidade	33
1.4. Justificativa	33
2. Objetivos	35
3. Metodologia	36
3.1. Estrutura física	36
3.2. Estrutura lógica	36
3.3. Montagem do banco de dados das seqüências peptídicas e nucleicas relacionadas ao reparo em <i>Escherichia coli</i>	36
3.4. Montagem do banco de dados das seqüências peptídicas e nucleicas relacionadas ao reparo em <i>G. diazotrophicus</i> , <i>G. oxydans</i> e <i>C. crescentus</i>	36
3.5. Confirmação dos ortólogos e parálogos via análise dos domínios presentes	38
3.6. Construção das árvores filogenéticas	38

4.	Resultados e discussão	40
4.1.	Reparo por excisão	40
a.	Reparo por excisão de nucleotídeos (NER)	40
b.	Reparo por excisão de bases (BER)	46
c.	Reparo de bases mal emparelhadas (MMR)	52
4.2.	Reparo direto	57
a.	Fotorreativação	57
b.	Alquiltransferência	58
c.	Reparo oxidativo direto	61
4.3.	Reparo recombinacional	63
4.4.	Outras proteínas relacionadas ao reparo	76
a.	Resposta SOS	76
b.	Ligases	79
c.	Helicases	79
4.5.	Considerações finais	82
5.	Conclusões	84
6.	Perspectivas	85
7.	Referências bibliográficas	86
8.	Anexos	94

## LISTA DE FIGURAS

- Figura 1.** Vias de reparo por excisão de base curta (SP-BER) e longa (LP-BER). As enzimas em azul correspondem às de procariotos e em vermelho às de eucariotos..... 14
- Figura 2.** Desenho esquemático do sistema de reparo de nucleotídeos. TCR refere-se ao Reparo Acoplado à Transcrição, na qual há a participação de Mfd. Em seguida no reparo genômico global (GGR), UvrA é o elemento primário para o reconhecimento da lesão..... 16
- Figura 3.** Mecanismo do reparo de bases mal emparelhadas direcionado por metilação em *E. coli*. Embora não mostrado, a DNA ligase restaura a ligação covalente da fita reparada após a holoenzima DNAPol III preencher a lacuna. As setas verdes indicam sinalização dependente de MutS e MutL entre os dois sítios de DNA envolvidos na reação..... 19
- Figura 4.** O DNA exposto à radiação ultravioleta (uv) resulta na dimerização covalente das pirimidinas adjacentes, ilustradas aqui como triângulos roxos. Essas lesões são reconhecidas pela fotoliase (verde), que absorve luz com comprimento de onda menor que 300 nm e monomeriza os dímeros formados. 21
- Figura 5.** Recombinação homóloga. A – reparo de quebras duplas na fita de DNA e B – reparo de uma lacuna na fita simples..... 24
- Figura 6.** Grupos organizados em clados. Grupos monofiléticos compartilham um ancestral comum. Um grupo é dito parafilético se ele contém um ancestral comum recente, porém não contém todos os descendentes desse ancestral. Já um grupo polifilético é aquele descendente de mais de um ancestral, podendo ser mais proximamente relacionado a outros organismos fora do grupo taxonômico..... 30
- Figura 7.** Árvore gerada pelo método de distância NJ para o gene *uvrA*. Os valores associados com cada nó correspondem aos valores gerados pela análise do bootstrap. Somente valores acima de 50% são mostrados nas árvores. Os organismos são indicados por nomes abreviados listados no anexo 1. Destaque em rosa onde consta a segunda cópia do gene *uvrA* de *G. oxydans*. A legenda: (A) para proteobactéria alfa, (B) para proteobactéria beta, (D) para proteobactéria delta, (G) para proteobactéria gama e (E) para proteobactéria épsilon. A seta indica o segundo ortólogo encontrado em *G. oxydans*..... 44
- Figura 8.** Árvore gerada pelo método de distância NJ para o gene *uvrD*. Os valores associados com cada nó correspondem aos valores gerados pela análise do bootstrap. Somente valores acima de 50% são mostrados nas árvores. Os organismos são indicados por nomes abreviados listados no anexo 1. A legenda:

(A) para proteobactéria alfa, (B) para proteobactéria beta, (D) para proteobactéria delta, (G) para proteobactéria gama e (E) para proteobactéria épsilon. As setas indicam os segundos ortólogos encontrados em *G. diazotrophicus* e *G. oxydans*..... 45

**Figura 9.** Alinhamento pelo Clustal das seqüências de Ung com o ortólogo da família 4 presente em *M. tuberculosis*. O motivo conservado que pode ser utilizado como indício característico para classificar as proteínas nessa família está destacado..... 49

**Figura 10.** Árvore não enraizada gerada pelo método de distância NJ para o gene *xthA*. Os valores associados com cada nó correspondem aos valores gerados pela análise do *bootstrap*. Somente valores acima de 50% são mostrados nas árvores. Os organismos são indicados por nomes abreviados listados no anexo 1. A legenda: (A) para proteobactéria alfa, (B) para proteobactéria beta e (G) para proteobactéria gama. As setas indicam os segundos ortólogos encontrados em *G. diazotrophicus* e *G. oxydans*..... 50

**Figura 11.** Árvore não enraizada gerada pelo método de distância NJ para o gene *mutS*. Os valores associados com cada nó correspondem aos valores gerados pela análise do *bootstrap*. Somente valores acima de 50% são mostrados nas árvores. Os organismos são indicados por nomes abreviados listados no anexo 1. As três caixas separam os grandes grupos de variantes da proteína MutS (MutS1, MutS2 e MutS3). A seta indica o segundo ortólogo encontrados em *G. diazotrophicus*..... 55

**Figura 12.** Representação esquemática dos padrões para a disposição dos domínios regulatórios das proteínas Ada e Ogt em bactérias. A porção regulatória de *ada* está representada pelos retângulos azuis e amarelos, respectivamente ligação ao DNA (sítio de ligação ada-zinco) e ativação transcricional (HTH-AraC). O domínio alquiltransferase, também presente na proteína Ogt, está representado pelos retângulos em laranja. O retângulo verde caracteriza o domínio AlkA, que está envolvido na remoção de grupos alquil pela via de reparo por excisão de bases..... 60

**Figura 13.** Alinhamento múltiplo das seqüências prováveis de RecB para Gd (*gldia\_recb*) e Go (*gloxy\_recb*) e RecB (*ecoli\_recb*) de *E. coli*. Os números após os nomes das seqüências representam o tamanho de cada ortólogo analisado. A coloração em verde indica os motivos conservados da helicase a conservação dentro dos grupos alinhados. Os gráficos abaixo representam, em seqüência, a taxa de conservação, a qualidade do alinhamento e o consenso gerado pelo programa Clustal. Os retângulos azuis localizados no final do alinhamento assinalam os três resíduos catalíticos importantes de RecB (D1067, D1080 e K1082)..... 65

**Figura 14.** Árvore não enraizada gerada pelo método de distância NJ para o gene *recD*. Os valores associados com cada nó correspondem aos valores

gerados pela análise do *bootstrap*. Somente valores acima de 50% são mostrados nas árvores. Os organismos são indicados por nomes abreviados listados no anexo 1. O retângulo rosa indica o grupo no qual os ortólogos de RecD em *G. diazotrophicus* apresentem maior similaridade..... 70

**Figura 15.** Organização gênica dos operons preditos codificadores dos componentes do reparo associado à Ku. A direção das setas indica o sentido da transcrição. Regiões distintas que codificam domínios no mesmo gene estão com cores diferentes (ligase em vermelho, primase em rosa e domínio de ligação ao ATP em verde claro). Os números dos *loci* são dados de acordo com a anotação genômica. Em Gd, o domínio ligase está truncado..... 72

**Figura 16.** Organização gênica de diferentes operons compostos pelos genes *dnaE*, *imuA* e *imuB*. Em *G. diazotrophicus* é possível encontrar a mesma organização presente em *C. crescentus* e em *Propionibacterium acnes*. Os demais organismos usados nessa comparação são: *Pseudomonas putida*, *Rhodospirellula báltica*, *Mycobacterium tuberculosis* e *Streptomyces avermitilis*. As setas indicam os genes e as linhas interrompidas sinalizam que os genes não estão próximos dentro do genoma. A interrogação marca um gene que pode ser codificador de uma proteína semelhante a RecA..... 78

**Figura 17.** Diagrama que apresenta o número de duplicações e de genes compartilhados por *G. diazotrophicus*, *G. oxydans* e *C. crescentus*..... 83

## LISTA DE TABELAS

<b>Tabela 1.</b> Componentes da via de reparo por excisão de bases em <i>E. coli</i> .....	13
<b>Tabela 2.</b> Principais genes que fazem parte da via de reparo por excisão de nucleotídeos (NER).....	43
<b>Tabela 3.</b> Principais genes que fazem parte da via de reparo por excisão de bases (BER).	51
<b>Tabela 4.</b> Principais genes que fazem parte da via de reparo de bases mal emparelhadas (MMR).....	56
<b>Tabela 5.</b> Principais genes que fazem parte da via de reparo direto.....	62
<b>Tabela 6.</b> Principais genes que fazem parte da via de reparo recombinacional.....	73
<b>Tabela 7.</b> Genes que fazem parte de outras vias relacionadas ao reparo .....	80



## RESUMO

*Gluconacetobacter diazotrophicus* é uma alfa-proteobactéria Gram-negativa, tolerante a meios ácidos, fixadora de nitrogênio atmosférico e foi a primeira bactéria diazotrófica endofítica isolada da cana-de-açúcar. Por sua vez, *Gluconobacter oxydans*, também alfa-proteobactéria Gram-negativa, possui a capacidade de oxidar incompletamente alcoóis e carboidratos. Ambas de interesse biotecnológico e industrial, essas bactérias tiveram seus genomas seqüenciados completamente em 2007. Desta forma, foi de interesse desse trabalho analisar e comparar os genes de reparo do DNA devido sua importância na manutenção da integridade genômica. Sendo assim, as vias de reparo presentes nos dois organismos foram identificadas, utilizando como base uma terceira alfa-proteobactéria, a *Caulobacter crescentus*, cujos genes de reparo foram descritos por um trabalho anterior e também os genes bem estabelecidos para o reparo do DNA em *Escherichia coli*. Para esse estudo, um banco de dados contendo ortólogos para os genes de reparo de DNA encontrados nos organismos foi criado e análises comparativas por similaridade usando o pacote Blast e o software Clustal foram feitas. Este estudo demonstrou que as principais vias de reparo ao DNA – reparos por excisão, reparo direto, reparo recombinacional e reparo pelo sistema SOS – estão presentes nos organismos analisados, demonstrando, na maioria das vezes, boa similaridade com *E. coli*. Interessantemente, foram encontradas duplicações gênicas nos quais uma das cópias estava presente no cromossomo e a outra, no plasmídeo, como no caso de UvrD, DnaE e Ssb, possivelmente caracterizando eventos de transferência lateral. Por fim, uma grande novidade foi a identificação de ortólogos para RecB em *G. diazotrophicus* e *G. oxydans* e de ortólogos duplicados de RecD em *G. diazotrophicus*. Até o momento, não havia sido relatada a presença de membros da via de iniciação RecBCD do reparo recombinacional em alfa-proteobactérias.

**Palavras chave:** *Gluconacetobacter diazotrophicus*, *Gluconobacter oxydans*, alfa-proteobactérias e reparo do DNA.

## 1. Introdução

A definição de genoma remete ao início do século passado, quando Winkler (1920) originalmente o utilizou para designar o somatório dos genes de uma célula haplóide de um organismo. Posteriormente, as seqüências de DNA não codificadoras foram incluídas nessa definição (Lederberg & McCray, 2001).

