

UFRB

Universidade Federal do
Recôncavo da Bahia

**UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS, AMBIENTAIS E BIOLÓGICAS
EMBRAPA MANDIOCA E FRUTICULTURA TROPICAL
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROBIOLOGIA AGRÍCOLA
CURSO DE MESTRADO**

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

FENÍCIA BRITO SANTOS

Cruz das Almas
2014

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

FENÍCIA BRITO SANTOS

Bacharel em Biologia na Universidade Federal do Recôncavo da Bahia, 2011

Dissertação submetida ao Colegiado do Programa de Pós-Graduação em Microbiologia Agrícola da Universidade Federal do Recôncavo da Bahia e Embrapa Mandioca e Fruticultura Tropical, como requisito parcial para obtenção do Grau de Mestre em Microbiologia Agrícola.

Orientador: Phellippe Arthur S. Marbach

Co-Orientador: Jorge Teodoro de Souza

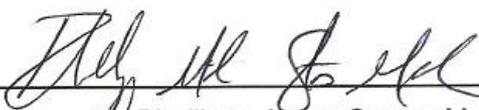
Cruz das Almas
2014

FICHA CATALOGRÁFICA

S231	<p>Santos, Fenícia Brito</p> <p>Diversidade estrutural e evolução da terceira e sexta etapas da biossíntese de purinas em procariontes / Fenícia Brito Santos. - Cruz das Almas, BA, 2014.</p> <p>111f.; il.</p> <p>Orientador: Prof .Phillippe Arthur S. Marbach.</p> <p>Dissertação (Mestrado) – Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias, Ambientais e Biológicas - CCAAB.</p> <p>1. Genética vegetal. 2. Filogenia.. 3. Genoma. 4. Biologia computacional. I.. Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias, Ambientais e Biológicas - CCAAB. II. Título.</p> <p>CDD: 581.35</p>
------	--

UNIVERSIDADE FEDERAL DO RECÔNCAO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS, AMBIENTAIS E BIOLÓGICAS
EMBRAPA MANDIOCA E FRUTICULTURA TROPICAL
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROBIOLOGIA AGRÍCOLA
CURSO DE MESTRADO

COMISSÃO EXAMINADORA DA DEFESA DE DISSERTAÇÃO DE
FENÍCIA BRITO SANTOS



Dr. Phellippe Arthur Santos Marbach

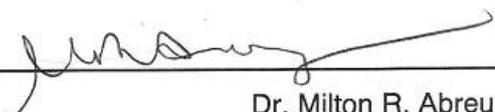
Universidade Federal do Recôncavo da Bahia - UFRB

(Orientador)



Dr. Artur Trancoso Lopo de Queiroz

Centro de Pesquisa Gonçalo Muniz, CPQGM, Brasil.



Dr. Milton R. Abreu Roque

Universidade Federal da Bahia - UFBA

CRUZ DAS ALMAS – BAHIA

JANEIRO – 2014

Aos meus pais, Oderiana e Agnaldo,
À minha irmã Dalma,
Dedico.

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 e melhor amiga, 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 Valter Magalhães e Daniele Diniz pela amizade, carinho e o apoio de sempre.

Às minhas amigas Gisele, Índira e Kali que mesmo distantes torceram por mim e me ajudaram da maneira que podiam. E a todos os amigos que durante o mestrado fizeram a minha caminhada mais fácil.

Ao meu orientador Phellippe Arthur Santos Marbach pela dedicação extraordinária e pela amizade incondicional.

À Karol Brito pela amizade e generosidade e por me ajudar mesmo quando eu não achava que precisava de ajuda.

À Adailson Feitoza pela dedicação e paciência. Você ajudou a dissipar a nuvem negra.

À Zayda Morales pela amizade e apoio.

Aos técnicos dos laboratórios do Bloco L pela ajuda. Em especial a “Lene” e Carol Yamamoto que sempre tinham a solução quando tudo parecia perdido.

Ao Profº Jorge Teodoro de Souza, Marcia Cazeta e Phelleppe Marbach por me ajudarem a continuar a trilhar o meu caminho.

À todos os mestres pelo conhecimento que adquiri.

À Universidade Federal do Recôncavo da Bahia pela formação.

De tudo ficaram três coisas: A certeza de que estamos começando, a certeza de que é preciso continuar e a certeza de que podemos ser interrompidos antes de terminar. Façamos da interrupção um caminho novo. Da queda, um passo de dança. Do medo, uma escada...

Fernando Sabino

Índice

CAPÍTULO I - Revisão Bibliográfica, Justificativa e Objetivos.....	9
INTRODUÇÃO.....	11
JUSTIFICATIVA.....	12
REVISÃO BIBLIOGRÁFICA.....	13
<i>Via Biossintética de Purinas.....</i>	<i>13</i>
<i>Origem das vias biossintéticas.....</i>	<i>20</i>
OBJETIVOS.....	25
REFERÊNCIAS.....	26
CAPÍTULO II - DIVERSIDADE ESTRUTURAL E EVOLUÇÃO DO TERCEIRO PASSO DA BIOSSÍNTESE DE PURINAS EM LINHAGENS PROCARIÓTICAS.....	30
RESUMO.....	33
ABSTRACT.....	34
INTRODUÇÃO.....	35
RESULTADOS.....	37
DISCUSSÃO.....	54
CONCLUSÃO.....	63
METODOLOGIA.....	63
REFERÊNCIAS.....	68
CAPÍTULO III - DIVERSIDADE ESTRUTURAL E EVOLUÇÃO DO SEXTO PASSO DA BIOSSÍNTESE DE PURINAS EM LINHAGENS PROCARIÓTICAS.....	70
RESUMO.....	72
ABSTRACT.....	73
INTRODUÇÃO.....	74
RESULTADOS.....	77
DISCUSSÃO.....	97
CONCLUSÃO.....	102
METODOLOGIA.....	103
REFERÊNCIAS.....	105
ANEXO I.....	108

Capítulo 1

Revisão Bibliográfica, Justificativa e Objetivos da Dissertação
intitulada:

Diversidade e Evolução da terceira e sexta etapas da Via
Biossintética de Purinas em Procariotos

Esse capítulo está formatado segundo as normas da Associação Brasileira de Normas
Técnicas (ABNT).

Numeração das figuras e tabelas segue a ordem dos capítulos.

Lista de Figuras

- Figura 1.1. Elementos doadores dos átomos que formam os anéis da purina. Fonte: Modificado de ZHANG; MORAR; EALICK, 2008. _____ 14
- Figura 1.2. Via biossintética de purinas. Fonte: ZHANG; MORAR; EALICK, 2008 _____ 15
- Figura 1.3. Modelos de evolução de vias metabólicas. (A) Retro-evolução; (B) Evolução direta. Fonte: Modificado de SCHMIDT et al., 2003. _____ 21
- Figura 1.4. Modelos de evolução de vias metabólicas. Fonte: Modificado de SCHMIDT et al., 2003. **A.** Especialização de enzima multifuncional. **B.** Especialização de enzima multifuncional. **C.** Recrutamento enzimático _____ 22
- Figura 1.5. Filogenia das vias metabólicas evidenciando que as vias biossintéticas de nucleotídeos foram, possivelmente, as primeiras rotas a surgirem. Fonte: CAETANO-ANOLLÉS; KIM; MITTENTHAL, 2007. _____ 24

INTRODUÇÃO

A Via Biossintética de Purinas (VBP) 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; ATKINS, 2002).

Como os genes da síntese *de novo* de purinas são essenciais no processo de duplicação do DNA, as enzimas da VBP também são importantes alvos para fármacos anticâncer ou antimicrobianos (KIRSCH; 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; MORAR; EALICK, 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; MORAR; EALICK, 2008).

As reações químicas da terceira e sexta etapas da VBP 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 (LI; FAST; BENKOVIC, 2009; NAGY; MCCORKLE; ZALKIN, 1993; SAXILD; JACOBSEN; NYGAARD, 1995; ZHANG; MORAR; EALICK, 2008). 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. Em organismos superiores a reação é

catalisada pela enzima PurE classe II. 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 (CHUNG et al., 1996; DANDANELL; SORENSEN, 1997; DRAZEK et al., 1995; FIRESTINE; DAVISSON, 1994; FIRESTINE et al., 1998; HOSKINS et al., 2007; MEYER et al., 1992; MUELLER et al., 1994).

A evolução de vias biossintéticas está 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.

JUSTIFICATIVA

O conhecimento atual sobre a diversidade estrutural da VBP 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 VBP, 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 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 anticâncer ou antimicrobianos (KIRSCH; 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.

REVISÃO BIBLIOGRÁFICA

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, CO₂ e NH₃. As vias de recuperação reutilizam as bases livres e os nucleotídeos liberados na quebra dos ácidos nucléicos, com menor custo energético (Figura 1.1). A Via de Degradação de nucleotídeos desempenha um papel central no metabolismo do nitrogênio (BERTHOMÉ et al., 2008).

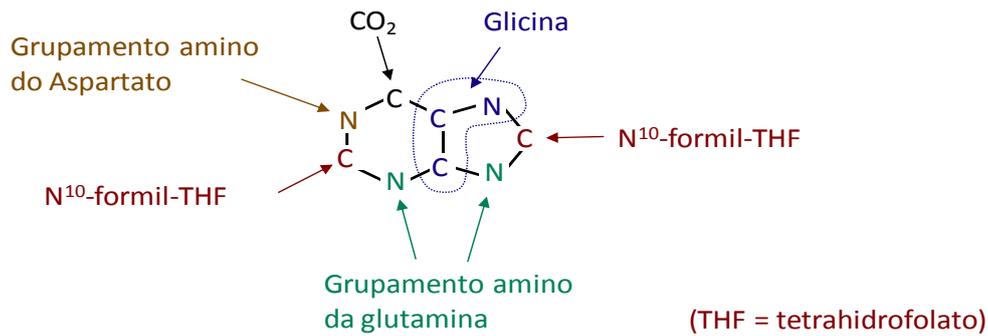


Figura 1.1. Elementos doadores dos átomos que formam os anéis da purina. Fonte: Modificado de ZHANG; MORAR; EALICK, 2008.

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*, foi possível determinar quais precursores deram origem aos átomos de carbono e nitrogênio do anel purínico. O estudo detalhado da biossintese de purinas possibilitou o desenvolvimento de um grande número de quimioterápicos antineoplásicos que têm sido amplamente utilizados nas últimas décadas (ZHANG; MORAR; EALICK, 2008).

A biossintese “de novo” de purinas ocorre através de 10-11 reações enzimáticas (Figura 1.2). 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; FRIEDMAN, 1960; ZHANG; MORAR; EALICK, 2008).

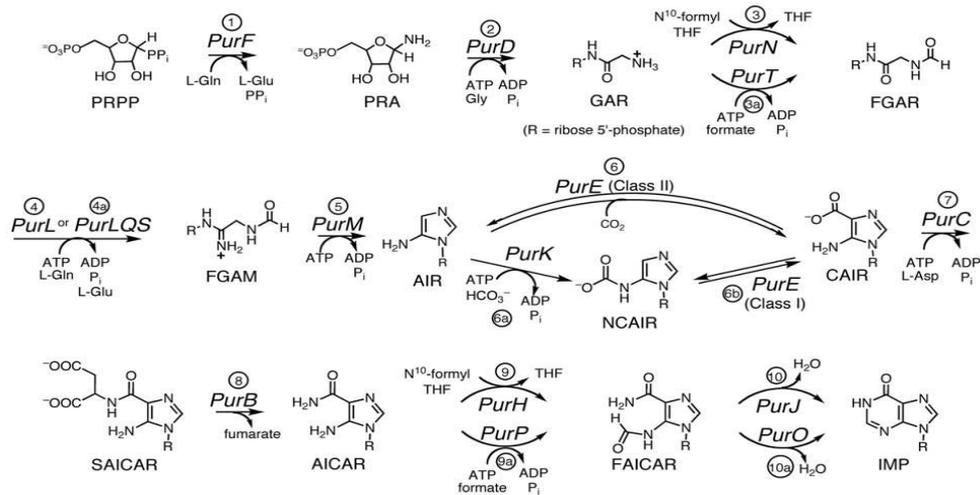


Figura 1.2. Via biossintética de purinas. Fonte: ZHANG; MORAR; EALICK, 2008

No primeiro passo da VBP, 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 NH₃ 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; MORAR; EALICK, 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; MORAR; EALICK, 2008).

O intermediário PRA resultante da primeira reação é altamente instável, com meia-vida de aproximadamente 30 segundos em pH 7,5 (ZHANG; MORAR; EALICK, 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; MORAR; EALICK, 2008).

Na terceira etapa da VBP, 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; MAROLEWSKI; BENKOVIC, 1996). Uma das enzimas, a Fosforibosilglicinamida formiltransferase (PurN), adiciona um grupo formil ao grupamento amino do GAR, utilizando como cofator o N¹⁰-formiltetrahidrofolato (N¹⁰-formilTHF) (NAGY; MCCORKLE; ZALKIN, 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 N¹⁰-formilTHF. O sítio de ligação do GAR é mais exposto ao solvente e é posicionado favorecendo o recebimento do grupo formil (WARREN; MAROLEWSKI; BENKOVIC, 1996; ZHANG; MORAR; EALICK, 2008). Em organismos superiores PurN compõe o domínio C-terminal de uma enzima trifuncional que também possui as atividades de PurD e PurM (ZHANG; MORAR; EALICK, 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 Mg²⁺, ATP e formato (LI; FAST; BENKOVIC, 2009; THODEN et al., 2002; ZHANG; MORAR; EALICK, 2008). 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; MORAR; EALICK, 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 VBP como a N5-Carboxiaminoimidazol ribonucleotídeo sintetase (PurK) e PurD

(KAPPOCK; EALICK; STUBBE, 2000; WARREN; MAROLEWSKI; BENKOVIC, 1996).

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; FAST; BENKOVIC, 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-Formilglicinamidina ribonucleotídeo (FGAM) (MORAR et al., 2008; ZHANG; MORAR; EALICK, 2008).

São conhecidas duas isoformas de PurL, codificadas por dois genes descritos como *purL small* e *purL large* (MORAR et al., 2008). 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; MORAR; EALICK, 2008). A PurL *large* é encontrada em eucariotos e bactérias Gram-negativas (MORAR et al., 2008; ZHANG; MORAR; EALICK, 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; MORAR; EALICK, 2008). PurQ é uma glutaminase responsável pela geração da amônia a partir da glutamina. PurS é homologa ao domínio N-terminal da PurL *large* (MORAR et al., 2008).

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).

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; MORAR; EALICK, 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; HOLDEN; FIRESTINE, 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 CO₂ e AIR como substrato, faz a adição direta do grupo carboxilato na posição C5 do anel imidazol, sem a participação da PurK (ZHANG; MORAR; EALICK, 2008; THODEN; HOLDEN; FIRESTINE, 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; HOLDEN; FIRESTINE, 2008). Em vertebrados PurE encontra-se fundida a PurC, formando uma enzima bifuncional

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; MORAR; EALICK, 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 ciclodrolase (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 ciclodrolase (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; MORAR; EALICK, 2008).

Em Archaeas a IMP ciclodrolase 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; XU; WHITE, 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 antimicrobiano (ZHANG; MORAR; EALICK, 2008).

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; CHALEIL; STERNBERG, 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 conhecida como 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 1.3A). Portanto, as primeiras enzimas a surgirem foram aquelas responsáveis pelas últimas reações, e aquelas que evoluíram mais

tardamente 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 1.3B). 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.

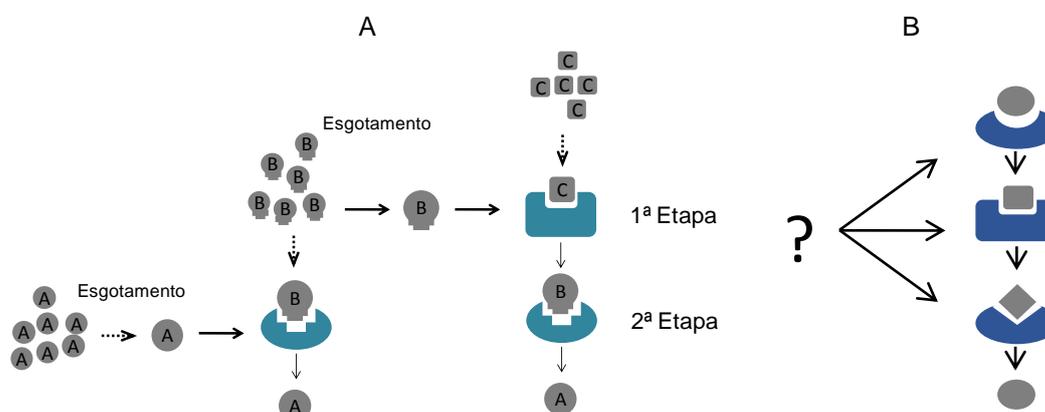


Figura 1.3. Modelos de evolução de vias metabólicas. (A) Retro-evolução; (B) Evolução direta. Fonte: Modificado de SCHMIDT et al., 2003.

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 (CAETANO-ANOLLÉS; KIM; MITTENTHAL, 2007; LAZCANO; MILLER, 1999).

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; MILLER, 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; LECOINTRE, 2002; LAZCANO; MILLER, 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; NAKATANI, 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 biosíntese de purinas e aminoácidos (LAZCANO; MILLER, 1999).

Jensen (1985) 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 1.4A). 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.

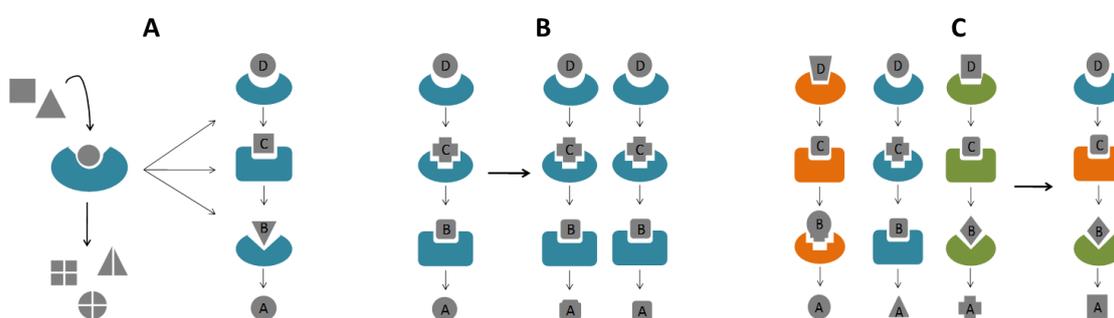


Figura 1.4. Modelos de evolução de vias metabólicas. Fonte: Modificado de SCHMIDT et al., 2003. **A.** Especialização de enzima multifuncional. **B.** Especialização de enzima multifuncional. **C.** Recrutamento enzimático

A partir do surgimento das primeiras vias, outras rotas poderiam surgir pela duplicação de um conjunto enzimático completo (Figura 1.4B), ou pelo recrutamento de enzimas de diferentes vias metabólicas (Figura 1.4C) (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 e colaboradores (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 (CAETANO-ANOLLÉS; KIM; MITTENTHAL, 2007; LAZCANO; MILLER, 1999). Além disso, a dispersão e diversificação das enzimas podem ter ocorrido por eventuais duplicações e mutações das enzimas disponíveis (ALVES; CHALEIL; STERNBERG, 2002).

Caetano-Anollés e colaboradores (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 1.5).

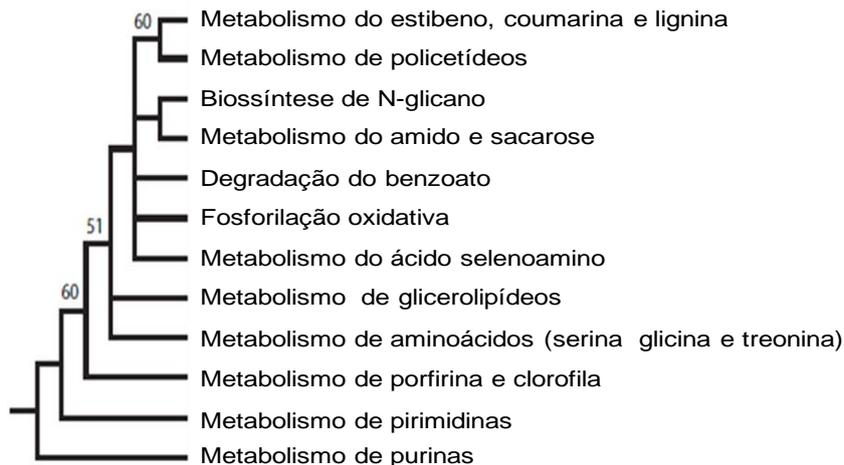


Figura 1.5. Filogenia das vias metabólicas evidenciando que as vias biossintéticas de nucleotídeos foram, possivelmente, as primeiras rotas a surgirem. Fonte: CAETANO-ANOLLÉS; KIM; MITTENTHAL, 2007.

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-ANOLLÉS; KIM; MITTENTHAL, 2007).

OBJETIVOS

Objetivos Gerais

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

Objetivos Específicos

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

Referências

ALVES, R.; CHALEIL, R. A. .; STERNBERG, M. J. . Evolution of Enzymes in Metabolism: A Network Perspective. **Journal of Molecular Biology**, v. 320, n. 4, p. 751–770, jul. 2002.

BERTHOMÉ, R. et al. pur4 mutations are lethal to the male, but not the female, gametophyte and affect sporophyte development in Arabidopsis. **Plant physiology**, v. 147, n. 2, p. 650–60, jun. 2008.

CAETANO-ANOLLÉS, G.; KIM, H. S.; MITTENTHAL, J. E. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 22, p. 9358–63, 29 maio. 2007.

CHUNG, S. O. et al. Genomic organization of purK and purE in *Brevibacterium ammoniagenes* ATCC 6872: purE locus provides a clue for genomic evolution. **FEMS microbiology letters**, v. 137, n. 2-3, p. 265–8, 1 abr. 1996.

CORDÓN, F. Tratado evolucionista de biología, **Aguilar**, Madrid, 1990.

CUNCHILLOS, C.; LECOINTRE, G. Early steps of metabolism evolution inferred by cladistic analysis of amino acid catabolic pathways. **Comptes rendus biologies**, v. 325, n. 2, p. 119–29, mar. 2002.

DANDANELL, G.; SORENSEN, I. S. Identification and sequence analysis of *Sulfolobus solfataricus* purE and purK genes. **FEMS microbiology letters**, v. 154, p. 173–180, 1997.

DRAZEK, E. S. et al. Deletion of purE attenuates *Brucella melitensis* 16M for growth in human monocyte-derived macrophages. **Infection and immunity**, v. 63, n. 9, p. 3297–301, set. 1995.

DURAN, M. et al. Inherited defects of purine and pyrimidine metabolism: laboratory methods for diagnosis. **Journal of inherited metabolic disease**, v. 20, n. 2, p. 227–36, jun. 1997.

FIRESTINE, S. M. et al. Biochemical role of the *Cryptococcus neoformans* ADE2 protein in fungal de novo purine biosynthesis. **Archives of biochemistry and biophysics**, v. 351, n. 1, p. 123–34, 1 mar. 1998.

FIRESTINE, S. M.; DAVISSON, V. J. Carboxylases in de novo purine biosynthesis. Characterization of the *Gallus gallus* bifunctional enzyme. **Biochemistry**, v. 33, n. 39, p. 11917–26, 4 out. 1994.

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

HOROWITZ, N. H. On the evolution of biochemical syntheses. **Proceedings of the National Academy of Sciences of the United States of America**, v. 31, n. 6, p. 153–157, 1945.

HOSKINS, A. A et al. N5-CAIR mutase: role of a CO₂ binding site and substrate movement in catalysis. **Biochemistry**, v. 46, n. 10, p. 2842–55, 13 mar. 2007.

JENSEN, R. A. Biochemical pathways in prokaryotes can be traced backward through evolutionary time. **Molecular biology and evolution**, v. 2, n. 2, p. 92–108, mar. 1985.

KAPPOCK, T. J.; EALICK, S. E.; STUBBE, J. Modular evolution of the purine biosynthetic pathway. **Current opinion in chemical biology**, v. 4, n. 5, p. 567–72, out. 2000.

KIRSCH, D. R.; WHITNEY, R. R. Pathogenicity of *Candida albicans* auxotrophic mutants in experimental infections. **Infection and immunity**, v. 59, n. 9, p. 3297–300, set. 1991.

LAZCANO, A; MILLER, S. L. On the origin of metabolic pathways. **Journal of molecular evolution**, v. 49, n. 4, p. 424–31, out. 1999.

LI, C. et al. X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution. **Structure (London, England : 1993)**, v. 7, n. 9, p. 1155–66, 15 set. 1999.

LI, H.; FAST, W.; BENKOVIC, S. J. Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. **Protein science : a publication of the Protein Society**, v. 18, n. 5, p. 881–92, maio. 2009.

