Robson Francisco de Souza

Estudos evolutivos do divisoma, um complexo
multiprotéico responsável pela divisão bacteriana

*Tese apresentada ao
Instituto de Química da Universidade de São Paulo para
obtenção do Título de
Doutor em Ciências Biológicas (Bioquímica)*

*Orientador: Frederico José Gueiros-Filho
Co-Orientador: Sandro José de Souza*

São Paulo
10/07/2007

Robson Francisco de Souza

Estudos evolutivos do divisoma, um complexo multiprotéico responsável pela divisão bacteriana

*Tese apresentada ao Instituto de Química da
Universidade de São Paulo para obtenção do Título de
Doutor em Ciências Biológicas (Bioquímica)*

Aprovado em: _____

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Resumo

de Souza, R. F. Estudos evolutivos do divisoma, um complexo multiprotéico responsável pela divisão bacteriana (2007). 140p. Tese de Doutorado, Programa de Pós-Graduação em Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo.

O mecanismo de divisão mais comum entre procariotos é a divisão binária, na qual a célula-mãe reparte seu genoma e conteúdo citoplasmático de forma igual entre duas células filhas. Esse processo é mediado por um complexo protéico especializado, chamado divisoma, composto por cerca de 20 proteínas, que promovem a constrição da parede celular e membrana citoplasmática, formando o septo de divisão. O complexo é organizado em torno do anel Z, uma estrutura em anel composta pela proteína FtsZ, um homólogo de tubulina presente na maioria dos procariotos e em algumas organelas de eucariotos. Partindo de um levantamento detalhado da distribuição dos genes do divisoma em genomas completos de procariotos, aplicamos métodos de máxima verossimilhança para inferência de estados ancestrais e reconstruímos o conteúdo gênico do divisoma no último ancestral comum das bactérias atuais. Estendendo essas análises com a aplicação de métodos filogenéticos, inferimos os eventos responsáveis pelas variações de composição deste complexo, observadas entre os diferentes grupos de bactérias. Nossos resultados mostram que o último ancestral comum de todas as bactérias já possuía a maior parte dos componentes conhecidos do divisoma, sugerindo a existência de uma parede de peptidoglicano e a presença de um aparato molecular tão ou mais complexo que o observado nas linhagens atuais, incluindo a presença de componentes considerados acessórios e de distribuição relativamente restrita, como as proteínas envolvidas na localização do anel Z (sistema Min)

e alguns efetores positivos da polimerização de FtsZ. Observamos também que a evolução do complexo não foi muito afetada por eventos de transferência lateral, mas apresenta vários exemplos de perda de genes, em especial em linhagens com genoma reduzido, o que sugere a redundância de vários componentes já presentes no ancestral e a freqüente redução da complexidade, pelo menos dos componentes centrais do divisoma. Episódios de expansão de famílias de componentes do divisoma em linhagens específicas e os mecanismos evolutivos responsáveis pela incorporação de tais variações são discutidos. A caracterização da história evolutiva detalhada do divisoma, aqui apresentada, poderá servir como ponto de partida para novas análises evolutivas e como base para elaboração de experimentos funcionais.

Palavras-chave: divisão celular, *Bacillus*, evolução molecular, estados ancestrais, máxima verossimilhança, divisoma.

Abstract

de Souza, R. F. Evolutionary studies of the divisome, a multiprotein complex responsible for bacterial division (2007). 140p. PhD Thesis, Programa de Pós-Graduação em Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo.

The most common cell division mechanism among prokaryotes is binary fission, where a mother cell partitions its cytoplasm and genome equally among two daughter cells. This process is mediated by a specialized protein complex, known as the divisome, composed of around 20 proteins, that promotes constriction of the cell wall and cytoplasmic membrane, thus forming the division septa. The complex is organized around the Z-ring, a ring-shaped structure composed by FtsZ, a tubulin homolog present in most prokaryotes and some eukaryotic organelles. After a detailed revision of the distribution of divisome genes among completely sequenced prokaryotic genomes, we applied maximum likelihood methods for the inference of ancestral states and reconstructed the gene content of the divisome in the last common ancestor of all extant bacteria. We then performed phylogenetic analysis of all cell division genes and inferred the series of events responsible for the observed variations of the complex's composition among bacterial lineages and their common ancestor. Our results show that the last common ancestor of all bacteria already possessed most of the known divisome components, thus suggesting the existence of a peptidoglycan cell wall and the presence of a molecular apparatus, perhaps more complex than those found in extant bacteria, including the presence of some accessory components with a somewhat restricted distribution, like the proteins involved in the localization of the Z-ring (Min system) and some positive effectors of FtsZ polymerization. We also observed that the

complex's evolution was almost never the subject of horizontal gene transfer events, but shows several examples of gene loss, specially in lineages displaying clear signs of genome reduction, thus suggesting the redundancy of several components in the ancestral divisome and a certain degree complexity reduction, at least for core components of the divisome. Lineage specific expansion of divisome component and the evolutionary mechanisms behind such processes are discussed. This characterization of the detailed evolutionary history of the divisome might serve as a starting point for new evolutionary analysis and as a basis for the design of functional experiments.

Keywords: cell division, *Bacillus*, molecular evolution, ancestral states, maximum likelihood, divisome.

SUMÁRIO

1.	Introdução.....	8
1.1	Divisão celular em procariotos.....	8
1.1.1	Proteínas do divisoma.....	9
1.1.2	Montagem do complexo.....	14
1.1.3	Seleção da posição do septo.....	15
1.2	Reconstrução de genomas ancestrais.....	19
1.3	Evolução de complexos protéicos.....	22
1.4	Evolução do divisoma.....	23
1.5	Objetivos.....	25
2.	Materiais e Métodos.....	27
2.1	Detecção de homólogos.....	27
2.2	Buscas iterativas com sequências intermediárias.....	28
2.3	Construção de alinhamentos e inferência filogenética.....	30
2.4	Contexto genômico.....	33
2.5	Inferência de estados ancestrais.....	35
3.	Resultados.....	39
3.1	Distribuição dos genes de divisão em genomas completos.....	39
3.2	Conteúdo gênico nos genomas ancestrais.....	42
3.3	Filogenias dos genes de divisão.....	50
3.3.1	FtsZ.....	50
3.3.2	FtsK.....	52
3.3.3	FtsX e FtsE.....	56
3.3.4	FtsQ.....	59
3.3.5	FtsA.....	61
3.3.6	FtsB e FtsL.....	65
3.3.7	FtsW e FtsI.....	66
3.3.8	ZapA.....	69
3.3.9	YlmF.....	69
3.3.10	SulA.....	72
3.3.11	FtsN.....	74
3.3.12	ZipA.....	76
3.3.13	EzrA.....	78
3.3.14	AmiC.....	79
3.3.15	SlmA, YyaA e MipZ (oclusão pelo nucleóide).....	85
3.3.16	DivIVA.....	90
3.3.17	Sistema Min.....	92
4.	Discussão.....	95
4.1	Um catálogo da distribuição dos genes de divisão.....	95
4.2	Divisoma mínimo.....	97
4.3	Divisoma ancestral.....	103
4.4	Considerações sobre os métodos utilizados.....	105
4.5	Transferência Horizontal de Genes (HGT).....	109
4.6	Perda de Genes.....	113
4.7	Genes exclusivos e duplicações de genes.....	117
5.	Conclusões.....	125
6.	Referências.....	127

1. Introdução

1.1 Divisão celular em procariotos

A maioria dos procariotos se reproduz através do mecanismo conhecido como fissão binária (Angert, 2005). A fissão binária consiste na separação, na maioria dos casos simétrica, dos componentes celulares da célula-mãe, incluindo seu material genético, em duas células-filhas viáveis. Esse seria, em princípio, um processo simples, exigindo apenas o crescimento da célula-mãe até o dobro do seu tamanho original e a partição da célula em duas metades iguais. No entanto, para que sejam bem-sucedidas, as células precisam garantir que a partilha dos componentes celulares seja feita sem erros e esteja coordenada com o ciclo celular, ocorrendo no momento oportuno após a replicação dos cromossomos. Essas restrições estimularam a evolução de sistemas sofisticados para execução da fissão binária (Guertin, Trautmann *et al.*, 2002; Angert, 2005). Em eucariotos, a divisão binária atinge esses objetivos através da re-organização do citoesqueleto e da formação de estruturas especializadas, como centríolos e fusos mitóticos, que garantem a simetria da partição das organelas e do material genético entre as células-filhas, e o anel de actina-miosina, que promove a força motora para citocinese (Guertin, Trautmann *et al.*, 2002). Em bactérias e arqueobactérias, proteínas especializadas, homólogas de proteínas do citoesqueleto de eucariotos (Michie e Lowe, 2006), cooperam para formar, na região mediana da célula, uma estrutura análoga ao anel de actina-miosina, conhecido como **septossomo** ou **divisoma**. O divisoma é responsável pela escolha do local e do plano onde ocorrerá a constrição e invaginação do envelope bacteriano (membrana citoplasmática, parede celular e, no caso de bactérias Gram-negativas, membrana externa) (Margolin, 2000). Apesar das semelhanças estruturais entre o divisoma bacteriano e o anel de actina-miosina, nenhum estudo realizado até o momento foi capaz de demonstrar um papel motor

para o divisoma, sendo possível que ele opere através do reposicionamento da maquinaria de síntese do peptidoglicano e mudança do plano de síntese do envelope (Errington, Daniel *et al.*, 2003; Vicente, Rico *et al.*, 2006).

1.1.1 Proteínas do divisoma

As primeiras proteínas de divisão celular descobertas em procariotos foram identificadas a partir do isolamento de mutantes que eram incapazes de se dividir se cultivados em certas condições, mas viáveis em outras (Rothfield, Justice *et al.*, 1999). Várias dessas proteínas receberam nomes iniciados com **fts** (“filamentous temperature sensitive”), como FtsA ou FtsK, por que a divisão nos mutantes condicionais estudados era bloqueada em temperatura maior que a permissiva. Na última década, várias outras proteínas de divisão foram identificadas em procariotos pela aplicação de métodos e tecnologias novos, como a biologia computacional, a microscopia de fluorescência, experimentos bioquímicos para detectar interações físicas entre proteínas, triagens para mutações sintéticas letais ou ensaios de duplo híbrido (Goehring e Beckwith, 2005).

A Figura 1 mostra modelos esquemáticos para os complexos de divisão celular em *Escherichia coli* e *Bacillus subtilis*, representados em associação com a membrana citoplasmática e a parede celular. Em todos os organismos já estudados (*e.g.* *E. coli*, *S. elongatus*, *S. aureus*, *S. coelicolor* ou *B. subtilis*), a formação do septo é iniciada pela polimerização da proteína FtsZ, uma GTPase da superfamília das tubulinas (Nogales, Downing *et al.*, 1998). Como as tubulinas, FtsZ é citoplasmática e forma polímeros que, durante a divisão, se organizam em uma estrutura conhecida como anel Z (painéis C e E na Figura 1), com a qual as demais proteínas de divisão interagem para compor um divisoma funcional. Normalmente, o anel Z se forma na região mediana da célula e permanece associado à face interna da membrana citoplasmática até o final da divisão (Bi e

Lutkenhaus, 1991). Esse anel não é rígido, mas mantido pela rápida reciclagem dos monômeros de FtsZ, que se desagregam quando o GTP é hidrolizado (Lowe e Amos, 1998). A polimerização de FtsZ pode ocorrer espontaneamente *in-vitro* e depende de interações entre o domínio C-terminal de uma subunidade e a região N-terminal de outra, a hidrólise de GTP ocorrendo na interface (Graumann, 2007). *In-vivo*, porém, a formação do anel Z é regulada por proteínas que ligam diretamente FtsZ. Em *Bacillus* e *E. coli*, a montagem do anel Z é auxiliada por FtsA, um homólogo de actina, e pela proteína acessória ZapA. ZapA é uma proteína pequena que promove a formação de feixes do polímero de FtsZ *in-vitro* e desempenha um possível papel regulatório durante a formação do anel (Gueiros-Filho e Losick, 2002). Em *E. coli*, uma terceira proteína, ZipA, é necessária para a maturação do divisoma e para a associação de FtsZ com a membrana (Pichoff e Lutkenhaus, 2005). A interação de FtsZ com ZipA e FtsA ocorre através de uma cauda não-estruturada na região C-terminal de FtsZ, mas a abundância relativa de FtsZ, presente na célula em quantidades muito maiores que seus parceiros, viabiliza a incorporação simultânea das duas proteínas ao divisoma (Margolin, 2005). Embora FtsA e ZipA possam induzir a formação do anel Z independentemente, as duas proteínas são necessárias para o recrutamento das proteínas subsequentes para o divisoma (Pichoff e Lutkenhaus, 2002), mas a substituição de um aminoácido em FtsA (R286W) torna ZipA dispensável (Geissler, Elraheb et al., 2003). Além disso, uma hélice anfipática na extremidade C-terminal de FtsA promove a associação de FtsA com a membrana e é necessária para a localização dessa proteína no sítio de divisão, podendo também contribuir para o direcionamento de FtsZ para a membrana (Pichoff e Lutkenhaus, 2005). Em *Bacillus*, a ausência de FtsA pode ser compensada pela super-expressão da proteína SepF (YlmF), um componente citoplasmático do divisoma de *Bacillus* que não existe em *E. coli* e que produz um fenótipo letal sintético, quando inativado em linhagens que não expressam FtsA (Ishikawa, Kawai et al., 2006).

As proteínas cuja interação com FtsZ discutimos acima (FtsA, YlmF, ZipA e ZapA) são, pelo menos em condições normais, efetores positivos da formação do anel. A regulação da montagem e dinâmica do anel Z, porém, também requer a participação de proteínas que atuam como efetores negativos, que bloqueiam ou limitam a formação do anel Z. Exemplos de efetores negativos são as proteínas SulA, EzrA e MinC. MinC forma um complexo com a ATPase MinD, associado à membrana, que inibe a formação do anel Z nos pólos da célula e será discutido mais abaixo (seção 1.1.3), quando considerarmos o papel do sistema Min e de outros inibidores (Noc e SlmA) na localização do anel Z. SulA é expressa durante a resposta SOS e inibe a formação do anel ao se ligar diretamente a FtsZ, bloqueando sua polimerização (Cordell, Robinson *et al.*, 2003). EzrA contém uma única região transmembrana e é homoganeamente distribuída em toda a membrana celular, mas localiza-se preferencialmente no septo, após sua formação (Gueiros-Filho, 2007). *In-vitro*, EzrA se liga a moléculas de FtsZ livres, impedindo sua polimerização, mas é incapaz de desfazer polímeros já formados (Haeusser, Schwartz *et al.*, 2004), o que parece explicar como a localização tardia de EzrA no septo não perturba o anel. O papel exato de EzrA na divisão ainda não é claro, mas é possível que a proteína atue reduzindo a tendência de FtsZ para formar polímeros (Levin, Kurtser *et al.*, 1999).

A proteína FtsK é membro de uma família de DNA translocases envolvidas em segregação de cromossomos e conjugação de elementos móveis (Lakshminarayan, Makarova *et al.*, 2004). Em *Bacillus*, a proteína ortóloga de FtsK é SpoIIIE, que é essencial para o processo de esporulação mas não durante a divisão vegetativa. Membros dessa família em geral apresentam quatro domínios transmembrana seguidos de um domínio citoplasmático que age como uma bomba de DNA alimentada pela hidrólise de ATP (Graumann, 2007). Os domínios transmembrana de FtsK têm papel relevante na formação do septo em *E. coli* e essa função pode ser isolada da atividade de DNA translocase do domínio citoplasmático (Draper, McLennan *et al.*, 1998; Bigot, Corre *et al.*, 2004). O

domínio citoplasmático, além de completar a segregação dos cromossomos que ficam presos no septo, atua, pelo menos em *E. coli*, no recrutamento das resolvases XerC e XerD, que fazem a separação de eventuais dímeros cromossômicos, formados durante a replicação (Aussel, Barre *et al.*, 2002). As proteínas dessa família constituem, portanto, uma ligação importante entre os sistemas de replicação e segregação do DNA e a citocinese.

Algumas das proteínas incluídas nos modelos da Figura 1 têm funções menos conhecidas. FtsE e FtsX são os dois componentes (par cognato) de um sistema transportador do tipo ABC (“ATP binding cassette”) que, quando inativado, prejudica a divisão. A atividade molecular (alvo do transporte) de FtsE/X ainda é desconhecida, mas é sabido que a inativação desses genes possa ser compensada em condições de alta osmolaridade (Schmidt, Peterson *et al.*, 2004). FtsQ, FtsB e FtsL são proteínas bitópicas com uma região curta exposta ao citoplasma, um único segmento transmembrana e um domínio periplasmático grande. Essas três proteínas formam um complexo ternário, pois a localização de uma delas no divisoma depende da localização das outras (Buddelmeijer e Beckwith, 2004). Embora este complexo possa desempenhar outras funções, o único papel proposto para ele é o de intermediário nas interações entre as proteínas do periplasma e os componentes citoplasmáticos do divisoma (Graumann, 2007).

Uma etapa fundamental na construção do septo é o redirecionamento da parede celular de peptidoglicano, que exige a atividade de enzimas especializadas em sua remodelagem, pois a parede deve ser simultaneamente clivada, reconstruída e estendida (Scheffers, 2007). A proteína FtsI é uma PBP (“penicillin binding protein”) da classe B (alto peso molecular e sem atividade de transglicosilase), que cataliza as reações de transpeptidação dos precursores nas vias de síntese do peptidoglicano, e atua especificamente no septo (Margolin, 2000). A atividade de FtsI está associada à atividade de FtsW, um possível transportador da família de proteínas SEDS (“shape, elongation, division and sporulation”), que alguns autores propuseram realizar o transporte dos substratos de FtsI através da

membrana celular (Vicente, Rico *et al.*, 2006). Os transportadores SEDS e as PBPs são comumente associados em pares e foram implicados em outros processos relacionados à parede celular de bactérias, como a formação da parede do esporo e o crescimento (Graumann, 2007). Da mesma forma que FtsI está especializada na síntese do peptidoglicano septal, a mureína hidrolase AmiC, de *E. coli*, está especializada na sua degradação. Esta enzima é o primeiro componente conhecido do divisoma e se localiza inteiramente no periplasma, para onde é transportada pelo sistema tat ("twin-arginine transport system") e se concentra na região do septo (Bernhardt e De Boer, 2003). AmiC catalisa a hidrólise das ligações entre o ácido N-acetilmurâmico e a L-alanina no interior do septo, que unem as pontes peptídicas ao esqueleto de glicano na parede celular, e sua inativação faz *E. coli* crescer em cadeias de células que se dividem mas não se separam. AmiC depende de outra proteína do divisoma com função pouco conhecida, FtsN, para se localizar. Esta última proteína, que não é encontrada em *Bacillus*, é essencial em *E. coli*.

Tanto em *Bacillus* como em *E. coli*, várias proteínas representadas do divisoma (FtsZ, FtsA, FtsQ, FtsL, FtsI e FtsW) são codificadas por genes localizados em um agrupamento de genes chamado **dcw**, que reúne também vários genes *mur*, da via de síntese do ácido N-acetilmurâmico, precursor do peptidoglicano (Vicente, Tamames *et al.*, 2004). Esse agrupamento não corresponde, porém, a um operon, pois várias origens de transcrição estão presentes nele em *E. coli* e sinais para término da transcrição são encontrados dentro dele em outros organismos (Vicente, Tamames *et al.*, 2004). A vizinhança dos genes encontrados no **dcw** de *E. coli* é preservada mesmo em linhagens de organismos extremamente divergentes (Tamames, González-Moreno *et al.*, 2001). Hipóteses sobre a relevância e a evolução da conservação do **dcw** são discutidas mais adiante.

1.1.2 Montagem do complexo

A montagem do divisoma a partir das proteínas descritas na seção anterior depende das interações entre esses componentes. Em *E. coli*, experimentos genéticos demonstraram a existência de vínculos para localização das proteínas de divisão no septo, com as proteínas atreladas a uma hierarquia de dependências que começa em FtsZ (painel B da Figura 1) (Errington, Daniel *et al.*, 2003). Em *B. subtilis*, os componentes do divisoma parecem ser incorporados simultânea e/ou cooperativamente, ou seja, com a maioria dos componentes dependendo de vários outros para a localização no septo (Graumann, 2007).

A sequência de eventos que levam à montagem do anel em *E. coli* começa com a polimerização de FtsZ e interações independentes com as proteínas que regulam a formação do anel *in-vivo* (FtsA, ZapA e ZipA). Essas proteínas formam juntas o que se convencionou chamar de “proto-anel citoplasmático” (Vicente e Rico, 2006), um sub-componente do divisoma composto por FtsZ e as proteínas que interagem diretamente com FtsZ. O proto-anel é montado nas etapas iniciais da maturação do divisoma (Vicente e Rico, 2006), às quais se segue a incorporação das proteínas FtsE, FtsX e FtsK, os primeiros componentes do anel com domínios que se inserem na membrana citoplasmática. Em *E. coli*, FtsK é incorporada ao divisoma antes de uma pausa na localização das proteínas seguintes (Aarsman, Piette *et al.*, 2005), permitindo uma distinção entre uma fase tardia e outra inicial na montagem do divisoma (Vicente e Rico, 2006). A montagem então prossegue com a incorporação das proteínas FtsQ, FtsB e FtsL, que formam um sub-complexo independente denominado “ponte periplasmática” (Vicente e Rico, 2006), que se supõe servir como conexão entre o proto-anel e o complexo transmembrana de síntese de peptideoglicano (“peptidoglycan factory”). Este último complexo é composto pelas enzimas FtsW, FtsI e AmiC, todas especializadas na síntese do peptideoglicano septal e dependentes da ponte periplasmática para localização. Muito embora, como salientado

acima, *Bacillus* aparentemente não apresenta a mesma hierarquia de vínculos entre os componentes do divisoma descrita para *E. coli*, a mesma classificação funcional, que separa componentes primariamente citoplasmáticos (“proto-anel”), integrados à membrana (“ponte periplasmática”) e associados ao peptidoglicano (“fábrica de peptidoglicano”) será considerada mais adiante em nossas discussões.

1.1.3 Seleção da posição do septo

Além da classificação funcional apresentada na seção anterior para as proteínas que integram o divisoma propriamente dito, uma outra classe de proteínas de divisão, responsável pela determinação do local onde será formado o anel Z, pode ser introduzida. A escolha do sítio de divisão é um dos passos críticos para a fissão binária, pois limita a capacidade do processo de gerar uma descendência viável (Lutkenhaus, 2007). Tanto em *Bacillus* como em *E. coli* foram identificados dois mecanismos para a seleção do sítio de divisão: o sistema Min e a oclusão pelo nucleóide. Duas proteínas do sistema Min, MinC e MinD, formam um heterodímero associado à membrana que inibe a formação do anel Z nos pólos da célula (Margolin, 2000). Essa inibição é exercida pelo domínio C-terminal de MinC, enquanto o domínio N-terminal é responsável pela interação com MinD (Shiomi e Margolin, 2007). MinD é membro da superfamília de ATPases que inclui ParA e proteínas relacionadas e se associa à membrana através de uma hélice anfipática localizada em sua extremidade C-terminal (Hu e Lutkenhaus, 2003). A distribuição do complexo MinCD determina as regiões livres do efeito inibidor de MinC, ou seja, que áreas da célula estão disponíveis para formação do anel Z (Figura 1, painéis C e E). Essa distribuição é determinada por proteínas e mecanismos diferentes em *E. coli* e *B. subtilis*. Em *B. subtilis* a proteína DivIVA se associa estaticamente aos pólos e com MinD, seqüestrando essa última para os pólos (Rothfield, Taghbalout *et al.*, 2005). Em *E. coli*, o complexo MinCD forma

uma capa de polímeros que cobrem um dos pólos da célula e se estendem para a região medial, mas esta estrutura oscila rapidamente entre os pólos, com uma periodicidade de aproximadamente ~40s, que começa reduzindo a região ocupada a partir do meio da célula, na direção do pólo. À medida que um pólo fica livre do polímero de MinCD, o outro pólo é ocupado pelas moléculas que foram liberadas (Lutkenhaus, 2007). A oscilação de MinCD é dependente da proteína MinE que, em *E. coli*, é codificada pelo terceiro gene do operon Min e também oscila, formando um anel a periferia da região ocupada por MinCD. DivIVA e MinE são ambos fatores que determinam a especificidade topológica do sistema Min, mas operam por mecanismos diferentes, em organismos diferentes, estimulando a questão da origem e ancestralidade de cada uma dessas alternativas.

O sistema Min só inibe a formação do anel Z nos pólos e a observação de que, sob condições normais, o anel Z não se forma em regiões da célula ocupadas pelo cromossomo levou à proposição de que um mecanismo adicional, a oclusão pelo nucleóide (“nucleoid occlusion” ou NO), impediria a formação do anel nessas regiões (Woldringh, Mulder *et al.*, 1991). Por muito tempo, o mecanismo por trás da oclusão pelo nucleóide permaneceu obscuro, até a descoberta de proteínas em *E. coli* (SlmA), *B. subtilis* (Noc) e *C. crescentus* (MipZ), que interagem com o DNA e inibem a polimerização do anel Z (Wu e Errington, 2004; Bernhardt e De Boer, 2005; Thanbichler e Shapiro, 2006). SlmA e Noc possuem domínios de interação não específica com o DNA, e são capazes impedir a formação do anel aparentemente sem o auxílio de outras proteínas. Já MipZ é um homólogo de ParA e, como essa proteína, depende de ParB para interagir com o cromossomo e promover a oclusão. Devido à interação com ParB, MipZ acompanha a segregação das origens de replicação, que se separam após a duplicação do cromossomo e são direcionadas para pólos opostos da célula. Assim, um gradiente de inibidor é formado dos pólos para o centro da célula. O gradiente de MipZ é outro mecanismo pelo qual uma região livre de inibidor é definida no centro da célula, análogo aos efeitos topológicos de MinE e DivIVA. MipZ

executa, portanto, um efeito de inibição polar sobre o anel Z e, ao mesmo tempo, media a oclusão pelo nucleóide, substituindo o papel do sistema Min e das proteínas de NO em *Caulobacter* (Thanbichler e Shapiro, 2006).

Figura 1: Modelo esquemático do complexo de divisão celular de *Escherichia coli* (A-C) e *Bacillus subtilis* (D,E). (A) e (D) mostram visões hipotéticas da organização do complexo de proteínas de divisão (divisoma) no sítio onde ocorrerá a formação do novo septo das células representadas em (C) e (E). (C) e (E) indicam também as localizações dos componentes do sistema Min (MinC, MinD e MinE), das proteínas SlmA e Noc (responsáveis pelo efeito oclusão do nucleóide) e da proteína SulA de *E. coli* (citoplasma), que forma um complexo com FtsZ, impedindo a formação do anel Z durante a resposta SOS. (B) relações de dependência para localização das proteínas do divisoma em *E. coli* (Margolin, 2000), indicadas por setas que apontam da proteína necessária para a dependente, indicando que esta última não se localiza no septo na ausência da primeira. Em (A) e (D), números indicam os sub-componentes do divisoma, propostos recentemente (Vicente e Rico, 2006): (1) Anel citoplasmático, composto pelas proteínas FtsZ, FtsA, ZapA, ZipA; (2) ponte periplasmática (FtsQ, FtsB e FtsL) e (3) fábrica de peptidoglicano (FtsW, FtsI, FtsN e AmiC).

1.2 Reconstrução de genomas ancestrais

A escassez de informações diretas e detalhadas sobre a história da vida têm frustrado os evolucionistas, pois para entender a evolução biológica é necessário saber não apenas os estados dos caracteres nos organismos atuais, mas também em seus ancestrais (Cunningham, Omland *et al.*, 1998). Embora os fósseis revelem exemplos de transformações evolutivas, muitas linhagens e tipos de caracteres são mal representados ou estão ausentes desse registro, levando os pesquisadores a utilizar métodos para inferência de caracteres nos ancestrais a partir da distribuição desses caracteres nos organismos atuais (Pagel, 1999a). Estudos comparativos baseados na comparação das espécies atuais com fósseis ou na reconstrução de caracteres ancestrais são chamados **comparações direcionais**, em contraposição às abordagens **não-direcionais**, que analisam as distribuições de caracteres por comparações diretas das espécies extantes ou de grupos relativamente independentes de espécies extantes, identificados com o auxílio de filogenias. Por permitirem, ao mesmo tempo, a avaliação tanto dos caracteres nas espécies extantes, quanto das transformações acumuladas dos ancestrais para os descendentes, os métodos direcionais permitem estabelecer a direção, a ordem e as taxas de transformação dos caracteres ao longo da evolução, além de correlações entre caracteres, que são o foco dos métodos não-direcionais (Harvey e Pagel, 1991). Pela riqueza de detalhes que podem fornecer, as comparações direcionais são os métodos ideais para a detecção de viéses associados a pressões seletivas ou a contingências históricas (Pagel, 1994). Exemplos de estudos que se beneficiam da análise dos estados ancestrais abordam problemas como a influência da adaptação a um estilo de vida parasitário sobre as taxa de evolução dos genes (Blanc, Ogata *et al.*, 2007) ou a convergência de estruturas complexas, como os olhos de artrópodes que parecem ter origens independentes em diferentes linhagens, ao invés de uma origem comum (Oakley e Cunningham, 2002). Em seus trabalhos, Harvey e Pagel (Harvey

e Pagel, 1991; Pagel, 1994; , 1999a) discutem os métodos empregados nesses estudos e dão vários exemplos de análises direcionais e não-direcionais, comparando-as às abordagens que não usam a filogenia dos organismos para distinguir entre correlações reais e as impostas pela descendência de um ancestral comum (Pagel, 1994). Embora a qualidade das inferências dos métodos comparativos direcionais seja sensível tanto aos dados quanto aos métodos de inferência de filogenias e reconstrução dos caracteres ancestrais (Cunningham, Omland *et al.*, 1998), a área tem experimentado desenvolvimentos expressivos nos últimos anos, estimulada principalmente pelo desenvolvimento e popularização de métodos baseados em modelos estatísticos explícitos, como os métodos de máxima verossimilhança e métodos bayesianos (Ronquist, 2004).

Além das análises morfológicas e ecológicas, nas quais são aplicados desde meados da década de 1970, os métodos de reconstrução de estados ancestrais encontraram, na última década, aplicações no estudo da evolução de genomas completos (Ouzounis, 2005). Desde a publicação do segundo genoma completo de bactérias (Fraser, Gocayne *et al.*, 1995) muitas análises comparativas de genomas completos foram publicadas, incluindo estudos que revelaram a presença de grandes rearranjos cromossômicos entre bactérias relacionadas (Eisen, Heidelberg *et al.*, 2000), variações de conteúdo gênico entre linhagens da mesma ou de espécies próximas (Perna, Plunkett *et al.*, 2001; Da Silva, Ferro *et al.*, 2002) e a conservação da ordem ou vizinhança em regiões curtas em torno dos genes em procariotos (Tamames, 2001). Esses e outros estudos foram pioneiros ao apontar a importância de processos antes tidos como marginais para a evolução do conteúdo gênico, como a transferência lateral (“horizontal gene transfer” ou HGT) (Eisen, 2000) e a evolução redutiva de genomas, incluindo a aceleração da perda de genes em linhagens de patógenos (Mira, Ochman *et al.*, 2001). As forças até então tidas como dominantes na evolução do conteúdo gênico eram a duplicação de genes e a formação de novos genes, seja por recombinação de domínios ou pelo recrutamento de regiões não codificantes (origem *de*

novo) (Long, Betran *et al.*, 2003). Embora indicassem a importância da HGT na evolução dos genomas, esses estudos não abordaram quantitativamente a contribuição de cada um desses processos para a composição dos genomas das espécies atuais, limitando-se a apontar diferenças no impacto da HGT e da perda de genes em certas linhagens, como *Escherichia* e *Chlamydia* (Perna, Plunkett *et al.*, 2001).

O primeiro estudo a avaliar quantitativamente, com base em dados de genomas completos, o impacto dos processos que determinam o ganho e a perda de genes, foi publicado por Snel *et al.* (Snel, Bork *et al.*, 2002). Esses autores identificaram genes ortólogos em genomas completos de proteobactérias e arqueobactérias e, mapeando sua distribuição na filogenia de cada grupo, reconstruíram os genomas ancestrais e estimaram a contribuição relativa da transferência lateral e dos demais processos nas variações de conteúdo gênico. Estudos mais recentes (Kunin e Ouzounis, 2003; Mirkin, Fenner *et al.*, 2003) estenderam estas análises para novos genomas e formalizaram a abordagem comparativa direcional empregada por Snel *et al.*, baseada na reconstrução dos estados ancestrais pelo princípio de máxima parcimônia (Hillis, Moritz *et al.*, 1996). Outros autores introduziram o uso de modelos estatísticos direcionais para o estudo da evolução do conteúdo gênico (Hahn, De Bie *et al.*, 2005), bem como programas e algoritmos, baseados nestes modelos, para estimar a composição dos genomas ancestrais e o impacto dos diferentes processos evolutivos (De Bie, Cristianini *et al.*, 2006). Modelos com um parâmetro adicional, que representa a taxa de aparecimento de novos genes, discriminam eventos de transferência lateral ou origem *de novo* dos eventos de duplicação gênica e são usados no estudo da evolução do tamanho das famílias gênicas (Karev, Wolf *et al.*, 2002).

A reconstrução de estados ancestrais para genomas completos é uma área em desenvolvimento, na qual técnicas e modelos estão sendo introduzidos não apenas para a evolução do conteúdo gênico, mas para processos como as variações na composição de cromossomos e aminoácidos (Lawrence e Ochman, 1998) e a evolução da regulação da

expressão gênica (Babu, Luscombe *et al.*, 2004; Teichmann e Babu, 2004). Esses estudos estão permitindo identificar as pressões seletivas que operam em níveis acima de genes individuais e que definem a arquitetura da informação genômica e a forma como os genes irão interagir uns com os outros.

1.3 Evolução de complexos protéicos

Paralelamente aos estudos de larga escala baseados em genomas completos, vários grupos realizaram estudos a inferência manual e detalhada da composição ancestral de complexos protéicos individuais e da série de eventos evolutivos que determinaram as variações na composição desses complexos entre espécies (Huynen, Gabaldon *et al.*, 2005). Este tipo de análise evita erros comuns em análises de larga escala, como as falhas na detecção de homólogos, permitindo maior precisão na identificação de ortólogos e na detecção de episódios de substituição por não-homólogos, que devem ser identificados com base na literatura e em dados experimentais. Complementando as análises de larga escala, esses estudos permitem desvendar a origem e pressões seletivas que determinam a evolução desses sistemas, bem como a identificação de subsistemas ou módulos, com histórias semi-independentes, contribuindo para o entendimento de seu funcionamento (Huynen, Gabaldon *et al.*, 2005).