Hoje, é sabido que num organismo, o genoma está representado por uma ou várias moléculas de ácido desoxirribonucléico (DNA) e, em alguns vírus, por uma molécula de ácido ribonucléico (RNA). A Genômica surgiu como uma área cujo objetivo é a compreensão global da estrutura e funcionamento dos diversos organismos por meio do seqüenciamento do DNA, do mapeamento de genes nos cromossomos e plasmídeos e análise de suas funções (Gregory, 2004).

Até recentemente, a descoberta dos genes e de suas funções era elucidada passo-a-passo pela combinação da genética, da biologia molecular e da bioquímica. Entender a função de um gene envolve a identificação de fenótipos mutantes, localização dos genes e seus perfis de expressão, além da confirmação de suas funções utilizando ensaios bioquímicos. No final dos anos 70, Sanger e seus colaboradores desenvolveram uma metodologia capaz de determinar a seqüência nucleotídica de qualquer DNA, sendo denominada de seqüenciamento enzimático. Sanger e colaboradores (1978) publicaram o primeiro genoma totalmente seqüenciado do bacteriófago  $\phi$ X174, utilizando a metodologia *plus and minus*, contendo aproximadamente 5374 nucleotídeos. Como conseqüência, o banco de dados genômicos passou a crescer e com grande velocidade.

Dado o presente volume e a velocidade de crescimento dos dados disponíveis sobre genomas, a tarefa de análise destes dados torna-se inviável pelos métodos laboratoriais conhecidos. Para enfrentar este problema, verifica-se na atualidade uma crescente interação entre a biologia molecular, a ciência da computação e a matemática. O reconhecimento de que somente tal interação poderá enfrentar com sucesso esse problema levou à moderna integração de especialistas dessas áreas, criando campos comuns de atividades: a bioinformática e a biologia computacional. O desenvolvimento de métodos computacionais eficientes e sensíveis para a análise de bancos de seqüências de DNA e de proteínas é hoje importante e muito necessário. A

análise teórica de seqüências compreende um largo espectro, desde algoritmos simples até os mais complexos que são desenvolvidos para processadores paralelos avançados.

Desde a publicação da seqüência genômica completa de *Haemophilus influenzae* (Fleischmann *et al.*, 1995) mais de 800 genomas microbianos completamente seqüenciados foram publicados e muitos mais estão em seu caminho (Liolios, *et al.*, 2007) – a lista pode ser verificada em [www.genomesonline.org](http://www.genomesonline.org). Porém, existe a tendência de focar a atenção em organismos de interesses particulares devido as suas propriedades, como os patógenos humanos, e agora com a disponibilidade de seqüências genômicas completas, tem-se mostrado um panorama melhor acerca da diversidade observada em procariotos cultiváveis.

No Brasil, a genômica obteve destaque internacional com o seqüenciamento do genoma da gama-proteobactéria *Xylella fastidiosa*, um fitopatógeno endofítico responsável pela clorose variegada dos citros (Simpson *et al.*, 2000). O projeto foi de iniciativa da FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) e realizado pela rede ONSA (*Organization for Nucleotide Sequence Analysis*), no qual 34 laboratórios do estado de São Paulo foram envolvidos. Com o sucesso do projeto, diversos outros organismos tiveram seus genomas seqüenciados com o auxílio da FAPESP, como *Xanthomonas citris* e *Schistosoma mansoni*. E por conta do reconhecimento internacional pela competência e qualidade do seqüenciamento, o Ministério da Ciência e Tecnologia juntamente com o Conselho Nacional para o Desenvolvimento da Ciência e Tecnologia (CNPq) instituíram o Projeto Genoma Brasileiro, a fim de expandir esta tecnologia ao restante do Brasil e, com isso, 25 laboratórios distribuídos pelo país foram selecionados para participar da iniciativa. O organismo escolhido para estreitar o Projeto Genoma Brasileiro foi a beta-proteobactéria *Chromobacterium violaceum*, que é um organismo de vida livre de enorme potencial biotecnológico, cuja característica marcante é a coloração roxa advinda da produção do pigmento púrpuro, a violaceína (McClellan *et al.*, 1997). Esse projeto capacitou os laboratórios a seqüenciarem paralelamente outros genomas, como foi o caso da linhagem *Gluconacetobacter diazotrophicus* PAL5, pelo Consórcio

Genômico RioGene em associação com o Laboratório Nacional de Computação Científica (LNCC).

Enquanto que a disponibilidade de um número cada vez maior de genomas completamente seqüenciados tem facilitado significativamente a busca por novos marcadores moleculares, existem outras maneiras pelas quais essas seqüências podem ser utilizadas na dedução de novos marcadores moleculares entre os *taxa*. Neste trabalho, será focada uma abordagem no conteúdo gênico.

Quando se compara o conteúdo gênico dos organismos, a maneira mais simples é considerar os genomas como “grupos de genes” e os conteúdos de diferentes “grupos” são comparados (Huynen & Bork, 1998). Uma definição operacional para ortologia é: genes ortólogos são aqueles que têm o mais alto grau de identidade de pareamento significativo quando comparados com as identidades de quaisquer genes entre todos os outros genes de outros genomas. Isto é, eles são identificados como aqueles genes homólogos que mostram a maior identidade dentro de uma gama de possibilidades estando acima de uma linha de corte (Bansal & Meyer, 2002; Coenye & Vandamme, 2003 e Huynen & Bork, 1998). Sendo assim, a identificação de genes ortólogos é crucial nessa abordagem (Wolf, *et al.*, 2002).

A partir das primeiras análises feitas com um número limitado de genomas, algumas tendências surgiram (Huynen & Bork, 1998 e Snel, Bork, & Huynen, 1999). Foi observado que (i) genomas grandes apresentam muitos genes em comum; (ii) o número de genes que dois genomas têm em comum depende de sua distância evolutiva; (iii) a fração de genes ortólogos compartilhados diminui rapidamente na evolução, mais rápido até do que a identidade entre os ortólogos compartilhados; e (iv) a evolução do conteúdo gênico mostra características que não se assemelham a uma árvore, como em espécies próximas relacionadas filogeneticamente não necessariamente compartilham genes ortólogos que uma delas compartilhe com uma terceira espécie filogeneticamente mais distante (Gevers & Coenye, 2005).

Snel e colaboradores (1999) também observaram que as árvores baseadas no conteúdo gênico não correlacionavam bem com as árvores baseadas em fenótipo. De certa forma, isso não é esperado, pois se tem assumido que o fenótipo observado era o resultado das diferenças no conteúdo gênico.

Diversas abordagens que apresentam diferenças sutis são utilizadas para comparar o conteúdo gênico. A presença e ausência de famílias de genes codificadores de proteínas em genomas seqüenciados têm sido utilizadas para reconstruir as relações entre organismos (Fitch, 2000 e House & Fitz-Gibbon, 2002). Nesta abordagem, proteínas são agrupadas em famílias gênicas se sua similaridade no pareamento for maior que um valor previamente determinado, eliminando assim a necessidade da identificação de ortólogos prováveis em cada genoma e a necessidade de alinhamentos específicos.

Os métodos que levam em consideração o conteúdo de genes compartilhados para reconstruir filogenias foram criticados devido à tendência existente da conversão do conteúdo gênico – por meio de transferência gênica horizontal e perda gênica. Contudo, as árvores baseadas no conteúdo gênico geralmente correspondem bem com as árvores baseadas nas seqüências de rRNA 16S – porém algumas discrepâncias observadas em vários estudos permanecem sem explicação. Isso indica que, apesar da transferência gênica horizontal, duplicação gênica e perda gênica, existe um forte sinal filogenético no conteúdo gênico (Huson & Steel, 2004 e Zhang & Gu, 2004).

Desta forma, o contexto genômico com suas possíveis interações inferidas a partir do conteúdo gênico e da vizinhança de um gene é o passo inicial que pode gerar muitas informações sobre a função e a regulação gênica (Huynen, *et al.*, 2000 e Lawrence & Ochman, 1998). Também, comparando-se genes de diferentes organismos, é possível obter pistas a respeito da evolução genômica e extrapolar para a evolução do próprio organismo ao longo do tempo, por meio, por exemplo, da identificação de eventos de transferência lateral e elementos de inserção (Lawrence & Ochman, 1998 e (Nelson *et al.*, 1999). Já a simples identificação de genes sem possibilidade de definir função por falta de homólogos em outros organismos pode ajudar a descrever a história de um organismo ou grupos relacionados (Graham *et al.*, 2000) como são *Gluconobacter oxydans* e *Gluconacetobacter diazotrophicus*, membros do grupo das alfa-proteobactérias.

## 1.1 Alfa-proteobactérias

O filo das Proteobactérias, antigamente conhecido como bactérias púrpuras, abriga um grande número de bactérias Gram-negativas muito diversas. A fotossíntese, relacionada a um pigmento púrpuro, é limitada a alguns organismos desse filo (Teyssier *et al.*, 2004).

As proteobactérias foram claramente definidas por Woese em 1985 com a construção de uma árvore filogenética feita a partir de seqüências do rDNA 16S. Essa análise permitiu que ele subdividisse as proteobactérias em cinco sub-grupos denominados: alfa, beta, gama, delta e épsilon. Já em 1988, o comitê de bacteriologia sistemática propôs uma substituição da denominação “bactéria púrpura” para *Proteobacteria*, fazendo dessa maneira uma referência à diversidade de forma, de fisiologia e de modos de vida dos organismos desse filo. E recentemente, Cavalier-Smith (2002) revisou os grandes táxons bacterianos e deu ao filo das proteobactérias o nível de divisão. De acordo com ele, as alfa-proteobactérias apresentam modos de vida variados e originais, pois, particularmente, são capazes de estabelecer relações estreitas com as células eucarióticas.

As alfa-proteobactérias possuem uma estrutura genômica atípica. Nelas foram descritos a presença de um cromossomo clássico – único e circular, a presença de cromossomos múltiplos e com menos freqüência linear. Existem também grandes plasmídeos com mais de 1 Mbp. Em termos de evolução genômica, provavelmente parece que um ancestral comum possuía um genoma formado de um cromossomo circular, de um ou alguns megaplasmídeos e de plasmídeos um pouco menores. Os eventos independentes que ocorreram em cada ramo conduziram a diferentes estruturas genômicas das alfa-proteobactérias atuais (Teyssier *et al.*, 2004).

Tendo em vista sua diversidade, as AP incluem as bactérias de solo, tais quais as envolvidas no ciclo global do nitrogênio (*Bradyrhizobium*, *Mesorhizobium*, *Nitrobacter*, *Sinorhizobium* and *Rhizobium*), as indutoras de tumor bacteriano em plantas (*Agrobacterium*), as encontradas freqüentemente na água (*Candidatus*, *Caulobacter*, *Erythrobacter*, *Jannaschia*, *Sphingopyxis*, *Silicibacter*), além de organismos patogênicos (*Anaplasma*, *Bartonella*, *Brucella*, algumas espécies de *Wolbachia*, *Rickettsia* and *Ehrlichia*) e de interesse ambiental, como os produtores de etanol (*Zymomonas*) ou

aqueles que têm a capacidade de degradar uma vasta gama de compostos orgânicos tóxicos, assumindo assim um papel na biorremediação (*Novosphingobium*, *Rhodopseudomonas* and *Rhodobacter*) (Martins-Pinheiro, *et al.*, 2007).