LI, S.-X. et al. Octameric structure of the human bifunctional enzyme PAICS in purine biosynthesis. **Journal of molecular biology**, v. 366, n. 5, p. 1603–14, 9 mar. 2007.

MEYER, E. et al. Purification and characterization of the purE, purK, and purC gene products: identification of a previously unrecognized energy requirement in the purine biosynthetic pathway. **Biochemistry**, v. 31, n. 21, p. 5022–32, 2 jun. 1992.

MOAT, A G.; FRIEDMAN, H. The biosynthesis and interconversion of purines and their derivatives. **Bacteriological reviews**, v. 24, n. 3, p. 309–39, set. 1960.

MORAR, M. et al. Formylglycinamide ribonucleotide amidotransferase from *Thermotoga maritima*: structural insights into complex formation. **Biochemistry**, v. 47, n. 30, p. 7816–30, 29 jul. 2008.

MORETT, E. et al. Systematic discovery of analogous enzymes in thiamin biosynthesis. **Nature biotechnology**, v. 21, n. 7, p. 790–5, jul. 2003.

MUCHMORE, C. R. et al. Crystal structure of glutamine phosphoribosylpyrophosphate amidotransferase from *Escherichia coli*. **Protein science : a publication of the Protein Society**, v. 7, n. 1, p. 39–51, jan. 1998.

MUELLER, E. J. et al. N5-Carboxyaminoimidazole Ribonucleotide: Evidence for a new intermediate and two new enzymatic activities in the de Novo Purine Biosynthetic Pathway of *Escherichia coli*. **Biochemistry**, v. 33, n. 8, p. 2269–2278, 1994.

NAGY, P. L.; MCCORKLE, G. M.; ZALKIN, H. purU, a source of formate for purT-dependent phosphoribosyl-N-formylglycinamide synthesis. **Journal of bacteriology**, v. 175, n. 21, p. 7066–73, nov. 1993.

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-beta-D-ribofuranosyl 5'-monophosphate synthetase. A new enzyme in purine biosynthesis. **The Journal of biological chemistry**, v. 280, n. 12, p. 10881–7, 25 mar. 2005.

RODIONOV, D. A et al. Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. **The Journal of biological chemistry**, v. 277, n. 50, p. 48949–59, 13 dez. 2002.

SAXILD, H. H.; JACOBSEN, J. H.; NYGAARD, P. Functional analysis of the *Bacillus subtilis* purT gene encoding formate-dependent glycinamide ribonucleotide transformylase. **Microbiology**, v. 141, p. 2211–2218, 1995.

SCHMIDT, S. et al. Metabolites: a helping hand for pathway evolution? **Trends in biochemical sciences**, v. 28, n. 6, p. 336–41, jun. 2003.

SHIMAOKA, M. et al. Effect of amplification of desensitized purF and prs on inosine accumulation in *Escherichia coli*. **Journal of bioscience and bioengineering**, v. 103, n. 3, p. 255–61, mar. 2007.

SMITH, P. M. C.; ATKINS, C. A. Purine biosynthesis. Big in cell division, even bigger in nitrogen assimilation. **Plant Physiology**, v. 128, n. 3, p. 793–802, 2002.

THODEN, J. B. et al. PurT-encoded glycinamide ribonucleotide transformylase. Accommodation of adenosine nucleotide analogs within the

active site. **The Journal of biological chemistry**, v. 277, n. 26, p. 23898–908, 28 jun. 2002.

THODEN, J. B.; HOLDEN, H. M.; FIRESTINE, S. M. Structural Analysis of the Active Site Geometry of N5-Carboxyaminoimidazole Ribonucleotide Synthetase from Escherichia coli. **Biochemistry**, v. 47, n. 50, p. 13346–13353, 2008.

TSAI, M. et al. Substrate and product complexes of Escherichia coli adenylosuccinate lyase provide new insights into the enzymatic mechanism. **Journal of molecular biology**, v. 370, n. 3, p. 541–54, 13 jul. 2007.

WARREN, M. S.; MAROLEWSKI, A. E.; BENKOVIC, S. J. A Rapid Screen of Active Site Mutants in Glycinamide Ribonucleotide Transformylase. **Biochemistry**, v. 35, n. 27, p. 8855–8862, 1996.

ZHANG, R. et al. Structure of SAICAR synthase from Thermotoga maritima at 2.2 angstroms reveals an unusual covalent dimer. **Acta crystallographica. Section F, Structural biology and crystallization communications**, v. 62, n. Pt 4, p. 335–9, 1 abr. 2006.

ZHANG, Y.; MORAR, M.; EALICK, S. E. Structural biology of the purine biosynthetic pathway. **Cellular and molecular life sciences : CMLS**, v. 65, n. 23, p. 3699–724, nov. 2008.

Capítulo 2

Diversidade estrutural e evolução da terceira etapa da Via
Biosintética de Purinas em procariotos

Esse capítulo está formatado pelas normas da revista *Biology Direct*

Lista de Figuras

- Figura 2.1. Terceira reação da via biossintética de purinas. Fonte: KAPPOCK; EALICK; STUBBE, 2000. 34
- Figura 2.2. Distribuição dos genes *purN* e *purT* nos 582/1165 gêneros/espécies de procariotos amostradas. 36
- Figura 2.3. Genômica comparativa dos genes *purN* e *purT* em procariotos (A) Distribuição e padrão de ocorrência dos genes *purN* e *purT* em procariotos. As colunas referentes a genômica comparativa indicam os possíveis padrões de ocorrência dos genes *purN* e *purT* nos taxa de procariotos. Os valores de genômica comparativa são apresentados para gênero/espécie. As células em amarelo indicam presença do gene no genoma e as células em branco ausência. (B) Análise de anticorrelação dos genes *purN* e *purT*. Os valores acima de 0,7 são considerados significativos e indicam a anticorrelação dos genes nos gêneros/espécies. 41
- Figura 2.4. Contexto genômico de *purN* e *purT*. Dos 582/1165 gêneros/espécies amostrados, o gene *purN* aparece em um contexto genômico com outros genes da VBP em 63%/69% dos gêneros/espécies. Enquanto em 37%/31% dos gêneros/espécies *purN* não está em contexto com os genes da VBP. O gene *purT* está no contexto com outros genes da biossíntese de purinas em 4%/3% dos gêneros/espécies. E na maioria dos casos (96%/97% dos gêneros/espécies) esse gene não está em contexto com os genes da VBP. Ao lado está representado o contexto genômico de *purN* e *purT*. As caixas vazias representam genes não relacionados com a biossíntese de purinas. 43
- Figura 2.5. Filogenia de *PurT*. Árvore filogenética estimada a partir das sequências de aminoácidos da *PurT* pelo método da Máxima Verossimilhança. (A) Árvore enraizada pelas sequências de *PurK*, o 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. (B) Árvore não enraizada destacando em vermelho as sequências de *PurK*. A árvore filogenética foi criada usando o PhyML 3.0 e editada com o MEGA6 (TAMURA et al., 2013). O valor de confiança dos ramos foi estimado pelo método aLRT-SH. 47
- Figura 2.6. Filogenia de *PurN*. Árvore filogenética estimada a partir das sequências de aminoácidos da *PurN* pelo método da Máxima Verossimilhança. A árvore enraizada foi enraizada no ponto médio, o 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 filogenética foi criada usando o PhyML 3.0 e editada com o MEGA6 (TAMURA et al., 2013). O valor de confiança dos ramos foi estimado pelo método aLRT-SH. 50
- Figura 2.7. Nível de expressão de *purN* e *purT* em *E. coli*. (A e B) *E. coli* cultivada em meio LB. (A) Resultados da repetição 1 e (B) Resultados da repetição 2. (C e D) *E. coli* cultivada em meio mínimo. (C) Resultados da repetição 1 e (D) Resultados da repetição 2. O gene *cysG* foi utilizado para

normalizar a quantidade inicial de cDNA das amostras e determinar o nível de expressão nas fases de crescimento exponencial (Log), estacionária (Est) e de declínio (Decl). O nível de expressão foi calculado tomando a fase exponencial de crescimento como referência. As barras de erro indicam os desvios padrão das triplicatas de cada amostra.	52
Figura 2.8. Sobreposição da estrutura terciária da PurT (PDB: 1KJ8) e PurK (PDB: 3ETJ) de Escherichia coli.	54
Figura 2.9. 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.	55
Figura 2.10. 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.	56
Figura 2.11. Evolução das enzimas da família ATP-grasp proposta por Zhang e colaboradores (2008). Fonte: (ZHANG; MORAR; EALICK, 2008)	61

Lista de Tabelas

Tabela 2.1. Filos e classes dos Domínios Archaea e Bactéria amostradas no estudo. Entre parênteses está destacado o número total de Filos e Classes (EUZÉBY, 1997).	38
Tabela 2.2. Espécies que possuem purT em contexto genômico com outros genes da VBP	44
Tabela 2.3. Valores cinéticos das enzimas PurN e PurT de Escherichia coli	58
Tabela 2.4. Relação de bactérias patogênicas para animais e plantas que possuem apenas o gene purT.	61
Tabela 2.5. Sequências dos oligonucleotídeos utilizados para a reação de PCR em Tempo Real.....	66

Resumo

Introdução

Apesar do vasto conhecimento bioquímico da biossíntese de purinas ainda existem lacunas sobre a sua evolução e diversidade estrutural em procariotos. Por exemplo, reações da terceira etapa desta via envolvem a participação dos genes *purN/purT*. O gene *purT* não está presente no genoma de linhagens eucarióticas, como plantas e metazoários, portanto, podem ser alvos para fármacos antimicrobianos, contudo, sua distribuição nas linhagens procarióticas ainda é desconhecida.

Resultados

Para compreender como duas enzimas foram selecionadas para desempenhar a mesma função durante a evolução dos procariotos foi realizada a genômica comparativa e a filogenia molecular das sequências de aminoácidos de PurT e PurN. Os resultados mostram que o gene *purN* está amplamente distribuído nas linhagens procarióticas e predomina nos genomas das espécies analisadas nesse estudo. Além disso, na maioria dos genomas analisados, *purN* possui em seu contexto genômico outros genes da biossíntese de purinas. O mesmo não é observado em relação a *purT*. A filogenia molecular da PurT mostrou que esse gene foi intensamente transferido lateralmente entre os taxa de Bactéria e Archaea. No entanto, a análise de expressão gênica desses genes mostrou que *purT* é mais expresso que *purN* o que pode ser um indício de que PurT pode estar envolvido em outro processo biológico além da síntese de purinas.

Conclusões

Mesmo desempenhando a mesma função os genes *purT* e *purN* não estão igualmente distribuídos nas linhagens procarióticas. Em conjunto os resultados indicam que a condição ancestral para a VBP é PurN realizando a terceira reação da via. A origem de *purT* ocorreu posteriormente e transferência horizontal de genes foi um fator relevante na distribuição de *purT* nas linhagens procarióticas.

Abstract

Background

Despite the wide biochemical knowledge of purine biosynthesis gaps still exist about its evolution and structural diversity in prokaryotes. The reaction in the third step of PBP involves the participation of *purN/purT* genes. The *purT* gene is absent in eukaryotic strains like plants and metazoan, thus, this genes can be targets for antimicrobial drugs, however, its distribution in prokaryotic strains is still unknown.

Results

To understand how the two enzymes were selected to perform the same function during the evolution of prokaryotes we performed comparative genomics and phylogeny analysis *purT* and *purN*. The results show that *purN* gene is widely distributed in prokaryotic lineages and predominates in the genomes of the species analyzed in this study. Moreover, in most genomes analyzed *purN* is in genomic context with other genes of purine biosynthesis. The same is not observed for the *purT*. Furthermore, the molecular phylogeny of *purT* showed that this gene was intensively laterally transferred between the Taxa Bacteria and Archaea. However, the gene expression analysis showed that *purT* gene is over expressed than *purN* which may be a sign that *purT* may be involved in other biological processes beyond purine biosynthesis.

Conclusions

Even performing the same function *purT* and *purN* genes are not equally distributed in prokaryotic lineages. Our results indicate that the VBP ancestral condition in the third step reaction is PurN performing the reaction. The origin of *purT* occurred later and horizontal gene transfer was a relevant factor in its distribution on prokaryotic lineages.

Introdução

A evolução de vias biosintéticas é uma questão que está 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 biosintéticas estão sendo elucidados usando os dados obtidos a partir de genomas totalmente sequenciados (MORETT et al., 2003; RODIONOV et al., 2002).

Na terceira reação da via biosintética de purinas ocorre a formilação do grupamento amino do intermediário GAR (Figura 2.1). Em procariotos já foram descritas duas enzimas capazes de catalisar essa reação, a PurT e a PurN. No entanto, essas enzimas empregam diferentes mecanismos e utilizam distintos cofatores para realizar essa reação (WARREN; MAROLEWSKI; BENKOVIC, 1996). PurN faz a formilação direta do GAR utilizando como cofator o N¹⁰-formiltetrahydrofolato (N¹⁰-formilTHF) (NAGY; MCCORKLE; ZALKIN, 1993). A mesma reação quando catalisada por PurT, tem como cofatores Mg²⁺, ATP e formato (LI; FAST; BENKOVIC, 2009). A maior parte do formato utilizado por PurT provem da quebra do N¹⁰-formilTHF em tetrahydrofolato e formato, realizada pela enzima PurU (NAGY; MCCORKLE; ZALKIN, 1993).

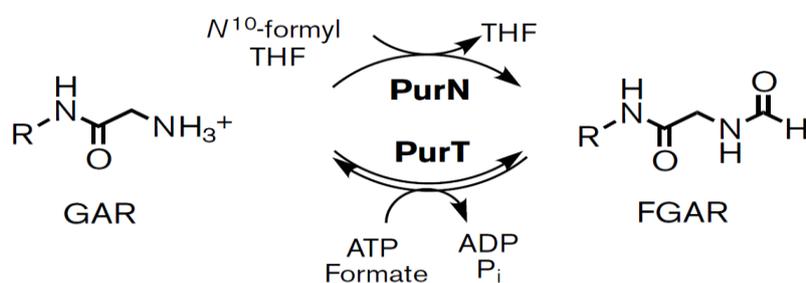


Figura 2.1. Terceira reação da via biosintética de purinas. Fonte: KAPPOCK; EALICK; STUBBE, 2000.

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 (SAMPEI et al., 2013). Ela deriva do fato deste gene estar presente no genoma da maioria dos organismos usados como modelo de estudo da via biosintética de purinas. Logo, a distribuição de *purN* e *purT* nas linhagens procarióticas ainda não foi completamente elucidada.

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 biosintética 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*.

Resultados

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

Foram amostrados 582/1165 gêneros/espécies com o genoma completamente sequenciado, nos quais foram realizadas as buscas pelos genes *purN* e *purT*. Dentre os gêneros/espécies amostrados 86%/84% possuem *purN* e 27%/31% possuem *purT* (Figura 2.2). Do total de gêneros/espécies 23%/26% possuem ambos os genes enquanto que 66%/57% possuem apenas *purN* e 5%/5% apenas *purT*. Em 11% dos gêneros/espécies os genes *purN* e *purT* não estão presentes. A amostragem realizada nesse trabalho é representativa, levando em consideração a diversidade de procariotos conhecida (Tabela 2.1).

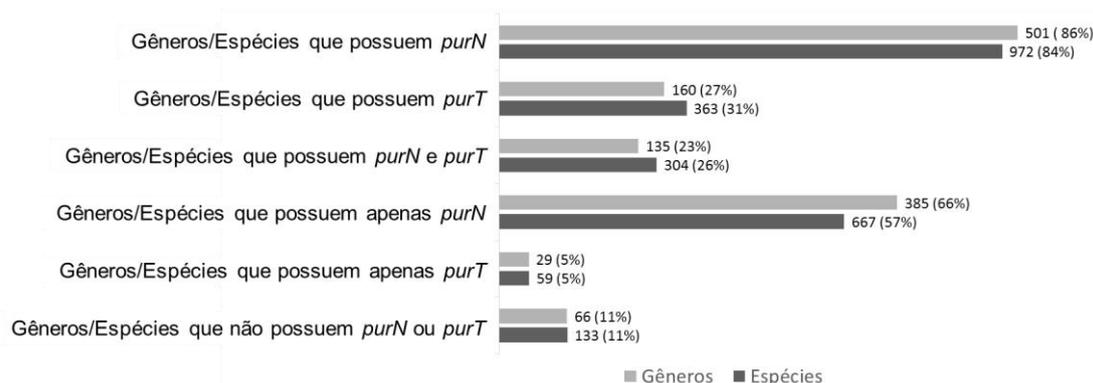


Figura 2.2. Distribuição dos genes *purN* e *purT* nos 582/1165 gêneros/espécies de procariotos amostradas.

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 dos genes *purN* e *purT* no genoma dos 11% gêneros/espécies mencionados anteriormente.

O relato de novos genes envolvidos em processos celulares amplamente conhecidos já foi relatado literatura (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.

Tabela 2.1. Filos e classes dos Domínios Archaea e Bactéria amostradas no estudo. Entre parênteses está destacado o número total de Filos e Classes (EUZÉBY, 1997).

Domínio	Filo	Classe	Sigla	Amostragem (Gênero/Espécie)	
Archaea	Crenarchaeota	Thermoprotei	CR		
		Euryarchaeota	EU	1/1	
		Archaeoglobi		2/4	
		Halobacteria		13/15	
		Methanobacteria		5/8	
		Methanococci		4/11	
		Methanomicrobia		13/18	
		Methanopyri		1/1	
		Thermococci		2/12	
	Thermoplasmata		2/3		
	Korarchaeota	?	KO	1/1	
	Thaumarchaeota	?	TA	1/1	
	Nanoarchaeota	?	NA	1/1	
Bacteria	Firmicutes	Bacilli	FB	25/107	
		Clostridia	FC	44/94	
		Erysipelotrichi	FE	2/2	
		Negativicutes	FN	5/6	
		Tenericutes	Mollicutes	FT	7/29
		Actinobacteria	Actinobacteria	AB	69/128
	Cyanobacteria	?	CB	11/27	
		Gloeobacteria		1/1	
		Aquificae	Aquificae	AQ	8/9
		Chloroflexi	Anaerolineae	CF	1/1
			Chloroflexi		3/6
			Dehalococcoidetes		2/6
			Thermomicrobia		2/2
		Deinococcus/thermus	Deinococci	DTH	6/14
		Acidobacteria	?	AC	2/2
			Acidobacteriia		3/3
	Solibacteres			1/1	
	Fusobacteria		Fusobacteriia	FU	5/5
	Planctomycetes		Planctomycetia	PLC	4/5
	Thermotogae		Thermotogae	THR	6/13
	Bacteroidetes	?	BC	3/3	
		Bacteroidia		9/17	
		Cytophagia		7/7	
		Flavobacteriia		18/26	
		Sphingobacteriia		7/8	
		Clorobi	Chlorobia	CL	5/10

Tabela 2.1. *Continuação.*

Domínio	Filo	Classe	Sigla	Amostragem
	Verrucomicrobia	?	VM	1/1
		Opitutae		2/2
		Verrucomicrobiae		1/1
	Dictyoglomi	Dictyoglomia	DC	1/2
	Elusimicrobia	Elusimicrobia	EM	1/1
	Nitrospirae	Nitrospira	NI	2/2
	Gemmatimonadetes	Gemmatimonadetes	GT	1/1
	Synergistetes	Synergistia	SY	3/3
	Deferribacteres	Deferribacteres	DF	4/4
	Chrysiogenetes	Chrysiogenetes	CS	1/1
	Fibrobacteres	Fibrobacteria	FR	1/1
	Thermodesulfobacteria	Thermodesulfobacteria	TB	2/2
	Chlamydiae	Chlamydia	CH	6/12
	Spirochaetes	Spirochaetia	SP	6/28
	Proteobacteria	Deltaproteobacteria	DT	25/42
		Epsilonproteobacteria	EP	11/24
		Alphaproteobacteria	AF	67/135
		Betaproteobacteria	BT	44/83
		Gammaproteobacteria	GM	84/181

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

A genômica comparativa mostrou que os genes *purN* e *purT* não estão igualmente distribuídos nos 38 fillos/classes procarióticos incluídos nessa análise. Com relação a presença desses genes, os seguintes padrões de ocorrência foram observados:

(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; e (IV) o organismo não possui nenhum dos genes em seu genoma (Figura 2.3).

Filos/Classes	(A)				(B)	
	Genômica Comparativa				Escores de Anticorrelação	
	purN	purN			Gêneros	Espécies
Crenarchaeota		2/8	9/14*	8/9	1	1
Euryarchaeota	1/1	27/34*	7/24	9/14	0,88	0,90
Korarchaeota				1/1	0	0
Thaumarchaeota		1/1			1	1
Nanoarchaeota				1/1	0	0
Firmicutes Bacilli	2/9	22/90*		4/8	0,76	0,75
Firmicutes Clostridia		42/90*		2/4	0,97	0,98
Firmicutes Erysipelotrichia		1/1		1/1	1	1
Firmicutes Negativicutes		5/6			1	1
Tenericutes				7/29	0	0
Actinobacteria	9/37	58/83*	2/7	1/1	0,72	0,42
Cyanobacteria	7/16*	7/12			0	-0,14
Aquificae	5/6*	3/3			-0,25	-0,33
Chloroflexi		7/14		1/1	1	1
Deinococcus-Thermus		6/14			1	1
Acidobacteria		6/6			1	1
Fusobacteria		5/5			1	1
Planctomycetes		4/5			1	1
Thermotogae		5/12		1/1	1	1
Bacteroidetes	9/16	34/41*	1/2	2/2	0,63	0,47
Chlorobi	3/5	3/5			0	0
Verrucomicrobia	1/1	3/3			0,50	0,50
Dictyoglomi		1/2			1	1
Elusimicrobia		1/1			1	1
Nitrospira		2/2			1	1
Gemmatimonadetes		1/1			1	1
Synergistetes		3/3			1	1
Deferribacteres		4/4			1	1
Chrysiogenetes		1/1			1	1
Fibrobacteres	1/1				-1	-1
Thermodesulfobacteria		2/2			1	1
Chlamydiae				6/12	0	0
Spirochaetes	1/3	4/14*		3/11	0,63	0,75
Deltaproteobacteria	5/16	21/24*		2/2	0,52	0,12
Epsilonproteobacteria	9/12*	2/10		1/2	0,64	0
Alphaproteobacteria	7/9	55/101*		8/25	0,79	0,86
Betaproteobacteria	26/53*	16/22	4/6	2/2	-0,14	-0,28
Gammaproteobacteria	49/121*	31/47	6/6	6/7	-0,13	-0,37

Figura 2.3. Genômica comparativa dos genes *purN* e *purT* em procariotos (A) Distribuição e padrão de ocorrência dos genes *purN* e *purT* em procariotos. As colunas referentes a genômica comparativa indicam os possíveis padrões de ocorrência dos genes *purN* e *purT* nos taxa de procariotos. Os valores de genômica comparativa são apresentados para gênero/espécie. As células em amarelo indicam presença do gene no genoma e as células em branco ausência. (B) Análise de anticorrelação dos genes *purN* e *purT*. Os valores acima de 0,7 são considerados significativos e indicam a anticorrelação dos genes nos gêneros/espécies.

O padrão I foi encontrado em 15 dos 38 filios/classes, e é predominante no filo Cyanobacteria, Aquificae e nas Classes Epsilon-, Beta- e Gama-proteobactéria. Nos demais filios/classes em que foi observado o padrão estrutural I esse não foi predominante ou o Taxa foi pouco amostrado, como é o caso de Fibrobacteres.

O padrão estrutural II ocorre em 33 dos 38 filios/classes amostradas, sendo predominante nos filios Euryarchaeota, Actinobacteria, Bacteroidetes e Spirochaetes e nas classes Bacilli, Clostridia, Delta- e Alfa-proteobacteria. No filo Chlorobi o número de gêneros/espécies que apresentam o padrão II é igual ao número de gêneros/espécies que possuem o padrão I. A maioria dos gêneros/espécies do filo Verrucomicrobia apresenta o padrão estrutural II. Porém, o número de gêneros/espécies amostrados nesse filo não permitem outras conclusões além de que os padrões estruturais I e II ocorrem nos gêneros/espécies deste filo (Figura 2.3).

O padrão estrutural III está presente em 6 filios/classes, porém, é predominante apenas no filo Crenarchaeota. Em todos os filios/classes onde *purT* é encontrado, *purN* também está presente, evidenciando a predominância desse gene em relação a *purT*.

Todas as espécies dos filios Tenericutes e Chlamydiae analisadas nesse estudo não possuem os genes *purN* e *purT*, por isso o *score* de anticorrelação desses taxa é 0. O mesmo ocorre nos filios Korarchaeota e Nanoarchaeota que são representados por apenas uma espécie. Esses filios são compostos por espécies que são parasitas

intracelulares e epissimbionte, respectivamente, logo, a ecologia das espécies destes filos pode explicar a ausência dos genes da biossíntese de purinas.

