



**UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS AMBIENTAIS E BIOLÓGICAS
TRABALHO DE CONCLUSÃO DE CURSO - TCC**

**DIVERSIDADE ESTRUTURAL E EVOLUÇÃO DA TERCEIRA E
SEXTA ETAPAS DA BIOSÍNTESE DE PURINAS EM PROCARIOTOS**

FENÍCIA BRITO SANTOS

FENÍCIA BRITO SANTOS

**DIVERSIDADE ESTRUTURAL E EVOLUÇÃO DA TERCEIRA E
SEXTA ETAPAS DA BIOSÍNTESE DE PURINAS EM PROCARIOTOS**

Trabalho de Conclusão de Curso
apresentado para obtenção do título de
bacharel em Ciências Biológicas.

Cruz das Almas
2011

S231 Santos, Fenícia Brito

Diversidade estrutural e evolução da terceira e sexta etapas da biossíntese de purinas em procariotos./ Fenícia Brito Santos._ Cruz das Almas - BA, 2011.
84 f.; il.

Orientador: Phelippe Arthur Santos Marbach

Monografia (Graduação) – Universidade Federal do Recôncavo da Bahia. Centro de Ciências Agrárias, Ambientais e Biológicas.

1. Genômica 2 Genética-Evolução I. Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias, Ambientais e Biológicas. II. Título.

CDD: 575.1

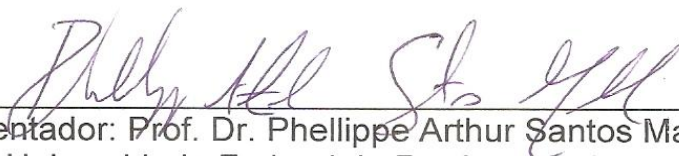
FOLHA DE APROVAÇÃO

Fenícia Brito Santos

DIVERSIDADE ESTRUTURAL E EVOLUÇÃO DA TERCEIRA E SEXTA ETAPAS DA BIOSÍNTESE DE PURINAS EM PROCARIOTOS

Aprovado em: 14 de Julho de 2011

BANCA EXAMINADORA



Orientador: Prof. Dr. Phellippe Arthur Santos Marbach
Universidade Federal do Recôncavo da Bahia



Membro Titular: Prof. Dra. Flora Maria de Campos Fernandes
Prof^a. da Universidade Federal da Bahia



Membro Titular: Prof. Dr. Jorge Teodoro de Souza
Prof^o. da Universidade Federal do Recôncavo da Bahia

Dedicatória

Dedico este trabalho a minha mãe, que sempre me incentivou e apoiou em todas as escolhas da minha vida. Foi você que segurando a minha mão me ensinou a escrever minhas primeiras palavras. Mas, mais que isso, você me ensinou a escrever minha própria história. Te amo.

Agradecimentos

Aos meus pais, Oderiana e Agnaldo pela dedicação e todos os sacrifícios que fizeram durante toda minha vida para que eu pudesse realizar os meus sonhos.

A minha irmã, Dalma, que sempre me ajudou nos momentos mais difíceis.

Aos meus familiares pelo apoio incondicional em todos os momentos dessa jornada. Pela compreensão e suporte nos maus momentos. E por vibrar comigo nos momentos de alegria.

À Gisele, Índira, Kaliane, Manu, Leo e Tio que se tornaram a minha segunda família. E a todos os outros companheiros de curso, que sempre me deram forças e que fizeram com que esse 4 anos e meio se tornassem inesquecíveis.

Ao meu orientador e amigo, Phellippe Arthur Santos Marbach pela dedicação extraordinária. Seus ensinamentos serão para sempre lembrados.

À Universidade Federal do Recôncavo da Bahia pela formação.

À todos os mestres pelo conhecimento que adquiri.

À Thatyane Lima, minha companheira de estágio, por todos os momentos e conhecimentos compartilhados.

À Diogo Ferreira Alves pela imensa ajuda no início desse trabalho.

Ao Profº Jorge Teodoro de Souza que disponibilizou os computadores da Pós-Graduação para que eu iniciasse as pesquisas.

"A utopia está lá no horizonte. Me aproximo dois passos, ela se afasta dois passos. Caminho dez passos e o horizonte corre dez passos. Por mais que eu caminhe, jamais alcançarei. Para que serve a utopia? Serve para isso: para que eu não deixe de caminhar"

(Eduardo Galeano)

RESUMO

A via biossintética de purinas (VBP) produz Inosina monofosfato em uma série de 10 reações enzimáticas. Esse nucleotídeo é precursor dos ácidos nucléicos, e outras moléculas de importância biológica. Na terceira etapa da VBP o composto GAR (Glicinamida ribonucleotídeo) é convertido em FGAR (N-Formilglicinamida ribonucleotídeo). Essa reação pode ser realizada por duas enzimas PurN ou PurT, que usam diferentes mecanismos e substratos. Na sexta etapa da VBP, o composto AIR (Aminoimidazol ribonucleotídeo) é convertido em CAIR (Carboxiaminoimidazol ribonucleotídeo), por dois mecanismos distintos. Em procariotos, plantas e fungos duas enzimas, PurK e PurE I, fazem a conversão de AIR em CAIR utilizando como substrato HCO_3^- e ATP. A mesma reação pode ser realizada pela PurE II que utiliza CO_2 como substrato. Essa reação é descrita somente para animais. A existência de dois mecanismos de reação na terceira e sexta etapas da VBP remete a questões sobre a evolução destas etapas, assim como a origem e distribuição destes genes nas linhagens procarióticas. Os resultados desse trabalho mostraram que *purN* é mais distribuído que *purT* nos filos e classes amostrados. A distribuição encontrada para *purT*, pode ser reflexo da intensa transferência horizontal. A genômica comparativa mostrou que existem espécies que possuem os genes *purE* e *purK*, enquanto outros possuem somente *purE*. A filogenia molecular mostrou que a *purE* dos organismos que não possuem *purK*, formam um clado contendo *purEs* de Archaeas, bactérias Gram positivas, Gram negativas e *purE* II funcionalmente caracterizadas. Portanto, estes resultados sugerem que a PurE II existe em procariotos e está amplamente distribuída. A filogenia da *purK* mostra a formação de dois grandes clados. Em um dos clados os genes estão organizados na ordem *purE/purK*, enquanto no outro predomina a ordem *purK/purE*. Esses dados dão uma nova perspectiva no entendimento da sexta etapa da VBP, visto que indicam que a distribuição da *purE* II em procariotos é maior que o reportado na literatura de forma que o modo de vida dos organismos pode ter determinado o tipo de reação a ocorrer na sexta etapa da biossíntese de purinas.

Palavras-chave: Genômica comparativa, contexto genômico, operon egoísta, filogenia molecular, vias biossintéticas.

ABSTRACT

The biosynthetic purines pathway produces inosine monophosphate in a series of 10 enzymatic reactions. This nucleotide is a precursor of nucleic acids and other molecules of biological importance. In the third step of the purine pathway the compound GAR (Glycinamide ribonucleotide) is converted into FGAR (N-Formylglycinamide ribonucleotide). This reaction can be carried out by two enzymes, PurN or PurT, which use different mechanisms and substrates. In the sixth step of the purine biosynthetic pathway, the compound AIR (Aminoimidazole ribonucleotide) is converted into CAIR (Carboxyaminoimidazole ribonucleotide), by two distinct mechanisms. In prokaryotes, plants and fungi two enzymes, PurK and PurE I, make the conversion of AIR to CAIR using ATP and HCO_3^- as substrates. The same reaction can be performed by PurE II that uses CO_2 as a substrate. This reaction is described in literature only for animals. The existence of two reaction mechanisms in the third and sixth steps of the purines biosynthetic pathway brings attention to issues about the evolution of these steps as well as the original distribution of these genes in prokaryotic lineages. The results of the present work showed that *purN* is more distributed than *purT* on the sampled phyla and classes. The distribution found for *purT* may be a result of the intense horizontal transfer. Comparative genomics has shown that there are species with *purE* and *purK* genes, while others have only *purE*. The *purK* phylogeny shows the formation of two major clades. In one of the clades genes are arranged in the order *purE/purK*, while in the other predominates the order *purK/purE*. Phylogenetic analysis showed that *purE* in the organisms that do not have *purK* form a clade containing *purEs* from Archaea, Gram positive bacteria, Gram negative and *purE* II functionally characterized. The data from this work provide a new perspective in understanding the sixth step of the purines biosynthetic pathway, as they indicate that *purE* II is more widespread in prokaryotes than previously reported in the literature, what can lead to the conclusion that the organism's way of life may have determined the kind of reaction occurring in the sixth step of the purines biosynthesis.

Key words: Comparative genomics, genomic context, selfish operon, molecular phylogeny, biosynthetic pathway.

LISTA DE FIGURAS

- Figura 1. Elementos doadores dos átomos que formam os anéis da purina. Fonte: Modificado de ZHANG et al., 2008. 17
- Figura 2. Via biossintética de purinas. Fonte: ZHANG et al., 2008. 18
- Figura 3. Modelos de evolução de vias metabólicas. (A) Retro-evolução; (B) Evolução direta. Fonte: Modificado de SCHMIDT et al., 2003. 24
- Figura 4. Modelos de evolução de vias metabólicas. Fonte: Modificado de SCHMIDT, 2003. 25
- Figura 5. Filogenia das vias metabólicas evidenciando que as vias biossintéticas de nucleotídeos foram, possivelmente, as primeiras rotas a surgirem. Fonte: CAETANO-ANOLLES et al., 2007. 27
- Figura 6. Terceira reação da via biossintética de purinas. Fonte: KAPPOCK et al., 2000. 29
- Figura 7. Distribuição de *purN* e *purT* nas 271/473 gêneros/espécies que possuem a via biossintética de purinas. 35
- Figura 8. Contexto genômico de *purN* e *purT* dos 271/473 gêneros/espécies que possuem a via biossintética de purinas. Em 184/321 gêneros/espécies *purN* aparece em um contexto genômico com outros genes da via biossintética de purinas. O contrário ocorre com o gene *purT*, em 91/170 gêneros/espécies esse gene é encontrado em um contexto genômico sem associação com outros genes da biossíntese de purinas. Ao lado esta representado o contexto genômico de *purN* e *purT*. As caixas vazias representam genes não relacionados com a biossíntese de purinas. 38
- Figura 9. Sobreposição da estrutura terciária da PurT e PurK de *Escherichia coli*. A sobreposição foi realizada utilizando o programa PyMol (DeLano et al., 2002). 40
- Figura 10. Cenário para evolução da PurT baseada na hipótese de recrutamento enzimático. Nesse cenário PurK que já atuava na síntese de purinas e duplicou. Uma das isoformas continuou com a sua função original e a outra, que deu origem a PurT, sofreu especiação e passou a atuar em outra etapa da via biossintética de purinas. 41
- Figura 11. Cenário da evolução das enzimas PurT e PurK. Esse cenário sugere que a enzima ancestral de PurT e PurK estava presente no ancestral de todos os seres vivos. Essa enzima dotada de ampla especificidade duplicou e as isoformas dessas enzimas se especificaram em diferentes substratos e passaram a atuar em diferentes etapas da via. 42
- Figura 12. Árvore filogenética com sequências de PurT e PurK. A árvore mostra que PurK (preto) e PurT (vermelho) formam clados distintos apesar de serem enzimas homólogas. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). 42
- Figura 13. Árvore filogenética não enraizada de PurT. Os clados que formam um grupo monofilético em nível de filo ou classe, foram colapsados. O nome filo/classe e o número de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). A identificação da sequência é composta pela abreviação do filo/classe a qual a espécie pertence, mais o número de identificação da sequência seguido pelo nome da espécie com o gênero abreviado. 44
- Figura 14. Sub-clado da árvore de PurT. **(A)** Espécies do gênero *Bifidobacterium* agrupando com Gram negativas. Os ramos em vermelho destacam as espécies de *Actinobacteria*. **(B)** Espécies de Firmicutes Bacilli que agrupam com espécies de Deltaproteobacteria. Os ramos em vermelho destacam as espécies da Classe Bacilli. O nome filo/classe e o número de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). 45

- Figura 15. Sub-clados da árvore de PurT. **(A)** Espécies do gênero *Micobacterium* agrupando com Betaproteobacteria. **(B)** Espécies do gênero *Corinebacterium* agrupando com espécies da classe Gamaproteobacteria. Os ramos em vermelho destacam as espécies de *Actinobacteria*. O nome filo/classe e o número de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). 46
- Figura 16. Sexta reação da via biossintética de purinas. Fonte: ZHANG et al., 2008. 51
- Figura 17. A filogenia da PurE mostra que a maioria das espécies que possuem somente PurE agrupam com as sequências de PurE II de Eucariotos formando o clado II (vermelho). Os outros clados são compostos por espécies que possuem PurE e PurK formando o clado I (preto). A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). 57
- Figura 18. Clado II da árvore filogenética da figura 20. Neste clado esta a maioria das espécies procarióticas que não possuem PurK. E as PurEs II de eucariotos (vermelho). A espécie *Tropheryma whipplei str Twist* e as Cianobactérias presentes nesse clado possuem PurK (preto). O nome filo/classe e o numero de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). 58
- Figura 19. Árvore não enraizada da PurE destacando as sequências que possuem alta taxa de divergência e por isso foram excluídas das análises posteriores. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). 59
- Figura 20. No clado I são encontradas espécies de organismos que possuem PurE e PurK. As exceções (vermelho) são espécies que possuem apenas PurE e que agrupam nesse clado possivelmente por causa da atração de ramos longos. O nome filo/classe e o numero de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). 62
- Figura 21. Sobreposição da estrutura da PurE II de *Homo sapiens* com as PurEs I das bactérias *Yersinia pestis*, *Escherichia coli*, *Bacillus anthracis*, *Thermotoga marítima*, e da Archaea *Methanocaldococcus jannaschii*, que não possui PurK. Essa sobreposição mostra que as PurEs I e II apresentam alto nível de semelhança estrutural. A sobreposição foi realiza com o programa PyMol (DeLano et al., 2002). 66
- Figura 22. Contexto genômico dos gêneros/espécies que possuem apenas *purE*, mostrando que esse gene pode ser encontrado com outros genes da via ou formando um operon onde não há outros genes da biossíntese de purinas. 67
- Figura 23. O contexto genômico mostra que das 203/359 gênero/espécies possuem *purE* e *purK*. Em 15/244 gêneros/espécies estes genes formam um operon sem outros genes da via, e em 29/83 gêneros/espécies esses genes estão em um contexto com outros outros genes da biossíntese de purinas. Em 22/32 gêneros/espécies *purE* forma um operon sem outros genes da biossíntese de purinas, e 21/31 gêneros/espécies é *purK* que esta em um contexto sem outros genes da via. Em apenas uma espécie o gene *purK* forma um operon com outro gene da via que não seja *purE*. 68
- Figura 24. Árvore filogenética de PurK não enraizada. No clado I (preto) são encontradas espécies que possuem o contexto *purE I/purK*. As espécies que compõem o clado II (vermelho) possuem o contexto *purK/purE I*. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). 69

LISTA DE TABELAS

Tabela 1. Filos e classes dos Domínios Archaea e Bacteria amostradas no estudo. Entre parênteses esta destacado o número total de Filos e Classes (EUZÉBY, 1997).	32
Tabela 2. (A) Distribuição e padrão de ocorrência dos genes <i>purN</i> e <i>purT</i> em procariotos. As três primeiras colunas indicam os padrões estruturais encontrados. Nas outras colunas estão listados os filios/classes de procariotos, indicando o número de gêneros/espécies que cada filo possui, os asteriscos destacam os padrões mais representativos. Amarelo indica presença do gene e branco ausência.	36
Tabela 3. Espécies que possuem <i>purT</i> em um contexto genômico com outros genes da biossíntese de purinas.	38
Tabela 4. Valores cinéticos das enzimas PurN e PurT de <i>Escherichia coli</i> .	47
Tabela 5. Relação de bactérias patogênicas para animais e plantas que possuem o gene <i>purT</i> .	48
Tabela 6. Distribuição e padrão de ocorrência dos genes <i>purE</i> e <i>purK</i> em procariotos. As duas primeiras colunas representam a ocorrência dos genes <i>purE</i> e <i>purK</i> . Nas outras colunas estão listados os filios de procariotos, indicando o número de gêneros/espécies que cada filo possui. Amarelo indica presença do gene e branco ausência.	54
Tabela 7. Espécies que possuem duas cópias do gene <i>purE</i> , uma delas agrupando no clado II e a outra no clado I. Abaixo do contexto são apresentados os números de acesso (GI) das <i>purEs</i> dessas espécies.	60
Tabela 8. Valores cinéticos das enzimas PurE II, PurE I e PurK.	64
Tabela 9. Espécies que foram capazes de crescer com o gene <i>purK</i> mutado.	64
Tabela 10. Tipos de contexto genômico de <i>purK</i> nos diferentes gêneros/espécies. Os filios/classes em branco não possuem <i>purK</i> .	68

SUMÁRIO

1.INTRODUÇÃO	15
2.JUSTIFICATIVA	16
3.REVISÃO BIBLIOGRÁFICA	17
3.1. <i>Via Biossintética de Purinas</i>	17
3.2. <i>Origem das vias biossintéticas</i>	23
4.OBJETIVO GERAL	27
CAPÍTULO I - DIVERSIDADE ESTRUTURAL E EVOLUÇÃO DO TERCEIRO PASSO DA BIOSSÍNTESE DE PURINAS EM LINHAGENS PROCARIÓTICAS	28
5.INTRODUÇÃO	29
6.OBJETIVO ESPECÍFICO	30
7.METODOLOGIA	30
7.1. <i>Busca pelos genes purN e purT da via biossintética de purinas</i>	30
7.2. <i>Análise da diversidade estrutural da terceira e etapa da via biossintética de purinas</i>	31
7.3. <i>Análise de anticorrelação</i>	33
7.4. <i>Contexto genômico</i>	33
7.5. <i>Alinhamento múltiplo e filogenia do gene purT</i>	33
8.RESULTADOS e DISCUSSÃO	34
8.1. <i>Distribuição de purN e purT nas linhagens procarióticas</i>	34
8.2. <i>Contexto genômico de purN e purT nas linhagens procarióticas</i>	35
8.3. <i>Genômica comparativa de purN e purT nas linhagens procarióticas</i>	37
8.4. <i>Análise de anticorrelação dos genes purN e purT</i>	38
8.5. <i>Filogenia da enzima PurT</i>	42
9.Conclusões	49

CAPÍTULO II - DIVERSIDADE ESTRUTURAL E EVOLUÇÃO DO SEXTO PASSO DA BIOSÍNTESE DE PURINAS EM LINHAGENS PROCARIÓTICAS 50

10.	INTRODUÇÃO	51
11.	OBJETIVO ESPECÍFICO	52
12.	METODOLOGIA	52
	12.1. <i>Busca pelos genes canônicos da via biossintética de purinas</i>	52
	12.2. <i>Análise da diversidade estrutural da sexta etapa da via biossintética de purinas</i>	52
	12.3. <i>Contexto genômico</i>	53
	12.4. <i>Alinhamento múltiplo e filogenia das enzimas purK e purE.</i>	53
13.	RESULTADO E DISCUSSÃO	53
	13.1. <i>Genômica comparativa de purE e purK nas linhagens procarióticas</i>	53
	13.2. <i>Filogenia de PurE</i>	56
	13.3. <i>Contexto genômico de purE e purK</i>	66
	13.4. <i>Filogenia de PurK</i>	68
14.	Conclusão	70
	REFERÊNCIAS	71
	ANEXO	78

1. INTRODUÇÃO

A via biossintética de purinas é um conjunto de reações que tem como produto final o IMP (Inosina Monofosfato), precursora de compostos purínicos e derivativos como ATP, GTP, cAMP, cGMP coenzima A, NAD, NADP e FAD. Estes compostos desempenham funções estruturais, carreadoras de energia, sinalizadoras e metabólicas essenciais. Em plantas, as purinas são precursores de citocininas (uma classe de fitorreguladores), de compostos envolvidos na fixação simbiótica e estocagem do nitrogênio e de alcalóides de grande importância comercial como teobromina e cafeína (SMITH et al., 2002).

Como os níveis de expressão dos genes da síntese de novo de purinas são maiores durante a duplicação do DNA, as enzimas da via biossintética de purinas também são importantes alvos para fármacos anti-câncer ou antimicrobianos (KIRSCH e WHITNEY, 1991). Além disso, o mau funcionamento das enzimas desta via causam importantes doenças hereditárias na espécie humana (DURAN et al., 1997).

Os 13 genes que catalisam reações na síntese de IMP são, *purF*, *purD*, *purN*, *purT*, *purL*, *purM*, *purE*, *purK*, *purC*, *purB*, *purH*, *purP* e *purO*. Destes, cinco possuem relação de analogia, codificando para proteínas que não são homólogas, mas são funcionalmente equivalentes (ZHANG et al., 2008). Os seguintes compostos são utilizados como doadores de átomos ou grupos funcionais para construção do IMP: Glicina, Glutamina, CO₂, Aspartato e N¹⁰-formiltetrahidrofolato (ZHANG et al., 2008).

As reações químicas da terceira e sexta etapas da via biossintética de purinas podem ser catalisadas por diferentes enzimas, dependendo do conjunto gênico presente no genoma do organismo. Na terceira etapa, o composto GAR (Glicinamida ribonucleotídio) é formilado e convertido a FGAR (N-Formilglicinamida ribonucleotídio). A reação pode ser catalisada pelas enzimas PurN ou PurT, que utilizam diferentes mecanismos (SAXILD et al., 1995; ZHANG et al., 2008; LI et al., 2009; NAGY et al., 1993). No entanto, a distribuição dessas enzimas nas linhagens procarióticas é desconhecida.

O mesmo ocorre durante a carboxilação do composto AIR (Aminoimidazol ribonucleotídio) para formar o CAIR (Carboxiaminoimidazol ribonucleotídio), na sexta etapa. Segundo dados encontrados na literatura em organismos superiores a reação

é catalisada pela enzima PurE classe II, enquanto em organismos inferiores, a PurK faz a carboxilação no grupamento amino na posição C4 do anel purínico e subsequentemente a enzima PurE classe I, com atividade mutase, transfere o grupo carboxila para a posição C5. Assim como na terceira etapa, a ocorrência dos genes que codificam as enzimas PurE I e PurK nas diferentes linhagens procarióticas é desconhecida (HOSKINS et al., 2007; MUELLER et al., 1994; DRAZEK et al. 1995; CHUNG et al., 1996; SORENSEN et al., 1997; FIRESTINE et al., 1998; THODEN et al. 1999; MEYER et al., 1992; CHEN et al., 1990; FIRESTINE e DAVISSON et al., 1994).

A evolução de vias biossintéticas esta diretamente relacionada com a evolução de procariotos e eucariotos e, com o advento da era genômica muitos aspectos sobre evolução das vias biossintéticas nestas linhagens celulares estão sendo elucidadas, permitindo responder algumas questões evolutivas relevantes sobre origem e diversificação da vida. Por exemplo, Morett e colaboradores (2003) e Rodionov e colaboradores (2002) usaram análises de genômica comparativa, para encontrar novos genes relacionados com a biossíntese de tiamina. Estes resultados contribuíram para ampliar o conhecimento sobre a diversidade estrutural da via biossintética de tiamina nas linhagens procarióticas e eucarióticas.

Apesar do vasto conhecimento bioquímico sobre a biossíntese de purinas, pouco se sabe sobre a sua diversidade estrutural e história evolutiva das enzimas envolvidas no processo. Contudo, a associação dos conhecimentos sobre diversidade estrutural, evolução e bioquímica da biossíntese de purinas constitui uma ferramenta poderosa no desenvolvimento de estudos que visem o controle de doenças infecciosas causadas por fungos e bactérias e para o tratamento do câncer.

2. JUSTIFICATIVA

O conhecimento atual sobre a diversidade estrutural da via biossintética de purinas em procariotos é limitado a trabalhos desenvolvidos em alguns organismos modelos, portanto, pode não refletir a diversidade existente nas linhagens procarióticas conhecidas.

À luz do conhecimento atual, a terceira etapa da via biossintética de purinas, envolve a participação dos genes *purN* ou *purT* enquanto que a sexta etapa pode ser executada pelo produto dos genes *purK/purE I* ou *purE II*. Contudo, a distribuição

destes genes nas linhagens procarióticas é uma incógnita, apesar dessa informação ter relevância básica e aplicada.

O conhecimento gerado sobre a distribuição e evolução destes genes nos organismos procarióticos poderá ampliar o conhecimento sobre a origem e diversificação da vida no planeta Terra. Por exemplo, inferir sobre o cenário que norteou os primeiros passos da evolução da via biossintética de purinas e sobre os fatores que determinaram os primeiros caminhos percorridos pela vida primitiva.

As enzimas da biossíntese de purinas são importantes alvos para fármacos anti-câncer ou antimicrobianos (KIRSCH e WHITNEY, 1991), portanto, conhecer a distribuição e diversidade dessas enzimas também poderá ter relevância para a saúde humana, orientando o desenvolvimento de novas abordagens na produção de fármacos que possuam tais enzimas como alvos.

3. REVISÃO BIBLIOGRÁFICA

3.1. Via Biossintética de Purinas

O metabolismo de purinas é uma rota complexa que consiste em três processos, sendo dois deles de síntese e um de degradação. A Via de Síntese “de novo” produz os nucleotídeos a partir dos precursores metabólicos – ribose (PRPP), aminoácidos, CO_2 e NH_3 . As vias de recuperação reutilizam as bases livres e os nucleotídeos liberados na quebra dos ácidos nucléicos, com menor custo energético (Figura 1). A Via de Degradação de nucleotídeos desempenha um papel central no metabolismo do nitrogênio (BERTHOMÉ, 2008).