Exemplos de reconstruções da historia evolutiva de sistemas biomoleculares incluem a análise da origem do peroxissomo (Gabaldon, Snel *et al.*, 2006), no qual as relações filogenéticas de proteínas direcionadas para essa organela foram usadas para avaliar a hipótese da origem endossimbiótica do peroxissomo (Latruffe e Vamecq, 2000). As filogenias das proteínas do peroxissomo mostraram que a maioria delas tem origem eucariótica e aquelas que apresentam maior similaridade com proteínas de α -proteobacterias são, na verdade, proteínas mitocondriais recrutadas para o peroxissomo. Em

outro estudo, Mushegian aplicou uma combinação de análises comparativas, dados estruturais e informações compiladas da literatura para inferir a composição do ribossomo no LUCA e propor a composição de um ribossomo mínimo hipotético e, supostamente, funcional (Mushegian, 2005). Exemplos adicionais incluem a reconstrução do complexo NADH:ubiquinona redutase ancestral, que revelou a origem mitocondrial de algumas subunidades e o aumento do número de componentes em eucariotos (Gabaldon, Rainey *et al.*, 2005), a reconstrução de vias metabólicas, como a via glicolítica, vias de síntese de isoprenóides e triptofano e o a via responsável pelo decaimento de mRNAs “truncados” (“non-sense” mediated decay ou NMD) (Huynen, Gabaldon *et al.*, 2005),

A reconstrução de sistemas biomoleculares ancestrais permite avaliar hipóteses quanto à origem e o destino dos genes que codificam os componentes desses sistemas. Cenários alternativos indicaram a perda ou a substituição de genes ausentes, bem como o recrutamento ou a duplicação de componentes para aumento da complexidade do sistema (Huynen, Gabaldon *et al.*, 2005). Tais cenários podem ser avaliados rigorosamente e a influência relativa de cada um dos tipos de fenômenos contabilizados, permitindo avaliar o grau de conservação das interações dos componentes, em geral extrapolada de estudos experimentais em um ou poucos sistemas modelo.

1.4 Evolução do divisoma

A evolução e a organização do divisoma já foram alvos de diferentes abordagens comparativas (Vicente, Tamames *et al.*, 2004). Uma das abordagens envolveu a análise detalhada da história das famílias de alguns genes de divisão com homólogos em outros sistemas (*ftsZ*, *ftsA* e *ftsK*). A semelhança estrutural de FtsZ à tubulina dos eucariotos motivou diversos trabalhos e levou alguns autores a propor um papel para *ftsZ* na origem da tubulina e do citoesqueleto dos eucariotos (Faguy e Doolittle, 1998; Erickson, 2007). O

debate em torno das relações filogenéticas e da função do ancestral de tubulinas e FtsZ continua a ser estimulado pela descoberta em bactérias de genes mais semelhantes à tubulina do que a FtsZ (Pilhofer, Rosati *et al.*, 2007). As relações filogenéticas dos homólogos de *ftsZ* presentes em genomas completos também foram analisadas (Vaughan, Wickstead *et al.*, 2004), permitindo a identificação de duplicações antigas e homólogos distantes, com similaridade reduzida e mutações em resíduos necessários para a hidrólise de GTP. A proteína FtsA também foi alvo de estudos evolutivos, por ser altamente conservada e apresentar similaridade à actina, outra proteína do citoesqueleto de eucariotos (Bork, Sander *et al.*, 1992). Análises recentes da estrutura e sequência de FtsA indicaram nos membros dessa família a presença de uma inserção, ausente em outros membros da superfamília das actinas, e que é provavelmente a região especializada nas interações de FtsA e outras proteínas do divisoma (Anantharaman e Aravind, 2004). A evolução da superfamília de ATPases que inclui os homólogos de *ftsK* foi caracterizada por Lakshminarayan e colaboradores. (Lakshminarayan, Makarova *et al.*, 2004), que propuseram um cenário evolutivo no qual mudanças estruturais adquiridas durante a diversificação dessa superfamília dá origem à variedade estrutural e às diferenças entre famílias desse grupo de ATPases.

Embora a reconstrução de estados ancestrais não tenha sido empregada diretamente na análise do conteúdo gênico do divisoma, estudos baseados numa abordagem semelhante buscaram identificar correlações entre a variação do conteúdo gênico e a morfologia das bactérias. Nesses estudos (Vicente, Tamames *et al.*, 2004), a morfologia das espécies foi comparada com agrupamentos produzidos pela aplicação de métodos de distância e parcimônia ao padrão de presença/ausência dos genes de divisão. Agrupamentos com excesso de espécies com determinada morfologia (cilindro, esfera ou espiral) indicariam um vínculo entre os genes presentes ou ausentes destes organismos e seu fenótipo (Tamames, González-Moreno *et al.*, 2001). Nenhuma correlação, porém, foi identificada

entre o conteúdo gênico do divisoma e a morfologia da célula, exceto nos casos já documentados de vínculo entre as proteínas do sistema Min e a morfologia esférica (Vicente, Tamames *et al.*, 2004). Em outro estudo, Tamames *et al.* empregaram os mesmos métodos de análise, mas substituíram o conteúdo gênico por uma matriz representando o padrão de vizinhança dos genes de divisão, ou seja, a conservação do **dcw** (Tamames, González-Moreno *et al.*, 2001). Dessa vez, eles constataram uma relação entre a conservação do **dcw** e o agrupamento de organismos com morfologias cilíndricas, o que os levou a propor a existência de um mecanismo de “genome channeling” para explicar as vantagens e a ampla conservação do **dcw** em bacilos (Mingorance, Tamames *et al.*, 2004). “Genome channeling” é o acoplamento da transcrição, tradução e transporte de produtos gênicos promovido pela proximidade de genes. Esse processo pode ser vantajoso em procariotos, onde a transcrição e a tradução são acoplados e ocorrem no mesmo compartimento. Nesses organismos, genes que atuam em um mesmo processo molecular e dependem de interações diretas e da co-localização das proteínas codificadas, podem (teoricamente) tirar vantagem de sua proximidade para aumentar a concentração local dessas proteínas, favorecendo as interações necessárias. Apesar de atraente, a hipótese de “genome channeling” dos genes de divisão carece de evidências experimentais.

1.5 Objetivos

Os estudos evolutivos do divisoma ainda não exploraram a questão da origem dos componentes desse sistema ou o destino de genes originados pela duplicação de seus componentes, que podem ou não continuar a interagir com a maquinaria de divisão. Ao contrário de outros sistemas biomoleculares e de alguns estudos sobre genomas inteiros, a incidência de eventos de HGT sobre o divisoma ainda não foi avaliada, mas poderia indicar

quais componentes desse sistema podem ser considerados lábeis, ou seja, substituíveis. Nosso objetivo neste trabalho é estudar a evolução dos componentes do divisoma, buscando uma justificativa biológica para como a distribuição desses genes entre os procariotos foi estabelecida. Além da sequência de eventos históricos, avaliamos o impacto de cada um dos processos evolutivos que alteraram a composição do divisoma desde o último ancestral comum a todas as bactérias. Definir o núcleo conservado de componentes do divisoma, o conjunto de proteínas de um complexo de divisão celular mínimo, associado à síntese de peptideoglicano, e reconstruímos a composição do divisoma no último ancestral comum a todas as bactérias atuais, que denominamos LCAB, para “last common ancessor of all bacteria”.

2. Materiais e Métodos

2.1 Detecção de homólogos

Após uma busca inicial por genes de procariotos associados pelo projeto GOA (Camon, Magrane *et al.*, 2004) ao termo “cytokinesis”, selecionamos, com base na literatura, os genes que codificam proteínas para as quais existem evidências experimentais, em pelo menos um procarioto, de localização preferencial no septo e/ou interação com outros genes de divisão (*amiC*, *ftsA*, *ftsB*, *ftsE*, *ftsI*, *ftsK*, *ftsL*, *ftsN*, *ftsQ*, *ftsW*, *ftsX*, *ftsZ*, *ylmF*, *zipA* e *zapA*). Incluímos também genes que, por interagirem diretamente ou formarem complexos com proteínas que interagem diretamente com proteínas que se localizam no complexo de divisão celular, afetam a montagem e a função do divisoma. Essa última categoria inclui os genes que codificam proteínas envolvidas na determinação do sítio de divisão (*minC*, *minD*, *minE*, *divIVA*, *ezrA*, *mipZ*, *noc* e *slmA*) e proteínas que transduzem sinais de “stress” para a maquinaria de divisão (*sulA*). Ao todo 24 famílias gênicas (segunda linha da Tabela 1) foram selecionadas para análise, com base na literatura de divisão celular em bactérias, todas com membros previamente caracterizados como componentes do divisoma em sistemas modelo como *E. coli*, *B. subtilis*, *C. crescentus*, *S. elongatus* e outros. A distribuição dos genes selecionados em organismos com genomas completamente seqüenciados foi baseada, inicialmente, na classificação de homólogos disponível no banco de dados STRING, versão 6.3 (Von Mering, Jensen *et al.*, 2005). Este banco de dados contém listas de genes ortólogos e parálogos (COGs) para 163 genomas completos de procariotos, a princípio importados do COG (Tatusov, Fedorova *et al.*, 2003) e depois ampliados pelo mesmo procedimento usado para construir os COGs, ou seja, pela identificação de trincas de melhores alinhamentos recíprocos (Von Mering, Jensen *et al.*, 2007). Todas as proteínas nos COGs selecionados foram em seguida inspecionadas, com

base na arquitetura de domínios identificada pelo banco de dados PFam, para caracterizar a estrutura de domínios típica das proteínas do divisoma em cada COG e descartar falsos positivos da classificação inicial. Os membros de cada COG que apresentavam a mesma arquitetura de domínios das proteínas do divisoma foram então usados como sementes (entradas) para buscas iterativas por similaridade (“iterated similarity searches” ou ISS) e os resultados validados com base em análises filogenéticas e de contexto genômico (ver abaixo).

2.2 Buscas iterativas com sequências intermediárias

Um dos problemas em nossa análise era a possibilidade de que homólogos de genes do divisoma apresentassem seqüências tão divergentes dos genes conhecidos que a detecção de suas relações com membros conhecidos do divisoma, usando as buscas por similaridade baseadas no BLAST e empregados na construção dos COGs, fosse inviável (Brenner, Chothia *et al.*, 1998). Tais circunstâncias levariam à classificação do gene como ausente em organismos nos quais ele se tornou irreconhecível, o que necessariamente afeta as inferências das taxas de ganho e perda dos genes e a composição dos genomas ancestrais (Zhaxybayeva, Nesbo *et al.*, 2007). Buscando evitar estes artefatos, empregamos duas estratégias para identificar eventuais homólogos remotos de genes presentes ns COGs selecionados na seção anterior: (i) para todos os genes de divisão para os quais identificamos uma arquitetura característica de domínios buscamos os homólogos com a mesma arquitetura usando os modelos do banco de dados PFam (Bateman, Coin *et al.*, 2004) e o pacote Hmmer (Eddy, 1998) e (ii) após a remoção de falsos positivos dos COGs que incluíam genes de divisão celular caracterizados experimentalmente (seção 2.1), usamos as seqüências desses COGs como sementes em buscas de similaridade iterativas baseadas no programa PSI-BLAST e os resultados dessas buscas foram utilizados em novas

Figura 2: Fluxograma do método empregado para a detecção de homólogos remotos, usado tanto para detecção de homólogos dos genes de divisão, como na caracterização dos homólogos do domínio tAMIC (seção 3.3.14, página 59). O esquema mostra como novas buscas com o programa PSI-BLAST são iniciadas usando regiões homólogas identificadas na busca anterior. Tais regiões (“aligned hit regions” na caixa ao centro) correspondem às seqüências intermediárias, referidas no texto. Os critérios que definem se uma seqüência similar à semente deve ser considerada homóloga são o *e-value* do alinhamento, a cobertura do alinhamento na semente e a sobreposição, na seqüência intermediária, entre a região incluída no alinhamento atual e a obtida na última iteração.

buscas, até que não fossem encontradas outras seqüências com similaridade significativa. Esse procedimento foi implementado em um programa escrito na linguagem Perl e é conhecido como busca iterativa com seqüências intermediárias (“intermediate sequence searches” ou ISS). Nossa implementação da ISS é essencialmente igual à descrita por John e Sali (John e Sali, 2004), mas acrescenta o critério de cobertura da semente (seqüência “query”), com o intuito de evitar a inclusão de homólogos com similaridade em apenas um ou uma parte dos domínios característicos das proteínas em estudo. A Figura 2 mostra o fluxograma da ISS. Os critérios adotados para definir novas seqüências como homólogos são (1) o *e-value*, ou seja, o número médio de alinhamentos aleatórios com pontuação (“score”) maior ou igual ao observado, que não pode exceder o limite para inclusão

Figura 3: Critério para transitividade das relações de homologia. **(A)** sobreposição dos alinhamentos e coordenadas usadas para cálculo do do escore ISS_{new} . As regiões da sequência intermediária (**I**) que alinham com a semente (usada na iteração anterior) e com o candidato a homólogo encontrado na iteração corrente devem corresponder a uma fração mínima (e.g. $\geq 90\%$) de todas as regiões alinhadas de **I**. **(B)** Fórmula usada no cálculo do escore ISS_{new} . Adaptado de (John e Sali, 2004).

("inclusion threshold") usado nas iterações do programa PSI-BLAST; (2) a razão entre o comprimento da região alinhada na semente e a alinhada na sequência do novo candidato e o comprimento total da semente, chamada cobertura da semente, deve exceder um valor mínimo (cobertura mínima), e (3) a região da sequência intermediária incluída no alinhamento com a semente deve ter um mínimo de sobreposição com a região da sequência intermediária alinhada com um novo candidato a homólogo (este é o critério ISS_{new} , uma medida introduzida por John e Sali (John e Sali, 2004) para assegurar a transitividade das relações de homologia, ou seja, que a região da sequência intermediária com similaridade ao novo candidato seja a mesma com similaridade à semente).

2.3 Construção de alinhamentos e inferência filogenética

Os métodos de detecção de homólogos descritos acima, por si só, não revelam as

relações entre as seqüências identificadas, tornando necessárias análises adicionais que dividem a família em grupos menores e revelam as relações entre esses subgrupos e seus membros, incluindo as relações de ancestralidade (filogenia) e similaridade. Esse tipo de análise depende da construção de alinhamentos múltiplos e da detecção de regiões conservadas e remoção de regiões mal-alinhadas ou altamente divergentes (Castresana, 2000). Com o intuito de aplicar tais análises às famílias de genes do divisoma e grupos menores dentro dessas famílias, desenvolvemos um “pipeline” automatizado e flexível, que permite a opção de diferentes fontes para os dados das seqüências e algoritmos para construção de alinhamentos, seleção de colunas, agrupamento e inferência de filogenias. O programa que automatiza esses procedimentos foi implementado na linguagem Perl, usando módulos disponíveis na biblioteca Bioperl (Stajich, Block *et al.*, 2002), e é controlado por arquivos no formato texto (Figura 4), que também servem para documentação da análise realizada. Ao aplicarmos nosso “pipeline” para as proteínas de divisão celular de procariotos, construímos alinhamentos usando os programas Clustalw (Thompson, Higgins *et al.*, 1994) ou Muscle (Edgar, 2004), com seleção das regiões conservadas e/ou sem inserções/deleções (“gaps”) usando o programa Gblocks (Castresana, 2000). Para proteínas com domínios conhecidos, empregamos também o programa hmmalign do pacote Hmmer (Eddy, 1998) e restringimos os alinhamentos para incluir apenas resíduos correspondentes a estados alinhados (“match states”), sem seleção de resíduos conservados. As análises filogenéticas foram realizadas usando métodos de máxima verossimilhança disponíveis no programa PhyML (Guindon e Gascuel, 2003)., com seleção do melhor modelo de substituição de aminoácidos pelo programa Prottest (Abascal, Zardoya *et al.*, 2005).

```

# FtsI
[pipeline]
project          = FtsI
save_tempfiles  = 1      # whether to save temporary files
concatenate     = 1      # whether to concatenate batch alignments
-make_batch_dir  # process each batch in a separate directory

[concatenate]
phylogeny = protein      # Build phylogeny for concatenated sequences

[domain]
source      = file                # source of domain descriptions
table      = COG0768_local.tsv    # table of domain regions
type       = protein              # coordinates refer to dna or protein
list       = FtsI = Transpeptidase PBP_dimer # domains to create sub-batches
include    = FtsI = Transpeptidase # regexp for target architectures
exclude    = FtsI = MecA_N        # regexp for non-target architectures

[alignment]
select_blocks = 0                # method to select conserved blocks
minimum_length = 40              # Smallest alignment to process
program       = hmmalign         # Alignment method (program) name
parameter fetch = Pfam_ls        # database of HMM models
parameter match_states_only = 1  # exclude residues aligned to
                                # insertion/deletion states
save_gapped   = 1                # preserve alignments with gaps

[phylogeny]
for_each_batch = 0                # set sequence type for batch phylogeny
program       = phyml             # inference method (program name)
select_model  = protest           # procedure to select an evolution model
parameter bootstrap = 100         # number of non-parametric bootstrap samples

[gblocks]
parameter max_contiguous_nonconserved = 20
parameter min_block_length            = 2
parameter similarity_matrix            = y
parameter save_results_and_parameters = y
parameter allowed_gaps                 = h

[database]
connect -dbconfig = /home/rfsouza/.stringdb
connect -verbose  = 0

[string]
restrict      = 1
full_sequences = 1

[data]
source      = string # method to access input data
batch      = FtsI = COG0768 # batch definition

```

Figura 4: Exemplo de arquivo de controle do “pipeline” para construção de alinhamentos e inferência filogenética. Neste exemplo as sequências das proteínas codificadas pelos homólogos do gene *ftsI*, presentes no grupo de homólogos COG0768, são carregadas da cópia local do banco de dados SQL do STRING (bloco [data]). As regiões de cada proteína correspondentes a domínios são selecionadas usando a descrição de domínios presente na tabela COG0768_local.tsv,

excluindo-se seqüências que possuam o domínio MecA_N, envolvidas com resistência a antibióticos. Um alinhamento é construído para cada domínio usando o programa hmalign e o alinhamento concatenado desses domínios é usado para inferência da filogenia pelo programa PhyML, após seleção do modelo de evolução pelo programa Prottest. Note que, nesse exemplo, o bloco [gblocks] é ignorado porque select_blocks no bloco [alignment] é igual a zero.

2.4 Contexto genômico

Classicamente, a inferência de funções para os genes de genomas completamente sequenciados é feita com base na similaridade das novas seqüências a seqüências conhecidas e já caracterizadas experimentalmente, ou seja, com base na premissa de que homologia ou, mais especificamente, ortologia, implica conservação da função (Eisen, 1998). Para uma proporção significativa dos genes em novos genomas, porém, não há disponibilidade de informação funcional em homólogos previamente caracterizados (Doolittle, 2002). As limitações da transferência de dados funcionais baseada em homologia e o rápido acúmulo de várias centenas de genomas completos nos bancos de dados motivou a introdução de novos métodos comparativos, que utilizam dados sobre a organização dos genomas e dos genes para inferir a função das proteínas codificadas (Marcotte, 2000). As inferências possibilitadas por esses métodos, que denominamos métodos de contexto genômico, compreendem a dedução de vínculos funcionais, como a participação em complexos protéicos, vias metabólicas ou redes regulatórias. Diferem, portanto, do tipo de inferência funcional normalmente deduzida com base na homologia, que tende a se concentrar principalmente na função molecular, como a reação ou classe de reações enzimáticas, que a proteína codificada deve desempenhar.

Dentre os métodos de contexto genômico, um dos mais populares é a detecção de correlações no padrão de herança dos genes, ou seja, na semelhança das distribuições de dois ou mais genes em um conjunto de genomas completos (Date e Marcotte, 2003). Essa

distribuição dos genes, denominada perfil filogenético (**PP** para “phylogenetic profile”), corresponde ao padrão de presença/ausência em vários genomas ou, nos casos mais sofisticados, à distribuição das relações de similaridade (escore do alinhamento ou e-value) dos genes em todos os genomas completos em relação a um genoma de referência (Date e Marcotte, 2003; , 2005). A detecção de perfis filogenéticos similares é baseada em medidas de correlação, como a função de informação mútua (Date e Marcotte, 2003). Outro método para detecção de vínculos funcionais, particularmente útil para predição de interações diretas entre proteínas, é a detecção de episódios de fusões de genes em um dos genomas sequenciados, método conhecido como “pedra da Rosetta” (“Rosetta Stone” ou **RS**) (Enright, Iliopoulos *et al.*, 1999; Marcotte, Pellegrini *et al.*, 1999). Da mesma forma que na análise de perfis filogenéticos, esse método também extrapola a análise inicial de homologia, necessária para detectar pares de homólogos fusionados e não-fusionados nos diferentes genomas, e permite a inferência de um papel em um complexo protéico para o produto de um gene hipotético. Outra abordagem é a detecção de conservação da vizinhança dos genes (“gene neighborhood” ou **GN**). Esse método tem origem na idéia de que a conservação da vizinhança, quando não da ordem dos genes, em vários genomas é um indicativo de interações entre as proteínas codificadas é antiga (Dandekar, Snel *et al.*, 1998) e possibilita a maior sensibilidade na detecção de interações conhecidas quando comparada aos outros métodos de contexto genômico (Huynen, Snel *et al.*, 2000). Depende, como os métodos anteriores, da detecção de homólogos, e da aplicação de algoritmos que permitam identificar grupos de genes com vizinhança conservada, como os implementados nos trabalhos de Tamames e colaboradores (Tamames, 2001) ou de Wolf *et al.* (Wolf, Rogozin, Kondrashov *et al.*, 2001).

A introdução dos métodos de contexto genômico e a disponibilidade de bancos de dados sobre interações entre proteínas, determinadas experimentalmente por meio de técnicas de larga escala (“high-throughput”), como experimentos de duplo híbrido, arranjos

de proteínas e TAP (“tandem affinity purification”) (Bader, Betel *et al.*, 2003), ou curadas da literatura (Karp, Riley *et al.*, 2002; Kanehisa, Goto *et al.*, 2004; Stark, Breitkreutz *et al.*, 2006), permitiram a criação de repositórios integrados, nos quais as relações funcionais derivadas dos métodos de contexto genômico e aquelas derivadas por experimentos são integradas. Tais repositórios permitem tanto a determinação de interações preditas para grandes conjuntos de genes hipotéticos (Doerks, Von Mering *et al.*, 2004), como ao usuário navegar o conjunto de predições associadas a um gene de interesse (Myers, Robson *et al.*, 2005). Dentre os vários servidores integrados de dados de contexto genômico (Mellor, Yanai *et al.*, 2002; Bowers, Pellegrini *et al.*, 2004; Date e Marcotte, 2005), escolhemos utilizar em nossas análises o banco de dados STRING, versão 6.3 (Von Mering, Jensen *et al.*, 2005). Essa escolha foi motivada tanto pelo número de organismos incluídos no STRING (180), o maior disponível na época, e pela integração do maior número de diferentes fontes de informação sobre vínculos funcionais. Outro fator foi a disponibilidade de uma filogenia de referência para os organismos deste banco de dados (Ciccarelli, Doerks *et al.*, 2006), inferida com base em uma seleção exaustiva de genes conservados. Nosso uso da estrutura e dos dados disponíveis no STRING se deram principalmente através da interface HTML desse banco de dados, mas também pela exploração de uma instância local do banco de dados SQL no qual essa interface é baseada e foi suplementada pela inspeção manual da anotação dos genomas analisados que, em algumas ocasiões, exigiram novas análises de similaridade e permitiram revelar erros de anotação, como no caso de *zapA* em *A. tumefaciens* e de *sulA* em cianobactérias (Tabela 1).

2.5 Inferência de estados ancestrais

Métodos para reconstrução de caracteres ancestrais se baseiam nas mesmas premissas metodológicas dos métodos de inferência filogenética, ou seja, na premissa de transições

contínuas ou descontínuas (por mutação) entre estados dos caracteres, e no uso de um critério objetivo para escolha da hipótese evolutiva mais plausível para a origem da distribuição atual dos caracteres. Os métodos mais usados são baseados em critérios objetivos como o princípio de máxima parcimônia (Oakley e Cunningham, 2000; Mirkin, Fenner *et al.*, 2003), a maximização da função de verossimilhança (Pagel, 1999a) e o cálculo da distribuição de probabilidades *a posteriori* (métodos bayesianos) para os diferentes cenários evolutivos (Pagel, Meade *et al.*, 2004; Ronquist, 2004). As duas últimas categorias compreendem métodos baseados em modelos probabilísticos explícitos para a evolução dos caracteres (Pagel, Meade *et al.*, 2004), que integram a informação sobre a divergência genética ou temporal e as relações de parentesco entre as espécies extantes (topologia da árvore) com os dados sobre a distribuição atual dos caracteres nessas espécies. A divergência entre as seqüências é representada pelos comprimentos de ramos da filogenia ou filogenias de referência (Pagel, Meade *et al.*, 2004), e as transformações dos caracteres ao longo da filogenia são modeladas por processos de Markov reversíveis e contínuos no tempo, como o modelo para evolução de caracteres binários introduzido por Pagel e Harvey (Harvey e Pagel, 1991; Pagel, 1999a). Esses métodos estatísticos têm mostrado desempenho superior aos métodos baseados no critério de parcimônia, tanto em simulações como na análise de histórias evolutivas conhecidas, *i.e.* criadas em laboratório, principalmente quando as taxas de evolução são elevadas ou a filogenia inclui ramos longos (Cunningham, 1999; Pagel, 1999a).

Em nossas análises, após curarmos manualmente a classificação dos homólogos dos genes de divisão, consideramos apenas o padrão de presença e ausência de quaisquer membros das famílias dos genes do divisoma. Essa restrição limita nossa discussão de estados ancestrais às primeiras aparições de uma família em cada linhagem, mas permite o uso do modelo introduzido por Harvey e Pagel para a evolução de caracteres binários ao longo de uma filogenia (Harvey e Pagel, 1991). Em princípio menos informativa que a

avaliação direta da história de cada família, essa abordagem foi motivada pela impossibilidade de inferência das relações filéticas entre genes das linhagens mais distantes (filos e/ou classes, ver seção 3.3, página 50). Essa mesma abordagem já foi aplicada para estimar a intensidade dos eventos de transferência lateral na definição do conteúdo gênico dos eucariotos (Lester, Meade *et al.*, 2006) e na inferência da história evolutiva do ribossomo (Mushegian, 2005).

Usando um “pipeline” semelhante ao criado para inferência semi-automática de filogenias, descrito na seção 2.3, e baseado no programa BayesMultistates (Pagel, Meade *et al.*, 2004), combinamos o padrão de presença/ausência dos genes do divisoma nos genomas completos do banco de dados STRING com a árvore da vida publicada por Ciccarelli *et al.* (Ciccarelli, Doerks *et al.*, 2006) (Figura 5). A filogenia de Ciccarelli e colaboradores foi obtida a partir da análise de genes amplamente conservados (componentes de ribossomos e tRNA sintetases). Essa filogenia é, para fins de análise, assumida correta e constitui uma fonte independente de informação sobre a história dos organismos analisados, o que nos permitiu usá-la como filogenia de referência ao avaliar as filogenias obtidas para cada gene do divisoma. Inferências do conteúdo gênico ancestral foram realizadas usando o método de máxima verossimilhança e duas variantes do modelo de evolução para caracteres independentes, uma com taxas iguais para a perda (α , $1 \rightarrow 0$) e ganho (β , $0 \rightarrow 1$) dos genes e outra sem restrições para essas taxas (Pagel, 1999b). O melhor modelo para cada gene foi escolhido usando um teste baseado na distribuição χ^2 do logaritmo da razão dos valores da função de verossimilhança (conhecido como “likelihood ratio test” ou LRT). Para avaliar o impacto da amostragem de táxons nas taxas de ganho e de perda de cada gene, reamostramos a distribuição dos genes pela remoção de algumas espécies, em uma análise permitindo que cada uma das espécies fosse removida e em outra que 20 (cerca de 12% do total de espécies) fossem removidas aleatoriamente do conjunto de dados.

Figura 5: Filogenia para os genomas completos incluídos neste estudo. Esta árvore foi usada na inferência da composição do divisoma ancestral e como filogenia de referência para avaliação das filogenias para cada gene. É uma versão reduzida da árvore da vida inferida por Ciccarelli *et al.* (Ciccarelli, Doerks *et al.*, 2006) com base no método de máxima verossimilhança e em um conjunto conservado de proteínas de tradução (proteínas ribossomais e tRNA sintetases). Os códigos dos organismos foram inspirados nos identificadores de espécies do banco de dados SWISSPROT (O'donovan, Martin *et al.*, 2002) e são relacionados aos nomes científicos das espécies na Tabela A1 (seção Anexos).

3. Resultados

3.1 Distribuição dos genes de divisão em genomas completos

A Tabela 1 mostra a distribuição dos genes que codificam proteínas homólogas de componentes conhecidos do divisoma nos genomas completos de procariotos presentes no banco de dados STRING, versão 6.3 (Von Mering, Jensen *et al.*, 2005). Os genes de divisão mais amplamente distribuídos são *ftsZ*, *ftsW* e *ftsI*. *ftsZ* é um homólogo distante de tubulina e pode também ser encontrado no genoma de vários eucariotos, nos quais participa da divisão de organelas de origem endossimbiótica, como cloroplastos e mitocôndrias (Vaughan, Wickstead *et al.*, 2004). Como notado anteriormente por Margolin (Margolin, 2000), as Crenarchaeaota, Planctomycetales, Chlamydiae e algumas Mollicutes são notórias por não possuírem nenhum homólogo de *ftsZ*. Chlamydiae e Mollicutes são endoparasitas e é possível que proteínas do hospedeiro sejam recrutadas para ajudar no processo de septação, complementando as funções de *ftsZ*. As Crenarchaeaota e as Planctomycetales, por outro lado, são organismos de vida livre e mecanismos de divisão muito diferentes podem estar operando sua divisão celular. Muito embora o processo de divisão por brotamento e a morfologia celular ímpar das Planctomycetales possa dispensar o anel de constrição formado por FtsZ (Angert, 2005), estudos recentes mostraram que pelo menos algumas espécies de Crenarchaeaota se dividem por fissão binária (Horn, Paulmann *et al.*, 1999) e, portanto, é provável que tais organismos possuam um complemento diferente de genes especializados em marcar o sítio de divisão e em promover o processo de constrição. Por sua vez, outras linhagens de arqueobactérias (últimos dois blocos da Tabela 1) possuem pelo menos duas ou três cópias de *ftsZ*, que correspondem a eventos de duplicação antigos, anteriores à diversificação dessas linhagens (Vaughan, Wickstead *et al.*, 2004). Levando

em conta a divergência entre estes parálogos, é possível que eles estejam envolvidos com diferentes fases do ciclo celular das arqueobactérias ou com funções não relacionadas à divisão, mas poderiam também cooperar na complementação de funções providas em bactérias por genes de divisão ausentes nos genomas de arqueobactérias e que atuam na polimerização de *ftsZ* e na montagem do anel Z (*ftsA*, *zapA*, *zipA* e *ylmF*).

FtsW e FtsI são homólogos das proteínas de morfologênese celular e crescimento RodA e Pbp2a, respectivamente. Genes codificando estas duas proteínas estão distribuídos entre todas as bactérias que sintetizam peptidoglicano e em Chlamydiae, apesar de não ter sido demonstrada a presença de peptidoglicano na parede celular de clamídias (Cloud-Hansen, Peterson *et al.*, 2006). Outra enzima envolvida na remodelagem do peptidoglicano que é amplamente distribuída entre as bactérias é AmiC, uma N-acetilmuramoil-L-alanina amidase, responsável pela hidrólise de N-acetilmuramoil-L-alanina a ácido N-acetilmurâmico e L-alanina, que ocorre durante a remodelagem da parede celular nos estágios finais da divisão celular de *E. coli* (Bernhardt e De Boer, 2003). Um gene codificando um homólogo de AmiC é encontrado no genoma da arqueobactéria *Methanosarcina mazei* (Deppenmeier, Johann *et al.*, 2002).

Os demais genes de divisão celular apresentam uma distribuição mais dispersa, sendo (1) característicos de certas linhagens, como *zipA* (Proteobacteria), *mipZ* (α -Proteobacteria) e *ezrA* (Firmicutes), (2) ausentes de grandes divisões taxonômicas, como *ftsK* (ausente nos membros do filo Cyanobacteria) e *ftsA* (filo Actinobacteria), ou (3) ausentes em vários grupos menores, dentro dos grupos taxonômicos principais, como *ftsQ*, *minC/D/E* e *ftsE/ftsX*. Os dois últimos exemplos citados (sistema Min e genes *ftsE* e *ftsX*) constituem exemplos de genes com distribuições entre as linhagens (perfis filogenéticos ou “phylogenetic profiles”) altamente correlacionadas, um reflexo dos vínculos funcionais já caracterizados entre esses genes (De Leeuw, Graham *et al.*, 1999).

Tabela 1: Distribuição dos genes que codificam proteínas homólogas de componentes conhecidos do complexo de divisão celular. A primeira linha define a classificação de sub-componentes proposta recentemente para o divisoma (Vicente e Rico, 2006). Células coloridas na última coluna correspondem aos grupos taxonômicos indicados na Figura 5. Dígitos em células com fundo cinza são o número de homólogos de cada família em cada genoma. Células vazias indicam que não foram encontrados homólogos na espécie referida pela linha.