A ausência desses genes também foi observada em espécies parasitas intracelulares ou simbioses pertencentes aos filos Crenarchaeota, Euryarchaeota, Actinobactéria, Chloroflexi, Thermotogae, Bacteroidetes e Spirochaetales, e nas classes Firmicutes Bacilli, Clostridia e Erysipelotrichia, Alfa-, Epsilon- e Gama-proteobacteria. 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.

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

Na maioria dos gêneros/espécies pesquisados (63%/69%) *purN* está em um contexto genômico associado com outros genes da biossíntese de purinas. Em 37%/31% dos gêneros/espécies o contexto genômico de *purN* não possui genes relacionados a síntese de purinas (Figura 2.4).

Em todas as espécies de Archaeas incluídas nesse estudo que possuem *purN* esse gene encontra-se em contexto com outros genes da biossíntese de purinas. Nas Gram-positivas esse padrão é encontrado em espécies de todos os filos, com exceção do filo Firmicutes que não possui a PBP. E *purN* sem outros genes da VBP em contexto está presente em 3 dos 6 filos de Gram-positivas, no entanto, esse padrão não é tão abundante quanto o outro. Dentre as Gram negativas ambos os contextos estão bem representados dentre nos filos/classes, no entanto, *purN* em contexto é predominante levando em consideração o número de gêneros/espécies.

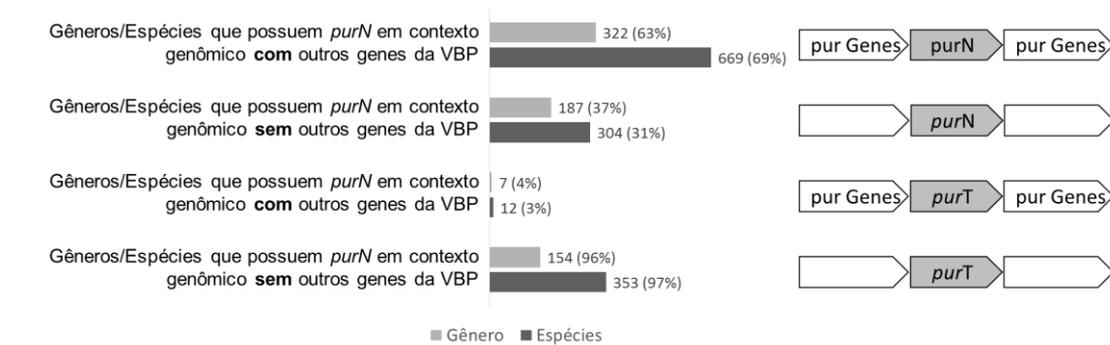


Figura 2.4. Contexto genômico de *purN* e *purT*. Dos 582/1165 gêneros/espécies amostrados, o gene *purN* aparece em um contexto genômico com outros genes da VBP em 63%/69% dos gêneros/espécies. Enquanto em 37%/31% dos gêneros/espécies *purN* não está em contexto com os genes da VBP. O gene *purT* está no contexto com outros genes da biossíntese de purinas em 4%/3% dos gêneros/espécies. E na maioria dos casos (96%/97% dos gêneros/espécies) esse gene não está em contexto com os genes da VBP. Ao lado está representado o contexto genômico de *purN* e *purT*. As caixas vazias representam genes não relacionados com a biossíntese de purinas.

O gene *purT* que está em um contexto genômico com outros genes da via biossintética de purinas em apenas 4%/3% gêneros/espécies amostradas (Figura 2.4). As espécies que possuem *purT* em contexto com outros genes da VBP pertencem aos filos Crenarchaeota e Actinobacteria e a classe Alphaproteobacteria (Tabela 2.2). É possível observar uma sintenia no contexto das espécies pertencentes a mesma família. Isso pode ser um indicativo de que a *purT* presente nessas espécies se originou em um evento de THL no ancestral dessas famílias.

Tabela 2.2. Espécies que possuem *purT* em contexto genômico com outros genes da VBP

Filo/Classe	Família	Espécies	Contexto genômico
Crenarchaeota	Desulfurococcaceae	<i>Desulfurococcus kamchatkensis</i> 1221n	
		<i>Ignisphaera aggregans</i> DSM 17230	
	Thermoproteaceae	<i>Caldivirga maquilingensis</i> IC-167	
		<i>Vulcanisaeta distributa</i> DSM 14429	
		<i>Vulcanisaeta moutnovskia</i> 768-28	
Actinobacteria	Bifidobacteriaceae	<i>Bifidobacterium adolescentis</i> ATCC 15703	
		<i>Bifidobacterium animalis lactis</i> AD011	
		<i>Bifidobacterium bifidum</i> PRL2010	
		<i>Bifidobacterium breve</i> UCC2003	
		<i>Bifidobacterium dentium</i> Bd1	
		<i>Gardnerella vaginalis</i> ATCC 14019	
Alphaproteobacteria	Sphingomonadaceae	<i>Sphingobium chlorophenolicum</i> L-1	

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

A análise de anticorrelação nada mais é do que uma avaliação da não correlação entre os genes *purN* e *purT* nos genomas das espécies estudadas. O valor significativo de anticorrelação indica que quando um gene está presente o outro tende a estar ausente no genoma dos organismos (MORETT et al, 2003). Os resultados da análise de anticorrelação mostraram que esses genes anticorrelacionam tanto em nível de gênero, quanto de espécie nos filo/classes de Crenarchaeota e Euryarchaeota, Firmicutes Bacilli, Clostridia e Negativicutes, Chloroflexi, Deinococcus-Thermus, Acidobacteria, Fusobacteria, Planctomycetes, Thermotogae, Synergistetes, Deferribacteres e Alfarproteobacteria. No filo Actinobacteria a anticorrelação é percebida apenas em nível de gênero e no filo Spirochaetes a anticorrelação só ocorre em nível de espécie, esse desvio se deve ao viés da amostragem de gêneros e espécies (Figura 2.3).

Nos filos/classes Cyanobacteria, Aquificae, Bacteroidetes, Chlorobi, Verrucomicrobia, Delta-, Epsilon-, Beta- e Gama-proteobacteria 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 (Figura 2.3).

Nos filos/classes Thaumarchaeota, Firmicutes Erysipelotrichia, Dictyoglomi, Elusimicrobia, Nitrospirae, Gemmatimonadetes, Chrysiogenetes e Thermodesulfobacteria os genes anticorrelacionam. Na maioria desses 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 que esta presente no filo ou classe é *purN*, mais uma vez evidenciando a predominância desse gene em relação a *purT* nos filos/classes amostrados.

Filogenia de PurT

A topologia da árvore filogenética da PurT mostra que a transferência lateral de genes (TLG) influenciou significativamente a distribuição desse gene nas linhagens procarióticas. Isso pode ser observado pela formação de clados contendo a PurT de espécies pertencentes a diferentes Taxa (Figura 2.5). A maiorias das espécies que possuem PurT pertencem a filos de bactérias Gram-negativas. As PurTs de espécies Gram-positivas não agrupam num único clado, mas sim com espécies de Gram-negativas.

As PurTs de Firmicutes Bacilli agrupam com as PurTs de Deltaproteobacteria pertencentes as famílias *Pelobacteraceae* e *Geobacteraceae* (Tabela Suplementar). As PurTs de Actinobacteria estão distribuídas em cinco grupos. Um dos grupos contém a PurT das famílias *Pseudonocardiaceae* e *Nocardiaceae* (Tabela Suplementar). Essas PurTs agrupam junto com parte das PurTs da família *Burkholderiaceae* do filo Betaproteobacteria. Cinco espécies da família Mycobacteriaceae *Mycobacterium abscessus* ATCC 19977 (PI:169628284), *Amycolicococcus subflavus* DQS3-9A1

(PI:404442360), *Mycobacterium smegmatis* str. MC2 155 (PI:118473801), *Mycobacterium gilvum* PYR-GCK (PI:145222627) e *Mycobacterium vanbaalenii* PYR-1 (PI:120405630) também agrupam com espécies da família *Burkholderiaceae*, que não estão no grupo citado anteriormente (Tabela Suplementar). As demais espécies de *Mycobacteriaceae*, a família *Corynebacteriaceae*, *Tsukamurellaceae* e *Gordoniaceae* formam um terceiro grupo de PurTs de Actinobacterias. Essas PurTs agrupam com PurT de espécies da classe Gamaproteobacteria, das famílias *Methylococcaceae* e *Moraxellaceae*. A espécie da família *Actinomycetaceae* agrupa com espécies da família *Acidithiobacillaceae*, que também pertencem a classe Gamaproteobacteria, e forma o quarto grupo de Actinobacterias. O último grupo de Actinobacterias é formado pelas espécies das famílias *Promicromonosporaceae* e *Bifidobacteriaceae* que agrupam com a família *Xanthomonadaceae*.

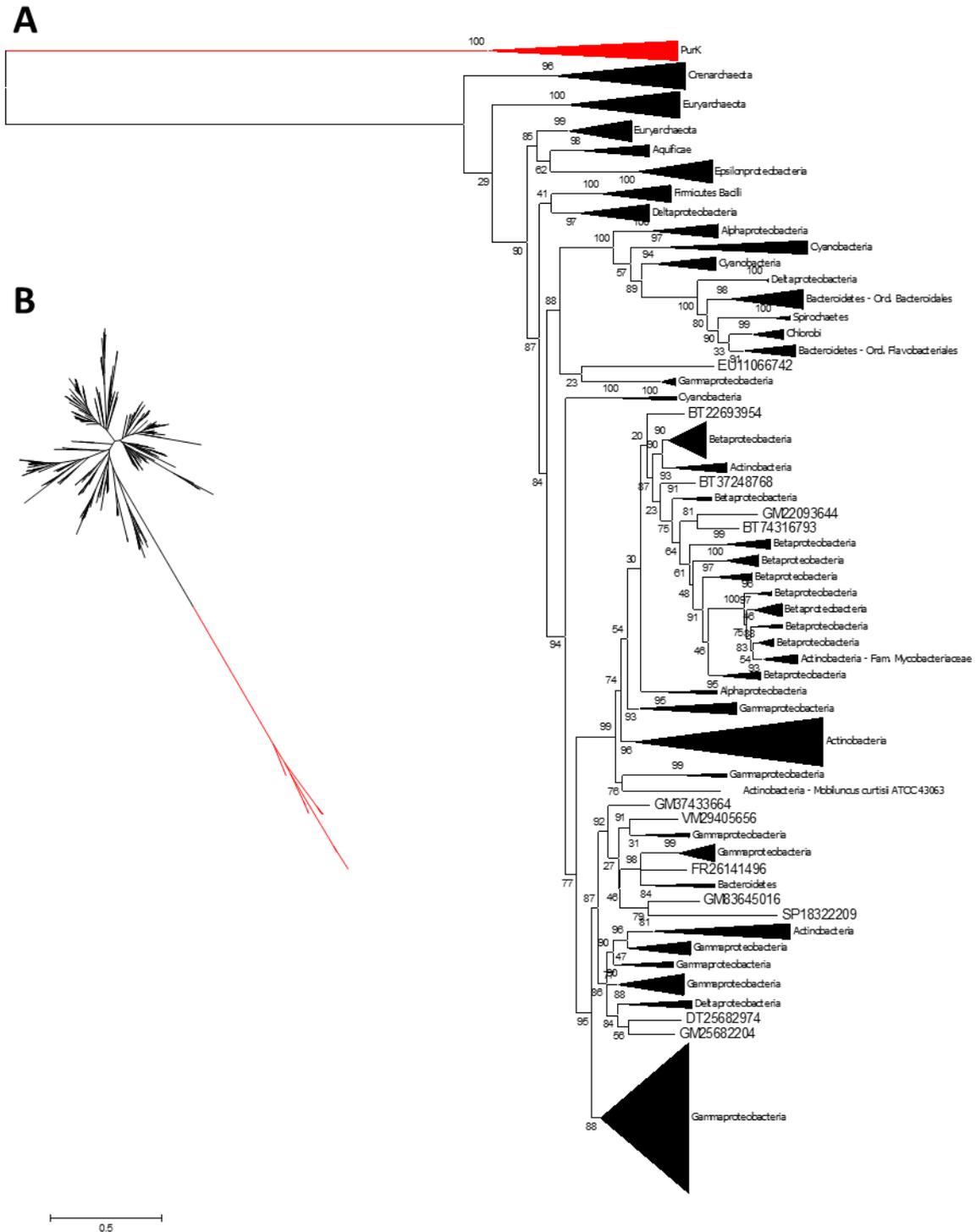


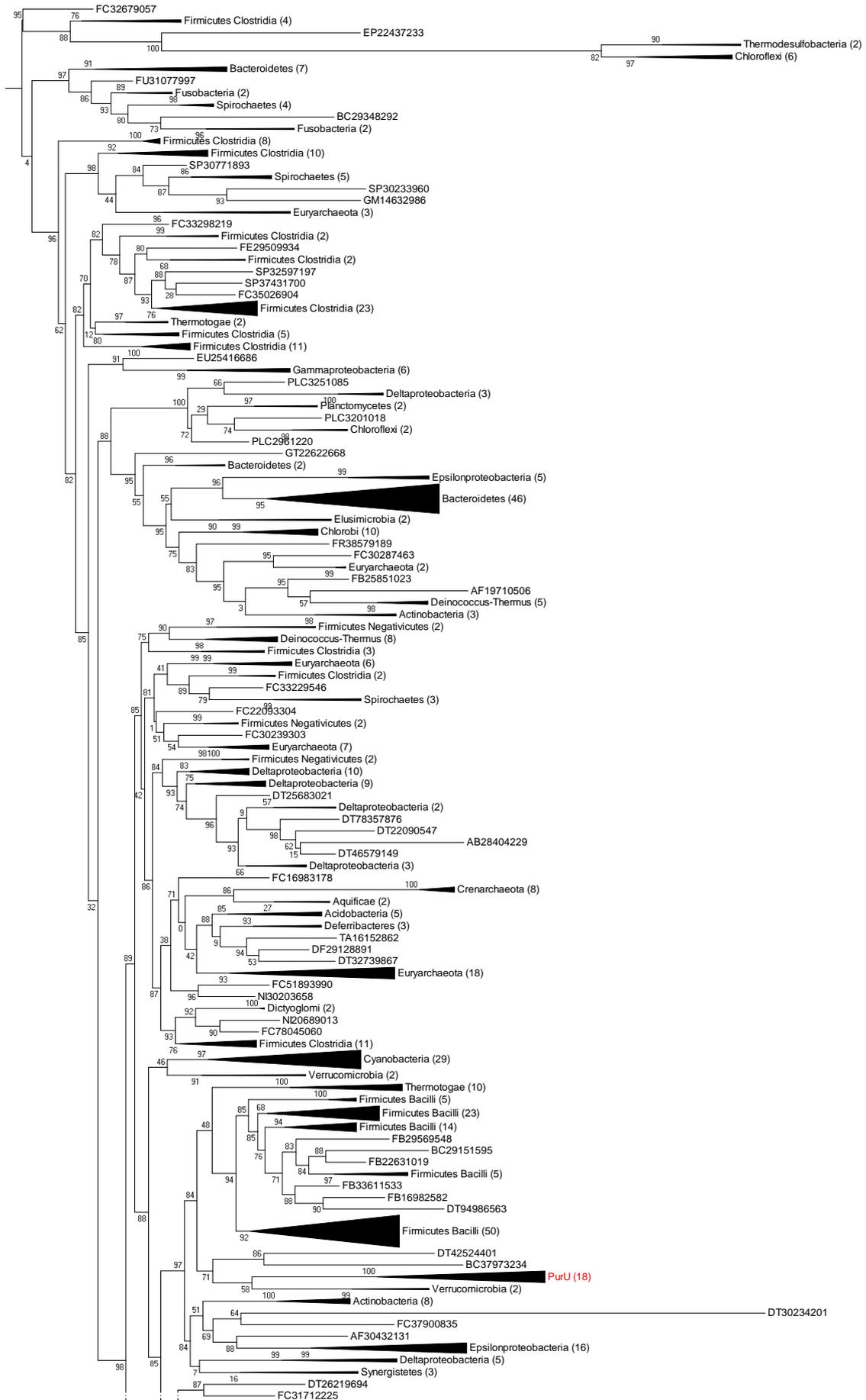
Figura 2.5. Filogenia de PurT. Árvore filogenética estimada a partir das sequencias de aminoácidos da PurT pelo método da Máxima Verossimilhança. (A) Árvore enraizada pelas sequencias de PurK, o clados que formam um grupo monofilético em nível de filo ou classe, foram colapsados. O nome filo/classe e o número de espécies em cada clado foi especificado em cada ramo. (B) Árvore não enraizada destacando em vermelho as sequencias de PurK. A árvore filogenética foi criada usando o PhyML 3.0 e editada com o MEGA6 (TAMURA et al., 2013). O valor de confiança dos ramos foi estimado pelo método aLRT-SH.

Dentro dos grupos de Gram-negativas também observa-se casos de TLG. Em Alfa-proteobacteria, cinco famílias possuem PurT. A PurT dessas famílias estão divididas em dois grupos (Tabela Suplementar). Um deles possui as PurTs das famílias *Rhodospirillaceae*, *Erythrobacteraceae* e *Sphingomonadaceae* que agrupa com as PurTs de Cianobacteria, Bacteroidetes, Clorobi, Spirochaetae e com a família Myxococcales da classe Deltaproteobacteria. O outro grupo de Alfacaproteobacteria é formado pelas PurTs de espécies das famílias *Caulobacteraceae* e *Phyllobacteriaceae* emergem dentro do grupo das PurTs de Betaproteobacterias (Tabela Suplementar).

As PurTs de Deltaproteobacteria apresentam um agrupamento em nível de ordem. Das oito ordens de Deltaproteobacteria amostradas, três possuem PurT. As PurTs da ordem Desulfovibrionales agrupam com Gamaproteobacteria, as da ordem Desulfuromonadales com Firmicutes Bacilli e as PurTs de Deltaproteobacteria pertencentes a ordem Myxococcales formam um clado com PurTs dos filos Bacteroidetes, Clorobi e Spirochaetae. As PurTs do domínio Archaea e das Gram-negativas Aquificae e Epsilonproteobacteria emergem junto ao clado das PurTs, formando o grupo mais basal de PurTs.

Filogenia de PurN

Assim como na filogenia de PurT a árvore filogenética de PurN também mostra que a TLG foi um evento frequente na história evolutiva dessa enzima (Figura 2.6). Na árvore filogenética de PurN as enzimas das Proteobacterias das classes Gama, Beta e Alfacaproteobacteria estão agrupadas no mesmo clado. A exceção são as PurNs das espécies de Alfacaproteobacteria *Parvularcula bermudensis* HTCC2503 e *Phenyllobacterium zucineum* HLK1, e as Gamaproteobacterias das famílias *Legionellaceae* e *Francisellaceae* (Tabela Suplementar).



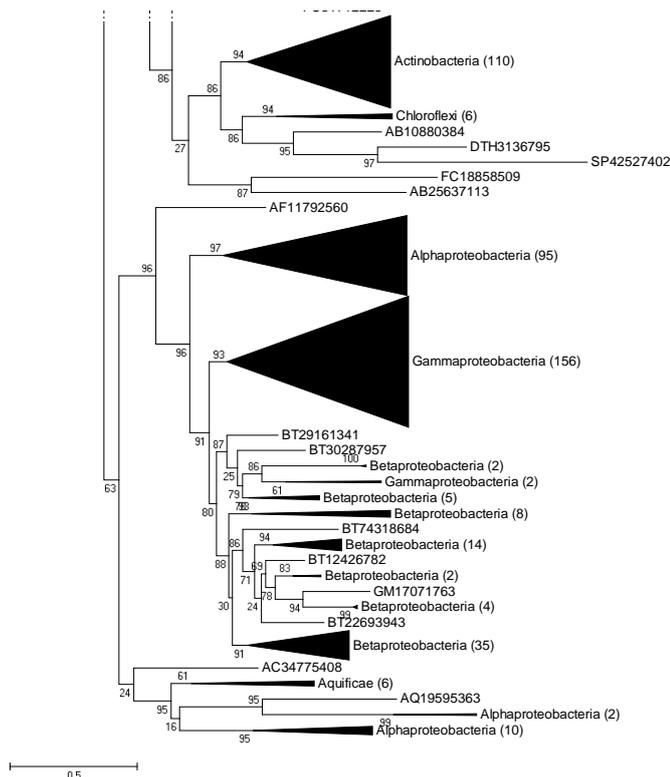


Figura 2.6. Filogenia de PurN. Árvore filogenética estimada a partir das sequências de aminoácidos da PurN pelo método da Máxima Verossimilhança. A árvore enraizada foi enraizada no ponto médio, o clados que formam um grupo monofilético em nível de filo ou classe, foram colapsados. O nome filo/classe e o número de espécies em cada clado foi especificado em cada ramo. A árvore filogenética foi criada usando o PhyML 3.0 e editada com o MEGA6 (TAMURA et al., 2013). O valor de confiança dos ramos foi estimado pelo método aLRT-SH.

Nas Gram-positivas da classe Firmicutes Bacilli, todas as PurNs agrupam-se em um único clado, com exceção da PurN de *Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446*. Nessa espécie a PurN não está em contexto genômico com outros genes da VBP como a PurN das outras espécies da Classe Bacilli. As PurNs de Actinobacteria também formam um único clado, a exceção são as enzimas da família *Coriobacteriaceae* e das espécies *Conexibacter woesei DSM 14684*. As espécies *Kitasatospora setae KM-6054*, *Frankia sp. CcI3*, *Frankia sp. EuI1c* possuem duas cópias da do gene *purN*. Nas referidas espécies de Actinoacteria, uma das cópias do gene *purN* tem em seu contexto genômico outros genes da biossíntese de purinas, essa cópia agrupa com as demais PurNs de Actinoacteria. A outra cópia do gene não

possuem em seu contexto genômico genes da VBP, e essas cópias agrupam com espécies de *Deinococcus-Thermus*. Outros casos de TLG entre filos/classes foram observados na filogenia de PurN. Esses casos envolvem a PurN de outras taxa como, Epsilon e Deltaproteobacteria, Bacteroidetes, Firmicutes Clostridiales entre outros.

À filogenia de PurN foram adicionadas sequências de PurU, enzima homóloga a PurN, com o objetivo utilizar essa segunda enzima como grupo externo da árvore filogenética. No entanto, a árvore não enraizada mostrou que as PurUs formam um clado dentro das PurNs e não um grupo mais distante. Assim a árvore apresentada foi enraizada no ponto médio.

Expressão gênica

Outro aspecto que também foi investigado em busca de uma possível explicação para o padrão de distribuição dessas enzimas foi o padrão de expressão gênica de *purT* e *purN*. Para isso a reação de PCR quantitativo em tempo real (qPCR) foi utilizada para comparar o nível de expressão dos dois genes nas fases do crescimento exponencial, estacionário e de declínio na bactéria *Escherichia coli*.

O ensaio de expressão mostrou que em todas as fases de crescimento da *E. coli* o nível de expressão de *purT* é maior do que o de *purN* (Figura 2.7).