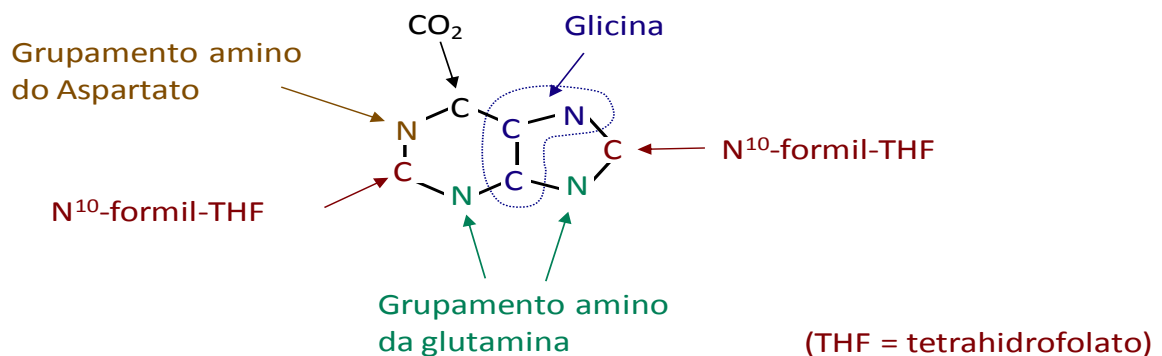


Figura 1. Elementos doadores dos átomos que formam os anéis da purina. Fonte: Modificado de ZHANG et al., 2008.

Um dos cientistas que mais contribuíram para a elucidação da via biossintética de purinas foi John Buchanan (1917 - 2007). Buchanan e seus colegas começaram a estudar a biossintese de purinas utilizando pombos, que excretam grandes quantidades de purinas através da urina. Empregando métodos isotópicos de marcação *in vivo*, eles determinaram quais precursores deram origem aos átomos de carbono e nitrogênio do anel purínico. O professor Buchanan fazia parte do Massachusetts Institute of Technology, em Cambridge, Massachusetts. Ele faleceu em 2007 aos 89 anos. Seu estudo detalhado da biossintese de purinas possibilitou o desenvolvimento de um grande número de quimioterápicos antineoplásicos que têm sido amplamente utilizados nas últimas décadas (CHEN, 2007; e ZHANG et al., 2008).

A biossintese “de novo” de purinas ocorre através de 10-11 reações enzimáticas (Figura 2). Das enzimas envolvidas na sequência de reações, seis (PurF, PurD, PurL, PurM, PurC e PurB) têm presença obrigatória na rota de diferentes organismos, como aves, mamíferos, leveduras e microrganismos (MOAT et al., 1960; ZHANG et al., 2008).

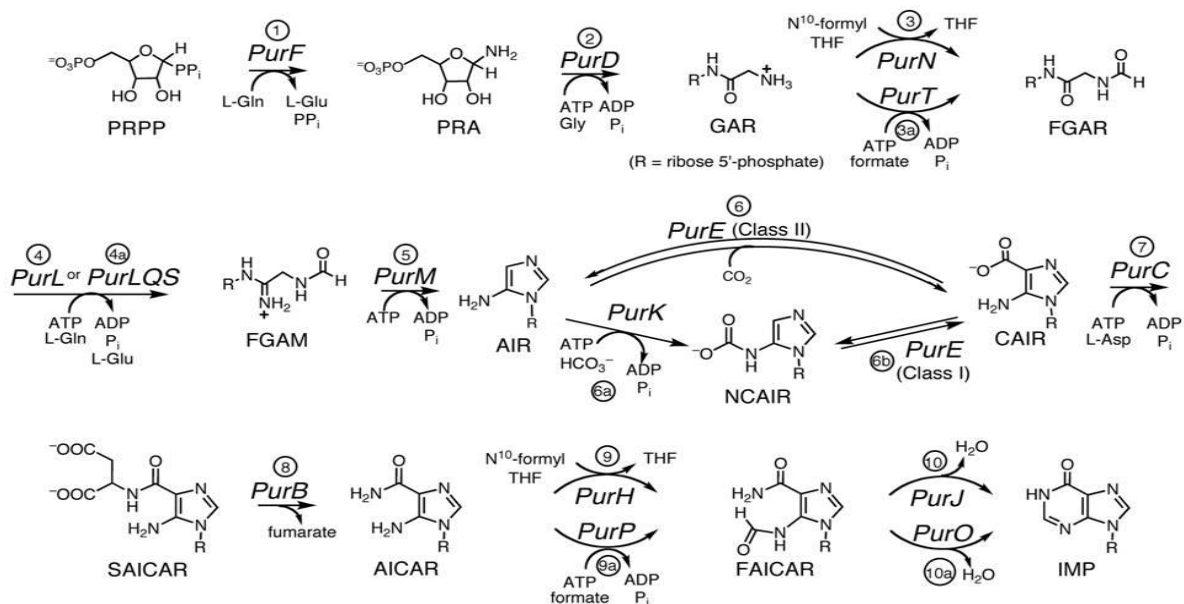


Figura 2. Via biossintética de purinas. Fonte: ZHANG et al., 2008.

No primeiro passo da via, a enzima amidofosforibosiltransferase (PurF) adiciona um grupo amino, doado pela glutamina, ao C1 do 5-fosforibosilpirofosfato (PRPP), convertendo-o em 5-fosforibosilamina (PRA) (SHIMAOKA et al., 2007). PurF possui dois domínios estruturais, um domínio N-terminal que hidrolisa a

glutamina e um domínio C-terminal que utiliza nitrogênio liberado da glutamina ou da NH_3 em solução para a síntese de PRA (MUCHMORE et al., 1998). A estrutura cristalográfica de PurF mostra que ela forma um tetrâmero, mas em solução ela também pode ser observada como um dímero (MUCHMORE et al., 1998; ZHANG et al., 2008). Além de sua função catalítica, PurF também está envolvida na regulação da biossíntese de purinas através da inibição alostérica dos nucleotídeos guanina e adenina (ZHANG et al., 2008).

O intermediário PRA resultante da primeira reação é altamente instável, com meia-vida de aproximadamente 30 segundos em pH 7,5 (ZHANG et al., 2008). No segundo passo da via, a enzima Fosforibosilglicinamida sintetase (PurD) adiciona três átomos da glicina ao grupamento amino do intermediário PRA, gerando Glicinamida ribonucleotídeo (GAR). Esta reação ocorre com o consumo de ATP. PurD é uma enzima monomérica, com três domínios (A, B e C) característico da superfamília de proteínas ATP-grasp. Ela é a única enzima citada na literatura que catalisa este passo da via, sendo, portanto encontrada em todos os organismos (ZHANG et al., 2008).

Na terceira etapa da via biossintética de purinas, a reação pode ser catalisada por duas enzimas distintas que realizam a formilação de GAR em FGAR, porém elas o fazem por meio de diferentes mecanismos e cofatores (WARREN et al., 1996). Uma das enzimas, a Fosforibosilglicinamida formiltransferase (PurN), adiciona um grupo formil ao grupamento amino do GAR, utilizando como cofator o N^{10} -formiltetrahidrofolato (N^{10} -formilTHF) (NAGY et al., 1993), o resultado dessa reação é a formação do N-Formilglicinamida ribonucleotídeo (FGAR). A estrutura da PurN pode ser dividida em dois domínios, um N-terminal e outro C-terminal, e no espaço formado entre eles localiza-se o sítio de ligação do N^{10} -formilTHF. O sítio de ligação do GAR é mais exposto ao solvente e é posicionado favorecendo o recebimento do grupo formil (ZHANG et al., 2008; WARREN et al., 1996). Em organismos superiores PurN compõe o domínio C-terminal de uma enzima tri-funcional que também possui as atividades de PurD e PurM (ZHANG et al., 2008).

A outra enzima que catalisa esta reação é a Fosforibosilglicinamida transformilase (PurT). Esta enzima pode ser encontrada juntamente com PurN em *Escherichia coli* e outros organismos relacionados. PurT catalisa a adição de um grupo formil ao GAR por um mecanismo que requer Mg^{2+} , ATP e formato (LI et al., 2009; ZHANG et al., 2008; THODEN et al., 2002). A PurT pertence a superfamília de

proteínas conhecida como ATP-grasp. A estrutura de PurT é formada por um dímero, em que cada monômero possui três domínios, designados como A, B, e C (THODEN et al., 2002). O sítio ativo de PurT contém subsítios para o ATP, o GAR e o formato. O sítio de ligação do ATP é formado principalmente pelos domínios A e B, enquanto os domínios de ligação do GAR são formados principalmente pelos domínios A e C (ZHANG et al., 2008). PurT tem quase o dobro do tamanho da PurN, no entanto, as duas enzimas não são homólogas (WARREN et al., 1996), por outro lado, PurT é homóloga a outras enzimas da via biossintética das purinas como a N5-Carboxiaminoimidazol ribonucleotídeo sintetase (PurK) e PurD (WARREN et al., 1996; KAPPOCK et al., 2000).

Apesar do baixo valor de identidade entre as sequências de aminoácidos, PurT e PurK compartilham uma grande similaridade estrutural, ambas pertencem à superfamília de proteínas ATP-grasp. Além disso, os mecanismos de ação empregados por elas são muito semelhantes, constituindo de clivagem de ATP, formação de um acil-fosfato reativo, e ataque nucleofílico no grupo amino dos intermediários (LI et al., 2009).

A enzima Fosforibosilformilglicinamida sintase (PurL) realiza uma reação ATP dependente na quarta etapa da via de biossíntese de purinas. PurL utiliza amônia derivada da glutamina como fonte de nitrogênio, adicionando um grupo amina ao FGAR, produzindo então N-Formilglicinamida ribonucleotídeo (FGAM) (ZHANG et al., 2008; MORAR et al., 2004).

São conhecidas duas isoformas de PurL, codificadas por dois genes descritos como *purL small* e *purL large* (MORAR et al., 2004). A PurL *large* possui três domínios em uma única sequência peptídica, um domínio glutaminase, um FGAM sintase, e um domínio N-terminal importante para formação do canal de amônia e acoplamento entre os outros dois domínios catalíticos (ZHANG et al., 2008). A PurL *large* é encontrada em eucariotos e bactérias Gram-negativas (MORAR et al., 2004; ZHANG et al., 2008). PurL *small* é encontrada em bactérias Gram-positivas e Archaeas e possuem a mesma atividade catalítica e homologia estrutural do domínio FGAM sintase da PurL *large*, porém nesses organismos outras duas enzimas são necessárias para produzir FGAM, estas enzimas são PurQ e PurS (ZHANG et al., 2008). PurQ é uma glutaminase responsável pela geração da amônia a partir da glutamina. PurS é homóloga ao domínio N-terminal da PurL *large* (MORAR et al., 2004).

O quinto passo da biossíntese de purinas é a conversão ATP dependente de FGAM em aminoimidazol ribonucleotídeo (AIR), catalisado pela enzima Fosforibosilformilglicinamidina ciclo-ligase (PurM). PurM catalisa uma reação quimicamente muito similar a da PurL, utiliza ATP para fosforilar um oxigênio, que então torna-se ativo para ataque nucleofílico do N1 do FGAM. Nessa reação ocorrem a desidratação e fechamento do anel imidazol (LI et al., 1999). A estrutura de PurM forma um dímero, cada monômero possui dois domínios N-terminal e C-terminal, chamados de domínios A e B. O sítio ativo localiza-se no espaço formado entre os domínios de cada monômero (LI et al., 1999; ZHANG et al., 2008).

O sexto passo da via, que resulta na formação de Carboxiaminoimidazol ribonucleotídeo (CAIR), essa reação pode ocorrer por diferentes mecanismos. Em bactérias duas enzimas são necessárias para realizar esta reação. O intermediário AIR é convertido a N5-Carboxiaminoimidazol ribonucleotídeo (NCAIR) pela enzima N5-Carboxiaminoimidazol ribonucleotídeo sintetase (PurK). PurK faz a ligação do bicarbonato ao grupo amino do AIR, numa reação dependente de ATP, produzindo NCAIR. Por ser um composto muito instável o NCAIR é rapidamente convertido em Carboxiaminoimidazol ribonucleotídeo (CAIR) pela enzima N5-Carboxiaminoimidazol ribonucleotídeo mutase (PurE I) (ZHANG et al., 2008).

PurK é um dímero e como outros membros da superfamília ATP-grasp o monômero pode ser dividido em três domínios A, B e C. Os domínios A e C formam um centro compacto e o domínio B se estende além do centro formando a alça flexível do sítio ativo. No sítio ativo encontram-se os subsítios de ligação do ATP, AIR e do bicarbonato (THODEN et al., 2008).

A reação catalisada pela PurE, pode ocorrer de duas maneiras. Em procariotos e fungos, após a adição do grupo carboxilato ao grupo amino exocíclico pela PurK, ele é transferido para a posição C5 do anel do imidazol por uma PurE classe I (HOSKINS et al., 2007). Em eucariotos superiores a conversão de AIR para CAIR acontece diretamente. Nesses organismos a enzima Carboxiaminoimidazol ribonucleotídeo carboxilase (PurE classe II), utilizando CO_2 e AIR como substrato, faz a adição direta do grupo carboxilato na posição C5 do anel imidazol, sem a participação da PurK (ZHANG et al., 2008; THODEN et al., 2008). A PurE pertence a superfamília ATP-grasp de proteínas. Essa enzima é formada por oito monômeros com os três domínios A, B e C, compondo, portanto um octamero. Cada um dos oito sítios ativos está localizado em uma fenda formada por três monômeros (THODEN

et al., 2008). Em vertebrados PurE encontra-se fundida a PurC, formando uma enzima bifuncional fosforibosilaminoimidazol carboxilase/fosforibosilaminoimidazol succinocarboxamida sintetase (PAICS) (LI et al., 2007).

A fosforibosilaminoimidazol succinocarboxamida sintase (PurC) catalisa o sétimo passo da via. Empregando ATP, essa enzima liga o grupo amino do L-aspartato ao grupo carboxilato da CAIR, dando origem ao intermediário N-Succinocarboxamida-aminoimidazol ribonucleotídeo (SAICAR). A estrutura do monômero de PurC pertence a superfamília de proteínas ATP-grasp (ZHANG et al., 2008).

A oitava reação da biossíntese de purinas é catalisada pela Adenilosuccinate liase (PurB). Nessa reação o SAICAR formado na reação anterior, é convertido a Aminoimidazole-4-carboxamida ribonucleotídeo (AICAR). PurB é a única enzima da via que tem dupla especificidade. A PurB é responsável por duas reações não consecutivas catalisadas no mesmo sítio ativo. Primeiro converte SAICAR em AICAR e posteriormente participa da conversão do IMP em AMP, nas duas reações ocorre a remoção do grupo fumarato do substrato em questão (ZHANG et al., 2006). PurB quebra a SAICAR em AICAR e fumarato, além disso desempenha um papel fundamental na regulação do metabolismo celular, pelo controle da quantidade de AMP livre (TSAI et al., 2007).

O gene *purH* codifica uma enzima bifuncional, Fosforibosilaminoimidazolcarboxamida formiltransferase (PurH) e IMP ciclohidrolase (IMPCH), presente em bactérias e eucariotos. A enzima PurH é responsável por transferir um grupo formil, utilizando como cofator N¹⁰-formilTHF, para AICAR gerando 5-Formamido-4-imidazolcarboxamida ribonucleotídeo (FAICAR), em seguida a IMP ciclohidrolase (PurJ), que fecha o segundo anel purínico gerando IMP, molécula base das purinas. Embora formado por uma cadeia polipeptídica única, PurH é formada por um monômero com dois domínios funcionalmente independentes. Cada domínio pode ser individual, expressa em um fragmento enzimaticamente ativo. Outra enzima a PurP, análoga a PurH, também pode realizar a síntese de FAICAR, transferindo o grupo formil para o AICAR utilizando o formato como cofator e com gasto de ATP (ZHANG et al., 2008).

Em Archaeas a IMP ciclohidrolase II (PurO), análoga ao domínio IMPCH da PurH, por um mecanismo distinto de ciclização converte FAICAR em IMP eliminando H₂O sem consumo de ATP (OWNBY et al., 2005).

Com exceção do ATP, a quantidade total de nucleotídeos nas células é muito pequena, cerca de 1% da quantidade necessária para sintetizar o DNA celular. Por isso, a síntese de nucleotídeos precisa ser constante durante a síntese de ácidos nucléicos, podendo até limitar a taxa de replicação e transcrição do DNA. Por ser de grande importância nos processos de multiplicação celular, os agentes que inibem a síntese de nucleotídeos são de particular interesse para a medicina. Nesse contexto, compreender os aspectos bioquímicos e evolutivos das enzimas da via biossintética de purinas pode fornecer importantes informações para o desenvolvimento de drogas anticâncer, antiviral e antimicrobianos (ZHANG et al., 2008).

3.2. Origem das vias biossintéticas

O metabolismo é um complexo conjunto de processos físico-químicos catalisados por enzimas, que permite a sobrevivência e a reprodução das células (ALVES et al., 2002). Atualmente há um grande interesse nos processos que governam a origem das vias biossintéticas, pois, como qualquer outro fenômeno biológico, as vias metabólicas resultam de processos evolutivos, logo, a compreensão da sua história evolutiva pode revelar novas pistas sobre a evolução e diversificação da vida.

A teoria da origem heterotrófica da vida sugere que os compostos presentes na sopa primitiva eram utilizados como nutrientes para manutenção das primeiras formas de vida (OPARIN, 1938). Com base neste cenário, em 1945, Horowitz visando explicar como as primeiras rotas metabólicas surgiram, postulou a hipótese da “retro-evolução” (HOROWITZ, 1945). Segundo esta hipótese, as células primitivas usavam determinados compostos presentes no ambiente, assim, com o intenso consumo destes compostos, eles se tornaram escassos. Células capazes de sintetizar estes compostos a partir de um precursor disponível no ambiente teriam vantagens seletivas em detrimento de outras que não fossem capazes de fazê-lo.

Assim de acordo com a teoria da retro-evolução, cada etapa de uma via biossintética foi selecionada de acordo com o esgotamento sucessivo de seu produto (Figura 3A). Portanto, as primeiras enzimas a surgirem foram aquelas responsáveis pelas últimas reações, e aquelas que evoluíram mais tardiamente foram as responsáveis pelas etapas iniciais da via (HOROWITZ, 1945).

Uma hipótese distinta foi proposta por Cordón (1990) para a evolução das vias catabólicas, a evolução direta (Figura 3B). De acordo com essa hipótese, as

primeiras formas de vida produziam energia pela degradação de compostos disponíveis no ambiente. Os organismos que desenvolveram uma reação complementar nessa degradação conseguiram obter mais energia a partir do composto original, dessa forma, esses organismos tiveram uma vantagem seletiva. Assim, as primeiras enzimas a surgirem nas vias de degradação são as mais precoces, enquanto as que surgirem mais tardiamente, são as últimas a realizarem a reação.

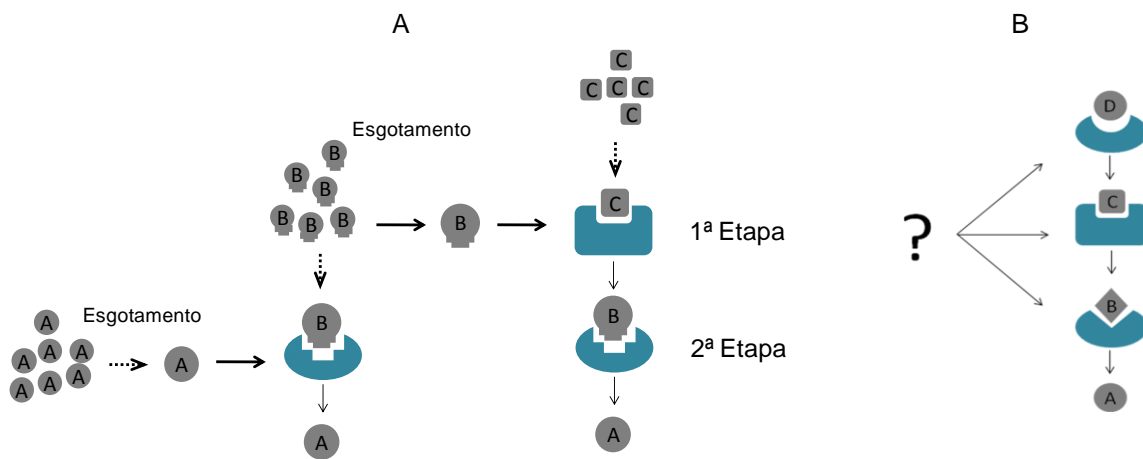


Figura 3. Modelos de evolução de vias metabólicas. (A) Retro-evolução; (B) Evolução direta. Fonte: Modificado de SCHMIDT et al., 2003.

A teoria da retro-evolução proposta por Horowitz pode ser aplicada em um segundo cenário onde os compostos disponíveis no ambiente sofreriam degradação não enzimática e os subprodutos se acumulariam no meio. Os organismos que desenvolveram enzimas capazes de reverter o processo de degradação, produzindo o composto original a partir dos subprodutos, teriam uma vantagem seletiva. Assim, surgiram primeiramente reações não enzimáticas seguidas pela aquisição da enzima (LAZCANO et al., 1999; CAETANO-ANOLLES et al., 2008).

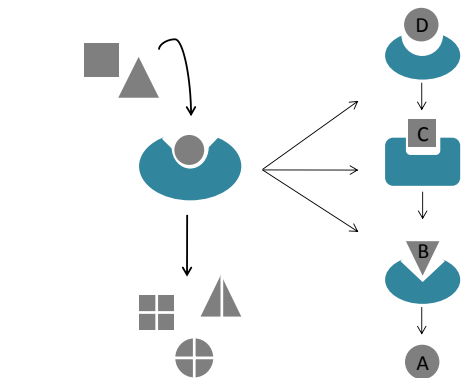
O modelo da retro-evolução estabelece uma conexão entre a evolução química pré-biótica e o desenvolvimento de vias metabólicas e pode ser aplicada para explicar algumas rotas metabólicas (LAZCANO et al., 1999). Porém, a origem de outras vias biossintéticas não pode ser entendida em termos de desenvolvimento retrogrado visto que produzem intermediários instáveis e a síntese e acumulação destes compostos no ambiente prebiótico seria improvável (CUNCHILLOS et al., 2002 e LAZCANO et al., 1999).

Outra hipótese para explicar a origem das vias biossintéticas sugere que a evolução dessas vias ocorreu sem adoção de enzimas existentes (GRANICK, 1957).

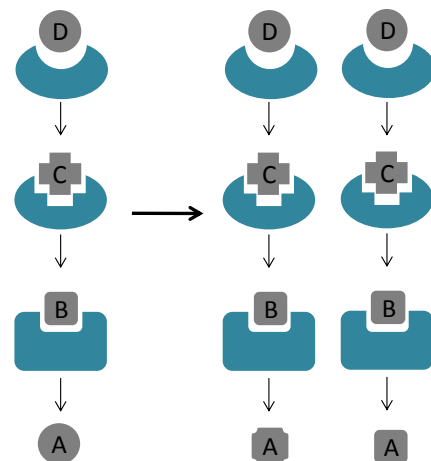
Neste cenário os compostos prebióticos não desempenhavam um papel essencial para evolução da rota. Desta forma as vias teriam surgido na mesma ordem em que a reação ocorre. Para isso, cada intermediário produzido seria acumulado, e o metabolismo se completaria com o desenvolvimento de novas enzimas. Essa teoria é útil para explicar a evolução de rotas como a síntese de clorofila e de isopreno (OURISSON et al., 1994), porém não pode ser utilizada para explicar a evolução de vias em que os intermediários não têm qualquer utilidade para a célula, como é o caso da biossíntese de purinas e aminoácidos (LAZCANO et al., 1999).

Jensen (1976) sugeriu que as rotas metabólicas evoluíram como resultado do recrutamento de uma série de enzimas relativamente ineficientes dotadas de ampla especificidade catalítica que poderiam reagir com uma grande variedade de substratos quimicamente relacionados (Figura 4A). A partir dessa enzima multifuncional a via poderia ter evoluído seguida pela duplicação e diversificação desta enzima precursora em enzimas mais específicas e eficientes que catalisam apenas um passo de cada vez.

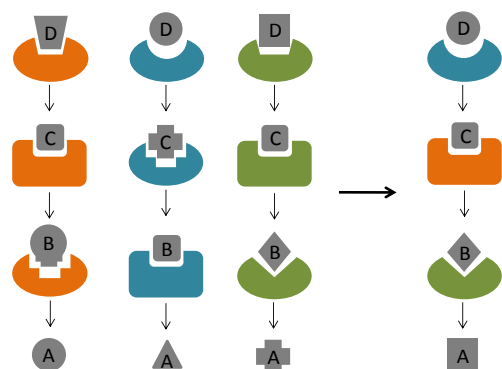
A partir do surgimento das primeiras vias, outras rotas poderiam surgir pela



A. Especialização de enzima multifuncional



B. Duplicação de via completa



C. Recrutamento enzimático

Figura 4. Modelos de evolução de vias metabólicas. Fonte: Modificado de SCHMIDT, 2003.

duplicação de um conjunto enzimático completo (Figura 4B), ou pelo recrutamento de enzimas de diferentes vias metabólicas (Figura 4C) (SCHMIDT et al., 2003). No entanto, a combinação de enzimas poderia funcionar somente após o surgimento da biossíntese de proteínas e o desenvolvimento de enzimas, ou seja, após o surgimento do DNA e das proteínas. Segundo LAZCANO et al. (1999), esse mecanismo não pode ser aplicado para as primeiras rotas, pois havia poucas enzimas disponíveis.