	Anel citoplásmico						ponte periplásmica				fábrica de peptidoglicano				Localização do anel				Oclusão pelo nucleóide			Inibidores anel-Z		Taxonomia		
espécie	PtsZ	PtsA	ZapA	ZipA	YnfF	PtsE	PtsX	PtsK	PtsQ	PtsB/L	PtsW	PtsI	AmiC	PtsN	MinC	MinD	MinE	DivIVA	MpZ	SimA	YyaA	BuA	EzrA	classe	filó	
WIGBR	1	1	1						1	2	2	2	1		1	1	1							γ	Proteobacteria	
CANBF	1	1	1					1	1	2	2	2	1		1	1	1									
BUCAF	1	1								2	1	1	1		1	1	1									
BUCEP	1	1									1	1	1		1	1	1									
PHOLL	1	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1			1		1				
YERPC	1	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1			1		1				
ECOLK	1	1	1	1	1	1	1	1	1	2	2	2	3	1	1	1	1			1		1				
HAEIN	1	1	1	1	1	1	1	1	1	2	2	2	1	1						1						
PASHU	1	1	1	1	1	1	1	1	1	2	2	2	1	1		1				1						
VIBCH	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1			1						
PHOPR	1	1	1	1	1	1	1	1	1	2	2	3	3	1	1	1	1			1						
SHEON	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1			1		1				
PSEAE	1	1	2	1	1	1	1	1	1	2	2	3	2	1	1	1	1					1				
XYLFA	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1									
COXBU	1	1	1	1	1	1	1	1	1	2	2	2	1													
BORPE	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1			1			β			
RALSO	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1			1						
NEDM	1	1	1	1	1	1	1	2	1	1	1	1	3	1	1	1	1									
NITEU	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	1									
RJCCM	1	1	1			1	1	1	1	1	2	2							1					α		
WOLPM	1	1	1			1	1	1	1	1	2	2														
BRIME	1	1	1			1	1	2	1	1	1	1	1		1	1	1									
RITLO	2	1	1			1	1	2	1	1	1	1	1		1	1	1									
RHDE	2	1	1			1	1	2	1	1	1	2	1		1	1	1									
AGRTW	2	1	1			1	1	2	1	1	1	2	1		1	1	1									
BRATA	2	1	1			1	1	1	1	1	1	1	2		1	1	1			1						
RHOPA	1	1	1			1			1	1	1	1	1	1						1						
CAJCR	1	1	1			1	1	1	1	1	3	2	1							1						
HELPI	1	1						1			2	2	1		1	1	1							ε		
CANJE	1	1						1			2	2	1													
BDEBA	1	1				1	1	1	1	1	2	3	1						1					δ		
GEOSL	1	1				1	1	1	1	1	2	2	3	1												
DESVH	1	1				1	1	1	1	1	2	2	1											Cyanobacteria		
SYNY3	1				1			1			1	1	3		1	1	1	1								
SYNEL	1				2	1	1				2	2	2		1	1	1	1								
PROHM	1				2			1			2	2	1		1	1	1	1								
PROMP	1				1			1			2	2	1		1	1	1	1								
GLOVI	1				1	1	1	1	1		2	2	3		1	1	1	1								
DEIRA	1	1				1	1	1	1		1	1	6		1	1	1	1								
THET2	1	1				1	1	1	1	1	2	2	2		1	1	1	1								
THEMA	1	2									2	2			1	1										
AQUAE	1	1						1			2	2	1		1	1										
FUSHN	1	1			2				1	1	2	2	1		1	1	1									
TREPA	1	1						1	1	1	2	2	1											Spirochaetes		
BORBU	1	1						1	1	1	2	2	2													
LEPID	1	1	1					1	1		2	2	1													
LEPTC	1	1	1					1	1	1	2	2	1											Planctomycetes		
RHOPA								1																		
TROWT	1					1	1	1	1	1	2	2	2						1					Actinobacteria		
BIFLO	1					1	1	1	1	1	1	2	2						2							
COREF	1					1	1	1	1	1	1	2	2	1					2							
CORLI	1					1	1	1	1	1	1	2	2	1					2							
MYCLE	1					1	1	1	1	1	1	2	2	2					2							
MYCPA	1					1	1	1	1	1	1	2	2	3					2							
STRCO	1					3	1	1	1	1	2	4	6	1					2							
BACTN	1	1				1	1	1	1	1	2	3	2											Bacteroidetes		
PORGI	1	1				1	1	1	1	1	2	2	1													
CHLTE	1	1				1	1	1	1	1	2	2	1													
CHLCV								1	1		2	2	1											Chlamydiae		
CHLTR								1	1		2	2	1													
MYCPN	1																							Firmicutes		
UREPA																										
MYCPU	1																			1						
MYCHS	1				1																					
ONYPE																										
LACPL	1	1	1			1			2	1	2	3	2	1		1	1				1		1			
LACJO	1	1				1			1	1	1	2	2							1			1			
STRPY	1	1				1	1	1	1	1	2	1	1										1			
STRPM	1	1				1	1	1	1	1	2	2	2										1			
LACLA	1	1				1	1	1	1	1	2	3	2										1			
ENTFA	1	1	1			1	1	1	1	1	2	4	2										1			
OCEID	1	2	1			1	1	1	2	1	2	3	3	7		1	1			1		1				
BACAN	1	1	1			1	1	1	2	1	3	4	7	5		1	1				1		1			
BACSU	1	1	1			1	1	1	2	1	2	3	5	5		1	1				1		1			
LISID	1	1	1			1	1	1	2	1	2	6	2	3		1	1				1		1			
STAEF	1	1	1			1			2	1	2	2	2	1							1		1			
CLOPE	1	1	1			1	1	1	1	1	1	3	6	4		1	1	1	2					Clostridia		
CLOTE	1	2	1			1	1	1	1	1	3	4	6							1						
CLOAB	2	2	1			1	1	2	1	1	1	3	8	1		1	1				1					
THETN	1	2	1			1	1	1	1	1	2	3	5	2		1	1	1	2		1			Euryarchaeota		
METMA	2												1													
ARCTU	4																									
METKA	3	1																								
METTH	1																									
PYRPU	4																									
SULSO																								Crenarchaeota		
AERPE																										
PYRAE																										
NANEQ	2																								Nanoarchaeota	

3.2 Conteúdo gênico nos genomas ancestrais

Para definir a composição do divisoma ancestral, classificamos os valores para a probabilidade de presença dos genes em intervalos, como mostrado na Tabela 2 e na Tabela 3 (códigos de cores na parte inferior, à direita). Cada coluna nestas duas tabelas representa a probabilidade de presença de um gene nos nós ancestrais da filogenia na Figura 5. Quando o valor da probabilidade de presença de um gene (coluna), em um determinado nó descendente (linha), pertence a um intervalo diferente daquele ao qual pertence a probabilidade no nó parental (linha superior), consideramos que houve um evento de ganho ou perda do gene. No caso da Tabela 2 e da Tabela 3 temos os resultados da inferência de estados ancestrais assumindo, respectivamente, probabilidades mínimas de 90 e 70% para a presença de um gene em um dado nó. A comparação das duas tabelas não revela mudanças significativas nas inferências, em especial no caso do LCAB (linha 3, Bactéria), de modo que optamos por definir os eventos com base nos intervalos mais estridentes da Tabela 2. Usando tais intervalos, inferimos que o complexo ancestral, presente no LCAB (Tabela 2, terceira linha), incluía a maior parte dos componentes atualmente conhecidos do divisoma, sendo composto pelas proteínas citoplasmáticas FtsZ e FtsA, pelos componentes da fábrica de peptideoglicano (FtsW/FtsI/AmiC), pelas proteínas YlmF, FtsK e FtsQ, pelo menos um ancestral das proteínas FtsB e FtsL, o par cognato de transportadores do tipo ABC, FtsE e FtsX, todos os componentes do sistema Min e por DivIVA. A presença de homólogos de ZapA e Noc (YyaA, uma proteína da família de ParB) no LCAB é duvidosa, pois a probabilidade de presença no LCAB, para estes genes, está no intervalo correspondente à nossa zona de incerteza para eventos de ganho (60 a 89%). O primeiro divisoma bacteriano tinha, portanto, um conjunto de genes diferente do encontrado em qualquer das linhagens atuais, embora muito semelhante à observada hoje em alguns membros da classe Clostridia, exceto pela possível ausência dos genes *zapA* e *yyaA*.

A semelhança do divisoma ancestral descrito acima com o de membros da classe Clostridiales nos levou a considerar a hipótese de que o conjunto de espécies selecionado para análise poderia afetar de forma crucial as estimativas de probabilidade de presença dos genes no LCAB. Como não dispúnhamos de outra amostra de genomas completos com uma filogenia de referência para comparação, avaliamos essa hipótese com base na distribuição da probabilidade de presença dos genes em genomas ancestrais inferidas usando sub-amostras do nosso conjunto de dados original, procedimento análogo ao método de “jackknife”, porém não equivalente, devido às dependências impostas pela filogenia das espécies aos elementos em nossa amostra, *i.e.* aos divisomas dos organismos analisados, os quais, portanto, não são independentes e identicamente distribuídos (i.i.d) (Miller, 1974; Wu, 1986). Nesta análise, removemos arbitrariamente alguns grupos de bactérias (Tabela 4) e repetimos a análise de estados ancestrais para a distribuição dos genes de divisão nas espécies remanescentes, usando a mesma filogenia de referência. Fizemos também a inferência dos estados ancestrais após exclusão de cada espécie analisada (ao invés de todo um grupo) e de 20 espécies (~12%) selecionadas ao acaso. Nesse última análise, a re-amostragem foi repetida 1000 vezes.

Essas análises exploratórias revelaram, para a maioria dos genes, apenas pequenas variações (<1%) dos valores para a probabilidade de presença nos genomas ancestrais, quando comparados aos valores inferidos usando todos os dados, independentemente do método de re-amostragem. Apenas os genes *zapA*, *ylmF*, *minE* e *yyaA* apresentaram mudanças na categoria de probabilidade de presença no LCAB, todos os quatro genes sendo afetados pela exclusão das Clostridiales (Tabela 4). A presença de homólogos de *yyaA* no ancestral de Firmicutes e de todas as bactérias não é suportada quando removemos a maioria ou todos os membros de Clostridiales da análise, mas sobe a 100% quando desconsideramos as arqueobactérias, sugerindo que a pouca divergência entre o LCAB e as Clostridiales e a grande divergência de arqueobactérias e bactérias exercem papel

determinante nas estimativas da taxa de perda e ganho para este gene. O gene *minE*, por sua vez, também é afetado pela remoção manual de todos membros da classe Clostridia, e pela exclusão de *T. tengcongensis* (ver Tabela 5). Nessas sub-amostras, observamos uma redução da probabilidade de presença de *minE* no LCAB e a inferência de ganho deste gene pelas outras bactérias em um ancestral mais recente, posterior à divergência dos ancestrais de Firmicutes e do ancestral das demais bactérias. A exclusão de *ylmF* do divisoma do LCAB e do ancestral das Firmicutes ocorre, mais uma vez, quando removemos as Clostridiales, mas também quando eliminamos genomas que sofreram redução de seu conteúdo gênico (Tabela 4). Em boa parte das amostras aleatórias nas quais a probabilidade de *ylmF* muda observamos também a exclusão de *F. nucleatum* e/ou *T. tengcongensis*. Esse padrão sugere que as probabilidades para presença de *ylmF* no LCAB são sensíveis à remoção das linhagens mais basais (Clostridiales) ou apicais (mais divergentes em relação ao LCAB) de bactérias, como as espécies de genoma reduzido. O gene *zapA* tem sua probabilidade de presença no LCAB, levando em conta todos os dados, próximo ao valor mínimo do intervalo no qual a presença do gene pode ser considerada (60%). Removendo qualquer uma dentre as espécies *T. tengcongensis*, *E. faecalis* e *L. plantarum*, que possuem homólogos de *zapA*, ou *B. aphidicola* str. *Bp* (*Baizongia pistaciae*), *N. equitans* e *L. johnsonii*, que não possuem homólogos, o valor da probabilidade de presença no LCAB cai para menos de 60%, intervalo no qual o gene é considerado ausente, sendo maior a queda quando retiramos as espécies citadas que possuem homólogos desse gene.

Outros genes para os quais a re-amostragem levou à inferência de variações maiores que 1% na probabilidade de presença no LCAB e outros ancestrais são listados na Tabela 5. Os genes *ftsN* e *ttk* (*slmA*) apresentaram apenas variações para a incerteza associada à ausência desses componentes no divisoma do ancestral das proteobactérias. A probabilidade de presença desses dois genes no ancestral das proteobactérias, porém, se mantêm abaixo de nosso valor mínimo (90%) para todas as re-amostragens, ao passo que as

probabilidades de presença no último ancestral comum das β e γ -proteobactérias, nas quais homólogos desses genes são encontrados, não se altera significativamente, implicando que, embora não possamos precisar exatamente quando estes genes aparecem nas proteobactérias, ambos estavam certamente presentes antes da divergência das β e γ -proteobactérias. No caso de *amiC*, A distribuição de probabilidades para amostras com exclusão aleatória de 20 espécies revelou a instabilidade dos valores para a probabilidade do gene *amiC* nos ancestrais de Firmicutes e Actinobactérias. Não foi possível identificar, para estes ancestrais, um conjunto de espécies que fosse responsável pela redução da probabilidade mas, em ambos os grupos, *amiC* não está presente em uma das linhagens descendente do último ancestral comum, mas é encontrado na linhagem irmã. Levando-se em conta que a probabilidade de presença nos ancestrais de Firmicutes e Actinobacteria são menores que no LCAB, mesmo quando todos os dados são usados (Tabela 2), concluímos que a presença de *amiC* nesses ancestrais não pode ser estabelecida apenas com base na reconstrução por máxima verossimilhança. Essa questão será avaliada novamente mais abaixo, levando em conta dados como a filogenia e a vizinhança do gene.

A Figura 6 mostra a distribuição dos valores da probabilidade de presença dos genes *zapA* e *minE* nos ancestrais dos filos com maior número de espécies analisadas, obtidos pela remoção aleatória de 20 espécies. Embora a probabilidade de presença, nesses dois exemplos, apresente ampla variação, com valores, em alguns ancestrais, oscilando entre 0 e 100%, a maior parte dos valores está concentrada em torno do valor inferido usando todos os dados, com uma distribuição similar à observada para os outros genes. Essa distribuição dos valores inferidos sugere que, se a amostra de espécies contém um número suficiente de representantes das linhagens principais (filos) de bactérias, a inferência não irá diferir significativamente da obtida usando todos os dados. Distribuições similares, com amplitudes menores, foram obtidos para os demais genes.

Tabela 2: Eventos de ganho e perda de genes de divisão, indicados pela probabilidade da presença dos genes nos nós internos (ancestrais) da filogenia mostrada na Figura 5. As linhas correspondem, de cima para baixo, à raiz da filogenia na Figura 1 e, a partir desse nó, no sentido anti-horário, a uma seleção dos outros nós ancestrais, até o ancestral de *Buchnera*. O esquema de cores é explicado na legenda abaixo da tabela e foi definido assumindo a presença de um gene quando sua probabilidade é maior ou igual que 90%. Clados em células verdes são exemplos conhecidos de genomas reduzidos (coluna Clade).

node	clade	FtsZ	FtsA	ZapA	ZipA	YlmF	FtsE	FtsX	FtsK	FtsQ	FtsBL	FtsW	FtsI	AmiC	FtsN	MinC	MinD	MinE	DivIVA	MipZ	Ttk	YyaA	SulA	EzrA
1	Root	99	84			100	100	100	100	100	100	100	100	71		100	100	100	100					
2	Archaea	99																						
3	Bacteria	100	100	60		100	100	100	100	100	100	100	100	100		100	100	100	100			79		
4	Firmicutes	100	100	100		100	100	100	100	100	100	100	100	92		100	100	100	100			100		
5	Bacillales Lactobacillales Mollicutes	100	89	83		100	100	100	100	100	100	100	100	0		100	100	0	100			91		
6	Mollicutes	95	0	0		100	0	0	0	0	0	0	0	0		0	0		100			0		
7	Bacillales Lactobacillales	100	100	100		100	100	100	100	100	100	100	100	93		100	100		100			100		100
8	Staphylococcus	100	100	100		100	0	0	100	100	100	100	100	100		0	0		100			100		100
9	Lactobacillaceae	100	100	83		100	0	0	100	100	100	100	100	79		100	100		100			100		100
10	Enterococcaceae Streptococcaceae	100	100	80		100	100	100	100	100	100	100	100	0		0	0		100			0		100
11	Streptococcaceae	100	100	0		100	100	100	100	100	100	100	100			0	0		100					100
12	Group A	100	100			100	100	100	100	100	100	100	100	100		100	100	100	100					
13	Group B	100	98			100	100	100	100	100	100	100	100	100		0	0	0	100					
14	Bacteroidetes Chlorobi Chlamydiae	97	96			0	100	100	100	100	100	100	100	100					0					
15	Chlamydiae	0	0			0	0	0	100	100	0	0	0	0										
16	Actinobacteria	100	0			100	100	100	100	100	100	100	100	87					100					
17	Spirochaetales	100	100	62		0	0	0	100	100	100	100	100	100					0					
18	Leptospiraceae	100	100	100		100	100	100	100	100	100	100	100	100										
19	Group C	100	100			100	100	100	100	100	100	100	100	100		100	100	100	100					
20	Aquificaeles/Thermotogales	100	100			0	0	0	0	100	0	100	100	68		100	100	0	0					
21	Deinococcus-Thermus	100	100			0	100	100	100	100	100	100	100	100		100	100	100	100					
22	Cyanobacteria	100	0			100	100	100	0	100	0	100	100	100		100	100	100	100					
23	Proteobacteria	100	100			0	100	100	100	100	100	100	100	100		100	100	100	100					
24	Delta	100	100			100	100	100	100	100	100	100	100	100		0	0	0	100					
25	Epsilon Alpha Beta Gamma	100	100	95		100	100	100	100	100	100	100	100	100		100	100	100	100			66		
26	Epsilon	100	100	0		0	0	100	0	0	0	100	100	100		100	100	100						
27	Alpha Beta Gamma	100	100	100	72	100	100	100	100	100	100	100	100	84		100	100	100				76		
28	Alpha	100	100	100		100	100	100	100	100	100	100	100	79		100	100	100				99		
29	Rhizobiales	100	100	100		100	100	100	100	100	100	100	100	100		100	100	100				41		
30	Rickettsiales	100	100	100		0	0	100	100	100	100	100	100	0		0	0	0						
31	Rickettsiaceae	100	100	100		100	100	100	100	100	100	100	100									100		
32	Beta Gamma	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100				95		82
33	Neisseriales	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100				0		
34	Gamma	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100				65		93
35	Xanthomonadales	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100						0
36	Pseudomonadales	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100						100
37	Vibrionales	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100				100		0
38	Pasteurellales	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0				100		
39	Enterobacteriales	100	100	100	96	100	100	100	100	100	100	100	100	100	98	100	100	100				98		98
40	Endosymbionts	100	100	74	0	0	0	100	100	100	100	100	100	0		100	100	100				0		0
41	Buchnera	100	100	0		0	0	0	0	100	100	100	100	100		100	100	100						

Clades

Group A: all bacteria, except Firmicutes (same as groups B + C)

Group B: Actinobacteria Planctomycetales Spirochaetes

Bacteroidetes Chlorobi Chlamydiae

Group C: Cyanobacteria Proteobacteria

Aquificaeles/Thermotogales Deinococcus-Thermus

Events

yellow vertical heritage (root of a group)

gray vertical heritage

red gene family loss (probability < 60)

gold gene family loss (uncertain, i.e. 60 <= probability < 90)

light blue gene gain (uncertain, i.e. 60 < probability < 90)

blue gene gain (probability >= 90)

empty family absent in most members of this lineage (probability < 60)

Tabela 3: Eventos de ganho e perda de genes de divisão, assumindo a presença de um gene em um nó ancestral quando sua probabilidade é maior ou igual a 70%. O esquema de cores é explicado na legenda abaixo, e os números das linhas se referem aos nós da filogenia na Figura 5, como na Tabela 2.

node	clade	FtsZ	FtsA	ZapA	ZipA	YlmF	FtsE	FtsX	FtsK	FtsQ	FtsBL	FtsW	FtsI	AmiC	FtsN	MinC	MinD	MinE	DivIVA	MipZ	SlmA	YyaA	SulA	EzrA
1	Root	99	84			100	100	100	100	100	100	100	100	71		100	100	100	100					
2	Archaea	99																						
3	Bacteria	100	100	60		100	100	100	100	100	100	100	100	100		100	100	100	100			79		
4	Firmicutes	100	100	100		100	100	100	100	100	100	100	100	92		100	100	100	100			100		
5	Bacillales Lactobacillales Mollicutes	100	89	83		100	100	100	100	100	100	100	100			100	100	0	100			91		
6	Mollicutes	95	0	0		100	0	0	0	0	0	0	0	0		0	0		100			0		
7	Bacillales Lactobacillales	100	100	100		100	100	100	100	100	100	100	100	93		100	100		100			100		100
8	Staphylococcus	100	100	100		100	0	0	100	100	100	100	100	100		0	0		100			100		100
9	Lactobacillaceae	100	100	83		100	0	0	100	100	100	100	100	79		100	100		100			100		100
10	Enterococcaceae Streptococcaceae	100	100	80		100	100	100	100	100	100	100	100	0		0	0		100			0		100
11	Streptococcaceae	100	100	0		100	100	100	100	100	100	100	100						100					100
12	Group A	100	100	0		100	100	100	100	100	100	100	100	100		100	100	100	100			0		
13	Group B	100	98			100	100	100	100	100	100	100	100	100		0	0	0	100					
14	Bacteroidetes Chlorobi Chlamydiae	97	96			0	100	100	100	100	100	100	100	100					0					
15	Chlamydiae	0	0				0	0	100	100	0	100	100	100										
16	Actinobacteria	100	0			100	100	100	100	100	100	100	100	87					100					
17	Spirochaetales	100	100	62		0	0	0	100	100	100	100	100	100					0					
18	Leptospiraceae	100	100	100					100	100	100	100	100	100										
19	Group C	100	100			100	100	100	100	100	100	100	100	100		100	100	100	100					
20	Aquificaeles/Thermotogales	100	100			0	0	0	0	100	0	100	100	65		100	100	0	0					
21	Deinococcus-Thermus	100	100			0	100	100	100	100	100	100	100	100		100	100	100	100					
22	Cyanobacteria	100	0			100	100	100	0	100	0	100	100	100		100	100	100	100					
23	Proteobacteria	100	100	51		0	100	100	100	100	100	100	100	100		100	100	100	100			47		
24	Delta	100	100				100	100	100	100	100	100	100	100		0	0	0	100					
25	Epsilon Alpha Beta Gamma	100	100	95		100	100	100	100	100	100	100	100	100	58	100	100	100	100	0		66		45
26	Epsilon	100	100	0			0	0	100	0	0	100	100	100		100	100	100						
27	Alpha Beta Gamma	100	100	100	72	100	100	100	100	100	100	100	100	99	84	100	100	100		58	76		55	
28	Alpha	100	100	100	0	100	100	100	100	100	100	100	100	79	0	100	100	100		99	0			
29	Rhizobiales	100	100	100		100	100	100	100	100	100	100	100	100		100	100	100		41				
30	Rickettsiales	100	100	100		0	0		100	100	100	100	100	0		0	0	0		47				
31	Rickettsiaceae	100	100	100					100	100	100	100	100							100				
32	Beta Gamma	100	100	100	100		100	100	100	100	100	100	100	100	100	100	100	100			95		82	
33	Neisseriales	100	100	100	100		100	100	100	100	100	100	100	100	100	100	100	100			0			
34	Gamma	100	100	100	100		100	100	100	100	100	100	100	99	100	100	100	100			65		93	
35	Xanthomonadales	100	100	100	100		100	100	100	100	100	100	100	100	100	100	100	100					0	
36	Pseudomonadales	100	100	100	100		100	100	100	100	100	100	100	100	100	100	100	100					100	
37	Vibrionales	100	100	100	100		100	100	100	100	100	100	100	100	100	100	100	100			100		0	
38	Pasteurellales	100	100	100	100		100	100	100	100	100	100	100	100	100	100	100	0	0			100		
39	Enterobacteriales	100	100	100	96		100	100	100	100	100	100	100	100	98	100	100	100			98		98	
40	Endosymbionts	100	100	74	0		0	0		100	100	100	100	100	0	100	100	100			0		0	
41	Buchnera	100	100	0					0	0	100	100	100	100		100	100	100						

Clades

- Group A: all bacteria, except Firmicutes (same as groups B + C)
Group B: Actinobacteria Planctomycetales Spirochaetes
Bacteroidetes Chlorobi Chlamydiae
Group C: Cyanobacteria Proteobacteria
Aquificaeles/Thermotogales Deinococcus-Thermus

Events

- yellow vertical heritage (root of a group)
gray vertical heritage
red gene family loss (probability < 40)
gold gene family loss (uncertain, i.e. 40 <= probability < 70)
light blue gene gain (uncertain, i.e. 40 < probability < 70)
blue gene gain (probability >= 70)
empty family absent in most members of this lineage (probability < 40)

Tabela 4: Probabilidades de presença no LCAB (alto) e taxas de ganho (alfa, no meio) e perda (beta, embaixo) inferidas após a remoção de alguns táxons (listados na coluna à direita). Os códigos de cores para probabilidade de presença no LCAB são os mesmos usados na Tabela 2. Note que a diferença entre as cores amarelo e azul na tabela abaixo se referem apenas à presença ($p > 90\%$) de cada gene no último ancestral universal, ou seja, no último ancestral comum a bactérias, arqueabactérias e eucariotos, que não está sendo considerado em nossa discussão, e não implicam variação da composição do divisoma no LCAB. A linha “graus de liberdade” (embaixo) identifica o modelo selecionado pelo teste LRT quando incluímos todos os genomas: (1) modelo com taxas de ganho e perda iguais, (2) modelo sem restrições para as duas taxas. As células cinzas identificam sub-amostras para as quais o melhor modelo de evolução diferiu do modelo selecionado usando todos os dados.

táxon(s) removido(s)	FtsZ	FtsA	ZapA	ZipA	YlmF	FtsE	FtsX	FtsK	FtsQ	FtsBL	FtsW	FtsI	AmiC	FtsN	MinC	MinD	MinE	DivIVA	MipZ	SlmA	YyaA	SulA	EzrA
probabilidade																							
Nenhum	100	100	60		100	100	100	100	100	100	100	100	100		100	100	100	100			79		
Archaea	100	100			100	100	100	100	100	100	100	100	100		100	100	100	100			100		
Firmicutes	100	100				100	100	100	100	100	100	100	100		100	100	100	100			NA		NA
Clostridia	100	100				100	100	100	100	100	100	100	98		100	100	73	100					
Chlamydiae	100	100	62		100	100	100	100	100	100	100	100	100		100	100	100	100			80		
Genomas reduzidos	100	100				100	100	100	100	100	100	100	100		100	100	100	100			73		
Archaea e reduzidos	100	100				100	100	100	100	100	100	100	100		100	100	100	100			100		
Archaea, Chlamydia e reduzidos	100	100				100	100	100	100	100	100	100	100		100	100	100	100			100		
Firmicutes e Archaea	100	100				100	100	100	100	100	100	100	100		100	100	100	100			NA		NA
alfa (α)																							
nenhum	0,18	0,08	0,13	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,29	0,03	0,00	0,00	0,00	0,00	0,04	0,06	0,03	0,07	0,03
Archaea	0,19	0,00	0,30	0,08	0,00	0,00	0,00	0,21	0,00	0,00	0,08	0,08	0,31	0,05	0,00	0,00	0,00	0,00	0,05	0,07	0,00	0,10	0,04
Firmicutes	0,11	0,21	0,11	0,04	0,10	0,00	0,00	0,18	0,00	0,00	0,07	0,07	0,23	0,04	0,00	0,00	0,00	0,02	0,05	0,08	NA	0,10	NA
Clostridia	0,19	0,08	0,12	0,03	0,15	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,30	0,03	0,00	0,00	0,04	0,00	0,04	0,06	0,03	0,08	0,03
Chlamydiae	0,15	0,18	0,13	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,30	0,03	0,00	0,00	0,00	0,00	0,04	0,06	0,03	0,08	0,03
Genomas reduzidos	0,11	0,21	0,19	0,03	0,15	0,00	0,00	0,10	0,00	0,00	0,07	0,07	0,26	0,04	0,00	0,00	0,00	0,00	0,05	0,06	0,03	0,09	0,03
Archaea e reduzidos	0,10	0,20	0,28	0,05	0,30	0,00	0,00	0,10	0,26	0,00	0,05	0,05	0,27	0,10	0,00	0,00	0,00	0,00	0,07	0,09	0,00	0,12	0,05
Archaea, Chlamydia e reduzidos	0,05	0,16	0,29	0,05	0,32	0,00	0,00	0,11	0,27	0,49	0,05	0,05	0,28	0,10	0,00	0,00	0,00	0,00	0,07	0,09	0,00	0,13	0,05
Firmicutes e Archaea	0,10	0,21	0,16	0,10	0,15	0,00	0,00	0,22	0,34	0,00	0,05	0,05	0,23	0,07	0,00	0,00	0,00	0,00	0,08	0,10	NA	0,14	NA
beta (β)																							
nenhum	0,18	0,30	0,57	0,37	0,98	1,01	1,01	0,31	0,36	0,63	0,12	0,12	0,29	0,79	1,10	0,95	1,14	0,99	1,60	4,71	1,56	3,53	0,03
Archaea	0,19	0,27	0,30	0,08	1,08	1,08	1,08	0,21	0,34	0,64	0,08	0,08	0,31	0,73	1,18	1,02	1,24	1,07	1,55	4,90	2,04	3,54	0,04
Firmicutes	0,11	0,21	0,11	0,38	0,10	0,94	0,93	0,18	0,36	0,61	0,07	0,07	0,23	0,81	0,93	0,76	0,94	1,21	1,63	4,76	NA	3,52	NA
Clostridia	0,19	0,31	0,48	0,37	0,15	1,07	1,07	0,32	0,37	0,60	0,13	0,13	0,30	0,79	1,16	1,01	0,98	1,06	1,60	4,72	1,09	3,54	0,03
Chlamydiae	0,15	0,18	0,58	0,37	0,94	0,96	0,96	0,32	0,38	0,59	0,13	0,13	0,30	0,79	1,10	0,96	1,14	0,95	1,60	4,72	1,57	3,53	0,03
Genomas reduzidos	0,11	0,21	0,19	0,03	0,15	0,78	0,78	0,10	0,29	0,59	0,07	0,07	0,26	0,40	1,03	0,87	1,18	0,85	2,14	3,80	1,03	2,44	0,03
Archaea e reduzidos	0,10	0,20	0,28	0,05	0,00	0,82	0,82	0,10	0,26	0,60	0,05	0,05	0,27	0,10	1,12	0,93	1,31	0,92	1,97	3,89	1,36	1,58	0,05
Archaea, Chlamydia e reduzidos	0,05	0,16	0,29	0,05	0,00	0,76	0,76	0,11	0,27	0,49	0,05	0,05	0,28	0,10	1,13	0,94	1,32	0,86	1,96	3,90	1,38	1,54	0,05
Firmicutes e Archaea	0,10	0,21	0,16	0,10	0,15	1,00	1,00	0,22	0,34	0,62	0,05	0,05	0,23	0,73	1,00	0,81	1,02	1,37	1,57	5,05	NA	3,75	NA
graus de liberdade	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	1

Genomas reduzidos: Mollicutes, Ricktsiales e endossimbiontes de insectos (*Buchnera*, *B. floridans* e *W. glossinidia*)

NA: não aplicável (todas as linhagens com homólogos foram removidas)

Figura 6: Distribuição de freqüências para os genes *zapA* (primeiras duas colunas à esquerda) e *minE* (colunas da direita), obtidas usando sub-amostras dos dados do STRING. Para produzir cada histograma, geramos 1000 re-amostragens das distribuições de genes pela remoção aleatória de 20 táxons e calculamos a probabilidade de presença dos genes no divisoma do último ancestral comum de cada grupo (identificados acima de cada painel). Para a maioria dos genes de divisão, as probabilidades re-amostradas estavam centradas na probabilidade inferida a partir do conjunto completo de dados, com uma variação de menos de 1% entre os valores mínimo e máximo da probabilidade. Os dois genes acima apresentaram as maiores variações da probabilidade durante a re-amostragem mas suas distribuições também estavam centradas em torno do valor inferido para o LCAB usando todos os dados (60% para *zapA* e 100% para *minE*).

Tabela 5: Genes com variação superior a 1% para a probabilidade de presença em genomas ancestrais. Os genes *ftsE* e *ftsX*, tiveram a probabilidade de presença no ancestral das cianobactérias reduzida a zero apenas quando todas as espécies com homólogos são removidas deste grupo. *ylmF* apresenta três valores para a presença no ancestral das Firmicutes: 100, 99 e 0%. Tais valores extremos decorrem da inferência de estados absorventes com taxas de ganho (α) e perda (β) iguais à zero. Os códigos entre parênteses na última coluna identificam as espécies que provocaram a variação, durante a remoção de espécies individuais (“cada espécie” na coluna Re-amostragem). A coluna “Total” mostra o valor da probabilidade (%) para a presença de cada gene, no ancestral à direita, quando usamos todos os dados.

Gene	Re-amostragem	Total	Intervalo	Clado afetado
<i>ftsE</i>	aleatória	100	0	Cyanobacteria
<i>ftsN</i>	aleatória	0	0-60	Proteobacteria
<i>ftsX</i>	aleatória	100	0	Cyanobacteria
<i>minE</i>	cada espécie	100	53	Bacteria (THETN)
<i>minE</i>	aleatória	100	0-100	Bacteria, Proteobacteria e Firmicutes
<i>Ttk</i>	aleatória	4	0-70	Proteobacteria
<i>ylmF</i>	cada espécie	100	33	Bacteria (FUSNN)
<i>ylmF</i>	aleatória	100	0-100	Bacteria e Firmicutes
<i>yyaA</i>	aleatória	79	0-100	Bacteria
<i>yyaA</i>	aleatória	100	0-70	Firmicutes
<i>zapA</i>	cada espécie	60	31-59	Bacteria (ENTFA, THETN, LACPL, LACJO, BUCBP, NANEQ)
<i>zapA</i>	aleatória	60 e 51	0-100	Bacteria e Proteobacteria
<i>amiC</i>	aleatória	100	60-100	Firmicutes
<i>amiC</i>	aleatória	100	75-100	Firmicutes e Actinobacteria

3.3 Filogenias dos genes de divisão

Uma vez que a inferência dos estados ancestrais para o padrão de presença/ausência dos componentes do divisoma, descrito na seção anterior, é baseado no mapeamento binário da distribuição dos genes, dois fenômenos evolutivos de interesse não foram contemplados: a duplicação de genes e a transferência lateral dentro de linhagens que já tinham homólogos das famílias gênicas estudadas ou que tenham seus representantes substituídos pelo homólogo de um organismo doador (substituição por xenólogo). Para identificar esses eventos na evolução do divisoma, realizamos a inferência das filogenias para cada um de seus genes componentes. As árvores obtidas apresentaram, para a maioria das linhagens e dos genes, pouca ou nenhuma resolução para níveis taxômicos mais elevados, como para as relações entre diferentes filos e, em alguns casos, classes de bactérias, quase sempre resultando em valores de suporte por “bootstrap” não-paramétrico inferiores a 70% e ramos internos curtos (por exemplo, para *ftsE* e *minC*). Para alguns dos genes mais problemáticos, porém, foi possível obter valores de suporte mais altos e comprimentos de ramos internos maiores para divisões taxômicas mais recentes, como classes, ordens e famílias, após a remoção das linhagens mais claramente divergentes, ou seja, com comprimentos de ramos muito maiores que a média das linhagens irmãs.

3.3.1 FtsZ

A Figura 7 mostra a filogenia dos homólogos de *ftsZ*. Uma análise inicial dos homólogos de *ftsZ* mostrou a presença de agrupamentos discordantes da filogenia de referência, porém com valores suportados muito baixos (<10%). Para evitar possíveis artefatos provocados pela presença de linhagens mais divergentes (“long branch

attraction”), a filogenia na Figura 7 não inclui homólogos que foram associados, nas primeiras análises, a ramos longos e valores de suporte baixos. Esse critério levou à exclusão de todas as cópias de *FtsZ* das arqueobactérias, Mollicutes, Bacteroidetes, Chlorobi, Thermotogales e também algumas sequências truncadas de Proteobactérias. A filogenia resultante é congruente com filogenias inferidas anteriormente para *ftsZ* (Vaughan, Wickstead *et al.*, 2004; Pilhofer, Rosati *et al.*, 2007), mas ainda não reproduz os agrupamentos presentes em nossa filogenia de referência e é incapaz de resolver as relações entre os filos estudados, como observado também por Pilhofer *et al.* (Pilhofer, Rosati *et al.*, 2007). As principais divergências com a filogenia de referência são a inclusão das Actinobactérias dentro das Firmicutes, o agrupamento de Spirochaetes com Cianobacteria e a posição anômala dos homólogos de Mollicutes. Todas essas discordâncias estão associadas a nós com baixos valores de suporte por “bootstrap” (<35%). Os parálogos de Rhizobiales, como descrito por Vaughan *et al.* (Vaughan, Wickstead *et al.*, 2004), formam um grupo que inclui *Brucella* e exclui os homólogos de *Bradyrhizobium*, que formam outro grupo. A relação entre os parálogos de *ftsZ* encontrados em α -Proteobactérias é, portanto, compatível com a duplicação do gene, independentemente, nos ancestrais das rizobiales e bradirizobiales, mas também pode ser explicada pela duplicação do gene em um ancestral comum a esses dois grupos, seguida pela perda do parálogo mais recente em *Brucella*. Com respeito ao outro caso de paralogia de *ftsZ* nos genomas que estudamos, a filogenia dos parálogos de *ftsZ* em arqueobactérias produziu uma topologia idêntica à figura 5 em (Vaughan, Wickstead *et al.*, 2004), com o adendo de posicionar os dois homólogos de *N. equitans* junto ao ancestral comum de *ftsZ2* e *ftsZ1* (dados não mostrados). Essa filogenia também apresenta sub-grupos monofiléticos e congruentes com a filogenia de referência para *ftsZ2* e *ftsZ3*, mas divide os ortólogos de *ftsZ1* em grupos distintos, sugerindo que esse gene poderia não ser monofilético. A origem parafilética de *ftsZ1* em nossa árvore de

máxima verossimilhança não foi suportado por valores elevados de “bootstrap” e não foi reproduzida em filogenias de máxima parcimônia e “neighbor-joining” por Vaughan *et. al*, o que levou esses autores a sugerir que esta porção da árvore está incorreta e que a duplicação de *ftsZ* no último ancestral comum a todas as arqueobactérias produziu *ftsZ1* (o mais amplamente distribuído dos três parálogos) e, em duplicações subsequentes, *ftsZ2* e *ftsZ3*.