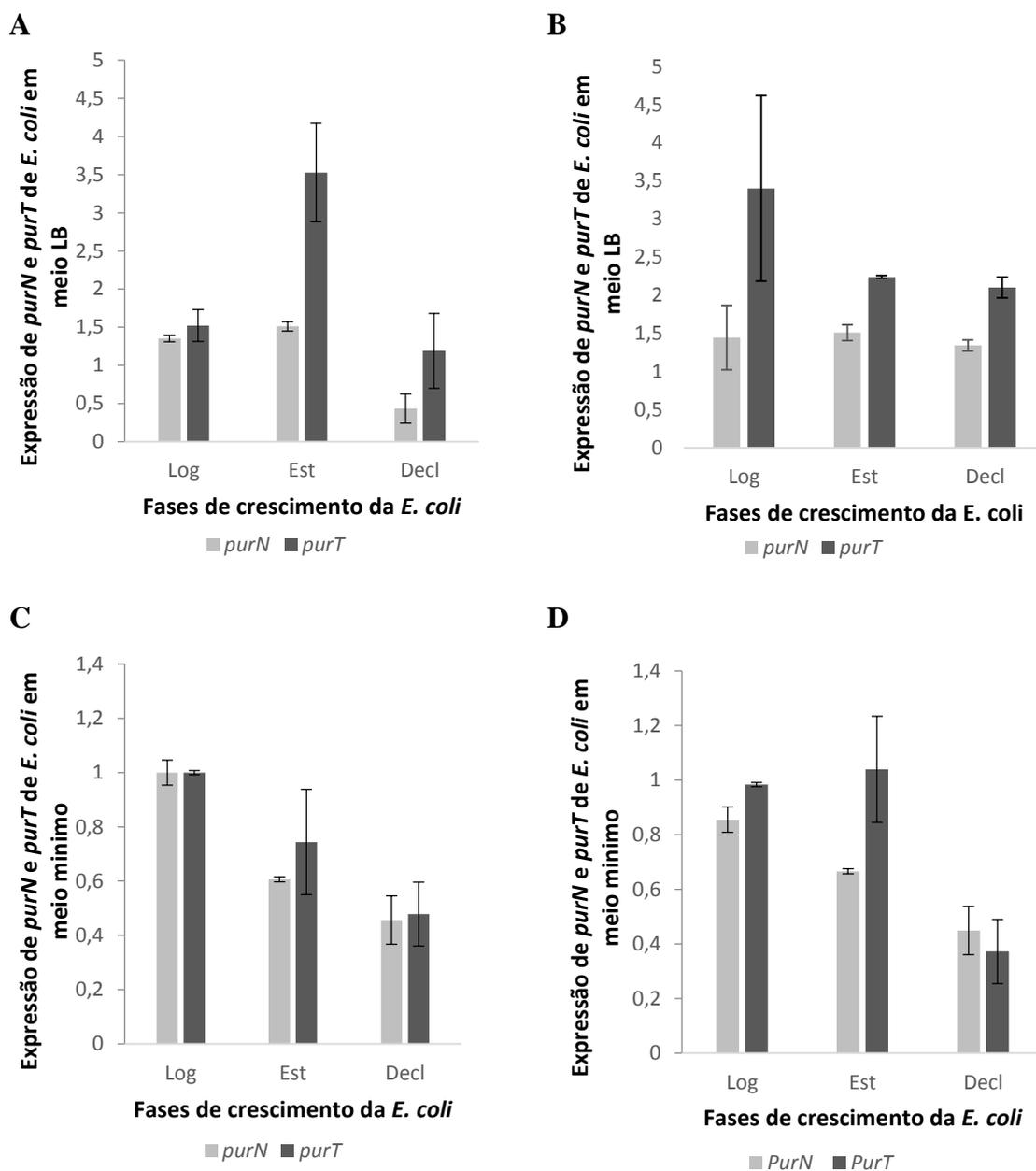


Figura 2.7. Nível de expressão de *purN* e *purT* em *E. coli*. (A e B) *E. coli* cultivada em meio LB. (A) Resultados da repetição 1 e (B) Resultados da repetição 2. (C e D) *E. coli* cultivada em meio mínimo. (C) Resultados da repetição 1 e (D) Resultados da repetição 2. O gene *cysG* foi utilizado para normalizar a quantidade inicial de cDNA das amostras e determinar o nível de expressão nas fases de crescimento exponencial (Log), estacionária (Est) e de declínio (Decl). O nível de expressão foi calculado tomando a fase exponencial de crescimento como referência. As barras de erro indicam os desvios padrão das triplicatas de cada amostra.

As Figuras 2.7A e 2.7B mostram o nível de expressão dos genes *purN* e *purT* em *E. coli* cultivada em meio rico (meio LB). Na Figura 2.7A o nível de expressão de *purN* não varia nas fases log e estacionária e decresce na fase de declínio. Enquanto o nível de expressão de *purT* apresentou um aumento da expressão da fase log para a estacionária, onde a expressão foi mais acentuada, e então reduz na fase de declínio. Na figura 2.7B, que mostra a repetição do tratamento com meio rico (repetição 2), a expressão de *purN* não apresenta uma variação significativa entre as fases, mostrando-se estável durante todas as fases de crescimento da *E. coli*. Por outro lado *purT* apresenta um alto nível de expressão na fase log, mas nas demais fases sua expressão também se mostra estável.

Nos tratamentos com meio mínimo (Figura 2.7C e 2.7D), a expressão de ambos os genes tiveram o maior nível de expressão na fase log e apresentaram uma redução gradual nas fases estacionária e de declínio. No entanto, na Figura B *purT* apresenta nível de expressão mais elevado na fase estacionária. Porém, considerando o desvio padrão o nível de expressão da *purT* na fase estacionária não diferiu da expressão na fase log.

Discussão

Apesar do viés amostral dos genomas completamente sequenciados em alguns grupos, o conjunto de resultados apresentados indicam 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. Os dados analisados destacam dois eventos marcantes na evolução dessa etapa da VBP. Duas enzimas, PurN e PurT, foram selecionadas para desempenhar a mesma função durante a evolução dos procariotos. Além disso, a PurN parece 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 (Figura 3), além disso, na maioria dos dos gêneros/espécies o contexto genômico de *purT* não envolve outros genes da via biossintética de purinas (Figura 4). Estes resultados, associados à filogenia de *purT* e *purK*, podem ser explicados por dois cenários.

Para compreender melhor origem da PurT é preciso investigar a sua relação com outras enzimas com função semelhante. A PurT é homóloga a outra enzima da VBP, a PurK. Essas enzimas pertencem a superfamília de proteínas conhecida como ATP-grasp. As enzimas pertencentes a essa família proteica compartilham semelhanças tanto em nível de sequência de aminoácidos, quanto na estrutura tridimensional (KAPPOCK; EALICK; STUBBE, 2000; LI; FAST; BENKOVIC, 2009; ZHANG; MORAR; EALICK, 2008). Assim, é razoável inferir que os genes que codificam essas enzimas têm uma história evolutiva comum. A homologia entre PurT e PurK é evidente quando observada a sobreposição da estrutura terciária dessas enzimas (Figura 2.8).

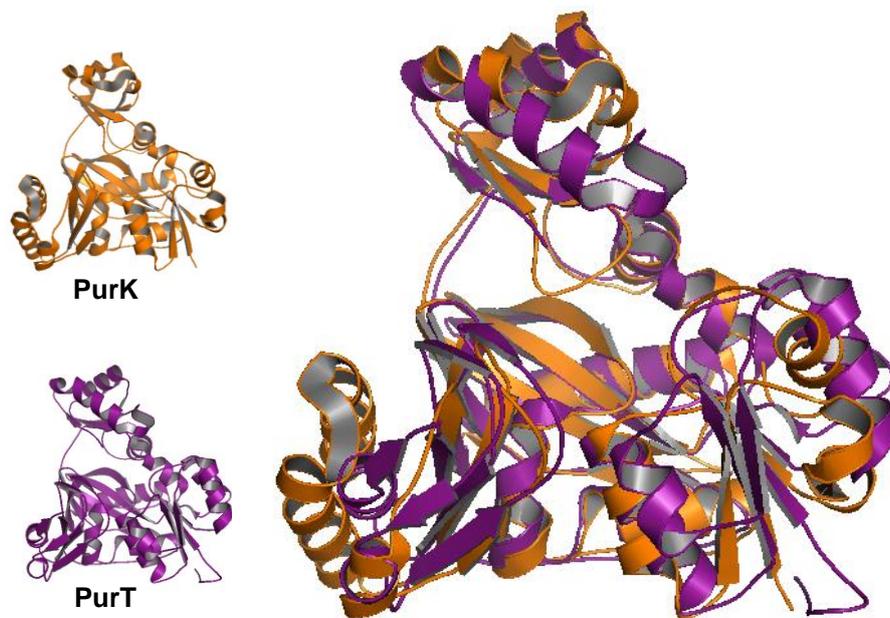


Figura 2.8. Sobreposição da estrutura terciária da PurT (PDB: 1KJ8) e PurK (PDB: 3ETJ) de *Escherichia coli*.

O primeiro cenário proposto para explicar a origem de PurT assume a ocorrência de um evento de recrutamento enzimático (SCHMIDT et al., 2003). Nesse caso, o gene *purT* seria mais recente, e teria surgido por um evento de duplicação do gene *purK* em alguma linhagem procariótica, por exemplo, o ancestral de Gama/Beta-proteobactérias, visto que são grupos irmãos e no qual *purT* é mais frequente que *purN*. 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 2.9). 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 as PurTs seriam descendentes das PurKs e numa árvore filogenética destas enzimas PurT emergiria dentro de um clado de PurK.

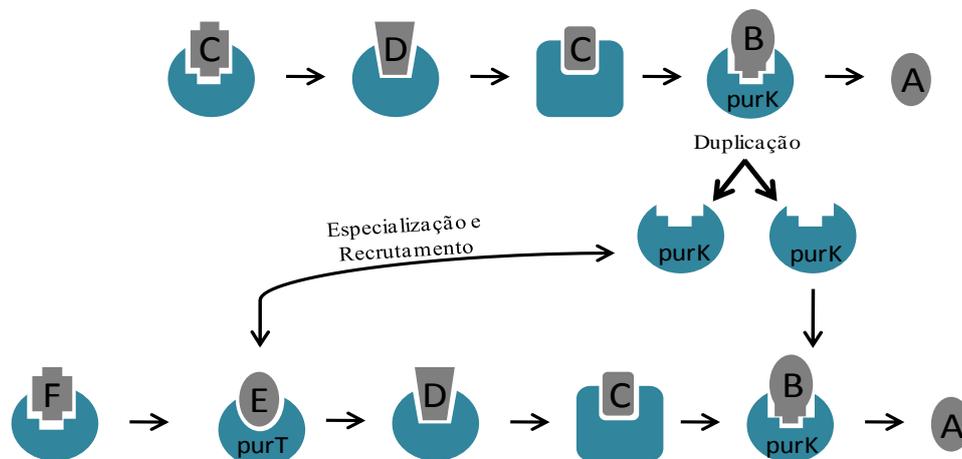


Figura 2.9. Cenário para evolução da PurT baseada na hipótese de recrutamento enzimático. Nesse cenário PurK que já atuava na síntese de purinas e duplicou. Uma das isoformas continuou com a sua função original e a outra, que deu origem a PurT, sofreu especiação e passou a atuar em outra etapa da via 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 2.10) 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. Este cenário é favorecido também pelo fato de que PurT e PurK pertencerem a mesma família de proteínas que inclui outras enzimas da PBP, como PurD e PurC (ZHANG; MORAR; EALICK, 2008).

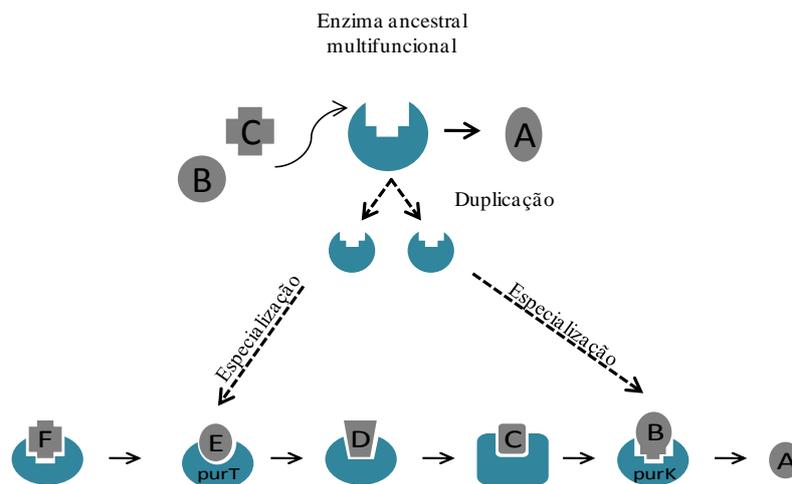


Figura 2.10. Cenário da evolução das enzimas PurT e PurK. Esse cenário sugere que a enzima ancestral de PurT e PurK estava presente no ancestral de todos os seres vivos. Essa enzima dotada de ampla especificidade duplicou e as isoformas dessas enzimas se especificaram em diferentes substratos e passaram a atuar em diferentes etapas da via.

A árvore filogenética da PurT enraizada com PurK corrobora o segundo cenário proposto neste trabalho para origem da PurT, visto que essa enzima não emerge dentro de um clado da PurK o que torna *purT* um gene tão primitivo quanto *purK*.

Doolittle e Brown (1994) laboraram 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. Portanto, a distribuição de PurT em procariotos seria resultante de eventos de TLG e de perda ou manutenção seletiva nas linhagens procarióticas.

As características bioquímicas de PurT e PurN também 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 2.3). 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 2.3. Valores cinéticos das enzimas PurN e PurT de *Escherichia coli*

Enzima	Substrato	K_m (μM)	K_{cat} (s^{-1})	K_{cat}/K_m	Fonte
PurN	GAR	19 ± 4	-	-	(NIXON; BENKOVIC, 2000) (WARREN; MAROLEWSKI; BENKOVIC, 1996)
	N^{10} -formilTHF	84.8 ± 5.9	13.5 ± 0.8	0.27	(NAGY et al., 1995)
PurT	GAR	10.1 ± 0.51	-	-	(MAROLEWSKI et al., 1997)
	Formato	319 ± 15	37.6 ± 0.8	0.84	(LI; FAST; BENKOVIC, 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 dezesseis dos 38 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. Em *E. coli* a dupla mutação *purN-purT* e *purN-purU* tornou a bactéria incapaz de sintetizar o FGAR, no entanto a deleção de apenas um dos genes não bloqueou a síntese do GAR (NAGY; MCCORKLE; ZALKIN, 1993). Isso mostra que apenas a PurT não é suficiente para a síntese do GAR. Nagy e colaboradores (1995) demonstraram que o composto Tetrahydrofolato (THF) é utilizado na síntese de glicina. Portanto, a quebra do FormilTHF, pela PurU regula o balanço de THF, e consequentemente a síntese de glicina. Adicionalmente, a glicina regula a atividade de PurU. Assim, não é arriscado supor que talvez PurT desempenhe uma função adicional relacionada com a manutenção do balanço de THF e FormilTHF. Uma evidencia que apoia essa hipótese é o resultado da comparação dos níveis de expressão de *purN* e *purT* (Figura 2.7). Em todos os ensaios a expressão de *purT* é maior que a de *purN*, isso pode ser um indicio de que esse gene

pode estar sendo recrutado para atuar em outro processo biológico, como o metabolismo de glicina.

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., 1995) 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 não seriam afetadas pelo composto inibidor.

Além disso, na filogenia de PurN foram incluídas sequências de PurU com o objetivo utilizar essa segunda enzima como grupo externo da árvore filogenética. No entanto, a árvore não enraizada mostrou que as PurUs formam um clado dentro das PurNs e não um grupo mais distante (Figura 2.6). O que é um indício de que a PurU se originou a partir da PurN. De fato, Sampei e colaboradores (2013), ao compararem o mecanismo de ação das enzimas PurN e PurU apontam que, por ambas apresentarem similaridades tanto enzimáticas quanto em nível de estrutura terciária e quaternária essas enzimas são evolutivamente relacionadas. Ainda segundo Sampei e colaboradores (2013), a PurU derivou de PurN e durante a sua evolução adquiriu mais um domínio proteico chamado de domínio ACT.

A explicação mais plausível para a predominância de PurN nas linhagens procarióticas é que essa enzima deriva de uma enzima ancestral que teve um papel central na evolução da VBP. De acordo com Sampei e colaboradores (2010) as enzimas PurD, PurT e PurK além de PurU surgiram a partir de uma enzima ancestral da PurN. Essas análises levando em consideração apenas a estrutura e o mecanismo de ação da PurN e das demais enzimas coloca a PurN em um ponto central da evolução de outras enzimas da VBP como as mencionadas anteriormente.

A hipótese levantada por Sampei e colaboradores ajuda a explicar não só a predominância de PurN, mas também evolução de outros genes da VBP. Assim, a PurN seria uma enzima mais antiga que PurU, PurD, PurT e PurK, e teria sido recrutada para atuar na terceira etapa da VBP. No entanto, admitindo-se que a VBP originou-se pelo modelo de retro-evolução (HOROWITZ, 1945) a origem de PurK não seria possível pela hipótese mencionada, uma vez que PurK atua na sexta etapa da VBP e por isso teria surgido primeiro. O mesmo valeria para PurT que, como já foi discutido, provavelmente originou-se a partir da PurK. A origem de PurD não seria contraditória, uma vez que essa enzima catalisa uma reação anterior a reação catalisada pela PurN. No entanto, a homologia entre PurK e PurD é evidente e já foi demonstrada em outros trabalhos (KAPPOCK; EALICK; STUBBE, 2000; LI; FAST; BENKOVIC, 2009; ZHANG; MORAR; EALICK, 2008). Portanto a hipótese de que enzima ancestral similar a PurN deu origem a PurU, PurD, PurT e PurK não parece improvável. Considerando que PurT e PurK atuam em etapas onde outras enzimas estão envolvidas a origem dessas enzimas a partir de PurN é bastante plausível. Dessa forma, é razoável inferir que uma enzima similar a PurN foi ancestral das atuais PurD e PurK, e que posteriormente essa última deu origem a PurT.

De fato as similaridades estruturais entre PurD, PurK, PurT e PurN já havia sido destacada por Zhang e colaboradores (2008). Além disso, esses autores propõem um modelo de divergência das enzimas da família ATP-grasp envolvidas na VBP. Nesse modelo as enzimas PurD, PurK e PurT formam um grupo, onde PurK e PurT são evolutivamente mais próximas e a PurD aparece como uma enzima mais primitiva (Figura 2.9).

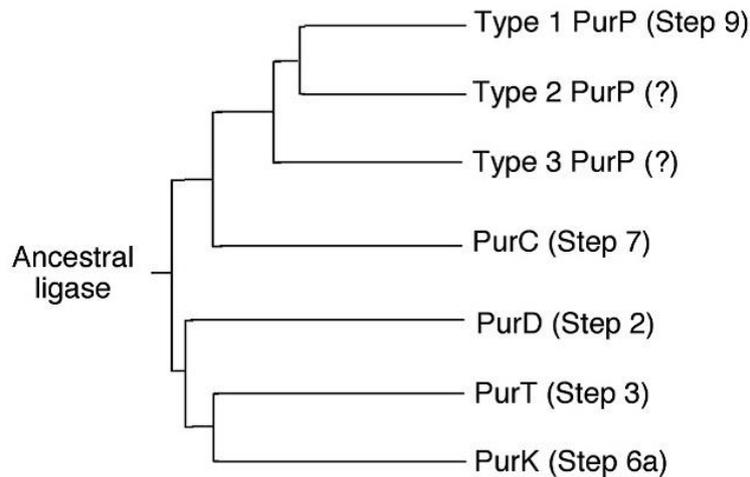


Figura 2.11. Evolução das enzimas da família ATP-grasp proposta por Zhang e colaboradores (2008). Fonte: (ZHANG; MORAR; EALICK, 2008)

A predominância de *purN* nos taxa procarióticos pode ser explicada pelo cenário exposto acima. A presença de *purT* ficou restrita a alguns filos/classes e o mais provável é que a TLG foi em grande parte responsável pela distribuição desse gene nas linhagens procarióticas. Apesar disso, *purT* esta presente em muitas espécies de bactérias causadoras de importantes infecções para animais e plantas (Tabela 2.4). Portanto, esse gene constitui um potencial alvo para fármacos antibacterianos, visto que *purT* não está presente em Eucariotos.

Tabela 2.4. Relação de bactérias patogênicas para animais e plantas que possuem apenas o gene *purT*.

Actinobacteria	<i>Bifidobacterium dentium</i> Bd1
	<i>Bifidobacterium longum</i> subsp. <i>infantis</i> ATCC 15697
	<i>Gardnerella vaginalis</i> ATCC 14019
Bacteroidia	<i>Prevotella denticola</i> F0289
	<i>Prevotella melaninogenica</i> ATCC 25845
Betaproteobacteria	<i>Bordetella pertussis</i> Tohama I
Gammaaproteobacteria	<i>Actinobacillus pleuropneumoniae</i> serovar 3 str. JL03
	<i>Haemophilus ducreyi</i> 35000HP

Conclusão

Os genes *purT* e *purN* não estão igualmente distribuídos nas linhagens procarióticas. A condição ancestral para a VBP é PurN realizando a terceira reação da via. A origem de *purT* ocorreu posteriormente e transferência horizontal de genes foi um fator relevante na distribuição de *purT* nas linhagens procarióticas.

Metodologia

Busca pelos genes *purT* e *purN*

As buscas pelos genes *purN* e *purT* incluíram 1490 genomas completamente sequenciados e anotados, disponíveis na base de dados do NCBI até março de 2011. Esses genomas representam 1165 espécies de 582 gêneros dos Domínios Archaea e Bactéria. As buscas foram feitas no banco de dados de proteínas, com os programas BLASTP e TBLASTN (ALTSCHUL et al., 1997). A matriz BLOSUM62 foi utilizada para as buscas. 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.

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. Aproximadamente 88% dos Filos e 82% das Classes de procariotos reconhecidas (EUZÉBY, 1997) estão representados neste estudo, contudo, não estão igualmente amostrados. Por isso, os resultados foram analisados em nível gênero e de espécie para minimizar o viés amostral dos

genomas completamente sequenciados. 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.

Análise de anticorrelação

A análise de anticorrelação dos genes *purT* e *purN* foi realizada de acordo com a metodologia proposta por Morett et al (MORETT et al., 2003). Para cada genoma presente na base de dados foram avaliadas a presença ou ausência dos genes, e a pontuação foi determinada como segue. Para os organismos que possuem a via completa, (i) foi atribuída pontuação 1 quando apenas um dos genes esta presente; (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) aqueles que não apresentam nenhum dos genes é atribuído valor 0; 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.

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 casos em que eles estavam flanqueados por genes da VBP.

Alinhamento múltiplo e filogenia dos genes *purT* e *purN*

As sequências de nucleotídeos homólogas aos genes *purN* e *purT* foram recuperadas na base de dados do NCBI. As sequências de nucleotídeos foram traduzidas para aminoácidos foram alinhadas códon a códon no programa MEGA6 (TAMURA et al., 2013). Os alinhamentos múltiplos foram feitos no servidor GUIDANCE usando o algoritmo MAFFT (PENN et al., 2010). Os alinhamentos foram editados com base nas estruturas terciária das proteínas, disponíveis na base de dados do PDB (<http://www.rcsb.org/pdb/home/home.do>). As sequências incompletas ou com erros de anotação foram excluídas. A visualização e sobreposição das estruturas terciárias das proteínas foram feitas no programa PyMOL (DeLANO, 2002).

A partir das sequências de aminoácidos foram construídas as árvores filogenéticas de Máxima Verossimilhança utilizando o modelo LG (LeC Gascuel) (LE; GASCUEL, 2008) empregando o PhyML (GUINDON et al., 2010). A confiança dos ramos foi calculada pelo método aLRT (Approximate Likelihood-Ratio Test) (ANISIMOVA; GASCUEL, 2006).

Condições de cultivo

Para as análises de expressão gênica de *purT* e *purN* foi utilizada a bactéria *E. coli* DH5 α . Para preparar o microrganismo para a extração de RNA a *E. coli* DH5 α foi cultivada em meio Luria Bertani (LB) e meio mínimo M9 sob agitação a 150 rpm a 37°C. O padrão de crescimento da bactéria foi monitorado num período de 24h por meio da técnica de contagem de unidades formadoras de colônias(UFC) a cada 2 horas para o meio LB e a cada 3 horas para as bactérias cultivadas em meio mínimo M9.

Extração de RNA e síntese de cDNA

O RNA total das células foi extraído em três fases do crescimento bacteriano: fase de crescimento exponencial, fase estacionária e fase de declínio. A extração de RNA e

tratamento com DNase foi realizada utilizando-se o kit SV Total RNA Isolation System (PROMEGA) seguindo as especificações do fabricante. Após extração as amostras de RNA foram armazenadas a -80°C até a etapa de síntese de cDNA. Para a síntese de cDNA foi realizada a transcrição reversa (RT). Utilizou-se o kit High Capacity cDNA Reverse Transcription (Applied Biosystems) seguindo as recomendações do fabricante. O produto da reação final de 20µL foi estocado a -20°C.

PCR em tempo real

A expressão dos genes *purT* e *purN* nas diferentes fases de crescimento da bactéria *Escherichia coli* foi avaliada usando a comparação por PCR em tempo real. Para a realização dessa reação foram utilizados 2µl de cDNA, 0,4µl de cada *primer*, 0,4µl de referência passiva (ROX Low - 50X) (Kapabiosystems, Boston, Massachusetts, USA), 10µl de KAPA SYBR® FAST qPCR Master Mix, o qual contem KAPA SYBR® DNA Polymerase, SYBR® Green I Dye, MgCl₂, dNTPs e tampão otimizado dos componentes (Kapabiosystems), e 6,8µl de H₂O livre de nucleases. Os primers usados nesse estudo estão listados na Tabela 5. Os *primers* dos genes *purT* e *purN* foram desenhados usando o programa Gene Runner versão 4.0.7 Beta. Os primers do gene *cysG* foi escolhido como controle interno para normalizar as reações de PCR em tempo real pois apresentou eficiência mais próximo de 100%. Esse gene essencial codifica uma enzima bifuncional que atua na síntese de siroheme (Tabela 5). A eficiência e especificidade de cada *primer* foi determinada por uma curva padrão e pela curva de *melting*, respectivamente, usando KAPA SYBR® FAST qPCR. As reações foram feitas com ABI PRISM® 7500 Sequence Detection System (Applied Biosystems) em placas de PCR de 96 poços com adesivos óticos. Para testar a eficiência dos primers foi utilizado o seguinte programa: pré-incubação a 95°C por 3 minutos; seguido por 40 ciclos, cada um consistindo de

desnaturação a 95°C por 40 segundos, anelamento a 61°C por 40 segundos e extensão a 72°C por 40 segundos.

Tabela 2.5. Sequências dos oligonucleotídeos utilizados para a reação de PCR em Tempo Real

Genes	Sequência dos <i>Primers</i> (5'-3')	Tamanho do Fragmento Amplificado (pb)	Eficiência da PCR (%)
<i>purT</i>	F: CCATGTCATTAATATGCTTGAT R: ACAACATTCAGTCCTTCCTCTT	138	104,4
<i>purN</i>	F: GGTTTTATGCGCATTCTCAG R: CTTTCATCGCCATTTTCCAG	133	105
<i>cysG*</i>	F: TTGTCGGCGGTGGTGATGTC R: ATGCGGTGAACTGTGGAATAAACG	105	105,9
<i>hcaT*</i>	F: GCTGCTCGGCTTCTCATCC R: CCAACCACGCTGACCAACC	86	79,9
<i>idnT*</i>	F: CTGTTTAGCGAAGAGGAGATGC R: ACAACGGCGGCGATAGC	90	77,6

*Genes utilizados como controle endógeno. Fonte: ZHOU et al., 2011.

Os níveis relativos de mRNA foi determinado pela quantificação relativa utilizando a expressão $2^{-\Delta\Delta Ct}$. Para cada tratamento é detectado o valor de Ct (cycle threshold), tanto para o gene alvo quanto para o controle endógeno. O valor de Ct indica ponto em que o sinal de amplificação é detectado. O valor de Ct do controle endógeno é subtraído do valor de Ct do gene alvo, obtendo-se o valor de ΔCt . O valor de ΔCt dos tratamentos é subtraído do valor de ΔCt da amostra controle (calibrador), tendo o valor de $\Delta\Delta Ct$. Este valor é utilizado na fórmula do nível de expressão, onde o número 2 representa a somatória da eficiência do gene alvo e do controle endógeno, considerando que ambos os genes possuem 100% de eficiência.

Referências

ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic acids research**, v. 25, n. 17, p. 3389–402, 1 set. 1997.

ANISIMOVA, M.; GASCUEL, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. **Systematic biology**, v. 55, n. 4, p. 539–52, ago. 2006.

DeLANO, W. L. The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA. 2002.

DOOLITTLE, W. F.; BROWN, J. R. Tempo, mode, the progenote, and the universal root. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, p. 6721–6728, 1994.

EUZÉBY, J. P. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. **International journal of systematic bacteriology**, v. 47, n. 2, p. 590–2, abr. 1997.

GUINDON, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. **Systematic biology**, v. 59, n. 3, p. 307–21, maio. 2010.

HOROWITZ, N. H. On the evolution of biochemical syntheses. **Proceedings of the National Academy of Sciences of the United States of America**, v. 31, n. 6, p. 153–157, 1945.

KAPPOCK, T. J.; EALICK, S. E.; STUBBE, J. Modular evolution of the purine biosynthetic pathway. **Current opinion in chemical biology**, v. 4, n. 5, p. 567–72, out. 2000.

LE, S. Q.; GASCUEL, O. An improved general amino acid replacement matrix. **Molecular biology and evolution**, v. 25, n. 7, p. 1307–20, jul. 2008.

LI, H.; FAST, W.; BENKOVIC, S. J. Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. **Protein science : a publication of the Protein Society**, v. 18, n. 5, p. 881–92, maio. 2009.

MAROLEWSKI, A. E. et al. Formyl Phosphate: A Proposed Intermediate in the Reaction Catalyzed by Escherichia coli PurT GAR Transformylase. **Biochemistry**, v. 36, n. 22, p. 6709–6716, 1997.

MORETT, E. et al. Systematic discovery of analogous enzymes in thiamin biosynthesis. **Nature biotechnology**, v. 21, n. 7, p. 790–5, jul. 2003.

NAGY, P. L. et al. Formyltetrahydrofolate hydrolase, a regulatory enzyme that functions to balance pools of tetrahydrofolate and one-carbon tetrahydrofolate adducts in Escherichia coli. **Journal of bacteriology**, v. 177, n. 5, p. 1292–8, mar. 1995.

NAGY, P. L.; MCCORKLE, G. M.; ZALKIN, H. purU, a source of formate for purT-dependent phosphoribosyl-N-formylglycinamide synthesis. **Journal of bacteriology**, v. 175, n. 21, p. 7066–73, nov. 1993.

NIXON, A E.; BENKOVIC, S. J. Improvement in the efficiency of formyl transfer of a GAR transformylase hybrid enzyme. **Protein engineering**, v. 13, n. 5, p. 323–7, maio. 2000.

PENN, O. et al. GUIDANCE: a web server for assessing alignment confidence scores. **Nucleic acids research**, v. 38, n. Web Server issue, p. W23–8, jul. 2010.

RODIONOV, D. A et al. Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. **The Journal of biological chemistry**, v. 277, n. 50, p. 48949–59, 13 dez. 2002.

SAMPEI, G. et al. Structures and reaction mechanisms of the two related enzymes, PurN and PurU. **Journal of biochemistry**, v. 154, n. 6, p. 569–79, dez. 2013.

SAMPEI, G.-I. et al. Crystal structures of glycinamide ribonucleotide synthetase, PurD, from thermophilic eubacteria. **Journal of biochemistry**, v. 148, n. 4, p. 429–38, out. 2010.

SCHMIDT, S. et al. Metabolites: a helping hand for pathway evolution? **Trends in biochemical sciences**, v. 28, n. 6, p. 336–41, jun. 2003.

TAMURA, K. et al. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. **Molecular biology and evolution**, v. 30, n. 12, p. 2725–9, dez. 2013.

THORNDIKE, J. et al. Inhibition of Glycinamide Ribonucleotide Formyltransferase and Other Folate Enzymes by Homofolate Polyglutamates in Human Lymphoma and Murine Leukemia Cell Extracts Inhibition of Glycinamide Ribonucleotide Formyltransferase and Other Folate Enzymes by Hom. p. 158–163, 1989.

WARREN, M. S.; MAROLEWSKI, A. E.; BENKOVIC, S. J. A Rapid Screen of Active Site Mutants in Glycinamide Ribonucleotide Transformylase. **Biochemistry**, v. 35, n. 27, p. 8855–8862, 1996.

WELIN, M. et al. Structural studies of tri-functional human GART. **Nucleic acids research**, v. 38, n. 20, p. 7308–19, nov. 2010.

ZHANG, Y.; MORAR, M.; EALICK, S. E. Structural biology of the purine biosynthetic pathway. **Cellular and molecular life sciences : CMLS**, v. 65, n. 23, p. 3699–724, nov. 2008.

Capítulo 3

Diversidade estrutural e evolução da sexta etapa da Via
Biossintética de Purinas em procariotos

Esse capítulo está formatado pelas normas da revista *Biology Direct*

Lista de Figuras

- Figura 3.1. Sexta reação da Via Biossintética de Purinas. Fonte: ZHANG; MORAR; EALICK, 2008. 74
- Figura 3.2. Genômica comparativa e contexto genômico dos genes *purE* e *purK* em procariotos (A) Distribuição e padrão de ocorrência dos genes *purE* e *purK* em procariotos. As colunas referentes a genômica comparativa indicam os possíveis padrões de ocorrência dos genes *purE* e *purK* nos taxa de procariotos. Os valores de genômica comparativa são apresentados para gênero/espécie. As células em amarelo indicam presença do gene no genoma e as células em branco ausência. (B) Contexto genômico dos genes *purE* e *purK*. Os genes *purE* e *purK* estão representados por caixas que indicam o tipo de contexto genômico encontrado nos genomas. 77
- Figura 3.3. Contexto genômico dos gêneros e espécies que possuem os genes *purE* e *purK*, ou apenas *purE*. (A) Contexto genômico dos gêneros/espécies que possuem apenas *purE*, nesses organismos esse gene pode ser encontrado com outros genes da biossíntese de purinas (104/157 gêneros/espécies) ou em um contexto sem outros genes da via (74/141 gêneros/espécies). (B) Contexto genômico dos gêneros/espécies que possuem *purE* e *purK*. Nesses organismos os genes *purE* e *purK* podem estar em contextos genômicos diferentes (44/76 gêneros/espécies), ou juntos no mesmo contexto. Quando esses genes estão no mesmo contexto genômico eles podem ser encontrados com outros genes da VBP (40/141 gêneros/espécies), ou sem outros genes da VBP (298/535 gêneros/espécies)..... 90
- Figura 3.4. Filogenia de *PurK*. Árvore filogenética não enraizada estimada a partir das sequências de aminoácidos da *PurK* pelo método da Máxima Verossimilhança. O clado contendo as *PurKs* que estão no arranjo *purK/purE* esta destacado em vermelho, o grupo que contem as *PurKs* no arranjo *purE/purK* esta em preto. A árvore filogenética foi criada usando o PhyML 3.0 e editada com o MEGA6 (TAMURA et al., 2013) 92
- Figura 3.5. Filogenia de *PurK*. Árvore filogenética não enraizada estimada a partir das sequências de aminoácidos da *PurK* e *PurT*. O clado contendo as *PurKs* que estão no arranjo *purK/purE* esta destacado em vermelho, o grupo que contem as *PurKs* no arranjo *purE/purK* esta em preto. O grupo de *PurTs* esta apresentado em verde. 93
- Figura 3.6. Filogenia de *PurE*. Árvore filogenética não enraizada estimada a partir das sequências de aminoácidos da *PurE* pelo método da Máxima Verossimilhança. As *PurEs* das espécies que possuem *purK* nos contextos *purK/purE* e *purE/purK* são mostradas em vermelho e rosa, respectivamente. Em conjunto, essas *PurEs* formam o Grupo I. As *PurEs* das espécies que não possuem *PurK* estão destacadas em verde (Grupo II). A árvore filogenética foi criada usando o PhyML 3.0 e editada com o MEGA6 (TAMURA et al., 2013). 95

Lista de Tabelas

- Tabela 3.1. Duplicações e fusões envolvendo os genes *purE* e *purK*. 79

Resumo

Introdução

A Via Biossintética de Purinas esta dentre as vias biossintéticas mais primitivas. Por isso o conhecimento sobre a diversidade e evolução dessa via pode ajudar a compreender importantes aspectos que permearam a evolução de procariotos e eucariotos. 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.

Resultados

Nós usamos análise genômicas e a filogenia molecular para investigar os padrões de distribuição dos genes *purE* e *purK* nas linhagens procarióticas de modo a ampliar nossa compreensão sobre a evolução destes genes e desta etapa da VBP. Os genes *purE* e *purK* foram encontrados na maioria das espécies analisadas nesse estudo. Em 89% das espécies analisadas nesse estudo esses genes estão no mesmo contexto genômico. Nós encontramos uma quantidade significativa de espécies que não possuem *purK* no seu genoma, indicando que essas espécies possuem uma PurE classe II. A filogenia de PurE mostrou que os organismos que não possuem *purK* formam um grupo distinto daquelas que possuem esse gene. Além disso, a PurE dos organismos que não possuem *purK* agrupam com PurEs classe II caracterizadas funcionalmente. A filogenia de *purK* mostrou a formação de dois clados distintos. Em um clado prevalece a *purK* com o arranjo genômico *purK/purE*, enquanto no outro prevalecem espécies cujo arranjo genômico é *purE/purK*.

Conclusões

Os resultados lançam uma nova perspectiva sobre a diversidade e evolução da sexta etapa da biossíntese de purinas em procariotos, principalmente por demonstrar a ampla distribuição de PurE classe II em procariotos e levanta a questão de qual teria sido a condição ancestral da terceira etapa da VBP em procariotos.

Abstract

The Purine Biosynthetic Pathway produces inosine monophosphate, the precursor of molecules of biological importance such as nucleotides, coenzyme A and Phytohormones. In the sixth step of the purine biosynthetic pathway, the compound aminoimidazole ribonucleotide (AIR) is converted to carboxyaminoimidazole ribonucleotide (CAIR) by two distinct mechanisms. In one of these mechanisms, two enzymes, PurK and PurE I, perform the carboxylation reaction that results in CAIR. PurK adds a carboxyl group from the HCO_3^- to AIR and produces the N^5 -CAIR. PurE I, a mutase, transfer the carboxyl group from N^5 -CAIR to the C4 position, resulting in CAIR. The same reaction can be performed by the PurE II, which uses CO_2 as a source of carboxyl group to synthesize CAIR. The existence of two reaction mechanisms in the third and sixth stages of PBP refers to questions about the evolution of these steps, as well as the origin and distribution of these genes in prokaryotic lineages.

Results

We analyzed the genomic and molecular phylogeny to investigate the distribution patterns of *purE* and *purK* genes in prokaryotic lineages in order to expand our understanding of the evolution of these genes and this stage of PBP. The *purE* and *purK* genes were found in most of the species analyzed in this study. In 89% of the species analyzed in this study, these genes are in the same genomic context. We found a significant number of species that do not have PurK in their genome, indicating that these species have a PurE class II. The phylogeny of *purE* revealed that organisms without *purK* form a clade containing Archaeal and bacterial *purEs* and *purE IIs* that are functionally characterized. The phylogenetic tree of *purK* showed two distinct clades, both containing Archaeal and bacterial *purKs*, in one clade prevails the genomic context *purE/purK*, while in the other prevails the genomic context *purK/purE*.

Conclusions

These data give a new perspective on the comprehension on the sixth stage of purine biosynthetic pathway, as they indicate a larger distribution of *purE II* in prokaryotes than previously reported in the literature and lead to a question about the ancestral condition of the sixth step in the PBP.

Introdução

A Via Biossintética de Purinas (VBP) tem como produto final a Inosina Monofosfato (IMP), precursora de compostos purínicos e derivativos que desempenham funções estruturais, carreadoras de energia, sinalizadoras e metabólicas essenciais.

São conhecidos treze genes que codificam enzimas que catalisam reações na síntese de IMP, dentre eles, *purF*, *purD*, *purL*, *purM*, *purC* e *purB* são ubíquos nas espécies procarióticas e eucarióticas. Em contraste, os genes *purN/purT* e *purH/purP/purO* codificam proteínas análogas não homólogas e possuem distribuição limitada a determinados grupos de organismos. Por exemplo, os genes *purP* e *purO* são considerados assinaturas do Domínio Archaea (OWNBY; XU; WHITE, 2005).

Dependendo do conjunto gênico presente no genoma do organismo a síntese da IMP pode ocorrer em uma série de 10 ou 11 reações enzimáticas (ZHANG; MORAR; EALICK, 2008). Esta diversidade estrutural é observada tanto nas linhagens procarióticas como nas linhagens eucarióticas e tem sido explorada na produção de fármacos antimicrobianos (FIRESTINE et al., 2009).

Um exemplo da diversidade estrutural da VBP entre os seres vivos é a reação enzimática que ocorre na sexta etapa da via, a conversão do intermediário Aminoimidazole ribonucleotide (AIR) em 4-Carboxi-5-aminoimidazole ribonucleotide (CAIR) (Figura 3.1) (LI; FAST; BENKOVIC, 2009; MEYER et al., 1999; MUELLER et al., 1994). Em procariotos, fungos e plantas a conversão de AIR em CAIR é realizada por duas enzimas, a PurK e a PurE classe I (FIRESTINE et al., 1998; MEYER et al., 1992; THODEN et al., 2010). A PurK, utiliza bicarbonato (HCO_3^-) e ATP para adicionar um grupo carboxila no grupo amino na posição cinco do AIR produzindo o N^5 -CAIR (MUELLER et al., 1994). Em seguida, a PurE classe I, usando uma atividade de mutase, transfere o carbono adicionado na reação anterior para a posição C4 do anel purínico em

construção, convertendo o N⁵-CAIR em CAIR, sem consumo de ATP (MUELLER et al. 1994).

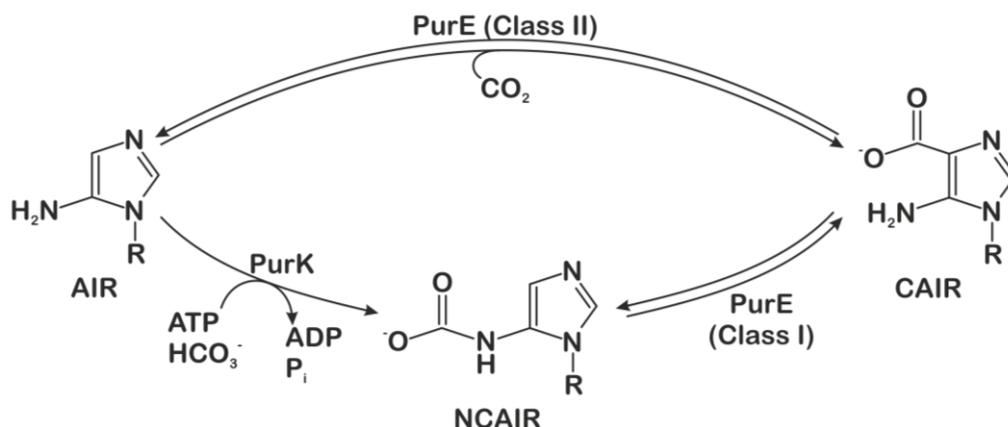


Figura 3.1. Sexta reação da Via Biossintética de Purinas. Fonte: ZHANG; MORAR; EALICK, 2008.

Em animais a conversão de AIR em CAIR é catalisada pela enzima PurE classe II, que adiciona o grupo carboxila diretamente na posição C4 do AIR sem consumo de ATP. A PurE classe II, em contraste com a PurE classe I, usa o CO₂ como fonte de carbono (FIRESTINE et al., 1994).

Apesar de catalisarem reações distintas as PurEs class I and II são proteínas homólogas e compartilham alto nível de similaridade estrutural, desde o nível primário até o quaternário (BOYLE et al., 2005; LI et al., 2007; MATHEWS et al., 1999; SCHAEFER et al., 2009; SCHWARZENBACHER et al., 2004). Como o alto nível de similaridade destas enzimas se estende também ao sítio ativo a questão como enzimas tão similares catalisam reações tão distintas tem sido abordada em inúmeros trabalhos desde a década de 1990, contudo esta questão continua a ser elucidada (DANDANELL; SORENSEN, 1997; DRAZEK et al., 1995; LI; FAST; BENKOVIC, 2009; MEYER et al., 1992; ZHANG; MORAR; EALICK, 2008).

Embora os produtos dos genes *purE* e *purK* sejam apontados como alvos para fármacos antibacterianos e antifúngicos (FIRESTINE et al., 1998, 2009), nosso conhecimento atual sobre a diversidade e distribuição destes genes está limitado a poucos organismos modelos. Por exemplo, a PurE classe II é descrita na literatura como uma enzima presente somente em animais (FIRESTINE; DAVISSON, 1994; LI et al., 2007). No entanto, Hamilton e Reeve (1985) demonstraram que as PurEs de *Methanothermobacter thermautotrophicus* e *Methanobrevibacter smithii* são capazes de complementar um mutante *purK* de *Escherichia coli*. Estas espécies de Archaea não possuem *purK* e os resultados dos ensaios de complementação sugerem que elas possuem uma PurE classe II. O mesmo já foi proposto para a PurE de *Archaeoglobus fulgidus* (MATHEWS et al., 1999). Tranchimand e colaboradores (2011) relataram que a *purE* de *Treponema denticola*, uma bactéria Gram negativa, tem atividade de carboxilase, típica da PurE classe II. Estes resultados sugerem que a PurE classe II não é exclusiva de metazoários e indicam que nossa compreensão sobre a distribuição e diversidade de *purK* e *purE* nas linhagens procarióticas pode estar subestimada.

O conhecimento sobre a distribuição e diversidade da PurE e PurK nas linhagens procarióticas é fundamental para o uso eficiente destas enzimas como alvos para fármacos antimicrobianos. Além disso, o metabolismo de purinas é basal na evolução metabólica (CAETANO-ANOLLÉS; KIM; MITTENTHAL, 2007) e a elucidação dos eventos evolutivos que nortearam a origem e diversificação da biossíntese de purinas pode fornecer novas pistas sobre a evolução precoce dos seres vivos.

Neste trabalho nós usamos análises genômicas e filogenia molecular para expandir nossa compreensão sobre a evolução e distribuição da diversidade estrutural da sexta etapa da VBP nos domínios Archaea e Bacteria. Os dados aqui apresentados dão uma nova perspectiva sobre a distribuição da diversidade estrutural da sexta etapa da VBP em

procariotos e levanta a questão de qual seria a origem e condição ancestral em sexto passo da VBP em procariotos.

Resultados

Distribuição de *purE* e *purK*

As espécies cujos genomas foram analisados neste estudo representam 88% dos Filos e 82% das Classes de procariotos reconhecidas (Tabela 1 – Capítulo 2). Portanto, os resultados apresentados aqui fornecem uma ampla compreensão da distribuição dos genes *purK* e *purE* nas linhagens procarióticas. Como estes grupos taxonômicos não foram igualmente amostrados os resultados das análises de genômica comparativa e contexto genômico são apresentados em nível de gênero e de espécie para minimizar um eventual viés amostral nos genomas completamente sequenciados.

O gene *purE* está presente no genoma de 547/1048 gêneros/espécies de microrganismos enquanto que o gene *purK* possui uma distribuição mais limitada, sendo encontrado em 376/750 gêneros/espécies (Figura 3.2). Todas as espécies que possuem o gene *purK* também possuem o gene *purE*, entretanto, em 171/298 gêneros/espécies apenas o gene *purE* foi encontrado (Figura 3.2). Não foram observadas espécies que possuem apenas o gene *purK*.

Filos/Classes	(A) Genômica Comparativa			(B) Contexto genômico		
	E	E		purE purK	purK purE	purE ... purK
	K					
Crenarchaeota	7/18	4/4	8/9	7/16		1/2
Euryarchaeota	17/21	28/51	1/1	1/1	11/13	5/7
Korarchaeota		1/1				
Thaumarchaeota	1/1					1/1
Nanoarchaeota			1/1			
Firmicutes Bacilli	22/98	1/1	3/8	22/98		
Firmicutes Clostridia	11/23	34/68	1/3	2/2	9/21	
Firmicutes Erysipelotrichia		1/1	1/1			
Firmicutes Negativicutes		5/5	1/1			
Tenericutes			7/29			
Actinobacteria	59/113	9/14	1/1	4/9	55/102	1/2
Cyanobacteria	12/28				1/1	11/27
Aquificae	5/6	3/3			1/2	4/4
Chloroflexi	4/7	3/7	1/1	2/3	2/4	
Deinococcus-Thermus	5/13	1/1		4/11		1/2
Acidobacteria	5/5	1/1		4/4		1/1
Fusobacteria	1/1	4/4		1/1		
Planctomycetes	4/5			4/5		
Thermotogae	2/7	4/5	1/1		2/7	
Bacteroidetes	33/39	9/20	2/2		26/28	9/11
Chlorobi	1/1	4/9			1/1	
Verrucomicrobia	2/2	2/2		2/2		
Dictyoglomi		1/2				
Elusimicrobia		1/1				
Nitrospira	1/1	1/1		1/1		
Gemmatimonadetes	1/1			1/1		
Synergistetes		3/3				
Deferribacteres		4/4				
Chrysiogenetes		1/1				
Fibrobacteres		1/1				
Thermodesulfobacteria	1/1	1/1			1/1	
Chlamydiae			6/12			
Spirochaetes	2/4	4/14	2/10	2/4		
Deltaproteobacteria	6/7	19/35		6/7		
Epsilonproteobacteria		11/22	1/2			
Alphaproteobacteria	59/110	1/1	7/24	55/100		4/10
Betaproteobacteria	40/79	2/2	2/2	38/75		2/4
Gammaproteobacteria	75/159	8/14	7/8	71/152	2/2	4/5

Figura 3.2. Genômica comparativa e contexto genômico dos genes *purE* e *purK* em procariotos (A) Distribuição e padrão de ocorrência dos genes *purE* e *purK* em procariotos. As colunas referentes a genômica comparativa indicam os possíveis padrões de ocorrência dos genes *purE* e *purK* nos taxa de procariotos. Os valores de genômica comparativa são apresentados para gênero/espécie. As células em amarelo indicam presença do gene no genoma e as células em branco ausência. (B) Contexto gênomico

dos genes *purE* e *purK*. Os genes *purE* e *purK* estão representados por caixas que indicam o tipo de contexto genômico encontrado nos genomas.

Na maioria dos genomas analisados foram encontrados apenas uma cópia dos genes *purK* e *purE*. No entanto, em algumas espécies foram encontradas mais de uma cópia destes genes ou eles formam arranjos gênicos atípicos (Tabela 3.1). Tais casos parecem ser linhagens específicas e podem ser usados como assinaturas do grupo em estudos de diversidade microbiana.

Por exemplo, a *purE* presente no genomas das espécies de Archaeas pertencentes à ordem Methanobacteriales contém dois domínios PurE, sugerindo que este gene se originou de um evento de duplicação seguido de fusão gênica no ancestral dessa família (Tabela 1). O domínio N-terminal é mais conservado que o domínio C-terminal e essas espécies não possuem o gene *purK*, portanto, Brown et al. (2011) especularam que ele poderia exercer a função da PurK nestes organismos, entretanto isto ainda não foi testado. A fusão de *purE* e *purD* encontrada em espécies das famílias *Desulfarculaceae*, *Desulfobacteraceae* e *Desulfobulbaceae* da Classe Delta-proteobacteria também parece ser específico desta famílias (Tabela 3.1). As Delta-proteobacterias desempenham importantes funções na recuperação de áreas degradadas por resíduos poluentes, sendo portanto organismos com potencial para atuação em processos de biorremediação (ACOSTA-GONZÁLEZ; ROSSELLÓ-MÓRA; MARQUÉS, 2013; SUN et al., 2010). Desse modo, a fusão *purE* e *purD* parece ser um marcador em potencial para a presença desse grupo de bactéria em estudos de diversidade microbiana e bioprospecção.

Tabela 3.1. Duplicações e fusões envolvendo os genes *purE* e *purK*.

Filo/Classe	Espécie	Duplicações e Fusões	
Fusões			
GI			
Actinobacteria	<i>Kocuria rhizophila</i>		184201397
Deltaproteobacteria	<i>Desulfarculus baarsii</i>		302341612
	<i>Desulfatibacillum alkenivorans</i>		218779902
	<i>Desulfobacterium autotrophicum</i>		224368044
	<i>Desulfococcus oleovorans</i>		158521438
	<i>Desulfobulbus propionicus</i>		320353910
	<i>Desulfotalea psychrophila</i>		51246040
	Euryarchaeota	<i>Methanobrevibacter smithii</i>	
<i>Methanosphaera stadtmanae</i>			84490261
<i>Methanothermobacter thermautotrophicus</i>			15679392
Duplicações			
<i>Duplicações envolvendo purK</i>			
		<i>purK GI</i>	<i>purK GI</i>
Firmicutes Bacilli	<i>Carnobacterium sp. 17-4</i>		328958672
	<i>Enterococcus faecalis</i>		
	<i>Lactobacillus buchneri</i>		
	<i>Lactobacillus casei ATCC 334</i>		
	<i>Lactobacillus casei BL23</i>		
	<i>Lactobacillus fermentum</i>		
	<i>Lactobacillus plantarum JDM1</i>		
	<i>Lactobacillus plantarum WCFS1</i>		
	<i>Lactobacillus reuteri DSM 20016</i>		
	<i>Lactobacillus reuteri JCM 1112</i>		
	<i>Lactobacillus rhamnosus GG</i>		
	<i>Lactobacillus rhamnosus HN001</i>		
	<i>Lactobacillus rhamnosus Lc 705</i>		
	<i>Lactobacillus sakei</i>		
	<i>Lactobacillus salivarius</i>		
	<i>Pediococcus pentosaceus</i>		
<i>Duplicações envolvendo purE</i>			
		<i>purE GI</i>	<i>purE GI</i>
Firmicutes Clostridia	<i>Clostridium cellulovorans</i>		
	<i>Thermoanaerobacter pseudethanolicus</i>		
	<i>Thermoanaerobacter sp. X514</i>		
	<i>Thermoanaerobacter sp. X513</i>		
Dictyoglomi	<i>Dictyoglomus thermophilum</i>		
Chrysiogenetes	<i>Desulfurispirillum indicum</i>		
Deltaproteobacteria	<i>Desulfomicrobium baculatum</i>		
	<i>Pelobacter carbinolicus</i>		
Betaproteobacteria	<i>Sideroxydans lithotrophicus</i>		
Gammaproteobacteria	<i>Aeromonas hydrophila</i>		
	<i>Aeromonas salmonicida</i>		
	<i>Aeromonas veronii</i>		
	<i>Tolomonas auensis</i>		
	<i>Ferrimonas balearica</i>		
	<i>Psychromonas ingrahamii</i>		
	<i>Kangiella koreensis</i>		
	<i>Shewanella loihica</i>		
	Euryarchaeota	<i>Methanosarcina mazei</i>	

Algumas espécies de bactérias pertencentes à Classe Bacilli possuem uma cópia de *purK* em contexto genômico com *purE*, e uma segunda cópia que codifica uma *purK* mais divergente quando comparadas as demais *PurKs*. Esta *purK* divergente foi

encontrada no genoma da única espécie da família *Carnobacteriaceae* amostrada, no genoma de uma das cinco espécies amostradas da família *Enterococcaceae* e no genoma de nove das vinte espécies amostradas da família *Lactobacillaceae* (Tabela 1). Na maioria dos casos esta segunda cópia de *purK* não está no contexto genômico com genes *pur*, exceto, nos genomas de *Lactobacillus rhamnosus*, *Lactobacillus sakei* e *Carnobacterium* sp. 17-4 (Tabela 1).

As buscas BLAST revelaram a existência de um gene que codifica uma proteína cuja região C-terminal possui similaridade de sequência com a PurE. Este gene está amplamente distribuído no genoma de espécies dos Domínios Archaea e Bacteria e são homólogos do gene *cpmA*, de *Synechococcus* sp. strain PCC 7942. A CpmA (*c*ircadian *p*hase *m*odifier) que regula o sistema circadiano em cianobactérias (DVORNYK, 2006; KATAYAMA et al., 1999). Devido a similaridade com a PurE na sua região C-terminal os ortólogos da CpmA estão erroneamente anotados como prováveis PurEs na maioria do genomas analisados neste estudo.

Os ortólogos da CpmA são mais similares às PurEs de espécies do Filo Euryarchaeota, contudo, esta similaridade de sequência não se estende à região C-terminal das PurEs que contém a última α -hélice desta enzima (KATAYAMA et al., 1999). Esta região está ausente na CpmA e seus ortólogos. Além disso, as quatro alças que formam o sítio ativo das PurEs não são conservadas nas CpmAs. As exceções são os resíduos de aminoácidos His45 e Arg46 na alça 2 (PurE de *Escherichia coli*) e os resíduos de aminoácidos Gly72, Ala73 e Leu76 na alça 3 (PurE de *Escherichia coli*), também conservados nas CpmAs (KATAYAMA et al., 1999). Portanto, estender a similaridade de sequência observada entre as PurEs e CpmAs para uma similaridade na atividade bioquímica é meramente especulativo.

Genômica comparativa de *purE* e *purK*

Os seguintes padrões de presença e ausência foram observados na genômica comparativa dos genes *purE* e *purK*: (I), *purE* e *purK* estão presentes no genoma dos organismos (II), apenas *purE* está presente no genoma, (III), *purE* e *purK* estão ausentes no genoma. Estes padrões serão referidos daqui em diante como estrutura I, II e III, respectivamente (Figura 3.2).

Os padrões estruturais I e II são os mais amplamente distribuídos nas linhagens procarióticas (Figura 3.2). Não foram encontradas espécies que possuem apenas o gene *purK*, o que sugere que sua presença está condicionada à presença do gene *purE* no genoma (Figura 3.2).

As espécies dos filos Tenericutes e Chlamydiae incluídas nesse estudo não possuem os genes *purE* e *purK* (Figura 3.2). Os demais genes da VBP 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.

A ausência de *purE* e *purK* também foi observado em espécies de dezesseis outros filios e classes (Figura 3.2). Estas espécies são parasitas intracelulares ou simbiotes e a ausência dos genes *purE* e *purK* nos genomas destas espécies indica que o modo de vida dos organismos é um fator determinante na estrutura desta etapa da VBP. Provavelmente, a ausência destes genes é resultado da redução do tamanho do genoma do organismo relacionada com seu modo de vida.

Todos os filios e classes das bactérias Gram positivas foram amostrados neste estudo, exceto a classe Thermolithobacteria do Filo Firmicutes. As análises de genômica comparativa de *purE* e *purK* no Filo Tenericutes já foram apresentados acima no texto. Os resultados para os demais filios e classes das bactérias Gram positivas serão apresentados a seguir.

A estrutura I da sexta etapa da VBP foi encontrada em 22 gêneros e 98 espécies da Classe Bacilli do Filo Firmicutes. Resultados similares foram encontrados para as espécies do Filo Actinobacteria onde 59/113 gêneros/espécies possuem a estrutura I. Apenas 9/14 gêneros/ espécies de Actinobacteria possuem a estrutura II da sexta etapa da VBP. Estes gêneros/espécies estão distribuídos apenas em três das trinta e sete famílias amostradas do Filo Actinobacteria sendo 1/5 gênero/espécies da família *Franckiaceae*, todos os 7/8 gêneros/espécies da família *Coriobacteriaceae* e a única espécie da família *Conexibacteriaceae* amostrada. (Tabela Suplementar). Estes resultados sugerem que a estrutura I é a condição ancestral da sexta etapa da VBP para as espécies do filo Actinobacteria e da Classe Bacilli do Filo Firmicutes.

As demais classes de Firmicutes analisadas apresentaram padrões variáveis da sexta etapa da VBP ou foram amostrados poucos gêneros e espécies, portanto os resultados devem ser vistos com cautela. Na Classe Clostridia a estrutura II foi a mais frequente em nível de gênero e espécie, contudo, sem a mesma predominância que do padrão estrutural I na classe Bacilli e Filo Actinobacteria. Enquanto 33/68 gêneros/espécies da Classe Clostridia possuem a estrutura II, 11/23 gêneros/espécies possuem a estrutura I (Figura 3.2). A estrutura II da sexta etapa da VBP é predominante em todas as famílias de Clostridia analisadas, e os 11/23 gêneros/espécies que possuem a estrutura I estão distribuídos nestas famílias (Tabela Suplementar). Isto sugere que a estrutura II da sexta etapa da VBP é a condição ancestral para as espécies da classe Clostridia.

O genoma de apenas duas espécies representando dois gêneros da classe *Erysipelotrichia* foram analisados. Ambas espécies pertencem à família *Erysipelotrichaceae*, contudo, enquanto a espécie *Erysipelothrix rhusiopathiae* não possui *purK* e *purE* a espécie *Eubacterium cylindroides* T2-87 possui apenas *purE*. Estes

resultados mostram a existência de variação para estrutura da sexta etapa da VBP nos membros desta família (Tabela Suplementar).

Apenas cinco espécies representando cinco gêneros da classe Negativicutes foram amostrados, e apenas *purE* foi encontrado no genomas delas (Figura 3.2). Contudo, estas espécies representam as duas famílias reconhecidas desta classe sugerindo que a ausência de *purK* pode ser a condição ancestral para as espécies deste taxon. As bactérias Gram negativas constituem o grupo mais diverso dentre os procariotos conhecidos, sendo reconhecidos 28 filos e 46 classes destes organismos. Destes 86% dos filos e 65% das classe reconhecidos destes organismos foram amostrados neste trabalho.

Nos Taxa de bactérias Gram negativas apenas a estrutura I foi encontrada em todos os gêneros e espécies de Cianobacteria, e, foi predominante nas espécies dos Filos Deinococcus-Thermus e Bacteroidetes e nas classes Alfa-proteobacteria, Beta-proteobacteria e Gama-proteobacteria (Figura 3.2). Os 12/28 gêneros/espécies do Filo Cyanobacteria incluídos nesta análise representam cinco das sete Ordens do grupo, logo, é razoável inferir que estrutura I pode ser a condição ancestral presente no ancestral comum desse Filo.

Os cinco gêneros e treze espécies do Filo Deinococcus-Thermus cujo genomas foram analisados representam todas as ordens, famílias e seis dos dez gêneros conhecidos deste filo (Tabela Suplementar). Apenas uma espécie, *Oceanithermus profundus* DSM 14977, possui a estrutura II do sexto passo da VBP, as demais possuem a estrutura I. Esta espécie pertence a família *Thermaceae* que possui seis gêneros reconhecidos e quatro deles foram amostrados nesse estudo (Tabela Suplementar). Apesar de apenas espécies de seis gêneros do Filo Deinococcus-Thermus terem sido amostrados em nossas análises de genômica comparativa, eles representam todos os taxa conhecidos do grupo em nível

de classe, ordem e família e a maioria do gêneros conhecidos. Isto sugere que a estrutura I da VBP é ancestral neste grupo.

Em Bacteroidetes foram analisados genomas de 61 espécies de um total de 44 gêneros que representam todas as classes e ordens deste filo (Figura 3.2). A estrutura I foi encontrada em 39 espécies em um total de 33 gêneros distribuídos em todas as classes e ordens. A estrutura II da sexta etapa da VBP foi encontrada em 20 espécies de nove gêneros, contudo, limitada a três das quatro famílias amostradas da Ordem Bacteroidales e a um gênero da família *Blattabacteriaceae*, pertencente à Ordem Flavobacteriales (Tabela Suplementar).

As Alfa-, Beta- e Gama-proteobacteria são classes estreitamente relacionados filogeneticamente (LANG; DARLING; EISEN, 2013), isto pode explicar por que a estrutura I da VBP é predominante nas espécies destas classes. Dentre as 135 espécies de Alpha-proteobacteria analisadas apenas *Magnetococcus marinus MC-1* possui a estrutura II da sexta etapa da VBP, única representante da sua classe e ordem analisada neste trabalho (Figura 3.2). As duas espécies de Beta-proteobacteria que possuem a estrutura II da sexta etapa da VBP pertencem à mesma ordem e família (Figura 3.2). Nas Gama-proteobacterias as 14 espécies que possuem a estrutura II estão distribuídas em 7 famílias de 6 ordens, onde a estrutura I é predominante nas demais espécies cujos genomas foram analisados (Tabela Suplementar). A exceção é a família *Coxiellaceae*, representada apenas pela espécie *Coxiella burnetii*, e que possui a estrutura II da sexta etapa da VBP.

Apesar da estrutura I da sexta etapa da VBP ser a mais frequente nas espécies dos Filos Aquificae, Chloroflexi, Acidobacteria e Planctomycetes analisadas ela não foi predominante. Além disso, poucas espécies destes taxa foram analisadas e o padrão de distribuição das espécies com a estrutura I ou II nas famílias e ordens não permite inferir sobre a condição ancestral da sexta etapa da VBP neste filós. Por exemplo, as nove

espécies de Aquificae cujos genomas foram analisados representam as três famílias conhecidas desse filo. Contudo, as estruturas I e II são observadas em espécies de duas famílias e em diferentes espécies da Família *Hydrogenothermaceae* (Figura 3.2).

Em Chloroflexi o cenário para a variação da sexta etapa da VBP é similar ao encontrado em Aquificae. Por exemplo, embora todas as espécies da Classe Dehalococcoidetes analisadas possuam a estrutura II elas representam apenas dois gêneros da mesma ordem e família. As Famílias *Sphaerobacteraceae* e *Thermomicrobiaceae* pertencem a ordens diferentes da classe Thermomicrobia e estão representadas apenas pelas espécies, *Sphaerobacter thermophilus* DSM 20745 e *Thermomicrobium roseum* DSM 5159, respectivamente. Enquanto a primeira possui a estrutura I da sexta etapa da VBP a última possui a estrutura II. Todas as espécies da Classe Chloroflexi possuem a estrutura I, contudo, a representação delas é limitada a três gêneros, duas famílias e duas ordens (Tabela Suplementar).

Dos seis genomas analisados de espécies do Filo Acidobacteria dois pertencem a espécies que estão na condição de *candidatus* e não possuem uma classificação definida. Apenas *Candidatus Koribacter versatilis* Ellin345 possui a estrutura II da sexta etapa da VBP, as demais espécies possuem a estrutura I contudo, pertencem a quatro gêneros que representam apenas duas classes, duas ordens e duas famílias (Tabela Suplementar). No Filo Planctomycetes a representatividade dos genomas analisados é ainda menor. Todas as cinco espécies cujos genomas foram analisados possuem a estrutura I da sexta etapa da VBP, entretanto, representam apenas uma classe, uma ordem e uma família (Tabela Suplementar).

Com base nos resultados apresentados acima não é possível inferir sobre a condição ancestral da sexta etapa da VBP nos Filos Aquificae, Chloroflexi, Acidobacteria e Planctomycetes, entretanto, mostram a existência de variação para estrutura da sexta

etapa da VBP nos membros destes taxa. Todas as vinte e quatro espécies analisadas da Classe Epsilon-proteobacteria possuem apenas estrutura II da sexta etapa da VBP. Estas espécies representam aproximadamente 20% das espécies conhecidas deste taxon e estão distribuídas nas três ordens, nas três das quatro famílias e em dez dos dezessete gêneros conhecidos desse filo. Diante do exposto é razoável inferir que a estrutura II da sexta etapa da VBP é a condição ancestral deste filo.

Assim como observado no Filo Epsilon-proteobacteria, apenas a estrutura II da sexta etapa da VBP foi encontrada nos genomas analisados das espécies dos Filos Dictyoglomi, Elusimicrobia, Synergistetes, Deferribacteres, Chrysiogenetes e Fibrobacteres, mas estes resultados devem ser visto com cautela. Estes filios estão representados por apenas uma a quatro espécies em nossas análises, além disso, estes taxa foram descritos ou levados a categoria de filo recentemente e poucas espécies destes são conhecidas ainda (EUZÉBY, 1997). Deste modo, não é possível fazer qualquer consideração sobre a variação e a condição ancestral da sexta etapa da VBP nestes taxa.

A estrutura II da sexta etapa da VBP é predominante nos genomas analisados das espécies dos Filos Fusobacteria, Chlorobi e Delta-proteobacteria, mas a representatividade das espécies analisadas difere entre estes taxa. As cinco espécies de Fusobacteria analisadas representam uma única ordem, as duas famílias e cinco dos nove gêneros conhecidos deste filo. A estrutura I da sexta etapa da VBP está presente apenas na espécie *Ilyobacter polytropus* DSM 2926, que pertence à Família *Fusobacteriaceae*. A outra espécie desta família cujo genoma foi analisado, *Fusobacterium nucleatum*, possui a estrutura II, assim como as três espécies da Família *Leptotrichiaceae* (Tabela Suplementar). Estes resultados mostram que existe variação na estrutura da sexta etapa da VBP nas espécies do Filo Fusobacteria, e sugere que a estrutura II da sexta etapa da VBP pode ser a condição ancestral para as espécies neste taxon.

O Filo Chlorobi possui duas classes, Chlorobia e Ignavibacteria, e ambas possuem apenas uma ordem e uma família. As dez espécies deste filo cujos genomas foram analisados neste trabalho pertencem à Classe Chlorobia, ordem Chlorobiales e representam cinco dos seis gêneros da família *Chlorobiaceae* (Tabela Suplementar). A espécie *Chloroherpeton thalassium ATCC 35110* é a única que possui a estrutura I da sexta etapa da VBP sugerindo que pelo menos para os membros da Classe Chlorobia a estrutura II é a condição ancestral da sexta etapa da VBP.

A estrutura II da sexta etapa da VBP é predominante nos vinte e cinco gêneros e quarenta e duas espécies da Classe Delta-proteobacteria cujos genomas foram analisados. Estas espécies representam todas as oito ordens e quinze das vinte e cinco famílias dessa Classe. As sete espécies que possuem a estrutura I da sexta etapa da VBP são das ordens Desulfobacterales, Desulfovibrionales e Myxococcales, dentre elas *Desulfomicrobium baculatum DSM 4028*, *Stigmatella aurantiaca DW4/3-1* e *Sorangium cellulosum 'So ce56'* são as únicas representantes das suas famílias analisadas neste trabalho. As demais espécies de Delta-proteobacteria com a estrutura I pertencem a famílias que também possuem espécies com a estrutura II (Tabela Suplementar).

A estrutura II da sexta etapa da VBP é predominante nas Delta-proteobacterias e está presente em todas as espécies da Classe Epsilon-proteobacteria analisadas neste trabalho. Estas duas classes do Filo Proteobacteria são estreitamente relacionadas filogeneticamente e basais na filogenia das proteobacteria (LANG; DARLING; EISEN, 2013). Assim, isso remete a questão de que a estrutura II da sexta etapa da VBP estaria presente no ancestral destas Classes do Filo Proteobacteria.

No domínio Archaea as estruturas I e II da sexta etapa da VBP são encontradas nas espécies do Filo Crenarchaeota e do Filo Euryarchaeota. Apenas uma espécie dos Filos Korarchaeota, Thaumarchaeota e Nanoarchaeota foi analisada, portanto não é possível

inferir sobre a condição ancestral nem sobre a existência de variação da sexta etapa da VBP nestes taxa.

No Filo Crenarchaeota a estrutura I é predominante, ela esta presente em 7/18 gênero/espécies, enquanto a estrutura II esta presente em apenas 4/4 gêneros/espécies.

É notável que a variação na sexta etapa da VBP nas Crenarchaeotas ocorre em nível de ordem. Todas as espécies com a estrutura I pertencem as Ordens Sulfolobales e Thermoproteales, e todas as quatro espécies que possuem a estrutura II são da Ordem Desulfurococcales. É possível observar que a maioria das espécies amostradas desta ordem possui a estrutura III da sexta etapa da VBP. De fato elas não possuem o conjunto mínimo de genes da VBP duas ordens amostradas.

No Filo Euryarchaeota é observado o contrário, a estrutura II predomina, sendo encontrada em 28/51 gêneros/espécies, enquanto que a estrutura I esta presente em 17/21 gêneros/espécies. Assim como observado no Filo Crenarchaeota, a maior parte da variação na sexta etapa da VBP nas espécies de Euryarchaeota cujos genomas foram analisados ocorre em nível de ordem. A estrutura I está presente em todas as espécies das Ordens Halobacteriales e Thermoplasmatales. A estrutura I também está presente no genoma de três das doze espécies da Ordem Thermococcales cujos genomas foram analisados, mas nestes casos a variação ocorre entre espécies do mesmo gênero.

A ampla distribuição de grupos de microrganismos procarióticos com *purE* e sem *purK* (estrutura II da sexta etapa da VBP) foi uma surpresa, pois implica que estes microrganismos teriam um análogo da *purK* ou teriam uma *purE* classe 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* classe II só tinha sido encontrada em eucariotos (CHEN; DIXON;

ZALKIN, 1990; CHUNG et al., 1996; DANDANELL; SORENSEN, 1997; DRAZEK et al., 1995; FIRESTINE; DAVISSON, 1994; FIRESTINE et al., 1998; HOSKINS et al., 2007; MEYER et al., 1992; MUELLER et al., 1994; THODEN et al., 1999). Uma terceira hipótese é que estes microrganismos conseguiriam sobreviver apenas com uma PurE classe I, em função de suas características ecológicas tais como ambiente nativo rico em CO₂ e baixa taxa de crescimento.

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. Os genomas destas Archaeas foram analisados neste trabalho e de acordo com a nossa genômica comparativa possuem apenas o gene *purE* em seus genomas. Em conjunto estes dados sugerem que a PurE destas archaeas seja classe II.

Adicionalmente, Tranchimand e colaboradores (2011) relataram que *Treponema denticola*, uma bactéria pertencente ao filo Spirochaetes, possui uma PurE classe II e, de acordo com nossa análise de genômica comparativa esta espécie de espiroqueta não possui *purK*. Mais uma vez os dados da genômica comparativa são congruentes com o tipo de PurE codificada nos genomas dos organismos.

Contexto genômico de *purE* e *purK*

O gene *purE* foi encontrado em contexto genômico com outros genes da VBP em 58%/53% dos gêneros/espécies que possuem a estrutura II da sexta etapa da VBP (Figura 3.3A).