Determinar qual o mecanismo envolvido na evolução de vias biossintéticas é uma tarefa que exige cautela. Talvez, a melhor maneira de entender este processo é por meio da combinação das diferentes hipóteses. Assim, algumas das primeiras vias podem ter surgido a partir do esquema de Horowitz, ou de Granick, e então, após o surgimento das primeiras enzimas, o recrutamento enzimático de Jensen pode ter proporcionado a diversificação de atividades enzimáticas e surgimento de novas rotas metabólicas (LAZCANO et al., 1999; CAETANO-ANOLLES et al., 2008). Além disso, a dispersão e diversificação das enzimas podem ter ocorrido por eventuais duplicações e mutações das enzimas disponíveis (ALVES et al., 2002).

Caetano-Anollés et al. (2007) fizeram a reconstrução filogenética de estruturas protéicas, para descobrir qual o padrão evolutivo dessas estruturas. Essa análise utilizou domínios estruturais comuns a 185 genomas completamente sequenciados, representando os três domínios da vida. As doze formas mais basais encontradas nesta filogenia, representam estruturas importantes que podem ser rastreadas até o ancestral comum dos três domínios da vida. Entre as estruturas mais basais, duas delas são responsáveis pelo metabolismo de nucleotídeos: purinas e pirimidinas, e as outras estão relacionadas com o metabolismo de porfirina e clorofila, dos aminoácidos serina, glicina e treonina, glicérolídeos, ácido selenoamino, enzimas envolvidas na fosforilação oxidativa, degradação do benzoato, metabolismo do amido e sacarose, biossíntese de N-glicano, policetídeos, e pelo metabolismo de estibeno (Figura 5).

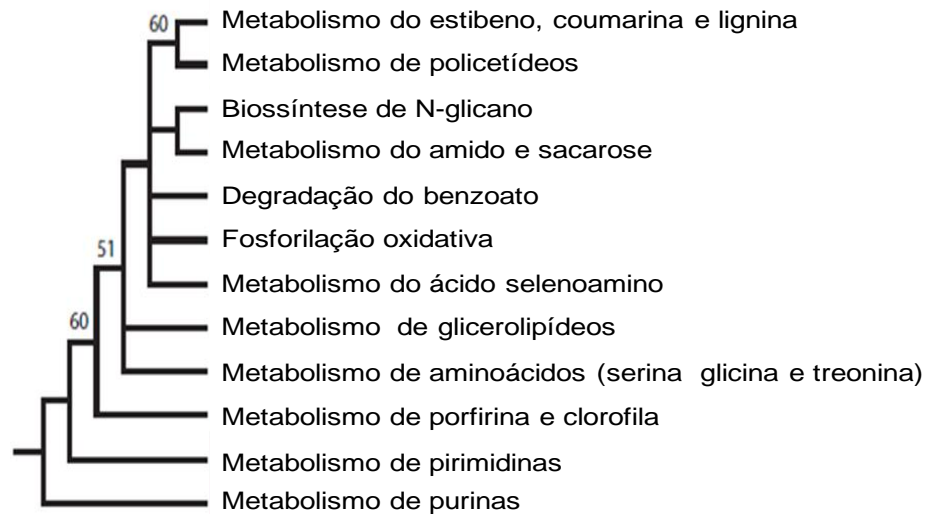


Figura 5. Filogenia das vias metabólicas evidenciando que as vias biossintéticas de nucleotídeos foram, possivelmente, as primeiras rotas a surgirem. Fonte: CAETANO-ANOLLES et al., 2007.

Estes dados sugerem que as vias metabólicas atuais se originaram a partir do metabolismo de nucleotídeos, provavelmente da via biossintética de purinas. Esse resultado é congruente com as condições primitivas nas quais surgiram as primeiras formas de vida, visto que os nucleotídeos têm relação direta com a produção de RNA, sendo, portanto, de fundamental importância no mundo de RNA, já que este teria sido o primeiro material informacional e catalítico no mundo primitivo, até o surgimento do DNA (CAETANO-ANOLLES et al., 2007).

4. OBJETIVO GERAL

- Compreender a distribuição e evolução dos genes *purE*, *purK*, *purN* e *purT* em procariotos.

CAPÍTULO I

DIVERSIDADE ESTRUTURAL E EVOLUÇÃO DO TERCEIRO PASSO DA BIOSSÍNTESE DE PURINAS EM LINHAGENS PROCARIÓTICAS

5. INTRODUÇÃO

A evolução de vias biossintéticas é uma questão que esta diretamente relacionada com a origem e evolução das linhagens procarióticas e eucarióticas. Os avanços na área de biotecnologia, em especial no desenvolvimento de técnicas de sequenciamento de DNA, tem permitido o sequenciamento em larga escala de genomas de organismos procarióticos e eucarióticos. Muitos aspectos sobre a evolução das vias biossintéticas estão sendo elucidados usando os dados obtidos a partir de genomas totalmente sequenciados (RODIONOV, 2002; MORETT, 2003).

Na terceira reação da via biossintética de purinas ocorre a formilação do grupamento amino do intermediário GAR (Figura 6). Em procariotos já foram descritas duas enzimas capazes de catalizar essa reação, a PurT e a PurN. No entanto, essas enzimas empregam diferentes mecanismos e utilizam distintos cofatores para realizar essa reação (WARREN et al., 1996). PurN faz a formilação direta do GAR utilizando como cofator o N¹⁰-formiltetrahydrofolato (N¹⁰-formilTHF) (NAGY et al., 1993). A mesma reação quando catalisada por PurT, tem como cofatores Mg²⁺, ATP e formate (LI et al., 2009). A maior parte do formate utilizado por PurT provem da quebra do N¹⁰-formilTHF em tetrahydrofolato e formate, realizada pela enzima PurU (NAGY et al., 1993).

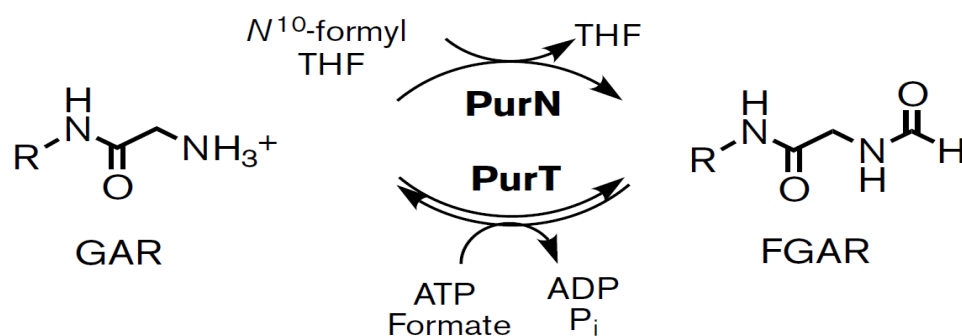


Figura 6. Terceira reação da via biossintética de purinas. Fonte: KAPPOCK et al., 2000.

De acordo com a literatura *purN* possui distribuição maior do que *purT*, contudo, esta afirmação não é baseada em análises sistemáticas, tendo essa questão como objeto de estudo (ZHANG et al., 2008). Ela deriva do fato deste gene estar presente no genoma da maioria dos organismos usados como modelo de

estudo da via biossintética de purinas. Logo, a distribuição de *purN* e *purT* nas linhagens procarióticas é uma icógnita.

A existência de pelo menos duas enzimas em linhagens procarióticas capazes de converter GAR em FGAR remete a questões sobre a evolução da terceira etapa da via biossintética de purinas como (i) qual a distribuição destes genes nas linhagens procarióticas?; (ii) PurT ou PurN estão igualmente distribuídas ou são preferencialmente utilizadas nas linhagens procarióticas? (iii) Como duas enzimas foram selecionadas para desempenhar a mesma função durante a evolução dos procariotos e não apenas uma?

A relevância de tais perguntas vai além do interesse sobre questões evolutivas, que por si só justifica tentar respondê-las. Na espécie humana a reação de formilação do GAR, é catalisada por uma enzima tri-funcional que possui os domínios de PurD, PurM e PurN (WELIN et al., 2010), PurT não está presente no genoma plantas, fungos e animais. Portanto, conhecer a distribuição de *purT* e *purN* nos genomas de procariotos é uma informação importante e pode ser aplicada no desenvolvimento de fármacos antibacterianos, visto que muitos microrganismos patogênicos para a espécie humana podem possuir o gene *purT*.

6. OBJETIVOS ESPECÍFICOS

- Fazer a genômica comparativa e contexto genômico dos genes *purN* e *purT* em procariotos;
- Fazer a análise de anticorrelação dos genes *purN* e *purT* em procariotos;
- Fazer a reconstrução filogenética da enzima PurT.

7. METODOLOGIA

7.1. Busca pelos genes *purT* e *purN* da via biossintética de purinas

Foram feitas buscas pelos genes *purN* e *purT* em 815 genomas completamente sequenciados e anotados disponíveis na base de dados do NCBI até 16 de fevereiro de 2009 e representando 535 espécies dos Domínios Archaea e Bacteria. Durante a busca foram usados os programas BLASTP e TBLASTN

(ALTSCHUL et al., 1997). O nível e extensão de similaridade, assim como a presença de domínios conservados típicos de cada produto gênico, foram utilizados como critério de homologia para a busca das sequências dos genes em questão. As sequências de *Escherichia coli* foram utilizadas como *query* para a busca dos genes. Esse organismo foi escolhido por possuir todos os genes de interesse nesse estudo caracterizados funcionalmente.

7.2. Análise da diversidade estrutural da terceira e etapa da via biossintética de purinas

Durante a busca pelos genes *purN* e *purT* foram anotados o padrão de presença/ausência e a quantidade de cópias dos genes canônicos da via biossintética de purinas. No entanto, os estudos de genômica comparativa e contexto genômico foram conduzidos apenas com os organismos que possuem o conjunto mínimo de genes conhecidos necessário para a síntese do IMP, exceto para o Filo Crenarchaeota. Os Filos e Classes incluídos na análise, assim como o número de gêneros e espécies amostradas em cada Filo/Classe são apresentados na Tabela 1.

Tabela 1. Filos e classes dos Domínios Archaea e Bacteria amostradas no estudo. Entre parênteses esta destacado o número total de Filos e Classes (EUZÉBY, 1997).

Domínio	Filo (30)	Classe (62)	Siglas	Amostragem (Gênero/Espécie)	
Archaea	Crenarchaeota	Thermoprotei	CR	11/16	
	Euriarchaeota	Archaeoglobi	EU	23/32	
		Halobacteria			
		Methanobacteria			
		Methanococci			
		Methanomicrobia			
		Methanopyri			
		Thermococci			
Thermoplasmata					
Bactéria	Gram positiva	Firmicutes	Bacilli	FB	15/51
			Clostridia	FC	16/28
		Tenericutes	Mollicutes	FT	6/20
		Actinobacteria	Actinobacteria	AB	21/43
	Gram negativa	Cyanobacteria	Gloeobacteria	CB	11/13
		Aquificae	Aquificae	AQ	3/3
		Chloroflexi	Chloroflexi	CF	4/7
			Dehalococcoidetes		
		Deinococcus-Thermus	Deinococci	DTH	2/3
		Acidobacteria	Solibacteres	AC	2/2
		Fusobacteria	Fusobacteria	FU	1/1
		Planctomycetes	Planctomycetacia	PLC	1/1
		Thermotogae	Thermotogae	THR	4/8
		Bacteroidetes	Bacteroidia	BC	10/13
			Cytophagia		
			Flavobacteria		
			Sphingobacteria		
		Chlorobi	Chlorobia	CL	5/10
		Verrucomicrobia	Verrucomicrobiae	VM	3/3
			Opitutae		
		Dictyoglomi	Dictyoglomia	DC	1/2
		Elusimicrobia	Elusimicrobia	EB	2/2
		Nitrospirae	Nitrospira	NT	1/1
		Chlamydiae	Chlamydiae	CH	3/7
	Spirochaetes	Spirochaetes	SP	3/12	
	Proteobacteria	Alphaproteobacteria	AF	46/78	
		Betaproteobacteria	BT	23/44	
Deltaproteobacteria		DT	13/20		
Epsilonproteobacteria		EP	7/13		
Gammaproteobacteria		GM	57/102		

7.3. Análise de anticorrelação

A análise de anticorrelação dos genes *purT* e *purN* foi realizada de acordo com a metodologia usada por Morett et al. (2003). Para cada genoma presente na base de dados foram avaliadas a presença ou ausência dos genes, e a pontuação seguiu o seguinte critério: dos organismos que possuem a via completa, aqueles que (i) possuem apenas um dos genes receberam pontuação 1; (ii) os que possuem ambos os genes foram penalizados com -1; e (iii) aqueles que não possuem nenhum dos genes não foram penalizados nem pontuados. No caso dos organismos que não tem a via (i) os organismos que não possuem a via e que não apresentam nenhum dos genes receberam a pontuação 1; e (ii) aqueles que apresentam um dos genes ou ambos, foram penalizados com -1.

O score de anticorrelação foi obtido a partir da soma do total de pontos, dividido pelo número de genomas avaliados. Os valores foram calculados para ao total de organismos da base de dados e separadamente entre os grupos de Gram positivas, Gram negativas e Archaea. Os valores acima de 0,7 foram considerados significativos, ou seja, indica que os genes *purN* e *purT* anticorrelacionam.

7.4. Contexto genômico

Durante a busca pelas sequências dos genes *purT* e *purN*, o contexto genômico dos organismos pesquisados foi anotado, permitindo acessar informações sobre a organização genômica dos genes estudados. Para cada gene pesquisado, foram registrados os genes que os flanqueiam, gerando uma figura onde foram representados os genes da via biossintética de purinas, e os genes que os flanqueiam.

7.5. Alinhamento múltiplo e filogenia do gene *purT*

As sequências dos genes recuperadas nas buscas realizadas nas bases de dados do NCBI foram alinhadas códon a códon no programa MEGA 4 (TAMURA et al., 2007). As sequências de nucleotídeos foram traduzidas para aminoácidos e os alinhamentos múltiplos foram editados com base no alinhamento estrutural das estruturas terciária das proteínas disponíveis na base de dados do PDB (<http://www.rcsb.org/pdb/home/home.do>). As sequências redundantes e incompletas foram excluídas. A visualização e sobreposição das estruturas terciárias das

proteínas foram feitas nos programas TOPMACH (SIPPL e WIEDERSTEIN, 2008) e PYMOL (DeLano et al., 2002).

As análises filogenéticas foram realizadas utilizando as sequências de aminoácidos da enzima PurT. Nessas análises não foram considerados os sítios com GAPs. A reconstrução filogenética foi realizada com base no método de *Neighbor-joining* (SAITOU e NEI, 1987), utilizando o programa ClustalX.

8. RESULTADOS E DISCUSSÃO

Aproximadamente 73% dos Filos e 63% das Classes de procariotos reconhecidas (EUZÉBY, 1997) estão representados neste estudo, contudo, não estão igualmente amostrados. Portanto, os resultados a seguir foram analisados em nível gênero e de espécie para minimizar o viés amostral dos genomas completamente sequenciados.

8.1. Distribuição de *purN* e *purT* nas linhagens procarióticas

Dentre os 271/473 gêneros/espécies que possuem o conjunto mínimo de genes da biossíntese de purinas, 96%/94% dos gêneros/espécies possuem *purN* e 35%/38% possuem *purT* (Figura 8). Dentre estes gêneros/espécies 29%/33% possuem ambos os genes enquanto que 68%/61% possuem apenas *purN* e 6%/5% apenas *purT*.

Os genes *purN* e *purT* não foram encontrados em 5/5 gêneros/espécies (Figura 7). Em vários genomas da base de dados do NCBI foi realizada apenas a anotação automática que ainda não foi checada manualmente. Nestes casos a busca com o programa BLASTP pode dar resultado negativo por que o gene em questão não foi anotado ou foi anotado incorretamente como pseudogene, devido a problemas na montagem dos *contigs* ou a erros de sequenciamento em regiões homopoliméricas.

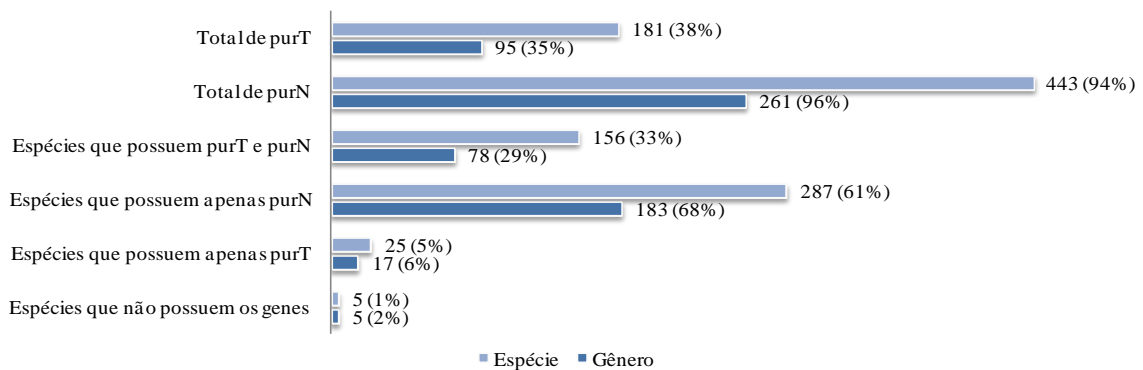


Figura 7. Distribuição de *purN* e *purT* nas 271/473 gêneros/espécies que possuem a via biossintética de purinas.

Para eliminar os problemas de erros de anotação todos os resultados de ausência do gene nas buscas realizadas com o BLASTP foram checados com uma nova análise usando o programa TBLASTN. Esta análise confirmou a ausência do *purN* e *purT* no genoma das 5/5 gêneros/espécies acima descritos. Estes gêneros/espécies pertencem ao Filo Euryarchaeota, são de vida livre e possuem os demais genes necessários para a síntese de purinas, portanto a ausência de *purN* e *purT* nestas espécies sugere a existência de um análogo destes genes ainda não descrito nestas espécies.

O relato de novos genes envolvidos em processos celulares amplamente conhecidos tem sido recorrente na literatura nos últimos anos (MORETT et al., 2003). Um fator que certamente contribui para isso é o aumento do número de genomas sequenciados disponíveis em bases de dados públicas, permitindo que extensas análises genômicas sejam desenvolvidas.

8.2. Genômica comparativa de *purN* e *purT* nas linhagens procarióticas

A genômica comparativa mostrou que os genes *purN* e *purT* não estão igualmente distribuídos nos 27 Filos/Classes procarióticas incluídos nessa análise. Os seguintes padrões estruturais foram encontrados: (I) o organismo possui ambos os genes presentes no genoma; (II) o organismo possui apenas o gene *purN* no genoma; (III) o organismo possui apenas o gene *purT* no genoma; (IV) o organismo não possui nenhum dos genes em seu genoma (Tabela 2).

A

Padrão estrutural	Genes pur		Filos/Classes																											
	N	T	CR	EU	FB	FC	FT	AB	CB	AQ	CF	DTH	AC	FU	PLC	THR	BC	CL	VM	DC	EB	NT	CH	SP	AF	BT	GM	DT	EP	
I				1/1	1/2			3/15	6/8*	2/2*							3/6	2/5						1/3	5/5	14/30*	31/65*	4/8	5/5*	
II			2/5	12/16*	15/46*	16/28		18/25*	5/5	1/1	4/7	2/3	2/2	1/1	1/1	4/8	4/4	3/5	3/3	1/2	2/2	1/1		1/1	40/63*	9/12	22/28	11/11*	2/6	
III			5/7*	4/8				1/3																		2/2	5/5			
IV			4/4	7/7	1/3		6/20										2/2							3/7	1/7	2/10		3/3		1/2

B

	Filos/Classes																											
	CR	EU	FB	FC	FT	AB	CB	AQ	CF	DTH	AC	FU	PLC	THR	BC	CL	VM	DC	EB	NT	CH	SP	AF	BT	GM	DT	EP	
Gênero	1	0,69	0,86	1	1	0,71	-0,10	-0,30	1	1	1	1	1	1	0,40	0,20	1	1	1	1	1	1	0,30	0,78	-0,21	-0,01	0,07	-0,40
Espécie	1	0,75	0,92	1	1	0,30	-0,23	-0,33	1	1	1	1	1	1	0,07	0	1	1	1	1	1	1	0,50	0,87	-0,36	-0,03	0,10	0,23

Tabela 2. (A) Distribuição e padrão de ocorrência dos genes *purN* e *purT* em procariotos. As três primeiras colunas indicam os padrões estruturais encontrados. Nas outras colunas estão listados os filos/classes de procariotos, indicando o número de gêneros/espécies que cada filo possui, os asteriscos destacam os padrões mais representativos. Amarelo indica presença do gene e branco ausência. (B) Análise de anticorrelação. Os valores acima de 0,7 são considerados significativos. As pontuações negativas ou inferiores a 0,7 não são significativas, nesses casos ambos os genes estão presentes na maioria dos gêneros/espécies. A pontuação 1 é a pontuação máxima de anticorrelação, significa que todos os gêneros/espécies possuem apenas um dos genes. Nos filo/classe Firmicutes Tenericutes e Chlamydiae a pontuação 1 (vermelho), indica que nenhum dos gêneros/espécies possuem os genes.

O padrão estrutural I foi encontrado em 13 Filos/Classes, embora seja predominante apenas nas Classes Epsilon, Beta e Gamma proteobactérias. Neste caso, o problema do viés amostral é evidente nas Epsilon-proteobactérias. Quando o número de espécies é considerado o padrão II é o predominante, sendo encontrado em seis espécies. Entretanto, quando o número de gêneros é levado em consideração o padrão estrutural I é predominante, sendo observado em cinco gêneros contra apenas dois com o padrão II. Nos demais Filos/Classes em que foi observado o padrão estrutural I não foi predominante ou o Taxa foi pouco amostrado, como no caso de Aquificae e Spirochaete.

O padrão estrutural II ocorre em 25 dos 27 Filo/Classes amostradas, sendo predominante nos filos/classes Euryarchaeota, Bacilli, Clostridia, Actinobacteria, Alfa e Deltaproteobacteria. No filo Chlorobi o número de gêneros que apresentam o padrão II é maior do que aqueles com o padrão estrutural I. Porém, o número de gêneros/espécies amostrados não permitem outras conclusões além de que os padrões estruturais I e II ocorrem nos gêneros/espécies deste Filo.

O padrão estrutural III está presente em 5 filos/classes, porém, é predominante apenas no Filo Crenarchaeota. Em todos os filos/classes onde *purT* é

encontrado, *purN* também está presente, evidenciando a predominância desse gene em relação a *purT*.

Todas as espécies dos filos Tenericutes e Clamydiae analisadas nesse estudo não possuem os genes *purN* e *purT*. Os demais genes da Biossíntese de purinas também não estão presentes nos genomas destas espécies, exceto o gene *purB* que foi encontrado no genoma de algumas espécies de Tenericutes. Estes Filos são compostos por organismos parasitas intracelulares, logo, a ecologia das espécies destes Filos pode explicar a ausência dos genes da biossíntese de purinas.

Esse padrão estrutural também foi observado em espécies parasitas intracelulares ou simbiotes pertencentes aos filios Crenarchaeota, Euryarchaeota, Firmicutes Bacilli, Bacteroidetes, Spirochaetales, Alfa, Epsilon e Gamaproteobacterias. Embora este padrão não seja o predominante nestes Filos/Classes indica que o modo de vida dos organismos é um fator determinante na estrutura da via biossintética de purinas.

8.3. Contexto genômico de *purN* e *purT* nas linhagens procarióticas

Na maioria dos gêneros/espécies (69%/72%) *purN* esta em um contexto genômico, associado com outros genes da biossíntese de purinas, nos demais gêneros/espécies o contexto genômico de *purN* não possui genes relacionado a síntese de purinas. O contrário ocorre com o gene *purT*, que está em um contexto genômico com outros genes da via biossintética de purinas em apenas 7%/6% gêneros/espécies (Figura 8). As espécies que possuem *purT* em um contexto com outros genes da via pertencem aos Filos/Classes Firmicutes Bacilli, Deltaproteobactéria, Crenarchaeota e Euriachaeota (Tabela 2).

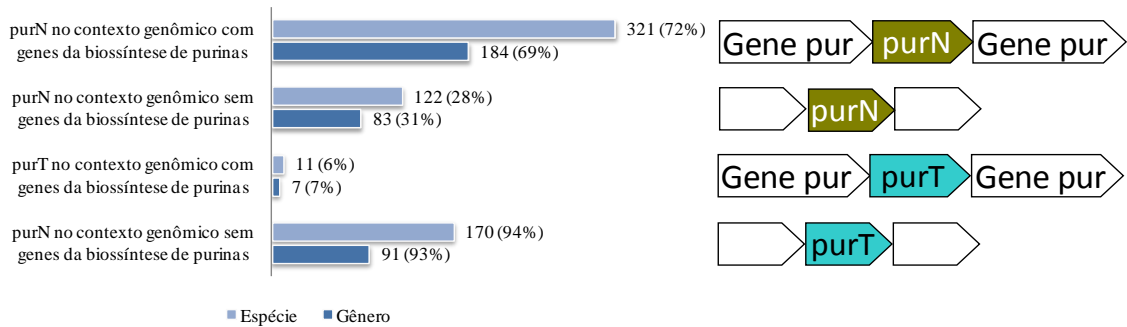


Figura 8. Contexto genômico de *purN* e *purT* dos 271/473 gêneros/espécies que possuem a via biossintética de purinas. Em 184/321 gêneros/espécies *purN* aparece em um contexto genômico com outros genes da via biossintética de purinas. O contrário ocorre com o gene *purT*, em 91/170 gêneros/espécies esse gene é encontrado em um contexto genômico sem associação com outros genes da biossíntese de purinas. Ao lado esta representado o contexto genômico de *purN* e *purT*. As caixas vazias representam genes não relacionados com a biossíntese de purinas.