3.3.2 FtsK

O gene *ftsK* pertence a uma família de DNA translocases envolvidas com segregação de cromossomos e conjugação de plasmídeos, amplamente distribuída entre as bactérias. A arquitetura típica das proteínas dessa família envolvidas em segregação e divisão celular consiste de um domínio citoplasmático com atividade de ATPase e DNA translocase e um domínio transmembrana, responsável pelas interações com outros membros do divisoma e capaz de promover a maturação do divisoma em *E. coli* mesmo na ausência do domínio citoplasmático (Draper, McLennan et al., 1998). Outras proteínas com domínios homólogos do domínio citoplasmático de FtsK atuam em processos de conjugação e apresentam arquiteturas de domínios diferente, com duas ou mais cópias do domínio citoplasmático.

A Figura 8 mostra a filogenia inferida para os homólogos de *ftsK* e *spoIIIE*. Essa filogenia não apresentou um resultado melhor que outros genes altamente conservados do divisoma, como *ftsZ* e *ftsI*, no que se refere à resolução de relações entre os filos de bactérias, mas resultou em alguns exemplo notáveis de concordância com a filogenia de referência, com sub-grupos que não foram observados nas filogenias dos outros genes, como o grupo formado por Spirochaetes, Planctomycetales e Chlamydia. A árvore da esquerda representa a filogenia, inferida por “neighbour-joining”, de todos os homólogos

do domínio citoplasmático, identificados com base na similaridade com o modelo SpoIIIE do banco de dados Pfam. A filogenia à direita corresponde ao sub-grupo destacado pela oval na figura da esquerda e compreende tanto os homólogos que comprovadamente atuam no processo de divisão celular em *B. subtilis* e *E. coli*, bem como parálogos que codificam proteínas com a mesma composição de regiões transmembrana e/ou alta similaridade a ortólogos de *ftsK*. A figura destaca (em negrito) casos de duplicação do gene *ftsK*, que ocorreram de forma independente nas α e β -Proteobactérias (no centro e no alto) e nas Firmicutes (genes *ytpT* e *spoIIIE* de Bacillales). Essa filogenia não reproduziu as relações entre as famílias de β e γ -Proteobactérias presentes na filogenia inferida por (Ciccarelli, Doerks et al., 2006), porém as relações desses sub-grupos nesta filogenia não são suportados por valores significativos (>70%) de “bootstrap” não-paramétrico, implicando que a inferência de eventos de transferência lateral com base em divergência com a filogenia de referência, *e.g.* entre Xanthomonadales e β -proteobactérias, não podem ser feitas com confiabilidade. O exemplo específico das β -proteobactérias e Xanthomonadales, porém, chama a atenção por corresponder a um agrupamento observado com frequência nas filogenias estudadas, ainda que quase sempre associado a valores de suporte reduzidos. Outras divergências com a filogenia de referência incluem a posição das Deinococcales, que aparecem como irmãs de Actinobacteria, e a posição de Bacteroidetes, que não formou o clado de referência com Actinobacteria.

Figura 7: Filogenia do gene *ftsZ*. As cores identificam os grupos destacados na Figura 5 e na Tabela 1. Essa filogenia reproduz a posição dos parálogos das Rhizobiales (grupos aos quais pertencem os homólogos de *Agrobacterium*, que aparecem sublinhados) e a maioria os clados descritos por Vaughan *et al.* (Vaughan, Wickstead *et al.*, 2004), mas difere da filogenia de referência (Figura 5) por não mostrar a monofilia das Firmicutes (em amarelo) e na posição das Spirochaetes e Actinobacteria. Essas discrepâncias, porém, não são suportadas por valores elevados de “bootstrap” não-paramétrico (números nos nós internos).

Figura 8: Filogenia dos homólogos de *ftsK*. A árvore à esquerda representa a filogenia de todos os homólogos do domínio citoplasmático, identificados com base na similaridade com o modelo SpoIIIE, do banco de dados Pfam, e inferida por “neighbour-joining”. A filogenia à direita corresponde ao sub-grupo destacado pela oval na figura da esquerda e compreende homólogos que comprovadamente atuam no processo de divisão celular em *B. subtilis* e *E. coli*, como parálogos com alta similaridade aos ortólogos de divisão. A figura destaca (em negrito) casos de duplicação gênica de *ftsK* nas α e β -Proteobactérias (no centro e no alto) e nas Firmicutes (genes *ytpT* e *spoIIIE* de Bacillales). Essa filogenia não reproduziu as relações entre as famílias de β e γ -Proteobactérias presentes na filogenia de referência..

3.3.3 FtsX e FtsE

Os genes *FtsX* e *ftsE* pertencem a famílias de componentes de transportadores do tipo ABC (“ATP-binding cassette” ou cassete de ligação de ATP), sendo que FtsX corresponde à permease do complexo, integrada à membrana, e FtsE é homólogo dos componentes (primariamente citoplasmáticos) que se ligam ao ATP e dirigem as mudanças conformacionais e a hidrólise do ATP, ambas necessárias para o transporte (Corbin, Wang *et al.*, 2007). Embora nem o ligante, nem o papel dessas proteínas no processo de divisão celular sejam conhecidos, vários experimentos demonstraram sua participação na divisão em *E. coli* (Schmidt, Peterson *et al.*, 2004). Nos genomas analisados os dois genes ocorrem em operon e um caso de fusão é observado no genoma de *Deinococcus radiodurans*, fenômeno que é observado para muitas proteínas que interagem diretamente (Doerks, Von Mering *et al.*, 2004). Em casos como esse, para o qual também foi demonstrada a interação entre as proteínas codificadas (Corbin, Wang *et al.*, 2007), semelhança entre as filogenias dos dois genes implicará co-evolução. A Figura 9 e a Figura 10 (próximas páginas) mostram as filogenias dos genes *ftsE* e *ftsX*, respectivamente. Ambas as filogenias mostram a monofilia dos filos de bactérias nas quais esses genes estão presentes, com a exceção da posição de *ftsX* presente em *G. sulfurreducens* (suporte por “bootstrap” não-paramétrico de apenas 4%) e dos homólogos de Bacteroidetes que, ainda que monofiléticos, são incluídos entre as proteobactérias. Outras divergências com a filogenia de referência incluem a posição de *C. tepidum* na filogenia de *ftsX*, a posição das β -proteobactérias na filogenia de *ftsE* e o agrupamento de Pseudomonadales e Xanthomonadales em ambas as filogenias. Mesmo considerando tais divergências, é grande a semelhança das duas filogenias, principalmente para os grupos com suporte elevado, o que é compatível com a hipótese de co-evolução.

Figura 9: Filogenia inferida por máxima verossimilhança para os homólogos do gene *ftsE*.

Figura 10: Filogenia dos homólogos de *ftsX*. O parálogo sem par cognato de *C. acetobutylicum* (Q97FR6_CLOAB) é um “in-parálogo”, como demonstrado por seu agrupamento com os membros do gênero *Clostridium*. A posição dos homólogos de Bacteroidetes (no alto) e de algumas outras seqüências (AAR35151_GEOSL de *Geobacterium* e Q8KFL1_CHLTE de *Chlorobium*) são discordantes da filogenia de referência. Porém, no caso destas linhagens, como para a maioria das árvores inferidas para os genes de divisão celular, os baixos valores de “bootstrap” e a presença de linhagens que evoluem a taxas diferentes de suas linhagens irmãs (e.g. α -Proteobactérias, no alto), implicam que as posição dessas linhagens são duvidosas.

3.3.4 FtsQ

A proteína FtsQ é uma proteína bitópica, com uma região N-terminal curta, exposta ao citoplasma, seguida de um segmento que atravessa a membrana citoplasmática e uma grande porção exposta ao meio extracelular. A porção extracelular contém, em sua extremidade N-terminal, um domínio da família POTRA (“POlypeptide-TRansport-Associated domain”), seguido de uma região recentemente caracterizada como um domínio autônomo, chamado domínio β (Robson e King, 2006), e terminando em uma região C-terminal não estruturada. O domínio POTRA ocorre em algumas outras famílias de proteínas transmembrana, que invariavelmente contêm um outro domínio com enovelamento do tipo barril- β (“ β -barrel”) (Sanchez-Pulido, Devos *et al.*, 2003). Já o domínio β é estruturalmente único, não apresentando semelhança com nenhuma outra família de proteínas. A Figura 11 mostra a filogenia dos homólogos do gene *ftsQ* que codificam proteínas com semelhança significativa aos dois domínios característicos desta família (POTRA_! e FtsQ, e-value ≤ 0.01 para o alinhamento contra os modelos do PFam). Nesta árvore, além da tendência característica para reproduzir a monofilia dos clássicos filos de bactérias (Firmicutes, Proteobactéria e, exceto pelo isolamento de *Streptomyces*, Actinobacteria), obtivemos valores de bootstrap consideravelmente altos na raiz desses filos. A relação entre os filos é semelhante à da filogenia de referência apenas em linhas gerais e a exceção mais notável a esse padrão de congruência é a posição de P9RWN7_DEIRA (*Deinococcus radiodurans*), que foi incluído entre as proteobactérias, porém com suporte reduzido. Outras discordâncias com a filogenia de referência incluem a posição de duas β -proteobactérias no clado das γ -proteobactérias e a relação de várias ordens de γ -proteobactérias, como o agrupamento de Pseudomonadales e Xanthomonadales e a posição das Pasteurellales.

Figura 11: Filogenia das proteínas codificadas pelos homólogos de *ftsQ*. Foram incluídas nessa análise apenas as proteínas com similaridade significativa (e-value ≤ 0.01) aos domínios POTRA_1 e FtsQ, do banco de dados PFam. Valores de suporte por “bootstrap” não-paramétrico são mostrados abaixo e à esquerda dos nós internos aos quais se referem. Filos e classes presentes na filogenia de referência são identificados pelas cores e pelas etiquetas à direita. Com exceção de Actinobacteria, β -Proteobacteria e *Deinococcus*, todos esses clados são monofiléticos e definidos por valores de “bootstrap” significativos.

3.3.5 FtsA

A proteína FtsA apresenta a típica estrutura dos homólogos de actina, com duas repetições do domínio RNase H-fold (Bork, Sander *et al.*, 1992). Difere, porém, de outras proteínas desse grupo, como actina, ARP, ParM, HSP70, MreB e algumas quinases de açúcares, por apresentar uma inserção no primeiro domínio RNase H-fold, que foi caracterizada como um domínio ou módulo independente por comparações estruturais e de sequência (Anantharaman e Aravind, 2004). Este domínio, denominado SHS2, atua principalmente como mediador de interações proteína-proteína ou na ligação de pequenas moléculas, sendo seu envolvimento nas interações de FtsA com outras proteínas do divisoma corroborado por ensaios de duplo híbrido e de localização (Rico, Garcia-Ovalle *et al.*, 2004). Parálogos de *ftsA* são observados em algumas Bacillales, Clostridiales e *Thermotoga* (ver Tabela 1). A Figura 12 mostra a relação provável desses parálogos com os ortólogos de *ftsA* e com algumas outras sub-famílias do grupo das actinas presentes em procariotos. O agrupamento do homólogo de *M. kandleri* (Q8TY55_METKA) com os demais parálogos de *ftsA* e a separação destes parálogos dos ortólogos de *ftsA* presentes em *C. perfringens*, *Thermotoga*, *Fusobacterium*, Bacteroidetes e nas demais bactérias é suportado por valores de “bootstrap” elevados se inferimos a filogenia apenas com *ftsA* e *ftsA2*, excluindo as outras sub-famílias. Esses parálogos não apenas são mais próximos dos ortólogos de *ftsA*, como apresentam a região 1C, correspondente ao domínio SH2S e uma assinatura na extremidade C-terminal, ambos característico dessa sub-família (Lowe e Van Den Ent, 2001). Considerando-se a evolução do ortólogo de *ftsA*, propriamente dito, nas linhagens de bactérias, vemos na Figura 13 a filogenia de *ftsA* para uma seleção desses ortólogos, na qual excluimos alguns táxons mais divergentes, como Bacterioidetes, *Thermotoga* e *C. perfringens*, visando obter uma melhor resolução da filogenia. Não foi

possível, porém, obter a monofilia das Proteobactérias, implicada pela filogenia de referência, dada a tendência dos ortólogos de α e ϵ -proteobactérias de formarem clados isolados das demais proteobactérias. Da mesma forma, os homólogos de *Staphylococcus*, *A. aeolicus*, *Chlorobium* e Deinococcales não se localizam em posições congruentes com a filogenia de referência. As posições relativas desses grupos, porém, não são suportadas por valores elevados de “bootstrap”, ao contrário do que é observado para as relações dos ortólogos dentro de todos os grupos, que são bem suportadas e congruentes com a filogenia de referência, demonstrando o predomínio da herança vertical de *ftsA* após a divergência dessas linhagens. Note-se ainda que a maioria das bactérias apresenta a cópia de *ftsA* em um contexto genômico preservado, muitas vezes dentro do agrupamento **dcw** e, mais frequentemente, entre *ftsQ* e *ftsZ*, ambos genes com filogenias nas quais as Proteobactérias aparecem como um grupo monofilético.

Figura 12: Posição relativa dos parálogos de *ftsA* e do homólogo encontrado em *M. kandleri*. Não foi possível obter uma filogenia mais bem definida (*i.e.* com valores de suporte maiores) mas a figura acima indica uma maior similaridade dos homólogo do grupo FtsA2 com os ortólogos de *ftsA*, do que com outros homólogos de actina presentes em bactérias. A seleção das sequências para inferência da filogenia acima foi feita usando buscas por similaridade iterativas (ISS).

Figura 13: Filogenia dos homólogos de *ftsA*. Apesar do alinhamento de 305 resíduos ser comparável ao dos genes *ftsQ* e *ftsZ*, a filogenia inferida por máxima verossimilhança resultou na divisão das seqüências de Proteobacteria em três clados. Essa topologia, bem como a posição de *Staphylococcus*, podem implicar a transferência lateral do gene *ftsA*, porém essa possibilidade não é suportada por valores elevados de “bootstrap”.

3.3.6 FtsB e FtsL

A classificação de homólogos do STRING define três famílias distintas para essas duas proteínas: uma para todos os homólogos de FtsB e outras duas para FtsL, a primeira incluindo homólogos de FtsL de várias linhagens, como Firmicutes e Actinobacteria, e outra apenas com os homólogos encontrados em γ -Proteobacteria e em algumas β -Proteobacteria. Desse modo, apenas duas linhagens muito divergentes apresentam proteínas FtsL e a reconstrução da distribuição filética baseada nessa classificação (dado não mostrado) implica origens independentes para as proteínas FtsB e FtsL de Firmicutes e Proteobacteria, bem como a ausência das duas proteínas no último ancestral comum de todas as bactérias. A reconstrução de estados ancestrais para a classificação revisada de FtsB e FtsL (Tabela 2), difere da reconstrução para os COGs separados, por resultar em alta probabilidade para a presença de membros da super-família FtsB/FtsL (clã “FtsL-like”) no ancestral comum de todas as bactérias. A Figura 14 mostra a filogenia de máxima verossimilhança para esta família, na qual as sequências de FtsL de Proteobacteria formam um grupo distante dos homólogos de FtsL em Firmicutes e divergem dos demais grupos numa posição intermediária, entre os homólogos de FtsB de Firmicutes e de Proteobacteria, parecendo mais relacionado ao grupo de FtsB de Proteobacteria que ao grupo formado por FtsL de Firmicutes. A filogenia da super-família sugere, portanto, duplicações independentes de um gene ancestral, nos ancestrais das linhagens de bactérias atuais que possuem um homólogo de FtsL. Outra possibilidade é a duplicação do gene ancestral na raiz da linhagem das bactérias, caso em que os agrupamentos observados na filogenia constituiriam um artefato, provocado pela excessiva divergência de FtsL na linhagem de γ -Proteobacteria. Essa última hipótese é corroborada pela sintenia típica da região onde estão localizados os homólogos de FtsL, quase sempre encontrados numa posição conservada

dentro do **dcw**, o que é compatível com a idéia de que a duplicação do gene ancestral de FtsL e FtsB ocorreu antes da aparição do **dcw**, ou seja, antes da divergência de Firmicutes e Proteobacterias. Assim, embora nossos dados não permitam precisar a origem da proteína FtsL de γ -Proteobacteria, sugerem fortemente a presença de um homólogo do clã FtsB/FtsL no último ancestral comum a todas as bactérias e a duplicação desse homólogo antes da divergência das principais linhagens nas quais são encontrados homólogos de FtsL

3.3.7 FtsW e FtsI

Os genes *ftsI* e *ftsW* são membros de famílias de parálogos grandes e com histórias complexas, envolvendo vários episódios de duplicação. As filogenias de FtsW e FtsI, na Figura 15, indicam duplicações antes da divergência das atuais linhagens de bactérias, que separaram esses genes das linhagens dos genes *rodA* e *pbp2*, respectivamente. Para ambos os genes, outras duplicações ocorreram mais tarde, independentemente, em Firmicutes e Actinobacteria. Os parálogos de FtsW em Actinobacteria (linhagem em vermelho, aproximadamente no centro do painel esquerdo), se agrupam com elevado valor de suporte a um grupo de seqüências de Clostridiales (Firmicutes) e podem constituir um caso de transferência lateral de RodA, muito embora a identificação do gene *rodA* em Actinobacteria seja difícil, pela pouca conservação do contexto genômico e falta de dados experimentais. As duplicações de *pbp2* em Actinobacteria e Firmicutes podem estar relacionadas à maior espessura da camada de peptidoglicano na parede celular desses organismos, que são bactérias Gram-positivas e provavelmente necessitam de uma taxa maior de incorporação dos precursores, ou à síntese da parede em condições específicas, como a formação de esporos, que ocorre em ambas as linhagens, ou ainda à resistência a antibióticos.

Figura 14: Filogenia, inferida de máxima verossimilhança (sem “bootstrap”), para todos os homólogos de *ftsB* e *ftsL* identificados usando o modelo para o clã FtsL (Bateman, Coin *et al.*, 2004). O alinhamento das seqüências foi feito usando o programa hmmpfam e a filogenia foi inferida usando o programa PHYML (Guindon e Gascuel, 2003). Nessa árvore, são notáveis o alto grau de divergência dos homólogos da proteína *ftsL* das γ -proteobactérias e o agrupamento dos homólogos *ftsL* e *ftsB* de Firmicutes à parte de quase todos os homólogos de *ftsB* e *ftsL* encontrados em proteobactérias e actinobactérias. A topologia acima pode ser explicada pela origem de *ftsB* e *ftsL* a partir da duplicação de um gene ancestral em Firmicutes ou no LCAB. A origem a partir de uma duplicação em Firmicutes implicaria a transferência do gene *ftsL* de Firmicutes para as δ -proteobactérias e *Streptomyces* e para o ancestral das γ/β -proteobactérias. Esta última transferência teria que estar associada à inserção do xenólogo *ftsL* no mesmo contexto genômico da *ftsL* de Firmicutes, ou seja, dentro do cluster **dcw** da proteobacteria ancestral (Tamames, 2001), pois este contexto é conservado na maioria das proteobactérias. Já a duplicação no LCAB implicaria a perda de *ftsL* nos ancestrais de várias linhagens, como cianobactérias, actinobactérias, Spirochaetes, δ , α e ϵ -proteobactérias. Dado que a inserção de um xenólogo em um contexto genômico específico é grandemente facilitada se ocorrer pela substituição de um homólogo pré-existente, a hipótese de duplicação no LCAB é compatível com a possibilidade de antigos eventos de transferência lateral, e.g. de Firmicutes para as δ -proteobactérias. Finalmente a grande divergência dos homólogos de *ftsL* em γ/β -proteobactérias pode implicar uma aceleração da taxa de evolução desse gene no ancestral dessa linhagem, que explicaria a grande divergência deste sub-grupo para os demais homólogos.

Figura 15: Filogenia para as proteínas FtsW e FtsI. Os códigos de cores identificam os taxons na Figura 5. O ponto no centro de cada filogenia indica a localização da raiz, deduzida com base na separação dos parálogos de FtsW e FtsI, RodA e Pbp2, respectivamente. Note as duplicações de Pbp2 em Firmicutes (amarelo) e Actinobacteria (vermelho). A linhagem de FtsW proveniente de Actinobacteria mais próxima da raiz é, na verdade, RodA, mas a filogenia, que apresenta baixo valores de bootstrap, falha em agrupar RodA de Actinobacteria com seus ortólogos nas outras linhagens.

3.3.8 ZapA

A proteína ZapA consiste de um único domínio conservado, com uma extensa região “coiled-coil” (α -hélices com repetições periódicas de aminoácidos hidrofóbicos e hidrofílicos) na extremidade C-terminal e de duas fitas β antiparalelas na extremidade N-terminal (Low, Moncrieffe *et al.*, 2004). A filogenia desse gene é mostrada na Figura 16, onde podemos notar a separação dos homólogos de Firmicutes e Proteobactérias em clados distintos. Filogenias que incluíram os homólogos de *Leptospira* posicionaram esses genes em um grupo irmão, porém distinto, das Proteobactérias. Os parálogos de *Pseudomonas*, em filogenias que incluíam apenas os homólogos de Proteobactérias, são incluídos no mesmo grupo de γ/β -Proteobactérias, com valores elevados de “bootstrap”, porém fora do grupo que inclui o ortólogo funcional de *P. aeruginosa*. Em todas as análises, uma seleção cuidadosa de linhagens que não apresentassem indícios de variação nas taxas de evolução melhora a filogenia de *zapA*, resultando numa topologia congruente com a filogenia das espécies, ou seja, que separa com altos valores de “bootstrap” os homólogos de Firmicutes e Proteobactérias. A classificação inicial de ortólogos (COGs) incluiu genes de Chlamydia no grupo de *zapA*, porém, o predomínio de ramos muito mais longos neste sub-grupos que no resto da filogenia, associados à ausência de um contexto genômico conservado e diferenças na estrutura das proteínas desse subgrupo (o único com proteínas que apresentam peptídeos sinal na extremidade N-terminal), nos levaram a excluir essas proteínas de clamídias da família de *zapA*.

3.3.9 YlmF

O gene *ylmF* está presente em Actinobacteria, Cyanobacteria e Firmicutes,

frequentemente localizado no agrupamento (“cluster”) **ylm**, entre o gene *divIVA* e o **dcw**. O agrupamento **ylm** pode ser considerado, pelo menos em algumas linhagens, uma extensão do **dcw**. Também conhecido como *sepF*, *ylmF* codifica uma proteína com um domínio conservado de 100 a 200 aminoácidos, capaz de compensar a ausência de FtsA. Por causa deste fenótipo e por depender apenas de FtsZ para se localizar, Ishikawa *et al.* propuseram que YlmF atua nas etapas iniciais de formação do anel (Ishikawa, Kawai *et al.*, 2006). A Figura 17 mostra a filogenia desse gene, na qual os homólogos de Firmicutes não formam um grupo monofilético. A parafilia dos homólogos de *ylmF*, como observado para os outros genes, não é suportada por valores de “bootstrap” altos e não é observada na árvore de máxima verissimilhança (sem “bootstrap”). A posição dos dois homólogos de *Fusobacterium* não é compatível com a filogenia de referência, mas é conciliável com a hipótese de HGT e/ou com a hipótese de que a verdadeira posição desse grupo na árvore da vida é como uma linhagem irmã de Mollicutes (Mira, Pushker *et al.*, 2004).

Figura 16: Filogenia de homólogos de ZapA. Esta árvore mostra a posição dos parálogos de *zapA* encontrados em *Pseudomonas* (*zapA2*). Muito embora os valores de suporte nesta filogenia sejam baixos, *zapA2* se agrupa com os outros homólogos de γ -proteobactérias, o que é confirmado, com valores de suporte mais altos, quando inferimos uma filogenia contendo apenas α , β e γ -proteobactérias.

Figura 17: Filogenia do gene *ylmF*. Os membros desta família codificam proteínas pequenas, resultando em alinhamentos curtos, com sinal filogenético reduzido. A filogenia final, por este motivo, apresenta valores de “bootstrap” sistematicamente baixos, exceto na separação de clados bem definidos (alto valor de suporte na filogenia de referência e do gene), como cianobactérias e actinobacterias. A monofilia dos homólogos de Firmicutes, não representada no consenso acima, é obtida na árvore de máxima verossimilhança.

3.3.10 Sula

Danos ao DNA, provocados por luz ultravioleta ou produtos químicos, induzem a chamada resposta SOS em bactérias, durante a qual a expressão de um conjunto de mais de 30 genes é induzida pela inativação de seu repressor comum, LexA (Foster, 2005). A

Figura 18: Filogenia dos genes *sulA* and *imuA*. *Pirelulla* é um membro do filo Planctomycetes e é usado para orientar a filogenia, embora não possamos apontá-lo como um grupo externo adequado, uma vez que este gene é muito mais semelhante aos homólogos de Proteobactéria do que esperado para um grupo externo e pode ter sido adquirido via HGT. Apesar disso, o homólogo de *Pirelulla* também é vizinho de *imuB* e, portanto, a filogenia acima é compatível com a origem de *sulA* a partir de *imuA*, como discutido no texto.

inativação de LexA é uma reação proteolítica que é amplificada por RecA, quando esta proteína estiver ligada a DNA de fita simples, comum quando há lesões. Em *E. coli*, o gene *sulA* é induzido pela resposta SOS e atua bloqueando a polimerização de FtsZ e, em consequência, a divisão celular, o que permite que a célula realize os reparos necessários em seu genoma antes de efetuar a próxima divisão celular (Cordell, Robinson *et al.*, 2003). Sula é um homólogo de RecA, mas apresenta uma distribuição muito mais restrita, limitado apenas a algumas linhagens de γ -proteobactérias (Tabela 1). A Figura 18 mostra a filogenia dos homólogos de *sulA* e sua relação com *imuA*, outro homólogo de RecA

encontrado em proteobactérias, um gene também induzido na resposta SOS, porém sem atividade sobre FtsZ (Galhardo, Rocha *et al.*, 2005). Essa filogenia é idêntica à descrita por (Abella, Erill *et al.*, 2004) e compatível com a origem de *imuA* a partir da duplicação de *recA*, primeiro no ancestral das $\alpha/\beta/\gamma$ -proteobactérias, sendo esse novo gene herdado verticalmente pelo ancestral das β/γ -proteobactérias, dando origem aos genes “*imuA*-like” desse grupo. Uma segunda duplicação, no ancestral das γ -proteobactérias, daria origem a *sulA*. O homólogo de *Pirelulla baltica*, uma Planctomycetes, é consistentemente agrupado com os homólogos de *imuA* e ocorre na mesma vizinhança genômica deste gene, flanqueado por homólogos de *dinP* e *dnaE*, como observado para todos os homólogos de *imuA* e membros do clado “*imuA*-like” (Abella, Erill *et al.*, 2004). A ampla divergência entre *Pirelulla* e as proteobactérias sugere que o gene nesta espécie foi adquirido horizontalmente (ver Figura 5).

3.3.11 FtsN

O gene *ftsN* é encontrado na maioria das β/γ -proteobactérias, constituindo um dos componentes tardios do divisoma. A estrutura da proteína FtsN de *E. coli*, foi caracterizada usando ressonância magnética nuclear e experimentos genéticos (Yang, Van Den Ent *et al.*, 2004), que demonstraram a presença de um segmento longo não estruturado, rico em glutamina, que compreende a maior parte da região exposta ao periplasma dessa proteína, seguida de um domínio SPOR, que contém dois sítios de ligação a moléculas de peptidoglicano (Mishima, Shida *et al.*, 2005). A região N-terminal de FtsN compreende um segmento citoplasmático curto (1 a 33 aminoácidos) e três hélices transmembrana, que antecendem a região não estruturada e ancoram a proteína na membrana interna. As buscas por ISS revelaram a similaridade de *ftsN* aos membros de outra família de genes comum em

Figura 19: Filogenia do domínio SPOR dos homólogos de *ftsN* e *dedD*. A posição dos homólogos de *Bordetella*, embora suportada por valores reduzidos de “bootstrap”, é corroborada pelo exame da vizinhança genômica dos dois genes, sugerindo que os homólogos desse gênero agrupados com *dedD* são “out-parálogos”, ao invés de originados por uma duplicação de *ftsN* em um ancestral do gênero *Bordetella*.

β/γ -proteobactérias, chamados *dedD* em *E. coli*, que codificam proteínas com uma estrutura de domínios idêntica à de FtsN, incluindo regiões similares a todos os segmentos descritos para FtsN de *E. coli*. A filogenia de *dedD* e *ftsN* é mostrada na Figura 19, na qual os dois genes formam clados isolados, com os parálogos de *Bordetella* (embaixo), que inicialmente haviam sido classificados como membros do grupo de ortólogos de *ftsN* (COG3087), incluídos na linhagem de *dedD* (COG3147). O padrão de bifurcação das linhagens, tanto

para *dedD* como *ftsN*, é compatível com nossa filogenia de referência, ainda que com algumas discordâncias associadas a baixos valores de suporte. Como esse dois genes são fracamente conservados, tendo apenas a região C-terminal (SPOR) como região de homologia, eles são facilmente confundidos com outros genes que contém o mesmo domínio e regiões de baixa complexidade como *rlpA*. Por isso, é possível que a distribuição da super-família *ftsN/dedD* seja mais ampla que o mostrado aqui e homólogos estejam presente em proteobactérias que não as das classes β e γ .

3.3.12 ZipA

O gene *zipA* é essencial para a formação do anel em *E. coli*, onde atua como effector positivo da polimerização de FtsZ. A proteína ZipA de *E. coli* consiste em uma região N-terminal altamente hidrofóbica, de aproximadamente 25 aminoácidos, que forma um domínio transmembrana e ancora a proteína à membrana citoplasmática (Moy, Glasfeld *et al.*, 2000). Essa região é seguida por uma região de baixa complexidade e por um domínio C-terminal mais conservado, com similaridade estrutural, mas não de sequência, a proteínas que contém um motivo de ligação a ribonucleotídeos (domínio RNP), comum em eucariotos (Moy, Glasfeld *et al.*, 2000). ZipA foi anteriormente considerado exclusivo de γ -proteobactérias (Rothfield, Justice *et al.*, 1999; Margolin, 2000), porém, usando buscas iterativas por homólogos remotos (ISS), identificamos homólogos dessa família em β -proteobactérias Tabela 1. A Figura 20 mostra a filogenia obtida a partir das regiões conservadas no alinhamento de todos esses homólogos, que apresenta valores de “bootstrap” sistematicamente elevados para todos os nós congruentes com a filogenia de referência.

Figura 20: Filogenia dos homólogos de *zipA*, incluindo os homólogos de β -proteobactérias, identificados usando buscas iterativas (ISS). Com exceção do homólogo de *N. europaea* (Q82SM8_NITEU), todos os homólogos de *zipA* incluídos estão localizados antes (5') de uma DNA ligase dependente de NAD, o que corrobora a idéia de que os novos homólogos identificados em β -proteobactérias codificam proteínas com o mesmo papel de ZipA.

Figura 21: Filogenia do gene EzrA. A única discordância entre esta filogenia e a filogenia de referência é a posição de *Staphylococcus* (EZRA_STAEP, EZRA_STAAW e EZRA_STAAM), que na filogenia de referência constitui um grupo separado de *Listeria* e *Bacillus*. Não há parálogos nesta família e as discordâncias observadas estão sempre associadas a nós com suporte reduzido, tanto na filogenia acima como na de referência.

3.3.13 EzrA

EzrA atua como um regulador do número de anéis em *B. subtilis*, está restrita aos membros das classes Bacillales e Lactobacillales, do filo Firmicutes, e apresenta similaridade remota com as proteínas eucarióticas quinesina, miosina e proteínas que ligam tubulina (PSI-BLAST, e-values entre $1e^{-7}$ e $1e^{-9}$ na segunda iteração), basicamente restrita às regiões “coiled-coil” que cobrem quase toda a proteína. A filogenia dos homólogos de *ezrA* é mostrada na Figura 21 e só difere da filogenia de referência dentro do grupo das Bacillales, porém os nós discordantes, tanto na filogenia do gene como na de referência, não são suportados pelo “bootstrap” não-paramétrico.

3.3.14 AmiC

Dentre os genes de divisão celular que analisamos, as N-acetilmuramoyl-L-alanina amidases, referidas genericamente como mureína hidrolases, compõem a família com maior diversidade estrutural, *i.e.* com maior variação em termos de estrutura de domínios, sendo os membros da família caracterizados pelo domínio Amidase_3 (Kuroda, Sugimoto *et al.*, 1992). Esse domínio corresponde é responsável pela atividade enzimática que cliva as ligações covalentes entre os resíduos de L-alanina e o ácido N-acetilmurâmico (MurNac ou ANM) nas ligações cruzadas entre unidades de N-acetilglucosamina e ácido N-acetilmurâmico do peptidoglicano (Smith, Blackman *et al.*, 2000). A Figura 22 mostra as estruturas de domínios para as proteínas codificadas por homólogos de *amiC*, ao lado de uma filogenia inferida usando um alinhamento das seqüências de aminoácidos do domínio Amidase_3. Nessa figura, os ramos com valor de suporte menor que 40% tiveram seu comprimento reduzido a zero e alguns grupos monofiléticos de proteínas, na maioria dos casos de um mesmo grupo de organismos (mesma cor), são identificados pelo nome de uma proteína conhecida de *Bacillus* ou *E. coli* que pertence ao grupo (como CwID ou AmiC). Essa filogenia, no entanto, não permite deduzir a relação entre as várias sub-famílias, uma vez que essa filogenia é, em essência, uma “estrela” ou politomia, na qual a maioria dos ramos que internos das sub-famílias partem do mesmo ponto, sem informação filogenética (veja a Figura 23a). Para várias das sub-famílias, porém, ramos basais com valores de suporte mais elevados ($\geq 40\%$) favorecem a hipótese de monofilia.