Além dessas, existem as bactérias pertencentes à família *Acetobacteraceae* que são acidófilas e têm a capacidade de crescer a pH 4,5 com um metabolismo oxidativo estrito (Sievers, *et al.*, 2005). A espécie *Gluconobacter oxydans* é de grande importância alimentícia e farmacológica por ser capaz de oxidar uma grande variedade de substratos (Gupta, Singh, Qazi, & Kumar, 2001) e *Gluconacetobacter diazotrophicus* por ser responsável pela maior parte do nitrogênio fixado em cana-de-açúcar em regiões tropicais e por sua associação endofítica com a cana-de-açúcar representar um sistema modelo para associações monocotiledônia-diazotrófica (Lery, *et al.*, 2008 e Muthukumarasamy, *et al.*, 2002).

#### **1.1.1. *Gluconacetobacter diazotrophicus***

*Gluconacetobacter diazotrophicus* foi a primeira bactéria fixadora de nitrogênio atmosférico (diazotrófica) endofítica isolada da cana-de-açúcar e é muito tolerante a meios ácidos. Pode ser encontrada em raízes, caules e folhas das plantas de cana-de-açúcar, bem como em gramíneas, batata-doce, café, banana e abacaxi. Ela tem distribuição ampla e é responsável por pelo menos metade do nitrogênio fixado que está disponível para a planta hospedeira (Stephan *et al.*, 1991).

Suas células apresentam morfologia semelhante a um bacilo com extremidades arredondadas cujas medidas vão de 0,7 a 0,9  $\mu\text{m}$  por 1 a 2  $\mu\text{m}$  e podem ser vistas ao microscópio como células únicas, aos pares ou ainda na forma de estrutura em cadeia sem endósporos (Muthukumarasamy, *et al.*, 2000).

Antes denominada de *Acetobacter diazotrophicus*, *G. diazotrophicus* pertence à divisão Proteobacteria, subdivisão  $\alpha$ -Proteobacteria, ordem Rhodospirillales e família *Acetobacteraceae*. Atualmente, esta família contém três gêneros fixadores de nitrogênio, compreendendo sete espécies denominadas *Acetobacter nitrogenifigens*, *Gluconacetobacter kombuchae*, *Gluconacetobacter johannae*, *Gluconacetobacter azotocaptans*, *Gluconacetobacter diazotrophicus*, *Swaminathania salitolerans* and *Acetobacter peroxydans* (Kerstens, *et al.*, 2006). Existem ainda nessa família, dois

patógenos humanos: *Asaia bogorensis*, causadora da peritonite e bacteremia (Snyder, *et al.*, 2004) e *Granulibacter bethesdensis* (Greenberg, *et al.*, 2006), associada à doença granulomatosa em humanos.

*G. diazotrophicus* também contribui positivamente na associação com a planta hospedeira, produzindo substâncias promotoras de crescimento (auxinas e giberelinas), bacteriocina (inibidor de crescimento de patógeno, como *Xanthomonas albilineans*) e apresenta atividade antagônica contra a fitotoxina produzida por fungos, como o *Colletotrichum falcatum*, o causador da podridão-vermelha – principal e mais temida doença na cana-de-açúcar (Muthukumarasamy, *et al.*, 2000).

*G. diazotrophicus* é umas das endofíticas estudadas como organismo modelo para avaliar as interações feitas planta hospedeira e bactéria (Cocking *et al.*, 2006). Dentro das diazotróficas já isoladas, *G. diazotrophicus* e *Azoarcus* sp. são as bactérias cujas provas acerca da capacidade de fixação de nitrogênio estão disponíveis atualmente (Saravanan, *et al.*, 2008).

O genoma de *G. diazotrophicus* foi seqüenciado completamente em 2007 por duas iniciativas distintas: uma brasileira liderada pela RioGene e LNCC e outra pelo DOE Joint Genome Institute. Para este trabalho, o genoma utilizado será aquele seqüenciado pelo Consórcio RioGene.

O genoma é composto por 3.944.163 pares de bases dispostos em um único cromossomo circular abrangendo 3.778 genes (Bertalan, *et al.*, 2007). Ele foi depositado no banco de genomas do NCBI referindo-se à *Gluconobacter diazotrophicus* PAI5 em dezembro de 2007 e atualizado em julho de 2008.

### **1.1.2. *Gluconobacter oxydans***

*Gluconobacter oxydans*, previamente conhecida como *Acetobacter suboxydans*, é uma bactéria de morfologia redonda ou oval cujas medidas vão de 0,5 a 0,8 mm por 0,9 a 4,2 mm (Gupta, *et al.*, 2001). O prefixo *oxy* deriva do Latim para ácido e *dans* significa dado. Elas tendem a ter um genoma pequeno por conta de suas habilidades metabólicas limitadas que incluem a oxidação parcial de alcoóis e carboidratos por



meio da fermentação oxidativa e também podem ser utilizadas para a síntese de vitamina C, ácidos D-glucônico e cetoglucônicos (Macauley, 2001).

*G. oxydans* são encontradas em flores, frutas, solo de jardim, bebidas alcoólicas e cidra, pois existem cepas capazes de crescer em meios com altas concentrações de açúcar e baixos valores de pH – o crescimento ótimo se dá entre 5,5 e 6,0. Embora seja capaz de crescer em condições extremas, sua *taxa* de crescimento é baixa e a sua importância se dá pela habilidade de oxidar incompletamente substratos carbônicos para o uso na indústria (Gupta *et al.*, 2001 e Deppenmeier & Ehrenreich, 2009). Esses produtos (aldeídos, cetonas e ácidos orgânicos) são secretados quase que em sua totalidade para o meio. As linhagens de *G. oxydans* não são patogênicas ao homem e outros animais, porém são capazes de causar apodrecimento de frutas, como maçãs e peras, nas quais surgem várias manchas de escurecimento (Prust *et al.*, 2005).

De acordo com a classificação taxonômica atual proposta por Garrity e colaboradores (2001), *Gluconobacter oxydans* pode ser incluída como: Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirales; Acetobacteraceae; *Gluconobacter*; *Gluconobacter oxydans*.

O genoma de *G. oxydans* foi seqüenciado completamente por Prust e colaboradores (2005) e é composto por 2.702.173 pares de bases dispostos em um único cromossomo circular abrangendo 2.432 genes. Ele foi depositado no banco de genomas do NCBI referindo-se à *Gluconacetobacter diazotrophicus* 621H em janeiro de 2005 e atualizado em julho de 2008.

### **1.1.3. *Caulobacter crescentus***

Como membro da subdivisão  $\alpha$  das proteobactérias, a bactéria *Caulobacter crescentus* cresce em ambientes aquáticos. Cada célula apresenta uma diferenciação e divisão assimétrica a cada ciclo, fazendo alusão a diversidade de organismos multicelulares. Trata-se de uma bactéria simples e altamente manipulável, sendo utilizada como sistema modelo de célula única para estudar a diferenciação, divisão assimétrica e sua coordenação com a progressão do ciclo celular (Skerker & Laub, 2004). Existem dois estágios celulares: um cuja bactéria apresenta-se pedunculada e

tem a capacidade de aderir-se a superfícies sólidas e outro estágio de motilidade, no qual a bactéria pedunculada atua como uma célula-tronco que dá origem a células móveis. Desta forma, a produção de células móveis minimiza a competição durante a fase de crescimento bacteriano, garantindo que elas irão colonizar um novo local (Nierman, *et al.*, 2001).

O genoma de *C. crescentus* foi seqüenciado completamente em 2001 (Nierman, *et al.*, 2001) e é composto por 4.016.942 pares de bases dispostos em um único cromossomo circular abrangendo 3.767 genes. Ele foi depositado no banco de genomas do NCBI referindo-se à *Caulobacter crescentus* CB15 em março de 2001 e atualizado em novembro de 2008. Sua classificação taxonômica é: Bacteria, Proteobacteria, Alphaproteobacteria, Caulobacterales, Caulobacteraceae, *Caulobacter*, *Caulobacter crescentus*.

Neste trabalho, os genes de reparo no genoma de *C. crescentus* previamente analisados por Martins-Pinheiro, *et al.* (2007) foram utilizados como guia para as análises realizadas por este. Isso se deveu ao fato de *C. crescentus* também ser uma  $\alpha$ -proteobactéria e por possuir uma base de dados mais acurada que outras bactérias proximamente relacionadas às *G. diazotrophicus* e *G. oxydans*.

## 1.2. Sistemas de Reparo

Diversos fatores são capazes de por em risco a integridade genômica. Toda e qualquer modificação na estrutura química da molécula de DNA é denominada de lesão: tais como quebras duplas ou simples na molécula de DNA, mudança química de bases, formação de sítios abásicos e adutos, ligações cruzadas inter e intra-fita, além do pareamento errôneo das bases são promovidos por mudanças químicas espontâneas e erros no metabolismo normal do DNA. Como consequência, efeitos diretos e indiretos nas células e organismos podem ser detectados, como por exemplo, mutações, recombinação genética, inibição ou alteração de processos celulares, aberrações cromossômicas, tumorigênese e também morte celular (Eisen & Hanawalt, 1999).

A lesão que ocorre na molécula de DNA modifica sua estrutura química, de forma que em um novo ciclo de replicação, a polimerase pode não reconhecer o par correto e assim inserir outro nucleotídeo, levando a uma modificação na seqüência nucleotídica. Essa modificação é denominada mutação, e uma vez instalada será passada para as futuras gerações celulares. Mutações podem ocorrer em nível gênico ou em cromossômico. A mutação gênica é o processo pelo qual os genes são alterados, podendo ocorrer ganho ou perda de função – seja pela substituição, inserção ou deleção de um único par de bases (mutação de ponto) ou de um segmento de pares de bases. Já as mutações cromossômicas são classificadas como alterações numéricas ou estruturais no cromossomo (Friedberg, *et al.*, 1995).

Embora a mutação consista em uma das principais fontes de variabilidade genética, sendo assim importante para a evolução, a estabilidade do material genético também é fundamental para a continuidade da vida. Desta forma, para manter a integridade genômica nas espécies, a atuação de sistemas de reparo de DNA é fundamental. Uma vez tendo a gama de lesões que podem ocorrer ao DNA, não surpreende o fato de existirem também diversas vias celulares capazes de reparar tais danos. Para isso, as funções dos sistemas de reparo incluem desde a correção de erros simples na replicação tendo como consequência a resistência à morte celular causada por agentes danosos (Cline & Hanawalt, 2003).

Os mecanismos de proteção ao material genético podem ser divididos em duas categorias: (i) a prevenção de erros e a (ii) o reparo destes. Na primeira categoria estão enquadrados os mecanismos que existem para evitar que haja reações lesivas ao DNA, como alguns sistemas enzimáticos que neutralizam compostos químicos nocivos antes que eles reajam com o DNA. Nessa classe estão incluídos, por exemplo, os sistemas de detoxicação de radicais superóxido (Michaels & Miller, 1992).

Provavelmente, os mecanismos de reparo de DNA surgiram no início da evolução dos seres vivos e, aparentemente, todas as células vivas apresentam mecanismos enzimáticos capazes de corrigir a maioria dos danos induzidos por diversos agentes mutagênicos. A grande diversidade dos mecanismos de reparo dentro da própria espécie permite a ela reconhecer e reparar muitas lesões, mas também ocorrem redundâncias uma vez que há sobreposição dentro de muitas vias. Já as diferenças intra-específicas mostram que alguns organismos não possuem certas vias de reparo, como é o caso do reparo por fotorreativação presente em espécies como *Escherichia coli*, porém ausente em outras, como em humanos, por exemplo. E, em outros casos, por mais que haja a mesma via de reparo em organismos diferentes, os mecanismos iniciadores e os tipos de lesões reparados por ela irão variar de acordo com a necessidade específica de cada um (Eisen e Hanawalt, 1999).