Firmicutes Bacilli	<i>Bifidobacterium adolescentis</i> 15703	
	<i>Bifidobacterium animalis subsp lactis</i> AD011	◁ purC purT ▷
	<i>Bifidobacterium longum</i> DJO10A	
Deltaproteobactéria	<i>Mycobacterium avium</i> 104	◁ purC purT ▷
	<i>Pelobacter carbinolicus</i> DSM 2380	◁ purC ▷ ▷ purT ▷
Crenarchaeota	<i>Caldivirga maquilingensis</i> IC-167	
	<i>Desulfurococcus kamchatkensis</i> 1221n	◁ purT purC purC purC purC purC purC purC purC purC purC ▷
	<i>Pyrococcus abyssi</i> GE5	◁ purC purT ▷
Euriarchaeota	<i>Pyrococcus furiosus</i> DSM 3638	◁ purC purT ▷
	<i>Pyrococcus horikoshii</i> OT3	◁ purC purT ▷
	<i>Thermococcus kodakarensis</i> KOD1	◁ purT purT ▷

8.4. Análise de anticorrelação dos genes *purN* e *purT*

Valor significativo de anticorrelação significa que quando um gene esta presente o outro tende a estar ausente no genoma dos organismos. A análise de anticorrelação dos genes *purN* e *purT* mostrou que esses genes anticorrelacionam tanto em nível de gênero, quanto de espécie nos filo/classes de Firmicutes Bacilli, Firmicutes Clostridia, Alfabroteobacteria, Crenarchaeota e Euryarchaeota (Tabela 8B). No filo Actinobacteria a anticorrelação é percebida apenas em nível de gênero, esse desvio se deve ao viés da amostragem de gêneros e espécies.

Nos filos/classes Cianobacteria, Bacteroidetes, Chlorobi, Spirochaeta, Delta-, Epsilon-, Beta- e Gammaproteobacteria o valor de anticorrelação é inferior a 0,7. Esse valor não é indicativo de anticorrelação. Nesses filos/classes na maioria dos gêneros/espécies ambos os genes estão presentes.

Nos filos/classes Aquificae, Chloroflexi, Deinococcus-Thermus, Acidobacteria, Fusobacteria, Planctomycetes, Thermotogae, Verrucomicrobia, Dictyoglomi, Elusimicrobia e Nitrospirae os genes anticorrelacionam. Nesses filos/classes apenas *purN* esta presente. No entanto, devido ao viés amostral encontrado nesses Taxa, os resultados de anticorrelação devem ser considerados com cautela.

Em todos os casos em que os genes anticorrelacionam o gene predominante no filo/classe é *purN*, mais uma vez evidenciando a dominância desse gene em relação a *purT* nos filos/classes amostrados.

Apesar do viés amostral dos genomas completamente sequenciados em alguns grupos, o conjunto de resultados apresentados acima indica que o gene *purN* é preferencialmente utilizado por procariotos em relação ao gene *purT* e, que estes genes não estão igualmente distribuídos nas linhagens procarióticas. Tais resultados remetem as seguintes questões:

- Como duas enzimas foram selecionadas para desempenhar a mesma função durante a evolução dos procariotos e não apenas uma?
- Qual a possível explicação para PurN ser preferencialmente utilizada na maioria das linhagens procarióticas analisadas neste estudo?

Para tentar resolver tais questões é necessário explorar as informações oriundas da bioquímica e filogenia molecular das enzimas codificadas por esses genes.

O gene *purT* é menos distribuído que *purN* nas linhagens procarióticas, além disso, em 93%/94% dos gêneros/espécies que possuem o conjunto mínimo de genes da biossíntese de purinas, o contexto genômico de *purT* não envolve outros genes da via biossintética de purinas. Estes resultados, associados à filogenia de *purT* e *purK*, podem ser explicados por dois cenários.

Sabendo-se que PurT e PurK são enzimas homólogas (LI et al., 2009; ZHANG et al., 2008; KAPPOCK et al., 2000), é razoável inferir que os genes que codificam essas enzimas têm uma história evolutiva comum. A homologia entre PurT e PurK é notada por meio da sobreposição da estrutura terciária dessas enzimas (Figura 9).

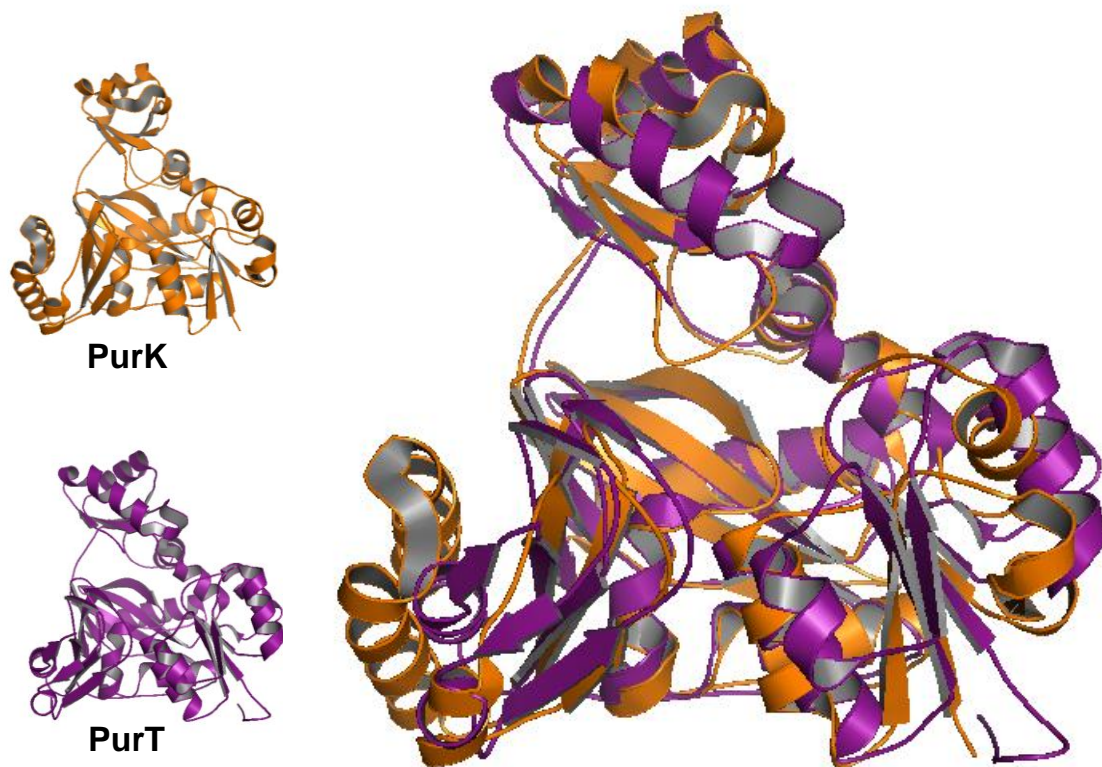


Figura 9. Sobreposição da estrutura terciária da PurT e PurK de *Escherichia coli*. A sobreposição foi realizada utilizando o programa PyMol (DeLano et al., 2002).

No primeiro cenário teria ocorrido um evento de recrutamento enzimático. Assim, o gene *purT* seria mais recente, ele teria surgido por um evento de duplicação do gene *purK* em alguma linhagem procariótica, por exemplo, o ancestral de γ/β -proteobactérias, visto que são grupos irmãos e no qual *purT* é mais abundante. Após a duplicação uma das cópias de *purK* teria divergido e adquirido especificidade por outro substrato, passando a atuar em outra etapa da via (Figura 10). Posteriormente, esse novo gene teria sido transferido horizontalmente para outras espécies procarióticas. Isto explicaria o fato de *purT* não estar presente em operons com outros genes da via biossintética de purinas na maioria das espécies procarióticas. Nesse cenário a topologia da árvore filogenética destas enzimas teria PurT emergindo dentro de um clado de PurK.

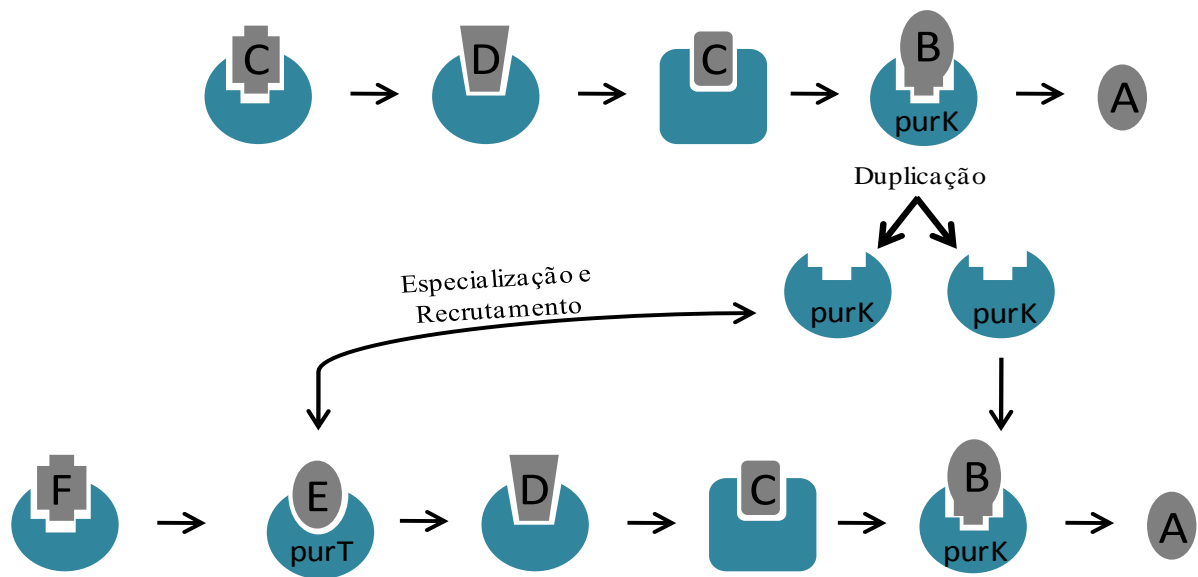


Figura 10. Cenário para evolução da PurT baseada na hipótese de recrutamento enzimático. Nesse cenário PurK que já atuava na síntese de purinas e duplicou. Uma das isoformas continuou com a sua função original e a outra, que deu origem a PurT, sofreu especiação e passou a atuar em outra etapa da via biossintética de purinas.

No entanto, *purT* está presente em genomas de espécies de Archaea, e bactérias Gram positivas, grupos monofiléticos que divergiram antes do surgimento de Proteobactéria. Isso sugere um segundo cenário (Figura 11) onde *purT* e *purK* evoluíram a partir da duplicação de um gene que codificava uma proteína com ampla especificidade cujas cópias divergiram funcionalmente originando os genes *purT* e *purK*, conforme proposto por Li et al. (2009) com base na comparação da estrutura terciária destas enzimas.

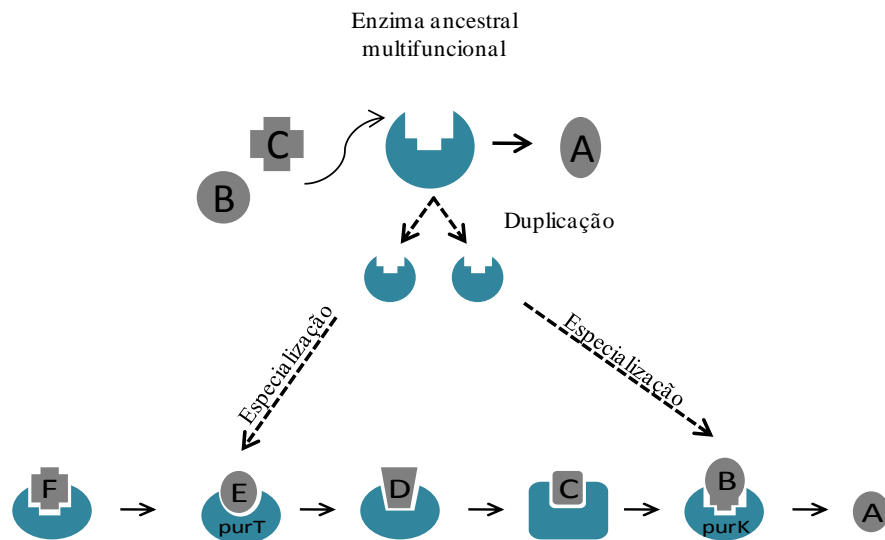


Figura 11. Cenário da evolução das enzimas PurT e PurK. Esse cenário sugere que a enzima ancestral de PurT e PurK estava presente no ancestral de todos os seres vivos. Essa enzima dotada de ampla especificidade duplicou e as isoformas dessas enzimas se especificaram em diferentes substratos e passaram a atuar em diferentes etapas da via.

8.5. Filogenia da enzima PurT

A topologia da árvore filogenética de PurT/PurK corrobora o segundo cenário proposto para origem dessa enzima, visto que PurT não emerge dentro de um clado da PurK o que torna *purT* um gene tão primitivo quanto *purK* (Figura 12).

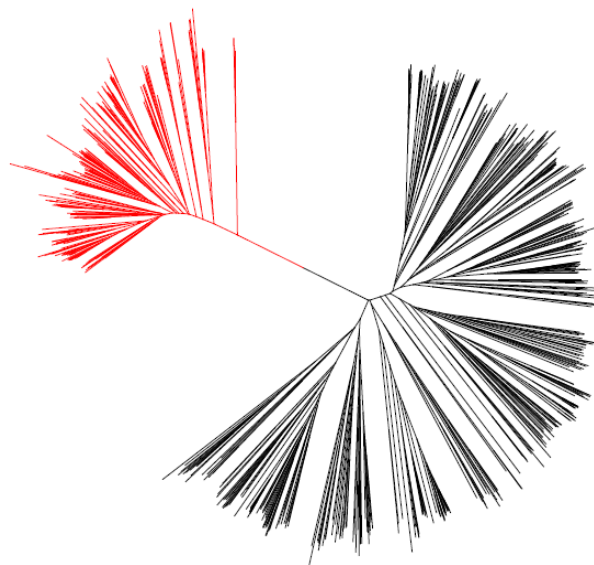


Figura 12. Árvore filogenética com seqüências de PurT e PurK. A árvore mostra que PurK (preto) e PurT (vermelho) formam cladros distintos apesar de serem enzimas homólogas. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007).

Doolittle e Brown (1994) elaboraram uma hipótese elegante para explicar as diferenças observadas na maquinaria informacional dos organismos dos domínios Archaea e Bacteria. Eles postularam que as linhagens Archaea e Bacteria divergiram do ancestral comum dos seres vivos no período, denominado por eles, de evolução darwiniana progressiva. Neste período ocorreu a fixação de mutações que aumentaram a precisão, velocidade e eficiência de transferência de informação e, assim, a adaptabilidade das células ou unidades pré-celulares existentes. Nessa era radicalmente diferente no que dizem respeito às condições ambientais, os componentes da maquinaria informacional eram compostos por proteínas não homólogas funcionalmente equivalentes. Estes análogos teriam sido posteriormente, seletivamente mantidas ou perdidas nas linhagens procarióticas Archaea e Bactéria após sua divergência do ancestral comum dos seres vivos.

Esta é uma hipótese factível para explicar a existência de dois genes desempenhando a mesma função na terceira etapa da biossíntese de purinas em procariotos. Isto implica que estes genes já estavam presentes no ancestral comum dos seres vivos, portanto, a sua distribuição em procariotos seria resultantes de eventos de perda ou manutenção seletiva nas linhagens procarióticas.

Contudo, a topologia da árvore filogenética da PurT mostra que a transferência lateral de genes (TLG) também influenciou significativamente a distribuição desse gene nas espécies procarióticas (Figura 13). As espécies de Gram positivas que possuem *purT* pertencem ao filo Actinobacteria e a classe Bacilli do Filo Firmicutes. Contudo, na filogenia de PurT esses grupos não formam um clado monofilético.

As espécies de Bacilli agrupam com espécies da classe Deltaproteobacteria, além disso, não possuem no seu contexto genômico outros genes da biossíntese de purinas (Figura 14A e Anexo 7). Isso sugere que as *purTs* dessa classe de Gram positivas foram adquiridas por TLG de espécie de Deltaproteobacteria.

As espécies do Filo Actinobacteria estão distribuídas em clados distintos na filogenia de PurT. As espécies do gênero *Bifidobacterium* formam um clado com espécies de Cianobacteria e outra espécie de *Actinobacteria* do gênero *Mobiluncus*. Esse clado agrupa com um clado de Gram negativas composto por espécies de *Bacteroidetes/Chlorobi*, Beta, Alfa e Gamaproteobacterias e *Cianobacteria* (Figura 14B).

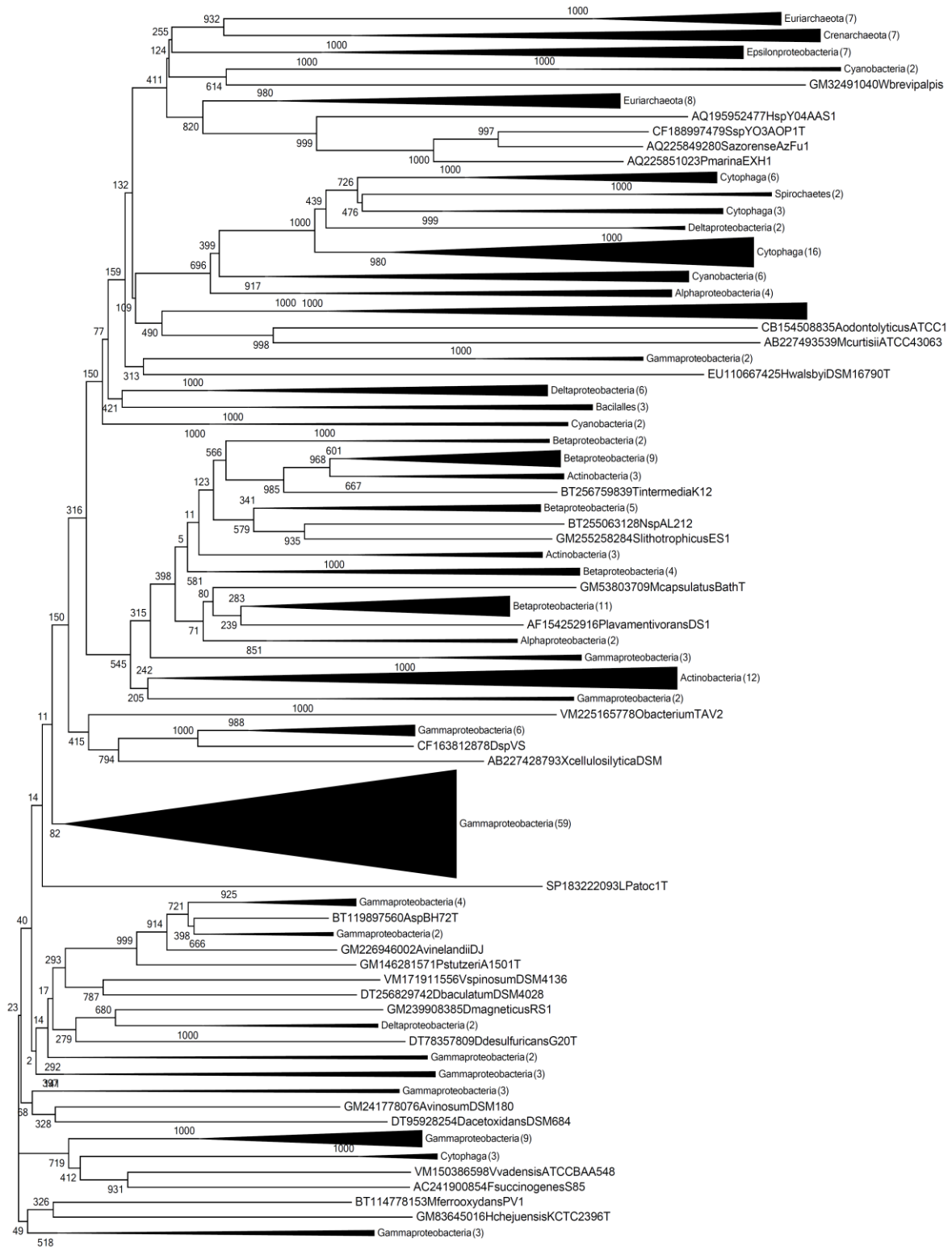


Figura 13. Árvore filogenética não enraizada de PurT. Os clados que formam um grupo monofilético em nível de filo ou classe, foram colapsados. O nome filo/classe e o número de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007). A identificação da sequência é composta pela abreviação do filo/classe a qual a espécie pertence, mais o número de identificação da sequência seguido pelo nome da espécie com o gênero abreviado.

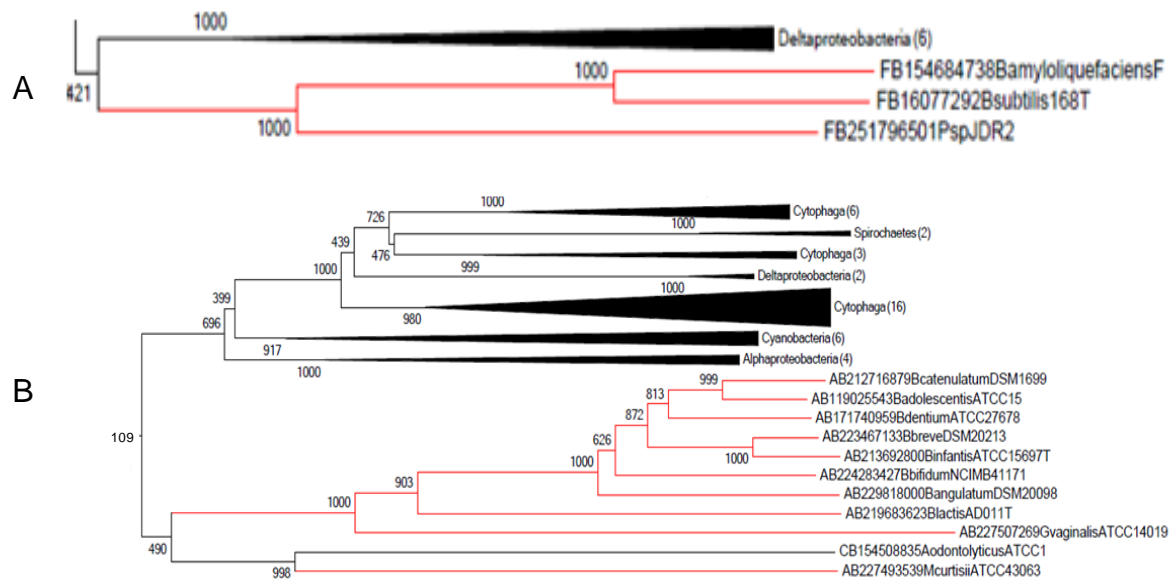


Figura 14. Sub-clado da árvore de PurT. **(A)** Espécies do gênero *Bifidobacterium* agrupando com Gram negativas. Os ramos em vermelho destacam as espécies de *Actinobacteria*. **(B)** Espécies de Firmicutes Bacilli que agrupam com espécies de Deltaproteobacteria. Os ramos em vermelho destacam as espécies da Classe Bacilli. O nome filo/classe e o numero de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007).

As espécies do gênero *Micobacterium*, são encontradas formando um clado com espécies de Betaproteobacteria (Figura 15A) enquanto que as espécies do gênero *Corinebacterium* agrupam com espécies de Gamaproteobacteria (Figura 15B). Estes resultados indicam que a *purT* dos gêneros *Bifidobacterium*, *Micobacterium* e *Corinebacterium* têm origem de diferentes Taxa de Gram negativa.

Resultado semelhante pode ser observado dentre as bactérias Gram negativas. A PurT de grupos reconhecidamente monofiléticos destas bactérias ficaram distribuídas em clados distintos, não relacionados filogeneticamente (Figura 14).

Estes resultados sugerem que a TLG foi um evento modelador na distribuição de *purT* nas linhagens procarióticas e pode ser usado para explicar a existência de *purT* no genoma de espécies que pertencem a grupos nos quais o padrão estrutural II, presença apenas de *purN*, é predominante. Por exemplo, em *Clostridia*, *Actinobacteria*, Alphaproteobacteria, *Euryarchaeota* e Deltaproteobacteria.