A diversidade estrutural nesta família consiste na variação dos domínios amino- e/ou carboxi-terminais que flanqueiam o domínio enzimático e são, na maioria dos casos, envolvidos com a interação com componentes da parede celular, como o peptidoglicano

(LysM, SPOR, CW_binding_2 e PG_binding_1), a camada S de bactérias Gram-positivas (SLH) ou outras proteínas (SH3). Além das arquiteturas compreendendo domínios de interação, alguns membros dessa família apresentam outros domínios com provável ou comprovada atividade catalítica, como o domínio Glucosaminidase, encontrado na proteína Q7NkW3, de *G. violaceus*, previamente descrito em hidrolases de peptidoglicano (Nambu, Minamino *et al.*, 1999). O domínio CHAP é encontrado em proteínas que atuam na clivagem de peptidoglicano, em proteínas de *Trypanosoma* envolvidas com síntese de triptationa, um composto essencial para a proteção desses organismos contra o estresse oxidativo, e na glutathionilperimidina sintase de *E. coli* (GSP), na qual o papel catalítico do domínio CHAP foi demonstrado experimentalmente (Bateman e Rawlings, 2003; Llull, Lopez *et al.*, 2006). Curiosamente, o único homólogo que contém o domínio CHAP, encontrado em *S. aureus*, apresenta grande similaridade com proteínas de fagos encontrados em outras cepas de *Staphylococcus*. É muito provável que esta proteína de *S. aureus* e outros parálogos dessa família não só não tenham relação com a divisão ou diferenciação celular, mas sejam o produto de eventos de transferência lateral. O domínio tAMIC, apresentado na figura, foi anteriormente descrito por (Bernhardt e De Boer, 2003), que demonstraram, por meio da deleção de porções do gene *amiC* em *E. coli* e microscopia de fluorescência de fusões com GFP, que esse domínio é responsável pela localização de AmiC no septo de divisão em *E. coli*. Esses autores, no entanto, não caracterizaram a distribuição desse domínio, sugerindo apenas que se tratava de um domínio restrito a enterobactérias. Como nosso objetivo era caracterizar a evolução de *amiC*, conhecer a distribuição do domínio tAMIC era necessário e, uma vez que esse domínio ainda não fora descrito nos bancos de dados especializados, caracterizamos sua distribuição filética, estrutura secundária e alinhamentos múltiplos. Tais análises permitiram a identificação e localização de regiões homólogas em amidases de outras linhagens de cianobactérias e em

outras famílias de proteínas, notadamente em PilQ e alguns transportadores de ferro de cianobactérias¹.

Após a caracterização do domínio tAMIC, buscamos definir a sequência de eventos de duplicação que deram origem aos parálogos *amiA*, *amiB* e *amiC* observados em *E. coli*. A Figura 23 mostra a filogenia inferida para os homólogos de *amiC* nas α , β e γ -proteobactérias. Como visto no painel A, estas três classes formam o grupo monofilético mais claramente definido na árvore que inclui todos os membros dessa família, muito embora com um valor de bootstrap menor que 70%. As classes δ e ϵ , por outro lado, não se agrupam com as outras proteobactérias, aparentemente por falta de sinal filogenético. O painel B mostra que os parálogos *amiC* e *amiB* estão distribuídos entre membros das classes β e γ , mas que *amiA* é restrito às enterobactérias. Embora a posição do grupo de *amiA* que esse gene divergiu cedo dos outros dois parálogos, os valores de “bootstrap” não-paramétrico não permitem dão suporte à ordem de aparecimento dos três genes implicada pela filogenia, que deve ainda ser considerada uma politomia na base dos três grupos. Para os parálogos presentes em *Pseudomonas* e *Xanthomonas*, não foi possível assinalar, com base na filogenia, relações de ortologia com *amiC* e *amiB*, mas a conservação da ordem dos genes em torno do *locus amiB* em *Pseudomonas* e, em menor medida, em Xanthomonadales e algumas β -proteobactérias (Figura 24), em especial a proximidade do gene *mutL*, sugerem que o gene ancestral desses parálogos ocupava um *locus* pelo menos parcialmente sintênico com o de *amiB*. Mais importante ainda: as presenças de tAMIC em α -proteobactérias, cianobactérias e em *G. sulfurreducens*, uma δ -proteobactéria, sugerem fortemente que o ancestral das amidases de proteobactérias possuía o domínio tAMIC, o que descarta a possibilidade da presença de *amiA* no último ancestral comum.

¹ Uma versão reduzida da apresentação desses resultados, exibida na forma de pôster na Gordon Research Conference on Structural, Functional and Evolutionary Genomics (2007) foi incluída em anexos.

Figura 22: (página anterior) Composição de domínios dos homólogos de AmiC. As linhas pretas horizontais representam as proteínas e as figuras geométricas (retângulos, elipses, etc.) identificam as regiões com similaridade a domínios catalogados no banco de dados PFam. A tabela à direita lista os nomes de acesso dos domínios. O domínio tAMiC (retângulo vermelho), não está descrito no PFam, mas foi caracterizado experimentalmente em *E. coli* por (Bernhardt e De Boer, 2003) e *in-silico* por nosso grupo (ver texto). A filogenia à esquerda foi inferida com base no domínio Amidase_3, característico dessa família, e usada para organizar as proteínas em diferentes sub-famílias. Não serve, porém, como referência para as relações entre as sub-famílias, que são agrupadas por ramos muito curtos (topologia “em estrela”) e com valores de bootstrap menores que 40%. Valores de suporte abaixo desse valor (40%) foram omitidos e os respectivos ramos internos reduzidos a zero, *i.e.* condensados, e alguns grupos de homólogos sem variação na composição de domínios (em negrito, *e.g.* CWID) foram colapsados, para reduzir a figura.

Figura 23: filogenia dos homólogos de *amiC*, presentes em α , β e γ -proteobactérias. **(A)** Representação radial da filogenia para os homólogos de *amiC* mostrada na Figura 22, destacando a monofilia dos homólogos de proteobactérias. **(B)** Filogenia inferida com base nas sequências de proteínas do domínio Amidase_3 dos homólogos de *amiC* provenientes das apenas de α , β e γ -proteobactérias. Restringindo a análise filogenética às sequências mostradas acima, obtivemos uma separação mais confiável (suporte por “bootstrap” não-paramétrico de 99%) dos homólogos de α e β/γ -proteobactérias, mas a árvore não permite dizer qual a ordem das duplicações que produziram os três parálogos dessa família observados em enterobactérias, pois a separação de AmiA e AmiB/AmiC é observada em apenas 59% das réplicas no “bootstrap”. Por outro lado, a análise da vizinhança genômica em torno do gene *amiB* de *E. coli*, pelo menos parcialmente conservada nos membros do grupo AmiB (Figura 24), permitiu identificar os homólogos de *Pseudomonas*, *Xanthomonas* e α -proteobactérias que apresentam uma vizinhança semelhante, sugerindo que o *locus* mais antigo da família é o correspondente a *amiB*.

Figura 24: Conservação da ordem e composição gênica (sintenia) na vizinhança do gene *amiB* de proteobactérias, obtida no servidor STRING, versão 7.0. As três classes com grupos de homólogos sintênicos, indicadas pelas barras verticais na última coluna à direita, são, de cima para baixo, γ , β e α . A filogenia à esquerda corresponde à filogenia de referência adotada pelo STRING, na qual grupos que não apresentam variação da vizinhança genômica entre espécies foram colapsados.

3.3.15 **SlmA, YyaA e MipZ (oclusão pelo nucleóide)**

A oclusão pelo nucleóide (“nucleoid occlusion” ou NO) corresponde à inibição da polimerização de FtsZ e, em consequência, da formação do anel Z em regiões da célula onde o citoplasma é ocupado pelo nucleóide (Rothfield, Taghbalout *et al.*, 2005). Por muito tempo, o mecanismo pelo qual a célula induz o NO permaneceu desconhecido, porém reconheceu-se recentemente, em três sistemas modelo, algumas proteínas que, por meio de interações com o nucleóide e com FtsZ, promovem esse efeito. Proteínas diferentes mediam a oclusão pelo nucleóide em Proteobactérias (Bernhardt e De Boer, 2005; Thanbichler e Shapiro, 2006) e Firmicutes (Wu e Errington, 2004), evidenciando que se trata de um sub-componente do divisoma que não é conservado entre as várias linhagens de bactérias (Tabela 1), o que é compatível com seu caráter não-essencial (Rothfield, Taghbalout *et al.*, 2005). A proteína SlmA, codificada em *ttk*, é responsável pelo efeito de oclusão pelo nucleóide em *E. coli* (Bernhardt e De Boer, 2005) e contém um domínio TetR_N, um motivo (“fold”) do tipo hélice-volta-hélice (“helix-turn-helix” ou HTH), encontrado na extremidade N-terminal da proteína TetR, repressor do gene de resistência à tetraciclina (*tetA*), bem como de outros reguladores de transcrição (Hillen e Berens, 1994). Muito embora o domínio N-terminal de SlmA, responsável pela interação com o nucleóide, seja bastante conservado, dificultando a distinção entre ortólogos de SlmA e outras proteínas não homólogas que contém esse domínio (Wintjens e Rooman, 1996; Aravind, Anantharaman *et al.*, 2005), buscas por ISS (“inclusion threshold” $\leq 1e^{-10}$) identificaram homólogos de *slmA* que apresentam homologia também na região C-terminal, que contém um segmento “coiled-coil” e parece constituir um domínio independente, ainda não caracterizado. A Figura 25 mostra a filogenia dos ortólogos de *ttk* identificados por ISS, que corresponde à mesma topologia da filogenia de referência para as espécies incluídas.

Figura 25: Filogenia dos ortólogos do gene *ttk* (*slmA*), idêntica à relação entre as espécies na filogenia de referência.

Figura 26: Representação circular para as relações de similaridade entre os homólogos da família CbiA. O agrupamento hierárquico acima foi produzido aplicando o método de “neighbour-joining” ao alinhamento das regiões homólogas à assinatura CbiA do PFam. O modelo de substituição de aminoácidos Poisson, com correção gama ($\alpha = 1$) (Kumar e Nei, 2000) foi usado para estimar as distâncias entre as proteínas. Os nomes em negrito identificam as sub-famílias. Homólogos de MipZ são destacados em um subgrupo da família ParA. Os sub-grupos de ParA em amarelo e laranja são encontrados em cromossomos e plasmídeos, respectivamente.

Figura 27: (página anterior) filogenia dos homólogos de *parB*. Foram incluídos apenas os genes localizados em cromossomos, excluindo todos os homólogos de plasmídeos, exceto o homólogo plasmidial de *E. coli* (*sopB*, embaixo), usado como grupo externo. O grupo destacado na parte de baixo da filogenia (amarelo) compreende todos os homólogos de *yjaA*, que em *B. subtilis* codifica a proteína Noc, mostrando sua monofilia e distribuição restrita a Firmicutes.

Em Firmicutes, a oclusão pelo nucleóide é mediada pela proteína Noc, codificada pelo gene *yyaA* (Wu e Errington, 2004), homólogo da proteína ParB que, como SlmA, contém motivos HTH, responsáveis pela interação com o DNA. A Figura 27 mostra a filogenia dos homólogos de *parB*, incluindo *yyaA*, que forma um grupo monofilético junto à base do grupo de homólogos localizados em cromossomos. Também no caso desta família, as relações entre filos não são resolvidas, resultando em agrupamentos com valores de bootstrap reduzido, que são mais elevados apenas para grupos mais recentes. Esse padrão permite diferenciar os ortólogos de *yyaA* dos seus “out-parálogos” em Firmicutes, denominados *spo0J* em *B. subtilis* e que participa, como comprovado para outros membros da família *parB*, da segregação do cromossomo (Yamaichi e Niki, 2000). Os homólogos de *yyaA* e *spo0J* se localizam em um agrupamento conservado, composto pelos genes *yyaA-soj-spo0J*, presente na maioria das Firmicutes (ausências notáveis para esse agrupamento são observadas em todas as Mollicutes e na maioria das Lactobacillales). Onde o agrupamento dos três genes está ausente mas pelo menos um homólogo de *parB* é encontrado, o par cognato da família *parA (soj)* também é encontrado no mesmo operon, na ordem *parA-parB* (Yamaichi e Niki, 2000). Esses dados combinados sugerem que *yyaA* se originou a partir de uma duplicação de um homólogo de *spo0J* em um ancestral comum a todas as Firmicutes.

A última proteína conhecida que participa do processo de oclusão pelo nucleóide é a proteína MipZ, presente em *Caulobacter crescentus* e outras α -proteobactérias. Essa proteína é, como *soj*, um homólogo de *parA* mas atua, além do processo de NO, inibindo a formação do anel nos pólos da célula e, portanto, complementando o papel do sistema Min, que não tem homólogos em *Caulobacter* (Thanbichler e Shapiro, 2006). Diferentemente de Soj, que tem a capacidade de interagir diretamente com o DNA e atua na regulação de proteínas de esporulação (Real, Autret *et al.*, 2005), MipZ depende de interações com ParB

Figura 28: Vizinhança dos homólogos de *divIVA* nos genomas completos do banco de dados STRING (versão 7.0). As últimas duas colunas à direita mostram a vizinhança conservada para o parâlogo *ypsB*. Os homólogos de cianobactérias e actinobactérias são vizinhos a membros do COG0669, diferindo, portanto, da vizinhança de *ypsB*. O ortólogo de *DivIVA* em cianobactérias não aparece na figura, mas está na vizinhança do operon *ylm*, como descrito por (Miyagishima, Wolk *et al.*, 2005) e validado por inspeção manual da anotação do genoma.

para se associar ao DNA e assim inibir a polimerização de FtsZ ao redor do nucleóide.

MipZ apresenta os dois motivos conservados, Walker-A e B, característicos de sua superfamília, que inclui homólogos de ParA, MinD, MRP, CobQ e CobB (Koonin, 1993), muito embora o modelo para a família MipZ, incluído na versão 22.0 recente do PFam, não incluía o motivo A (13-19), que é o mais conservado. Usando o programa hmalign e o HMM CbiA (Eddy, 1998), do banco de dados PFam, foi possível construir um alinhamento

de qualidade para os membros desta superfamília de proteínas presentes nos genomas completos do STRING. A Figura 26 mostra a árvore de “neighbor-joining”, inferida usando o programa MEGA (Kumar e Nei, 2000), a partir desse alinhamento, e permite verificar que MipZ, embora realize uma função análoga à de MinD, é muito divergente deste último e provavelmente se originou da duplicação e divergência de um gene semelhante aos genes *para* presentes em plasmídeos de várias bactérias (subgrupo da família *para*, em laranja, dentro da elipse).

3.3.16 DivIVA

A proteína DivIVA é uma proteína bifuncional, com atividades como determinante topológico do complexo MinC/D e em processos de esporulação (Thomaides, Freeman *et al.*, 2001). DivIVA forma oligômeros *in-vitro* e é caracterizada pela presença de um longa hélice “coiled-coil” e uma curta região estruturada na extremidade N-terminal, uma topologia semelhante à encontrada na proteína tropomiosina de eucariotos (Stahlberg, Kutejova *et al.*, 2004). Sua distribuição nos genomas completos que analisamos é caracterizada pela posição conservada nos genomas de Firmicutes (Hamoen, Meile *et al.*, 2006), cianobactérias (Miyagishima, Wolk *et al.*, 2005) e actinobactérias (Flardh, 2003), junto ao operon **ylm** e próximo à extremidade do agrupamento **dcw**, característica que nos levou a manter todas as sequências classificadas como membros do COG3599 na análise, mesmo quando sua similaridade com o modelo do PFam para esta família ficava abaixo do valor mínimo exigido para outras famílias (e-value > 0.01). Por outro lado, a vizinhança dos parálogos de DivIVA, em Firmicutes e actinobactérias não é tão conservada, sendo característica de cada parálogo (Figura 28). A Figura 29 mostra a filogenia dos homólogos de *divIVA* com similaridade significativa ao modelo do PFam. Como outras filogenias dos

Figura 29: Filogenia do gene *divIVA* e seu parálogo em Firmicutes (*ypsB*). A aparente parafilia dos homólogos de *B. longum* (BIFLO) carece de suporte estatístico (os valores de “bootstrap” são mostrados abaixo dos nós internos), o mesmo sendo observado para a posição dos homólogos de δ -proteobactérias (Delta) e Deinococci, incluídos entre as Firmicutes. Grupos como as Bacillales e Lactobacillales, além de todos os homólogos de *ypsB*, no entanto, mostram-se bem definidos (monofiléticos).

genes de divisão, elas mostram um padrão de herança vertical nos grupos em que há sinal filogenético satisfatório e sugerem que a duplicação do gene *ypsB* ocorreu pelo menos no último ancestral comum de Bacillales e Lactobacillales, mas pode ter ocorrido antes, no ancestral de todas as Firmicutes. Os parálogos de cianobactérias e actinobactérias, excessivamente divergentes para serem incluídos nessa análise, parecem ter origens

independentes.

3.3.17 Sistema Min

Os genes que codificam os componentes do sistema Min (*minC*, *minD* e *minE*) formam um operon conservado. MinD é quase inteiramente constituída pela região homóloga às ATPases da família ParA, contendo apenas mais uma região C-terminal característica, com potencial para formar uma hélice anfipática responsável pela interação com a membrana citoplasmática (Hu e Lutkenhaus, 2003). MinC apresenta dois domínios, distintos funcional e estruturalmente: o domínio N-terminal, responsável pela inibição da polimerização de FtsZ, e o domínio C-terminal, responsável pela dimerização e interação com MinD (Hu e Lutkenhaus, 2000; Cordell, Anderson *et al.*, 2001). MinE também forma dímeros e apresenta dois domínios funcionalmente independentes, sendo a extremidade N-terminal responsável pelo efeito inibidor sobre o complexo MinC/MinD e a extremidade C-terminal necessária para a formação de dímeros e a manutenção da especificidade topológica (King, Rowland *et al.*, 1999). A Figura 30 mostra a filogenia inferida por máxima verissimilhança para os homólogos de *minE*, na qual as proteínas de organismos pertencentes a um mesmo filo formam grupos suportados por valores elevados de “bootstrap”, reproduzindo até mesmo a monofilia das proteobactérias, que não foi obtida na análise de genes de divisão maiores, como *ftsA*. A relação de MinD com outras ATPases da mesma família é mostrada na Figura 26, onde fica evidente a divergência dos homólogos de MinD, separados dos outros grupos por um longo ramo interno. A Figura 31 mostra a filogenia dos ortólogos de *minD* e *minC* encontrados em bactérias. Na árvore de *minD* (esquerda), com exceção dos homólogos de Clostridiales, ϵ -proteobactérias e algumas proteínas isoladas, todas as proteínas formam grupos monofiléticos, equivalentes aos filos, as exceções não

Figura 30: Filogenia do gene *minE*, inferida a partir do alinhamento feito pelo programa hmalign com base no modelo (HMM) desta família disponível no Pfam. Não só a topologia é congruente com a filogenia de referência, mas os comprimentos dos ramos, pelo menos para as proteobactérias, parecem ter as mesmas proporções nas duas árvores.

apresentando suporte estatístico. Na filogenia da direita, inferida com base em um alinhamento do domínio conservado C-terminal de MinC, essencialmente os mesmos grupos são observados, embora com valores de “bootstrap” sistematicamente mais baixos e com uma posição anômala do homólogo de *L. plantarum*, suportada por valores de “bootstrap” baixos, porém não muito distantes do limite de significância (70%).

Figura 31: Filogenia dos homólogos de *minD* (esquerda) e *minC* (direita). Na análise de *minD*, incluímos todas as sequências incluídas do grupo monofilético de MinD na Figura 26. A filogenia de *minC* foi inferida a partir do alinhamento da região carboxi-terminal (domínio MinC_C do banco de dados PFam).

4. Discussão

4.1 Um catálogo da distribuição dos genes de divisão

Nosso levantamento da distribuição dos genes de divisão celular atualiza análises anteriores da distribuição desses genes em procariotos, apresentados por (Rothfield, Justice *et al.*, 1999; Margolin, 2000; Vicente, Tamames *et al.*, 2004). Dentre os resultados obtidos durante esse levantamento, é interessante ressaltar a detecção de alguns homólogos remotos de genes de divisão celular, não necessariamente envolvidos com o processo de divisão em si, que não podemos inferir apenas com base nas análises *in silico*, mas interessantes por indicarem os mecanismos evolutivos envolvidos na diversificação do divisoma. Os homólogos de *ftsQ* em Chlamydiae (Tabela 1), por exemplo, não haviam sido identificados em nenhum dos trabalhos citados acima, devido à pouca semelhança em termos de sequência, que também inviabilizou sua inclusão nas análises filogenéticas feitas para esse gene. Apesar da baixa similaridade, os homólogos de *ftsQ* em Chlamydiae estão localizados próximos a genes que codificam enzimas da via de síntese dos precursores do peptidoglicano e em outros genomas são encontrados no **dcw** (*murC*, *murD*, *murF*, *murG*, *mraY* e *ftsW*) (Tamames, 2001; Mingorance, Tamames *et al.*, 2004). Embora apenas no genoma de *Protochlamydia amoebophyla*, que não foi incluído neste estudo, este homólogo seja codificado por um gene na mesma fita dos genes **mur**, em todos os genomas de clamídias sequenciados ele aparece ao lado desses genes, sugerindo um rearranjo genômico durante a evolução das clamídias, e a possibilidade de que estes homólogos desempenham um papel na divisão celular e no desenvolvimento de corpos reticulados na fase vacuolar do ciclo de vida clamidial merece ser estudado (McCoy e Maurelli, 2006). O ISS ainda detectou homólogos de *ftsQ* que não foram incluídos no COG1589 em mais oito outras espécies: *Bdellovibrio*, *Fusobacterium*, *Aquifex*, *Leptospira* e membros do grupo

Bacteoidetes-Chlorobi. Em comum, esses organismos parecem apresentar um elevado grau de divergência das linhagens mais próximas nas quais o gene *ftsQ* foi identificado pelo COG (Figura 5), o que pode implicar que o alto grau de divergência de *ftsQ* acompanha uma tendência geral de seus genomas. O uso de buscas iterativas permitiu ainda a detecção de homólogos que não haviam sido identificados pelo COG para os genes *ftsZ*, *zapA*, *zipA* e *ftsN*.

Em alguns outros casos, porém, as revisões de COGs com base nos critérios de sintenia (similaridade da vizinhança no genoma), estrutura de domínios e filogenia levaram à exclusão de parálogos que formavam agrupamentos claramente distintos, como no caso de *ftsK*. Os homólogos de *minD* encontrados em arqueabactérias (grupo denominado “minD-like” na Figura 26) foram citados por Rothfield em seu levantamento da distribuição dos genes de divisão (Rothfield, Taghbalout *et al.*, 2005) como ortólogos de *minD* nesses genomas, o que leva naturalmente à hipótese de que essas proteínas desempenham um papel semelhante na inibição da polimerização de FtsZ em arqueabactérias. Apesar da topologia da filogenia para os membros da superfamília ParA posicionar os genes “minD-like” em um grupo distinto, junto com os genes *minD* de bactérias, esses genes não são tão claramente distintos de outros grupos de homólogos de ParA como MinD ou as enzimas CobQ e CobB, o que é evidenciado pelo comprimento de ramo curto na base desse grupo na Figura 26. Como efeito, variando o critério de similaridade (“inclusion threshold”) nas buscas de ISS que usavam os homólogos de *minD* como sementes, não foi possível identificar um valor máximo para esse critério que fosse capaz de isolar homólogos de *minD* e genes “minD-like”, sem retornar também membros da subfamília ParA envolvidos em conjugação e segregação. Finalmente, os homólogos em Archaea são encontrados na vizinhança de genes relacionados a atividades metabólicas ou com a segregação dos cromossomos, o que sugere que não estão diretamente relacionados à septação. Esse fatores nos levaram a desconsiderar os homólogos de *minD* presentes em arqueabactérias na

inferência dos genomas ancestrais. Com relação à presença de outro gene de divisão em arqueobactérias, também apontado por Rothfield (Rothfield, Taghbalout *et al.*, 2005), nenhum homólogo de *ftsL* foi detectado em *Methanobacterium thermoautotrophicum* usando buscas iterativas de similaridade (ISS) ou usando os modelos disponíveis no banco de dados PFam para o “clã” de FtsL. A ausência de um agrupamento **dcw** em arqueobactérias também impediu a inferência da presença de tais genes nesses organismos e, a partir de tais evidências, sustentamos que, com exceção de *ftsZ*, alguns homólogos remotos de *ftsA* e do homólogo de *amiC* em *M. mazei*, nenhum outro homólogo dos genes de divisão bacteriana foi identificado nos genomas completamente seqüenciados de arqueobactérias e, por este motivo, tais organismos não serão considerados mais adiante, quando estivermos discutindo as variações de composição do divisoma.

4.2 Divisoma mínimo

O primeiro aspecto da distribuição dos genes de divisão celular em bactérias que devemos enfatizar é que, ao contrário de outros sistemas como o ribossomo (Mushegian, 2005), nenhuma proteína do divisoma pode ser considerada universal, *i.e.* nenhum gene de divisão celular é encontrado em todas as linhagens atuais de bactérias. Como indicado na seção anterior, alguns componentes citoplasmáticos do divisoma bacteriano (Tabela 1), que são homólogos de proteínas encontradas nos três domínios da vida (Arqueobactéria, Bactéria e Eucariotos), como FtsZ (tubulina) e FtsA (actina), estão ausentes de algumas linhagens muito divergentes (*e.g.* Planctomycetales) e de genomas reduzidos, *e.g.* Mollicutes (Moran, 2002). Este padrão implica que não poderemos definir um divisoma bacteriano mínimo, que sirva de modelo para todos os grupos de bactérias, a partir de um conjunto universal de proteínas especializado em definir e remodelar o plano de divisão celular. Podemos ressaltar, no entanto, que as Planctomycetales e Mollicutes, que apresentam genomas desprovidos de todos ou da maioria dos genes de divisão celular,

também são desprovidos de paredes celulares compostas de peptideoglicano, o que sugere que, com a exceção de FtsZ e seus efetores - positivos e negativos - de polimerização, os genes de divisão conhecidos em procariotos são intimamente ligados, mesmo que indiretamente, à síntese do peptidoglicano septal. Portanto, levando esse fator em conta, não basearemos nossa definição do divisoma bacteriano mínimo em grupos de organismos que não apresentem paredes celulares compostas por peptidoglicano.

Considerando bactérias patogênicas, as clamídias são únicas por possuírem genes que codificam as enzimas da via de síntese dos precursores do peptideoglicano e homólogos de genes de divisão associados com a síntese de peptideoglicano septal (*ftsW*, *ftsI* and *amiC*), mas não apresentarem qualquer gene semelhante a *ftsZ*, o mais ubíquo dos genes de divisão bacterianos (Mccoy e Maurelli, 2006). Considerando o papel essencial de FtsZ para marcar a região onde ocorrerá a constrição, é provável que as clamídias possuam outro gene que codifique uma proteína com funções similares às de FtsZ, de forma análoga aos casos conhecidos de substituição de genes do ribossomo bacteriano por genes não-homólogos de arqueabactérias (Mushegian, 2005). Além das Chlamydiae e Mollicutes, *Campylobacter jejuni* é o patógeno com o menor complemento de genes do divisoma, apresentando apenas homólogos de *ftsA*, *ftsZ*, genes dos componentes da maquinaria de síntese do peptideoglicano e *ftsK*. *C. jejuni* é, portanto, a bactéria patogênica que mais se aproxima do que pretendemos definir como um divisoma mínimo funcional.

Focando organismos de vida livre com parede baseada em peptideoglicano, o genoma que mais se aproxima do que pode ser o conteúdo gênico de um divisoma mínimo é *Thermotoga maritima* (Nelson, Clayton et al., 1999). Tomando como base a classificação em sub-complexos proposta recentemente para genes do divisoma (Vicente e Rico, 2006), observamos que *Thermotoga* não possui nenhum gene que codifique para componentes da ponte periplasmática (*ftsQ*, *ftsB* e *ftsL*). Somem-se a essa observação as evidências que apontam vários episódios de perda deste complexo durante a evolução das bactérias (ver

abaixo) e podemos concluir que as funções da ponte periplasmática não são essenciais ou podem ser supridas por outros genes. Os genes que determinam a localização do anel Z (*minE* and *divIVA*) também estão ausentes em *Thermotoga* e, embora este organismo em particular possua os outros componentes do sistema Min (*minC* e *minD*), este sistema é considerado necessário apenas para a divisão simétrica em bactérias cilíndricas e não precisa ser considerado parte de um divisoma mínimo. *Thermotoga* não possui homólogos de *ftsK*, a translocase de DNA responsável por garantir a segregação dos cromossomos durante os estágios finais da divisão celular, mas foi sugerido recentemente (Lakshminarayan, Makarova *et al.*, 2004) que um homólogo remoto de *ftsK* encontrado em arqueobactérias, chamado *herA*, complementa sua função de translocase em cianobactérias e bactérias termófilas. Muito embora uma DNA translocase possa ser necessária para garantir a segregação e evitar a quebra dos cromossomos durante a divisão da maioria das bactérias (Errington, Daniel *et al.*, 2003), *ftsK* não é essencial para o crescimento vegetativo em *B. subtilis* e está ausente nos endossimbiontes do gênero *Buchnera*, que não têm nenhum homólogo de *ftsK* e *herA*. Este último exemplo parece ter se originado em paralelo às perdas de vários genes envolvidos com reparo de DNA em *Buchnera*, os quais não apresentam homólogos de proteínas que interagem com FtsK, como as recombinases XerC e XerD, o que sugere que estes organismos são capazes de sobreviver com uma maquinaria de replicação, recombinação e reparo que não interage com o complexo de divisão celular. Isto talvez tenha se tornado possível por causa da presença de várias cópias do genoma em *Buchnera* (Moran e Degnan, 2006). Levando-se em conta que (1) a maioria dos homólogos de HerA não possui as regiões transmembrana na extremidade N-terminal que são necessárias e suficientes para a localização septal de FtsK (Draper, McLennan *et al.*, 1998), (2) que não há evidências de que algum homólogo de HerA possa complementar o papel de FtsK no divisoma em *Thermotoga* e em outras bactérias e que (3) *spoIIIE*, o único homólogo de *ftsK* com localização septal confirmada, não é essencial à divisão, sua deleção

produzindo fenótipos normais de crescimento e divisão, sugerimos que FtsK não precisa ser contabilizada como componente de um divisoma mínimo.

A inclusão de um homólogo de *ftsA* no divisoma mínimo pode parecer contraditória, uma vez que esta proteína não está presente em grupos como cianobactérias e actinobactérias. No entanto, homólogos de *ftsA* estão presentes em todas as proteobactérias com divisomas reduzidos (*Buchnera*, *Rickettsia* e *Campylobacter*) e em *Thermotoga*. FtsA desempenha vários papéis, participando dos estágios iniciais da montagem do anel Z, nos quais promove a polimerização de FtsZ (Geissler, Elraheb *et al.*, 2003), recrutando as proteínas seguintes ao divisoma nascente (Pichoff e Lutkenhaus, 2002) e facilitando o direcionamento de FtsZ para a face interna da membrana celular (Pichoff e Lutkenhaus, 2005). A observação de que *ftsA* e *ftsZ*, mas não *ftsQ* e os genes *mur*, estão ambos ausentes nos genomas de clamídias, indica que FtsA não é necessário na ausência de FtsZ. Por outro lado, embora esta proteína não seja essencial em *B. subtilis* e tanto YlmF (Ishikawa, Kawai *et al.*, 2006) como ZapA (Gueiros-Filho e Losick, 2002) apresentem funções que se sobrepõem parcialmente a alguns dos papéis desempenhados por FtsA, ressaltamos que as funções desempenhadas por FtsA na divisão celular parecem ser sempre necessárias e supridas, quando não por FtsA propriamente dita, por um outro componente do divisoma. Novas proteínas, como YlmF, ou novos domínios de interação, como o encontrado no homólogo de FtsW de *Mycobacterium tuberculosis*, que permite uma interação direta de FtsW e FtsZ neste organismo (Datta, Dasgupta *et al.*, 2002), podem explicar a ausência de FtsA nas actinobactérias e cianobactérias. Assim, sugerimos que um análogo de FtsA deve ser considerado integrante do divisoma mínimo.

Os divisomas reduzidos descritos acima sugerem que um divisoma mínimo será composto de uma proteína que indica o sítio onde se formará o septo (FtsZ), de uma ou mais proteínas e/ou domínios de interação com funções análogas às de FtsA, *i.e.* que atuem como intermediários dos contatos entre FtsZ e a maquinaria de síntese do peptídeoglicano,

e de uma maquinaria especializada na síntese do peptideoglicano septal (FtsW e FtsI). Além desses componentes, muito embora *Thermothoga* e *Rickettsia* sejam exemplos de divisomas reduzidos sem homólogos de AmiC, a atividade catalítica desta e de outras autolisinas é necessária para completar a separação das células filhas, pelo menos em *E. coli* (Heidrich, Templin *et al.*, 2001) e *B. subtilis* (Ohnishi, Ishikawa *et al.*, 1999). Mais importante: a síntese e remodelagem da parede celular no septo requerem a atividade coordenada de mureína sintases (FtsI/PBP2) e transglicosilases (PBPs da classe A, de alto peso molecular) e estudos têm demonstrado que as hidrolases de mureína formam um complexo que interage e se co-localiza com as mureína sintases no septo de divisão (Scheffers e Errington, 2004; Scheffers, Jones *et al.*, 2004), constituindo, portanto, componentes do divisoma. Muito embora nenhuma PBP da classe A tenha sido incluída em nossas análises de estados ancestrais, homólogos destas proteínas estão presentes nos genomas com divisomas reduzidos e acreditamos que estas proteínas devem também ser consideradas membros do divisoma bacteriano mínimo associado à parede celular de peptideoglicano.

Figura 32: Modelos esquemáticos para os divisomas mínimo (A,B) e ancestral (C,D). OS números nos painéis (A) e (C) identificam os sub-componentes do divisoma, como indicado na Figura 1 (pág. 18). O divisoma hipotético mínimo, ilustrado acima, foi inferido com base na distribuição dos componentes mais conservados em divisomas reduzidos. O divisoma ancestral foi inferido aplicando-se modelos de máxima verossimilhança sobre a mesma distribuição de genes e se mostrou consideravelmente mais complexo e completo que um complemento mínimo de genes conservado, indicando que o LCAB já se reproduzia usando mecanismos semelhantes ao encontrados na maioria das bactérias atuais.

4.3 Divisoma ancestral

As variações observadas na composição do divisoma sugerem que, em torno dos componentes obrigatórios de um divisoma mínimo, a evolução introduziu alguma variação entre as linhagens de bactérias, enquanto preservava importantes contatos e mecanismos de interação entre componentes centrais do divisoma. Uma vez que o conhecimento detalhado dos eventos evolutivos, *e.g.* perda e ganho de genes, que determinaram as mudanças na composição de divisoma permitiriam elaborar hipóteses sobre os mecanismos de divisão em bactérias, aplicamos métodos de inferência filogenética (Pagel, 1999a) para reconstruir essa série de eventos evolutivos. A reconstrução destes eventos exigiu a inferência da composição do divisoma em todos os ancestrais das linhagens atuais de bactérias, que foi obtida mapeando a distribuição dos genes de divisão numa tabela binária e inferindo o padrão de presença/ausência desses genes nos nós internos da árvore da vida proposta por Ciccarelli et al. (Ciccarelli, Doerks *et al.*, 2006) (Figura 5). Tal mapeamento binário foi imposto pelas limitações do modelo de máxima verossimilhança usado e direciona o foco de nossas inferências para as primeiras aparições de cada família gênica em cada linhagem (Lester, Meade *et al.*, 2006). As reconstruções de máxima verossimilhança foram depois complementadas pela análise filogenética de cada família gênica. Notavelmente, a inclusão de linhagens sem uma parede celular composta por peptidoglicano não teve um impacto significativo nas inferências da composição do divisoma no último ancestral comum das bactérias (Tabela 4).