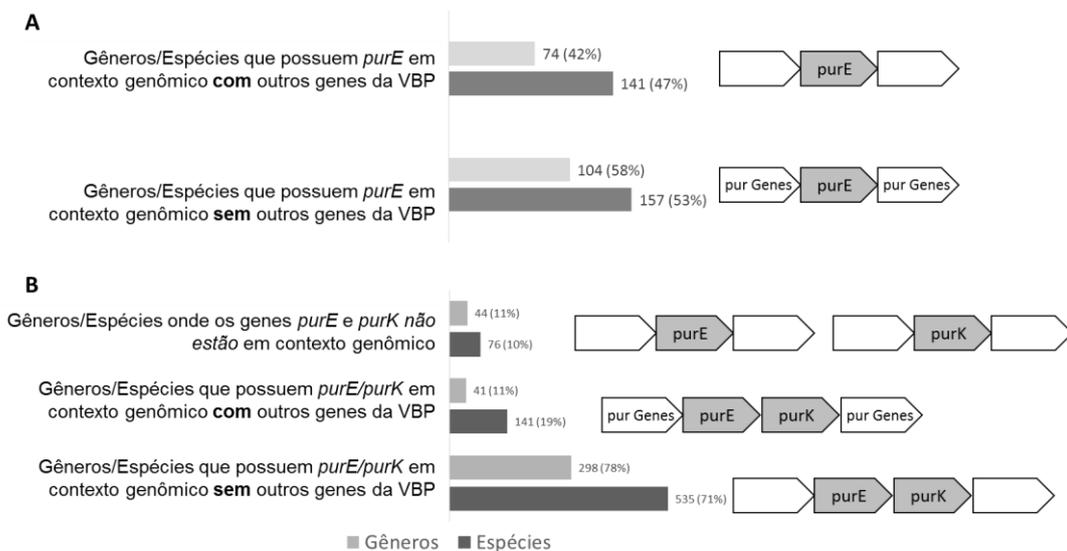


Figura 3.3. Contexto genômico dos gêneros e espécies que possuem os genes *purE* e *purK*, ou apenas *purE*. (A) Contexto genômico dos gêneros/espécies que possuem apenas *purE*, nesses organismos esse gene pode ser encontrado com outros genes da biossíntese de purinas (104/157 gêneros/espécies) ou em um contexto sem outros genes da via (74/141 gêneros/espécies). (B) Contexto genômico dos gêneros/espécies que possuem *purE* e *purK*. Nesses organismos os genes *purE* e *purK* podem estar em contextos genômicos diferentes (44/76 gêneros/espécies), ou juntos no mesmo contexto. Quando esses genes estão no mesmo contexto genômico eles podem ser encontrados com outros genes da VBP (40/141 gêneros/espécies), ou sem outros genes da VBP (298/535 gêneros/espécies).

Em 90% das espécies que possuem a estrutura I da sexta etapa da VBP os genes *purE* e *purK* estão associados em nível de contexto genômico no arranjo *purE/purK* ou *purK/purE* (Figura 3.3B). Contudo, na maioria destes gêneros/espécies esses genes não estão associados em nível de contexto genômico com os outros genes da VBP (Figura 3.3B). A mesma situação é observada nos gêneros/espécies em que *purE* e *purK* não estão associados em nível de contexto genômico (Figura 3.3B).

Os arranjos *purE/purK* ou *purK/purE* estão amplamente distribuídos nas linhagens procarióticas e sua distribuição e predominância nos filos analisados não refletem necessariamente a filogenia de procariotos (Figura 3.2 e 3.3). A exceção é a predominância do arranjo *purE/purK* nas espécies das Classes Alfa-, Beta- e Gama-Proteobacteria, estas taxa são estreitamente relacionados filogeneticamente. Estes

resultados indicam uma forte associação entre *purE* e *purK* em nível de contexto genômico nas linhagens procarióticas, independente dos demais genes da VBP.

Em conjunto, os resultados apresentados até aqui dão suporte a hipótese da existência de *purE* classe II em procariotos e de que o gene *purE* presente nos genomas de organismos com *purK* codifica uma PurE classe I. Adicionalmente nossas análises de genômica comparativa sugerem que estas enzimas estão amplamente distribuídas nos Domínios Archaea e Bacteria.

Filogenia molecular de PurK

Nas espécies que possuem *purE* e *purK* em contexto genômico estes genes apresentam-se nos arranjos *purE/purK* ou *purK/purE* (Figura 3.2). Estes arranjos estão presentes em espécies dos Domínios Archaea e Bacteria e sua distribuição não reflete necessariamente a filogenia de procariotos, ou seja, espécies de um grupo monofilético possuem arranjos diferentes.

Estes arranjos podem ser resultados de rearranjos gênicos que ocorreram recentemente ou de antigos operons que foram sendo transferidos lateralmente nas linhagens procarióticas. Para investigar a origem desses arranjos nós fizemos uma análise filogenética com as sequências de proteínas das PurKs recuperadas na análise de genômica comparativa.

Na árvore filogenética não enraizada as *purKs* que estão no contexto *purE/purK* ficam em um grupo distinto das *purKs* que estão no contexto *purK/purE* (Figura 3.4). As *purKs* que não estão associadas em nível de contexto genômico com *purE* ficam distribuídas entre os dois clados. Estes resultados mostram que os arranjos *purE/purK* ou *purK/purE* resultam de eventos antigos na evolução procariótica.

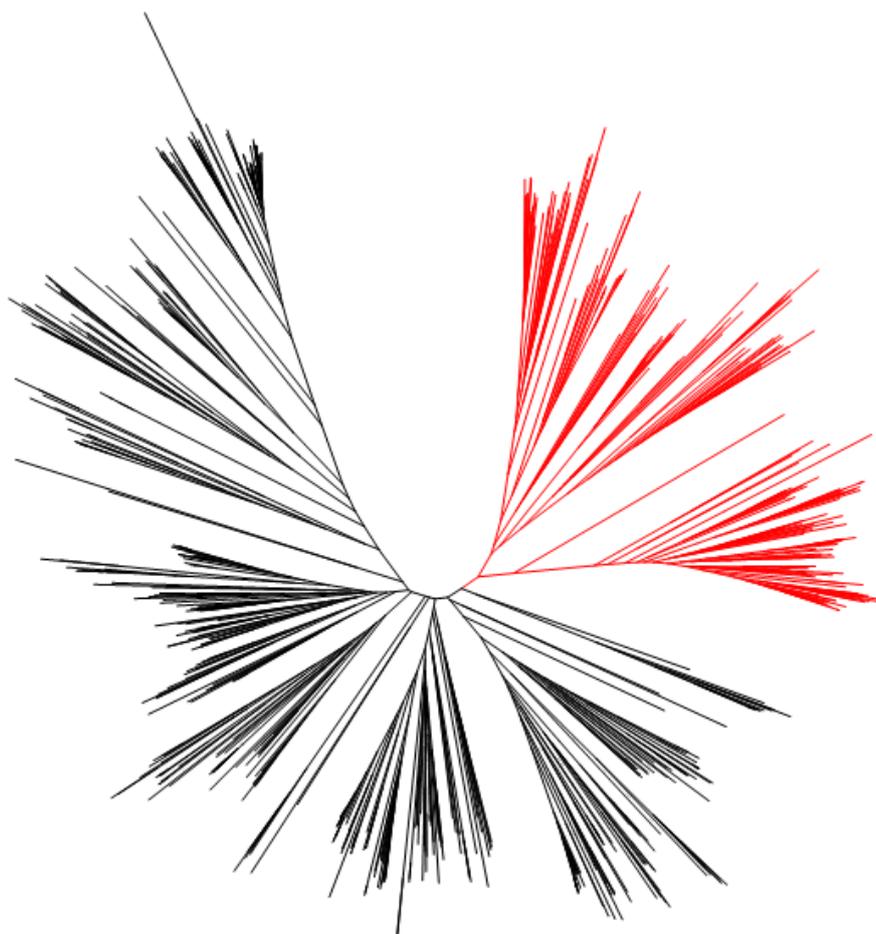


Figura 3.4. Filogenia de PurK. Árvore filogenética não enraizada estimada a partir das sequências de aminoácidos da PurK pelo método da Máxima Verossimilhança. O clado contendo as PurKs que estão no arranjo *purK/purE* está destacado em vermelho, o grupo que contém as PurKs no arranjo *purE/purK* está em preto. A árvore filogenética foi criada usando o PhyML 3.0 e editada com o MEGA6 (TAMURA et al., 2013)

Para expandir nossa compreensão sobre a origem desses arranjos nós enraizamos a árvore filogenética da *purK* com a *purT*, um parálogo da *purK* (Figura 3.5). A PurT cataliza a terceira reação da via biossintética de purinas e compartilha considerável similaridade em nível de estrutura primária e terciária com a PurK. Os clados basais da árvore filogenética enraizada da *purK* são compostos por sequências que estão arranjo *purE/purK* (Figura 3.5). Além disso, todas as *purKs* que participam do arranjo *purK/purE* estão em um clado que emerge mais internamente na árvore filogenética, dentro de um

clado contendo *purK*s que estão arranjo *purE/purK*. Este padrão de ramificação indica que o arranjo *purE/purK* é ancestral do arranjo *purK/purE*.

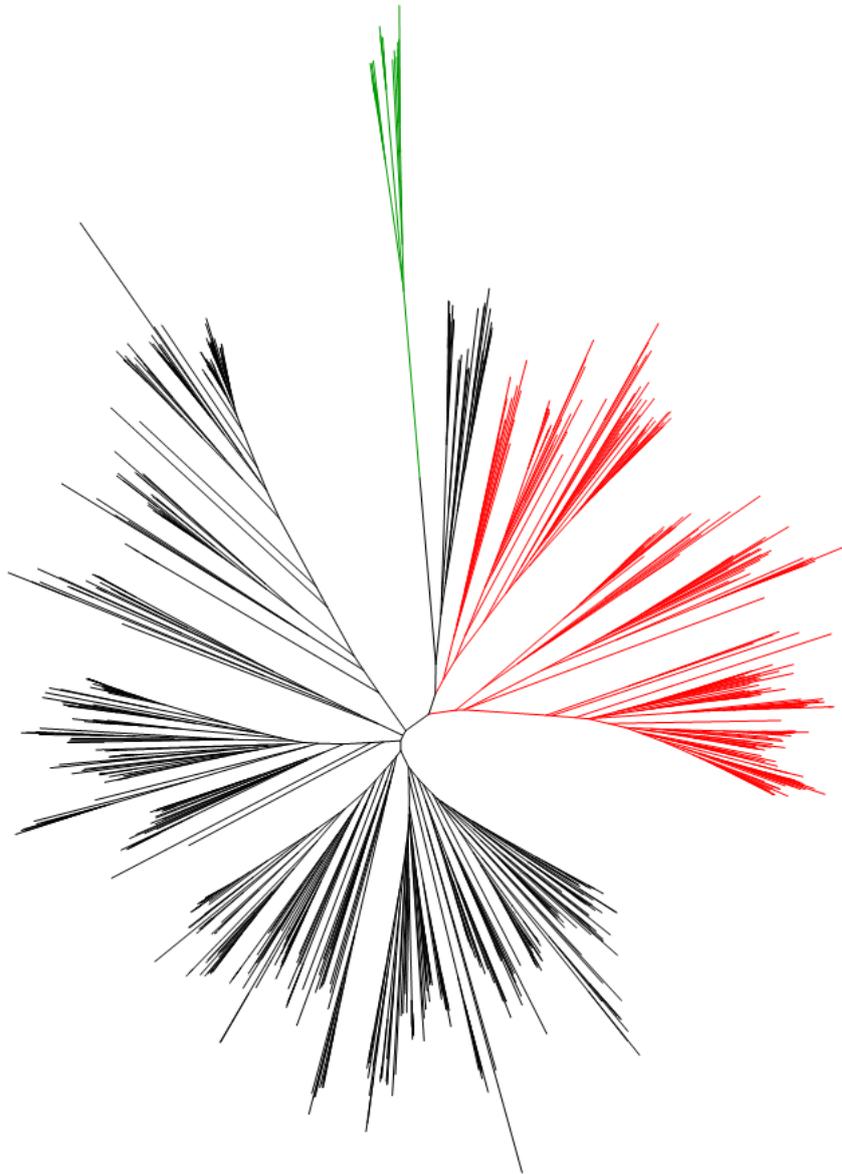


Figura 3.5. Filogenia de PurK. Árvore filogenética não enraizada estimada a partir das sequencias de aminoácidos da PurK e PurT. O clado contendo as PurKs que estão no arranjo *purK/purE* esta destacado em vermelho, o grupo que contem as PurKs no arranjo *purE/purK* esta em preto. O grupo de PurTs esta apresentado em verde.

Nós não encontramos um parálogo do gene *purE*. Isto impossibilitou obter uma árvore filogenética enraizada deste gene de modo não arbitrário. Contudo, a árvore da *purE* enraizada arbitrariamente no ponto médio ou no ramo que conecta as *purEs* dos

organismos que contém *purK* com as *purEs* dos organismos sem *purK* possui uma topologia congruente com a árvore enraizada da *purK*.

Filogenia molecular de PurE

A existência de grupos de microrganismos procarióticos com *purE* e sem *purK* sugere que estes microrganismos ou possuem um análogo da *purK* ou uma *purE* classe II ou conseguem sobreviver apenas com uma PurE classe I. Estas hipóteses são igualmente plausíveis, contudo, algumas evidências experimentais indicam que algumas espécies do Domínio Bacteria e Domínio Archaea possuem PurE classe II (HAMILTON; REEVE, 1985; TRANCHIMAND et al., 2011). Estes resultados apontam para a segunda hipótese e, adicionalmente, nossas análises de genômica comparativa e contexto genômico sugerem ainda que a PurEs classe I e II estão amplamente distribuídas nos Domínios Archaea e Bacteria.

Estes resultados remetem a questões sobre as relações evolutivas entre as PurEs das linhagens procarióticas. Portanto, nós realizamos uma análise filogenética com as sequências de nucleotídeos (primeira e segunda bases dos códons – dados não apresentados) e de aminoácidos das PurEs recuperadas na análise de genômica comparativa e as PurEs classe II de metazoários caracterizadas funcionalmente.

As árvores filogenéticas com sequências de nucleotídeos e de aminoácidos geradas possuem topologias similares. Em ambas as árvores as PurEs de organismos sem *purK* formam um grupo distinto e inclui todas as PurEs com evidências experimentais de que são classe II. Por exemplo, as PurEs de *Methanothermobacter thermautotrophicus*, *Methanobrevibacter smithii*, *Treponema denticola* e as PurEs classe II de metazoários (Figura 3.3). Este grupo será referido daqui em diante como grupo II. Os demais grupos da árvore filogenética contém as PurEs de organismos que possuem PurK, incluindo todas

as PurEs classe I caracterizadas funcionalmente (Figura 3.6). Estes grupos, em conjunto, serão referidos daqui em diante como grupo I.

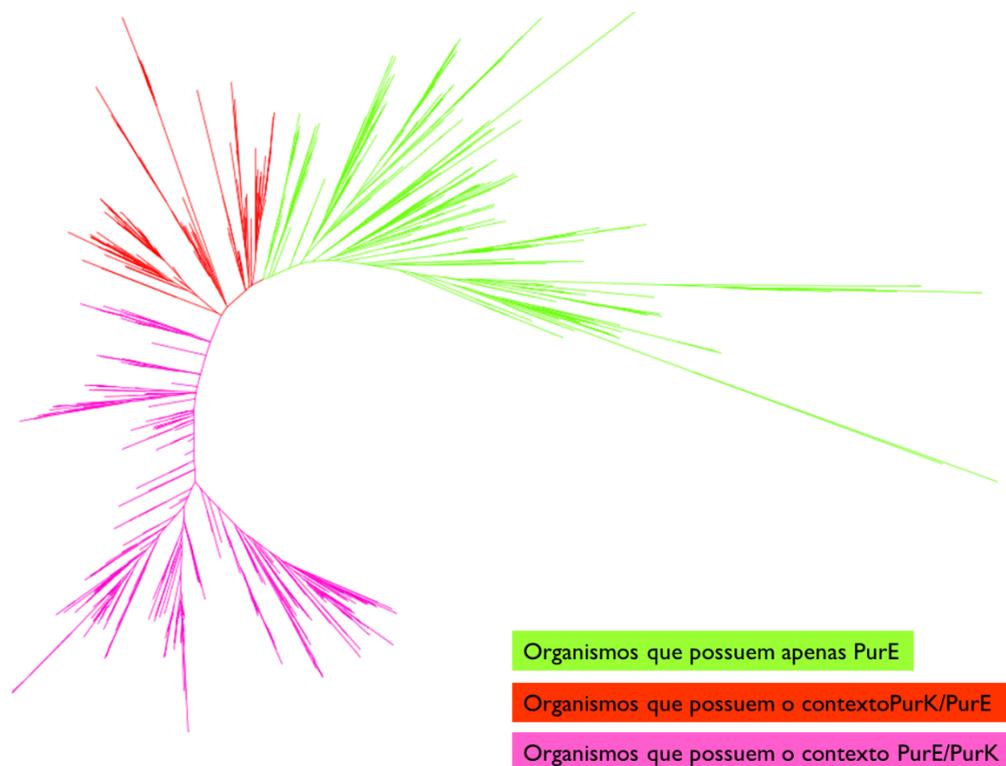


Figura 3.6. Filogenia de PurE. Árvore filogenética não enraizada estimada a partir das sequências de aminoácidos da PurE pelo método da Máxima Verossimilhança. As PurEs das espécies que possuem *purK* nos contextos *purK/purE* e *purE/purK* são mostradas em vermelho e rosa, respectivamente. Em conjunto, essas PurEs formam o Grupo I. As PurEs das espécies que não possuem PurK estão destacadas em verde (Grupo II). A árvore filogenética foi criada usando o PhyML 3.0 e editada com o MEGA6 (TAMURA et al., 2013).

Algumas espécies de bactérias possuem duas cópias do gene *purE* que ficam em grupos distintos na árvore filogenética, uma no grupo I e a outra no grupo II. As isoformas de *purE* que estão no grupo I, estão em um contexto genômico com o gene *purK* enquanto que as isoformas que ficam no grupo II não possuem em seus contextos genômicos genes da biossíntese de purinas. A exceção é a *purE* do grupo II de *Thermoanaerobacter pseudethanolicus* 33223, que está em um contexto genômico com outros genes da

biossíntese de purinas. Estes resultados são uma evidência de que as *purEs* I e II estão sujeitas a Transferência Lateral de Genes (TLG).

A topologia das árvores filogenéticas de PurE obtidas com sequências de aminoácidos e com sequência de nucleotídeos são similares. Além disso, a distribuição das PurEs funcionalmente caracterizada nas reconstruções filogenéticas é uma evidência de que as PurEs que estão nos grupos I e II são, respectivamente, candidatas a serem PurEs classe I e PurEs classe II. Estes dados, junto com as análises genômicas, também suportam a hipótese de que as PurEs classe I e classe II estão amplamente distribuídas em procariotos.

Discussão

Genômica comparativa e filogenia da PurE

A via biossintética de purinas é uma das vias biossintéticas mais antigas, e compreender sua origem e diversificação ajuda a expandir nossa compreensão sobre os eventos que nortearam a evolução da vida. Além disso, conhecer a diversidade e eventuais variações funcionais das enzimas desta via biossintética é fundamental para usá-las como alvos para fármacos anticâncer e antimicrobianos. Assim, buscamos neste artigo fornecer um quadro sobre a diversidade e evolução dos genes *purK* e *purE*, ambos envolvidos na sexta etapa da VBP.

As análises de genômica comparativa mostram os que os genes *purE* e *purK* tem três padrões de distribuição nas linhagens procarióticas incluídas neste estudo. Estes resultados associados com a filogenia destes genes tornaram possível inferir sobre a condição ancestral da sexta etapa da VBP em algumas linhagens dos Domínios Archaea e Bactéria.

As análises de genômica comparativa também revelaram que inúmeras espécies de Bactérias e Archaeas possuem *purE* mas não possuem *purK*. Este resultado associado

à filogenia da *purE* sugere que as *purEs* dos organismos sem *purK* são filogeneticamente relacionadas. Contudo, a similaridade entre as sequências do grupo I é maior do que a observada entre as sequências do grupo II. Além disso, o comprimento médio dos ramos das sequências do grupo I é menor quando comparado com o grupo II (Figura .6). Estes dados poderiam sugerir que o agrupamento das *purEs* dos organismo sem *purK* seja um artefato oriundo de Long Branch Attraction (LBA).

Entretanto, todas as PurEs caracterizadas experimentalmente como classe I ficam dentro do grupo I enquanto que as PurEs caracterizadas experimentalmente como classe II ficam dentro do grupo II (Figura 3.6). Adicionalmente, a topologia da árvore filogenética da PurE é congruentes com os dados da genômica comparativa.

É importante salientar que a topologia dentro dos grupos I e II reflete parcialmente a filogenia de procariotos. Por exemplo, as PurEs de filos aonde todas as espécies amostradas não possuem *purK* ficam dentro de um mesmo grupo. Estes dados suportam a hipótese de que os grupos I e II da árvore filogenética da PurE contém PurEs classe I e II, respectivamente.

Ordem de recrutamento de PurE e PurK para a via VBP

Em 78/71% dos gêneros/espécies que possuem *purE* e *purK* esses genes encontram-se no contexto genômico *purE/purK* ou *purK/purE*, que estão amplamente distribuídos nas linhagens procarióticas (Figura 3.3). É notável que, em geral, *purK* e *purE* não estão associados em nível de contexto genômico com outros genes da biossíntese de purinas.

A filogenia molecular das PurKs e de PurEs indica que estes genes estão claramente separados quanto aos arranjos *purE/purK* ou *purK/purE* (Figura 3.5 e 3.6). Além disso, as topologias dos clados que contém as *purKs* e *purEs* pertencentes ao mesmo arranjo são congruentes tanto na filogenia de PurE quanto na filogenia de PurK. Estes

dados mostram uma associação particular entre estes genes, independente dos demais genes da via biossintética de purinas.

Embora os arranjos *purK/purE* e/ou *purE/purK* sejam amplamente distribuídos em procariotos, a fusão de *purK* e *purE* é um evento raro neste organismos. Ele só foi observado no genoma de *Kocuria rhizophila*, uma Actinobacteria, dentre as 1048 espécies analisados neste estudo que possui esse gene (Tabela 3.1). De fato, eventos de fusão gênica envolvendo *purE* ou *purK* não foram favorecidos durante a evolução procariótica (Tabela 3.1), em contraste com o observado em linhagens eucarióticas (LI et al., 2007; SMITH; ATKINS, 2002; THODEN et al., 2010).

Entretanto, uma vez que PurK e PurE atuam na mesma etapa da VBP uma pergunta natural é quem foi recrutada primeiro? De acordo com a hipótese da “retro-evolução” proposta por Horowitz (1945) as células primitivas usavam inicialmente compostos presentes no ambiente. Com o intenso consumo destes compostos, eles se tornaram escassos e as células que possuíam enzimas capazes de sintetizar estes compostos a partir de um precursor disponível no ambiente teriam vantagens seletivas em detrimento daquelas que não possuísem tais enzimas.

Portanto, as enzimas das atuais vias biossintéticas teriam sido selecionadas de acordo com o esgotamento sucessivo de seus produtos. As primeiras enzimas que foram recrutadas durante a origem de uma via Biossintética teriam sido aquelas responsáveis pelas últimas etapas da via e as enzimas responsáveis pelas etapas iniciais teriam sido recrutadas posteriormente.

A hipótese de recrutamento inicial da PurK não se ajusta ao modelo da retro-evolução. Além disso, implica em assumir que o aumento na quantidade de N⁵-CAIR proporcionado pelo PurK favoreceu aumento da quantidade de CAIR suficiente para permitir o crescimento celular. Contudo, o N⁵-CAIR e CAIR rapidamente se decompõem

em AIR e a síntese não enzimática de CAIR a partir de N⁵-CAIR é de baixa eficiência (MUELLER et al., 1994).

Em espécies que possuem os genes *purE* e *purK*, mutações no gene *purE* resultam em auxotrofia para purinas, ou seja, nestas espécies apenas o gene *purK* funcional não é suficiente para o crescimento celular (CONSTANTINE et al., 2006; DRAZEK et al., 1995; WATANABE et al., 1989). É importante salientar que, dentre as 1165 espécies dos domínios Archaea e Bactéria analisados neste estudo não foram encontrados organismos que possuem apenas *purK* (Figura 3.2).

Em contraste, mutações no gene *purK* geralmente não são letais ou resultam em auxotrofia condicional em várias espécies procarióticas. Por exemplo, mutantes de *purK* de *Escherichia coli*, uma Gama-proteobactéria, conseguem crescer em meio mínimo em altas concentrações de CO₂ (5%) ou após o cultivo prolongado por vários dias em concentrações atmosféricas de CO₂ (SØRENCEN e DANDANELL, 1997). Resultados similares foram observados para mutantes do gene *purK* em *Synechococcus sp. strain PCC 7942*, uma Cianobactéria, *Brucella melitensis 16M*, uma Alfa-proteobactéria e em *Vibrio cholerae* e *Yersinia pestis*, ambas Gama-proteobactérias, (CHIANG; MEKALANOS, 1998; DRAZEK et al., 1995; FLASHNER et al., 2004; SCHWARZ et al., 1993).

Estas espécies foram incluídas neste estudo e pertencem a filos ou classes distintos cujas *purKs* estão, inclusive, em diferentes arranjos, o que mostra que estes resultados não são uma particularidade de grupo específico de organismos. Assim, a produção de CAIR pela PurE I a partir de N⁵-CAIR sintetizado não enzimaticamente *in vivo* (FIRESTINE; DAVISSON, 1994) pode explicar porque mutações no gene *purK* geralmente não são letais ou resultam em auxotrofia condicional.

Estes resultados são congruentes com a hipótese de que o recrutamento da PurK favoreceu o aumento do pool de N⁵-CAIR disponível para a PurE. Portanto, a argumentação exposta aqui suporta o cenário onde a PurK evoluiu como uma enzima acessória para PurE I, com recrutamento para a VBP posterior ao da PurE I.

O contexto genômico e filogenia de *purK* e *purE* indicam que os arranjos *purE/purK* e *purK/purE* encontrados nos genomas procarióticos, mantiveram-se estáveis desde a sua formação ao longo da evolução procariótica e que o arranjo *purE/purK* é ancestral do arranjo *purK/purE*. Em algumas espécies estes arranjos já foram de fato caracterizados como operons (NILSSON; KILSTRUP, 1998). Se a PurK já possuía atividade de carboxilase do AIR a formação deste provável operon ancestral ocorreu de acordo com o modelo de correção e implica em assumir que a PurE ancestral tinha atividade de mutase ou era multifuncional.

Mas, se a PurK adquiriu essa atividade posterior à formação desses operons significa que a formação de um operon egoísta (LAWRENCE; ROTH, 1996) envolvendo os genes *purK* e *purE* foi o evento central no recrutamento do gene *purK* para a biossíntese de purinas. Sendo assim, os arranjos destes genes nos genomas procarióticos atuais resultariam de operons egoístas ancestrais que foram sendo transferidos horizontalmente nas linhagens procarióticas. A elucidação da atividade catalítica presente no ancestral das PurEs irá ajudar a resolver essa questão de como a PurK foi recrutada para a VBP.

A hipótese para o recrutamento do gene *purK* e a ampla distribuição da provável *purE* classe II em procariotos remetem a questões sobre a condição ancestral de *purE*. Os dados apresentados torna inda torna difícil inferir se as PurEs I e II resultam de um evento de duplicação gênica ou se uma delas divergiu funcionalmente em alguma linhagem e depois foi transferida lateralmente para outras linhagens procarióticas. Contudo, nossos

resultados indicam que as PurEs I e II estão amplamente distribuídas nos domínios Archaea e Bactéria. Isto sugere que estas enzimas estavam presentes no ancestral comum dos seres vivos.

Contudo, a hipótese apresentada por Sampei e colaboradores propõe que a PurK é uma enzima que derivou de outra enzima que atua na terceira etapa da VBP, a PurN. Levando em consideração que a VBP evoluiu segundo o modelo de retro-evolução (HOROWITZ, 1945) a condição ancestral da sexta etapa da VBP seria a reação catalisada apenas pela PurE. Isso caracterizaria essa enzima como uma PurE classe II, que tem função carboxilase. A partir do surgimento da *purK* essa enzima teria sido recrutada para atuar junto com a PurE e esta última poderia ter sofrido especialização tornando-se uma mutase (PurE classe I) A manutenção desses genes e a formação de um operon pode ter sido favorecida pela eficiência da reação realizada por ambos os genes.

O conhecimento da distribuição dos genes *purE* e *purK* em procariotos tem uma importante aplicação na área médica, uma vez que a Via Biossintética de Purinas é um alvo para drogas anticâncer e antimicrobianas (BANERJEE; NANDAGOPAL, 2009; FIRESTINE et al., 2009). De fato, um grande número de espécies que possuem *purE* e *purK* são reconhecidos patógenos para animais e plantas (Tabela 3.2).

Conclusão

Ao contrário do conhecimento corrente, a PurE classe II está amplamente distribuída entre os filos/classes de procariotos. O recrutamento da PurE e PurK foi determinado pelo aumento do *pool* de CAIR e esse aumento teria sido a principal pressão de seleção no recrutamento de PurE para VBP. Dessa forma após seu surgimento a PurK evoluiu como uma enzima acessória da PurE, com seu recrutamento para a VBP sendo um precursor para surgimento de uma PurE I. Em muitos organismos a associação em nível genômico desses genes pode ter gerado operons de *purE* e *purK*, onde arranjos

purE/purK e *purK/purE* são operons egoístas que foram seletivamente mantidos ou perdidos nas diferentes linhagens procarióticas

Metodologia

Busca pelos genes *purK* e *purE*

As buscas pelos genes *purK* e *purE* incluíram 1490 genomas completamente sequenciados e anotados, disponíveis na base de dados do NCBI até março de 2011. Esses genomas representam 1165 espécies de 582 gêneros dos Domínios Archaea e Bacteria. As buscas foram feitas no banco de dados de proteínas, com os programas BLASTP e TBLASTN (ALTSCHUL et al., 1997). A matriz BLOSUM62 foi utilizada para as buscas. 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.

Análise da diversidade estrutural da sexta e 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. Aproximadamente 88% dos Filos e 82% das Classes de procariotos reconhecidas (EUZÉBY, 1997) estão representados neste estudo, contudo, não estão igualmente amostrados. Por isso, os resultados foram analisados em nível gênero e de espécie para minimizar o viés amostral dos genomas completamente sequenciados. Os Filos e Classes incluídos na análise, assim como o número de gêneros e espécies amostradas em cada Filo/Classe são apresentados na Tabela 2.

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 casos em que eles estavam flanqueados por genes da VBP.

Alinhamento múltiplo e filogenia de PurK e PurE

As sequências de nucleotídeos homólogas aos genes *purE* e *purK* foram recuperadas na base de dados do NCBI. As sequências de nucleotídeos foram traduzidas para aminoácidos foram alinhadas códon a códon no programa MEGA6 (TAMURA et al., 2013). Os alinhamentos múltiplos foram feitos no servidor GUIDANCE usando o algoritmo MAFFT (PENN et al., 2010). Os alinhamentos foram editados com base nas estruturas terciária das proteínas, disponíveis na base de dados do PDB (<http://www.rcsb.org/pdb/home/home.do>). As sequências incompletas ou com erros de anotação foram excluídas. A visualização e sobreposição das estruturas terciárias das proteínas foram feitas no programa PyMOL (DeLANO, 2002).

A partir das sequências de aminoácidos foram construídas as árvores filogenéticas de Máxima Verossimilhança utilizando o modelo LG (LeC Gascuel) (LE; GASCUEL, 2008) empregando o PhyML (GUINDON et al., 2010). A confiança dos ramos foi calculada pelo método aLRT (Approximate Likelihood-Ratio Test) (ANISIMOVA; GASCUEL, 2006).

Referências

- ACOSTA-GONZÁLEZ, A.; ROSSELLÓ-MÓRA, R.; MARQUÉS, S. Characterization of the anaerobic microbial community in oil-polluted subtidal sediments: aromatic biodegradation potential after the Prestige oil spill. **Environmental microbiology**, v. 15, n. 1, p. 77–92, jan. 2013.
- ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic acids research**, v. 25, n. 17, p. 3389–402, 1 set. 1997.
- ANISIMOVA, M.; GASCUEL, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. **Systematic biology**, v. 55, n. 4, p. 539–52, ago. 2006.
- BANERJEE, D.; NANDAGOPAL, K. Phylogenetic analysis and in silico characterization of the GARS-AIRS-GART gene which codes for a tri-functional enzyme protein involved in de novo purine biosynthesis. **Molecular biotechnology**, v. 42, n. 3, p. 306–19, jul. 2009.
- BOYLE, M. P. et al. Crystal structure of PurE (BA0288) from *Bacillus anthracis* at 1.8 Å resolution. **Proteins**, v. 61, n. 3, p. 674–6, 15 nov. 2005.
- CAETANO-ANOLLÉS, G.; KIM, H. S.; MITTENTHAL, J. E. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 22, p. 9358–63, 29 maio. 2007.
- CHEN, Z.; 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. **Biochemistry**, v. 87, p. 3097–3101, 1990.
- CHIANG, S. L.; MEKALANOS, J. J. Use of signature-tagged transposon mutagenesis to identify *Vibrio cholerae* genes critical for colonization. **Molecular microbiology**, v. 27, n. 4, p. 797–805, mar. 1998.
- CHUNG, S. O. et al. Genomic organization of purK and purE in *Brevibacterium ammoniagenes* ATCC 6872: purE locus provides a clue for genomic evolution. **FEMS microbiology letters**, v. 137, n. 2-3, p. 265–8, 1 abr. 1996.
- CONSTANTINE, C. Z. et al. Biochemical and Structural Studies of N 5 - Carboxyaminoimidazole Ribonucleotide Mutase from the Acidophilic Bacterium *Acetobacter aceti*. **Biochemistry**, v. 45, n. 27, p. 8193–8208, 2006.
- DANDANELL, G.; SORENSEN, I. S. Identification and sequence analysis of *Sulfolobus solfataricus* purE and purK genes. **FEMS microbiology letters**, v. 154, p. 173–180, 1997.
- DRAZEK, E. S. et al. Deletion of purE attenuates *Brucella melitensis* 16M for growth in human monocyte-derived macrophages. **Infection and immunity**, v. 63, n. 9, p. 3297–301, set. 1995.
- DVORNYK, V. Subfamilies of cpmA, a gene involved in circadian output, have different evolutionary histories in cyanobacteria. **Microbiology (Reading, England)**, v. 152, n. Pt 1, p. 75–84, jan. 2006.

EUZÉBY, J. P. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. **International journal of systematic bacteriology**, v. 47, n. 2, p. 590–2, abr. 1997.

FIRESTINE, S. M. et al. Reactions catalyzed by 5-aminoimidazole ribonucleotide carboxylases from *Escherichia coli* and *Gallus gallus*: a case for divergent catalytic mechanisms. **Biochemistry**, v. 33, n. 39, p. 11927–34, 4 out. 1994.

FIRESTINE, S. M. et al. Biochemical role of the *Cryptococcus neoformans* ADE2 protein in fungal de novo purine biosynthesis. **Archives of biochemistry and biophysics**, v. 351, n. 1, p. 123–34, 1 mar. 1998.

FIRESTINE, S. M. et al. Identification of inhibitors of N5-carboxyaminoimidazole ribonucleotide synthetase by high-throughput screening. **Bioorganic & medicinal chemistry**, v. 17, n. 9, p. 3317–23, 1 maio. 2009.

FIRESTINE, S. M.; DAVISSON, V. J. Carboxylases in de novo purine biosynthesis. Characterization of the *Gallus gallus* bifunctional enzyme. **Biochemistry**, v. 33, n. 39, p. 11917–26, 4 out. 1994.

FLASHNER, Y. et al. Generation of *Yersinia pestis* Attenuated Strains by Signature-Tagged Mutagenesis in Search of Novel Vaccine Candidates. v. 72, n. 2, p. 908–915, 2004.

GUINDON, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. **Systematic biology**, v. 59, n. 3, p. 307–21, maio. 2010.

HAMILTON, P. T.; REEVE, J. N. Sequence divergence of an archaeobacterial gene cloned from a mesophilic and a thermophilic methanogen. **Journal of molecular evolution**, v. 22, n. 4, p. 351–60, jan. 1985.

HOROWITZ, N. H. On the evolution of biochemical syntheses. **Proceedings of the National Academy of Sciences of the United States of America**, v. 31, n. 6, p. 153–157, 1945.

HOSKINS, A. A et al. N5-CAIR mutase: role of a CO₂ binding site and substrate movement in catalysis. **Biochemistry**, v. 46, n. 10, p. 2842–55, 13 mar. 2007.

KATAYAMA, M. et al. *cpmA*, a gene involved in an output pathway of the cyanobacterial circadian system. **Journal of bacteriology**, v. 181, n. 11, p. 3516–24, jun. 1999.

LANG, J. M.; DARLING, A. E.; EISEN, J. A. Phylogeny of Bacterial and Archaeal Genomes Using Conserved Genes : Supertrees and Supermatrices. **PloS one**, v. 8, n. 4, 2013.

LAWRENCE, J. G.; ROTH, J. R. Evolution of coenzyme B₁₂ synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex. **Genetics**, v. 142, n. 1, p. 11–24, jan. 1996.

LE, S. Q.; GASCUEL, O. An improved general amino acid replacement matrix. **Molecular biology and evolution**, v. 25, n. 7, p. 1307–20, jul. 2008.

LI, H.; FAST, W.; BENKOVIC, S. J. Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. **Protein science : a publication of the Protein Society**, v. 18, n. 5, p. 881–92, maio. 2009.

- LI, S.-X. et al. Octameric structure of the human bifunctional enzyme PAICS in purine biosynthesis. **Journal of molecular biology**, v. 366, n. 5, p. 1603–14, 9 mar. 2007.
- MATHEWS, I. I. et al. Crystal structure of Escherichia coli PurE, an unusual mutase in the purine biosynthetic pathway. **Structure (London, England : 1993)**, v. 7, n. 11, p. 1395–406, 15 nov. 1999.
- MATHEWS, I. I. et al. Crystal structure of Escherichia coli PurE, an unusual mutase in the purine biosynthetic pathway. **Structure (London, England : 1993)**, v. 7, n. 11, p. 1395–406, 15 nov. 1999.
- MEYER, E. et al. Purification and characterization of the purE, purK, and purC gene products: identification of a previously unrecognized energy requirement in the purine biosynthetic pathway. **Biochemistry**, v. 31, n. 21, p. 5022–32, 2 jun. 1992.
- MEYER, E. et al. Purification and characterization of the purE, purK, and purC gene products: identification of a previously unrecognized energy requirement in the purine biosynthetic pathway. **Biochemistry**, v. 31, n. 21, p. 5022–32, 2 jun. 1992.
- MEYER, E. et al. Evidence for the direct transfer of the carboxylate of N5-carboxyaminoimidazole ribonucleotide (N5-CAIR) to generate 4-carboxy-5-aminoimidazole ribonucleotide catalyzed by Escherichia coli PurE, an N5-CAIR mutase. **Biochemistry**, v. 38, n. 10, p. 3012–8, 9 mar. 1999.
- MUELLER, E. J. et al. N5-Carboxyaminoimidazole Ribonucleotide: Evidence for a new intermediate and two new enzymatic activities in the de Novo Purine Biosynthetic Pathway of Escherichia coli. **Biochemistry**, v. 33, n. 8, p. 2269–2278, 1994.
- NILSSON, D.; KILSTRUP, M. Cloning and expression of the Lactococcus lactis purDEK genes, required for growth in milk. **Applied and environmental microbiology**, v. 64, n. 11, p. 4321–7, nov. 1998.
- OWNBY, K.; XU, H.; WHITE, R. H. A Methanocaldococcus jannaschii archaeal signature gene encodes for a 5-formaminoimidazole-4-carboxamide-1-beta-D-ribofuranosyl 5'-monophosphate synthetase. A new enzyme in purine biosynthesis. **The Journal of biological chemistry**, v. 280, n. 12, p. 10881–7, 25 mar. 2005.
- PENN, O. et al. GUIDANCE: a web server for assessing alignment confidence scores. **Nucleic acids research**, v. 38, n. Web Server issue, p. W23–8, jul. 2010.
- SCHAEFER, J. et al. Multiple active site histidine protonation states in Acetobacter acetii N5-carboxyaminoimidazole ribonucleotide mutase (PurE) detected by REDOR NMR. **Biochemistry**, v. 46, n. 33, p. 9507–9512, 2009.
- SCHWARZ, R. et al. Phenotypic Complementation of High CO₂-Requiring Mutants of the Cyanobacterium Synechococcus sp. Strain PCC 7942 by Inosine 5'-Monophosphate. **Plant physiology**, v. 100, p. 1987–1993, 1993.
- SCHWARZENBACHER, R. et al. Crystal structure of a phosphoribosylaminoimidazole mutase PurE (TM0446) from Thermotoga maritima at 1.77-Å resolution. **Proteins**, v. 55, n. 2, p. 474–8, 1 maio. 2004.

SMITH, P. M. C.; ATKINS, C. A. Purine biosynthesis. Big in cell division, even bigger in nitrogen assimilation. **Plant Physiology**, v. 128, n. 3, p. 793–802, 2002.

SUN, H. et al. Complete genome sequence of *Desulfarculus baarsii* type strain (2st14). **Standards in genomic sciences**, v. 3, n. 3, p. 276–84, jan. 2010.

TAMURA, K. et al. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. **Molecular biology and evolution**, v. 30, n. 12, p. 2725–9, dez. 2013.

THODEN, J. B. et al. Three-dimensional structure of N5-carboxyaminoimidazole ribonucleotide synthetase: a member of the ATP grasp protein superfamily. **Biochemistry**, v. 38, n. 47, p. 15480–92, 23 nov. 1999.

THODEN, J. B. et al. Structural and Functional Studies of *Aspergillus clavatus* N5-Carboxyaminoimidazole Ribonucleotide Synthetase. **Biochemistry**, v. 49, n. 4, p. 752–760, 2010.

TRANCHIMAND, S. et al. *Treponema denticola* PurE Is a bacterial AIR carboxylase. **Biochemistry**, v. 50, n. 21, p. 4623–37, 31 maio. 2011.

WATANABE, W. et al. Identification and sequence analysis of *Escherichia coli* *purE* and *purK* genes encoding 5'-phosphoribosyl-5-amino-4-imidazole carboxylase for de novo purine biosynthesis. **Journal of bacteriology**, v. 171, n. 1, p. 198–204, jan. 1989.

ZHANG, Y.; MORAR, M.; EALICK, S. E. Structural biology of the purine biosynthetic pathway. **Cellular and molecular life sciences : CMLS**, v. 65, n. 23, p. 3699–724, nov. 2008.

Anexo 1

Tabela 1. Relação de bactérias patogênicas para animais e plantas que possuem os genes *purE* e *purK*.

Filo/Classe	Espécie
Patógenos de Animais	
Bacilli	<i>Bacillus anthracis</i> str. Ames <i>Bacillus cereus</i> <i>Listeria ivanovii</i> subsp. <i>ivanovii</i> PAM 55 <i>Listeria monocytogenes</i> EGD-e <i>Staphylococcus aureus</i> RF122 <i>Staphylococcus epidermidis</i> ATCC 12228 <i>Staphylococcus haemolyticus</i> JCSC1435 <i>Staphylococcus lugdunensis</i> HKU09-01 <i>Staphylococcus pseudintermedius</i> ED99 <i>Staphylococcus saprophyticus</i> ATCC15305 <i>Aerococcus urinae</i> ACS-120-V-Col10a <i>Enterococcus faecalis</i> V583 <i>Enterococcus faecium</i> Aus0004 <i>Lactococcus garvieae</i> ATCC 49156 <i>Streptococcus agalactiae</i> 2603V/R <i>Streptococcus dysgalactiae</i> subsp. <i>equisimilis</i> GGS_124 <i>Streptococcus equi</i> subsp. <i>equi</i> 4047

Streptococcus gallolyticus subsp. *gallolyticus* ATCC 43143
Streptococcus gordonii str. Challis substr. CH1
Streptococcus mitis B6
Streptococcus mutans UA159
Streptococcus oralis Uo5
Streptococcus parasanguinis ATCC 15912
Streptococcus parauberis KCTC 11537
Streptococcus pasteurianus ATCC 43144
Streptococcus pneumoniae 70585
Streptococcus pyogenes MGAS315
Streptococcus salivarius 57.I
Streptococcus sanguinis SK36
Streptococcus suis BM407
Streptococcus uberis 0140J
Actinobacteria
Tropheryma whippelii str. Twist
Corynebacterium aurimucosum ATCC 700975
Corynebacterium diphtheriae NCTC 13129
Corynebacterium jeikeium K411
Corynebacterium kroppenstedtii DSM 44385
Corynebacterium pseudotuberculosis 1/06-A
Corynebacterium resistens DSM 45100
Corynebacterium ulcerans 809
Corynebacterium urealyticum DSM 7109
Kytococcus sedentarius DSM 20547
Gordonia bronchialis DSM 43247
Renibacterium salmoninarum ATCC 33209
Rothia dentocariosa ATCC 17931
Rothia mucilaginosa DY-18
Mycobacterium abscessus ATCC 19977
Mycobacterium avium 104
Mycobacterium bovis AF2122/97
Mycobacterium canettii CIPT 140010059
Mycobacterium intracellulare MOTT-64
Mycobacterium leprae TN
Mycobacterium marinum M
Mycobacterium sp. JDM601
Mycobacterium tuberculosis KZN 1435
Mycobacterium ulcerans Agy99
Nocardia cyriacigeorgica GUH-2
Nocardia farcinica IFM 10152
Rhodococcus equi 103S
Nocardiosis dassonvillei subsp. *dassonvillei* DSM 43111
Propionibacterium acnes KPA171202
Saccharomonospora viridis DSM 43017
Segniliparus rotundus DSM 44985
Tsakamurella paurometabola DSM 20162

Spirochaetia *Leptospira borgpetersenii* serovar *Hardjo-bovis* L550
Leptospira interrogans serovar *Lai* str. 56601

Alphaproteobacteria *Bartonella bacilliformis* KC583
Bartonella clarridgeiae 73
Bartonella grahamii as4aup
Bartonella henselae str. *Houston-1*
Bartonella quintana str. *Toulouse*
Bartonella tribocorum CIP 105476
Brucella abortus bv. 1 str. 9-941
Brucella canis ATCC 23365
Brucella melitensis ATCC 23457
Brucella microti CCM 4915
Brucella ovis ATCC 25840
Brucella pinnipedialis B2/94
Brucella suis ATCC 23445
Ochrobactrum anthropi ATCC 49188
Granulibacter bethesdensis CGDNIH1
Anaplasma centrale str. *Israel*
Anaplasma marginale str. *St. Maries*
Anaplasma phagocytophilum HZ
Ehrlichia canis str. *Jake*
Ehrlichia chaffeensis str. *Arkansas*
Ehrlichia ruminantium str. *Welgevonden*
Neorickettsia risticii str. *Illinois*
Neorickettsia sennetsu str. *Miyayama*

Betaproteobacteria *Bordetella avium* 197N
Bordetella bronchiseptica RB50
Bordetella parapertussis 12822
Bordetella pertussis *Tohama I*
Taylorella asinigenitalis MCE3
Burkholderia cenocepacia J2315
Burkholderia mallei ATCC 23344
Burkholderia multivorans ATCC 17616
Burkholderia pseudomallei MSHR346
Ralstonia pickettii 12D
Laribacter hongkongensis HLHK9
Neisseria gonorrhoeae FA 1090
Neisseria meningitidis FAM18

Gammaproteobacteria *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966
Aeromonas salmonicida subsp. *salmonicida* A449
Aeromonas veronii B565
Alkalilimnicola ehrlichei MLHE-1
Citrobacter koseri ATCC BAA-895
Cronobacter sakazakii ATCC BAA-894
Cronobacter turicensis z3032
Edwardsiella ictaluri 93-146

Edwardsiella tarda EIB202
Enterobacter aerogenes KCTC 2190
Enterobacter cloacae SCF1
Klebsiella pneumoniae 342
Pectobacterium wasabiae WPP163
Photobacterium asymbiotica subsp. ATCC 43949
Proteus mirabilis HI4320
Salmonella bongori NCTC 12419
Salmonella enterica subsp. Typhi str. CT18
Salmonella typhimurium LT2
Shigella boydii CDC 3083-94
Shigella dysenteriae Sd197
Shigella flexneri 5 str. 8401
Shigella sonnei Ss046
Yersinia enterocolitica subsp. enterocolitica 8081
Yersinia pestis KIM
Yersinia pseudotuberculosis IP 32953
Legionella pneumophila str. Corby
Actinobacillus pleuropneumoniae serovar 3 str. JL03
Aggregatibacter aphrophilus NJ8700
Haemophilus influenzae PittEE
Haemophilus parainfluenzae T3T1
Haemophilus parasuis SH0165
Pasteurella multocida subsp. multocida str. Pm70
Acinetobacter baumannii SDF
Acinetobacter calcoaceticus PHEA-2
Moraxella catarrhalis RH4
Pseudomonas aeruginosa PAO1
Pseudomonas syringae pv. phaseolicola 1448A
Francisella philomiragia subsp. philomiragia ATCC 25017
Francisella sp. TX077308
Francisella tularensis subsp. tularensis WY96-3418
Vibrio cholerae M66-2
Vibrio parahaemolyticus RIMD 2210633
Vibrio vulnificus CMCP6

Patógenos de Plantas

Actinobacteria	<i>Clavibacter michiganensis</i> subsp. <i>sepedonicus</i> <i>Leifsonia xyli</i> subsp. <i>xyli</i> str. CTCB07
Alphaproteobacteria	<i>Agrobacterium tumefaciens</i> str. C58 <i>Agrobacterium vitis</i> S4 <i>Candidatus Liberibacter asiaticus</i> str. psy62 <i>Candidatus Liberibacter solanacearum</i> CLso-ZC1
Betaproteobacteria	<i>Burkholderia gladioli</i> BSR3 <i>Burkholderia glumae</i> BGR1 <i>Ralstonia solanacearum</i> GMI1000 <i>Acidovorax citrulli</i> AAC00-1

Gammaproteobacteria *Dickeya dadantii* Ech703
 Dickeya zeae Ech1591
 Erwinia amylovora ATCC 49946
 Erwinia pyrifoliae Ep1/96
 Erwinia sp. Ejp617
 Pectobacterium carotovorum subsp. *carotovorum* PC1
 Xanthomonas axonopodis pv. *citri* str. 306
 Xanthomonas campestris pv. *campestris* str. ATCC 33913
 Xanthomonas oryzae pv. *oryzae* KACC10331