As diferenças citadas anteriormente tanto em especificidade quanto em tipos de mecanismos de reparo podem ter importantes efeitos biológicos. Características fenotípicas que podem variar entre indivíduos, linhagens e espécies decorrentes de diferenças no reparo de lesões no DNA incluem aumento/diminuição na frequência do surgimento de cânceres, longevidade, patogênese, utilização preferencial de códons e conteúdo GC, *taxas* evolutivas, sobrevivência em ambientes extremos e especiação. E para entender tamanhas diferenças, o estudo das diferenças no reparo é fundamental (Cline & Hanawalt, 2003).

A caracterização dos sistemas de reparo é importante para a compreensão da evolução dos processos e proteínas que os compõem, não só porque o reparo constitui um dos grandes processos celulares, mas sim as informações decorrentes desse estudo

proverão uma perspectiva evolutiva interessante, pois permitirá um foco em como e o porquê similaridades e diferenças surgiram.

1. **Reparo por excisão** (reparo por excisão de bases, reparo por excisão de nucleotídeos e reparo de bases mal emparelhadas);
2. **Reparo direto** (fotorreativação, alquiltransferência e desmetilação oxidativa);
3. **Reparo recombinacional**;
4. **Outras proteínas relacionadas ao reparo** (sistema SOS, ligases e helicases).

### 1.2.1 Reparo por excisão

A maioria dos danos é corrigida por meio do reparo por excisão – um processo no qual a falha em um segmento numa fita de DNA é excisada e então substituída pela síntese do reparo para replicação, utilizando a fita complementar não danificada como molde.

A redundância da informação genética inerente à fita dupla de DNA é essencial para o reparo por excisão bem como para a replicação genômica semiconservativa. A via ubíqua do reparo por excisão de nucleotídeos (NER) lida com uma grande variedade de distorções na dupla-hélice de DNA estruturalmente não relacionadas. NER foi descoberto há 45 anos (Cleaver, 1968) a partir do tratamento do DNA de bactérias com ultra-violeta (UV) e as características básicas do reparo por excisão de base (BER) e do reparo de bases mal emparelhadas (MMR) foram identificadas logo em seguida (Hanawalt & Spivak, 2008).

As seções seguintes darão mais detalhes sobre os mecanismos de reparo por excisão.

#### a. Reparo por excisão de bases (BER)

A via de reparo por excisão de base (BER) tem sido encontrada em todos os organismos vivos e envolve diferentes proteínas com funções de DNA glicosilase e endonucleases (Martins-Pinheiro, *et al.*, 2007). Ela é a principal via para a correção de bases modificadas, sítios nos quais há perda da base, quebras de fita simples e pequenas lacunas no DNA.

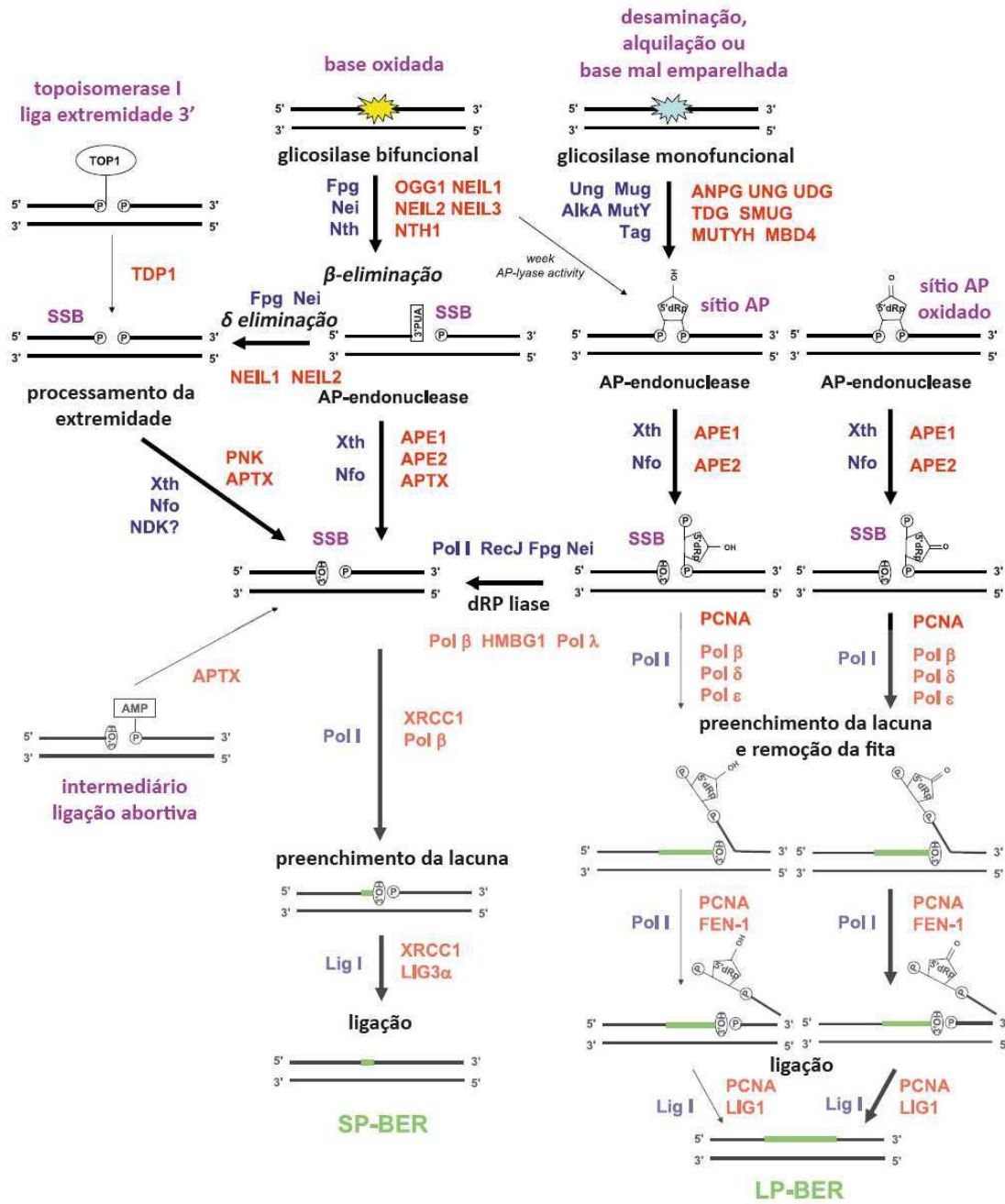
Existem duas enzimas principais para essa via: as DNA glicosilases e AP endonucleases. A DNA glicosilase é utilizada para quebrar a ligação  $\beta$ -N glicosídica, criando um sítio abásico. Assim, a AP endonuclease reconhece esse sítio e cliva o DNA danificado na extremidade 5' do sítio abásico e com isso gera uma ponta 3'-OH. A DNA polimerase sintetiza o DNA a partir da extremidade livre 3'-OH usando sua atividade exonucleásica para preencher o nucleotídeo da base danificada, seguido da atividade da DNA ligase que une a porção nucleotídica recém sintetizada ao DNA (Nickoloff & Hoekstra, 1998).

Ensaio bioquímicos indicaram que BER pode ocorrer como duas sub-vias que são diferenciadas tanto pelo tamanho da lacuna existente como pelas enzimas envolvidas. Essas sub-vias são denominadas como “via BER de único nucleotídeo” ou “via curta (SP)” e “via BER de múltiplos-nucleotídeos” ou “via longa (LP)” (Jaiswal, *et al.*, 2002). Em ambas as vias, o tipo de lesão age como checkpoint na seleção da via de BER (Almeida & Sobol, 2007).

As enzimas que fazem parte da via BER em *E. coli* podem ser encontradas na tabela 1 e um esquema mostrando as vias curta e longa está ilustrado pela figura 1.

**Tabela 1.** Componentes da via de reparo por excisão de bases em *E. coli*.

Genes em <i>E. coli</i>	Descrição
<b>Glicosilases Monofuncionais</b>	<i>alkA</i> Enzima induzível 3-metil-adenina DNA glicosilase II
	<i>mutY</i> Adenina glicosilase
	<i>tag</i> Enzima constitutiva 3-metil-adenina DNA glicosilase I
	<i>ung</i> ou <i>udg</i> Uracil DNA glicosilase
<b>Glicosilases Bifuncionais</b>	<i>fpg/mutM</i> Formamidopirimidina DNA glicosilase
	<i>nei</i> Endonuclease VIII e DNA N-glicosilase com atividade AP liase
	<i>nth</i> Endonuclease III, específica para sítios apurínicos e/ou apirimidínicos
<b>AP endonucleases</b>	<i>nfo</i> Endonuclease IV
	<i>nfi</i> Endonuclease V
	<i>xthA</i> Exodesoxirribonuclease III



**Figura 1.** Vias de reparo por excisão de base curta (SP-BER) e longa (LP-BER). As enzimas em azul correspondem às de procariontes e em vermelho às de eucariontes (adaptado de Krwawicz *et al.*, 2007).

## **b. Reparo por excisão de nucleotídeos (NER)**

O reparo por excisão de nucleotídeos (NER) é um mecanismo versátil, complexo e consiste numa série de reações enzimáticas capazes de remover um grande espectro de alterações estruturalmente não relacionadas na fita dupla de DNA, pois esse mecanismo utiliza endonucleases mais genéricas.

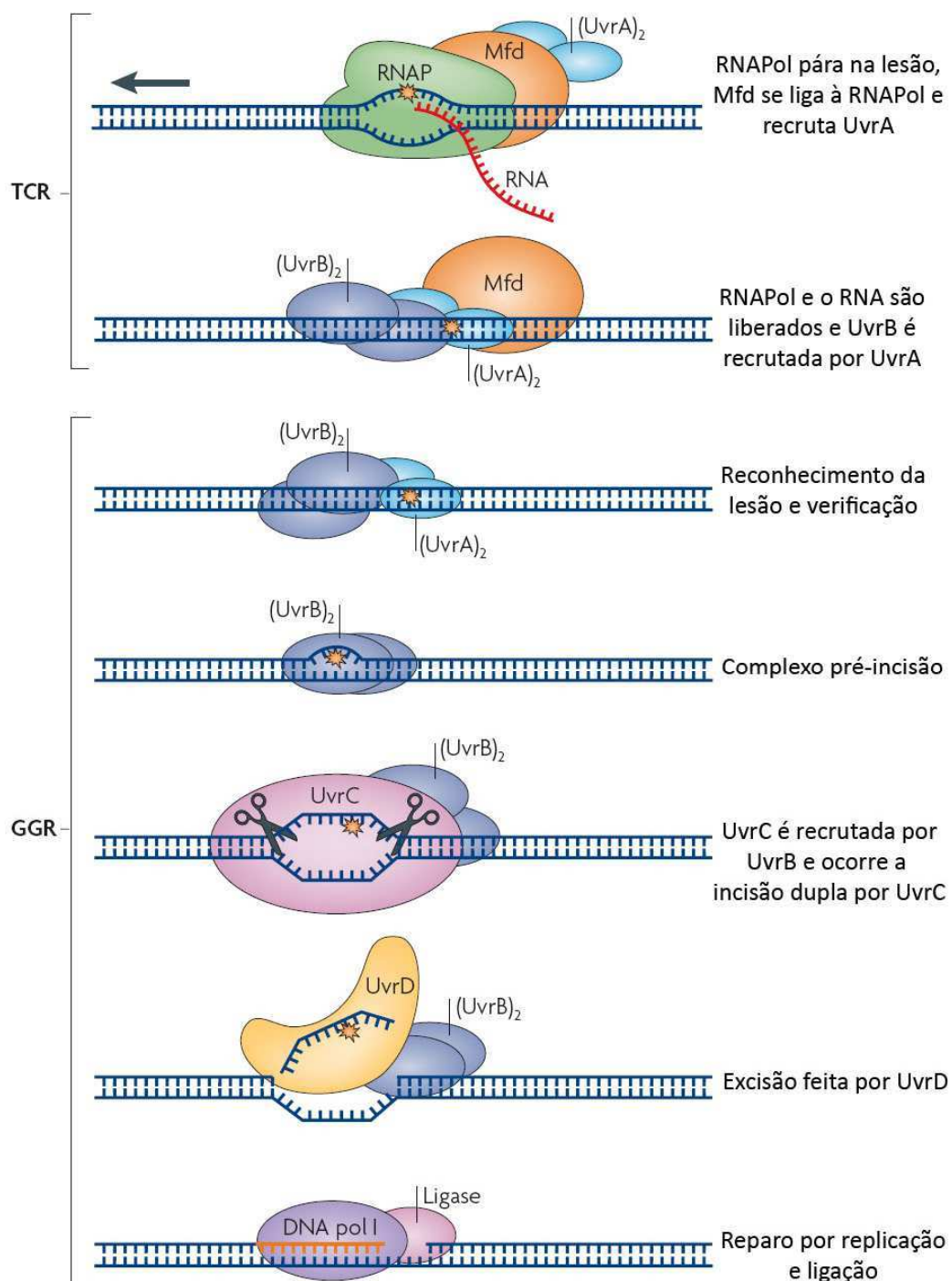
Ilustrado pela figura 2, NER consiste no reconhecimento da lesão no DNA, seguido pelas incisões feitas em ambos os lados da fita danificada e após a remoção do pequeno pedaço da fita de DNA contendo a lesão, há nova síntese nucleotídica utilizando a fita de DNA não danificada como molde e finalmente a ligação do novo oligonucleotídeo em 3' com a fita existente. A via NER pode ser dividida em duas subvias: o reparo genômico global (GGR) e o reparo acoplado à transcrição (TCR) (Hanawalt & Spivak, 2008).