O divisoma inferido para o LCAB é muito mais complexo que o apresentado para o divisoma mínimo (seção anterior) e consideravelmente mais complexo que a composição, em termos de genes conhecidos, dos divisomas mais reduzidos em organismos com peptidoglicano (*Thermotoga* e vários parasitas). Isto implica que, enquanto as linhagens de bactérias divergiam umas das outras, ocorreu uma redução do divisoma, pela simples perda

de genes, ou a substituição dos genes presentes no LCAB por genes específicos de cada linhagem, gerando divisomas característicos de cada grande grupo de bactérias (Tabela 2). A semelhança do divisoma ancestral com aquele das linhagens mais antigas de bactérias (Clostridiales), inferida mesmo na ausência deste grupo, é outro indicador da provável natureza do divisoma ancestral, um complexo protéico essencialmente completo já na primeira bactéria. A presença de complexos protéicos elaborados nos ancestrais mais remotos também foi reportada em outros estudos, como na reconstrução do primeiro ribossomo no LUCA (Mushegian, 2005), do “spliceossomo” no primeiro eucarioto (Collins e Penny, 2005) e do complexossomo em eucariotos (Ceulemans, Beke *et al.*, 2006). Importante notar que muitos dos genes mais conservados do divisoma (*ftsZ*, *ftsA*, *ftsW*, *ftsI*) se localizam, na maioria das linhagens, no agrupamento **dcw** e, uma vez que a conservação e antiguidade desse agrupamento gênico é subordinada à antiguidade dos genes, é interessante especular que o próprio agrupamento **dcw** corresponda à organização desses genes no LCAB, como proposto por Tamames e colaboradores (Vicente, Tamames *et al.*, 2004). Esses últimos autores reportaram (ver Introdução) a existência de uma correlação entre transições para morfologias não-cilíndricas, como cocos e hélices, e a tendência para perda da integridade do agrupamento **dcw** e propuseram (Tamames, González-Moreno *et al.*, 2001; Mingorance, Tamames *et al.*, 2004; Vicente, Tamames *et al.*, 2004) que a degradação do **dcw** seria viabilizada pelo relaxamento da pressão seletiva negativa sobre o agrupamento em células não-cilíndricas (cocos e hélices), por que apenas em bacilos ele representaria uma vantagem adaptativa. A teoria proposta por esses autores sustenta que a vantagem adaptativa do agrupamento **dcw** deriva do acoplamento de transcrição, tradução, secreção e atividade enzimática de seus produtos gênicos durante a divisão. A idéia central é que este acoplamento aumenta significativamente a síntese de peptideoglicano septal, permitindo que a atividade de síntese de precursores e do próprio peptideoglicano no septo supere a mesma atividade na parede lateral do cilindro, mudando o plano de síntese do

peptideoglicano e redirecionando a síntese da parede do crescimento (aumento do comprimento) para a divisão celular. Nossos dados sobre a distribuição dos genes e composição dos divisomas ancestrais são compatíveis com o modelo de Tamames. São também compatíveis com a hipótese de que um processo semelhante de relaxamento da pressão seletiva opere sobre a variação de composição gênica do divisoma, como sugerido por simulações da evolução do conteúdo gênico em organismos de genoma reduzido (Pal, Papp *et al.*, 2006).

4.4 Considerações sobre os métodos utilizados

A possibilidade de testar hipóteses evolutivas pelo mapeamento de caracteres aos nós internos de uma filogenia atrai grande interesse, mas os desafios associados com a reconstrução dos estados ancestrais para um único caractere são maiores que os encontrados na inferência de filogenias com base em milhares de caracteres (Cunningham, 1999). Tais desafios se tornaram evidentes com a introdução de métodos de inferência baseados em máxima verossimilhança, que revelaram o alto grau de incerteza associado às reconstruções (Cunningham, Omland *et al.*, 1998). No caso dos divisomas ancestrais, tentamos avaliar criticamente tais incertezas perturbando a distribuição dos genes por re-amostragem (seção 3.2). A remoção manual e aleatória de vários grupos e espécies de bactérias e arqueobactérias mostrou que, para a maioria dos genes, os resultados variam muito pouco. As exceções a esse padrão, mostradas na Figura 6 e na Tabela 5, são os genes *zapA*, *minE* e *ylmF*, encontrados apenas em linhagens distantes entre si na filogenia de referência. Para *ylmF* e *minE*, a análise de todas as espécies leva à inferência da presença desses genes no LCAB com 100% de probabilidade. Para *zapA* a probabilidade fica no limite da região de incerteza (60%) e a presença desse gene no ancestral é duvidosa. Se aceitarmos o cenário em que esses genes estão ausentes no ancestral universal, inferimos que um ou mais dos

ganhos independentes nas linhagens com homólogos correspondem a eventos de transferência lateral entre os ancestrais dessas linhagens. As análises filogenéticas, porém, sugerem fortemente que a evolução dos três genes em cada linhagem, após sua primeira aparição, foi exclusivamente vertical e duplicações observadas na evolução de *ylmF* e *zapA* são todas linhagem específicas e não foram implicadas em HGT (Figura 16, Figura 17 e Figura 30). Este padrão é diferente do observado para famílias do divisoma nas quais há casos conhecidos de transferência lateral, como *mecA*, que não apresentam evidências da monofilia de suas subfamílias de parálogos (próxima seção). O gene *minE* é encontrado em linhagens muito divergentes (cianobactérias, *Deinococcus-Thermus*, proteobactérias, fusobactérias e firmicutes). A remoção de *T. tengcongensis* e de todas as Clostridiales resulta em redução da probabilidade de presença de *minE* no LCAB para 53% e 73%, respectivamente, mas a identificação de *minE* em *Symbiobacterium* (Ueda, Yamashita *et al.*, 2004), pertencente a um ramo basal das Firmicutes, e em outros genomas de Clostridia, parecem implicar uma distribuição mais ampla desse gene na base das Firmicutes. Já o gene *ylmF* está restrito a quatro linhagens (cianobactérias, firmicutes, actinobacteria e *F. nucleatum*) e, embora a remoção de qualquer uma delas induza a inferência de ganhos independentes para as linhagens remanescentes, outras perturbações não afetam a presença desse gene no LCAB (Tabela 2). Diferente desses dois genes, porém, *zapA* é encontrado em uma espécie isolada (*Leptospira interrogans*), além de estar presente em Firmicutes e Proteobacteria. A presença dos homólogos de *zapA* apenas em *Leptospira*, mas não em outras Spirochaetes, favorece a hipótese de transferência lateral, mas essa evidência precisa ser corroborada pela análise dos genomas de outras Spirochaetes. As incertezas para a presença dos três genes no LCAB, introduzidas pela re-amostragem, foram sensivelmente menores para *minE*, mas não podem ser ignoradas para *ylmF* e *zapA*. No caso de *ylmF*, porém, a alta probabilidade inferida para a presença no LCAB quando usamos todos os dados e a conservação da vizinhança genômica para este gene, localizado junto a *divIVA* e

próximo ao **dcw**, ou seja, junto a vários genes que estavam presentes no divisoma ancestral, sugere que os efeitos da re-amostragem podem ser artefatuais. Por esses motivos contabilizaremos apenas *zapA* como um possível exemplo de transferência lateral antiga entre grupos muito distantes de bactérias. O único outro exemplo de HGT antigo é *ftsK* (ver próxima seção).

Além das limitações intrínsecas aos métodos de inferência de estados ancestrais, a reconstrução do conteúdo gênico para genomas ou complexos protéicos ancestrais esbarra nas dificuldades para identificação dos caracteres homólogos nos diferentes organismos, um problema análogo ao do alinhamento de resíduos ambíguos em sequências muito divergentes. Nesse estudo, análises filogenéticas e buscas exaustivas por genes similares (ISS) foram empregadas com o objetivo de minimizar o número de falsos negativos, *i.e.* o não reconhecimento de um homólogo válido, e falsos positivos (filogenias, arquitetura e contexto genômico), ou seja, que COGs abarcando grandes famílias gênicas não induzissem a inclusão errônea de um gene não relacionado à divisão celular. Embora nosso método para busca de homólogos remotos (ISS) não garanta a detecção de todos os homólogos e vários outros métodos propostos para esse fim tenham sido descritos (Fariselli, Rossi *et al.*, 2007), uma comparação qualitativa de resultados obtidos para o domínio tAMIC usando nosso “pipeline” e usando o servidor HHsenser (Soding, Remmert *et al.*, 2006), que é baseado na seleção dos resultados de buscas com sequências intermediárias pela comparação de “sequence profiles”, sugere que estes dois métodos produzem resultados semelhantes (dados não mostrados). Uma comparação sistemática do desempenho do ISS contra um conjunto de dados de sequências homólogas remotas e não-homólogas conhecido é necessária, para que se possa avaliar suas limitações e o número de falsos negativos introduzidos pelo método. Apesar das incertezas sobre o desempenho do ISS, o procedimento permitiu identificar alguns genes homólogos aos genes de divisão em organismos nos quais estes genes não haviam sido reconhecidos anteriormente, como os

homólogos de *FtsQ* em *Chlamydia* e o parálogo de *ftsN*, *dedD*.

A identificação de eventos de transferência lateral e da história das duplicações de genes no divisoma também revelou limitações metodológicas. A árvore de *ftsQ* (Figura 11) é o exemplo típico do resultado obtido para a maioria dos grupos de homólogos com apenas um representante por genoma (*ftsE*, *ftsQ*, *ftsX*, *minC*, *minD*, *minE*, *ezrA*, *sulA*, *zapA*, *zipA*), caracterizada por uma filogenia congruente com a de referência, principalmente no que se refere ao agrupamento de membros dos clados mais bem definidos, *i.e.* clados suportados por altos valores de “bootstrap” na filogenia de referência (Ciccarelli, Doerks *et al.*, 2006) e já descritos na literatura (Wolf, Rogozin, Grishin *et al.*, 2001). Porém, na maior parte das árvores, muitos agrupamentos apresentaram valores reduzidos de suporte por “bootstrap” não-paramétrico e, em alguns casos, não conseguiram reproduzir agrupamentos importantes (como nas filogenias de *ftsZ* e *ftsA*). Tais resultados já foram observados em outros estudos, que mostraram as dificuldades para a inferência de filogenias baseadas no estudo de genes isolados (Castresana, 2007), que são sujeitas à relação desfavorável entre o número de táxons analisados e o comprimento limitado do alinhamento. Diante dessas limitações, decidimos considerar a reprodução dos agrupamentos principais da filogenia de referência nas filogenias de cada gene como indicio da ausência de eventos de transferência lateral na evolução dessas famílias de proteínas (ver seção 4.5). Discussões sobre as duplicações de genes, por se concentrarem em episódios de duplicação mais recente, basearam-se sempre em agrupamentos para os quais foi possível obter valores elevados de suporte (seção 4.7).

Finalmente, é importante notar que algumas transferências de genes podem ter sido perdidas pela amostragem de táxons ou pelo método usado para identificar esses eventos (filogenias). A análise de novos genomas completos e um uso mais sistemático da identificação de ilhas genômicas com base em variações de composição (GC%) (Lio e Vannucci, 2000; Ochman, Lawrence *et al.*, 2000) e da presença de elementos móveis e seqüências que favorecem a integração desses elementos (Moreira, De Souza *et al.*, 2005),

poderão indicar mais eventos, inclusive em famílias não implicadas em HGT neste estudo. Deve-se ter em mente ainda a possibilidade de ter ocorrido substituição, em algumas linhagens, por análogos funcionais ainda desconhecidos, provenientes de linhagens nas quais o divisoma ainda não foi bem caracterizado. Esse pode ser, por exemplo, o processo que levou à perda de vários genes de divisão em *Thermotoga* e *Aquifex*, para os quais a transferência de um grande número de genes de arqueobactérias já foi descrita (Nelson, Clayton *et al.*, 1999).

4.5 Transferência Horizontal de Genes (HGT)

Os objetivos principais das análises filogenéticas das famílias de cada gene de divisão foram: (i) avaliar mais profundamente a possibilidade de que genes de divisão celular tivessem sido alvos de transferências horizontais, um processo que pode ser perdido pelo mapeamento binário requerido pelo método de máxima verossimilhança empregado na inferência do conteúdo gênico no ancestral (seção 3.2), e (ii) mapear eventos de duplicação gênica aos nós internos da filogenia de referência. Ao avaliar individualmente a reconstrução dos ancestrais e a filogenia dos genes de divisão, identificamos **6** famílias gênicas com genes que podem ter sido alvos de transferência lateral em algum momento de sua evolução: *ftsZ*, *ftsK*, *ftsI*, *amiC*, *ezrA* e *zapA*. Para os demais genes de divisão, as filogenias reproduzem a filogenia de referência e uma origem única é inferida pela reconstrução de estados ancestrais para todos os homólogos em todas as linhagens (Tabela 2), corroborando a hipótese de herança vertical.

A hipótese de transferência de *zapA* foi proposta para explicar as possíveis origens independentes desse gene em Firmicutes e Proteobactérias e teria ocorrido apenas entre os ancestrais dessas linhagens (seção anterior). A origem de *ezrA*, um regulador da montagem do anel Z em *Bacillus* que codifica uma proteína com as mesmas características topológicas

de ZipA mas nenhuma similaridade com essa ou outras proteínas de bactérias (Errington, Daniel *et al.*, 2003) pode ter envolvido a transferência de genes a partir de uma linhagem não-bacteriana. Homólogos de EzrA são encontrados apenas em membros das ordens Bacillales e Lactobacillales, mas apresentam similaridade significativa com proteínas de eucariotos (ver Resultados, seção 3.3.13). A aparição de *ezrA* no genoma do último ancestral comum de Bacillales e Lactobacillales pode ser explicada por uma origem independente (“de novo”) ou pela aquisição do fragmento de um gene de eucariotos. Novas análises de similaridade de sequências e a estrutura do domínio citoplasmático serão necessárias para estabelecer a procedência da segunda hipótese.

Dentre os eventos de HGT, a substituição por xenólogos já foi proposta para explicar a complementaridade da distribuição de duas proteínas, *ftsK* e *herA* (Lakshminarayan, Makarova *et al.*, 2004), em bactérias e arqueobactérias: *ftsK* está ausente em Cyanobacteria, Thermotogales, Aquificaeles e Fusobacteria, os mesmos táxons onde existe um homólogo de *herA*. Não existem, no entanto, evidências de que algum membro da família HerA se localize no septo e é possível que essas proteínas possuam apenas uma atividade de DNA translocase, comum a vários membros da superfamília dessas ATPases, incluindo FtsK. Quanto à evolução de *ftsK* em bactérias, nenhum indício de transferência lateral foi encontrado por Lakshminarayan *et al.* ou em nossas análises.

Eventos de transferência lateral podem ter afetado também famílias de componentes do divisoma diretamente envolvidos com o processamento do peptidoglicano (*ftsI* e *amiC*). Alguns dos homólogos do gene *amiC*, cuja filogenia e estrutura de domínios são representados na Figura 23, estão localizadas em plasmídeos, elementos de inserção e profagos, sendo, muito provavelmente, transferidos lateralmente por conjugação e/ou transfecção (Loessner, Maier *et al.*, 1997). Nosso método de análise, porém, buscou transferências laterais mais antigas e não permite identificar eventos desprovidos de sinal filogenético, pois os agrupamentos com valores reduzidos de suporte por “bootstrap” foram

considerados equivalentes à filogenia de referência e não havia meios de aumentar a confiabilidade nesses agrupamentos. Uma revisão cuidadosa da vizinhança das amidases e de outros genes de divisão, que identifique genes localizados em ilhas genômicas, poderá detectar HGT entre espécies muito próximas.

O outro caso de transferência lateral de um homólogo de enzimas do divisoma envolvidas com remodelagem do peptideoglicano é a propagação de *mecA*, um homólogo de *ftsI/pbp2*, entre cepas hospitalares de *Staphylococcus* resistentes à meticilina (MRSA) (Hanssen e Sollid, 2006). A proteína codificada por *mecA* permite a continuidade da divisão celular na presença de meticilina, pois sua baixa afinidade a esse e outros β -lactâmicos permite que essa proteína continue a síntese de peptideoglicano mesmo na presença do inibidor, que é capaz de bloquear a atividade regular de *ftsI* (Pinho, Filipe *et al.*, 2001; Lim e Strynadka, 2002). Esse gene fica localizado numa ilha genômica em *Staphylococcus* MRSA, denominada SCCmec (de “staphylococcal chromosomal cassette”), que contém transposases e elementos de inserção e é considerada um elemento móvel (Hanssen e Sollid, 2006). Vários estudos indicam a aquisição do gene *mecA* por populações anteriormente sensíveis à meticilina (Enright, Robinson *et al.*, 2002). A origem do gene *mecA*, no entanto, ainda não foi estabelecida (Hanssen e Sollid, 2006), embora nossos dados mostrem a presença de genes com a mesma arquitetura de domínios em outras Firmicutes e em Actinobacteria, incluindo o domínio N-terminal MecA_N que não está presente nos ortólogos de *ftsI* e *pbp2*. Por sua importância no surgimento e propagação de cepas de *Staphylococcus* resistentes à meticilina, o gene *mecA* constitui o exemplo mais notável e bem estabelecido de HGT para homólogos do divisoma e um caso no qual a pressão seletiva (antibiótico) que estimula a propagação do gene é conhecida. A interação direta de MecA com o complexo de divisão, porém, ainda não foi demonstrada.

Outro caso de HGT envolvendo parálogos de proteínas presentes no divisoma do LCAB envolveu homólogos remotos de *ftsZ* amplamente distribuídos em megaplasmídeos de

espécies aparentadas a *B. cereus* (Tabela 1) e em *Paenibacillus* (Vaughan, Wickstead *et al.*, 2004). O gene ortólogo de *B. anthracis*, chamado *repX*, codifica uma proteína com apenas 17% de similaridade à cópia de FtsZ presente no cromossomo (Vaughan, Wickstead *et al.*, 2004) e foi recentemente caracterizado como um iniciador da replicação do plasmídeo pXO1, que não possui outros iniciadores de replicação (Tinsley e Khan, 2006). A extremidade C-terminal dos homólogos de RepX não apresenta similaridade com a mesma região em FtsZ, sugerindo esse gene não participa ativamente da maturação do divisoma (Tinsley e Khan, 2006). *repX* forma um grupo com a subfamília mais divergente de parálogos de *ftsZ* em arqueobactérias (FtsZ3), juntamente com um homólogo distante encontrado em *C. acetobutylicum* (dados não mostrados). A elevada divergência dessas seqüências, porém, resulta em valores reduzidos de “bootstrap” e pode induzir a atração de ramos muito divergentes, mas não relacionados, na filogenia, o que nos impede de afirmar que *repX* ou o homólogo em *C. acetobutylicum* tenham de fato se originado de *ftsZ3*.

Os episódios de transferência lateral descritos acima sugerem que a maioria dos genes do divisoma não são alvos comuns para a transferência lateral. Excluindo os casos de transferências muito antigas (*zapA*, *ftsK* e *ezrA*), os genes implicadas em HGT são membros de famílias grandes, com ampla distribuição e históricos de duplicações gênicas (ver próxima seção). Infelizmente, não há evidências experimentais de que os genes dessa última categoria preservem as interações com a maquinaria de divisão celular que os parálogos que lhe deram origem mantiveram.

Ainda mais interessante é a observação de que nenhum gene no divisoma parece foi alvo de substituição por gene ortólogo. O impacto reduzido desse tipo de HGT pode indicar um mecanismo de evolução correlacionada para os genes do divisoma. Nesse cenário os genes de divisão celular acumulariam mutações compensatórias, que preservam as interações entre proteínas de uma dada linhagem, mas tornam esses componentes progressivamente incompatíveis com os de outras linhagens, reduzindo a frequência das

transferências bem sucedidas. Esta possibilidade é compatível com experimentos mostrando que FtsA and FtsZ não interagem apropriadamente com outros componentes do divisoma quando expressos em um sistema heterólogo (Ma, Sun *et al.*, 1997) e que, para viabilizar o funcionamento de FtsZ num sistema heterólogo, tanto o gene como o genoma receptor precisaram ser amplamente modificados (Osawa e Erickson, 2006).

4.6 Perda de Genes

A presença da maior parte dos genes de divisão conhecidos no genoma do último ancestral comum a todas as bactérias implica a perda de genes durante a evolução desse sistema. Tais eventos são destacados e relacionados à diversificação dos principais grupos de bactérias na Tabela 2 (células em vermelho, seção 3.2, página 7). Uma incidência mais elevada de perdas de genes de divisão é correlacionada com a redução de genomas, como no caso de endosimbiontes e parasitas. O melhor exemplo de redução acentuada do divisoma concomitante a um processo de redução generalizada do genoma é o caso dos membros da classe Mollicutes. Algumas espécies deste grupo (*Ureaplasma*, *M. pulmonis* and Candidatus *Phytoplasma asteris*) perderam todos os genes de divisão celular presentes no último ancestral comum a todas as Firmicutes, do qual as mollicutes descendem (Tabela 2). Outras mollicutes ainda preservam homólogos de *ftsZ* e *divIVA* e um homólogo de *ftsA* foi identificado recentemente em uma mollicutes associada a plantas (Zhao, Hammond *et al.*, 2004), sugerindo que a perda dos genes do divisoma em Mollicutes é um processo gradual. Outros dois exemplos nos quais a composição do divisoma pode estar sendo afetada pela redução do genoma são os endosimbiontes de insetos, da classe γ -Proteobacteria, e as α -proteobacterias da ordem Rickettsiales. As Rickettsiales perderam o par cognato de transportadores ABC *ftsE/ftsX*, o operon *min* (dispensável nesses organismos de morfologia não-cilíndrica) e o gene *amiC*, embora o caráter gradual dessa

redução, sugerido por outros estudos (Blanc, Ogata *et al.*, 2007), não seja aparente em nossos dados, pois todos as espécies analisados neste grupo apresentam o mesmo conjunto de genes do divisoma (Tabela 1). Quando comparados ao complexo de divisão celular das *Rickettsia*, os divisomas dos endosimbiontes da classe γ -Proteobacteria (*Buchnera* e táxons relacionados) experimentaram um impacto muito maior da perda de genes que, como em Mollicutes, parece ter se acumulado gradualmente nesses organismos: *B. floridans* e *W. brevipalvis* possuem um número maior de genes de divisão, tendo perdido apenas *ftsE/ftsX* e alguns genes específicos de γ/β -proteobactérias (*ftsN*, *sulA* e *zipA*). *Buchnera*, por outro lado, perdeu também *ftsK*, *zapA* e, nas linhagens mais divergentes, todos os componentes da ponte periplasmática (*ftsQ* e *ftsB/ftsL*). A perda de *ftsK* em *Buchnera* pode ter sido facilitada pela poliploidia dos membros destes organismos (Moran, Plague *et al.*, 2003), que reduziria a vantagem seletiva dessa proteína ao tornar redundante sua função de garantir a segregação dos cromossomos.

Entre as perdas de genes não diretamente associadas a uma redução generalizada do genoma, os exemplos mais interessantes são perdas independentes do sistema Min e a perda de *ftsA* em Cianobacteria e Actinobacteria. A ausência de homólogos de *ftsA* em Cianobacteria e Actinobacteria, como apontado durante a discussão do divisoma mínimo (seção 4.2), pode ter sido compensada nestas linhagens pela presença do gene *ylmF*, cujo homólogo complementa a atividade de *ftsA* em *B. subtilis* (Ishikawa, Kawai *et al.*, 2006). Perdas repetidas do operon *min* estão associadas a transições para morfologias não cilíndricas, circunstância na qual este sistema pode se tornar dispensável (Vicente, Tamames *et al.*, 2004). Notavelmente, tanto *divIVA* quanto *minE*, responsáveis pela especificidade topológica do sistema Min em bactérias Gram+ e Gram-, respectivamente, são considerados componentes do divisoma do LCAB (Tabela 2). A inferência de estados ancestrais, portanto, sugere que a determinação da distribuição do complexo MinCD e, conseqüentemente, das regiões disponíveis para formação do anel Z, era redundante no

LCAB. Outra possibilidade é que a proteína DivIVA, cujos homólogos em vários organismos foram implicados em outros processos celulares (Thomaides, Freeman *et al.*, 2001; Akiyama, Inouye *et al.*, 2003; Flardh, 2003), adquiriu seu papel topológico na linhagem das bactérias Gram-positivas, mas executava um papel diferente e desconhecido no LCAB. Adicionalmente, Cianobacteria e Clostridia são exemplos de linhagens atuais que incluem espécies com homólogos dos dois determinantes topológicos e algumas bactérias Gram- (δ-proteobactérias) são exemplos de organismos que não possuem sistema Min mas possuem homólogos de *divIVA*, o que reforça a idéia de que esta proteína é importante em vários processos não relacionados à especificidade topológica.

As perdas dos componentes da maquinaria de síntese de peptidoglicano, *ftsW* e *ftsI* (ver Tabela 1) também podem estar relacionadas a processos a mudanças na morfologia celular ou no mecanismo que determina a manutenção da morfologia. Os homólogos destes dois genes perdidos em *Streptococcus pyogenes* e Rhizobiales não foram os genes do divisoma propriamente ditos, mas os parálogos destes genes especializados no crescimento celular, *rodA* e *pbp2*, respectivamente. A perda do par cognato *rodA/pbp2* nessas duas linhagens foi acompanhada pela perda de *mreB*, um homólogo de actina que, em *E. coli*, ancora RodA junto à membrana celular e tem papel central na determinação da morfologia celular (Kruse, Bork-Jensen *et al.*, 2005). Assim, a perda de *rodA/pbp2* em *Streptococcus* pode estar relacionada a transições da morfologia cilíndrica para esférica, análogas às observadas em certos mutantes de *Bacillus* (Cheung e Vitkovic, 1988). No caso das Rickettsia, a perda desses genes pode estar ligada à existência de um sistema de crescimento (alongamento) de células cilíndricas independente de *rodA/pbp2* ou a mudanças na morfologia durante as associações simbióticas ou parasíticas com as plantas hospedeiras (Vicente, Tamames *et al.*, 2004).

A ausência de genes de divisão em linhagens onde o sistema ainda não foi bem caracterizado pode ser um artefato provocado pela grande divergência dos homólogos de

alguns desses genes nos organismos mais distantes. Para evitar esse tipo de artefato empregamos buscas exaustivas de homólogos remotos, mas é possível que alguns desses genes tenham divergido além da capacidade de detecção de nossos métodos. Outra possibilidade é a substituição dos genes de divisão ausentes por genes análogos, *i.e.* genes não-homólogos capazes de desempenhar a mesma função (“non-orthologous gene displacement”) (Galperin, Walker *et al.*, 1998). Embora a substituição por genes não-ortólogos seja muito interessante do ponto de vista da evolução do divisoma e já tenha sido observado em estudos semelhantes de outros complexos protéicos (Mushegian, 2005), esse é, infelizmente, um fenômeno difícil de detectar sem evidências experimentais.

Mecanismos que determinam a perda de genes podem envolver fatores seletivos e aleatórios. Complexos protéicos ou vias metabólicas responsáveis por processos fisiológicos indispensáveis, como nutrientes que o ambiente (ou o hospedeiro) não fornece, irão constituir subsistemas essenciais que impõem a necessidade de preservar um número mínimo dos genes que participam desses processos. Essa pressão seletiva limitará as perdas que um dado conjunto de genes pode sofrer mas, enquanto há redundância nesses sistemas essenciais, estudos baseados em simulações sugerem que sua composição exata será determinada por contingências históricas, *i.e.* pela ordem, basicamente aleatória, na qual os componentes de um determinado sistema são perdidos (Pal, Papp *et al.*, 2006). Ciccarelli *et al.*, no mesmo trabalho em que inferiram a árvore da vida usada neste estudo, descreveram uma correlação negativa entre a distância até a raiz da filogenia e o tamanho do genoma, com uma tendência dos genomas menores evoluírem mais rapidamente (Ciccarelli, Doerks *et al.*, 2006). Nossas estimativas de máxima verossimilhança para a composição do divisoma ancestral mostram que os efeitos de redução do genoma também afetam a maquinaria de divisão celular, embora a correlação entre o tamanho do genoma e o tamanho do divisoma não seja forte, como ilustrado, por exemplo, pela conservação dos genes não-essenciais *ftsE* e *ftsX* e outros genes de divisão no patógeno humano *Tropheryma*

whipplei, um exemplo conhecido de redução do genoma (Bentley, Maiwald *et al.*, 2003).

4.7 Genes exclusivos e duplicações de genes

Estudos anteriores sobre a evolução do conteúdo gênico atribuíram um papel dominante para a transferência lateral entre as possíveis fontes de novos genes (Gogarten e Townsend, 2005). A influência da transferência lateral na evolução dos genes de divisão, discutida na seção 4.5, contabilizou eventos de vários tipos, como as transferências de parálogos mediadas por elementos móveis (*repX* e *mecA*) ou a substituição por genes não-ortólogos (*ftsK/herA*). A origem e a expansão das outras famílias gênicas do divisoma foram determinadas por mecanismos diferentes, como a origem *de novo* e a duplicação gênica. Origem *de novo* é a criação de novos genes a partir de regiões intergênicas (não codificantes), um processo que é considerado raro, pelo menos para genes inteiros (Long, Betran *et al.*, 2003). Nos grupos onde observamos inovações do conteúdo gênico, o alto grau de divergência dos genomas eliminou a informação originalmente presente nas seqüências não-codificantes sobre a origem *de novo* dos componentes do complexo e, em virtude disso, podemos apenas especular que esse processo esteve envolvido na origem dos genes de divisão presentes no LCAB e que não possuem homólogos em outros sistemas, famílias gênicas ou organismos, como *ftsQ*, *minC* e *ylmF*.

Entre os genes de divisão encontrados apenas em certos grupos de bactérias (que chamamos linhagem-específicos), a origem de *zipA* foi o evento para o qual encontramos menos informação, pois este gene é o único com distribuição restrita para o qual nenhum homólogo ou seqüência remotamente similar pôde ser identificada fora do grupo das β/γ -proteobactérias. Para outros genes linhagem-específicos, a similaridade a homólogos remotos sugere origens a partir de eventos de duplicação gênica, como no caso dos genes *sula* e *slmA* (β/γ -Proteobactérias), *mipZ* (algumas α -Proteobactérias) e *noc* (*yyaA* de

Firmicutes). *SulA* apresenta similaridade significativa com genes presentes em α - e β -Proteobactérias e com *RecA* (Cordell, Robinson *et al.*, 2003) (Figura 18). O homólogo de *Caulobacter crescentus* (*imuA*) parece não estar diretamente envolvido com divisão celular, mas com mecanismos de reparo do DNA (Galhardo, Rocha *et al.*, 2005) e pode ter sido transferido lateralmente de proteobactérias para *R. báltica* (Planctomycetes) (Figura 18). Homólogos de *sulA* e *imuA* em β -Proteobactérias, *Pseudomonas*, *Vibrio* e Xanthomonadales são encontrados na mesma vizinhança gênica que os homólogos de *imuA* em α -Proteobactérias, sugerindo um cenário no qual *imuA* é originado por duplicação de *recA* e depois dá origem a *sulA*, que evolui rapidamente para novas funções, como observado em outros casos de paralogia (Lynch e Conery, 2000). Cenários similares podem ser evocados para *mipZ* e *minD*, ambos membros da superfamília ParA/CbiQ (Figura 26) que atuam nos processos de inibição da polimerização de FtsZ nos pólos e na região ocupada pelo nucleóide. Ainda considerando a evolução do mecanismo de oclusão pelo nucleóide, é possível que o recrutamento de parálogos de famílias gênicas distintas em diferentes linhagens (*mipZ* em α -proteobactérias, *slmA* γ/β -proteobactérias e *noc* em Firmicutes) tenha sido facilitado pela presença do sistema Min nos ancestrais de cada uma dessas linhagens (Tabela 2), pois constatou-se recentemente que o efeito inibidor de MinC se estende além dos pólos, ainda que com intensidade reduzida (Rothfield, Taghbalout *et al.*, 2005).

Alguns parálogos de genes já presentes no divisoma do LCAB podem ter preservado interações com outros componentes desse sistema, passando elas próprias a desempenhar um novo papel nesse sistema. Um exemplo desse processo é a evolução dos genes *ftsB* e *ftsL*, ambos componentes atuais da ponte periplasmática que as análises de homologia baseadas em comparações de HMMs (ver Resultados) indicaram ter origem na duplicação de um gene ancestral comum. Após essa duplicação, que já estava presente no LCAB, os dois genes foram perdidos independentemente em várias linhagens e divergiram

acentuadamente entre si. Porém, em algumas linhagens nas quais ambos os genes estão presentes, eles ainda preservaram suas interações e formam um sub-complexo estável e dependente com FtsQ (Buddelmeijer e Beckwith, 2004). Outro caso interessante de duplicação que possivelmente preservou funções relacionadas à divisão celular é observado na família de *divIVA*. Os parálogos desse gene em Firmicutes, conhecidos como *ypsB*, aparecem pelo menos no último o último ancestral comum de Bacillales e Lactobacillales, como indicado pela similaridade das seqüências e a conservação da vizinhança genômica. Embora *ypsB* apresente uma deleção dentro da região homóloga à região de *divIVA* que codifica a α -hélice “coiled-coil” necessária para oligomerização (Stahlberg, Kutejova *et al.*, 2004), resultados preliminares em nosso laboratório (J.R. Tavares e F.J. Gueiros-Filho, resultados não publicados) mostraram que o padrão de localização de YpsB é muito semelhante ao de DivIVA, sugerindo que YpsB ainda desempenha um papel na divisão celular.

A origem do gene *mecA*, homólogo de *ftsI* do qual já falamos ao discutir episódios de transferência lateral, envolveu a duplicação de um homólogo de *ftsI* em uma linhagem desconhecida e a incorporação do domínio MecA_N, mas preservou as interações da proteína MecA com o divisoma, essenciais para seu papel como gene de resistência à meticilina. É possível que outros parálogos desses dois genes, muitos dos quais resultados de duplicações recentes (dados não mostrados) ainda preservem interações e um papel no processo de divisão celular, porém, ambas as famílias incluem exemplos de duplicações que levaram genes de divisão a adotar funções em outros processos celulares. Os genes cognatos *rodA* e *pbp2*, por exemplo, são parálogos de *ftsW* e *ftsI*, respectivamente, e já estavam presentes no LCAB (Figura 15), mas atuam especificamente no crescimento (alongamento) da célula e não há indícios de que interajam com o divisoma. Outros parálogos foram implicados em processos de esporulação e participam da síntese de uma parede celular diferenciada durante a formação do endósporo (Scheffers, Jones *et al.*,

2004). Os vários parálogos de *ftsW* presentes em *Listeria*, por outro lado, devem estar envolvidos em outros processos, uma vez que este organismo não forma endósporos. A história detalhada das duplicações de *ftsW* e *ftsI* ainda requer mais análises, tanto experimentais como *in-silico*.