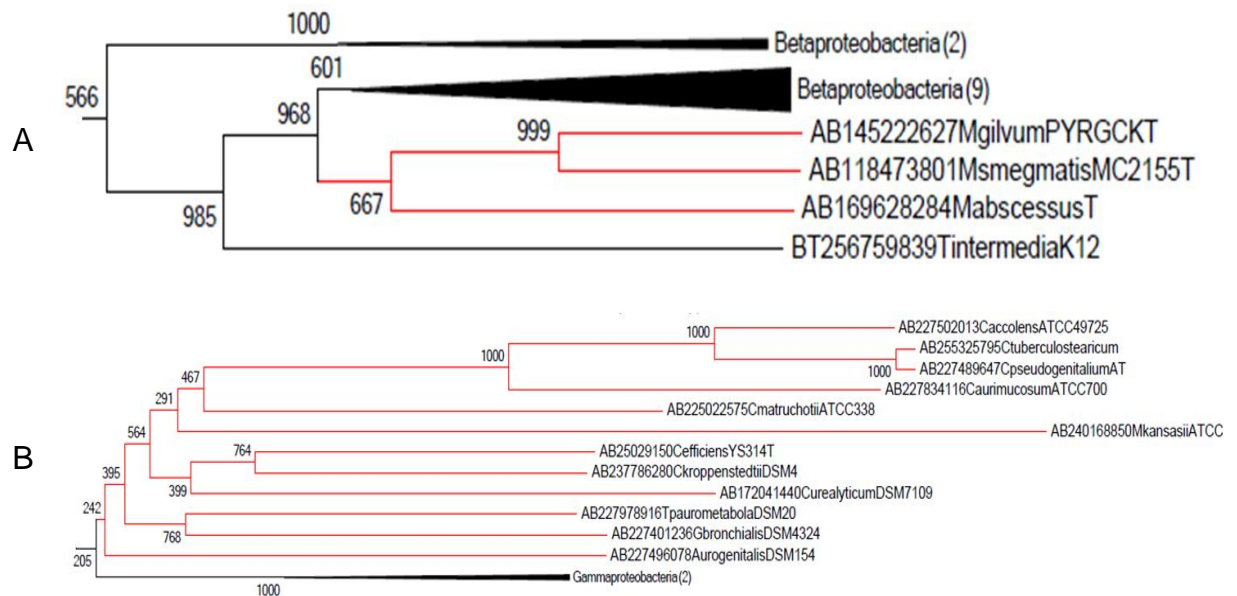


Figura 15. Sub-clados da árvore de PurT. **(A)** Espécies do gênero *Micobacterium* agrupando com Betaproteobacteria. **(B)** Espécies do gênero *Corinebacterium* agrupando com espécies da classe Gamaproteobacteria. Os ramos em vermelho destacam as espécies de *Actinobacteria*. O nome filo/classe e o número de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007).

As características bioquímicas de PurT e PurN podem ajudar a explicar a prevalência de *purN* nos genomas procarióticos. Estas enzimas atuam na terceira etapa da via biossintética de purinas, onde o composto intermediário GAR é formilado e convertido em FGAR. PurN cataliza essa reação utilizando o N¹⁰-formiltetrahydrofolato como fonte de formato enquanto que PurT, além de consumir ATP ao catalizar a sua reação, utiliza o formato oriundo da quebra do N¹⁰-formilTHF em tetrahydrofolato e formato, realizada pela enzima PurU .

O valor de K_m para o intermediário GAR é maior para PurN (19 μM) do que para PurT (10 μM) (Tabela 4). Ou seja, a afinidade de PurT pelo intermediário GAR é aproximadamente duas vezes maior que para PurN. O número máximo de reações realizadas por segundo é maior para PurT ($K_{cat} = 37,6 \pm 0,8$). No entanto, a eficiência catalítica da PurN é maior ($K_{cat}/K_m = 0,16$) quando comparada a eficiência catalítica da PurT ($K_{cat}/K_m = 0,12$). Isso indica que apesar de PurT ser mais eficiente ao se ligar ao GAR, PurN executa a reação de forma mais eficiente. Ademais, PurN não depende de uma terceira enzima para gerar o formato a partir do N¹⁰-formilTHF, ela é capaz de realizar esta reação e em seguida formilar GAR diretamente, e sem o gasto de ATP. Ao contrário de PurT que além de necessitar de uma enzima para

disponibilizar o formato, a PurU, ainda consome ATP para a ligação do grupo formil, formando então o FGAR.

Tabela 4. Valores cinéticos das enzimas PurN e PurT de *Escherichia coli*

Enzima	Substrato	K_m	K_{cat}	K_{cat}/K_m	Fonte
PurN	GAR	19 ± 4 (μM)	-	-	Nixon et al., 2000 e Warren et al., 1996
	N^{10} -formilTHF	84.8 ± 5.9 (μM)	13.5 ± 0.8 (s^{-1})	0.16	Nagy et al., 1994
PurT	GAR	10.1 ± 0.51 (μM)	-	-	Marolewski et al., 1994 e 1997
	Formato	319 ± 15 (μM)	37.6 ± 0.8 (s^{-1})	0.12	Li et al., 2009

Com base no exposto acima infere-se que ao longo da sua história evolutiva os organismos procarióticos priorizaram uma enzima que produz de FGAR em uma reação direta, sem o gasto de ATP, ao invés de uma reação que emprega duas enzimas e com gasto de energia pela célula.

Mesmo a inferência acima sendo verdadeira e o fato de a TLG ter contribuído para a distribuição de *purT*, ela não explica a distribuição de *purT* em treze dos 26 Filos/Classes incluídos neste estudo. Uma hipótese tentadora é que *purT* possui outra função além da já conhecida na biossíntese de purinas. O formato utilizado por PurT é disponibilizado pela quebra do Formiltetrahydrofolato (FormilTHF), reação catalisada pela PurU. Nagy et al. (1994) demonstraram que o composto Tetrahydrofolato (THF) é utilizado na síntese de glicina. Portanto, a quebra do Formiltetrahydrofolato, pela *purU* regula o balanço de THF, e conseqüentemente a síntese de glicina. Adicionalmente, a glicina regula a atividade de PurU. Talvez PurT desempenhe uma função adicional relacionada com a manutenção do balanço de THF e FormilTHF.

Thorndike e colaboradores (1989), demonstraram que Homofolato é eficiente na inibição de PurN na síntese de purinas. Visto que PurU é homólogo a PurN (NAGY et al., 1994) esse composto também pode inibir essa enzima. Como PurU faz a quebra do FormilTHF em THF e formato, conseqüentemente o Homofolato também pode afetar a atividade de PurT. No entanto, como a PurU não é a única fonte de formato das células, as espécies que possuem PurT podem não seriam afetadas pelo composto inibidor.

Os resultados mostram que *purN* é predominante nos gêneros/espécies dos Domínios Archaea e Bactéria incluídos nesse estudo. Apesar disso, *purT* esta

presente em muitas espécies de bactérias causadoras de importantes infecções para animais e plantas (Tabela 5). Portanto, esse gene constitui um potencial alvo para fármacos antibacterianos, visto que *purT* não está presente em Eucariotos.

Tabela 5. Relação de bactérias patogênicas para animais e plantas que possuem o gene *purT*.

Bactérias patogênicas para animais	
<i>Bordetella pertussis</i>	Coqueluche
<i>Corynebacterium diphtheriae</i>	Difteria
	Infecções do trato urinário
<i>Escherichia coli</i>	Diarréia
	Meningite
<i>E. coli</i> (O157:H7)	Colite hemorrágica
<i>Francisella tularensis</i>	Tularemia
<i>Leptospira interrogans</i>	Leptospirose
<i>Mycobacterium tuberculosis</i>	Tuberculose
	Infecção do trato Gastrointestinal, urinário e respiratório, da pele, olhos e ouvidos
<i>Pseudomonas aeruginosa</i>	Gastrointestinal tract infection
	Infecção do Sistema Nervoso Central
	Pneumonia secundária
	Infecções dos ossos e articulações
	Febre tifóide
<i>Salmonella typhi</i>	Disintéria
	Colite
<i>Salmonella typhimurium</i>	Salmonelose com gastroenterite e enterocolite
<i>Shigella sonnei</i>	Desinteira
<i>Treponema pallidum</i>	Sífilis
<i>Vibrio cholerae</i>	Cólera
Bactérias Fitopatogênicas	
<i>Erwinia carotovora</i> subsp. <i>atroseptica</i> SCRI1043	Podridão mole (Tomate)
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> str. 306	Cancro cítrico
<i>Pseudomonas syringae</i> pv. <i>phaseolicola</i> 1448 ^a	Podridão

9. CONCLUSÕES

- Os genes *purT* e *purN* não estão igualmente distribuídos nas linhagens procarióticas;
- O gene *purN* é preferencialmente utilizado pelas linhagens procarióticas;
- A transferência horizontal de genes foi um fator relevante na distribuição de *purT* nas linhagens procarióticas.

CAPÍTULO II

DIVERSIDADE ESTRUTURAL E EVOLUÇÃO DO SEXTO PASSO DA
BIOSSÍNTESE DE PURINAS EM LINHAGENS PROCARIÓTICAS

10. INTRODUÇÃO

Na sexta etapa da via biossintética de purinas o intermediário AIR é carboxilado na posição 5 no segundo do anel purínico produzindo CAIR. Esta reação pode ocorrer por diferentes mecanismos. Em animais a reação é catalisada por uma única enzima, a PurE classe II, denominada daqui em diante de PurE II. Essa enzima utiliza CO_2 para carboxilar o AIR. Em plantas, fungos e procariotos duas enzimas são responsáveis pela reação de carboxilação do AIR. Primeiro a PurK, utilizando bicarbonato (HCO_3^-) e ATP faz a adição do grupo carboxila no grupo amino livre do AIR. Em seguida uma PurE classe I, denominada daqui em diante de PurE I, transfere o grupo carboxila para a posição C4 do anel purínico em construção, formando o CAIR (Figura 16).

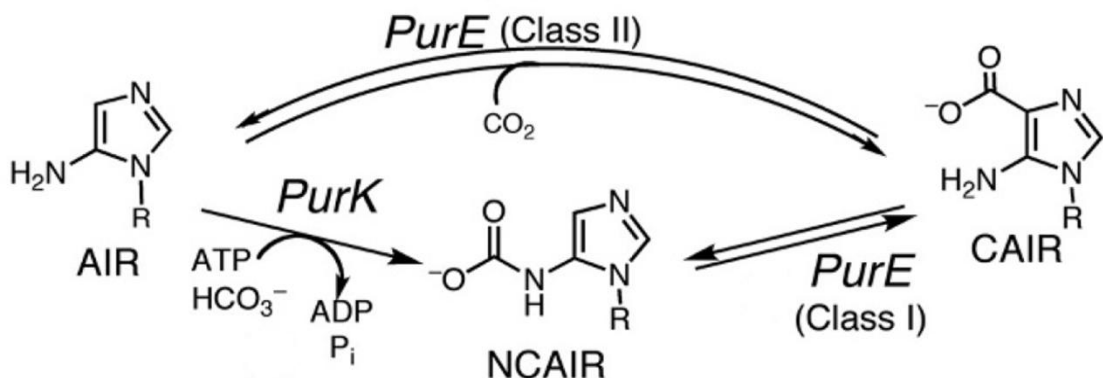


Figura 16. Sexta reação da via biossintética de purinas. Fonte: ZHANG et al., 2008.

A PurE II é descrita na literatura como uma enzima presente somente em animais. Não é conhecida a presença de PurK em metazoários. Portanto, a condição PurK/PurE I é uma característica dos organismos inferiores. Contudo, alguns relatos na literatura sugerem que a PurE II não é uma exclusividade de metazoários. Por exemplo, Hamilton e Reeve (1985) demonstraram que a PurE de *Methanothermobacter thermotrophicus* e *Methanobrevibacter smithii* é capaz de complementar *E. coli* com o gene *purK* nocauteado.

Devido à importância da biossíntese de purinas, e sabendo-se que diferentes conjuntos enzimáticos catalisam a reação em animais, procariotos e fungos, o conhecimento da distribuição dos genes *purE* e *purK* nesses organismos, mostra-se importante do ponto de vista farmacológico. As informações obtidas partir do estudo

dos mecanismos de ação desses genes, aliados ao conhecimento sobre sua distribuição nas linhagens procarióticas, podem ser utilizados no desenvolvimento e uso dirigido de drogas com ação antibacterianas e antifúngicas.

11. OBJETIVOS ESPECÍFICOS

- Fazer a genômica comparativa e contexto genômico dos genes *purK* e *purE* em procariotos;
- Fazer a reconstrução filogenética das enzimas PurK e PurE.

12. METODOLOGIA

12.1. Busca pelos genes canônicos da via biossintética de purinas

Foram feitas buscas pelos genes *purE* e *purK* em 815 genomas completamente sequenciados e anotados disponíveis na base de dados do NCBI até 16 de fevereiro de 2009, representando 535 espécies dos Domínios Archaea e Bacteria. Durante a busca foram usados os programas BLASTP e TBLASTN (ALTSCHUL et al., 1997). O nível e extensão de similaridade, assim como a presença de domínios conservados típicos de cada produto gênico, foram utilizados como critério de homologia para a busca das sequências dos genes em questão. As sequências de *Escherichia coli* foram utilizadas como *query* para a busca dos genes. Esse organismo foi escolhido por possuir todos os genes de interesse nesse estudo caracterizados funcionalmente.

12.2. Análise da diversidade estrutural da sexta etapa da via biossintética de purinas

Durante a busca pelos genes *purK* e *purE*, foram anotados o padrão de presença/ausência e a quantidade de cópias dos genes canônicos da via biossintética de purinas. No entanto, os estudos de genômica comparativa e contexto genômico foram conduzidos apenas com os organismos que possuem o conjunto mínimo de genes conhecidos necessário para a síntese do IMP, exceto para o Filo Crenarchaeota. Os Filos e Classes incluídos na análise, assim como o

número de gêneros e espécies amostradas em cada Filo/Classe são apresentados na Tabela 1.

12.3. Contexto genômico

Durante a busca pelas sequências dos genes *purE* e *purK*, o contexto genômico dos organismos pesquisados foi anotado, permitindo acessar informações sobre a organização genômica dos genes estudados. Para cada gene pesquisado, foram registrados os genes que os flanqueiam, gerando uma figura onde foram representados os genes da via biossintética de purinas, e os genes que os flanqueiam.

12.4. Alinhamento múltiplo e filogenia das enzimas *purK* e *purE*.

As sequências dos genes recuperadas nas buscas realizadas nas bases de dados do NCBI foram alinhadas códon a códon no programa MEGA 4 (TAMURA et al., 2007). Os alinhamentos múltiplos foram editados com base no alinhamento estrutural das estruturas terciária das proteínas disponíveis na base de dados do PDB (<http://www.rcsb.org/pdb/home/home.do>). As sequências redundantes e incompletas foram excluídas. A visualização e sobreposição das estruturas terciárias das proteínas foram feitas nos programas TOPMACH (SIPPL e WIEDERSTEIN, 2008) e PYMOL (DeLano et al., 2002).

As análises filogenéticas foram realizadas utilizando as sequências de aminoácidos oriundas da tradução das sequências de nucleotídeos dos genes *purK* e *purE*. Nessas análises não foram considerados os sítios com GAPs. A reconstrução filogenética foi realizada com base no método de *Neighbor-joining* (SAITOU e NEI, 1987), utilizando o programa ClustalX.

13. RESULTADOS E DISCUSSÃO

13.1. Genômica comparativa de *purE* e *purK* nas linhagens procarióticas

O gene *purE* está presente em todos os 271/473 gêneros/espécies de microrganismos que possuem o conjunto mínimo de genes da biossíntese de purinas (Tabela 6). Por suas vez, o gene *purK* é encontrado em 203/359

gêneros/espécies que possuem a via. Não foram encontradas espécies que possuem apenas o gene *purK*. No genoma de algumas espécies foram encontradas mais de uma cópia dos genes *purE* e *purK* (ANEXOS 3 e 5). Estes casos serão discutidos mais adiante.

Todas as espécies dos Filos Tenericutes e Clamydiae analisadas nesse estudo não possuem os genes *purE* e *purK* (Tabela 6). Como exposto no capítulo anterior, esses Filos são compostos por organismos parasitas intracelulares. Este padrão estrutural também foi observado nas espécies parasitas intracelulares ou simbiontes pertencentes aos filios Crenarchaeota, Euryarchaeota, Firmicutes Bacili, Bacteroidetes, Spirochaetae, Alfa, Epsilon e Gamaproteobacteria (Tabela 6).

Padrão estrutural	Genes pur		Filos/Classes																											
	E	K	CR	EU	FB	FC	FT	AB	CB	AQ	CF	DTH	AC	FU	PLC	THR	BC	CL	VM	DC	EB	NT	CH	SP	AF	BT	GM	DT	EP	
I	5/10	7/10	15/48	6/6			20/41	11/13	3/3	3/5	2/3	1/1				2/4	4/5	1/1	1/1					1/3	43/67	22/44	52/91	2/2		
II	2/2	15/20		11/22			1/2			1/2		1/1	1/1	1/1		3/4	4/6	4/9	2/2	1/2	2/2	1/1		1/1	1/1		6/7	11/18	7/11	
III																														
IV			4/4	2/2	1/3		7/20										2/2							3/7	2/8	3/10		3/4		1/2

Tabela 6. Distribuição e padrão de ocorrência dos genes *purE* e *purK* em procariotos. As duas primeiras colunas representam a ocorrência dos genes *purE* e *purK*. Nas outras colunas estão listados os filios de procariotos, indicando o número de gêneros/espécies que cada filo possui. Amarelo indica presença do gene e branco ausência.

A genômica comparativa mostra que os genes *purE* e *purK* apresentam dois padrões estruturais de distribuição: (I), onde ambos os genes estão presentes no genoma dos organismos e (II), apenas *purE* esta presente no genoma (Tabela 6). O padrão estrutural I é encontrado em dezenove Filios/Classes, e o padrão estrutural II é encontrado em vinte dos vinte sete Filios/Classe amostrados (Tabela 6). Contudo, os resultados a seguir devem ser analisados com parcimônia, considerando o viés amostral dos genomas sequenciados.

Todos os Filios de Bactérias Gram positivas estão amostrados neste estudo, contudo, apenas os Filios Actinobacteria e Firmicutes possuem organismos com o conjunto mínimo de genes da biossíntese de purinas. Deste último, duas das três Classes estão representadas, Bacilli e Clostridia, incluindo espécies de todas as ordens da Classe Bacilli e de três das quatro ordens da Classe Clostridia. O Filio Actinobacteria possui apenas uma Classe. Essas taxa são reconhecidamente grupos monofiléticos.

Nas Gram positivas a estrutura I foi encontrada em todos os gêneros/espécies da Classe Bacilli e em 20/41 gêneros/espécies do Filo Actinobacteria. Neste último a estrutura II foi encontrada em apenas 1/2 gênero/espécie (Tabela 6). Na Classe Clostridia a estrutura II foi a mais freqüente em nível de gênero e espécie, contudo, sem a mesma predominância do padrão estrutural I nos demais taxa de Gram positivas. Enquanto 11/22 gêneros/espécies da Classe Clostridia possuem a estrutura II, 6/6 gêneros/espécies possuem a estrutura I (Tabela 6).

Nos Taxa de bactérias Gram negativas a estrutura I foi encontrada em todos os gêneros/espécies de Cianobacteria, Aquificae, Deinococcus-Thermus e Betaproteobacteria e, foi predominante nos gêneros/espécies de Alfa e Gamaproteobacteria (Tabela 6). As Alfa, Beta e Gamaproteobacterias são taxa estreitamente relacionados do ponto de vista filogenético, o que pode explicar a similaridade observada entre eles quanto à estrutura predominante nestes grupos bacterianos. Os resultados observados para o Filo Aquificae devem ser vistos com cautela, uma vez que apenas três espécies de diferentes gêneros deste grupo foram incluídas neste estudo. O mesmo pode ser dito para o Filo Deinococcus-Thermus, onde apenas 2/3 gêneros/espécies foram incluídas nas análises. Os 11/13 gêneros/espécies do Filo Cyanobacteria incluídos nesta análise representam cinco das sete Ordens do grupo, logo, é razoável inferir que estrutura I é a condição ancestral e estava presente no ancestral comum desse Filo.

A estrutura II foi encontrada em todos os gêneros/espécies de Fusobacteria, Plantomicetales, Dictioclomi, Elusimicrobia, Nitrospira e Epsilonproteobacteria e foi predominante em Chlorobi e Deltaproteobacterias. No entanto, na maioria destes filios/classes é observado um viés amostral (Tabela 6). Os Filos Fusobacteria, Plantomicetales e Nitrospira possuem apenas uma espécie amostrada. No filo Dictioclomi apenas um gênero, representado por duas espécies, foi amostrado, enquanto que no filo Elusimicrobia apenas duas espécies de dois gêneros foram amostrados.

No filo Bacteroidetes a estrutura II esta distribuída em seis espécies, enquanto a estrutura I esta presente em cinco. No entanto, em nível de gênero, ambas as estruturas estão igualmente distribuídas (Tabela 6). Portanto, a predominância da estrutura II em nível de espécie é claramente resultante do viés amostral dos genomas completamente sequenciados.

Na Classe Epsilonproteobacteria os 7/11 gêneros/espécies que possuem o conjunto mínimo de genes da biossíntese de purinas possuem estrutura II, enquanto que em Deltaproteobacteria 11/18 gêneros/espécies apresentam a estrutura II, e apenas dois gêneros/espécies possuem o padrão I (Tabela 6). Estas Classes de Proteobacteria são estreitamente relacionadas filogeneticamente, o que indica que a predominância da estrutura II nessas taxa é resultado da relação evolutiva entre eles.

No domínio Archaea as estruturas I e II foram encontradas nos dois Filos considerados nessa análise. Em Crenarchaeota a estrutura I é predominante, ela está presente em 5/10 gênero/espécies, enquanto a estrutura II está presente em apenas dois gêneros/espécies. No Filo Euryarchaeota é observado o contrário, a estrutura II é encontrada em 15/20 gêneros/espécies, e a estrutura I está presente em 7/10 gêneros/espécies.

A existência de grupos de microrganismos procarióticos com *purE* e sem *purK* foi uma surpresa, pois implica que estes microrganismos ou teriam um análogo da *purK* ou teriam uma *purE* II, que é capaz de carboxilar o AIR diretamente no carbono 4. Ambas as hipóteses são igualmente surpreendentes. A primeira porque implica na existência de genes ainda desconhecidos envolvidos naquela que é, certamente, uma das vias biossintéticas mais estudadas e, a segunda porque a *purE* II só tinha sido encontrada em eucariotos (HOSKINS, 2007; MUELLER, 1994; DRAZEK, 1995; CHUNG, 1996; SORENSEN, 1997; FIRESTINE, 1998; THODEN et al., 1999; MEYER, 1992; CHEN, 1990; FIRESTINE e DAVISSON, 1994).

Hamilton e Reeve (1985) constataram por meio de ensaios de complementação que a *purE* das Archaeas *Methanothermobacter thermautotrophicus* e *Methanobrevibacter smithii* são capazes de complementar mutantes para os genes *purE* e *puK* de *E. coli* auxotróficos para purinas. Estas espécies de Archaea estão incluídas neste trabalho e de acordo com a genômica comparativa possuem apenas o gene *purE* em seus genomas. Em conjunto estes resultados sustentam a hipótese da existência de *purE* II em procariotos.

13.2. Filogenia de PurE

Para testar a hipótese da existência de *purE* II em procariotos foi realizada a reconstrução filogenética do gene *purE*. Nesta análise estão incluídas sequências de aminoácidos de PurEs de procariotos com e sem o gene *purK*, e PurEs II de

metazoários caracterizadas funcionalmente. Dentre as PurEs procarióticas também estão incluídas PurE I caracterizadas funcionalmente como a PurE I de *Escherichia coli*.

Na topologia da árvore filogenética a maioria das *purEs* de organismos sem *purK* ficaram dentro do mesmo clado, incluindo as PurEs de *Methanothermobacter thermautotrophicus* e *Methanobrevibacter smithii* e as PurEs II de metazoários (Figura 17). Este clado será referido daqui em diante como clado II (Figura 18).

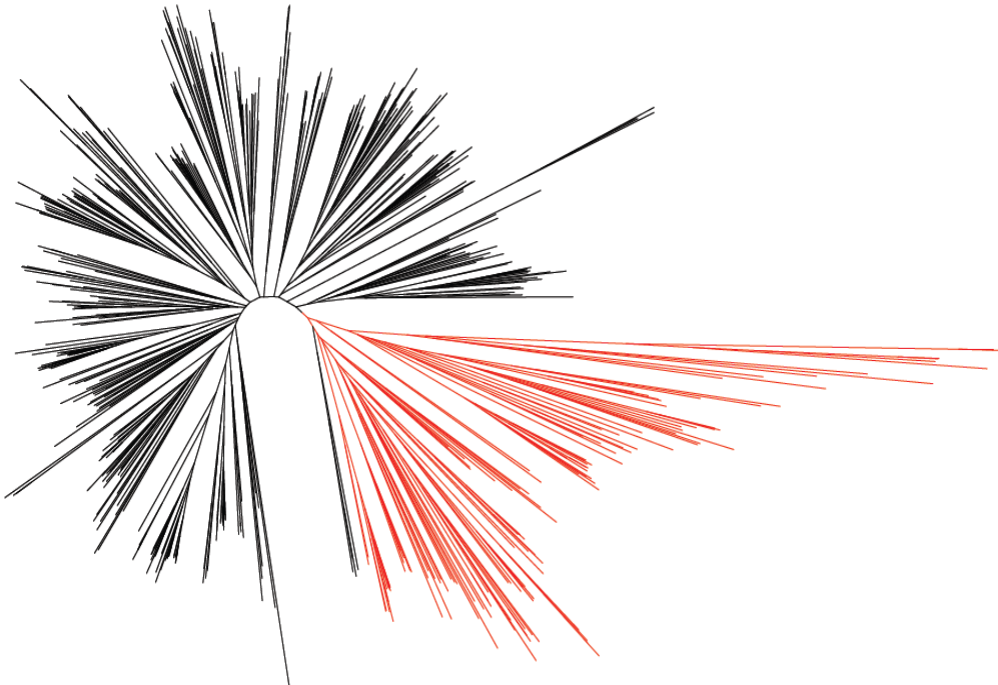


Figura 17. A filogenia da PurE mostra que a maioria das espécies que possuem somente PurE agrupam com as sequências de PurE II de Eucariotos formando o clado II (vermelho). Os outros clados são compostos por espécies que possuem PurE e PurK formando o clado I (preto). A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007).

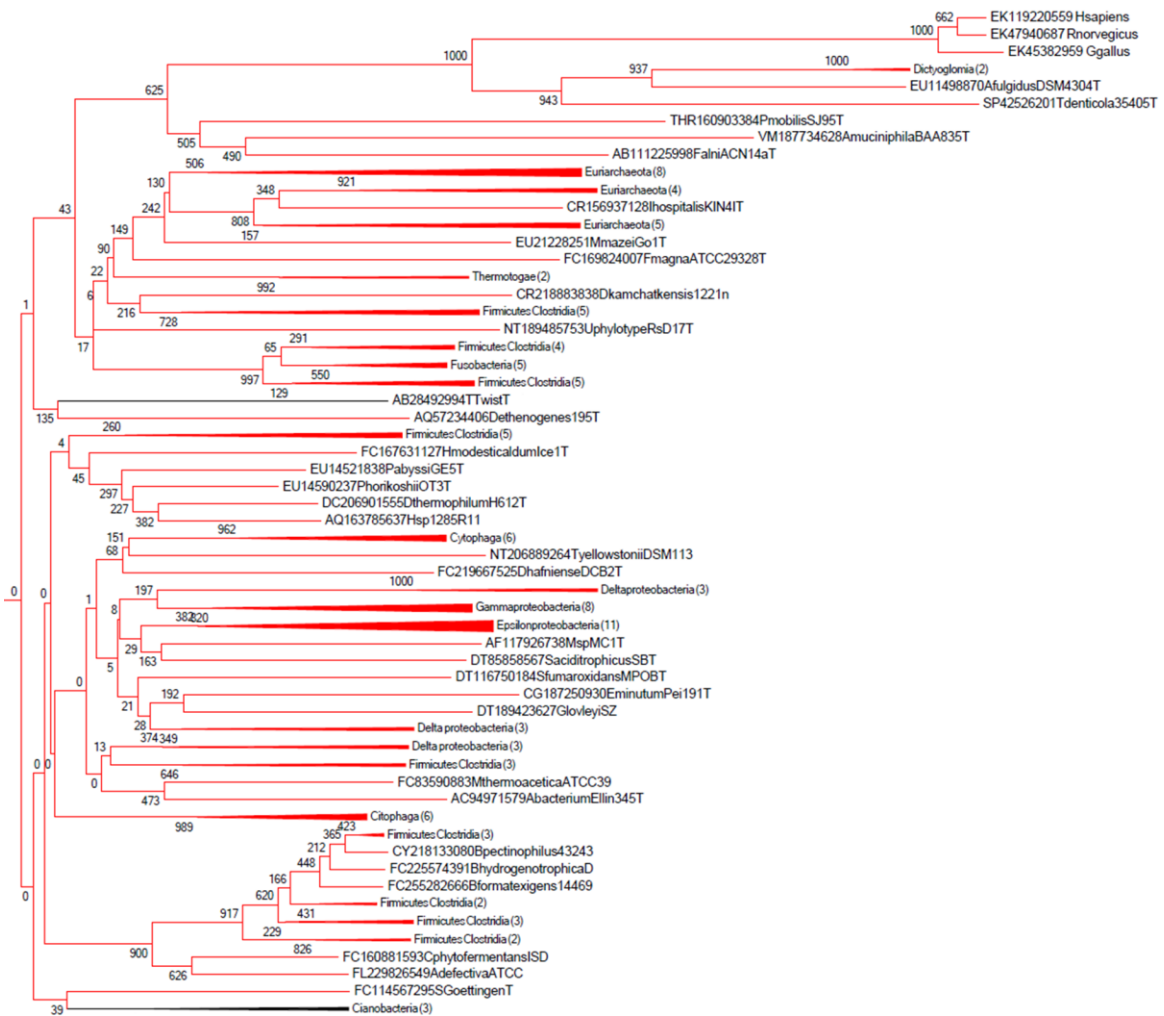


Figura 18. Clado II da árvore filogenética da figura 20. Neste clado esta a maioria das espécies procarióticas que não possuem PurK. E as PurEs II de eucariotos (vermelho). A espécie *Tropheryma whipplei str Twist* e as Cianobactérias presentes nesse clado possuem PurK (preto). O nome filo/classe e o numero de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007).

A PurE das espécies de Cianobactérias *Prochlorococcus marinus AS960*, *Prochlorococcus marinus NATL1A* e *Prochlorococcus marinus subsp. pastoris CCMP1986* e a Actinobactéria *Tropheryma whipplei str Twist*, aparecem no clado II, porém, essas espécies possuem PurK (Figura 18). No entanto, quando a análise filogenética é feita somente com espécies de procariotos a PurE desses organismos agrupam no clado I junto com as outras espécies de Cianobacteria. Isso indica que a presença da PurE dessas espécies no clado II pode ser um artefato de Atração de Ramos Longos (ARL), causado pela PurE II de eucariotos, que são muito divergentes das PurEs procarióticas e estão presente nesse clado.

A análise de genômica comparativa do gene *purE* revelou que algumas espécies possuem mais de uma cópia desse gene no genoma. Em algumas espécies uma destas isoformas possui alta taxa de evolução, conforme observado na topologia da árvore filogenética (Figura 19 e Anexo 3), portanto, foram excluídas das análises filogenéticas posteriores para evitar um eventual viés causado pela atração dos ramos longos (ARL). Além disso, as sequências de espécies que são proximamente relacionadas também foram removidas com o objetivo de reduzir a quantidade de sequências redundantes (Anexo 4).

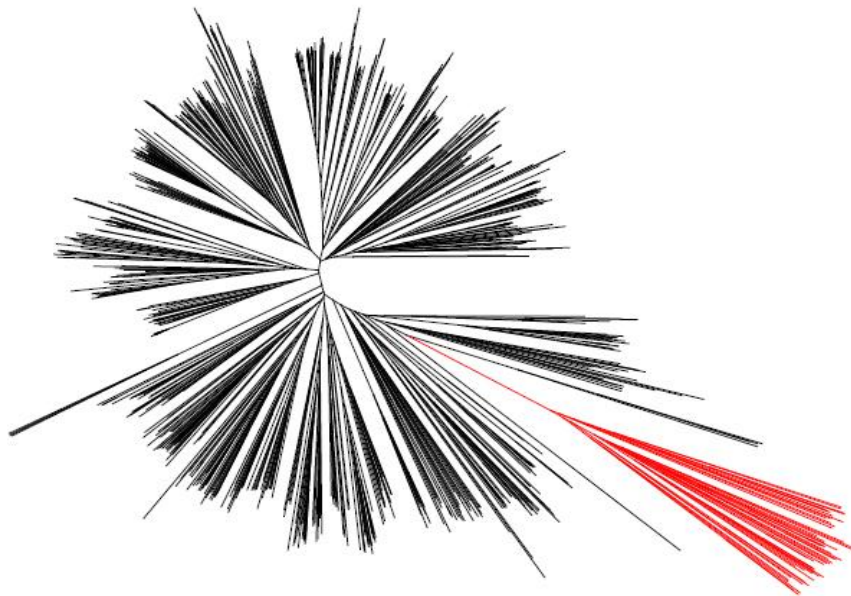
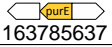
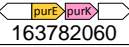
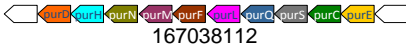

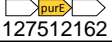
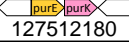
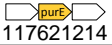

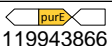
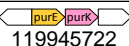


Figura 19. Árvore não enraizada da PurE destacando as sequências que possuem alta taxa de divergência e por isso foram excluídas das análises posteriores. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007).

As duas cópias de *purE* das espécies listadas na Tabela 4 foram mantidas nas análises filogenéticas. Estas isoformas não possuem elevada taxa de evolução e ficam em clados distintos na árvore filogenética, uma no clado I e a outra no clado II. A análise do contexto genômico dessas espécies, mostra que as isoformas do gene que agrupam com as PurEs I, estão em um contexto com o gene *purK*. No entanto, as isoformas que agrupam com as PurEs II não possuem em seus contexto genômico genes da biossíntese de purinas. Exceto em *Thermoanaerobacter pseudethanolicus* 33223 cuja isoforma que agrupa com as PurEs II está em um

contexto com outros genes da biossíntese de purinas, mas sem *purK* (Tabela 7). Estes resultados são uma forte evidência de que as *purEs* I e II existem em procariotos e que estão sujeitas a TLG.

Tabela 7. Espécies que possuem duas cópias do gene *purE*, uma delas agrupando no clado II e a outra no clado I. Abaixo do contexto são apresentados os números de acesso (GI) das *purEs* dessas espécies.

	Cópia que agrupa do clado II	Cópia que agrupa no clado I
AQ Hydrogenivirga sp. 128-5-R1-1	 163785637	 163782060
FC Thermoanaerobacter pseudethanolicus 33223	 167038112	 167038197
GM Shewanella loihica PV-4	 127512162	 127512180
GM Aeromonas hydrophila subsp. hydrophila 7966	 117621214	 117619504
GM Psychromonas ingrahamii 37	 119943866	 119945722

Os clados restantes possuem PurEs de organismos com *purK*, dentre elas as PurEs I caracterizadas funcionalmente, como a PurE de *Escherichia coli*. Estes clados serão referidos daqui em diante como clado I (Figura 20). As espécies *Opitutus terrae* PB90-1 e *Halothermothrix orenii* H 168, são os únicos organismos que não possuem o gene *purK* e cuja PurE está no clado I. Estas bactérias são anaeróbicas de vida livre (VAN PASSEL et al., 2011; MAVROMATIS et al., 2009), portanto, ou existe um análogo de *purK* nestes organismos ou a PurE codificada em seus genomas é capaz de catalizar a conversão de AIR em CAIR. Ou seja, são bons modelos de estudo para ampliar o conhecimento sobre a diversidade e evolução da biossíntese de purinas. Outra exceção é a espécie *Anaeromyxobacter dehalogenans* 2CP-C, no entanto, na filogenia com espécies apenas de procariotos essa espécie aparece no clado II, indicando o agrupamento encontrado também pode ser um efeito da ARL.

Os resultados das análises genômicas e filogenia molecular associados permitem inferir que *purE* II está amplamente distribuído nos Domínios Archaea e Bacteria. Durante a preparação deste trabalho Tranchimand e colaboradores (2011) publicaram a caracterização da PurE de *Treponema denticola* como o primeiro relato de uma PurE II em procariotos.

De fato, a PurE desta espécie de espiroqueta fica no clado II da filogenia da PurE realizada neste trabalho, assim como as PurEs das Archaeas *Methanothermobacter thermautotrophicus* e *Methanobrevibacter smithii*. Hamilton e Reeve (1985) constataram por meio de ensaios de complementação que o gene

purE destas Archaeas são capazes de complementar mutantes para os genes *purE* e *purK* de *E. coli* auxotróficos para purinas. Portanto, historicamente estes autores foram os primeiros a descreverem PurEs II em procariotos. O problema é que nessa época não era de conhecimento da comunidade científica a existência de PurEs I e II.

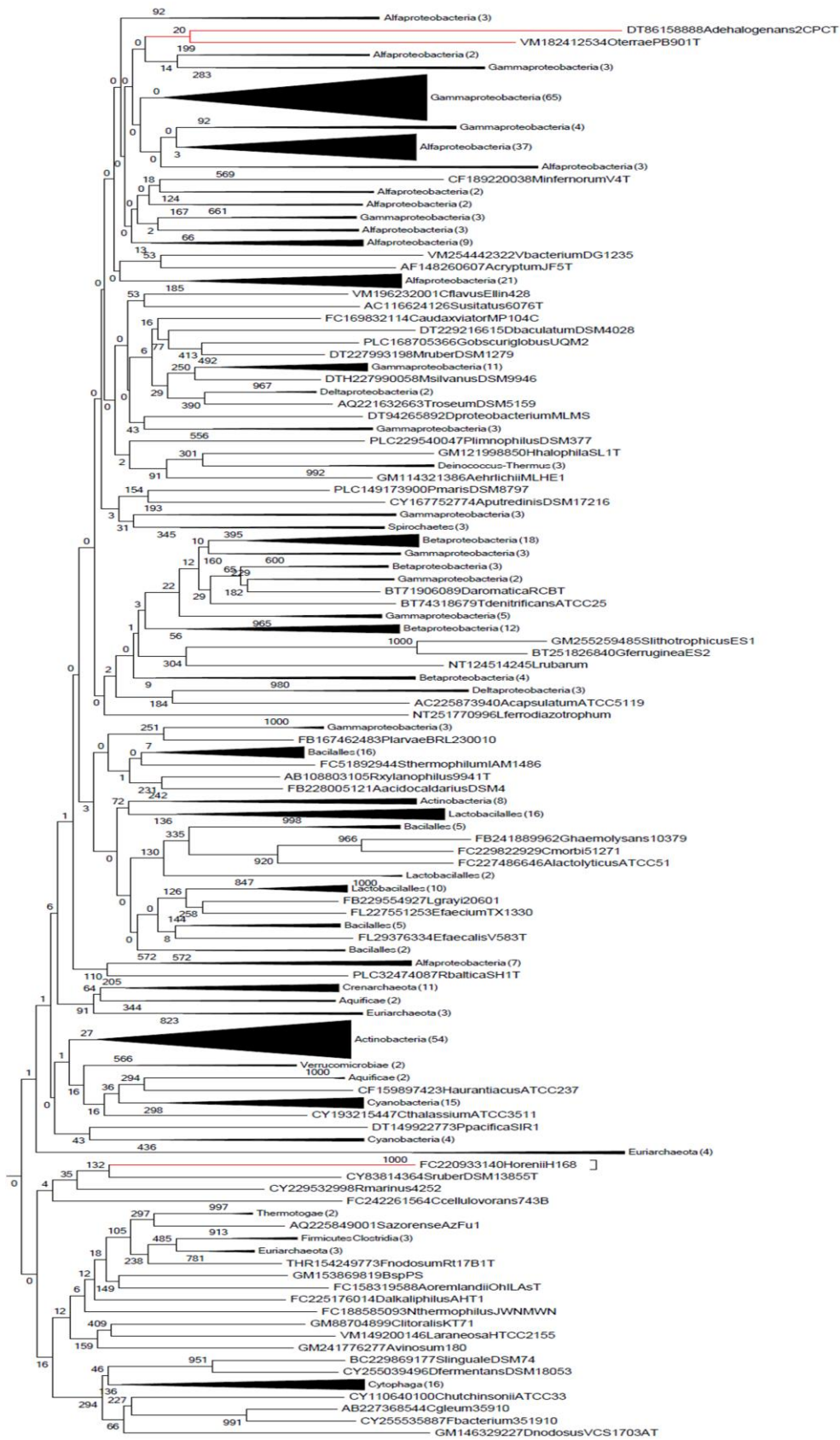


Figura 20. No clado I são encontradas espécies de organismos que possuem PurE e PurK. As exceções (vermelho) são espécies que possuem apenas PurE e que agrupam nesse clado possivelmente por causa da atração de ramos longos. O nome filo/classe e o número de espécies em cada clado foi especificado em cada ramo. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007).

A existência de PurE I e II em procariotos remete à questão sobre a condição primitiva do sexto passo da biossíntese de purinas. Pelo menos três cenários são factíveis. No primeiro, a PurE ancestral era uma carboxilase, ou seja, classe II e a PurE I evoluiu a partir dela. Em um cenário Alternativo, a PurE ancestral era uma mutase, classe I, com a PurE II evoluindo a partir dela. Por último, a PurE ancestral era multifuncional, com função de mutase e carboxilase, e, posteriormente, divergiu dando origem às PurEs I e II. A seguir estes cenários serão explorados à luz das informações oriundas deste trabalho e da literatura científica.

A ampla distribuição das PurEs I e II nas linhagens procarióticas sugerem que estas enzimas estavam presentes no ancestral comum dos seres vivos, entidade que viveu em um período radicalmente diferente do tempos atuais quanto ao tempo e modo. Neste período estima-se que a atmosfera primitiva da terra rica era CO₂ (Lichtenegger et al., 2010; Martin et al., 2007), conseqüentemente a quantidade de CO₂ dissolvido na água era maior do que a atual.

Este gás, quando dissolvido, reage com a água formando ácido carbônico (H₂CO₃), um ácido fraco que dissocia produzindo íons H⁺ e carbonato (HCO₃⁻). Estas reações ocorrem lentamente de modo que o CO₂ na forma dissolvida é predominante, o que é favorável para o primeiro cenário, visto que a PurE II usa CO₂ para carboxilar diretamente o AIR (KAPPOCK et al., 2000; EALICK, 2008). Além disso, esta enzima executa essa reação sozinha e, sem gasto de ATP, em contraste com a PurE I cujo substrato é o N⁵-CAIR produzido pela PurK, enzima que consome ATP e usa HCO₃⁻ como fonte de carbono para carboxilar o AIR

As características bioquímicas das PurEs I e II também favorecem o primeiro cenário. O K_m da PurE II para o AIR e o CO₂ é 76 µM e 0,8 mM, respectivamente (FIRESTINE e DAVISSON, 1994). Em contraste, o K_m da PurE I para o N⁵-CAIR é de 140 µM (Mueller et al., 1994). Portanto, a PurE II tem maior afinidade pelos seus substrato do que PurE I. A eficiência catalítica para a reação realizada pela PurE II (K_m/K_{cat}) é 3×10^7 (FIRESTINE e DAVISSON, 1994), superior a eficiência catalítica apresentada para PurE I ($K_m/K_{cat} = 330 \pm 70 \mu\text{M}^{-1} \text{min}^{-1}$) (FIRESTINE e DAVISSON, 1994). Os valores cinéticos das enzimas PurE II, PurE I e PurK podem ser encontrados na Tabela 8.

Tabela 8. Valores cinéticos das enzimas PurE II, PurE I e PurK

Enzima	Substrato	Km	Kcat	V _{max}	Kcat/Km	Fonte
PurE II	AIR	76 (μM)	2400 (min ⁻¹)		3 x 10 ⁷ (M ⁻¹ min ⁻¹)	Firestine e Davisson et al., 1994
	CO ₂	< 0.8 (mM)				
	HCO ₃	23 (mM)	1900 (min ⁻¹)		7 x 10 ⁴ (M ⁻¹ min ⁻¹)	
PurE I (<i>E. coli</i>)	AIR	110* / 430** (μM)		16* - 70** (μmol min ⁻¹ mg ⁻¹)		Mueller et al., 1994
	N5-CAIR	98 / 140 (μM)		70 (μmol min ⁻¹ mg ⁻¹)	2.2 x 10 ⁷ (M ⁻¹ min ⁻¹)	Firestine e Davisson et al., 1998
	HCO ₃	110 (mM)	1200 (min ⁻¹)			Mueller et al., 1994 Meyer et al., 1992
PurE I (Fungo)	N5-CAIR	33.6 ± 4.1 (μM)			1.6 x 10 ⁸ (M ⁻¹ min ⁻¹)	Firestine e Davisson et al., 1998
	AIR	26 (μM)	52 (s ⁻¹)			
PurK (<i>E. coli</i>)	HCO ₃					Meyer et al., 1992
	ATP	90 (μM)			2 (μM ⁻¹ s ⁻¹)	
PurK (Fungo)	AIR	6.6 ± 0.6 (μM)			2.4 x 10 ⁸ (M ⁻¹ min ⁻¹)	Firestine e Davisson et al., 1998
	ATP	16.5 ± 1.7 (μM)			6.4 x 10 ⁷ (M ⁻¹ min ⁻¹)	

MEYER e colaboradores (1992) demonstraram que em condições de altas concentrações de CO₂, o AIR é convertido não enzimaticamente em CAIR. Além disso, existem relatos de mutantes para *purK* de espécies pertencentes a taxa distintos (Tabela 9) que tiveram a taxa de crescimento atenuada ou são auxotróficos condicionais, sendo capazes de crescer sob altas concentrações de CO₂. Isto demonstra que esse gene não é essencial na síntese de purinas (DRAZEK et al., 1995; FLASHNER, 2004; GOTS et al. 1977; SCHWARZ, 1992; MEYER, 1992; CHIANG, 1998).

Tabela 9. Espécies que foram capazes de crescer com o gene *purK* mutado.

Espécie	Filo ¹ ou Classe ²	Referência
<i>Brucella melitensis</i>	Alfaproteobacteria ²	Drazek et al. (1995)
<i>Escherichia coli</i>	Gamaproteobacteria ²	Gots et al. (1997)
<i>Yersinia pestis</i>	Gamaproteobacteria ²	Flashner et al. (2004)
<i>Vibrio cholerae</i>	Gamaproteobacteria ²	Chiang e Mekalanos (1998)
<i>Synechococcus sp. PCC 7942</i>	Cyanobacteria ¹	Schwarz et al. (1992)

Já que AIR é convertido de modo não enzimático diretamente em CAIR, qualquer entidade viva que possuísse uma enzima que acelerasse essa reação teria uma grande vantagem evolutiva. A primeira vista, estes dados também favorecem o primeiro cenário de evolução das PurEs, visto que a PurE II cataliza diretamente a conversão de AIR em CAIR, em contraste com a PurE I, que atua como mutase tendo como substrato o N⁵-CAIR produzido pela PurK.

Porém, dois anos depois Mueller e colaboradores (1994) mostraram que durante a conversão não enzimática o AIR é primeiro convertido em N⁵-CAIR e este em seguida em CAIR. Aparentemente este dado é congruente com o segundo cenário, onde a PurE I é a condição primitiva, mas, neste mesmo trabalho os autores

constatarem que o N⁵-CAIR é instável e se decompõem rapidamente em AIR, o que impossibilitaria seu acúmulo no ambiente. O N⁵-CAIR funciona como um carregador de CO₂. A sua função é disponibilizar o grupo carboxila, permitindo que a PurE I o transfira para a posição C4 no segundo anel purínico, e assim produza o CAIR (HOSKINS, 2007).

Assim, a maior parte dos dados expostos acima corroboram o primeiro cenário proposto para a evolução das PurEs, com a PurE II sendo a forma primitiva dessa gene e a PurE I evoluindo a partir dela. Contrário ao proposto por Zhang e colaboradores (2008), que defendem o segundo cenário, PurE II evoluindo a partir da PurE I, como o mais factível.

Contudo, os resultados da filogenia molecular contrariam ambos os cenários. As PurEs I e II formam clados distintos na árvore filogenética (Figura 17), sugerindo que evoluíram a partir de uma enzima ancestral multifuncional com atividade de carboxilase e mutase, conforme o terceiro cenário proposto para a evolução desse gene. Portanto, a evolução do sexto passo da biossíntese de purinas ajusta-se melhor ao modelo evolutivo proposto por Jensen (1976) para evolução de rotas metabólicas.

Outros resultados desse trabalho, aliados aos dados da literatura corroboram o terceiro cenário. Os resultados da genômica comparativa mostram que dezenove dos 27 Filos/Classes amostradas possuem *purE* I e *purK*. Porém, em 20 Filos/Classes foram encontradas espécies com *purE* e sem *purK*, sugerindo que nessas espécies a PurE faz a carboxilação do AIR diretamente na posição C4. Além disso, os dois tipos de PurE são encontradas espécies dos Filos/Classes de Crenarchaeota, Euryarchaeota, Actinobacteria, Chloroflexi, Acidobacteria, Thermotogae, Bacteroidetes, Chlorobi, Verrucomicrobia, Spirochaetes, Alpha, Gamma, Deltaproteobacteria, Ou seja, as PurEs I e II estão amplamente distribuídas nas linhagens procarióticas, o que posiciona o evento de duplicação que deu origem as duas isoformas no ancestral comum do seres vivos.

Outra característica a favor desse cenário é a semelhança estrutural compartilhada pelas PurEs I e II (Figura 21). A sobreposição estrutural das PurEs de espécies de Bactéria, Archaea e Eucarioto, mostra que a estrutura terciária dessas enzimas é altamente conservada, mesmo considerando que entre essas enzimas estão estruturas de PurE I e II.

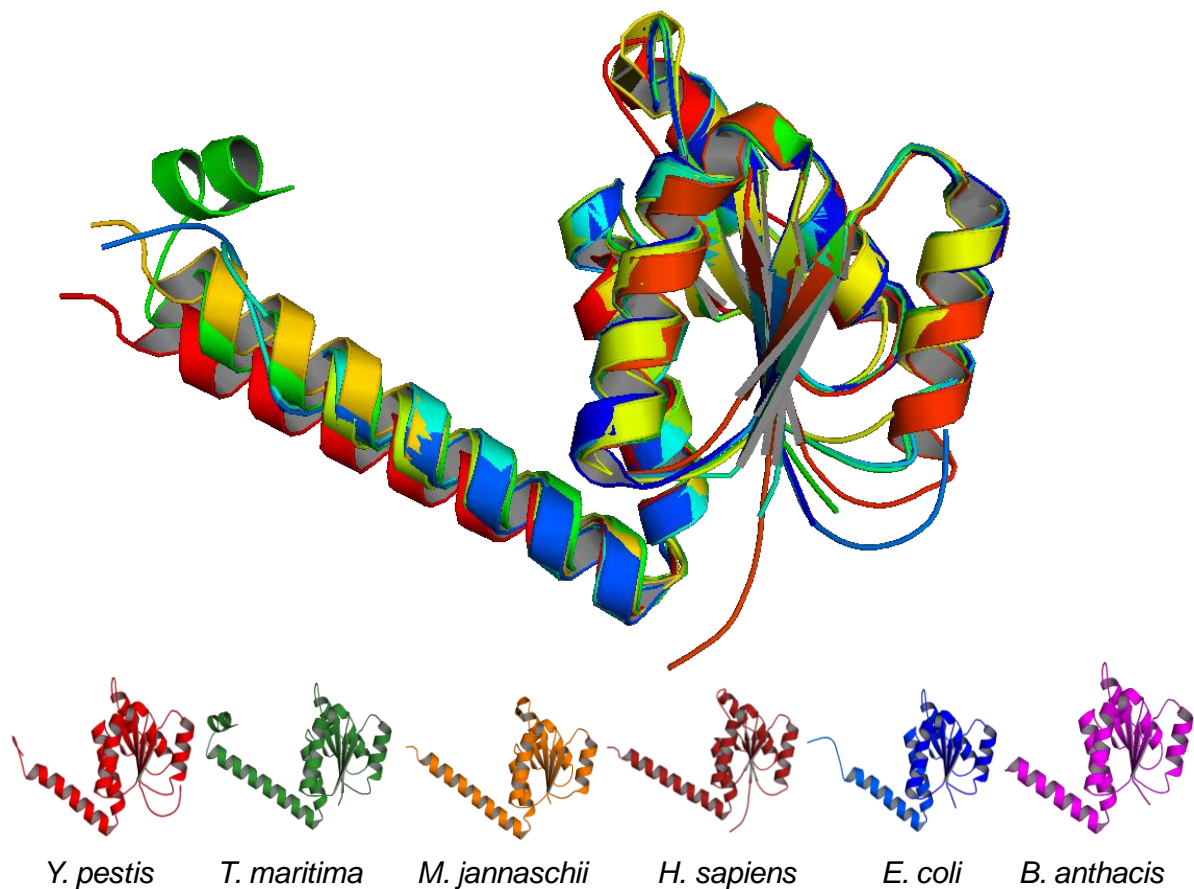


Figura 21. Sobreposição da estrutura da PurE II de *Homo sapiens* com as PurEs I das bactérias *Yersinia pestis*, *Escherichia coli*, *Bacillus anthracis*, *Thermotoga marítima*, e da Archaea *Methanocaldococcus jannaschii*, que não possui PurK. Essa sobreposição mostra que as PurEs I e II apresentam alto nível de semelhança estrutural. A sobreposição foi realizada com o programa PyMol (DeLano et al., 2002).

13.3. Contexto genômico de *purE* e *purK*

Dos 271/473 gêneros/espécies que possuem o conjunto mínimo de genes da biossíntese de purinas, são encontradas espécies que possuem *purE* e *purK* e espécies que possuem apenas *purE*. Nos 73/114 gêneros/espécies que possuem apenas *purE*, esse gene foi encontrado ou em contexto com outros genes da biossíntese de purinas (44%/47% dos gêneros/espécies) ou sem contexto genômico com outros genes da biossíntese de purinas (56%/53% dos gêneros/espécies) (Figura 22).

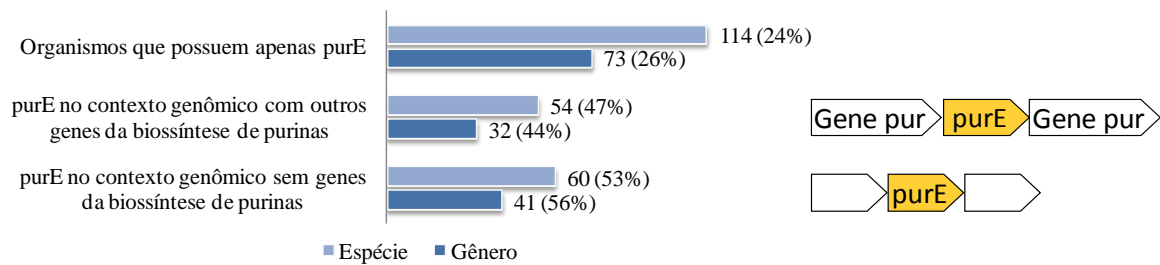


Figura 22. Contexto genômico dos gêneros/espécies que possuem apenas *purE*, mostrando que esse gene pode ser encontrado com outros genes da via ou formando um operon onde não há outros genes da biossíntese de purinas.

Os genes *purE* I e *purK* são encontrados em 75%/76% dos gêneros/espécies que possuem o conjunto mínimo dos genes da biossíntese de purinas. O contexto genômico de *purE* I e *purK* mostrou que em 75%/68% gêneros/espécies esses genes formam um operon onde não são encontrados outros genes da biossíntese de purinas. Em 14%/23% dos gêneros/espécies os genes *purE* I e *purK* formam um operon com outros genes envolvidos na síntese de purinas.

Nos gêneros/espécies onde *purE* I e *purK* não estão juntos no mesmo operon, *purE* I é encontrado em um contexto não relacionado com outros genes da síntese de purinas em 11%/9% dos gêneros/espécies, e *purK* forma um operon sem outros genes da via em 10%/9% dos gêneros/espécies. A Cyanobacteria *Trichodesmium erythraeum* IMS101, é o único caso onde o gene *purK* forma um operon com outro gene da via, este gene é *purD* (Figura 23). Nos gêneros/espécies que possuem *purE* I e *purK* não foram encontradas espécies com *purE* formando um contexto genômico com outros genes da biossíntese de purinas. Estes resultados indicam uma forte associação destes genes em nível de contexto genômico.

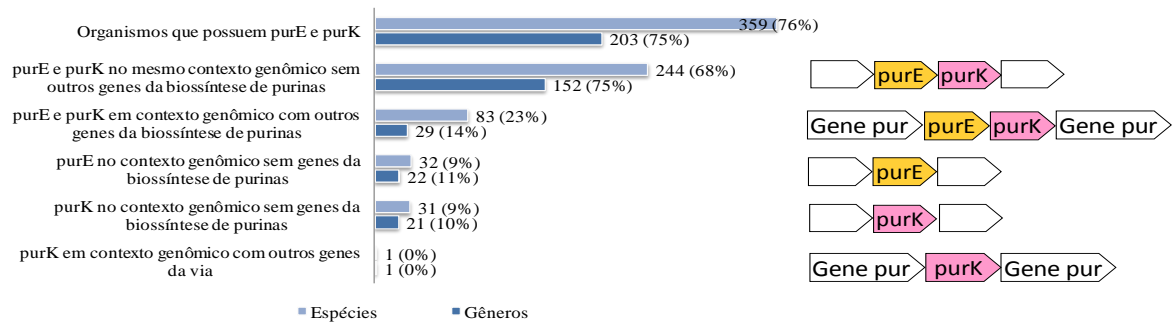


Figura 23. O contexto genômico mostra que das 203/359 gênero/espécies possuem *purE* e *purK*. Em 15/244 gêneros/espécies estes genes formam um operon sem outros genes da via, e em 29/83 gêneros/espécies esses genes estão em um contexto com outros genes da biossíntese de purinas. Em 22/32 gêneros/espécies *purE* forma um operon sem outros genes da biossíntese de purinas, e 21/31 gêneros/espécies é *purK* que está em um contexto sem outros genes da via. Em apenas uma espécie o gene *purK* forma um operon com outro gene da via que não seja *purE*.

13.4. Filogenia de PurK

Nas espécies que possuem *purE* e *purK* em contexto genômico estes genes apresentam-se no arranjo *purE* | *purK* ou *purK* | *purE* (Tabela 10). Estes arranjos estão amplamente distribuídos nas linhagens procarióticas e sua distribuição não reflete necessariamente a filogenia de procariotos. Isto sugere que estes arranjos podem ser operons egoístas que estão sendo transferidos horizontalmente nas linhagens procarióticas. Esta hipótese foi testada com a filogenia da PurK.

Na topologia as PurKs que possuem o contexto *purE* | *purK* agrupam em sua maioria em um grande clado, chamado daqui em diante de clado I. As espécies que possuem o contexto *purK* | *purE* agrupam em outro clado, chamado daqui em diante de clado II. As espécies onde *purK* não está associado com *purE* ficam distribuídas entre os dois cladros (Figura 24). Portanto, a filogenia da PurK corrobora a hipótese de que estes arranjos são operons egoístas.

Contexto genômico	Filos/Classes																										
	CR	EU	FB	FC	FT	AB	CB	AQ	CF	DTH	AC	FU	PLC	THR	BC	CL	VM	DC	EB	NT	CH	SP	AF	BT	GM	DT	EP
<i>purE</i> <i>purK</i>	5/9	4/4	14/47	3/3		3/7	1/1		1/2	1/2	1/1			2/4	4/5	1/1	1/1					1/3	38/55	21/42	50/86	2/2	
<i>purK</i> <i>purE</i>		1/3	1/1	4/5		17/34			2/3	1/1					4/6	4/9							2/5		4/4		7/11
<i>purK</i>	1/1	3/3					10/12	3/3															4/7	1/2	1/1		

Tabela 10. Tipos de contexto genômico de *purK* nos diferentes gêneros/espécies. Os filios/classes em branco não possuem *purK*.

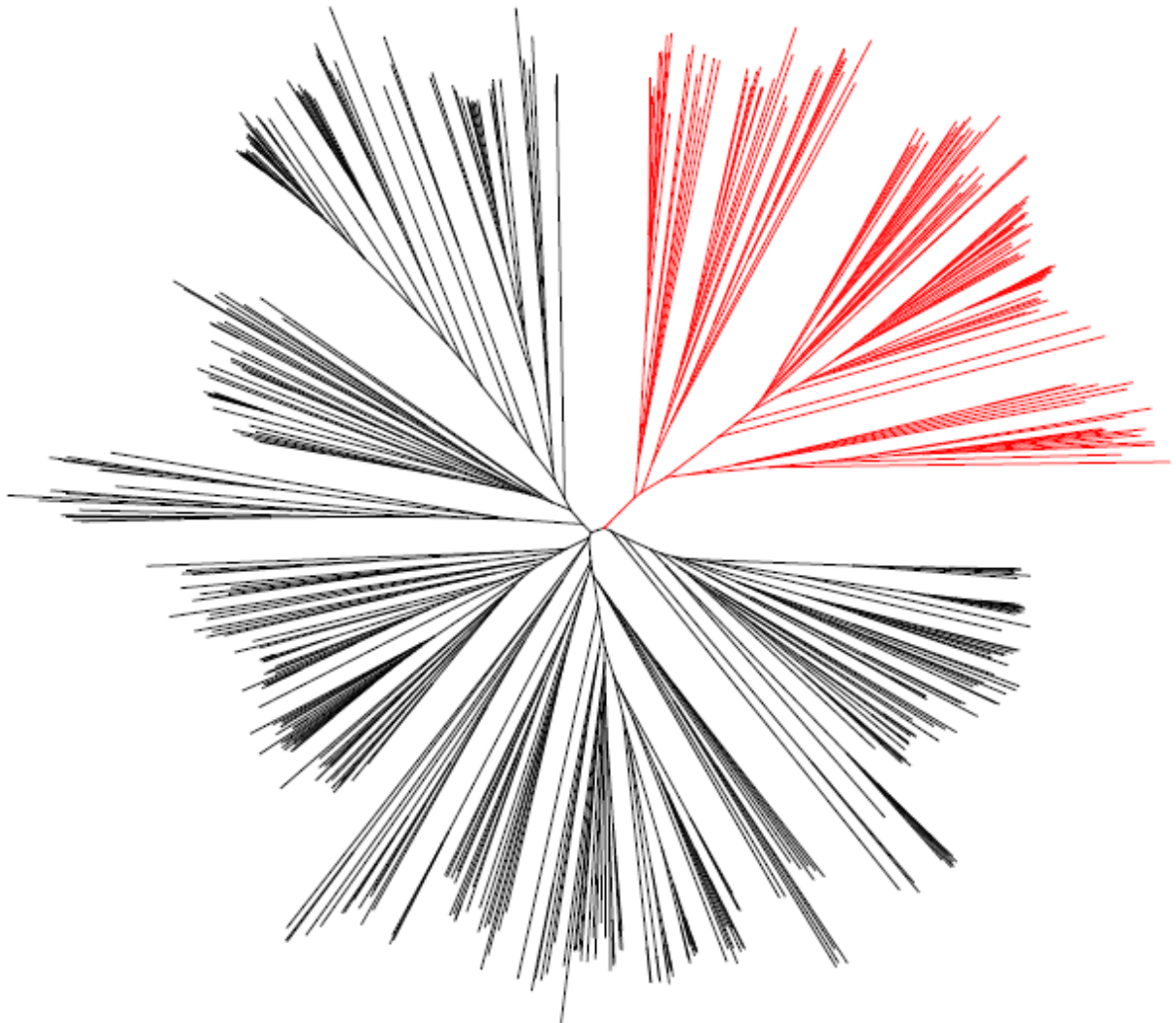


Figura 24. Árvore filogenética de PurK não enraizada. No clado I (preto) são encontradas espécies que possuem o contexto *purE I/purK*. As espécies que compõem o clado II (vermelho) possuem o contexto *purK/purE I*. A árvore foi editada no programa MEGA 5 (TAMURA et al., 2007).

Os genes *purE I* e *purK* são interdependentes, isto está bem demonstrado na genômica comparativa deste genes, onde não foram encontradas espécies apenas com PurK (Tabela 6). Portanto, os operons egoístas *purE I/purK* e *purK/purE I* podem ter sido a estratégia selecionada durante a evolução procariótica para garantir a presença simultânea de ambos os genes nos genomas. Duas hipóteses podem ser usadas para explicar a distribuição destes operons no domínio Archaea e Bacteria. Na primeira estes operons estavam presentes no ancestral comum dos seres vivos e foram seletivamente perdidos ou mantidos nas linhagens procarióticas.

Alternativamente, estes operons podem ter sido formados após a divergência das principais linhagens procariótica e posteriormente foram sendo seletivamente perdidos ou mantidos durante a evolução dos procariotos.

14. CONCLUSÕES

- PurE II esta amplamente distribuída entre os filos/classes de procariotos;
- A origem das enzimas PurE I e PurE II se deu através de uma enzima ancestral multifuncional, que se especializou em diferentes mecanismos de reação;
- Os arranjos *purE/purK* e *purK/purE* são operons egoístas que foram seletivamente mantidos ou perdidos nas diferentes linhagens procarióticas.

REFERÊNCIAS

- ALTSCHUL S. F.; MADDEN T. L.; SCHÄFFER A. A.; ZHANG J.; ZHANG Z.; MILLER W.; LIPMAN D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**. Vol. 25, No. 17, pp. 3389-3402, Sep. 1997.
- ALVES et al., R.; CHALEIL, R. A. G.; STERNBERG, M. J. E. Evolution of enzymes in metabolism: a network perspective. **J. Mol. Biol.** Vol. 320, No. 4, pp. 751–770, Jul. 2002.
- BERTHOMÉ, R.; THOMASSET, M.; MAENE, M.; BOURGEOIS, N.; FROGER, N.; BUDAR, F. *pur4* mutations are lethal to the male, but not the female, gametophyte and affect sporophyte development in arabidopsis. **Plant Physiology**, Vol. 147, pp. 650-660, Jun. 2008.
- CAETANO-ANOLLÉS, G.; KIM, S. K.; MITTENTHAL, J. E. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. **PNAS**. Vol. 104, pp. 9358-9363, Mai. 2007.
- CAETANO-ANOLLÉS, G.; SUN, F. S.; WANG, M.; YAFREMAVA, L. S.; ARISH, A.; KIM, H. S.; KNUDSEN, V.; CAETANO-ANOLLES et al., D.; MITTENTHAL, J. E. Origins and evolution of modern biochemistry: insights from genomes and molecular structure. **Frontiers in Bioscience**. Vol. 13, pp. 5212-5240, Mai. 2008.
- CHEN, L.; HARTMAN, S.; ZETTER, B.; ZHANG et al., S. IN MEMORIAM: John M. Buchanan. *Protein Science*. Vol. 16, pp. 2578–2579, 2007.
- CHEN, Z.D.; DIXON, J.E.; ZALKIN, H. Cloning of a chicken liver cDNA encoding 5-aminoimidazole ribonucleotide carboxylase and 5-aminoimidazole-4-N-succinocarboxamide ribonucleotide synthetase by functional complementation of *Escherichia coli pur* mutants. **Proc Natl Acad Sci U S A**. Vol. 87, No. 8, pp. 3097-3101, Abr. 1990.
- CHUNG, S.O.; LEE, J.H.; LEE, S.Y.; LEE, D.S. Genomic organization of *purK* and *purE* in *Brevibacterium ammoniagenes* ATCC 6872: *purE* locus provides a clue for genomic evolution. **FEMS Microbiol Lett**. Vol. 137, No. 2-3, pp. 265-268, Abr. 1996.
- CORDÓN, F. Tratado evolucionista de biología, **Aguilar**, Madrid, 1990.
- CUNCHILLOS et al., C.; LECOINTRE, G. Early steps of metabolism evolution inferred by cladistic analysis of amino acid catabolic pathways. **C. R. Biologies**. Vol. 325, pp. 119-129, Fev. 2002.
- DeLANO, W. L. The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA. 2002.
- DRAZEK, E.S.; HOUNG, H.S.; CRAWFORD, R.M.; HADFIELD, T.L.; HOOVER, D.L.; WARREN, R.L. Deletion of *purE* attenuates *Brucella melitensis* 16M for growth in

human monocyte-derived macrophages. **Infect Immun.** Vol. 63, No. 9, pp. 3297–3301, Set. 1995.

DURAN M, DORLAND L, MEULEMAN EE. Inherited defects of purine and pyrimidine metabolism: laboratory methods for diagnosis. **J. Inherit. Metab. Dis.;** Vol. 20, No. 2, pp. 227-236, Jun. 1997.

FIRESTINE, S.M.; DAVISSON, V.J. Carboxylases in de novo purine biosynthesis. Characterization of the *Gallus gallus* bifunctional enzyme. **Biochemistry.** Vol. 33, No. 39, pp. 11917-11926, Out. 1994.

FIRESTINE, S.M.; MISIALEK, S.; TOFFALETTI, D.L.; KLEM, T.J.; PERFECT, J.R.; DAVISSON, V.J. Biochemical role of the *Cryptococcus neoformans* ADE2 protein in fungal de novo purine biosynthesis. **Arch Biochem Biophys.** Vol. 351, No. 1, pp. 123-134, Mar. 1998.

FLASHNER, Y.; MAMROUD, E.; TIDHAR, A.; BER, R.; AFTALION, M.; GUR, D.; LAZAR, S.; ZVI, A.; BINO, T.; ARIEL, N.; VELAN, B.; SHAFFERMAN, A.; COHEN, S. Generation of *Yersinia pestis* attenuated strains by signature-tagged mutagenesis in search of novel vaccine candidates. **Infect Immun.** Vol. 72, No. 2, pp 908-915, Feb, 2004.

GOTS, J.S.; BENSON, C.E.; JOCHIMSEN, B.; KODURI, K.R. Microbial models and regulatory elements in the control of purine metabolism. **Ciba Foundation Symposium.** No. 48, pp. 23-41, 1977.

GRANICK, S. Speculations on the origins and evolution of photosynthesis. **Ann NY Acad Sci.** Vol. 69, No. 2, pp. 292-308, Ago. 1957.

HAMILTON, P.T.; REEVE, J.N. Sequence divergence of an archaeobacterial gene cloned from a mesophilic and a thermophilic methanogen. **J Molecular Evolution.** Vol. 22, No.4, pp. 351-360, 1985.

HAN, M.V.; ZMASEK, C.M. PhyloXML: XML for evolutionary biology and comparative genomics. **BMC Bioinformatics.** Vol. 10, No. 356, pp. 1-6, Out. 2009.

HOROWITZ, N.H. On the Evolution of Biochemical Syntheses. **Proc. Natl. Acad. Sci. USA.** Vol. 31, No. 6, pp. 153-157, Jun. 1945.

HOSKINS, A.A.; MORAR et al., M.; KAPPOCK et al., T.J.; MATHEWS, I.I.; ZAUGG, J.B.; BARDER, T.E.; PENG, P.; OKAMOTO, A.; EALICK, S.E.; STUBBE, J. N5-CAIR mutase: role of a CO₂ binding site and substrate movement in catalysis. **Biochemistry.** Vol. 46, No. 10, pp. 2842-2855, Mar. 2007.

JENSEN, R.A. Enzyme recruitment in evolution of new function. **Annu Rev Microbiol.** Vol. 30, pp. 409-425, 1976.

KAPPOCK et al., T.J.; EALICK, S.E.; STUBBE, J. Modular evolution of the purine biosynthetic pathway. *Curr. Opin. Chem. Biol.* Vol. 4, No. 5, pp. 567-572, Out. 2000.

KIRSCH D. R.; WHITNEY R. R. Pathogenicity of *Candida albicans* auxotrophic mutants in experimental infections. **Infection and Immunity**, Vol. 59, No. 9, pp. 3297–3300, Set. 1991.

KOONIN E. V. Evolution of genome architecture. **The International Journal of Biochemistry & Cell Biology**. Vol. 41, n. 2, pp. 298-306, Feb. 2009.

LABEDAN, B.; XU, Y.; NAUMOFF, D.G.; GLANSDORFF, N. Using Quaternary Structures to Assess the Evolutionary History of Proteins: The Case of the Aspartate Carbamoyltransferase. **Mol Biol Evol**. Vol.21, No.2, pp. 364-373, 2004.

LAZCANO et al., A.; MILLER, S. M. On the Origin of Metabolic Pathways. **J. Mol. Evol**. Vol. 49, pp. 424-431. Set. 1999.

LEHNINGER, A.L.; NELSON, D.L.; COX, M.M. Princípios de Bioquímica. 4ed. São Paulo: **SARVIER**, 2006.

LI et al., C.; KAPPOCK et al., T.J.; STUBBE, J.; WEAVER, T.M.; EALICK, S.E. X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution. **Structure**. Vol.7, No. 9, pp. 1155-1166, Set. 1999.

LI et al., H.; FAST, W.; BENKOVIC, S.J. Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. **Protein Science**. Vol. 18, No. 5, pp. 881-892, Mai. 2009.

LI et al., S.X.; TONG, Y.P.; XIE, X.C.; WANG, Q.H.; ZHOU, H.N.; HAN, Y.; ZHANG et al., Z.Y.; GAO, W.; LI et al., S.G.; ZHANG et al., X.C.; BI, R.C. Octameric structure of the human bifunctional enzyme PAICS in purine biosynthesis. **J Mol Biol**. Vol. 366, No. 5, pp. 1603-14, Mar. 2007.

LICHTENEGGER, H.I.M. LAMMER, H.; GRIEßMEIER, J.M.; KULIKOV, YU.N.; VON PARIS, P.; HAUSLEITNER, W.; KRAUSS, S.; RAUER, H. Aeronomical evidence for higher CO₂ levels during Earth's Hadean epoch. **Icarus**. Vo. 210, No. 1, pp. 1-7, Nov. 2010.

MAROLEWSKI, A.; SMITH, J.M.; BENKOVIC, S.J. Cloning and characterization of a new purine biosynthetic enzyme: a non-folate glycinamide ribonucleotide transformylase from *E. coli*. **Biochemistry**. Vol. 33, No. 9, pp. 2531-2537, Mar. 1994.

MAROLEWSKI, A.E.; MATTIA, K.M.; WARREN, M.S.; BENKOVIC, S.J. Formyl phosphate: a proposed intermediate in the reaction catalyzed by *Escherichia coli* PurT GAR transformylase. **Biochemistry**. Vol. 26, No. 22, pp. 6709- 6716, Jun. 1997.

MAVROMATIS, K.; IVANOVA, N.; ANDERSON, I.; LYKIDIS, A.; HOOPER, S.D.; SUN, H.; KUNIN, V.; LAPIDUS, A.; HUGENHOLTZ, P.; PATEL, B.; KYRPIDES, N.C. Genome analysis of the anaerobic thermohalophilic bacterium *Halothermothrix orenii*. **PLoS One**. Vol. 4, No. 1, Jan. 2009.

MEYER, E.; LEONARD, N.J.; BHAT, B.; STUBBE, J.; SMITH, J.M. Purification and characterization of the *purE*, *purK*, and *purC* gene products: identification of a

previously unrecognized energy requirement in the purine biosynthetic pathway. **Biochemistry**. Vol. 31, No. 21, pp. 5022-5032, Jun. 1992.

MOAT et al., A.G.; FRIEDMAN, H. The biosynthesis and interconversion of purines and their derivatives. **Bacteriol. Rev.** Vol.24, No.3, pp. 309-339, Set. 1960.

MORAR et al., M.; HOSKINS, A.A.; STUBBE, J.; EALICK, S.E. Formylglycinamide ribonucleotide amidotransferase from *thermotoga maritima*: structural insights into complex formation. **Biochemistry**. Vol. 47, No. 30, Jun. 2004.

MORETT, E.; KORBEL, J.O.; RAJAN E.; SAAB-RINCON G.; OLVERA L.; OLVERA M.; SCHMIDT S.; SNEL B.; BORK P. Systematic discovery of analogous enzymes in thiamin biosynthesis. **Nature Biotechnology**. Vol. 21, No. 7, pp. 790-795, Jun. 2003.

MORETT, E.; KORBEL, J.O.; RAJAN, E.; SAAB-RINCON, G.; OLVERA, L.; OLVERA, M.; SCHMIDT et al., S.; SNEL, B.; BORK, P. Systematic discovery of analogous enzymes in thiamin biosynthesis. **Nat Biotechnol**. Vol. 21, No. 7, pp. 790-795, Jun. 2003.

MUCHMORE, C.R.; KRAHN, J.M.; KIM, J.H.; ZALKIN, H.; SMITH, J.L. Crystal structure of glutamine phosphoribosylpyrophosphate amidotransferase from *Escherichia coli*. **Protein Sci**. Vol. 7, No. 1, pp. 39-51, Jan. 1998.