No reparo genômico global em *E. coli*, o primeiro passo é o reconhecimento da lesão feito pelo complexo UvrA<sub>2</sub>. Em seguida, UvrB liga-se ao complexo UvrA<sub>2</sub>+DNA, para aumentar a interação. O grupo UvrA<sub>2</sub>B transloca-se para o sítio danificado. Ocorre a dissociação de UvrA<sub>2</sub> e o recrutamento de UvrC para ligar-se ao complexo UvrB-DNA. Uma incisão dupla na fita de DNA é realizada por UvrBC. Ela se dá no quarto ou quinto nucleotídeo a 3' da lesão e no sétimo nucleotídeo a 5' da lesão. Os sítios catalíticos 3' e 5' da lesão se encontram na endonuclease UvrC. UvrD, também chamada de DNA helicase II, desloca UvrC e separa as fitas para liberar a porção nucleotídica contendo o dano. Por fim, a DNA polimerase I preenche a lacuna na fita simples e a DNA ligase sela a incisão (Verhoeven, 2000).

Já outra subvia de NER remove lesões das fitas molde de DNA de genes ativamente transcritos. Neste caso, genes expressos são escaneados por RNA polimerases, que detectam de forma sensível o DNA danificado e iniciam o reparo acoplado à transcrição (TCR) (Hanawalt & Spivak, 2008). Os dímeros de pirimidina ciclobutano (CPD) localizados na fita molde constituem um bloqueio absoluto para a transcrição (Nickoloff & Hoekstra, 1998). Em bactérias, a parada da RNA polimerase impede imediatamente o reconhecimento do CPD pelos fatores UvrA e UvrB. Entretanto, a RNA polimerase está ligada ao fator de reparo acoplado à transcrição



Mfd, que a libera e recruta UvrA. A partir daí, a via transcorre identicamente ao reparo via NER (Hanawalt & Spivak, 2008).



**Figura 2.** Desenho esquemático do sistema de reparo de nucleotídeos. TCR refere-se ao Reparo Acoplado à Transcrição, na qual há a participação de Mfd. Em seguida no reparo genômico global (GGR), UvrA é o elemento primário para o reconhecimento da lesão (adaptado de Hanawalt & Spivak, 2008)

### c. Reparo de bases mal emparelhadas (MMR)

O DNA não danificado é normalmente replicado de maneira muito precisa devido à grande seletividade replicativa da DNA polimerase e sua habilidade exonucleásica 3' para 5' de excisar erros que são produzidos durante a replicação (Kunkel & Erie, 2005). Porém, quando esses erros não são reconhecidos e corrigidos pelo sistema de verificação da polimerase, a responsabilidade para tal tarefa recai sobre o MMR. Ele age como um editor, removendo e corrigindo pares de bases errados introduzidos no DNA de forma restaurar o genótipo parental (Stojic, *et al.*, 2004).

MMR é altamente conservado entre quase todos os organismos e está muito bem descrito em *E. coli* (Kunkel, 2008). Mal emparelhamentos na fita dupla de DNA podem ser reparados principalmente por três sistemas: 1) Reparo via *very short patch* (VSR), que é considerado uma variante do MMR geral; 2) Reparo dependente de MutY, glicosilase de BER que reconhece alguns pareamentos errados; e 3) Reparo de bases mal emparelhadas direcionado pela DNA adenina metiltransferase (Dam). Os primeiros dois sistemas têm especificidade restrita aos substratos (emparelhamentos errôneos T•G e A•G, respectivamente), enquanto que o terceiro sistema é capaz de reparar 11 de 12 possibilidades de mal emparelhamento de bases, considerando recalcitrante apenas o par errôneo C•C (Nickoloff & Hoekstra, 1998).

Em *E. coli*, o DNA é metilado na posição N6 dos resíduos de adenina nas seqüências dGATC, porém a fita-filha recém replicada é transientemente não metilada nessas seqüências. São essas seqüências hemimetiladas GATC que permitem ao reparo identificar na fita nascente em qual local há um erro (Li, 2005). Três proteínas são essenciais para detectar pareamentos errados e dirigir a maquinaria de reparo ao local do erro. Elas são MutS, MutH e MutL. MutS forma um dímero (MutS<sub>2</sub>) que reconhece a base errada na fita filha e se liga ao DNA. Em seqüência, MutL<sub>2</sub> é recrutado e forma com MutS<sub>2</sub> um complexo juntamente ao heterodúplex de DNA (complexo ternário). A montagem desse complexo ativa a endonuclease latente MutH. O sítio 5' ou 3' ao local do erro é localizado e submetido à ação combinada de MutS, MutL, MutH e ATP. MutH

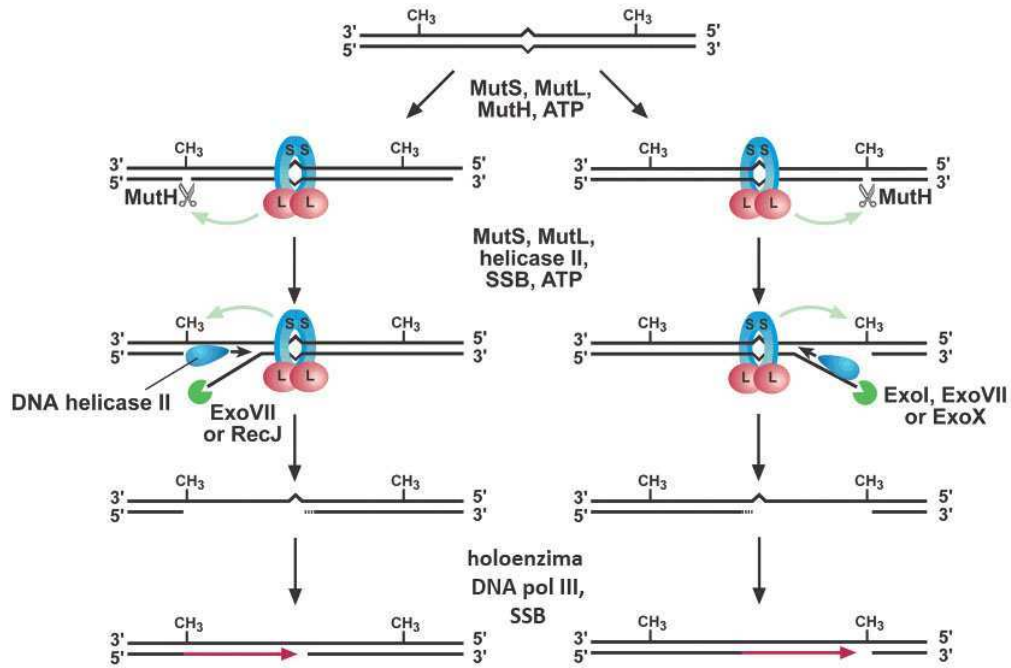
faz uma incisão no sítio hemimetilado dGATC e esta incisão fita-específica é o ponto de partida para a excisão da base mal pareada.

MutS e MutL ativam UvrD que logo alcança a incisão e desenrola o duplex a partir dela em direção ao erro, gerando uma fita simples de DNA. Nela as proteínas de ligação à fita simples (Ssb) rapidamente ligam-se e a protegem do ataque de nucleases. Dependendo da posição da quebra relativa ao local do erro, ExoI ou ExoX (exonucleases 3'→5'), ou ainda ExoVII ou RecJ (exonucleases 5'→3') excisam a fita cortada a partir do sítio de clivagem dGATC até um pouco depois do erro (Li, 2008). Tanto o sítio no qual há o erro quanto os nucleotídeos ao redor dele são completamente excisados. A fenda criada pela exonuclease pode ser então reparada pela DNA polimerase III – com a ajuda proteína de ligação à fita-simples (Ssb) – que usa a outra fita como molde, e em seguida, unida pela ligase. Por fim, Dam metilase rapidamente metila a fita filha (Kunkel & Erie, 2005).

#### **i. Reparo *very short patch* (VSP)**

Em *E. coli*, pareamentos errôneos T/G resultantes de uma desaminação da 5-metilcitosina para timina são convertidos para um pareamento correto CG por um processo conhecido como reparo *very short patch* (VSP).

O processo é iniciado pela ação de uma única endonuclease, Vsr, que corta em 5' da timina mal colocada. DNA polimerase I remove um pequeno fragmento de nucleotídeos a 3' da incisão e então completa a lacuna, inserindo uma citosina oposta à guanina. DNA ligase completa o procedimento de reparo selando novamente a fita de DNA (Monastiriakos, *et al.*, 2004).



**Figura 3.** Mecanismo do reparo de bases mal emparelhadas direcionado por metilação em *E. coli*. Embora não mostrado, a DNA ligase restaura a ligação covalente da fita reparada após a holoenzima DNAPol III preencher a lacuna. As setas verdes indicam sinalização dependente de MutS e MutL entre os dois sítios de DNA envolvidos na reação (adaptado de Iyer *et al.*, 2006).

### 1.2.2. Reparo direto

A maioria dos danos ao DNA é reparada por processos nos quais há remoção de bases danificadas seguida por uma nova síntese da região excisada. Porém, algumas lesões no DNA podem ser reparadas pela reversão direta do dano, o que se torna uma forma mais eficiente de lidar com tipos de danos específicos de ocorrência freqüente no DNA – tais como, dímeros de pirimidina decorrentes de exposição à luz UV e resíduos de guaninas alquiladas modificadas pela adição de grupamentos metil ou etil na posição O<sup>6</sup> do anel purínico (Goosen, *et al.*, 2008).

Os três principais mecanismos do reparo direto são: a fotorreativação, o reparo de bases alquiladas – alquiltransferência e, o mais recente identificado, a desmetilação oxidativa.

#### a. Fotorreativação

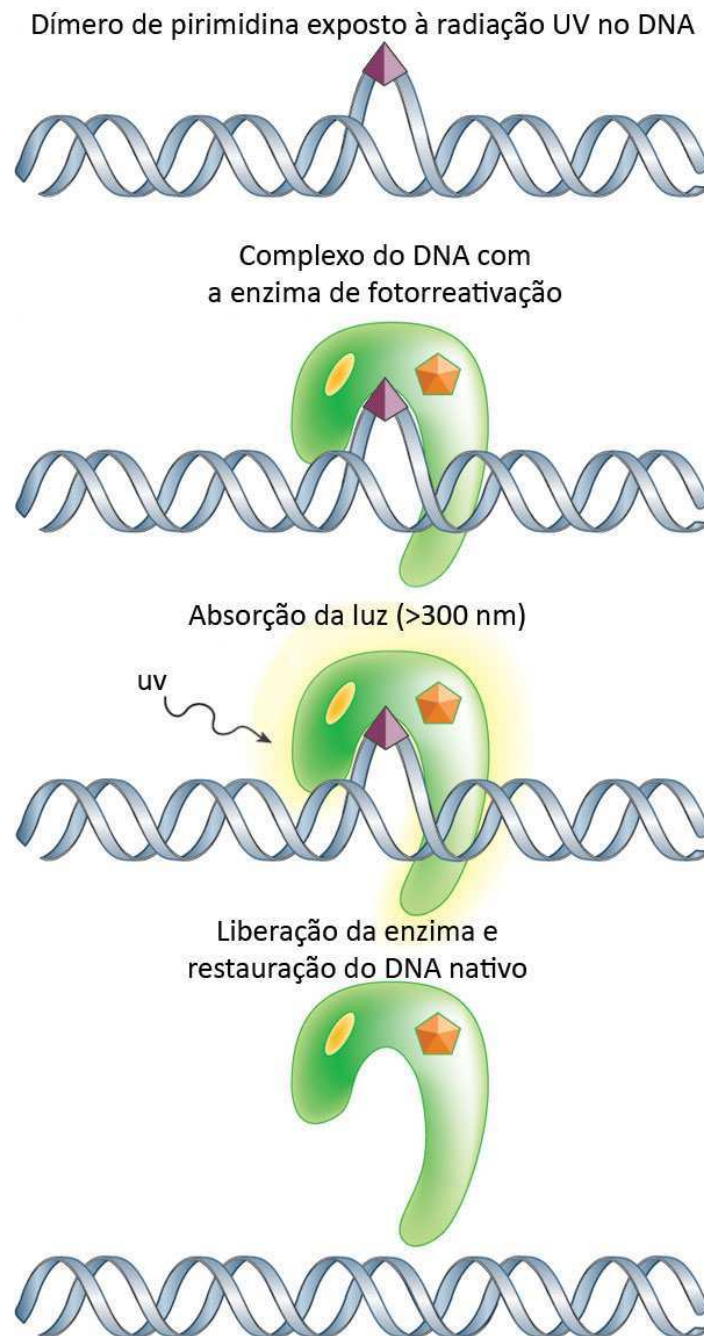
Os fotoprodutos são os principais alvos do reparo por fotorreativação. Dentre eles estão os dímeros de ciclobutanos (CPD) e as lesões 6,4 – pirimidina 6-4 pirimidona. A formação desses dímeros causa uma grande distorção na estrutura do DNA, impedindo que processos como a replicação e transcrição ocorram. Com o bloqueio, o reparo do DNA nesses organismos está proximamente relacionado com sua habilidade das células sobreviverem à radiação UV (Goosen & Moolenaar, 2008).

A fotorreativação surge como um mecanismo de reversão direta dos danos por meio de uma reação que utiliza a luz visível de 300 a 500 nm para corrigir a dimerização.

Encontrada de bactérias a mamíferos aplacentários, a enzima fotoliase é uma flavoproteína que possui dois cromóforos: um responsável por captar o fóton e transferir a energia ao outro, que a utiliza para quebrar a ligação covalente entre as duas pirimidinas. O reconhecimento da lesão pela enzima se dá na ausência de luz. Com o fim do processo, a pirimidina original permanece no DNA, agora restaurada ao seu estado normal (Sancar, 2008).

A radiação solar UV é uma das maiores fontes de danos ao DNA para diversos tipos celulares e o reparo de dímeros de pirimidina pela fotorreativação é comum a uma variedade de células procarióticas e eucarióticas. Entretanto, esse

mecanismo de reparo não é universal e muitas espécies, como a humana, não o possuem (Goosen & Moolenaar, 2008).



**Figura 4.** O DNA exposto à radiação ultravioleta (uv) resulta na dimerização covalente das pirimidinas adjacentes, ilustradas aqui como triângulos roxos. Essas lesões são reconhecidas pela fotoliase (verde), que absorve luz com comprimento de onda menor que 300 nm e monomeriza os dímeros formados.

### **b. Alquiltransferência**

Esta modalidade de reparo direto lida com o dano resultante de reações entre agentes alquilantes e o DNA. Esses agentes são compostos reativos que podem transferir grupos metil ou etil às bases de DNA, desta forma modificando-as. Um tipo particular de dano é a metilação na posição O<sup>6</sup> da guanina, pois o seu produto O<sup>6</sup>-metilguanina (O<sup>6</sup>metG) forma uma base complementar com a timina ao invés da citosina (Morohoshi, *et al.*, 1990).

Essa lesão pode ser reparada pela O<sup>6</sup>-metilguanina metiltransferase que transfere o grupo metil da O<sup>6</sup>metG para o resíduo de cisteína em seu sítio ativo. Essa reação não enzimática suicida faz com que a enzima perca sua função e a potencial modificação química mutagênica é então removida e a guanina original, restaurada (Morohoshi, *et al.*, 1990).

Em *E. coli*, os produtos dos genes *ada* e *ogt* são as proteínas que exercem essa atividade. O gene *ada* de *E. coli* apresenta dois domínios, integra o regulon Ada e em sua região N-terminal está a região responsável pela alquiltransferência. Quando Ada recebe um grupo metil, ela sofre mudança conformacional e se torna o regulador transcricional do seu próprio promotor e do operon Alk – formado pelos genes *alkA* e *alkB*. Esse processo levará ao início do reparo de danos por alquilação do DNA, ativando outras vias (Drabløs, *et al.*, 2004).

De forma geral, as enzimas que catalisam o reparo direto estão distribuídas tanto em procariotos quanto em eucariotos, incluindo em humanos.

### **c. Reparo oxidativo direto**

Begley e Samson (2003) descreveram um mecanismo que consistia numa terceira via de remoção direta dos danos por alquilação. O gene *alkB* – que faz parte do operon *alk* regulado pelo produto do gene *ada* – atua em uma reação oxidativa, liberando o grupo metil em formaldeído, revertendo desta forma a lesão para a base não modificada.

### 1.2.3. Reparo recombinacional

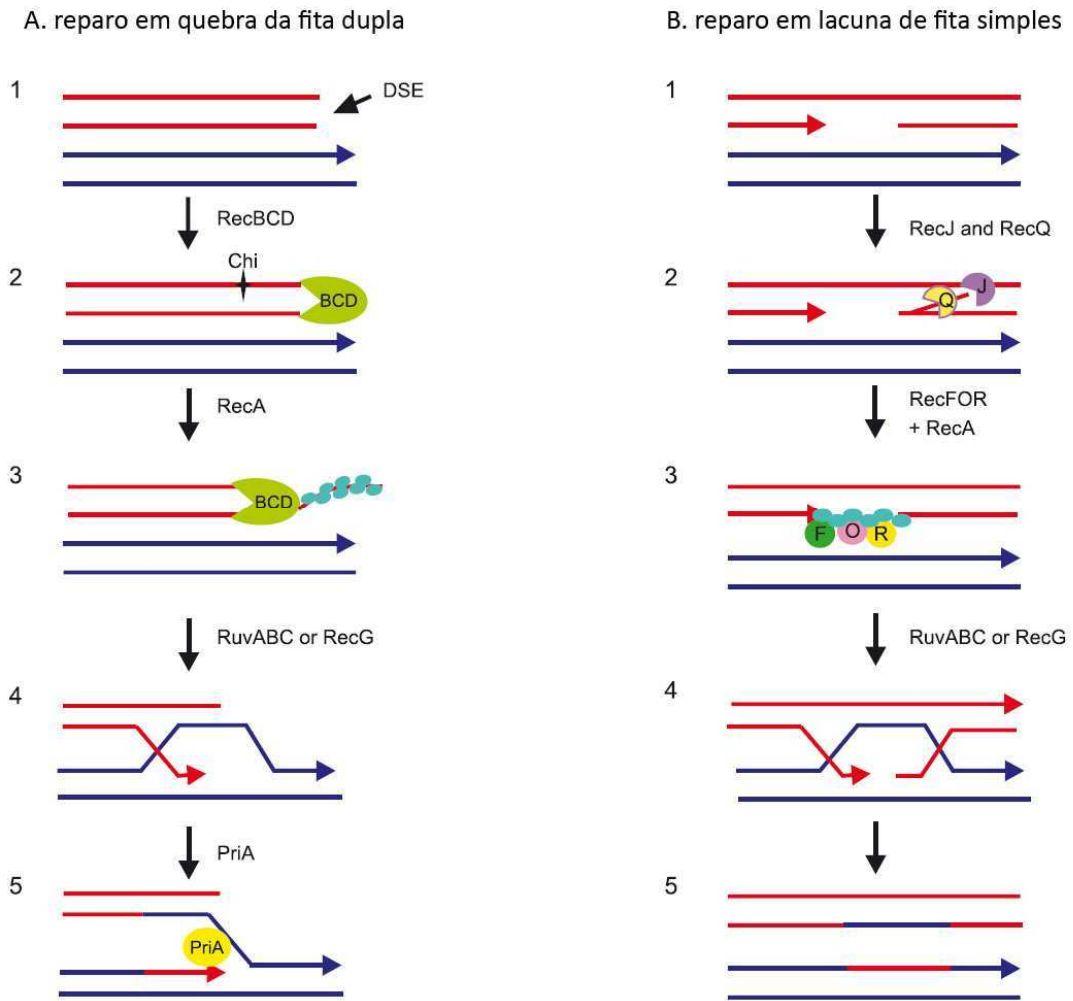
A resolução de quebras duplas na fita do DNA (DSB), ligações cruzadas entre fitas e proteínas ligadas covalentemente requerem diferentes tipos de vias de reparo que agem em conjunto: essas incluem a recombinação homóloga, anelamento de fita simples e *end-joining* não homólogo. A recombinação também é essencial para restaurar a continuidade genômica das forquilhas de replicação em colapso que ocorrem quando a replicação encontra uma quebra simples (SSB) ou dupla na fita de DNA parental (Hanawalt & Spivak, 2008).

A recombinação homóloga (figura 5) consiste de três estágios que são comuns aos procariotos e eucariotos: (i) pré-sinapse – onde as DSB ou a lacuna são formadas e a extremidade resultante no DNA é preparada para a recombinação; (ii) sinapse – fase na qual a conexão física entre o substrato recombinogênico e o duplex de DNA homólogo intacto é gerado levando à formação de moléculas heteroduplex (híbrido); e (iii) pós-sinapse: a síntese do DNA a partir da região 3' invasora ocorre seguida pela resolução dos intermédios da junção (Nowosielska, 2007).

Em *E. coli*, existem dois mecanismos principais de recombinação homóloga: a via RecBCD, que repara DSB, e a via RecFOR, para quebras em fita simples. Ambas requerem a recombinase RecA para o reconhecimento de homologia e troca de fitas (Rocha, *et al.*, 2005).

RecA é uma proteína multifuncional envolvida na regulação da resposta SOS e na atividade das polimerases que facilita a replicação do DNA quando há lesão. Em *E. coli*, o conjunto de moléculas formadas por RecA são resolvidos tanto pelo complexo RuvABC bem como por uma ação da helicase RecG de forma ainda não conhecida. As proteínas RuvAB e RuvC catalisam a migração do braço e a resolução da junção Holliday, respectivamente. Essas três proteínas possivelmente interagem num complexo *resolvase*, no qual RuvABC varrem o DNA, com RuvC sendo capaz de buscar seqüências cliváveis. Finalmente, o cruzamento entre as fitas homólogas é desfeito e a recombinação mediada pelo complexo RecBCD é sempre acoplada com o reinício da replicação dependente da proteína PriA (Nowosielska, 2007 e Rocha, *et al.*, 2005).





**Figura 5.** Recombinação homóloga. A – reparo de quebras duplas na fita de DNA e B – reparo de uma lacuna na fita simples (adaptado de Nowosielska, 2007).

#### 1.2.4. Outros mecanismos de reparo

Existem ainda diversas proteínas que são importantes para o reparo do DNA. Dentre elas estão as que compõem o sistema SOS, as ligases e as helicases.

##### a. Resposta SOS

No fim da década de 60, Miroslav Radman encontrou descobertas que o levaram a concluir que em *E. coli* havia um sistema de reparo de DNA dependente das proteínas LexA e RecA, ativado quando o DNA se encontrava severamente danificado com sua síntese bloqueada; e a indução desse sistema estava conectada com a indução de mutações. Radman o denominou de reparo SOS e replicação SOS fazendo alusão ao envio do sinal SOS em código Morse por um telégrafo internacional: três pontos, três traços, três pontos (Bridges, 2005). Trabalhos posteriores confirmaram a hipótese de Radman e outros sistemas que se assemelhavam à resposta SOS de *E. coli* foram encontrados em células eucarióticas e também em outras bactérias, porém as repostas bacteriana e eucarionte são substancialmente diferentes (Janion, 2008).

O regulon SOS é um conjunto de respostas fisiológicas aos danos no DNA e, em *E. coli*, o sistema envolve a participação de mais de 40 genes (Courcelle, *et al.*, 2001). Muitos produtos desses genes são utilizados para reparar o DNA com BER, NER ou o reparo recombinacional. Os genes do sistema SOS também estão envolvidos em desencadear a divisão celular que ocorre apenas após a replicação total do genoma, assegurando a divisão celular (Schlacher, *et al.*, 2006).

Quando as células bacterianas sofrem grandes danos no DNA suficientes para induzir a resposta SOS, o principal evento consiste no colapso das forquilhas de replicação assim que elas encontram uma grande quantidade de lesões incluídas à fita molde de DNA. O objetivo primordial do sistema SOS é reiniciar produtivamente a replicação.

Os genes de resposta SOS são encontrados distribuídos por todo o cromossomo de *E. coli* como genes únicos situados em operons únicos. Seis deles *umuDC* (juntos montam a Pol V), *ruvAB* (catalisam a migração do ramo de DNA nas estruturas de Holliday) e *ysdAB* (de função desconhecida) são codificados por pares

de genes que fazem um operon. Geralmente, apenas um box SOS está presente em um operon. As exceções são os genes *lexA* e *ydjM* que contêm dois boxes cada (separados por uma e duas bases, respectivamente) e *recN* contendo três boxes SOS. As seqüências dos boxes em um gene são diferentes (Janion, 2008).

Em condições normais, o produto do gene *lexA* reprime a expressão dos genes SOS, ligando-se aos promotores destes. Uma das propostas para a ativação da resposta SOS diz que quando há uma situação de estresse elevado associada à intensa atuação do reparo de excisão e, como consequência, a geração de oligonucleotídeos de DNA fita simples, a proteína RecA é ativada, e assim, atua como coprotease, facilitando a clivagem autocatalítica da proteína repressora LexA. Com a saída do repressor LexA, os promotores dos genes do sistema SOS são expressos (Schlacher, *et al.*, 2006).

Uma vez ativado o regulon SOS, a proteína UmuD – codificada pelo gene *umuD* – é expressa e, na forma de dímero se complexa com *UmuC* – produto do gene *umuC* – formando assim, a DNA polimerase V. Dá-se início à síntese trans-lesão. O complexo UmuD<sub>2</sub>C atua como polimerase e é capaz de incorporar um nucleotídeo frente à lesão e permite a continuidade da replicação. DNA-Pol V continua a síntese, porém como apresenta baixa processividade, é logo substituída por DNA-Pol III. A interação que ocorre entre as subunidades catalíticas da DNA-Pol III e o complexo UmuD<sub>2</sub>C, é de grande importância para que ocorra a síntese trans-lesão (Goodman, 2002).

Embora não seja uma proteína de reparo, MutT, membro da superfamília Nudix-hidrolase, é importante juntamente com MutM e MutY (ambas de BER) compõe o sistema de reparo GO. Esse sistema é o responsável por prevenir mutações advindas da principal lesão oxidativa ao DNA – a 2' desoxi-7,8-diidro-8-oxoguanosina (8-oxodG ou GO) que resulta de transversões G→T (Nickoloff & Hoekstra, 1998). MutT é a responsável por hidrolisar 8-oxodG, evitando sua incorporação durante a síntese de DNA (Michaels & Miller, 1992).

### **b. Ligases**

A correta ligação das extremidades do DNA é um passo crítico em muitas reações fundamentais na célula, incluindo replicação, recombinação e reparo do DNA. Uma classe específica de enzimas, as DNA ligases, desempenha a maior parte dessas reações de ligação das extremidades.

As DNA ligases unem quebras no esqueleto fosfodiéster das moléculas de DNA e são usadas em muitas reações essenciais dentro da célula. Todas as DNA ligases seguem o mesmo mecanismo de reação, mas elas podem usar tanto ATP quanto  $\text{NAD}^+$  como cofator. As eubactérias contêm DNA ligases dependentes de  $\text{NAD}^+$  e algumas delas possuem genes putativos para DNA ligases dependentes de ATP. Estas últimas são mais proximamente relacionadas às Arqueas e aos vírus. Análises filogenéticas sugerem que todas as DNA ligases dependentes de  $\text{NAD}^+$  são proximamente relacionadas, mas as dependentes de ATP devem ter sido adquiridas por genomas bacterianos em um número de ocasiões distintas. (Wilkinson e Bowater, 2001)

### **c. Helicases**

As helicases mudam a estrutura do DNA e RNA para permitir o acesso da maquinaria transcricional. O gene *lhr* codifica uma helicase grande e sua distribuição restrita nas bactérias e ampla difusão em Arqueas sugere que esse gene pode ser uma aquisição recente por meio de transferência horizontal.

Outra helicase e ATPase DNA-dependente é a proteína DinG, um homólogo para helicases humanas que são conhecidas por participar no reparo do DNA e mutações nela têm sido relacionadas a algumas doenças genéticas (Martins-Pinheiro, 2007).

### 1.3. Comparação gênica e filogenia

Inicialmente, os pesquisadores usavam abordagens diretas para comparar genomas em termos de seqüências. Esses métodos buscavam por (i) homólogos, motivos (como de ligação ao ou regulação do DNA) comuns a oligonucleotídeos e oligopeptídeos (Karlin, *et al.*, 1996); (ii) ortólogos (Tatusov, *et al.*, 1997); (iii) duplicações gênicas (Riley, 1997) e (iv) a ocorrências de famílias conservadas em genomas diferentes (Clayton, *et al.*, 1997).

Vários métodos semi ou totalmente automatizados também têm sido desenvolvidos para comparar seqüências genômicas inteiras com bases de dados múltiplas, que compreendem desde a simples comparação estrutural usando a seqüência primária das proteínas às comparações considerando os valores da expressão do mRNA. Cada um desses métodos fornece informações únicas no que diz respeito ao genoma e como eles são comparados ou se contrastam. No entanto, a comparação genômica baseada na seqüência e estrutura protéica é vantajosa, pois se tratam de estruturas bem conservadas entre os organismos e a relação entre estrutura e função é bem definida (Das, 2001).

Adicionalmente às análises baseadas na função e nas seqüências, diversos estudos genômicos têm avaliado os genomas em nível de sistemas, mais especificamente análises filogenéticas e de vias metabólicas. No caso das vias metabólicas, elas estão organizadas em grupamento de seqüências de proteínas e suas análises podem também revelar informações sobre a fisiologia do organismo. É possível agrupar genomas baseado no critério de compartilhamento ou não de uma via metabólica específica e, com isso, os pesquisadores que trabalham com genomas microbianos têm criado mapas metabólicos abrangentes (Bono, *et al.*, 1998). As vias metabólicas também podem ser comparadas em nível de propriedade de enzimas e modos elementares (Galperin & Koonin, 1998 e Dandekar, *et al.*, 1999). Diversas bases de dados metabólicas estão disponíveis na Internet. A base de dados KEGG – *Kyoto Encyclopedia of Genes and Genomes* – tem uma coleção de mais de 300 vias metabólicas de referência. EcoCyc ([www.ecocyc.org](http://www.ecocyc.org)) e EcoGene ([www.ecogene.org](http://www.ecogene.org)) possuem uma base de dados específica para *E. coli*, apresentando informações detalhadas sobre as vias conhecidas nessa bactéria.

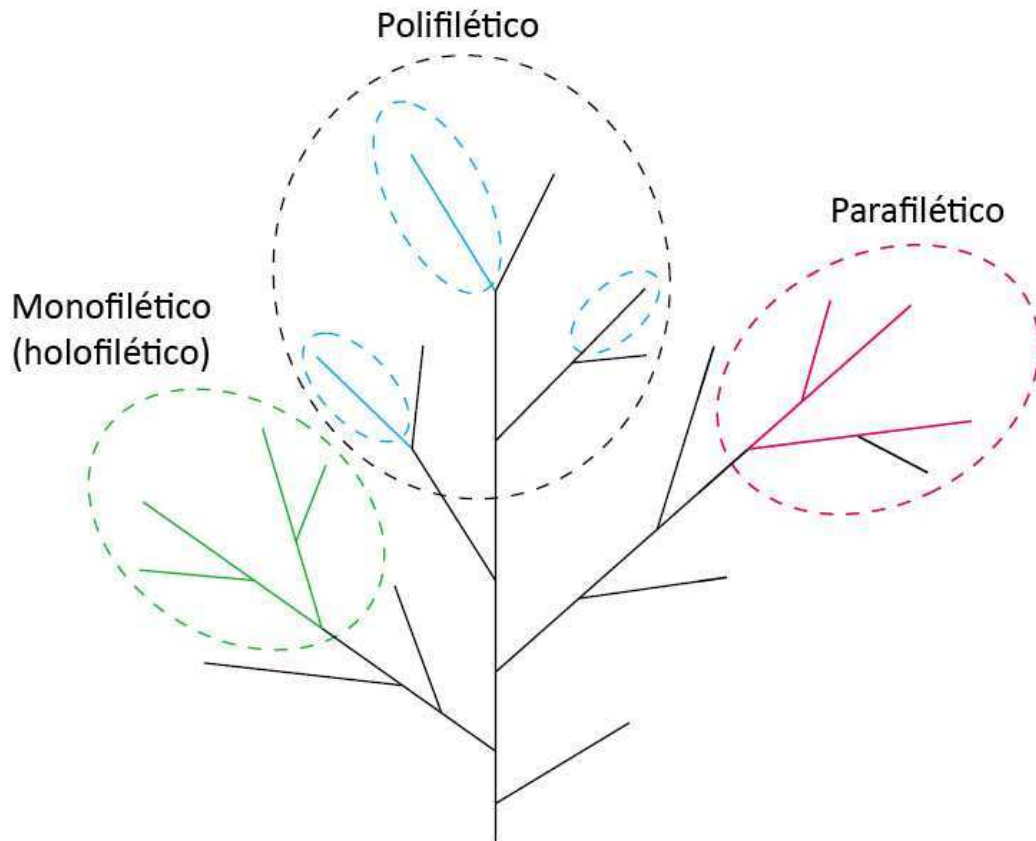
Existem outras grandes áreas de estudo nas quais os genomas são comparados em níveis de sistema, tais como a comparação filogenética, análises de expressão em relação a várias funções celulares, localização e eventos.

A Filogenética é a ciência que estima o passado evolutivo, no caso da filogenia molecular, baseado na comparação das seqüências de DNA ou proteínas. A idéia de representar essa hipótese como árvores provavelmente data da época de Darwin, mas o uso de cálculos numéricos e métodos quantitativos, ambos associados para isso, é relativamente recente e sua aplicação em dados moleculares ainda mais (Baldauf, 2003). A abordagem da filogenética molecular aponta para uma organização a partir da qual é possível relacionar objetivamente distintos organismos e, com isso, interpretar as correntes evolutivas bem como as maquinarias metabólicas que constituem a diversidade dos organismos.

### **1.3.1. Grupos**

Um nó e tudo mais que surge a partir dele é um clado ou um grupo monofilético (figura 5). Este é um grupo natural; todos os membros são derivados de um único ancestral em comum (com base no resto da árvore) e tem um conjunto de caracteres comuns únicos intrínsecos a ele. Um grupo excluindo alguns de seus descendentes é um grupo parafilético (como, por exemplo, animais excluindo humanos).

A distinção entre ortólogos – genes que surgiram via evento de especiação – dos parálogos – genes que surgiram via duplicação dentro do mesmo genoma – é crítica para a biologia comparativa. Isso se deve ao fato que a ortologia é a base na qual a função molecular em um organismo pode ser mais bem inferida a partir de dados experimentais em outros organismos modelos. Genes ortólogos são descendentes de um único gene no último ancestral comum de duas espécies. Eles provavelmente compartilham a função conservada do gene ancestral em comum (Goodstadt, 2006).



**Figura 6.** Grupos organizados em clados. Grupos monofiléticos compartilham um ancestral comum. Um grupo é dito parafilético se ele contém um ancestral comum recente, porém não contém todos os descendentes desse ancestral. Já um grupo polifilético é aquele descendente de mais de um ancestral, podendo ser mais proximamente relacionado a outros organismos fora do grupo taxonômico. (imagem adaptada de Baldauf, 2003)

### 1.3.2. Árvores filogenéticas

Uma árvore filogenética é um diagrama que se assemelha a uma árvore e retrata relações evolutivas entre diferentes genes ou organismos. As árvores objetivam reunir grupos e todas têm ramos (braços) e nós. Nós externos são freqüentemente chamados de unidades operacionais taxonômicas (OTU) e estão nos pontos terminais onde a história evolutiva é passível de verificação. Os nós internos são os pontos onde os braços separam-se uns dos outros e podem ser chamados de unidades taxonômicas hipotéticas para enfatizar que são os progenitores hipotéticos das OTUs, também referidas como o último ancestral comum de tudo o que surge a partir dele (Baldauf, 2003).

As árvores podem ser compostas por famílias multigênicas (árvores gênicas) ou por um único gene vindo de muitas *taxa* (árvores de espécies, pelo menos teoricamente) ou pela combinação das duas. No primeiro caso os nós internos correspondem a eventos de duplicação gênica, no segundo, a eventos de especiação. Uma árvore não enraizada apenas posiciona *taxa* individualmente relativos a outros sem indicar a direção do processo evolutivo. Em árvores enraizadas, a raiz representa um ancestral em comum de todas as seqüências analisadas e, desta forma, implica na ordem dos braços no resto da árvore. Uma árvore pode ser enraizada caso uma ou mais OTUs formem um ponto externo de referência e, são conhecidas (ou acredita-se) como sendo as mais distantemente relacionadas das OTUs – como um grupo externo (Gevers & Coenye, 2005).

### **1.3.3. Análise comparativa**

Para uma análise comparativa é necessário estudar seqüências homólogas – que compartilham uma origem comum – sendo essa a premissa básica para a análise filogenética. Além das bases de dados já descritas, como KEGG, EcoCyc e Ecogene, existe o *Entrez* do NCBI (*National Center for Biotechnology Information* – [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

Na reconstrução da filogenia a partir de alinhamentos de seqüências, não existem métodos unicamente corretos; muitos métodos existem e raramente é possível verificar aquele que obteve a verdadeira árvore filogenética. Existem basicamente quatro metodologias primárias para a construção de filogenias a partir do alinhamento de seqüências de ácidos nucleicos e proteínas: (i) o método baseado em distância *Neighbour-Joining* (NJ); (ii) o método baseado no caráter incluindo a Máxima Parcimônia (*Maximum Parsimony* – MP); (iii) Máxima Verossimilhança (*Maximum Likelihood* – ML) e (iv) Inferência Bayesiana.

#### **a. Métodos baseados em distância**

Eles são relativamente simples e diretos; começando a partir de um alinhamento de seqüências, as diferenças entre todas as combinações de pareamento das OTUs (seqüências de DNA ou de proteínas) são convertidas numa matriz de



distância que representa uma distância estimada entre as seqüências, aplicando um modelo denominado de substituição ou evolucionário. Essas distâncias são então montadas em uma árvore. Uma desvantagem dos modelos de distâncias é que eles reduzem a informação filogenética a um único número. A maior vantagem é que eles exigem muito menos trabalho computacional, o que é importante quando muitos *taxa* têm que ser comparados.

#### **b. Métodos baseados no caráter**

Os métodos baseados em caráter examinam cada coluna do alinhamento separadamente – cada posição na seqüência alinhada é denominada caráter – e procura por árvores que melhor ajustam toda essa informação.

Métodos de máxima parcimônia (MP) selecionam a árvore que explica os dados observados em termos de um número mínimo de substituições. E em muitos casos há diversas árvores, com diferenças sutis, que são consistentes com o mesmo número de eventos e, por conseqüência, igualmente parcimoniosas. Esse método se tornou menos popular principalmente pelo fato que a MP geralmente utiliza um modelo simples de substituição de seqüências (todas as mudanças são igualmente prováveis). Adicionalmente, tem sido mostrado que para seqüências mais divergentes, MP é sujeito a recuperar árvores incorretas (Kolaczowski & Thornton, 2004).

Os métodos de máxima verossimilhança (ML) buscam identificar a árvore mais provável e única numa base estatística, dado o modelo da evolução da seqüência. Este é um método que permite a correção de eventos mutacionais múltiplos entre as seqüências que têm sido separadas há muito tempo ou que evoluem rapidamente. Em ML, todas as vias mutacionais possíveis que são compatíveis com os dados são consideradas e a árvore que tem a maior das probabilidades de produzir as seqüências observadas é preferida (Huelsenbeck, 1995).

O principal obstáculo de ML é o processamento computacional, como, por exemplo, o número de diferentes topologias de árvores que têm que ser avaliadas aumenta enormemente em função do número de seqüências. Mesmo com adaptações do método ML às mais modernas tecnologias computacionais, construir árvores com mais de quarenta seqüências torna-se muito impraticável (Gevers & Coenye, 2005).

A estimação Bayesiana (BE) da filogenia é baseada em uma quantidade denominada probabilidade posterior da distribuição de árvores, na qual é a probabilidade de uma árvore condicionada nas observações. O condicionamento é obtido usando o teorema de Bayes. A definição matemática pode ser encontrada em no trabalho de Huelsenbeck e colaboradores (2001).

#### **1.3.4. *Bootstrap* – Análise de confiabilidade**

Uma forma comum de calcular a robustez da topologia de uma árvore é pelo método de confiabilidade não paramétrica (Felsenstein, 1985). A análise de confiabilidade (*bootstrap*) testa o quanto o conjunto de dados completo está dando suporte à árvore, ou se a árvore é apenas uma vencedora marginal entre muitas alternativas proximamente iguais. Na prática, isso segue: a partir do alinhamento original, colunas no alinhamento das seqüências são escolhidas randomicamente, até um novo alinhamento ser construído com o mesmo tamanho do original. Para cada grupo de dados, uma árvore é construída e comparada com a árvore baseada no alinhamento original. O número de vezes que um agrupamento, assim definido na árvore original, é também encontrado na árvore de *bootstrap*, os valores resultantes do *bootstrap* são armazenados e sobrepostos na árvore original.

Uma dificuldade com esta análise é a interpretação precisa de quais valores o *bootstrap* representa (Soltis & Soltis, 2003). Quanto maiores os valores, melhor e alguns cientistas concluíram que os valores acima de 70% parecem ser indicadores confiáveis dos agrupamentos (Hillis, *et al.*, 1993). Porém esses valores nem sempre são aceitos (Felsenstein & Kishino, 1993).

A maior vantagem da técnica de *bootstrap* é que ela pode ser aplicada a basicamente todos os três métodos de construção de árvores, entretanto deve-se ter em mente que aplicar o método do *bootstrap* multiplica o tempo computacional necessário pelo número de replicações requeridas.

Os métodos de análise podem levar ao erro não importando a quantidade de dados se tem. Entretanto não existe garantia que se pode produzir uma árvore com a topologia correta. Com relação à confiabilidade na topologia da árvore resultante, algumas sugestões devem ser observadas:

- a. Aplicar diferentes métodos na construção da árvore no banco de dados;
- b. Usar modelos evolucionários diferentes para a construção da matriz;
- c. Adicionar ou remover uma ou mais OTUs para ver se há influência na topologia da árvore;
- d. Aplicar a análise por *bootstrap* ao banco de dados.

Uma árvore deve ser considerada robusta e então confiável apenas quando diferentes métodos abrangentes inferirem topologias idênticas ou similares e tais topologias sejam corroboradas por bons valores de *bootstrap* (Gevers & Coenye, 2005).

#### **1.4. Justificativa**

Mesmo com grande número de genomas seqüenciados, poucos são os organismos que tiveram seus sistemas de reparo identificados. *E. coli* surgiu como um modelo inicial e, partir daí, outras análises foram feitas. Alguns exemplos são *Chromobacterium violaceum* ( $\beta$ -proteobactéria de vida livre), *Caulobacter crescentus* ( $\alpha$ -proteobactéria aquática e de vida livre), *Mycoplasma sinoviae* (patógeno responsável pela gripe aviária e que apresenta genoma reduzido) e *Deinococcus radiodurans* (bactéria extremófila radiorresistente). Mesmo com toda essa informação, pouco se sabe sobre o reparo em  $\alpha$ -proteobactérias e menos ainda sobre estes sistemas nas bactérias acido-acéticas.

Tendo em vista a abordagem da genômica comparativa como uma ferramenta capaz de dar pistas sobre características metabólicas e fisiológicas e considerando a importância estabilidade do genoma que é garantida em grande parte pelos sistemas de reparo de DNA, esse trabalho surge com a proposta de detalhar as principais vias de reparo do DNA das bactérias acido-acéticas *G. diazotrophicus* e *G. oxydans*.