A proteína FtsK já foi alvo de estudos evolutivos (Lakshminarayan, Makarova *et al.*, 2004), que descreveram a relação dessa proteína com outros membros da mesma superfamília de ATPases, que atuam em processos como geração de força mecânica para transporte de DNA durante a conjugação (TrwB) (Moncalian, Cabezon *et al.*, 1999) e secreção de proteínas (*e.g.* ESAT-6) (Pallen, 2002). Como discutimos ao tratar de transferência lateral, Lakshminarayan *et al.* demonstraram a provável substituição de *ftsK* pelo gene *herA* em algumas linhagens. Nossa análise filogenética complementa essas análises ao demonstrar a existência de parálogos recentes de FtsK em algumas β -Proteobactérias, em Rhizobiales e Brucellaceae (α -Proteobacteria) e em Bacillales e *Lactobacillus plantarum* (Firmicutes). Os parálogos de *ftsK* em Firmicutes (*ytpT*) não têm hélices transmembrana em suas extremidades N-terminais, regiões necessárias para a localização de *ftsK* no septo durante a divisão (Yu, Tran *et al.*, 1998). Apesar disso, a maioria dos homólogos de *ytpT* estão localizados perto de homólogos de *murC*, que codificam a enzima UDP-N-acetilmuramoyl-L-alanina sintetase, responsável pela síntese de precursores do peptidoglicano. A conservação da proximidade com *murC* sugere que *ytpT* pode ainda desempenhar papel em processos de divisão celular e crescimento, talvez através de interações com FtsK. Duplicações do gene *ftsK* em α -Proteobactérias podem ter ocorrido mais de uma vez (dados não mostrados) e uma função hipotética para estes genes não pode ser proposta com base no contexto genômico, devido à inexistência de uma vizinhança conservada. Entre as β -Proteobactérias, *Ralstonia*, *Neisseria* e *Chromobacterium* possuem duplicações de *ftsK* mas, ao contrário dos outros parálogos de *ftsK*, estes genes codificam proteínas que possuem uma região com hélices transmembrana,

que, em *FtsK*, são responsáveis pela interação com outras proteínas do divisoma. Comparando a filogenia de referência e a dos genes, inferimos que a duplicação nas β -Proteobactérias provavelmente precede a divergência dos três gêneros, implicando a perda deste parálogo no ancestral do gênero *Bordetella* (ver figura Figura 8). Como no caso dos outros parálogos de *ftsK* citados, não há evidências experimentais ou dados de contexto genômico que sirvam como evidência para inferência das funções desses parálogos em β -Proteobactérias, mas a preservação do domínio transmembrana e a alta similaridade com os ortólogos de *ftsK* torna factível a idéia de que esses genes ainda preservam regiões de interação com outros componentes do divisoma e pode ainda desempenhar um papel nesse processo.

A família de *amiC* é outro exemplo no qual episódios de duplicação e a inovação através da recombinação de domínios produziram uma ampla variedade de genes, alguns dos quais podem ter sido alvos de transferência lateral. A filogenia para essa família, representada na Figura 22, não permita discernir as relações entre as subfamílias de *amiC*, mas deixa claro que é comum a herança vertical dos parálogos dentro de um amplo grupo de bactérias, ou seja, para membros de mesmo filo ou classe (identificados pelas cores na figura). A Figura 23 mostra esse padrão de herança vertical em Proteobacteria, onde duas duplicações produziram os parálogos encontrados atualmente em *E. coli* (*amiA*, *amiB* e *amiC*). Experimentos de localização e a perda do domínio tAMIC sugerem que *amiA* não atua no processo de divisão (Bernhardt e De Boer, 2003). Novos experimentos serão necessários para verificar se outros membros desta família, incluindo os presentes em organismos modelo como *E. coli* e *Bacillus*, atuam no processo de divisão ou são exemplos de parálogos especializados em outras funções.

Os genes *zapA* e *ylmF* pertencem a famílias com distribuição, abundância e diversidade funcional menores que a das famílias descritos acima (*ftsK*, *ftsI*, *ftsW* e *amiC*). Apesar disso, exemplos de duplicações são observados no ancestral dos atuais membros do gênero

Pseudomonas. Estes parálogos apresentam mutações nos resíduos necessários para a interação de ZapA com FtsZ (resultados não mostrados) e é provável que tenham perdido a capacidade de auxiliar a polimerização e o enfeixamento de FtsZ. Parálogos de *ylmF* foram identificados em cianobactérias, actinobactérias e *Fusobacterium* e correspondem a duplicações independentes em cada uma dessas linhagens. Uma das duplicações, ocorrida em um ancestral do gênero *Streptomyces* e compartilhada por todas as espécies analisadas desse gênero constitui um exemplo interessante de alteração da taxa de evolução em parálogos, como se vê pelo comprimento do ramo dos homólogos desse gênero que estão na base da Figura 17.

O gene *ftsA* está presente nos genomas analisados ao lado de outros homólogos de actina, como *mreB* e *hsp70* (Bork, Sander *et al.*, 1992). Essas famílias, porém, compreendem homólogos distantes, dos quais o ramo de *ftsA* se separou durante ou antes da evolução do LCAB. Em nossa análise identificamos novos parálogos que correspondem a membros das famílias de *ftsA* propriamente dita, presentes em algumas firmicutes, *Thermotoga* e na arqueabactéria *M. kandleri*. Duas hipóteses podem explicar a distribuição de *ftsA2* entre algumas poucas bactérias e sua presença em *M. kandleri*: transferência lateral de ou para *M. kandleri* ou a perda desse parálogo nos ancestrais das linhagem nas quais ele não é encontrado. Tanto a posição de Q8TY55_METKA, que diverge quase simultaneamente de *ftsA* e *ftsA2*, como a probabilidade de presença de um homólogo de *ftsA* no ancestral universal (>80%), são compatíveis com a hipótese de origem vertical para o homólogo de *M. kandleri*, embora esta hipótese implique muitos eventos de perda. Análises do conteúdo de GC% desses genes, realizadas usando os perfis de composição disponíveis no “Comprehensive Microbial Resource” (Peterson, Umayam *et al.*, 2001), não indicaram desvios de composição da *ftsA2* em relação à composição do genoma de firmicutes, mas não foram conclusivos com relação ao homólogo da arqueabactéria *Methanopyrus* que pode, portanto, ter sido adquirido lateralmente. Se esta hipótese de HGT

para *M. kandleri* estiver correta, o cenário mais parcimonioso para a origem de *ftsA2* seria a duplicação de *ftsA* em um ancestral das firmicutes. Quanto à função desse gene, é interessante notar que nas espécies de firmicutes nas ele é encontrado, também possuem genes associados à esporulação e a resposta em condições de estresse que só são encontrado nessas linhagens (dados não mostrados). É possível, portanto, que este parálogo atue nessas condições, talvez direcionando proteínas especializadas para o septo de divisão.

Os ultimo exemplos de duplicação de gene do divisoma que caracterizamos foram os parálogo *ftsZ*, que aparecem independentemente em algumas linhagens. Parálogos desse gene em arqueobactérias remontam ao último ancestral comum das Euryarchaeota, como descrito por Vaughan *et al.* (Vaughan, Wickstead *et al.*, 2004). Outros episódios de duplicação de *ftsZ* são observadas em Rhizobiales (Figura 7) e outras proteobactérias, como *Magnetospirillum magnetotacticum* e *Colwellia psychrerythraea* (Vaughan, Wickstead *et al.*, 2004). As filogenias de Vaughan e colaboradores e as que inferimos para os homólogos de FtsZ presentes nos genomas do STRING são congruentes com a nossa filogenia de referência, exceto para os homólogos de *repX*, que circulam em plasmídeos de *Bacillus cereus* e podem ter sido transferidos de/para arqueobactérias (ver seção anterior).

Os eventos de duplicação gênica identificados em famílias de genes do divisoma incluem parálogos que atuam em especializações da maquinaria de divisão celular - como os parálogos de genes da maquinaria de síntese do peptideoglicano que atuam especificamente na esporulação – e em funções completamente não relacionadas, como parece ser o caso dos parálogos de *ftsK* encontrados no cromossomo linear das Rhizobiales. Adicionalmente, a provável origem de alguns genes do divisoma, exclusivos de certas linhagens, na duplicação de genes que não interagiam diretamente com componentes do divisoma sugere que novos genes de divisão celular usualmente representam um ganho de função em famílias que, antes do aparecimento desses novos genes, não possuíam membros que interagiam com o divisoma. Tais fenômenos provavelmente implicam o ganho de

novos contatos intermoleculares, que substituem ou acrescentam funções acessórias ao sistema, como a inibição da formação do anel nos pólos, a oclusão pelo nucleóide ou novos pontos de regulação do divisoma. Finalmente, o fato de que apenas as maiores famílias gênicas do divisoma parecem incluir exemplos de transferência lateral sugerem uma relação entre a incidência de duplicações e transferência lateral. Muito embora seja óbvio esperar que famílias que cujos membros são transferidos lateralmente apresentem mais parálogos, simplesmente porque seu número aumenta a cada nova aquisição, esse não é o padrão que observamos para as famílias do divisoma. A maior parte dos membros adicionais das famílias com mais de um representante por genoma são originados de duplicações do homólogo que participa da divisão celular e o único caso previsto de substituição por xenólogo, *ftsK/herA*, ainda não teve confirmado o envolvimento do gene adquirido com divisão. Assim, a transferência de genes de divisão parece depender quase sempre da duplicação de genes do divisoma e, considerando que boa parte dos genes linhagem-específicos se originam uma única vez (*slmA*, *noc*, *mipZ*) e tendem a não der transferidos, a HGT parece raramente desempenhar um papel relevante na evolução do divisoma exceto, talvez, no que se refere à aquisição de genes de resistência (*mecA*).

5. Conclusões

A premissa para a inferência de função a partir dos dados de contexto genômico é o reconhecimento de que os genes identificados em novos organismos fazem parte de módulos funcionais conservados e previamente caracterizados em outros organismos (Ceulemans, Beke *et al.*, 2006). Com base na mesma premissa, buscamos definir os componentes conservados e os módulos do divisoma bacteriano pela reconstrução de sua história evolutiva. Nossos resultados indicam que a primeira bactéria já possuía um divisoma essencialmente completo, associado ao peptideoglicano, mais completo ou redundante que os divisomas de muitas bactérias atuais. A alteração de conteúdo gênico do divisoma foi, na maioria das linhagens, um processo dominado pela perda ou substituição por genes não-homólogos. Os casos mais prováveis de transferência lateral (*ftsZ*, *ftsI* e *amiC*) correspondem a famílias abundantes e funcionalmente diversas, com exemplos de duplicações que provavelmente foram pré-requisito para a invenção dos genes móveis. Somente um desses genes móveis (*mecA*) comprovadamente atua na divisão celular, sugerindo que o fluxo de genes móveis que se associam ao divisoma é relativamente raro. Em nossa discussão sobre o divisoma mínimo de organismos que possuem paredes celulares com camadas de peptideoglicano, inferimos que os principais componentes de um divisoma mínimo seriam um marcador citoplasmático para o sítio de formação do septo (FtsZ) e uma maquinaria de remodelagem do peptideoglicano. O conceito de um divisoma mínimo foi proposto e pode ser usado na nascente área de biologia sistêmica (“systems biology”) como modelo de um divisoma simplificado. Nossos resultados indicam que o aparato de divisão celular ancestral evoluiu após a separação das bactérias da linhagem arqueabactéria-eucarioto, e foi desenvolvido especificamente para lidar com a parede celular bacteriana. Finalmente, a raridade da incorporação ou substituição de componentes

por HGT sugere que o divisoma evolui de forma altamente correlacionada, acumulando mutações que o tornam incompatível para intercâmbio de componentes.

6. Referências

- Aarsman, M. E., A. Piette, *et al.* Maturation of the Escherichia coli divisome occurs in two steps. Mol Microbiol, v.55, n.6, Mar, p.1631-45. 2005.
- Abascal, F., R. Zardoya, *et al.* ProtTest: selection of best-fit models of protein evolution. Bioinformatics, v.21, n.9, p.2104--2105. 2005.
- Abella, M., I. Erill, *et al.* Widespread distribution of a *lexA*-regulated DNA damage-inducible multiple gene cassette in the Proteobacteria phylum. Mol Microbiol, v.54, n.1, Oct, p.212-22. 2004.
- Akiyama, T., S. Inouye, *et al.* Novel developmental genes, *fruCD*, of *Myxococcus xanthus*: involvement of a cell division protein in multicellular development. J Bacteriol, v.185, n.11, Jun, p.3317-24. 2003.
- Anantharaman, V. e L. Aravind. The SHS2 module is a common structural theme in functionally diverse protein groups, like Rpb7p, FtsA, GyrI, and MTH1598/TM1083 superfamilies. Proteins, v.56, n.4, Sep 1, p.795-807. 2004.
- Angert, E. R. Alternatives to binary fission in bacteria. Nat Rev Microbiol, v.3, n.3, Mar, p.214-24. 2005.
- Aravind, L., V. Anantharaman, *et al.* The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev, v.29, n.2, Apr, p.231-62. 2005.
- Aussel, L., F. X. Barre, *et al.* FtsK Is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. Cell, v.108, n.2, Jan 25, p.195-205. 2002.
- Babu, M. M., N. M. Luscombe, *et al.* Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol, v.14, n.3, Jun, p.283-91. 2004.
- Bader, G. D., D. Betel, *et al.* BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res, v.31, n.1, Jan 1, p.248-50. 2003.
- Bateman, A., L. Coin, *et al.* The Pfam protein families database. Nucl. Acids Res., v.32, n.90001, p.D138-141. 2004.
- Bateman, A. e N. D. Rawlings. The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. Trends Biochem Sci, v.28, n.5, May, p.234-7. 2003.
- Bentley, S. D., M. Maiwald, *et al.* Sequencing and analysis of the genome of the Whipple's disease bacterium *Tropheryma whippelii*. Lancet, v.361, n.9358, Feb 22, p.637-44. 2003.
- Bernhardt, T. G. e P. A. De Boer. The Escherichia coli amidase AmiC is a periplasmic

septal ring component exported via the twin-arginine transport pathway. Mol Microbiol, v.48, n.5, Jun, p.1171-82. 2003.

_____. SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over Chromosomes in *E. coli*. Mol Cell, v.18, n.5, May 27, p.555-64. 2005.

Bi, E. F. e J. Lutkenhaus. FtsZ ring structure associated with division in *Escherichia coli*. Nature, v.354, n.6349, Nov 14, p.161-4. 1991.

Bigot, S., J. Corre, *et al.* FtsK activities in Xer recombination, DNA mobilization and cell division involve overlapping and separate domains of the protein. Mol Microbiol, v.54, n.4, Nov, p.876-86. 2004.

Blanc, G., H. Ogata, *et al.* Reductive genome evolution from the mother of Rickettsia. PLoS Genet, v.3, n.1, Jan 19, p.e14. 2007.

Bork, P., C. Sander, *et al.* An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. Proc Natl Acad Sci U S A, v.89, n.16, Aug 15, p.7290-4. 1992.

Bowers, P. M., M. Pellegrini, *et al.* Prolinks: a database of protein functional linkages derived from coevolution. Genome Biol, v.5, n.5, p.R35. 2004.

Brenner, S. E., C. Chothia, *et al.* Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships. Proc Natl Acad Sci U S A, v.95, n.11, May 26, p.6073-8. 1998.

Buddelmeijer, N. e J. Beckwith. A complex of the *Escherichia coli* cell division proteins FtsL, FtsB and FtsQ forms independently of its localization to the septal region. Mol Microbiol, v.52, n.5, Jun, p.1315-27. 2004.

Camon, E., M. Magrane, *et al.* The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. Nucl. Acids Res., v.32, n.90001, p.D262-266. 2004.

Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol Biol Evol, v.17, n.4, p.540--552. 2000.

_____. Topological variation in single-gene phylogenetic trees. Genome Biol, v.8, n.6, Jun 13, p.216. 2007.

Ceulemans, H., L. Beke, *et al.* Approaches to defining the ancestral eukaryotic protein complexome. Bioessays, v.28, n.3, Mar, p.316-24. 2006.

Cheung, H. Y. e L. Vitkovic. Coccus-shaped *Bacillus subtilis* cells are inhibited at stage 0 of sporulation. Can J Microbiol, v.34, n.5, May, p.583-7. 1988.

Ciccarelli, F. D., T. Doerks, *et al.* Toward automatic reconstruction of a highly resolved tree of life. Science, v.311, n.5765, Mar, p.1283--1287. 2006.

Cloud-Hansen, K. A., S. B. Peterson, *et al.* Breaching the great wall: peptidoglycan and microbial interactions. Nat Rev Microbiol, v.4, n.9, Sep, p.710-6. 2006.

Collins, L. e D. Penny. Complex spliceosomal organization ancestral to extant eukaryotes. Mol Biol Evol, v.22, n.4, Apr, p.1053-66. 2005.

Corbin, B. D., Y. Wang, *et al.* Interaction between cell division proteins FtsE and FtsZ. J Bacteriol, v.189, n.8, Apr, p.3026-35. 2007.

Cordell, S. C., R. E. Anderson, *et al.* Crystal structure of the bacterial cell division inhibitor MinC. Embo J, v.20, n.10, May 15, p.2454-61. 2001.

Cordell, S. C., E. J. Robinson, *et al.* Crystal structure of the SOS cell division inhibitor SulA and in complex with FtsZ. Proc Natl Acad Sci U S A, v.100, n.13, Jun 24, p.7889-94. 2003.

Cunningham, C. W. Some Limitations of Ancestral Character-State Reconstruction When Testing Evolutionary Hypotheses. Systematic Biology, v.48, n.3, p.665-674. 1999.

Cunningham, C. W., K. E. Omland, *et al.* Reconstructing ancestral character states: a critical reappraisal. Trends in Ecology & Evolution, v.13, n.9, p.361-366. 1998.

Da Silva, A. C., J. A. Ferro, *et al.* Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature, v.417, n.6887, p.459-63. 2002.

Dandekar, T., B. Snel, *et al.* Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci, v.23, n.9, Sep, p.324-8. 1998.

Date, S. V. e E. M. Marcotte. Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. Nature Biotechnology, v.21, n.9, Sep, p.1055-1062. 2003.

_____. Protein function prediction using the Protein Link EXplorer (PLEX). Bioinformatics, v.21, n.10, May 15, p.2558-2559. 2005.

Datta, P., A. Dasgupta, *et al.* Interaction between FtsZ and FtsW of *Mycobacterium tuberculosis*. J Biol Chem, v.277, n.28, Jul 12, p.24983-7. 2002.

De Bie, T., N. Cristianini, *et al.* CAFE: a computational tool for the study of gene family evolution. Bioinformatics, v.22, n.10, May 15, p.1269-71. 2006.

De Leeuw, E., B. Graham, *et al.* Molecular characterization of *Escherichia coli* FtsE and FtsX. Mol Microbiol, v.31, n.3, Feb, p.983-93. 1999.

Deppenmeier, U., A. Johann, *et al.* The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J Mol Microbiol Biotechnol, v.4, n.4, Jul, p.453-61. 2002.

Doerks, T., C. Von Mering, *et al.* Functional clues for hypothetical proteins based on genomic context analysis in prokaryotes. Nucl. Acids Res., v.32, n.21, p.6321-6326. 2004.

Doolittle, R. F. Biodiversity: microbial genomes multiply. Nature, v.416, n.6882, Apr 18, p.697-700. 2002.

Draper, G. C., N. McLennan, *et al.* Only the N-terminal domain of FtsK functions in cell division. J Bacteriol, v.180, n.17, Sep, p.4621-7. 1998.

Eddy, S. R. HMMER User's Guide: Biological sequence analysis using profile Hidden Markov Models, version 2.1.1: Washington University School of Medicine 1998.

Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics, v.5, Aug, p.113. 2004.

Eisen, J. A. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res, v.8, n.3, Mar, p.163-7. 1998.

_____. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. Curr Opin Genet Dev, v.10, n.6, Dec, p.606-11. 2000.

Eisen, J. A., J. F. Heidelberg, *et al.* Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biology, v.1, n.6. 2000.

Enright, A. J., I. Iliopoulos, *et al.* Protein interaction maps for complete genomes based on gene fusion events. Nature, v.402, n.6757, Nov 4, p.86-90. 1999.

Enright, M. C., D. A. Robinson, *et al.* The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc Natl Acad Sci U S A, v.99, n.11, May 28, p.7687-92. 2002.

Erickson, H. P. Evolution of the cytoskeleton. Bioessays, v.29, n.7, Jul, p.668-77. 2007.

Errington, J., R. A. Daniel, *et al.* Cytokinesis in bacteria. Microbiol Mol Biol Rev, v.67, n.1, Mar, p.52-65, table of contents. 2003.

Faguy, D. M. e W. F. Doolittle. Cytoskeletal proteins: the evolution of cell division. Curr Biol, v.8, n.10, May 7, p.R338-41. 1998.

Fariselli, P., I. Rossi, *et al.* The WWWH of remote homolog detection: the state of the art. Brief Bioinform, v.8, n.2, Mar, p.78-87. 2007.

Flardh, K. Essential role of DivIVA in polar growth and morphogenesis in *Streptomyces coelicolor* A3(2). Mol Microbiol, v.49, n.6, Sep, p.1523-36. 2003.

Foster, P. L. Stress responses and genetic variation in bacteria. Mutat Res, v.569, n.1-2, Jan 6, p.3-11. 2005.

Fraser, C. M., J. D. Gocayne, *et al.* The minimal gene complement of *Mycoplasma*

genitalium. Science, v.270, n.5235, Oct 20, p.397-403. 1995.

Gabaldon, T., D. Rainey, *et al.* Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). J Mol Biol, v.348, n.4, May 13, p.857-70. 2005.

Gabaldon, T., B. Snel, *et al.* Origin and evolution of the peroxisomal proteome. Biol Direct, v.1, p.8. 2006.

Galhardo, R. S., R. P. Rocha, *et al.* An SOS-regulated operon involved in damage-inducible mutagenesis in *Caulobacter crescentus*. Nucleic Acids Res, v.33, n.8, p.2603-14. 2005.

Galperin, M. Y., D. R. Walker, *et al.* Analogous enzymes: independent inventions in enzyme evolution. Genome Res, v.8, n.8, Aug, p.779-90. 1998.

Geissler, B., D. Elraheb, *et al.* A gain-of-function mutation in *ftsA* bypasses the requirement for the essential cell division gene *zipA* in *Escherichia coli*. Proc Natl Acad Sci U S A, v.100, n.7, Apr 1, p.4197-202. 2003.

Goehring, N. W. e J. Beckwith. Diverse paths to midcell: assembly of the bacterial cell division machinery. Current Biology, v.15, p.514--526. 2005.

Gogarten, J. P. e J. P. Townsend. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol, v.3, n.9, Sep, p.679-87. 2005.

Graumann, P. Bacillus : cellular and molecular biology. Norfolk: Caister Academic Press. 2007. xv, 454 p. p.

Gueiros-Filho, F. Cell Division. In: P. Graumann (Ed.). Bacillus : cellular and molecular biology. Wymondham: Caister Academic Press, 2007. Cell Division

Gueiros-Filho, F. J. e R. Losick. A widely conserved bacterial cell division protein that promotes assembly of the tubulin-like protein FtsZ. Genes Dev, v.16, n.19, Oct 1, p.2544-56. 2002.

Guertin, D. A., S. Trautmann, *et al.* Cytokinesis in eukaryotes. Microbiol Mol Biol Rev, v.66, n.2, Jun, p.155-78. 2002.

Guindon, S. P. e O. Gascuel. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol, v.52, n.5, Oct, p.696--704. 2003.

Haeusser, D. P., R. L. Schwartz, *et al.* EzrA prevents aberrant cell division by modulating assembly of the cytoskeletal protein FtsZ. Mol Microbiol, v.52, n.3, May, p.801-14. 2004.

Hahn, M. W., T. De Bie, *et al.* Estimating the tempo and mode of gene family evolution from comparative genomic data. Genome Res., v.15, n.8, p.1153-1160. 2005.

Hamoen, L. W., J. C. Meile, *et al.* SepF, a novel FtsZ-interacting protein required for a late step in cell division. Mol Microbiol, v.59, n.3, Feb, p.989-99. 2006.

Hanssen, A. M. e J. U. E. Sollid. SCCmec in staphylococci: genes on the move. Fems Immunology and Medical Microbiology, v.46, n.1, Feb, p.8-20. 2006.

Harvey, P. H. e M. D. Pagel. The comparative method in evolutionary biology. Oxford ; New York: Oxford University Press. 1991. viii, 239 p. p. (Oxford series in ecology and evolution)

Heidrich, C., M. F. Templin, *et al.* Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. Mol Microbiol, v.41, n.1, Jul, p.167-78. 2001.

Hillen, W. e C. Berens. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. Annu Rev Microbiol, v.48, p.345-69. 1994.

Hillis, D. M., C. Moritz, *et al.* Molecular systematics. Sunderland, Mass.: Sinauer Associates. 1996. xvi, 655 p. p.

Horn, C., B. Paulmann, *et al.* In vivo observation of cell division of anaerobic hyperthermophiles by using a high-intensity dark-field microscope. J Bacteriol, v.181, n.16, Aug, p.5114-8. 1999.

Hu, Z. e J. Lutkenhaus. Analysis of MinC reveals two independent domains involved in interaction with MinD and FtsZ. J Bacteriol, v.182, n.14, Jul, p.3965-71. 2000.

_____. A conserved sequence at the C-terminus of MinD is required for binding to the membrane and targeting MinC to the septum. Mol Microbiol, v.47, n.2, Jan, p.345-55. 2003.

Huynen, M., B. Snel, *et al.* Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res, v.10, n.8, Aug, p.1204-10. 2000.

Huynen, M. A., T. Gabaldon, *et al.* Variation and evolution of biomolecular systems: searching for functional relevance. FEBS Lett, v.579, n.8, Mar 21, p.1839-45. 2005.

Ishikawa, S., Y. Kawai, *et al.* A new FtsZ-interacting protein, YlmF, complements the activity of FtsA during progression of cell division in *Bacillus subtilis*. Mol Microbiol, v.60, n.6, Jun, p.1364-80. 2006.

John, B. e A. Sali. Detection of homologous proteins by an intermediate sequence search. Protein Sci, v.13, n.1, Jan, p.54-62. 2004.

Kanehisa, M., S. Goto, *et al.* The KEGG resource for deciphering the genome. Nucl. Acids. Res., v.32, n.90001, p.D277-280. 2004.

Karev, G. P., Y. I. Wolf, *et al.* Birth and death of protein domains: A simple model of evolution explains power law behavior. BMC Evolutionary Biology, v.2, n.18. 2002.

Karp, P. D., M. Riley, *et al.* The MetaCyc Database. Nucleic Acids Res, v.30, n.1, Jan 1,

p.59-61. 2002.

King, G. F., S. L. Rowland, *et al.* The dimerization and topological specificity functions of MinE reside in a structurally autonomous C-terminal domain. Mol Microbiol, v.31, n.4, Feb, p.1161-9. 1999.

Koonin, E. V. A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. J Mol Biol, v.229, n.4, Feb 20, p.1165-74. 1993.

Kruse, T., J. Bork-Jensen, *et al.* The morphogenetic MreBCD proteins of Escherichia coli form an essential membrane-bound complex. Mol Microbiol, v.55, n.1, Jan, p.78-89. 2005.

Kumar, S. e M. Nei. Molecular Evolution and Phylogenetics. New York: Oxford University Press. 2000

Kunin, V. e C. A. Ouzounis. The Balance of Driving Forces During Genome Evolution in Prokaryotes. Genome Research, n.13, p.1589--1594. 2003.

Kuroda, A., Y. Sugimoto, *et al.* Genetic structure, isolation and characterization of a Bacillus licheniformis cell wall hydrolase. Mol Gen Genet, v.234, n.1, Jul, p.129-37. 1992.

Lakshminarayan, M. I., K. S. Makarova, *et al.* Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res, v.32, n.17, p.5260--5279. 2004.

Latruffe, N. e J. Vamecq. Evolutionary aspects of peroxisomes as cell organelles, and of genes encoding peroxisomal proteins. Biol Cell, v.92, n.6, Sep, p.389-95. 2000.

Lawrence, J. G. e H. Ochman. Molecular archaeology of the Escherichia coli genome. Proc. Nat. Acad. Sci., v.95, p.9413--9417. 1998.

Lester, L., A. Meade, *et al.* The slow road to the eukaryotic genome. Bioessays, v.28, n.1, Jan, p.57--64. 2006.

Levin, P. A., I. G. Kurtser, *et al.* Identification and characterization of a negative regulator of FtsZ ring formation in Bacillus subtilis. Proc Natl Acad Sci U S A, v.96, n.17, Aug 17, p.9642-7. 1999.

Lim, D. e N. C. Strynadka. Structural basis for the beta lactam resistance of PBP2a from methicillin-resistant Staphylococcus aureus. Nat Struct Biol, v.9, n.11, Nov, p.870-6. 2002.

Lio, P. e M. Vannucci. Finding pathogenicity islands and gene transfer events in genome data. Bioinformatics, v.16, n.10, Oct, p.932-40. 2000.

Llull, D., R. Lopez, *et al.* Skl, a novel choline-binding N-acetylmuramoyl-L-alanine amidase of Streptococcus mitis SK137 containing a CHAP domain. FEBS Lett, v.580, n.8, Apr 3, p.1959-64. 2006.

- Loessner, M. J., S. K. Maier, *et al.* Three *Bacillus cereus* bacteriophage endolysins are unrelated but reveal high homology to cell wall hydrolases from different bacilli. J Bacteriol, v.179, n.9, May, p.2845-51. 1997.
- Long, M., E. Betran, *et al.* The origin of new genes: glimpses from the young and old. Nat Rev Genet, v.4, n.11, Nov, p.865-75. 2003.
- Low, H. H., M. C. Moncrieffe, *et al.* The crystal structure of ZapA and its modulation of FtsZ polymerisation. J Mol Biol, v.341, n.3, Aug 13, p.839-52. 2004.
- Lowe, J. e L. A. Amos. Crystal structure of the bacterial cell-division protein FtsZ. Nature, v.391, n.6663, Jan 8, p.203-6. 1998.
- Lowe, J. e F. Van Den Ent. Conserved sequence motif at the C-terminus of the bacterial cell-division protein FtsA. Biochimie, v.83, n.1, Jan, p.117-20. 2001.
- Lutkenhaus, J. Assembly dynamics of the bacterial MinCDE system and spatial regulation of the Z ring. Annu Rev Biochem, v.76, p.539-62. 2007.
- Lynch, M. e J. S. Conery. The evolutionary fate and consequences of duplicate genes. Science, v.290, p.1151--1155. 2000.
- Ma, X., Q. Sun, *et al.* Interactions between heterologous FtsA and FtsZ proteins at the FtsZ ring. J. Bacteriol., v.179, n.21, p.6788-6797. 1997.
- Marcotte, E. M. Computational genetics: finding protein function by nonhomology methods. Curr Opin Struct Biol, v.10, n.3, Jun, p.359-65. 2000.
- Marcotte, E. M., M. Pellegrini, *et al.* Detecting protein function and protein-protein interactions from genome sequences. Science, v.285, n.5428, Jul 30, p.751-3. 1999.
- Margolin, W. Themes and variations in prokaryotic cell division. FEMS Microbiology Reviews, v.24, n.4, October, p.531--548. 2000.
- _____. FtsZ and the division of prokaryotic cells and organelles. Nat Rev Mol Cell Biol, v.6, n.11, Nov, p.862-71. 2005.
- Mccoy, A. J. e A. T. Maurelli. Building the invisible wall: updating the chlamydial peptidoglycan anomaly. Trends Microbiol, v.14, n.2, Feb, p.70-7. 2006.
- Mellor, J. C., I. Yanai, *et al.* Predictome: a database of putative functional links between proteins. Nucleic Acids Res, v.30, n.1, Jan 1, p.306-9. 2002.
- Michie, K. A. e J. Lowe. Dynamic filaments of the bacterial cytoskeleton. Annu Rev Biochem, v.75, p.467-92. 2006.
- Miller, R. G. The Jackknife--A Review. Biometrika, v.61, n.1, p.1-15. 1974.
- Mingorance, J., J. Tamames, *et al.* Genomic channeling in bacterial cell division. J Mol

- Recognit, v.17, n.5, Sep-Oct, p.481-7. 2004.
- Mira, A., H. Ochman, *et al.* Deletional bias and the evolution of bacterial genomes. Trends Genet, v.17, n.10, Oct, p.589-96. 2001.
- Mira, A., R. Pushker, *et al.* Evolutionary relationships of *Fusobacterium nucleatum* based on phylogenetic analysis and comparative genomics. BMC Evol Biol, v.4, Nov 26, p.50. 2004.
- Mirkin, B. G., T. I. Fenner, *et al.* Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evolutionary Biology, v.3, n.2. 2003.
- Mishima, M., T. Shida, *et al.* Solution structure of the peptidoglycan binding domain of *Bacillus subtilis* cell wall lytic enzyme CwlC: characterization of the sporulation-related repeats by NMR. Biochemistry, v.44, n.30, Aug 2, p.10153-63. 2005.
- Miyagishima, S. Y., C. P. Wolk, *et al.* Identification of cyanobacterial cell division genes by comparative and mutational analyses. Mol Microbiol, v.56, n.1, Apr, p.126-43. 2005.
- Moncalian, G., E. Cabezón, *et al.* Characterization of ATP and DNA binding activities of TrwB, the coupling protein essential in plasmid R388 conjugation. J Biol Chem, v.274, n.51, Dec 17, p.36117-24. 1999.
- Moran, N. A. Microbial minimalism: genome reduction in bacterial pathogens. Cell, v.108, n.5, Mar 8, p.583-6. 2002.
- Moran, N. A. e P. H. Degnan. Functional genomics of *Buchnera* and the ecology of aphid hosts. Mol Ecol, v.15, n.5, Apr, p.1251-61. 2006.
- Moran, N. A., G. R. Plague, *et al.* A genomic perspective on nutrient provisioning by bacterial symbionts of insects. Proc Natl Acad Sci U S A, v.100 Suppl 2, Nov 25, p.14543-8. 2003.
- Moreira, L. M., R. F. De Souza, *et al.* Comparative analyses of *Xanthomonas* and *Xylella* complete genomes. OMICS, v.9, n.1, p.43-76. 2005.
- Moy, F. J., E. Glasfeld, *et al.* Solution structure of ZipA, a crucial component of *Escherichia coli* cell division. Biochemistry, v.39, n.31, Aug 8, p.9146-56. 2000.
- Mushegian, A. Protein content of minimal and ancestral ribosome. RNA, v.11, n.9, Sep, p.1400--1406. 2005.
- Myers, C. L., D. Robson, *et al.* Discovery of biological networks from diverse functional genomic data. Genome Biol, v.6, n.13, p.R114. 2005.
- Nambu, T., T. Minamino, *et al.* Peptidoglycan-hydrolyzing activity of the FlgJ protein, essential for flagellar rod formation in *Salmonella typhimurium*. J Bacteriol, v.181, n.5, Mar, p.1555-61. 1999.
- Nelson, K. E., R. A. Clayton, *et al.* Evidence for lateral gene transfer between Archaea and

bacteria from genome sequence of *Thermotoga maritima*. Nature, v.399, n.6734, May 27, p.323-9. 1999.

Nogales, E., K. H. Downing, *et al.* Tubulin and FtsZ form a distinct family of GTPases. Nat Struct Biol, v.5, n.6, Jun, p.451-8. 1998.

O'donovan, C., M. J. Martin, *et al.* High-quality protein knowledge resource: SWISS-PROT and TrEMBL. Brief Bioinform, v.3, n.3, p.275--284. 2002.

Oakley, T. H. e C. W. Cunningham. Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. Evolution, v.54, n.2, Apr, p.397-405. 2000.

_____. Molecular phylogenetic evidence for the independent evolutionary origin of an arthropod compound eye. Proc Natl Acad Sci U S A, v.99, n.3, Feb 5, p.1426-30. 2002.

Ochman, H., J. G. Lawrence, *et al.* Lateral gene transfer and the nature of bacterial innovation. Nature, v.405, p.299--304. 2000.