MUELLER, E.J.; MEYER, E.; RUDOLPH, J.; DAVISSON, V.J.; STUBBE, J. N5-carboxyaminoimidazole ribonucleotide: evidence for a new intermediate and two new enzymatic activities in the de novo purine biosynthetic pathway of *Escherichia coli*. **Biochemistry**. Vol. 33, No. 8, pp. 2269-2278, Mar. 1994.

NAGY et al., P.L.; MAROLEWSKI, A.; BENKOVIC, S.J.; ZALKIN, H. Formyltetrahydrofolate hydrolase, a regulatory enzyme that functions to balance pools of tetrahydrofolate and one-carbon tetrahydrofolate adducts in *Escherichia coli*. **J Bacteriol**. Vol. 177, No. 5, pp. 1292-1298, Mar. 1994.

NAGY et al., P.L.; MCCORKLE, G.M.; ZALKIN, H. *purU*, a source of formate for *purT*-dependent phosphoribosyl-N-formylglycinamide synthesis. **J Bacteriol**. Vol. 175, No. 21, pp. 7066-7073, Nov. 1993.

NIXON, A.E.; BENKOVIC, S.J. Improvement in the efficiency of formyl transfer of a GAR transformylase hybrid enzyme. **Protein Eng**. Vol. 13, No. 5, pp. 323-327, Mai. 2000.

OPARIN, A.I. The origin of life. **Macmillan**, New York, 1938.

OURISSON, G.; NAKATANI, Y. The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. **Chem Biol**. Vol. 1, No. 1, pp. 11-23, Set. 1994.

OWNBY K.; XU H.; WHITE R. H. A *Methanocaldococcus jannaschii* Archaeal Signature Gene Encodes for a 5-Formaminoimidazole-4-carboxamide-1- β -D-ribofuranosyl 5'-Monophosphate Synthetase: A NEW ENZYME IN PURINE BIOSYNTHESIS. **The Journal of Biological Chemistry**, Vol. 280, No. 12, pp. 10881-10887, Mar. 2005.

POLLACK, J.D.; MYERS, M.A.; DANDEKAR, T.; HERRMANN, R. Suspected utility of enzymes with multiple activities in the small genome *Mycoplasma* species: the replacement of the missing "household" nucleoside diphosphate kinase gene and activity by glycolytic kinases. **OMICS**. Vol. 6, No. 3, pp. 247-58, 2002.

RISON, S.C.G.; THORNTON, J.M. Pathway evolution, structurally speaking. **Curr. Opin. Struct. Biol.** Vol. 12, No. 3, pp. 374–382, Jun. 2002.

RODIONOV D. A.; VITRESCHAK A. G.; MIRONOV A. A.; GELFAND M. S. Comparative genomics of thiamin biosynthesis in procaryotes: New genes and regulatory mechanisms. **Journal of Biological Chemistry**. Vol. 277, No. 50, pp. 48949-48959, Dec. 2002.

SAITOU, N. e NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**. Vol. 4, No. 4, pp. 406-425, Jul. 1987.

SAXILD, H.H.; JACOBSEN, J.H.; NYGAARD, P. Functional analysis of the *Bacillus subtilis* purT gene encoding formate-dependent glycinamide ribonucleotide transformylase. **Microbiology**. Vol. 141, No. 9, pp. 2211-2218, Set. 1995.

SCHMIDT et al., S.; SUNYAEV, S.; BORK, P.; DANDEKAR, T. Metabolites: a helping hand for pathway evolution? **TRENDS in Biochemical Sciences**. Vol.28. No.6, pp. 336-341, Jun. 2003.

SCHWARZ, R.; LIEMAN-HURWITZ, J.; HASSIDIM, M.; KAPLAN, A. Phenotypic Complementation of High CO₂-Requiring Mutants of the Cyanobacterium *Synechococcus* sp. Strain PCC 7942 by Inosine 5'-Monophosphate. **Plant Physiology**. Vol. 100, No. 4, pp. 1987-1993, Dez. 1992.

SHIMAOKA, M.; TAKENAKA, Y.; KURAHASHI, O.; KAWASAKI, H.; MATSUI, H. Effect of amplification of desensitized purF and prs on inosine accumulation in *Escherichia coli*. **J Biosci Bioeng**. Vol.103, No. 3, pp. 255-61, Mar. 2007.

SIPPL, M.J. On distance and similarity in fold space. *Bioinformatics*. Vol. 24, No. 6, pp. 872-873, Mar. 2008.

SIPPL, M.J.; WIEDERSTEIN, M. A note on difficult structure alignment problems. *Bioinformatics*. Vol. 24, No. 3, pp. 426-427, Fev. 2008.

SMITH P. M.; ATKINS C. A. Purine Biosynthesis. Big in Cell Division, Even Bigger in Nitrogen Assimilation. **Plant Physiology**. Vol. 128, No. 3, pp. 793–802, Mar. 2002.

SØRENSEN, I.S.; DANDANELL, G. Identification and sequence analysis of *Sulfolobus solfataricus* purE and purK genes. **FEMS Microbiol Lett**. Vol. 154, No. 2, pp. 173-180, Set. 1997.

TAMURA K.; DUDLEY J.; NEI M.; KUMAR S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Molecular Biology and Evolution**. Vol. 24, No. 8, pp. 1596-1599, Aug. 2007.

THODEN et al., J.B.; FIRESTINE, S.; NIXON, A.; BENKOVIC, S.J.; HOLDEN, H.M. Molecular structure of Escherichia coli PurT-encoded glycinamide ribonucleotide transformylase. **Biochemistry**. Vol. 39, No. 30, pp. 8791-8802, Ago. 2000.

THODEN et al., J.B.; FIRESTINE, S.M.; BENKOVIC, S.J.; HOLDEN, H.M. PurT-encoded Glycinamide ribonucleotide transformylase. **The Journal of Biological Chemistry**. Vol. 277, No. 26, pp. 23898-23908, Jun. 2002.

THODEN et al., J.B.; HOLDEN, H.M.; FIRESTINE, S.M. Structural Analysis of the Active Site Geometry of N5-Carboxyaminoimidazole Ribonucleotide Synthetase from Escherichia coli. **Biochemistry**. Vol. 47, No. 50, pp. 13346-13353, Dec. 2008.

THODEN et al., J.B.; KAPPOCK et al., T.J.; STUBBE, J.; HOLDEN, H.M. Three-dimensional structure of N5-carboxyaminoimidazole ribonucleotide synthetase: a member of the ATP grasp protein superfamily. **Biochemistry**. Vol. 38, No. 47, pp. 15480-15492, Nov. 1999.

THORNDIKE, J.; GAUMONT, Y.; KISLIUK, R.L.; SIROTNAK, F.M.; MURTHY, B.R.; NAIR, M.G.; PIPER, J.R. Inhibition of glycinamide ribonucleotide formyltransferase and other folate enzymes by homofolate polyglutamates in human lymphoma and murine leukemia cell extracts. **Cancer Research**. Vol. 49, No. 1, pp. 158-163, Jan. 1989.

TRANCHIMAND, S.; STARKS, C.M.; MATHEWS, I.I.; HOCKINGS, S.C.; KAPPOCK et al., T.J. Treponema denticola PurE Is a Bacterial AIR Carboxylase. **Biochemistry**. Vol. 50, pp. 4623-4637, Mai. 2011.

TSAI et al., M.; KOO, J.; YIP, P.; COLMAN, R.F.; SEGALL, M.L.; HOWELL, P.L. Substrate and product complexes of Escherichia coli adenylosuccinate lyase provide new insights into the enzymatic mechanism. **J Mol Biol**. Vol. 370, No. 3, pp. 541-554, Jul. 2007.

VAN PASSEL, M.W.; KANT, R.; PALVA, A.; COPELAND, A.; LUCAS, S.; LAPIDUS, A.; GLAVINA DEL RIO, T.; PITLUCK, S.; GOLTSMAN, E.; CLUM, A.; SUN, H.; SCHMUTZ, J.; LARIMER, F.W.; LAND, M.L.; HAUSER, L.; KYRPIDES, N.; MIKHAILOVA, N.; RICHARDSON, P.P.; JANSSEN, P.H.; DE VOS, W.M.; SMIDT, H. Genome sequence of the verrucomicrobium Opiritutus terrae PB90-1, an abundant inhabitant of rice paddy soil ecosystems. **Journal of Bacteriology**. Vol. 193, No. 9, pp. 2367-2368, Mai. 2011.

WARREN, M.S.; KMATTIA, K.M.; MAROLEWSKI, A.E.; BENKOVIC, S.J. The transformylase enzymes of de novo purine biosynthesis. **Pure & Appl. Chem**. Vol. 68, No. 11, pp. 2029-2036, 1996.

WARREN, M.S.; MAROLEWSKI, A.E.; BENKOVIC, S.J. A rapid screen of active site mutants in glycinamide ribonucleotide transformylase. **Biochemistry**. Vol. 35, No. 27, pp. 8855-8862, Jul. 1996.

WELIN et al., M.; GROSSMANN, J.G.; FLODIN, S. NYMAN, T. STENMARK, P.; TRÉSAUGUES, L.; KOTENYOVA, T.; JOHANSSON, I. NORDLUND, P. LEHTIÖ, L.

Structural studies of tri-functional human GART. **Nucleic Acids Res.** Vol. 38, No. 20, pp. 7308-7319, Nov. 2010.

ZHANG Y.; MORAR M.; EALICK S. E. Structural biology of the purine biosynthetic pathway. **Cellular and Molecular Life Sciences.** Vol. 65, No. 23, pp. 3699-3724, Nov. 2008.

ZHANG et al., R.; SKARINA, T.; EVDOKIMOVA, E.; EDWARDS, A.; SAVCHENKO, A.; LASKOWSKI, R.; CUFF, M.E.; JOACHIMIAK, A. Structure of SAICAR synthase from *Thermotoga maritima* at 2.2 angstroms reveals an unusual covalent dimer. **Acta Crystallogr Sect F Struct Biol Cryst Commun.** Vol. 62, No. 4, pp. 335-339, Apr. 2006.

ANEXOS

Anexo 1. Sequências de PurT que foram removidas por serem divergentes. As espécies que têm mais de uma cópia, apenas a cópia divergente foi excluída.

Espécie	Qtd	Removida	Mantida
<i>Mycobacterium abscessus</i>	2	169631323	169628284
<i>Mycobacterium gilvum</i> PYR-GCK	2	145220827	145222627
<i>Mycobacterium smegmatis</i> str. MC2 155	2	118472682	118473801
<i>Actinobacillus pleuropneumoniae</i> serovar 3 str. JL03	2	165976529	165976532
<i>Mycobacterium avium</i> 104	1	118464706	
<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i> K-10	1	41409969	
<i>Mycobacterium bovis</i> AF2122/97	1	31791565	
<i>Mycobacterium tuberculosis</i> H37Rv	1	15607530	
<i>Mycobacterium marinum</i> M	1	183980709	
<i>Mycobacterium kansasii</i> ATCC 12478	1	240168850	

Anexo 2. Espécies que possuem cópias redundante de PurT e que foram removidas das análises filogenéticas.

Nº de indentificação	Espécie
33591890	<i>Bordetella pertussis</i> Tohama I
33603474	<i>Bordetella bronchiseptica</i> RB50
189501462	<i>Chlorobium phaeobacteroides</i> BS1
60681705	<i>Bacteroides fragilis</i> NCTC 9343
107023257	<i>Burkholderia cenocepacia</i> AU 1054
206560774	<i>Burkholderia cenocepacia</i> J2315
115352409	<i>Burkholderia ambifaria</i> AMMD
53718746	<i>Burkholderia pseudomallei</i> K96243
91782528	<i>Burkholderia xenovorans</i> LB400
187479583	<i>Bordetella avium</i> 197N
121606459	<i>Polaromonas naphthalenivorans</i> CJ2
152989880	<i>Nitratiruptor</i> sp. SB155-2
152991751	<i>Sulfurovum</i> sp. NBC37-1
17228794	<i>Nostoc</i> sp. PCC 7120
16330395	<i>Synechocystis</i> sp. PCC 6803
254430849	<i>Cyanobium</i> sp. PCC 7001
254375109	<i>Francisella novicida</i> GA99-3548
78224384	<i>Geobacter metallireducens</i> GS-15
190573284	<i>Stenotrophomonas maltophilia</i> K279a
78046825	<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> str. 85-10
165976532	<i>Actinobacillus pleuropneumoniae</i> serovar 3 str. JL03
146312064	<i>Enterobacter</i> sp. 638
114048592	<i>Shewanella</i> sp. MR-7
76811548	<i>Burkholderia pseudomallei</i> 1710b
74311828	<i>Shigella sonnei</i> Ss046
153950504	<i>Yersinia pseudotuberculosis</i> IP 31758

Anexo 3. Espécies que possuem mais de uma cópia de PurE. As cópias divergentes foram removidas das análises filogenéticas.

Espécies	Qtd	Removida	Mantida
<i>Geobacillus kaustophilus</i> HTA426	2	56420499	56418792
<i>Geobacillus thermodenitrificans</i> NG80-2	2	138895515	138893913
<i>Staphylococcus saprophyticus</i> subsp. <i>saprophyticus</i> 15305	2	73661484	73663035
<i>Pediococcus pentosaceus</i> ATCC 25745	2	116493421	116493205
<i>Alkaliphilus metalliredigens</i> QYMF	2	150388317	150388747
<i>Caldicellulosiruptor saccharolyticus</i> DSM 8903	2	146296723	146296640
<i>Clostridium beijerinckii</i> NCIMB 8052	2	150015450	150015940
<i>Clostridium botulinum</i> A str. ATCC 19397	2	153931808	153932727
<i>Desulfitobacterium hafniense</i> DCB-2	2	219668936	219667525
<i>Desulfotomaculum reducens</i> MI-1	2	134298027	134300210
<i>Heliobacterium modesticaldum</i> Ice1	2	167629721	167631127
<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223	3	167037942	167038112 167038197
<i>Thermoanaerobacter tengcongensis</i> MB4	2	20807081	20807254
<i>Rubrobacter xylanophilus</i> DSM 9941	2	108805837	108803105
<i>Anabaena variabilis</i> ATCC 29413	2	75908030	75909163
<i>Cyanothece</i> sp. PCC 7424	2	218440730	220909655
<i>Cyanothece</i> sp. PCC 8801	2	218247440	218247928
<i>Microcystis aeruginosa</i> NIES-843	2	166369007	166362749
<i>Nostoc punctiforme</i> PCC 73102	2	186685874	186685339
<i>Prochlorococcus marinus</i> str. AS9601	2	123969009	123967722
<i>Prochlorococcus marinus</i> str. NATL2A	2	72382684	124024954
<i>Synechococcus elongatus</i> PCC 6301	2	56750391	56751086
<i>Synechococcus</i> sp. CC9311	2	113953897	113954304
<i>Synechocystis</i> sp. PCC 6803	2	16332169	16332049
<i>Thermosynechococcus elongatus</i> BP-1	2	22298732	161485770
<i>Trichodesmium erythraeum</i> IMS101	2	113478096	113477875
<i>Akkermansia muciniphila</i> ATCC BAA-835	2	187736083	187734628
<i>Dictyoglomus thermophilum</i> H-6-12	2	206901330	206901555 206901546
<i>Dictyoglomus turgidum</i> DSM 6724	2	217967927	217966703
<i>Desulfatibacillum alkenivorans</i> AK-01	2	218777972	218779902
<i>Pelobacter carbinolicus</i> DSM 2380	2	77919824	77918638
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> 81-176	2	121612475	57237005
<i>Gluconacetobacter diazotrophicus</i> PAI 5	2	162147509	209545616
<i>Aeromonas hydrophila</i> subsp. <i>hydrophila</i> ATCC 7966	2		117621214 117619504
<i>Psychromonas ingrahamii</i> 37	2		119943866 119945722
<i>Shewanella loihica</i> PV-4	2		127512162 127512180

Anexo 3. Espécies que possuem mais de uma cópia de PurE. As cópias divergentes foram removidas das análises filogenéticas.

Espécie	Qtd	Removida	Mantida
<i>Haloarcula marismortui</i> ATCC 43049	2	55379220	55378228
<i>Haloquadratum walsbyi</i> DSM 16790	2	110667927	110667595
<i>Methanococcus aeolicus</i> Nankai-3	2	150401863	150400541
<i>Methanococcus vanniellii</i> SB	3	150399071 150399031	150400035
<i>Methanoculleus marisnigri</i> JR1	2	126179529	126178007
<i>Methanosarcina mazei</i> Go1	3	21227718	21228251 21228462
<i>Methanospirillum hungatei</i> JF-1	2	88604149	88603715

Anexo 4. Espécies que possuem cópias redundantes de PurE e que foram removidas das análises filogenéticas.

Nº de identificação	Espécie
152991952	<i>Sulfurovum</i> sp. NBC37-1
152991256	<i>Nitratiruptor</i> sp. SB155-2
89055843	<i>Jannaschia</i> sp. CCS1
99080445	<i>Ruegeria</i> sp. TM1040
49476091	<i>Bartonella henselae</i> str. Houston-1
110635739	<i>Chelativorans</i> sp. BNC1
170740910	<i>Methylobacterium</i> sp. 4-46
39937234	<i>Rhodopseudomonas palustris</i> CGA009
86748572	<i>Rhodopseudomonas palustris</i> HaA2
91975903	<i>Rhodopseudomonas palustris</i> BisB5
148258056	<i>Bradyrhizobium</i> sp. BTAi1
146338302	<i>Bradyrhizobium</i> sp. ORS278
85716100	<i>Nitrobacter</i> sp. Nb-311A
167644870	<i>Caulobacter</i> sp. K31
82543009	<i>Shigella boydii</i> Sb227
30062007	<i>Shigella flexneri</i> 2a str. 2457T
56479676	<i>Shigella flexneri</i> 2a str. 301
146310638	<i>Enterobacter</i> sp. 638
68249793	<i>Haemophilus influenzae</i> 86-028NP
16273505	<i>Haemophilus influenzae</i> Rd KW20
190149968	<i>Actinobacillus pleuropneumoniae</i> serovar 7 str. AP76
113460817	<i>Haemophilus somnus</i> 129PT
152995107	<i>Marinomonas</i> sp. MWYL1
113971310	<i>Shewanella</i> sp. MR-4
120599919	<i>Shewanella</i> sp. W3-18-1
126173280	<i>Shewanella baltica</i> OS155
148654001	<i>Psychrobacter</i> sp. PRwf-1
197334194	<i>Vibrio fischeri</i> MJ11
50086600	<i>Acinetobacter</i> sp. ADP1
213968528	<i>Pseudomonas syringae</i> pv. tomato T1
66048272	<i>Pseudomonas syringae</i> pv. syringae B728a
170719325	<i>Pseudomonas putida</i> W619
167036371	<i>Pseudomonas putida</i> GB-1
110670468	<i>Francisella tularensis</i> subsp. tularensis FSC198
54296250	<i>Legionella pneumophila</i> str. Paris
148358428	<i>Legionella pneumophila</i> str. Corby
119897669	<i>Azoarcus</i> sp. BH72
188591313	<i>Cupriavidus taiwanensis</i> LMG 19424
24375058	<i>Shewanella oneidensis</i> MR-1
54293224	<i>Legionella pneumophila</i> str. Lens

Anexo 5. Espécies que possuem mais de uma cópia de PurK. As cópias divergentes foram removidas das análises filogenéticas.

Espécie	Qtd	Removidas	Mantidas
<i>Enterococcus faecalis</i> V583	2	29376860	29376333
<i>Lactobacillus casei</i> ATCC 334	2	116494618	116495235
<i>Lactobacillus casei</i> BL23	2	191638055	191638744
<i>Lactobacillus fermentum</i> IFO 3956	2	90962322	184154595
<i>Lactobacillus plantarum</i> WCFS1	2	28377921	28379222
<i>Lactobacillus reuteri</i> JCM 1112	2	184154012	184152784
<i>Lactobacillus rhamnosus</i> HN001	2	199590793	199598015
<i>Lactobacillus sakei</i> subsp. <i>sakei</i> 23K	2	81429169	81428268
<i>Lactobacillus salivarius</i> UCC118	2	909622322	90961492
<i>Pediococcus pentosaceus</i> ATCC 25745	2	116429169	116493204

Anexo 6. Espécies que possuem cópias redundantes de PurK e que foram removidas das análises filogenéticas.

Nº de indentificação	Espécie
148655955	<i>Roseiflexus</i> sp. RS-1
55980482	<i>Thermus thermophilus</i> HB8
148654002	<i>Psychrobacter</i> sp. PRwf-1
146310637	<i>Enterobacter</i> sp. 638
170717541	<i>Haemophilus somnus</i> 2336
117921594	<i>Shewanella</i> sp. ANA-3
57239521	<i>Ehrlichia ruminantium</i> str. Welgevonden
15789831	<i>Halobacterium</i> sp. NRC-1
170288289	<i>Thermotoga</i> sp. RQ2
148269613	<i>Thermotoga petrophila</i> RKU-1
116669750	<i>Arthrobacter</i> sp. FB24
145294884	<i>Corynebacterium glutamicum</i> R
19551941	<i>Corynebacterium glutamicum</i> ATCC 13032
254776680	<i>Mycobacterium avium</i> subsp. <i>avium</i> ATCC 25291
15610412	<i>Mycobacterium tuberculosis</i> H37Rv
118618095	<i>Mycobacterium ulcerans</i> Agy99
220910512	<i>Cyanothece</i> sp. PCC 7425
148239542	<i>Synechococcus</i> sp. WH 7803
33595872	<i>Bordetella parapertussis</i> 12822
53718440	<i>Burkholderia pseudomallei</i> K96243
91790455	<i>Polaromonas</i> sp. JS666
30248871	<i>Nitrosomonas europaea</i> ATCC 19718
114321385	<i>Alkalilimnicola ehrlichii</i> MLHE-1
170740911	<i>Methylobacterium</i> sp. 4-46
148258057	<i>Bradyrhizobium</i> sp. BTAi1
90425436	<i>Rhodopseudomonas palustris</i> BisB18
167644869	<i>Caulobacter</i> sp. K31
49476092	<i>Bartonella henselae</i> str. Houston-1
82775635	<i>Shigella dysenteriae</i> Sd197
74311087	<i>Shigella sonnei</i> Ss046

Anexo 6. Espécies que possuem cópias redundantes de PurK e que foram removidas das análises filogenéticas.

Nº de indentificação	Espécie
213692839	<i>Bifidobacterium longum subsp. infantis</i> ATCC 15697
118476040	<i>Bacillus thuringiensis str. Al Hakam</i>
218530198	<i>Methylobacterium chloromethanicum</i> CM4
148559862	<i>Brucella ovis</i> ATCC 25840
16763913	<i>Salmonella enterica subsp. enterica serovar Typhimurium str. LT2</i>
156501790	<i>Francisella tularensis subsp. holarctica</i> FTNF002-00
41409492	<i>Mycobacterium avium subsp. paratuberculosis</i> K-10
116617831	<i>Leuconostoc mesenteroides subsp. mesenteroides</i> 8293
49479102	<i>Bacillus thuringiensis serovar konkukian str. 97-27</i>
56417093	<i>Anaplasma marginale str. St. Maries</i>
33592618	<i>Bordetella pertussis</i> Tohama I
78067423	<i>Burkholderia sp. 383</i>
23502602	<i>Brucella suis</i> 1330
195953397	<i>Hydrogenobaculum sp. Y04AAS1</i>
163782061	<i>Hydrogenivirga sp. 128-5-R1-1</i>
21243407	<i>Xanthomonas axonopodis pv. citri str. 306</i>
104784375	<i>Pseudomonas entomophila</i> L48
66048271	<i>Pseudomonas syringae pv. syringae</i> B728a
50086599	<i>Acinetobacter sp. ADP1</i>
152995106	<i>Marinomonas sp. MWYL1</i>
152999636	<i>Shewanella baltica</i> OS185
119717712	<i>Nocardioides sp. JS614</i>
227489819	<i>Corynebacterium pseudogenitalium</i> ATCC 33035
108798269	<i>Mycobacterium sp. MCS</i>
170077231	<i>Synechococcus sp. PCC 7002</i>
83720003	<i>Burkholderia thailandensis</i> E264
121596221	<i>Acidovorax sp. JS42</i>
119897670	<i>Azoarcus sp. BH72</i>
89055844	<i>Jannaschia sp. CCS1</i>
99080444	<i>Ruegeria sp. TM1040</i>
110635740	<i>Chelativorans sp. BNC1</i>
190893356	<i>Rhizobium etli</i> CIAT 652
227823358	<i>Sinorhizobium fredii</i> NGR234

Anexo 7. Contexto genômico resumido. A quantidade é representada da seguinte maneira: nº de espécies que possui o contexto / total de espécies.

Espécie	Contexto Genômico	Quant
Firmicutes Bacilli		41/49
		2/49
Firmicutes Clostridia		3/28
		20/28
		5/28
Actinobacteria		41/43
		14/43
		4/43
		38/43
Cyanobacteria		11/13
Aquificae		3/3
Chloroflexi		5/7
		2/7
Deinococcus-Thermus		2/3
		1/3
Acidobacteria		1/2
		1/2
Fusobacteria		1/1
Planctomycetes		1/1
Thermotogae		4/8
		4/8
Bacteroidetes		3/12
		3/12
Chlorobi		4/11
Verrucomicrobia		2/3
		1/3
Dictyoglomi		2/2
Elusimicrobia		1/2
		1/2
Nitrospirae		1/1
Spirochaetes		3/4
		1/4
Deltaproteobacteria		10/20
		2/20
Epsilonproteobacteria		5/11
		6/11
Alphaproteobacteria		56/69
		12/69
Betaproteobacteria		31/44
Gammaproteobacteria		55/98
Crenarchaeota		5/12
		5/12
Euriarchaeota		9/30
		8/30