Ohnishi, R., S. Ishikawa, *et al.* Peptidoglycan hydrolase LytF plays a role in cell separation with CwlF during vegetative growth of *Bacillus subtilis*. J Bacteriol, v.181, n.10, May, p.3178-84. 1999.

Osawa, M. e H. P. Erickson. FtsZ from divergent foreign bacteria can function for cell division in *Escherichia coli*. J Bacteriol, v.188, n.20, Oct, p.7132-40. 2006.

Ouzounis, C. A. Ancestral state reconstructions for genomes. Curr Opin Genet Dev, v.15, n.6, Dec, p.595-600. 2005.

Pagel, M. Detecting Correlated Evolution on Phylogenies - a General-Method for the Comparative-Analysis of Discrete Characters. Proceedings of the Royal Society of London Series B-Biological Sciences, v.255, n.1342, Jan 22, p.37-45. 1994.

_____. Inferring the historical patterns of biological evolution. Nature, v.401, n.6756, Oct, p.877--884. 1999a.

_____. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Systematic Biology, v.48, n.3, Sep, p.612-622. 1999b.

Pagel, M., A. Meade, *et al.* Bayesian estimation of ancestral character states on phylogenies. Syst Biol, v.53, n.5, Oct, p.673--684. 2004.

Pal, C., B. Papp, *et al.* Chance and necessity in the evolution of minimal metabolic networks. Nature, v.440, n.7084, Mar 30, p.667-70. 2006.

Pallen, M. J. The ESAT-6/WXG100 superfamily -- and a new Gram-positive secretion system? Trends Microbiol, v.10, n.5, May, p.209-12. 2002.

Perna, N. T., G. Plunkett, *et al.* Genome sequence of enterohaemorrhagic *Escherichia coli*

O157:H7. Nature, v.409, n.6819, p.529--533. 2001.

Peterson, J. D., L. A. Umayam, *et al.* The Comprehensive Microbial Resource. Nucleic Acids Res, v.29, n.1, Jan 1, p.123-5. 2001.

Pichoff, S. e J. Lutkenhaus. Unique and overlapping roles for ZipA and FtsA in septal ring assembly in *Escherichia coli*. Embo J, v.21, n.4, Feb 15, p.685-93. 2002.

_____. Tethering the Z ring to the membrane through a conserved membrane targeting sequence in FtsA. Mol Microbiol, v.55, n.6, Mar, p.1722-34. 2005.

Pilhofer, M., G. Rosati, *et al.* Coexistence of Tubulins and ftsZ in Different *Prostheco bacter* Species. Mol Biol Evol, Apr 4. 2007.

Pinho, M. G., S. R. Filipe, *et al.* Complementation of the essential peptidoglycan transpeptidase function of penicillin-binding protein 2 (PBP2) by the drug resistance protein PBP2A in *Staphylococcus aureus*. J Bacteriol, v.183, n.22, Nov, p.6525-31. 2001.

Real, G., S. Autret, *et al.* Cell division protein DivIB influences the Spo0J/Soj system of chromosome segregation in *Bacillus subtilis*. Mol Microbiol, v.55, n.2, Jan, p.349-67. 2005.

Rico, A. I., M. Garcia-Ovalle, *et al.* Role of two essential domains of *Escherichia coli* FtsA in localization and progression of the division ring. Mol Microbiol, v.53, n.5, Sep, p.1359-71. 2004.

Robson, S. A. e G. F. King. Domain architecture and structure of the bacterial cell division protein DivIB. Proc Natl Acad Sci U S A, v.103, n.17, Apr 25, p.6700-5. 2006.

Ronquist, F. Bayesian inference of character evolution. Trends in Ecology and Evolution, v.19, n.9, 9, p.475--481. 2004.

Rothfield, L., S. Justice, *et al.* Bacterial cell division. Annu Rev Genet, v.33, p.423-48. 1999.

Rothfield, L., A. Taghbalout, *et al.* Spatial control of bacterial division-site placement. Nat Rev Microbiol, v.3, n.12, Dec, p.959--968. 2005.

Sanchez-Pulido, L., D. Devos, *et al.* POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins. Trends Biochem Sci, v.28, n.10, Oct, p.523-6. 2003.

Scheffers, D.-J. The Cell Wall of *Bacillus subtilis*. In: P. Graumann (Ed.). Bacillus : cellular and molecular biology. Wymondham: Caister Academic Press, 2007. The Cell Wall of *Bacillus subtilis*

Scheffers, D. J. e J. Errington. PBP1 is a component of the *Bacillus subtilis* cell division machinery. J Bacteriol, v.186, n.15, Aug, p.5153-6. 2004.

Scheffers, D. J., L. J. Jones, *et al.* Several distinct localization patterns for penicillin-

- binding proteins in *Bacillus subtilis*. Mol Microbiol, v.51, n.3, Feb, p.749-64. 2004.
- Schmidt, K. L., N. D. Peterson, *et al.* A predicted ABC transporter, FtsEX, is needed for cell division in *Escherichia coli*. J Bacteriol, v.186, n.3, Feb, p.785-93. 2004.
- Shiomi, D. e W. Margolin. The C-terminal domain of MinC inhibits assembly of the Z ring in *Escherichia coli*. J Bacteriol, v.189, n.1, Jan, p.236-43. 2007.
- Smith, T. J., S. A. Blackman, *et al.* Autolysins of *Bacillus subtilis*: multiple enzymes with multiple functions. Microbiology, v.146 (Pt 2), Feb, p.249-62. 2000.
- Snel, B., P. Bork, *et al.* Genomes in Flux: The Evolution of Archaeal and Proteobacterial Gene Content. Genome Research, v.12, n.1, p.17--25. 2002.
- Soding, J., M. Remmert, *et al.* HHsenser: exhaustive transitive profile search using HMM-HMM comparison. Nucleic Acids Res, v.34, n.Web Server issue, Jul 1, p.W374-8. 2006.
- Stahlberg, H., E. Kutejova, *et al.* Oligomeric structure of the *Bacillus subtilis* cell division protein DivIVA determined by transmission electron microscopy. Mol Microbiol, v.52, n.5, Jun, p.1281-90. 2004.
- Stajich, J. E., D. Block, *et al.* The Bioperl Toolkit: Perl Modules for the Life Sciences. Genome Res., v.12, n.10, p.1611--1618. 2002.
- Stark, C., B. J. Breitkreutz, *et al.* BioGRID: a general repository for interaction datasets. Nucleic Acids Res, v.34, n.Database issue, Jan 1, p.D535-9. 2006.
- Tamames, J. Evolution of gene order conservation in prokaryotes. Genome Biology, v.2, n.6. 2001.
- Tamames, J., M. González-Moreno, *et al.* Bringing gene order into bacterial shape. Trends in Genetics, v.17, n.3, 3, p.124--126. 2001.
- Tatusov, R. L., N. D. Fedorova, *et al.* The COG database: an updated version includes eukaryotes. BMC Bioinformatics, v.4, Sep 11, p.41. 2003.
- Teichmann, S. A. e M. M. Babu. Gene regulatory network growth by duplication. Nat Genet, v.36, n.5, May, p.492-6. 2004.
- Thanbichler, M. e L. Shapiro. MipZ, a spatial regulator coordinating chromosome segregation with cell division in *Caulobacter*. Cell, v.126, n.1, Jul 14, p.147-62. 2006.
- Thomaides, H. B., M. Freeman, *et al.* Division site selection protein DivIVA of *Bacillus subtilis* has a second distinct function in chromosome segregation during sporulation. Genes Dev, v.15, n.13, Jul 1, p.1662-73. 2001.
- Thompson, J. D., D. G. Higgins, *et al.* CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, v.22, n.22, p.4673--4680.

1994.

Tinsley, E. e S. A. Khan. A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in *Bacillus anthracis*. J Bacteriol, v.188, n.8, Apr, p.2829-35. 2006.

Ueda, K., A. Yamashita, *et al.* Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. Nucleic Acids Res, v.32, n.16, p.4937--4944. 2004.

Vaughan, S., B. Wickstead, *et al.* Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryota. J Mol Evol, v.58, n.1, Jan, p.19--29. 2004.

Vicente, M. e A. I. Rico. The order of the ring: assembly of *Escherichia coli* cell division components. Mol Microbiol, v.61, n.1, Jul, p.5--8. 2006.

Vicente, M., A. I. Rico, *et al.* Septum enlightenment: assembly of bacterial division proteins. J Bacteriol, v.188, n.1, Jan, p.19--27. 2006.

Vicente, M., J. Tamames, *et al.* Molecules in Time and Space: Bacterial Shape, Division and Phylogeny. New York: Kluwer Academic / Plenum Publishers, v.1. 2004. 275 p.

Von Mering, C., L. J. Jensen, *et al.* STRING 7--recent developments in the integration and prediction of protein interactions. Nucleic Acids Res, v.35, n.Database issue, Jan, p.D358-62. 2007.

_____. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Research, v.33, n.1S, p.D433-437. 2005.

Wintjens, R. e M. Rooman. Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. J Mol Biol, v.262, n.2, Sep 20, p.294-313. 1996.

Woldringh, C. L., E. Mulder, *et al.* Toporegulation of bacterial division according to the nucleoid occlusion model. Res Microbiol, v.142, n.2-3, Feb-Apr, p.309-20. 1991.

Wolf, Y. I., I. B. Rogozin, *et al.* Genome trees constructed using five different approaches suggest new major bacterial clades. BMC Evolutionary Biology, v.1, n.8. 2001.

_____. Genome alignment, evolution of prokaryotic genome organization and prediction of gene function using genomic context. Genome Research, v.11, p.356--372. 2001.

Wu, C. F. J. Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. The Annals of Statistics, v.14, n.4, p.1261-1295. 1986.

Wu, L. J. e J. Errington. Coordination of cell division and chromosome segregation by a nucleoid occlusion protein in *Bacillus subtilis*. Cell, v.117, n.7, Jun 25, p.915-25. 2004.

Yamaichi, Y. e H. Niki. Active segregation by the *Bacillus subtilis* partitioning system in

Escherichia coli. Proc Natl Acad Sci U S A, v.97, n.26, Dec 19, p.14656-61. 2000.

Yang, J. C., F. Van Den Ent, *et al.* Solution structure and domain architecture of the divisome protein FtsN. Mol Microbiol, v.52, n.3, May, p.651-60. 2004.

Yu, X. C., A. H. Tran, *et al.* Localization of cell division protein FtsK to the Escherichia coli septum and identification of a potential N-terminal targeting domain. J Bacteriol, v.180, n.5, Mar, p.1296-304. 1998.

Zhao, Y., R. W. Hammond, *et al.* Cell division gene cluster in Spiroplasma kunkelii: functional characterization of ftsZ and the first report of ftsA in mollicutes. DNA Cell Biol, v.23, n.2, Feb, p.127-34. 2004.

Zhaxybayeva, O., C. L. Nesbo, *et al.* Systematic overestimation of gene gain through false diagnosis of gene absence. Genome Biol, v.8, n.2, p.402. 2007.

7. Anexo

Tabela A1: Nomes, códigos e informações taxonômicas das espécies analisadas.

código	Nome científico	Reino	Filo	Classe	Ordem	Família
AERPE	<i>Aeropyrum pernix</i>	Archaea	Crenarchaeota	Thermoprotei	Desulfurococcales	Desulfurococcaceae
AGRTc	<i>Agrobacterium tumefaciens</i> C58 Cereon	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
AGRTw	<i>Agrobacterium tumefaciens</i> C58 Uwash	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
AQUAE	<i>Aquifex aeolicus</i>	Bacteria	Aquificae		Aquificales	Aquificaceae
ARCFU	<i>Archaeoglobus fulgidus</i>	Archaea	Euryarchaeota	Archaeoglobi	Archaeoglobales	Archaeoglobaceae
BACAN	<i>Bacillus anthracis</i> Ames	Bacteria	Firmicutes		Bacillales	Bacillaceae
BACC1	<i>Bacillus cereus</i> ATCC 10987	Bacteria	Firmicutes		Bacillales	Bacillaceae
BACCR	<i>Bacillus cereus</i> ATCC14579	Bacteria	Firmicutes		Bacillales	Bacillaceae
BACHD	<i>Bacillus halodurans</i>	Bacteria	Firmicutes		Bacillales	Bacillaceae
BACSU	<i>Bacillus subtilis</i>	Bacteria	Firmicutes		Bacillales	Bacillaceae
BACTN	<i>Bacteroides thetaiotaomicron</i> VPI-5482	Bacteria	Bacteroidetes	Bacteroidetes	Bacteroidales	Bacteroidaceae
BDEBA	<i>Bdellovibrio bacteriovorus</i>	Bacteria	Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bdellovibrionaceae
BIFLO	<i>Bifidobacterium longum</i>	Bacteria	Actinobacteria	Actinobacteridae	Bifidobacteriales	Bifidobacteriaceae
BORBR	<i>Bordetella bronchiseptica</i>	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae
BORPA	<i>Bordetella parapertussis</i>	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae
BORPE	<i>Bordetella pertussis</i>	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae
BORBU	<i>Borrelia burgdorferi</i>	Bacteria	Spirochaetes		Spirochaetales	Spirochaetaceae
BRAJA	<i>Bradyrhizobium japonicum</i>	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae
BRUME	<i>Brucella melitensis</i>	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Brucellaceae
BRUSU	<i>Brucella suis</i> 1330	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Brucellaceae
BUCBP	<i>Buchnera aphidicola</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
BUCAP	<i>Buchnera aphidicola</i> Sg	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
BUCAI	<i>Buchnera</i> sp	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
CAMJE	<i>Campylobacter jejuni</i>	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae
CANBF	Candidatus <i>Blochmannia floridanus</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
CAUCR	<i>Caulobacter crescentus</i>	Bacteria	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae
CHLMU	<i>Chlamydia muridarum</i>	Bacteria	Chlamydiae		Chlamydiales	Chlamydiaceae

CHLTR	<i>Chlamydia trachomatis</i>	Bacteria	Chlamydiae		Chlamydiales	Chlamydiaceae
CHLCV	<i>Chlamydophila caviae</i>	Bacteria	Chlamydiae		Chlamydiales	Chlamydiaceae
CHLPa	<i>Chlamydophila pneumoniae</i> AR39	Bacteria	Chlamydiae		Chlamydiales	Chlamydiaceae
CHLPc	<i>Chlamydophila pneumoniae</i> CWL029	Bacteria	Chlamydiae		Chlamydiales	Chlamydiaceae
CHLPj	<i>Chlamydophila pneumoniae</i> J138	Bacteria	Chlamydiae		Chlamydiales	Chlamydiaceae
CHLPt	<i>Chlamydophila pneumoniae</i> TW 183	Bacteria	Chlamydiae		Chlamydiales	Chlamydiaceae
CHLTE	<i>Chlorobium tepidum</i> TLS	Bacteria	Chlorobi	Chlorobia	Chlorobiales	Chlorobiaceae
CHRVO	<i>Chromobacterium violaceum</i>	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae
CLOAB	<i>Clostridium acetobutylicum</i>	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae
CLOPE	<i>Clostridium perfringens</i>	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae
CLOTE	<i>Clostridium tetani</i> E88	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae
CORDI	<i>Corynebacterium diphtheriae</i>	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae
COREF	<i>Corynebacterium efficiens</i> YS-314	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae
CORGb	<i>Corynebacterium glutamicum</i> ATCC 13032 Bielefeld	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae
CORGk	<i>Corynebacterium glutamicum</i> ATCC 13032 Kitasato	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae
COXBU	<i>Coxiella burnetii</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Legionellales	Coxiellaceae
DEIRA	<i>Deinococcus radiodurans</i>	Bacteria	Deinococcus-Thermus	Deinococci	Deinococcales	Deinococcaceae
DESVH	<i>Desulfovibrio vulgaris</i> Hildenborough	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfiovibrionales	Desulfovibrionaceae
ENTFA	<i>Enterococcus faecalis</i> V583	Bacteria	Firmicutes		Lactobacillales	Enterococcaceae
ECOL6	<i>Escherichia coli</i> CFT073	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
ECOLk	<i>Escherichia coli</i> K12	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
ECO57	<i>Escherichia coli</i> O157H7	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
ECOLe	<i>Escherichia coli</i> O157H7 EDL933	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
FUSNN	<i>Fusobacterium nucleatum</i>	Bacteria	Fusobacteria		Fusobacteriales	Fusobacteriaceae
GEOSL	<i>Geobacter sulfurreducens</i>	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	Geobacteraceae
GLOVI	<i>Gloeobacter violaceus</i>	Bacteria	Cyanobacteria	Gloeobacteria	Gloeobacteriales	
HAEDU	<i>Haemophilus ducreyi</i> 35000HP	Bacteria	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
HAEIN	<i>Haemophilus influenzae</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
HALSA	<i>Halobacterium</i> sp	Archaea	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae
HELHP	<i>Helicobacter hepaticus</i>	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacteriales	Helicobacteraceae
HELpy	<i>Helicobacter pylori</i> 26695	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacteriales	Helicobacteraceae
HELpJ	<i>Helicobacter pylori</i> J99	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacteriales	Helicobacteraceae
LACJO	<i>Lactobacillus johnsonii</i> NCC 533	Bacteria	Firmicutes		Lactobacillales	Lactobacillaceae
LACPL	<i>Lactobacillus plantarum</i>	Bacteria	Firmicutes		Lactobacillales	Lactobacillaceae
LACLA	<i>Lactococcus lactis</i>	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
LEPIC	<i>Leptospira interrogans</i> serovar Copenhageni	Bacteria	Spirochaetes		Spirochaetales	Leptospiraceae

LEPIN	<i>Leptospira interrogans</i> serovar Lai	Bacteria	Spirochaetes		Spirochaetales	Leptosiraceae
LISIN	<i>Listeria innocua</i>	Bacteria	Firmicutes		Bacillales	Listeriaceae
LISMO	<i>Listeria monocytogenes</i>	Bacteria	Firmicutes		Bacillales	Listeriaceae
LISMF	<i>Listeria monocytogenes</i> 4b F2365	Bacteria	Firmicutes		Bacillales	Listeriaceae
RHILO	<i>Mesorhizobium loti</i>	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae
METTH	<i>Methanobacterium thermoautotrophicum</i>	Archaea	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae
METJA	<i>Methanococcus jannaschii</i>	Archaea	Euryarchaeota	Methanococci	Methanococcales	Methanocaldococcaceae
METMP	<i>Methanococcus maripaludis</i> S2	Archaea	Euryarchaeota	Methanococci	Methanococcales	Methanococcaceae
METKA	<i>Methanopyrus kandleri</i>	Archaea	Euryarchaeota	Methanopyri	Methanopyrales	Methanopyraceae
METAC	<i>Methanosarcina acetivorans</i>	Archaea	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae
METMA	<i>Methanosarcina mazei</i>	Archaea	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae
MYCPA	<i>Mycobacterium avium paratuberculosis</i>	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae
MYCBO	<i>Mycobacterium bovis</i>	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae
MYCLE	<i>Mycobacterium leprae</i>	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae
MYCTc	<i>Mycobacterium tuberculosis</i> CDC1551	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae
MYCTh	<i>Mycobacterium tuberculosis</i> H37Rv	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae
MYCGA	<i>Mycoplasma gallisepticum</i>	Bacteria	Firmicutes	Mollicutes		Mycoplasmataceae
MYCGE	<i>Mycoplasma genitalium</i>	Bacteria	Firmicutes	Mollicutes		Mycoplasmataceae
MYCMO	<i>Mycoplasma mobile</i> 163K	Bacteria	Firmicutes	Mollicutes		Mycoplasmataceae
MYCMS	<i>Mycoplasma mycoides</i>	Bacteria	Firmicutes	Mollicutes		Mycoplasmataceae
MYCPE	<i>Mycoplasma penetrans</i>	Bacteria	Firmicutes	Mollicutes		Mycoplasmataceae
MYCPN	<i>Mycoplasma pneumoniae</i>	Bacteria	Firmicutes	Mollicutes		Mycoplasmataceae
MYCPU	<i>Mycoplasma pulmonis</i>	Bacteria	Firmicutes	Mollicutes		Mycoplasmataceae
NANEQ	<i>Nanoarchaeum equitans</i>	Archaea	Nanoarchaeota			
NEIMB	<i>Neisseria meningitidis</i> MC58	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae
NEIMA	<i>Neisseria meningitidis</i> Z2491	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae
NITEU	<i>Nitrosomonas europaea</i>	Bacteria	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae
ANASP	<i>Nostoc</i> PCC7120	Bacteria	Cyanobacteria		Nostocales	Nostocaceae
OCEIH	<i>Oceanobacillus iheyensis</i>	Bacteria	Firmicutes		Bacillales	Bacillaceae
ONYPE	Onion yellows <i>Phytoplasma</i> OY-M	Bacteria	Firmicutes	Mollicutes	Acholeplasmatales	Acholeplasmataceae
PASMU	<i>Pasteurella multocida</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae
PHOPR	<i>Photobacterium profundum</i> SS9	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae
PHOLL	<i>Photorhabdus luminescens</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
RHOBA	<i>Rhodopirellula baltica</i> SH1	Bacteria	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae
PORGI	<i>Porphyromonas gingivalis</i> W83	Bacteria	Bacteroidetes	Bacteroidetes		
PROMA	<i>Prochlorococcus marinus</i> CCMP1375	Bacteria	Cyanobacteria	Prochlorales	Prochlorococcaceae	Prochlorococcus

PROMP	<i>Prochlorococcus marinus</i> MED4	Bacteria	Cyanobacteria	Prochlorales	Prochlorococcaceae	Prochlorococcus
PROMM	<i>Prochlorococcus marinus</i> MIT9313	Bacteria	Cyanobacteria	Prochlorales	Prochlorococcaceae	Prochlorococcus
PSEAE	<i>Pseudomonas aeruginosa</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
PSEPK	<i>Pseudomonas putida</i> KT2440	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
PSSEM	<i>Pseudomonas syringae</i> tomato DC3000	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
PYRAE	<i>Pyrobaculum aerophilum</i>	Archaea	Crenarchaeota	Thermoprotei	Thermoproteales	Thermoproteaceae
PYRAB	<i>Pyrococcus abyssi</i>	Archaea	Euryarchaeota	Thermococci	Thermococcales	Thermococcaceae
PYRFU	<i>Pyrococcus furiosus</i>	Archaea	Euryarchaeota	Thermococci	Thermococcales	Thermococcaceae
PYRHO	<i>Pyrococcus horikoshii</i>	Archaea	Euryarchaeota	Thermococci	Thermococcales	Thermococcaceae
RALSO	<i>Ralstonia solanacearum</i>	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae
RHOPA	<i>Rhodopseudomonas palustris</i> CGA009	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae
RICCN	<i>Rickettsia conorii</i>	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae
RICPR	<i>Rickettsia prowazekii</i>	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae
SALTt	<i>Salmonella enterica</i> subespécie <i>enterica</i> serovar typhi	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
SALTY	<i>Salmonella enterica</i> serovar typhimurium LT2	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
SALTe	<i>Salmonella enterica</i> subespécie <i>enterica</i> serovar typhi2	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
SHEON	<i>Shewanella oneidensis</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae
SHIFL	<i>Shigella flexneri</i> 2a linhagem 301	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
SHIFt	<i>Shigella flexneri</i> 2a linhagem 2457T	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
RHIME	<i>Sinorhizobium meliloti</i>	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
STAAM	<i>Staphylococcus aureus</i> Mu50	Bacteria	Firmicutes		Bacillales	
STAAW	<i>Staphylococcus aureus</i> MW2	Bacteria	Firmicutes		Bacillales	
STAAN	<i>Staphylococcus aureus</i> N315	Bacteria	Firmicutes		Bacillales	
STAEP	<i>Staphylococcus epidermidis</i> ATCC 12228	Bacteria	Firmicutes		Bacillales	
STRA5	<i>Streptococcus agalactiae</i> 2603	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
STRA3	<i>Streptococcus agalactiae</i> NEM316	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
STRMU	<i>Streptococcus mutans</i>	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
STRR6	<i>Streptococcus pneumoniae</i> R6	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
STRPN	<i>Streptococcus pneumoniae</i> TIGR4	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
STRPY	<i>Streptococcus pyogenes</i> M1 GAS	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
STRPm	<i>Streptococcus pyogenes</i> MGAS315	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
STRP8	<i>Streptococcus pyogenes</i> MGAS8232	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
STRPs	<i>Streptococcus pyogenes</i> SSI-1	Bacteria	Firmicutes		Lactobacillales	Streptococcaceae
STRAW	<i>Streptomyces avermitilis</i>	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Streptomycineae
STRCO	<i>Streptomyces coelicolor</i>	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Streptomycineae
SULSO	<i>Sulfolobus solfataricus</i>	Archaea	Crenarchaeota	Thermoprotei	Sulfolobales	Sulfolobaceae

SULTO	<i>Sulfolobus tokodaii</i>	Archaea	Crenarchaeota	Thermoprotei	Sulfolobales	Sulfolobaceae
SYNEL	<i>Synechococcus elongatus</i> PCC6301	Bacteria	Cyanobacteria		Chroococcales	
SYNPX	<i>Synechococcus</i> sp WH8102	Bacteria	Cyanobacteria		Chroococcales	
SYNY3	<i>Synechocystis</i> PCC6803	Bacteria	Cyanobacteria		Chroococcales	
THETN	<i>Thermoanaerobacter tengcongensis</i>	Bacteria	Firmicutes	Clostridia	Thermoanaerobacteriales	Thermoanaerobacteriaceae
THEAC	<i>Thermoplasma acidophilum</i>	Archaea	Euryarchaeota	Thermoplasmata	Thermoplasmatales	Thermoplasmataceae
THEVO	<i>Thermoplasma volcanium</i>	Archaea	Euryarchaeota	Thermoplasmata	Thermoplasmatales	Thermoplasmataceae
THEMA	<i>Thermotoga maritima</i>	Bacteria	Thermotogae		Thermotogales	Thermotogaceae
THET2	<i>Thermus thermophilus</i> HB27	Bacteria	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae
TREDE	<i>Treponema denticola</i> ATCC 35405	Bacteria	Spirochaetes		Spirochaetales	Spirochaetaceae
TREPA	<i>Treponema pallidum</i>	Bacteria	Spirochaetes		Spirochaetales	Spirochaetaceae
TROW8	<i>Tropheryma whipplei</i> TW08 27	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Micrococcineae
TROWT	<i>Tropheryma whipplei</i> Twist	Bacteria	Actinobacteria	Actinobacteridae	Actinomycetales	Micrococcineae
UREPA	<i>Ureaplasma urealyticum</i>	Bacteria	Firmicutes	Mollicutes		Mycoplasmataceae
VIBCH	<i>Vibrio cholerae</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae
VIBPA	<i>Vibrio parahaemolyticus</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae
VIBVU	<i>Vibrio vulnificus</i> CMCP6	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae
VIBVY	<i>Vibrio vulnificus</i> YJ016	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae
WIGBR	<i>Wigglesworthia brevipalpis</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
WOLPM	<i>Wolbachia</i> (endosymbiont of <i>Drosophila melanogaster</i>)	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae
WOLSU	<i>Wolinella succinogenes</i>	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacteriales	Helicobacteraceae
XANCP	<i>Xanthomonas campestris</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
XANAC	<i>Xanthomonas citri</i>	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
XYLFA	<i>Xylella fastidiosa</i> 9a5c	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
XYLFT	<i>Xylella fastidiosa</i> Temecula1	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
YERPm	<i>Yersinia pestis</i> biovar Mediaevails	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
YERPc	<i>Yersinia pestis</i> CO92	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
YERPk	<i>Yersinia pestis</i> KIM	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae

8. Trabalho apresentado na GRC on Structural, Functional and Evolutionary Genomics (2007)

Universidade de São Paulo
Instituto de Química

IN-SILICO CHARACTERIZATION OF A NOVEL DOMAIN OF THE CELL DIVISION PROTEIN AMIC

Robson F. de Souza^{1,2}, Sandro J. de Souza² and Frederico José Gueiros Filho¹

¹ Departamento de Bioquímica – Instituto de Química - Universidade de São Paulo – São Paulo – Brazil

² Ludwig Institute for Cancer Research – São Paulo Branch – São Paulo – Brazil

Abstract

We describe a novel protein domain, named IAMIC, for targeting of AmIC, due to its involvement in localization of *E. coli*'s AmIC protein to the division septum. This domain is found in homologs of N-acetylglucosaminyl-L-alanine amidase and the fimbrial outer membrane protein pilQ. We suggest this new domain mediates protein-protein interaction of periplasmic and extracellular proteins.

Introduction

In *E. coli*, the last step of cell division involves the complete separation of the daughter cells, which is accomplished through the localized enzymatic activity of AmC (N-acetylglucosaminyl-L-alanine amidase), a periplasmic component of the septal ring. Bernhardt and de Boer (2003) showed that the proteins AmC and AmA are exported to the periplasm through the twin-arginine transport system. Once in the periplasm, AmC is directed to the division septum, where it forms part that can be visualized using GFP-fusions. Bernhardt and de Boer also demonstrated that, in order to be restricted to a septal localization, AmC depends on the cell division protein FtsN and a ~150aa region located at its N-terminal, just after the export signals. The N-terminal region alone is also capable of localizing GFP fusions to the division septum, without AmC's enzymatic domain, a property that prompted us to explore whether this region could be an autonomously evolved domain.

Figure 1: Multiple sequence alignment of AmC homologs for representative Proteobacteria, including all three *E. coli* homologs (AmkA, AmkB and AmcC). The N-acetylglucosaminyl-L-alanine amidase protein in this family are characterized by a catalytic domain of the family Amidase_3 (see Figure 6). Sub-families are identified by colors on the left: AmkB (red), AmcC (blue), Xanthomonas and Pseudomonas AmcC homologs (green), AmkA (yellow) and amidases found in α-Proteobacteria (light blue). The N-terminal region of AmA homologs is not homologous to IAMIC, i.e. it is not recovered by iterated PSI-BLAST searches (ISS procedure, Figure 2).

Figure 2: Pipeline for detection of remote homologs used to scan for IAMIC family members. The panel on the left shows how new PSI-BLAST runs are initiated using homologous regions identified by PSI-BLAST using the last round of seeds. Criteria to select homologous PSI-BLAST alignments are (1) the returned e-value, which must not exceed PSI-BLAST's inclusion threshold, (2) the new alignment should have a minimum query coverage, i.e. the alignment must span at least 2% of the query sequence length, and (3) the region of the query sequence included in the new alignment must have a minimum overlap to the alignment used to select the query in the last ISS iteration (this is the score introduced by John and Sali (2004) to ensure transitivity for homology relationships).

Methods

While inspecting a Muscle based alignment of AmC homologs from members of the classes α-, β-, and γ-Proteobacteria (Figure 1), we noted the presence of a conserved region near the N-terminal end of the AmkB and AmcC sub-clades. Based on this alignment, we selected residues 31-145 of the *E. coli* AmC protein and used this sequence as a seed for a pipeline for intermediate sequence searches (ISS) (John and Sali, 2004), based on PSI-BLAST. The procedure implemented in our pipeline is shown schematically in Figure 2. Homologous sequences identified by ISS were extracted and aligned using ClustalW and Muscle. The alignments and selected individual sequences were used as input for the secondary structure prediction programs Jnet and PredictProtein. Disordered regions of *E. coli*'s AmC were predicted using DISOPRED. Phylogenetic trees were inferred using PhyML and substitution models selected by ProtTest.

Results

Using iterated PSI-BLAST searches (Figure 2), we demonstrated that segments homologous to the N-terminal region of AmC are present in several members of three families of periplasmic and extracellular proteins: PilQ (secretin), AmC/AmB (amidases), and a family of iron receptors from Cyanobacteria. Homology was also detected in a few other unrelated proteins (Figure 6). Proteins bearing IAMIC are found among Cyanobacteria, Proteobacteria, Clostridiales (Firmicutes) and Actinobacteria. Such a distribution could be either the result of lateral gene transfers among ancestors of these clades or vertical inheritance. Predictions of secondary structure and disordered regions suggest IAMIC homologs fold in a conformation comprising 8 β-strands, organized in two homologous sub-domains of 4 β-strands each and separated by a loop (Figures 3 and 4 and Table 1). Most of the proteins exhibiting the IAMIC domains have detectable signal peptides and/or domains known to be characteristic of extracellular proteins.

Figure 3: Predicted secondary structure for homologs of IAMIC found in iron-receptors of Cyanobacteria. Although this sub-family includes some of the less similar sequences, when compared to the seed AmC from *E. coli*, it shows the same structural features as those predicted for AmC and other sub-families by Jnet and other prediction programs.

Table 1: The N- and C-terminal regions of IAMIC are homologous, as demonstrated by the symmetry of ISS searches. The first row shows the e-value of the match of *E. coli*'s AmC full length IAMIC domain to itself, without the need of intermediate sequences. Rows below show the e-values for the matches to each IAMIC region by an intermediate sequence when the other region is used as seed for ISS.

Initial query sequence	Start	End	e-value	Number of intermediate sequences
IAMIC	27	145	8e-56	0
IAMIC_C	88	145	1e-23	0
IAMIC_N	31	77	7e-09	1
IAMIC_N	34	93	5e-22	0
IAMIC_N	98	159	2e-12	2

Figure 4: Domain architectures of IAMIC homologous proteins. Labeled domains are from the Pfam database and were detected using Hmmer and an e-value cutoff <= 0.01. The column Protein indicates family assignments and likely functional roles. Proteins and domains are drawn to scale.

REFERENCES

- Bernhardt, T.G. and de Boer, P.A.J (2003) The Escherichia coli amidase AmC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Molecular Microbiology*, 48 (5), 1171-1182.
- John, B and Sali A (2004) Detection of homologous proteins by an intermediate sequence search. *Protein Science* 13:54-62.

Discussion

Like AmC in *E. coli*, a PilQ family member from *Myxococcus xanthus* (Figure 6) is another example of a IAMIC containing protein which is known to display a polar distribution. It is thus possible that IAMIC acts as a polar targeting domain for both AmC and PilQ in these and other organisms. Notably, *Pseudomonas aeruginosa* was reported to produce polar type IV pili and has pilQ genes encoding proteins with IAMIC domains, but the localization of *Pseudomonas* PilQ homologs has not been determined. Interestingly, although some hemorrhagic strains of *E. coli* produce polar pili, their PilQ proteins lack regions of homology to IAMIC. Therefore, current experimental data on the distribution of IAMIC containing proteins is insufficient to establish a general role in protein localization.

Bernhardt and de Boer (2003) reported that AmC depends on FtsN to localize. Nevertheless, the target recognized by IAMIC to localize AmC in *E. coli* remains unknown and could be either FtsN or some septal component modified by FtsN. Another evidence suggesting that IAMIC localization might be based on protein interactions is the observation that some *Bacillus* amidases that localize to the septum have prokaryotic homologs of the SH3 domain – a typical protein-protein interaction domain – at the position where IAMIC is found in their Gram- counterparts.

Figure 5: Disordered regions predicted for *E. coli*'s AmC. Two strongly supported disordered regions, predicted by DISOPRED, separate the catalytic and IAMIC domains and the two sub-domains of IAMIC. The conservation of this pattern among different homologs helped the definition of domain boundaries.

Figure 6: Multiple sequence alignment of the N- and C-terminal regions of sequences from several sub-families of IAMIC homologs. This alignment highlights the conserved signatures observed in most members of the family (represented by the sequence logo shown above).

WORK SUPPORTED BY: