

Universidade de São Paulo

Instituto de Química

**Seqüenciamento de parte do genoma de *Xylella*
fastidiosa e análise do metabolismo de carboidratos
no genoma anotado**

Ari José Scattone Ferreira

Dissertação de Mestrado

Orientador: Prof. Dr. Hamza El-Dorri

São Paulo

2000

*“Seqüenciamento de parte do Genoma de Xylella fastidiosa e
Análise do Metabolismo de Carboidratos no Genoma
Anotado”*

ARI JOSÉ SCATTONE FERREIRA

**Dissertação de Mestrado submetida ao
Instituto de Química da Universidade de São Paulo
como parte dos requisitos necessários à obtenção do
grau de Mestre em Ciências - Área: Bioquímica.**

Aprovado por:

Prof. Dr. HAMZA FAHMI ALI EL DORRY
IQ - USP
(Orientador e Presidente)

Profa. Dra. ALINE/MÁRIA DA SILVA
IQ - USP

Profa. Dra. ELIZABETH ANGÉLICA LEME MARTINS
IB

SÃO PAULO
08 DE MAIO DE 2000.

Agradecimentos

Agradeço ao Prof. Dr. Hamza El-Dorry não somente a orientação nesse projeto, mas também as inúmeras perguntas instigantes, que me pegavam geralmente de surpresa e que fizeram com que eu aprendesse ainda mais. Fico grato também pela confiança que depositou em mim para coordenar esse projeto e pela amizade que cresceu resultante dele.

Ao Prof. Dr. Roberto Vicente Santelli devo muito do interesse pela Biologia Molecular, visto ter sido meu professor na graduação e agradeço-lhe o apoio e ajuda nesse projeto, em que foi meu coorientador, e a amizade desde que o conheci.

A ajuda e o apoio que recebi de meus amigos e colegas de seqüenciamento foram fundamentais para que esse projeto tivesse sucesso e por isso agradeço aos que diretamente estiveram envolvidos: Luci Navarro-Cattapan, Eric Bonaccorsi e Wilton Lima e aos que, trabalhando em outros laboratórios do Projeto Genoma *Xylella fastidiosa*, tornaram-se meus amigos: Adriana Matsukuma, Cássia Docena, Mara Silvestri, Anamaria Camargo e Cássio Baptista.

Foram indispensáveis os debates, as consultas e a colaboração com diversos pesquisadores que estiveram trabalhando no Projeto Genoma *Xylella fastidiosa*. Assim sendo, fica meu agradecimento às Profas. Dras. Ana Cláudia Rasera, Aline Maria Silva, Marilis Marques, Suely Gomes e aos Profs. Drs. Ronaldo Quaggio, Sérgio Verjovski-Almeida, Paulo Arruda e Andrew Simpson. Ao último agradeço imensamente a confiança depositada em minha pessoa para realizar o trabalho de anotação da categoria de Metabolismo Intermediário. Em particular fica meu agradecimento aos pesquisadores do Laboratório de Bioinformática: Profs. Drs. João Meidanis, João Carlos Setúbal, João Paulo Kitajima, aos alunos Felipe da Silva e Vagner Okura e aos demais envolvidos na bioinformática que disponibilizaram o grande número de ferramentas utilizadas durante todo esse projeto, desde a parte experimental até a confecção dessa dissertação. Agradeço também a todos os outros participantes do Projeto Genoma *Xylella fastidiosa* pelas contribuições diretas ou indiretas a esse projeto.

Um projeto que se realiza depende, além do trabalho árduo, de um ambiente que favoreça-o onde a amizade e a preocupação de alguns para com outros seja sempre presente. Assim, agradeço a Daniella, Lia, Nathalie, Tatiana, Zilda, Augusto, Euclides, Felipe, Ribamar e a Profa. Carla Columbano e seus alunos, pela ajuda, preocupação e dicas que me deram no laboratório. Agradeço também a Dirce por ter me ensinado os primeiros passos num laboratório de Biologia Molecular e me ajudar a andar com minhas próprias pernas nessa nova e desafiante trilha do conhecimento.

Agradeço pela paciência, compreensão e alegria dos últimos tempos que me foram dadas por aqueles que hoje considero mais do que amigos, mas parte de minha família: Alessandra, Alexandra, Joice, Elaine, João Henrique e Eric.

Não há melhor crítica sobre seu trabalho além daquela omitida por alguém que o apoia condicionalmente, não tem o mínimo pudor em delatar seus erros e defeitos, mas tem uma imensa preocupação com sua vida, ou seja, uma irmã. Por todas essas características, as discussões e os risos que tivemos juntos agradeço a minha irmã Ana Luiza.

Uma pessoa é a soma de seus conhecimentos e experiências adquiridos durante toda uma vida. Tais conhecimentos e experiências são depositados sobre um cerne moldado desde sua geração por um par de pessoas que se eximiu de seus sonhos para poder ver forjados os sonhos de um novo outro. A meus pais, Aristides e Zenaide, que nunca mediram esforços para que pudesse ter o que consideraram como o melhor do conhecimento, inúmeras experiências de vida e ser o que sou hoje, não posso somente agradecer, mas dedicar-lhes esse novo sonho realizado e retribuir-lhes sempre o amor que têm por mim.

Agradeço ao Conselho Nacional de Pesquisa (CNPq) pela bolsa de mestrado a mim concedida durante o desenvolvimento desse projeto.

Índice

I. INTRODUÇÃO	4
I A. CITRICULTURA E A CLOROSE VARIEGADA DOS CITROS (CVC)	6
I B. <i>Xylella fastidiosa</i>	10
I B 1. Localização e características nutricionais do habitat	13
I B 2. Patogenicidade	14
I B 3. Transmissão	16
I C. O PROJETO GENOMA <i>Xylella fastidiosa</i>	19
II. OBJETIVOS	20
III. MATERIAIS E MÉTODOS	21
III A. MEIOS DE CULTURA	21
III B. CEPAS DE BACTÉRIAS E VETORES UTILIZADOS	21
III C. SEQÜENCIAMENTO DO GENOMA DE <i>Xylella fastidiosa</i>	21
III D. OBTENÇÃO DE BIBLIOTECAS DE “SHOTGUN” A PARTIR DE COSMÍDEOS	23
III E. PREPARAÇÃO DE DNA DE PLASMÍDEOS RECOMBINANTES	24
III F. SEQÜENCIAMENTO DE INSERTOS EM PLASMÍDEOS PUC18	25
III G. MONTAGEM DE COSMÍDEOS A PARTIR DE LEITURAS DE BIBLIOTECAS DE “SHOTGUN”	27
III H. SEQÜENCIAMENTO TOTAL DE INSERTO DE PLASMÍDEO DE “SHOTGUN” GENÔMICO	29
III I. DETERMINAÇÃO DE ORFS NOS INSERTOS DE COSMÍDEOS	30
III J. ANOTAÇÃO DE ORFS	31
III K. BUSCA DE VIAS METABÓLICAS E ANOTAÇÃO DE CATEGORIAS DE GENES	33
IV. RESULTADOS	36
IV A. SEQÜENCIAMENTO E ANOTAÇÃO DE INSERTOS DE COSMÍDEOS	36
IV A 1. Cosmídeo X0QH-07A10	36
IV A 2. Cosmídeo X0QH-02E06	39
IV A 3. Cosmídeo X0QH-06A08	42
IV A 4. Cosmídeo X0QH/QS-02G12	45
IV A 5. Cosmídeo X0QH/QS-05D08	47
IV B. FECHAMENTO DE “GAPS”	49
IV B 1. GFS X0QH-00J42	49
IV C. SEQÜENCIAMENTO DE CLONES DE BIBLIOTECAS DE “SHOTGUN” GENÔMICO	50
IV D. ANOTAÇÃO DE REGIÕES DE COSMÍDEOS SEQÜENCIADOS POR OUTROS LABORATÓRIOS	51
IV E. ANOTAÇÃO DA CATEGORIA I: METABOLISMO INTERMEDIÁRIO	52
V. DISCUSSÃO	57
V A. ESTRATÉGIAS DE SEQÜENCIAMENTO	57
V B. CONSIDERAÇÕES SOBRE O METABOLISMO DE CARBOIDRATOS DE <i>Xylella fastidiosa</i>	62

VI A. MAPA DO PLASMÍDEO PUC18	70
VI B. DISPOSIÇÃO DOS CLONES SEQÜENCIADOS NO CROMOSSOMO DE <i>Xylella fastidiosa</i>	71
VI C. COBERTURA DO SEQÜENCIAMENTO DO CROMOSSOMO DE <i>Xylella fastidiosa</i>	72
VI D. MAPAS METABÓLICOS COMPARATIVOS	73
VI E. COSMÍDEO X0QH-02E06	91
VI E 1. Relatório de finalização de seqüenciamento	91
VI E 2. Cronograma de estágios do seqüenciamento	91
VI E 3. Digestão do cosmídeo X0QH-02E06 com enzimas de restrição e digestão eletrônica	92
VI F. COSMÍDEO X0QH/QS-02G12	93
VI F 1. Relatório de finalização do seqüenciamento	93
VI F 2. Cronograma de estágios de seqüenciamento	93
VI F 3. Digestão do cosmídeo X0QH/QS-02G12 com enzima de restrição e digestão eletrônica	94
VI G. FRAGMENTO 05D08	96
VI G 1. Relatório de finalização do seqüenciamento	96
VI G 2. Cronograma de estágios de seqüenciamento	96
VI H. FRAGMENTO 15D08	97
VI H 1. Relatório de finalização do seqüenciamento	97
VI H 2. Cronograma de estágios de seqüenciamento	97
VI I. COSMÍDEO X0QH-06A08	98
VI I 1. Relatório de finalização do seqüenciamento	98
VI I 2. Cronograma de estágios de seqüenciamento	98
VI I 3. Digestão do cosmídeo X0QH-06A08 com enzimas de restrição e digestão eletrônica	99
VI J. COSMÍDEO X0QH-07A10	101
VI J 1. Relatório de finalização do seqüenciamento	101
VI J 2. Cronograma de estágios de seqüenciamento	101
VI J 3. Observação com relação à extremidade quimérica do cosmídeo X0QH-07A10	102
VI J 4. Digestão do cosmídeo X0QH-07A10 com enzimas de restrição e digestão eletrônica	102
VI K. GFS X0QH-00J42 (X0QH1013G02)	103
VI K 1. Relatório de finalização do seqüenciamento	103
VI K 2. Cronograma de estágios de seqüenciamento	103
VI L. LISTA DE ORFS ANOTADAS NA CATEGORIA I: METABOLISMO INTERMEDIÁRIO	104
VI M. LISTA DE ORFS ANOTADAS NA CATEGORIA DE PROCESSOS CELULARES: TRANSPORTE DE CARBOIDRATOS, ÁCIDOS ORGÂNICOS E ÁLCOOIS	111
VI N. BLASTP DA FRUTOSE-1,6-BISFOSFATO ALDOLASE DE <i>Xylella fastidiosa</i>	112
VI O. ALINHAMENTO DE FRUTOSE-1,6-BISFOSFATO ALDOLASES COM A SUPOSTA ENZIMA DE <i>Xylella fastidiosa</i>	113
VI P. INFORMAÇÕES SOBRE O GENOMA DE <i>Xylella fastidiosa</i>	115

Resumo

Em 1987 uma nova doença de citros, conhecida popularmente como “Amarelinho” e denominada posteriormente de Clorose Variiegada dos Citros (CVC), foi identificada nos estados de São Paulo e Minas Gerais. Atualmente 34,1% de laranjeiras apresentam sintomas da CVC. Em 1989 foi identificada nos vasos do xilema de tecidos de plantas afetadas pela CVC a bactéria *Xylella fastidiosa*.

Xylella fastidiosa é uma bactéria Gram-negativa limitada aos vasos do xilema de uma grande diversidade de plantas hospedeiras que causa perdas em diversas culturas.

Em 1997 a FAPESP, conjuntamente ao FUNDECITRUS, desenvolveu o projeto de seqüenciamento do genoma do fitopatógeno *Xylella fastidiosa*.

Em nosso laboratório foram seqüenciados cinco cosmídeos, sendo dois parcialmente, representando linearmente 148.642 pb, em torno de 5,5 % do cromossomo de *Xylella fastidiosa*. Também foram seqüenciados clones das bibliotecas de “shotgun”. Por fim foi realizado o seqüenciamento completo de um inserto em plasmídeo, para o fechamento de um “gap” entre contíguos genômicos.

Foram realizadas as anotações de cosmídeos seqüenciados em nosso e em outros laboratórios e verificações de anotações na Categoria I, Metabolismo Intermediário, assim como buscas às proteínas importantes dentro dessa categoria que não estavam presentes no banco de dados de *Xylella fastidiosa*. Foram identificados como completos a glicólise, o ciclo dos ácidos tricarboxílicos e a cadeia de transporte de elétrons. Estão aparentemente incompletas a via das pentoses e a gliconeogênese. Não foram identificadas enzimas do complexo de beta-oxidação de ácidos graxos, da via do glioxilato ou relacionadas à fermentação alcoólica.

I. Introdução

Na última década a Biologia, em especial a área de Biologia Molecular, passou por uma grande mudança com o advento de projetos de seqüenciamento genômico de diversos organismos. Bactérias¹, arqueabactérias² e eucariotos³, uni e pluricelulares, já tiveram seus genomas completamente seqüenciados e o número de projetos envolvendo tal objetivo tem crescido incessantemente (para uma lista sempre atualizada consultar referência 4).

O interesse na determinação da seqüência completa do genoma de um organismo baseia-se na possibilidade de identificação de genes que concedam as características únicas do mesmo e que tenham importância econômica, científica ou social. Sempre há conseqüentemente uma contribuição para o entendimento cada vez maior e melhor dos processos químicos, bioquímicos e biológicos, em todas suas nuances, que geram a enorme diversidade existente em nosso meio ambiente.

Dentre as diversas linhas de pesquisa nessa área, a da determinação de seqüências do genoma de microorganismos patógenos tem se destacado pela necessidade do entendimento e controle dos mesmos. É esperado que dentro de alguns anos genomas de espécies bacterianas patógenas de humanos, animais e plantas já terão sido seqüenciados, havendo em muitos casos mais de uma seqüência por espécie ou por gênero⁵.

O seqüenciamento do genoma de um organismo provê um dado experimental inicial sobre o qual pode-se especular e inferir sobre a importância de determinados processos metabólicos para o mesmo. Tal especulação baseia-se na anotação de um genoma, composta de duas fases: 1) a determinação computacional de regiões que possivelmente codificam proteínas, denominadas fases abertas de leitura ou ORFs (Open Reading Frames); 2) a determinação de similaridade de seqüência das proteínas codificadas nessas ORFs através de comparação da seqüência protéica ou nucleotídica das mesmas com bancos de dados de proteínas e genes já conhecidos ou anotados.

A anotação tem um papel intermediário importante no estudo de um organismo cujo genoma foi seqüenciado uma vez que fornece informações iniciais sobre a provável presença ou ausência de determinados processos metabólicos e a importância relativa desses processos para o organismo, pressupondo-se que quão maior a parte do genoma dedicada a tais processos, maior sua importância. Por exemplo, quase 5% do conteúdo do genoma de *Mycobacterium genitalium* é dedicado a um gene de adesina e a seqüências relacionadas, que poderiam ser usadas para produzir variantes antigênicas da proteína com o objetivo de evitar a resposta imune. Por outro lado, não foram identificadas em várias bactérias, por exemplo *Haemophilus influenzae*, enzimas do ciclo dos ácidos tricarbóxicos, o que pode representar, ou não, a existência de vias metabólicas alternativas nesses organismos⁶.

Terminada a etapa da determinação da estrutura do genoma de um organismo e anotação de suas ORFs, as informações obtidas são de extrema utilidade para estudos laboratoriais posteriores denominados genoma funcional. Nesses estudos é verificada a veracidade das hipóteses levantadas na anotação com o objetivo de relacioná-las às características biológicas do organismo. Em combinação com experimentos para determinação de função, estratégias baseadas em genoma podem favorecer linhas de pesquisa direcionadas para a elucidação de mecanismos de patogenicidade bacteriana⁶.

Em 1997 a Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), conjuntamente o Fundo de Defesa da Citricultura de São Paulo (FUNDECITRUS), desenvolveu o projeto de seqüenciamento do genoma do fitopatógeno *Xylella fastidiosa*, causador da doença Clorose Variegada dos Citros (CVC) que aflige a citricultura paulista. Para tal foi definida uma rede virtual de 33 laboratórios com um laboratório de coordenação do projeto, dois laboratórios de seqüenciamento centrais, 29 laboratórios de seqüenciamento e um centro de bioinformática, denominada rede ONSA (Organization for Nucleotide Sequence and Analysis). Dentre os diversos motivos relacionados para a realização desse projeto de seqüenciamento estavam a inexistência de um modelo de interação entre o patógeno e os citros e questões como: qual a diversidade genética entre as diferentes cepas que compõem a espécie *Xylella fastidiosa*, qual o tamanho real do genoma, se haveria a ocorrência de plasmídeos na cepa que ataca os citros e a possível associação destes com a patogenicidade, e se haveria genes de virulência ou avirulência que estariam relacionados aos sintomas da CVC ou aos vetores⁷.

I A. Citricultura e a Clorose Variegada dos Citros (CVC)

A citricultura vem se destacando desde a década de 1980 como a mais importante cultura na agropecuária do estado de São Paulo. Numa área de aproximadamente 630.200 hectares há por volta de 164 milhões de árvores que produzem anualmente 374 milhões de caixas (40,8 kg cada) de laranjas, o que corresponde a 87% da produção brasileira e 30% da mundial⁷.

A produção de suco de laranja concentrado consome 70% da produção, sendo que o Estado de São Paulo e o Estado da Flórida, nos Estados Unidos da América, são responsáveis por 90% da produção mundial de suco concentrado. Tal atividade econômica emprega, no Estado de São Paulo, aproximadamente 400.000 pessoas em 204 municípios, 20.000 produtores e produz um saldo anual de exportação da ordem de US\$ 1,2 a 1,4 bilhões⁷.

Nos últimos anos a produção brasileira tem aumentado devido à expansão da lavoura pelo Estado de São Paulo e outros, entretanto a produtividade, quando comparada à americana, continua baixa: 2,0 caixas/árvore/ano no Brasil e 6,0 caixas/árvore/ano na Flórida. As principais razões dessa baixa produtividade têm sido doenças, déficit hídrico e de fatores nutricionais⁷.

Em 1987 uma nova doença de citros, conhecida popularmente como “Amarelinho” e denominada posteriormente de Clorose Variegada dos Citros (CVC), foi identificada nos estados de São Paulo e Minas Gerais. Inicialmente encontrada em pomares de Colina, município do centro-oeste do estado de São Paulo, a doença foi posteriormente detectada nas regiões norte e nordeste do estado de São Paulo e no estado de Minas Gerais⁸.

No início da década de 1990 a CVC teve uma rápida disseminação, sendo que em um período de 6 anos já estava presente em 83% da principal área de citricultura do estado de São Paulo: fora detectada em 134 municípios do estado, além de ter sido encontrada em outros estados da União⁸. Enquanto em 1992 pesquisas demonstravam que em 64% dos pomares havia menos de 1 árvore com sintomas detectáveis, em 1996 foram encontradas plantas, com idade entre 3 e 5 anos, com sintomas severos em 88% dos pomares.

Os dados mais recentes (1997) mostram que do total estimado de laranjeiras 34,1% apresentam sintomas da CVC, principalmente nas regiões norte, noroeste e central⁷. As conseqüentes perdas econômicas na citricultura do estado de São Paulo tornaram a CVC a doença de maior impacto e a ameaça mais preocupante a essa atividade econômica na época.

Os sintomas da doença incluem: clorose foliar, apresentando manchas amarelas na face superior das folhas, que correspondem a manchas marrons na face inferior ou em ambas as faces, que podem se coalescer (Figura 1); redução no tamanho da laranjeira, com a ocorrência de copa plana, quando esta é infectada ainda jovem, ou redução da velocidade de crescimento devido aos efeitos sobre os ramos em desenvolvimento; e, o principal e com conseqüências econômicas de maior relevância, a ocorrência de frutos pequenos (algumas vezes até três vezes menores do que o normal), com sabor muito ácido, de amadurecimento precoce, coloração amarela intensa com manchas marrons, que podem vir a tomar toda sua superfície, e constituição rígida (Figura 2). Tais características torna os frutos impróprios tanto para o mercado de frutas frescas como para a produção de suco concentrado⁷. A produção de uma laranjeira é severamente reduzida de 70 quilogramas em média para 17 quilogramas⁸.



Figura 1: Primeiro sintoma da Clorose Variegada dos Citros (CVC): o aparecimento de manchas amarelas nas folhas dos ramos infectados. Na primeira fotografia, estão dispostas as duas faces de uma mesma folha afetada pela CVC, mostrando os pontos marrons na face inferior, à esquerda, correspondentes às manchas amarelas na face superior, à direita (V. Rossetti). Um ramo com sintomas moderados da CVC é mostrado na segunda fotografia (H.Purcell).

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo



Figura 2: Comparação entre frutos provenientes de laranjeiras afetada pela CVC, á direita, e não afetada pela doença, à esquerda (H. Purcell).

Em 1989 foi identificada nos vasos do xilema de tecidos de plantas afetadas pela CVC uma bactéria similar ao gênero *Xylella*. Especialistas do Institute National de Recherches Agronomiques (INRA) de Bordeaux, França, sugeriram que a bactéria seria *Xylella fastidiosa* Wells (Figura 3). A presença da bactéria *Xylella fastidiosa* Wells em plantas afetadas pela CVC foi confirmada por pesquisadores do Instituto Biológico de São Paulo através de microscopia eletrônica (Figura 4). A bactéria também foi confirmada como um tipo de *Xylella fastidiosa* Wells por serologia. Em 1993, pesquisadores do INRA e do Instituto Biológico realizaram o isolamento da bactéria em meio de cultura que favorecia um crescimento acelerado da bactéria⁸.

Para determinar se uma bactéria isolada em meio de cultura é realmente a causadora de uma doença, é realizado um inóculo da mesma num organismo hospedeiro saudável em ambiente controlado, para a inexistência de qualquer outra infecção. Acompanha-se então o desenvolvimento do mesmo. Ocorrendo o aparecimento dos sintomas da doença, é realizado então novo isolamento do patógeno para verificação de que realmente é o que foi inoculado no organismo. Em caso afirmativo, diz-se que o patógeno reisolado completou os postulados de Koch, ou seja, é o real causador dos sintomas desenvolvidos pelo hospedeiro de onde foi inicialmente isolado. No ano de 1993, foram obtidos com sucesso inóculos e reisolamentos da bactéria em citros, completando os postulados de Koch para *Xylella fastidiosa*⁸.

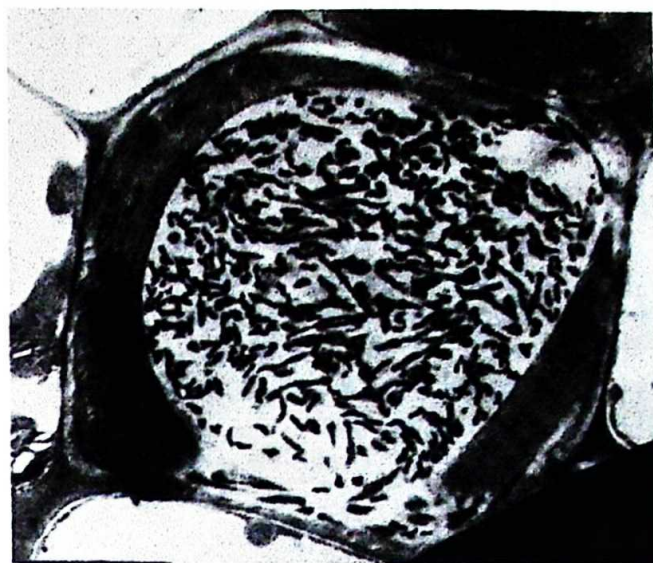


Figura 3: Primeira observação da bactéria *Xylella fastidiosa* no xilema de folhas de citros afetadas pela CVC (M. Garnier)



Figura 4: Observação da bactéria *Xylella fastidiosa* no xilema de pecíolos de folhas através de microscopia eletrônica de varredura (J.E.O. Lima).

I B. *Xylella fastidiosa*

*Xylella fastidiosa*⁹ é uma bactéria Gram-negativa limitada aos vasos do xilema de uma grande diversidade de plantas hospedeiras. É considerada atualmente espécie única do gênero *Xylella*, grupo *Xanthomonas*, família Lysobacteriaceae, ordem Lysobacteriales, subdivisão Gama das Eubactérias.

Xylella fastidiosa é a espécie na qual estão incluídas todas as bactérias limitadas a xilema e é derivada do termo XLB, ou em inglês, xylem-limited bacteria. Esse termo foi utilizado inicialmente para descrever patógenos vegetais procarióticos difíceis de serem isolados por procedimentos bacteriológicos convencionais. Tais organismos necessitam de meios complexos para seu crescimento e ocorrem somente no xilema de plantas infectadas, são transmitidos por insetos que se alimentam do xilema e causam doenças de difícil controle em diversas culturas economicamente importantes como a citricultura^{9,10}.

As bactérias limitadas ao xilema têm formato cilíndrico com paredes celulares enrugadas, não são flageladas nem formam esporos. Medem de 0,3 a 0,5 μm de diâmetro e de 1 a 5 μm de comprimento (Figura 5). Crescem bem a temperatura de 20 a 25 °C e pH variando de pH 6,7 a 7,0. Os tamanhos dos genomas estariam nas faixas de $1,4 \pm 0,2$ e entre 1,92 e 2,42 Mpb⁷. Informações atualizadas sobre o genoma de *Xylella fastidiosa* estão resumidas no Anexo, página 113⁴³.

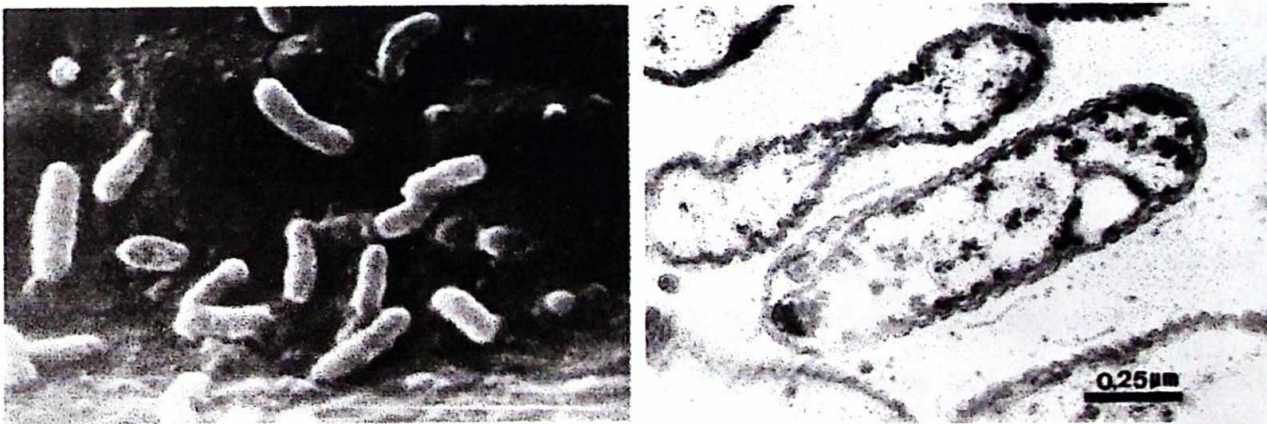


Figura 5: Micrografias de *Xylella fastidiosa* no xilema de pecíolos de folhas de citros (J.E.O. Lima), à esquerda, e mostrando a parede celular típica, à direita (x89.000, C.M. Chagas)

Estudos iniciais, em 1985, indicavam que o conteúdo GC do genoma variava de 50,1 a 54,0 % . Tal dado forneceu base para a distinção entre as bactérias Gram-negativas limitadas ao xilema (XLB) e o gênero *Rickettsiae* sp. (GC por volta de 30%). Estudos bioquímicos, sorológicos, de enzimas extracelulares, de ácidos graxos, de DNA e RNA de 25 isolados de XLB de videira, pessegueiro, amendoeira, ameixeira, olmo, plátano, carvalho e amoreira concluíram que todo o grupo era homogêneo e que haviam diferenças de crescimento *in vitro* entre os isolados¹¹.

Diferentes cepas de XLB eram conhecidas por afetarem diversas culturas causando doenças como: a doença de Pierce em videiras, a escaldadura foliar em ameixeiras, uma doença de má formação em pessegueiros, entre diversas outras. Em estudos de relação genética entre 5 cepas de XLB, as cepas não puderam ser distinguidas no nível de espécie devido à elevada homologia de DNA (75 a 100%) determinada por hibridizações^{12,13}.

A homologia entre a cepa causadora da doença de Pierce e a causadora da escaldadura foliar em ameixeiras era suficientemente baixa (75%) para sugerir uma distinção no nível varietal e possivelmente de subespécie. Dados sorológicos e de análise de ácidos graxos fortaleceram a conclusão de que as XLB constituíam um grupo homogêneo^{12,13}.

A homologia entre a cepa da escaldadura foliar em ameixeira e várias cepas bacterianas de espécies não limitadas ao xilema foi de menos de 5%. Somando-se a esse fato que as XLB não se relacionam a organismos que tenham conteúdo GC fora da faixa de 45 a 56 mol% ou genoma fora da faixa de 1,5 a 3,0 Mb concluiu-se que não eram relacionadas no nível de gênero a espécies de *Agrobacterium*, *Erwinia*, *Pseudomonas* e *Xanthomonas*^{12,13}.

A análise das seqüências de RNA ribossomal 16S determinou similaridades às de outras bactérias, sugerindo que as XLB pertenciam ao subgrupo Gama das Eubactérias, que inclui os fitopatógenos do gênero *Xanthomonas*. Entretanto as XLB diferiam genotípica e fenotipicamente das espécies do gênero *Xanthomonas* para pertencerem a um gênero separado. Foi proposto então o nome de *Xylella fastidiosa* para estabelecer um novo gênero com uma única espécie incluindo todas as cepas de bactérias Gram-negativas fastidiosas limitadas ao xilema (XLB ou FXLB)^{12,13}.

Há entretanto suficientes diferenças para uma separação no nível de subespécie em *Xylella fastidiosa*. Baseando-se em diferentes necessidades nutricionais, as cepas de *Xylella fastidiosa* podem ser separadas em dois grupos distintos: o grupo a que pertence a cepa causadora da doença de Pierce e o grupo a que pertence a cepa causadora da má formação de pêssegos. Cepas que ocorrem em videira, amendoeira, olmo e amoreira têm tempo de duplicação mais rápido, entre 0,5 e 1,6 dias, enquanto que as cepas que ocorrem em pessegueiro e ameixeiras têm o mais lento, até 2,0 dias. Teste de transmissão recíproca indicam que a doença de Pierce em videiras, a escaldadura foliar em amendoeiras e o nanismo da alfafa são causados pela mesma cepa, enquanto que a doença de má formação em pessegueiros e escaldadura foliar em ameixeiras são causadas por uma mesma cepa diferente da causadora do primeiro grupo de doenças^{12,13}. Uma análise utilizando a técnica de RAPD (Randomly Amplified Polymorphic DNA) entre 11 cepas de *Xylella fastidiosa* de ameixeira, amendoeira, amoreira, citros, olmo e videira sugeriu a existência de 5 grupos de bactérias de acordo com a origem: citros, ameixeira - olmo, videira – *Ambrosia* sp., amendoeira e amoreira.

A cepa detectada em citros (causadora da CVC no Brasil e da “Pecosita” na Argentina) e no café (no Brasil) parece diferir genética e biologicamente das outras cepas encontradas principalmente na América do Norte¹¹.

Proteínas de duas cepas de *Xylella fastidiosa* presentes em citros foram analisadas através de eletroforese para comparação com cepas presentes em videiras, *Ambrosia* sp., plátano e ameixeira, sendo detectadas similaridades. Estudos sorológicos entre cepas de *Xylella fastidiosa* utilizando a técnica DIBA (Dot Immuno Binding Assay) mostraram detecções positivas em amostras da cepa causadora da CVC e nas do grupo do pessegueiro, e com menor intensidade nas do grupo da doença de Pierce. Além disso compatibilidade sorológica entre cepas de CVC e da doença de Pierce foi detectada pela técnica DAS-ELISA (Double Antibody Sandwich – Enzyme Linked Immunosorbent Assay), além de haver reatividade total com as do grupo do pessegueiro. Tais dados sugerem que a cepa causadora da CVC possa ser uma intermediária entre esses dois grupos¹¹.

I B 1. Localização e características nutricionais do habitat

Xylella fastidiosa forma normalmente agregados nos vasos do xilema dos pecíolos, nervuras foliares e do tronco. Esses agregados aparentemente se mantêm coesos através de fibras de polissacarídeos extracelulares produzidas pela bactéria. Tais fibras são descritas como microfibrilas e fimbrias densas em elétrons, sendo mais abundantes nas extremidades do corpo cilíndrico das bactérias e assemelhando-se à massa entrelaçada de fibras de polissacarídeos que se estende da superfície de várias outras bactérias denominada glicocálix. Entre as funções do glicocálix se destacam: a capacidade de adesão entre as bactérias e destas a um tecido do hospedeiro em um ambiente favorável, a possível conservação e concentração de enzimas digestivas e o possível papel de resina de troca iônica, uma vez que tem densidade eletrônica elevada, o que tornaria o glicocálix um reservatório nutricional ligando íons ao agregado bacteriano e tornando-os disponíveis num ambiente nutricional diluído. Há evidências de que as fibras das cepas de *Xylella fastidiosa* causadoras da doença de Pierce funcionem como um glicocálix^{12,13}.

Visto que *Xylella fastidiosa* é uma espécie de bactéria limitada ao xilema, cabe aqui uma descrição das condições nutricionais que a bactéria encontra, e que podem estar relacionadas ao seu crescimento e patogenicidade.

Os elementos traqueiais do xilema das plantas hospedeiras de *Xylella fastidiosa* oferecem um ambiente físico e nutricional único para seu crescimento. A seiva bruta, suspensão de nutrientes proveniente das raízes que corre pelo xilema em direção às folhas, é um meio nutricionalmente diluído, embora contenha uma grande variedade de nutrientes essenciais. A quantidade e composição de tais nutrientes varia com o ciclo circadiano e sazonal, a localização, a idade, a saúde e a espécie da planta^{12,13}.

Dados sobre a seiva bruta de *Vitis rotundifolia*, espécie de videira, que também pode ser afetada por *Xylella fastidiosa*, indicam que ela consiste de uma solução pouco concentrada de moléculas orgânicas pequenas (aminoácidos, outros ácidos orgânicos e açúcares), ânions e cátions de micro e macroelementos, e concentrações muito baixas de reguladores de crescimento como citocininas e giberelinas. As concentrações totais de aminoácidos e ácidos orgânicos determinadas foram praticamente iguais (2 a 7 mmol/L) e os açúcares somaram uma fração menor do perfil total (menos de 0,2 mmol/L). Glutamina representava cerca de 80% do nitrogênio orgânico e 70% do nitrogênio total transportado pelo xilema¹⁴.

Todos os micro e macronutrientes essenciais analisados (NH_4^+ , NO_3^- , P, K, Ca, Mg, Mn, Fe, Cu, Zn, Na e Cl) foram detectados, sendo cálcio e potássio os macronutrientes mais concentrados e cobre o micronutriente menos concentrado. Todos os aminoácidos, exceto triptofano foram detectados, sendo glutamina cerca de 80% do total (tal concentração representa a soma das concentrações da amida e do ácido glutâmico devido ao método de análise). Histidina, tirosina, cisteína, metionina e lisina ocorreram em níveis muito baixos. As concentrações de poliaminas e de proteínas foram consideradas desprezíveis. Foram identificados 7 outros ácidos orgânicos, entre os quais os de maior concentração foram os ácidos cítrico, tartárico e málico. Na pequena fração referente aos açúcares foram detectados glicose, frutose e sorbitol, ou manitol (o método de análise utilizado não distinguia entre ambos), sendo glicose e frutose em maior parte. Os compostos orgânicos em maior concentração foram glutamina (2,3 a 5,0 mmol/L), ácido málico (0,4 a 2,8 mmol/L), ácido tartárico (0,6 a 1,7 mmol/L) e ácido cítrico (0,3 a 2,0 mmol/L)¹⁴.

I B 2. Patogenicidade

A patogenicidade de *Xylella fastidiosa* parece estar relacionada tanto à capacidade de multiplicação nos vasos do xilema como à habilidade de movimentação sistemática de um vaso para outro^{12,13}. As fibras de polissacarídeos podem estar envolvidas na patogenicidade, parecendo haver uma conexão entre agregação e virulência, ou agressividade^{12,13}. Cepas virulentas e avirulentas relacionadas à doença de Pierce têm taxas de multiplicação inicial iguais, entretanto após 14 dias do inóculo no xilema há um rápido declínio das populações avirulentas. Cepas avirulentas podem não ser capazes de se moverem para vasos adjacentes aos infectados, visto que o número de vasos infectados por secção do pecíolo é aproximadamente 10 vezes maior com cepas virulentas do que com cepas avirulentas. Células da cepa avirulenta raramente se movem de um pecíolo infectado para a nervura foliar, assim como são incapazes de se moverem para fora da região de entrenós inoculada, em severo contraste com as células da cepa virulenta¹⁵.

A taxa de dispersão de um patógeno vascular no xilema é limitada pelo comprimento do vaso e o tempo requerido para que sejam rompidas as membranas pontuadas que delimitam o mesmo. As membranas pontuadas nos vasos do xilema servem como microfiltros que permitem a passagem livre da seiva bruta no fluxo transpiratório mas que previnem a movimentação de patógenos como bactérias entre os vasos. Tais membranas são constituídas em princípio de parede celular contendo celulose, hemicelulose, pectinas, ligninas e proteínas^{12,13}.

O mecanismo pelo qual *Xylella fastidiosa* consegue se movimentar de vaso para vaso ainda não é conhecido, mas pode envolver digestão enzimática das membranas pontuadas. Foi sugerido que as enzimas bacterianas presentes em baixa concentração poderiam ser conservadas e concentradas nas microcolônias devido à agregação resultante da presença do glicocálix, agindo então sobre as membranas pontuadas. A cepa de *Xylella fastidiosa* causadora da doença de Pierce produz peroxidases e lisofosfolipases em cultura, as últimas com influência sobre membranas celulares¹⁵. Também foi detectada a produção de proteases, sendo em baixa quantidade e sem correlação aparente com a virulência das cepas testadas¹⁶. Há também as possibilidades de que enzimas de degradação da parede celular sejam produzidas e liberadas no xilema pela própria planta hospedeira como resposta à presença da bactéria, ou ainda, que a pressão criada pela bactéria dentro dos vasos do xilema possa levar ao rompimento mecânico das membranas pontuadas, permitindo a movimentação sistemática da bactéria^{12,13}.

Os sintomas causados pela bactéria *Xylella fastidiosa* podem ser agrupados em dois tipos: escaldadura foliar com efeito parcial ou total sobre a planta; ou a redução no tamanho da planta¹¹. Estudos realizados principalmente com relação à doença de Pierce em videiras, à escaldadura foliar em ameixeiras e à doença de má formação de pêssegos mostram que os principais sintomas: necrose foliar marginal, abscisão foliar, crescimento retardado na primavera e declínio do vigor da planta podendo levar à morte, sugerem uma disfunção do sistema de condução de água nos vegetais^{12,13}.

Células bacterianas, tiloses e gomas bloqueiam os vasos do xilema de plantas infectadas por *Xylella fastidiosa*. Foi determinado que a deposição de goma nos vasos é o primeiro sintoma interno e que a formação de tiloses nos vasos também aumenta, ambos ocorrendo antes do desenvolvimento dos sintomas externos da doença de Pierce e da doença de má formação de pêssego^{12,13}.

Cortes de pecíolos e nervuras foliares de videiras afetadas pela doença de Pierce demonstraram que diferentes vasos do xilema estavam bloqueados em diferentes secções, portanto havendo oclusões suficientes no comprimento do pecíolo ou da nervura foliar para produzir um estresse hídrico resultando em necrose foliar marginal. Isso se tornaria mais provável se as gomas nos vasos também interferissem no movimento lateral da água, limitando a possibilidade de desvio da oclusão^{12,13}. Estudos também mostraram que a resistência média à passagem forçada de fluxo no xilema do pecíolo e de um nó seria de 60 a 200 vezes maior em videiras infectadas com a doença de Pierce do que em plantas saudáveis^{12,13}.

Além do estresse hídrico, a doença de Pierce também resulta na parada do crescimento vegetativo, redução na transpiração e fotossíntese, aumento das concentrações de ácido abscísico, frutose, glicose, Ca^{2+} e Mg^{2+} , e diminuição da concentração de K^+ nas folhas afetadas. Muitas das mudanças fisiológicas são também similares a senescência foliar como clorose, altos níveis de prolina e ácido abscísico, e resistência estomatal aumentada, o que levou à conclusão de que os sintomas da doença de Pierce resultam de um estresse hídrico relativamente moderado mas prolongado que acelera a senescência foliar. Entretanto não são eliminados outros mecanismos de estarem relacionados a parte da síndrome da doença^{12,13}.

IB 3. Transmissão

Xylella fastidiosa depende necessariamente de insetos vetores para sua disseminação natural e penetração em tecidos vegetais susceptíveis. A bactéria permanece em maior parte confinada aos elementos traqueiais, traqueídes ou vasos, do xilema de seus hospedeiros, sendo ocasionalmente encontrada em espaços intercelulares. Assim, embora o número de vetores conhecidos seja bastante grande, todos são insetos suctoriais que se alimentam do xilema. Essa gama de vetores inclui membros da subfamília Cicadellinae e membros da família Cercopidae^{12,13}.

Em 1991 foram iniciados estudos sobre a transmissão de *Xylella fastidiosa* entre laranjeiras no campo usando pares de plantas, sendo uma delas protegida do contato com insetos e térmitas. Os resultados mostraram indiretamente que ocorria o envolvimento de vetores na transmissão da CVC. Prováveis vetores foram então estudados, havendo preferência por insetos que se alimentam do xilema. Cigarras das famílias Cicadellidae (pelo menos 19 espécies da subfamília Cicadellinae) foram observadas em citros. Foi confirmado que 3 espécies das mais comuns cigarras: *Acrogonia gracilis*, *Oncometopsis facialis* e *Dilobopteus costalimai*, transmitem a bactéria. Tais espécies são consideradas como os vetores de maior importância devido a suas ocorrências freqüentes em plantas cítricas, sendo originárias de arbustos ao redor de pomares. Novas espécies têm sido identificadas como vetores e outras ainda são consideradas como vetores potenciais¹⁷.

Ainda não há informações detalhadas sobre o mecanismo de transmissão de *Xylella fastidiosa* para citros pelas cigarras, mas é possível que tal mecanismo seja similar ao de outras cepas com pequenas variações como período de latência, eficiência de transmissão, etc. A maior parte dos trabalhos sobre vetores foi realizada com relação à doença de Pierce e a de má formação de pêssago, mas as cigarras são consideradas como prováveis transmissoras da maioria, senão de todas as cepas de *Xylella fastidiosa*^{12,13}.

Provavelmente todas as cepas sobrevivem na natureza e iniciam relações de parasita-hospedeiro através da transmissão para novas plantas hospedeiras por insetos vetores. Os vetores de *Xylella fastidiosa* têm uma gama de hospedeiros vasta, que somada aos seus hábitos sazonais resultam na exposição freqüente de hospedeiros potenciais à bactéria^{12,13}.

A aquisição de *Xylella fastidiosa* pelo inseto vetor ocorre pela alimentação a partir de plantas infectadas. Esses insetos são normalmente atraídos pela cor amarela, sendo portanto atraídos a plantas infectadas que apresentam folhas amareladas devido à doença. O principal vetor de *Xylella fastidiosa* em videiras, *Graphocephala atropunctata*, pode adquirir a bactéria em somente uma hora de alimentação em uma planta infectada, passando a transmiti-la de imediato, não havendo assim período de latência. A eficiência de transmissão aumenta com períodos mais longos de aquisição e inoculação¹⁷.

A infectividade dos adultos de *Graphocephala atropunctata* dura por 60 dias, podendo perdurar até a sua morte. Não há provas da passagem da bactéria às progenes do vetor através dos ovos. As ninfas do vetor também podem transmitir *Xylella fastidiosa* quando infectadas como os adultos, mas perdem sua infectividade após a ecdise¹⁷. São desconhecidas quaisquer anomalias ou doenças relacionadas à presença da bactéria nos dutos de alimentação dos vários estágios de desenvolvimento dos insetos vetores.

Xylella fastidiosa foi observada colonizando a bomba cibarial e o revestimento do esôfago dos insetos. Estudos mostraram que a bactéria se multiplica e permanece no canal de alimentação desses insetos (precibarium) e na câmara de ingestão (cibarium) na região pré-ventral, ocorrendo o acúmulo de bactérias nesses órgãos com o passar do tempo. Tal fato sugere que a multiplicação da bactéria seja propagativa mas não circulativa no vetor. A propagação no vetor explica o fato de um inseto adulto permanecer infectivo durante toda sua vida¹⁷. A força da ação succionante e o fluxo do líquido durante a alimentação do inseto a partir do xilema provavelmente desaloja células bacterianas das paredes dos canais de alimentação que são então egeridas através do canal alimentício nos elementos traqueiais da planta hospedeira, infectando-a^{12,13}.

I C. O projeto Genoma *Xylella fastidiosa*

O projeto Genoma *Xylella fastidiosa*, financiado pela Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) e pelo Fundo de Defesa da Citricultura de São Paulo (FUNDECITRUS), teve como objetivo principal a determinação da seqüência nucleotídica do fitopatógeno causador da Clorose Variegada dos Citros (CVC) a partir do isolado 8.1.b clone 9.a.5.c, do isolado original que cumpriu os postulados de Koch, fornecido pelo Dr. Bové do Institute National de Recherches Agronomiques (INRA) de Boredaux, França.

Estiveram envolvidos no seqüenciamento, montagem e anotação do genoma 33 laboratórios do Estado de São Paulo (32 laboratórios de seqüenciamento e 1 laboratório de bioinformática) organizados através de uma rede virtual denominada ONSA.

As estratégias de seqüenciamento¹⁸ desenvolvidas pela coordenação e pelos laboratórios centrais visaram a maior flexibilidade possível para incorporar uma extensa redundância e diversas estratégias secundárias à estratégia central.

A estratégia central se baseava num mapeamento detalhado do genoma e montagem de contíguos (consensos obtidos a partir da sobreposição de várias seqüências devido a regiões de identidade entre elas) antes do seqüenciamento final a qual seriam incorporados desde o início duas estratégias de seqüenciamento de bibliotecas de “shotgun” (fragmentos genômicos randômicos pequenos clonados em plasmídeos) para maximizar o progresso.

O seqüenciamento de *Xylella fastidiosa*, finalizado em fevereiro de 2000, forneceu uma série de informações sobre o genoma desse fitopatógeno relacionadas no Anexo, página 113. Também será mantida uma página na “Wolrd Wide Web” com informações tanto sobre o genoma como sobre as ORFs (“Open Reading Frames”) anotadas no genoma⁴³.

II. Objetivos

O presente trabalho teve como objetivo principal a maior contribuição possível ao desenvolvimento do projeto genoma de *Xylella fastidiosa* realizado pela rede ONSA.

Mais especificamente foram adotados como objetivos o seqüenciamento de aproximadamente 590.000 bases do genoma de *Xylella fastidiosa*, dentre seqüências obtidas a partir de bibliotecas de “shotgun” genômico e de cosmídeos designados ao nosso laboratório; a anotação no genoma completo do grupo de ORFs classificadas como pertencentes a categoria de metabolismo intermediário e, por fim, uma comparação entre as vias metabólicas relacionadas ao metabolismo de carboidratos identificadas na anotação do genoma de *Xylella fastidiosa* e as descritas para outros organismos com genomas seqüenciados e anotados.

III. Materiais e Métodos

III A. Meios de cultura

Foram utilizados os meios de cultura LB (Luria Broth¹⁹), 2YT¹⁹ e Terrific Broth¹⁹, contendo ampicilina 100 µg/mL (para a seleção de clones contendo plasmídeo pUC18) ou kanamicina 30 µg/mL (para a seleção de clones contendo cosmídeo Lawrist4). Para placas de cultura foi utilizado o meio LB-Agar¹⁹ contendo os mesmos antibióticos quando necessário.

III B. Cepas de bactérias e vetores utilizados

Foram utilizadas as seguintes cepas de *Escherichia coli*:

DH5α (Gibco, BRL): *endA1 recA1 relA1 deoR thi-1 supE44 λ-gyrA96 F⁻ hsdR17(r_K- m_K⁺) φ80dlacZΔM15 Δ(lacZYA-argF)U169*²⁰;

DH10B® (Gibco, BRL): *endA1 recA1 relA1 deoR thi-1 supE44 λ-gyrA96 merA F⁻ Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 Δ(lacZYA-argF)U169*²¹;

XL1-Blue® (Stratagene): *endA1 recA1 relA1 thi-1 supE44 gyrA96 hsdR17 lac[F['] proAB lacI^qZΔM15 Tn10 (Tet^r)]*;

JM109 (Stratagene): *endA1 recA1 thi-1 gyrA96 hsdR17(r_K- m_K⁺) e14-(McrA⁻) Δ(lac-proAB)[F⁻ traD36 proAB lacI^qZΔM15]*.

Foram utilizados os vetores pUC18 (Anexo, página 70) e Lawrist4²².

III C. Seqüenciamento do genoma de *Xylella fastidiosa*

Dentre as diversas estratégias¹⁸, desenhadas pela coordenação, foram adotadas principalmente o seqüenciamento de cosmídeos e de bibliotecas de “shotgun” genômico. Os cosmídeos seriam escolhidos para o seqüenciamento com base num mapa de hibridizações entre todos os disponíveis, esperando-se então minimizar a sobreposição entre cosmídeos e favorecer a formação de contíguos extensos.

O seqüenciamento de bibliotecas de “shotgun” serviria para a confirmação tanto da montagem dos contíguos como determinar a seqüência genômica em regiões que não fossem clonadas em cosmídeos, ou cujos clones não fossem determinados (“gaps”).

A biblioteca de cosmídeos de *Xylella fastidiosa* foi feita na Alemanha pela Dra. Anete Pereira Souza no laboratório do Dr. Hoheizel em Heidelberg a partir de DNA total de *Xylella fastidiosa* extraído de uma cultura realizada no Laboratório de Genética de Câncer do Instituto Ludwig de Pesquisa sobre o Câncer. A biblioteca foi feita através de digestão parcial do DNA total com a enzima de restrição *MboI* e clonagem dos fragmentos no sítio de *BamHI* do vetor Lawrist4²².

Foi realizada a validação da biblioteca através de hibridização de Southern “blots” de digestões de cosmídeos com DNA total de *Xylella fastidiosa* marcado utilizando a técnica de “random primers”. Também foi realizada uma comparação entre os DNAs extraídos de cultura realizada no Instituto Ludwig e de outra realizada no INRA através da técnica de RAPD. Os padrões de amplificação obtidos com 10 primers arbitrários diferentes indicaram que os DNA eram praticamente idênticos e que todas as regiões amplificadas estavam presentes em ambos os isolados.

Foram feitas diversas bibliotecas de “shotgun” genômico de *Xylella fastidiosa*. Algumas delas são descritas na Tabela I.

Tabela I: Bibliotecas de “shotgun” genômico de *Xylella fastidiosa*.

Biblioteca	Vetor	Inserto	Grupo responsável
05	pUC18 / sítio <i>SmaI</i>	Fragmentos de DNA nebulizado	Dr. Andrew Simpson LICR-São Paulo
07	pUC18 / sítio <i>SmaI</i>	Fragmentos de DNA nebulizado	Prof. Dr. Jesus Ferro UNESP-Jaboticabal
08	pUC18 / sítio <i>SmaI</i>	Fragmentos de 1500 a 2500 pb de DNA total nebulizado	Prof. Dr. Jesus Ferro UNESP-Jaboticabal
10	pUC18 / sítio <i>SmaI</i>	Fragmentos de 2000 a 4000 pb de DNA total nebulizado	Prof. Dr. Jesus Ferro UNESP-Jaboticabal
11	m13 mp19 / sítio <i>SmaI</i>	Fragmentos de 1500 a 3000 pb de DNA total nebulizado	Prof. Dr. Fernando Reinach USP-São Paulo
12	m13 mp19 / sítio <i>SmaI</i>	Fragmentos de 500 a 1500 pb de DNA total nebulizado	Prof. Dr. Fernando Reinach USP-São Paulo
13	Janus / <i>SmaI</i>	Fragmentos 1500 a 3000 pb de DNA total nebulizado	Prof. Dr. Fernando Reinach USP-São Paulo
14	Janus / <i>SmaI</i>	Fragmentos 500 a 1500 pb de DNA total nebulizado	Prof. Dr. Fernando Reinach USP-São Paulo
15	pUC18 / sítio <i>SmaI</i>	Fragmentos de DNA nebulizado	Dr. Andrew Simpson LICR-São Paulo
18	pUC18 / sítio <i>SmaI</i>	Continuação da biblioteca 08	Prof. Dr. Jesus Ferro UNESP-Jaboticabal

III D. Obtenção de bibliotecas de “shotgun” a partir de cosmídeos

Para o seqüenciamento de cosmídeos nosso laboratório recebeu da coordenação uma placa de cultura contendo um espalhamento de parte do estoque de um determinado cosmídeo em meio seletivo YT/kanamicina 30 µg/mL. Foi realizada a verificação da inexistência de contaminação com outro cosmídeo através de análise do padrão de digestão por enzimas de restrição (normalmente *EcoRI* ou *HindIII*), do DNA cosmidial extraído de 5 colônias isoladas. Não havendo contaminação, foram preparados estoques em glicerol (40% v/v final) a partir de uma das 5 pré-culturas que foi utilizada como inóculo para a preparação em grande quantidade do DNA do cosmídeo.

A obtenção do DNA de cosmídeos em grande quantidade foi feita por método tradicional¹⁹. O DNA resultante foi purificado por ultracentrifugação em gradiente de densidade de cloreto de cério¹⁹.

Realizava-se então fragmentação mecânica do DNA do cosmídeo (100 µg) por nebulização. Os fragmentos de DNA de 1,5 kb a 2,5kb foram selecionados por eletroforese em gel de agarose preparativa¹⁹. Após a extração do DNA da agarose, foi realizado tratamento com fragmento Klenow da DNA polimerase I e a T4 polinucleotídeo quinase para a obtenção de fragmentos de DNA com extremidades cegas fosforiladas para a ligação no sítio de *SmaI* do plasmídeo pUC18 linearizado (kit SureClone[®], Amersham Pharmacia).

Após a ligação, bactérias competentes foram transfectadas, através de eletroporação (BioRad Gene Pulser®, campo de 2,5 V, utilizando cubetas de eletroporação BioRad Gene Pulser® Cuvette de 0,2 cm), seguindo as recomendações do aparelho, ou transfecção com cloreto de cálcio²³. Foram selecionados clones contendo plasmídeos com inserto através da expressão de *lacZ*, após plaqueamento em meio sólido LB-Agar contendo ampicilina 100 µg/mL, isopropil-β-tiogalactopiranosídeo (IPTG) e 5'-bromo-4-cloro-3-indolil-galactopiranosídeo (X-gal)^{19,24}.

As bibliotecas de “shotgun” dos cosmídeos seqüenciados em nosso laboratório foram preparadas em conjunto (X0QH-06A08 e X0QH/QS-02G12), ou inteiramente (X0QH-07A10, X0QH-02E06, X0QH/QS-05D08/15D08), pela técnica do laboratório do Prof. Dr. Roberto Santelli, Luci Navarro-Cattapan.

III E. Preparação de DNA de plasmídeos recombinantes

A obtenção de DNA de plasmídeos derivados da subclonagem de cosmídeos ou de bibliotecas de “shotgun” de DNA genômico de *Xylella fastidiosa* foi realizada através de diversos tipos de minipreparação:

Lise alcalina¹⁹ modificada pelo Prof. Dr. Marcelo Bento Soares (University of Iowa, EUA), para preparação a partir das cepas DH5 α e XL1-Blue® de *Escherichia coli*;

Concert® Miniprep kit (Gibco, BRL), para preparação a partir de todas as cepas de *Escherichia coli*;

“Boiling prep” com a utilização de microondas desenvolvida pelo Prof. Dr. Marcelo Bento Soares (University of Iowa, EUA), para preparação a partir da cepa DH10B® de *Escherichia coli*, descrita a seguir.

Realizar inóculo de colônias isoladas em placas de acrílico de 96 poços para realização de miniprep em microondas (Sharp) contendo 1,0 mL de meio de cultura contendo ampicilina 100 $\mu\text{g/mL}$ em cada poço (foram utilizados os meios de cultura 2YT e Terrific Broth). Deixar a placa sob agitação de 300 rpm a 37°C para o crescimento das bactérias.

Ao fim do período de crescimento centrifugar a placa para a sedimentação das bactérias por 8 minutos a 2.700 rpm (centrífuga Sigma 4K15C Qiagen, rotor Qiagen 09100/09366). Desprezar o sobrenadante e drenar com a placa invertida sobre papel toalha por 5 minutos. Adicionar então 25 μL de água desionizada estéril a cada poço e ressuspender os sedimentados por agitação em vórtex por 4 a 6 minutos. Adicionar 70 μL de solução MWTween 20 e agitar em vórtex por 15 segundos. Deixar a placa em repouso a temperatura ambiente por 5 minutos.

Realizar 2 ou 3 sucessivos aquecimentos da placa em microondas com prato giratório na potência máxima, parando-os logo que houver fervura e invertendo a posição da placa entre um aquecimento e outro. Imediatamente após o último aquecimento, adicionar 300 μL de água desionizada estéril a cada poço, agitar em vórtex por 15 segundos e incubar em banho de gelo por 10 minutos. Centrifugar então por 30 minutos a 4000 rpm (centrífuga Sigma 4K15C Qiagen, rotor Qiagen 09100/09366).

Ao fim da centrifugação transferir de 50 a 100 μ L do sobrenadante, com o cuidado de não ressuspender o sedimentado, para placas de estoque de 96 poços com fundo em U, para estocar o DNA em freezer a -20°C ou -80°C .

Para um volume aproximado de 100 mL de solução MWTween 20 são utilizados 100 mL de solução STET/Tween 20, 2,4 mL de suspensão de RNase 10mg/mL (fervida por 10 minutos) e 50 mg de Lisozima. Após o preparo, no mesmo dia em que será utilizada, a solução MWTween 20 deve ser imediatamente guardada em geladeira até o momento do uso e posteriormente descartada.

Para um volume de 100 mL de solução STET/Tween 20 são utilizados 5,0 mL de Tween 20; 2,0 mL de NaCl 5,0 mol/L; 1,6 mL de Tris-HCl 1,0 mol/L pH 8,0; 200 μ L de EDTA 0,5 mol/L pH 8,0; 91,2 mL de água desionizada estéril. Após o preparo, a solução deve ser filtrada através de membrana 0,22 μm (Millex 0,22 μm , Milipore) e pode ser guardada em geladeira.

III F. Seqüenciamento de insertos em plasmídeos pUC18

A reação de seqüenciamento cíclico utilizada no projeto é baseada no método de extensão de primer terminada por didesoxinucleotídeos trifosfatos (ddNTPs), método de Sanger²⁵, consistindo de etapas sucessivas de desnaturação do molde (plasmídeo onde está inserido o fragmento a ser seqüenciado), pareamento do primer a ser estendido e extensão do primer pela ação de uma DNA polimerase termoestável, utilizando-se para isso um termociclador, resultando numa amplificação linear dos produtos de extensão.

Obtidos os plasmídeos contendo o DNA a ser seqüenciado foram realizadas as reações de seqüenciamento cíclico utilizando kit de ddNTPs covalentemente ligados a agentes fluorescentes específicos para cada ddNTP: ABI Prism[®] Big Dye[™] Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq[®] DNA Polymerase, FS (Applied Biosystems, Perkin-Elmer) e primer universal (M13 -21) ou reverso (M13 Reverse). Os termocicladores utilizados para o desenvolvimento das reações de seqüenciamento cíclico foram RoboCycler (Stratagene) e GeneAmp PCR System 9700 (Applied Biosystems, Perkin-Elmer). As condições da reação de extensão utilizadas em nosso laboratório, aprimoradas pelo Prof. Dr. Arthur Gruber (Faculdade de Medicina Veterinária e Zootecnia, USP), foram as seguintes:

Mistura reacional:

1,0 μL BigDye™ premix;

3,0 μL de 2,5x Buffer (200 mmol/L Tris-HCl pH 9,0; 5,0 mmol/L MgCl_2);

3,0 μL de suspensão de DNA molde (plasmídeo; 200 a 300 ng);

3,0 μL de primer (M13 -21 ou Reverse; 3,2 pmol totais);

Condições de extensão:

1 ciclo de 1 minuto a 96°C;

40 ciclos das etapas: 20 segundos a 96°C, 20 segundos a 52°C, 4 minutos a 60°C.

A purificação dos produtos das reações de seqüenciamento foi realizada através de precipitação com etanol ou isopropanol¹⁹. As amostra purificadas e concentradas foram suspensas em 2,0 μL de tampão de aplicação (1 parte de Blue Dextran em solução de EDTA 25 mmol/L para 5 partes de formamida desionizada) e aplicadas (1,0 μl) em gel desnaturante de seqüenciamento (36,0 cm de comprimento), sendo a eletroforese desenvolvida em seqüenciador automático ABI 377 DNA Sequencer (Applied Biosystems, Perkin-Elmer).

O géis de seqüenciamento foram preparados utilizando-se a mistura concentrada de acrilamida/bisacrilamida Long Ranger™ 50% (FMC). Os géis tinham composição final de acrilamida 5% (% m/v), tampão TBE¹⁹ 1% (% v/v) e uréia 0,36 g/mL (5,0 mol/L).

Para a obtenção dos dados de seqüenciamento no seqüenciador automático foram utilizados os módulos de programação do seqüenciador para eletroforeses desenvolvidas em gel de 36,0 cm, química de seqüenciamento BigDye™, tanto a 1200 V para 7 horas de aquisição de dados (Seq Run 36E-1200) quanto a 2400 V para 3,5 horas de aquisição de dados (Seq Run 36E-2400).

Os dados do seqüenciamento automático foram coletados por programa dedicado ABI Prism Data Collection (Applied Biosystems, Perkin-Elmer) em um microcomputador PowerMacintosh 7300/200 (Apple Microcomputer, Inc.). As análises dos dados do géis de seqüenciamento foram realizadas pelo programa Sequencing Analysis v3.3 (Applied Biosystems, Perkin-Elmer) em um computador PowerMacintosh G3 (Apple Microcomputer, Inc.), que forneceu para cada amostra seqüenciada um cromatograma. Após a análise visual dos cromatogramas para descarte daqueles de baixa qualidade, os restantes foram submetidas ao laboratório de bioinformática.

III G. Montagem de cosmídeos a partir de leituras de bibliotecas de “shotgun”

Os cromatogramas considerados de boa qualidade que faziam parte de uma biblioteca de “shotgun” de um cosmídeo foram utilizados para a montagem da seqüência do inserto do cosmídeo em questão. Inicialmente foi utilizado o programa SenquencherTM instalado em um PowerMacintosh G3 (Apple Microcomputer, Inc.) sob sistema MacOS 8.0 (Apple Microcomputer, Inc.), principalmente na montagem do primeiro cosmídeo seqüenciado. Visto às dificuldades encontradas e aos problemas que o programa apresentou com a grande quantidade de dados utilizada para a montagem de seqüências de DNA ao redor de 40 quilobases, passou a ser utilizado o pacote de programas de bioinformática phred/phrap/crossmatch^{26,27} e consed²⁸, instalados em estação Sun (Sun Microsystems) sob sistema UNIX, do Departamento de Bioquímica da USP, no laboratório do Prof. Dr. Sérgio Verjovski-Almeida (IQ-USP).

A montagem da seqüência do inserto de um cosmídeo consiste na junção ordenada dos cromatogramas dos plasmídeos seqüenciados baseada na similaridade entre as seqüências e critérios de qualidade de seqüenciamento para a formação de contíguos (contigs) de seqüências. A qualidade do seqüenciamento é dada para cada base de uma seqüência através da análise dos cromatogramas pelo programa phred. A comparação das seqüências, já com qualidade determinada para cada base, e a montagem de contíguos de seqüências são realizadas pelos programas crossmatch e phrap. O programa consed é utilizado para a visualização gráfica dos arquivos de montagem de contíguos e determinação de regiões das seqüências dos contíguos com qualidade baixa, seqüenciamento em uma única fita ou discrepâncias de alta qualidade.

Durante a montagem, até a obtenção de um único contíguo representando o inserto do cosmídeo sendo seqüenciado, foram utilizados cromatogramas referentes somente a clones de bibliotecas provenientes desse cosmídeo. Após tal estágio, foram utilizados cromatogramas provenientes do seqüenciamento de clones de bibliotecas de cosmídeos que se sobrepunham ao que estava sendo seqüenciado para a melhoria da qualidade total do seqüenciamento do inserto.

Quando foram obtidos contíguos de tamanho condizente com o do inserto do cosmídeo, o arquivo de montagem do inserto do cosmídeo foi submetida ao centro de bioinformática para validação e checagem de discrepâncias quanto aos critérios de qualidade estabelecidos para o seqüenciamento de insertos de cosmídeo. Tais critérios, determinados pelo coordenação e laboratório de bioinformática, foram:

- (1) os cromatogramas de todas as seqüências usadas na montagem do inserto deveriam estar depositados no laboratório de bioinformática;
- (2) o programa phrap deveria ser capaz de montar o inserto do cosmídeo usando somente as seqüências referidas em (1), tal montagem sendo realizada no laboratório de bioinformática;
- (3) a montagem referida em (2) deveria ter as seguintes características:
 - a) quando realizada sem varredura contra Lawrist4, as seqüências contendo regiões do vetor deveriam aparecer somente nas extremidades da montagem, flanqueando o inserto;
 - b) quando realizada com varredura contra Lawrist4, a montagem deveria ter as seguintes características:
 - b-abi-1) taxa de erro esperada para a montagem menor ou igual a 1 em cada 10.000 bases (0,01%);
 - b-abi-2) cada base (ou espaço) do consenso da montagem deveria ser confirmado por pelo menos uma seqüência em cada fita;
 - b-abi-3) a seqüência consenso da montagem não deveria ter regiões com qualidade menor ou igual a 20, de acordo com o programa phred;
 - b-abi-4) em cada seqüência não deveriam haver bases com qualidade igual ou superior a 40, de acordo com o programa phred, que discordassem da posição correspondente na seqüência consenso;
 - c) todas as “frameshifts” em ORFs deveriam ser checadas e corrigidas, ou validadas.

Estando a montagem de acordo com tais critérios, o cosmídeo em questão era considerado terminado (“finished”) e cessavam os trabalhos de seqüenciamento do inserto do cosmídeo em questão. Em caso de ainda haver discrepâncias quanto aos critérios de qualidade, o cosmídeo era classificado como: um contíguo (“one contig”), no caso de muitas discrepâncias; ou quase terminado (“close to finished”), no caso de um número pequeno de discrepâncias.

Para a solução das discrepâncias encontradas na checagem do laboratório de bioinformática foram utilizadas as estratégias de seqüenciamento direcionado e de seqüenciamento com primers específicos utilizando como molde plasmídeos da biblioteca de “shotgun” do cosmídeo, ou o próprio cosmídeo.

Com o intuito de confirmar experimentalmente que a seqüência do inserto obtida realmente era a do cosmídeo seqüenciado, foi realizada uma digestão computacional da seqüência do inserto devidamente inserida no vetor Lawrist4 (a orientação do inserto no vetor é dada pelas seqüências que o flanqueiam) e o resultado comparado com a digestão inicialmente utilizada para determinar se havia ou não contaminação na placa do cosmídeo.

III H. Seqüenciamento total de inserto de plasmídeo de “shotgun” genômico

Na parte final do projeto de seqüenciamento do genoma de *Xylella fastidiosa*, foram identificadas várias regiões que não haviam sido clonadas nos cosmídeos. Como passara-se a seqüenciar também bibliotecas de “shotgun” genômico de *Xylella fastidiosa*, foram identificados muitas vezes clones em plasmídeos cujas seqüências das extremidades eram similares às extremidades de contíguos de cosmídeos (seqüências resultantes de sobreposições das seqüências dos insertos dos cosmídeos). Para validar o seqüenciamento de tais regiões, os insertos de tais plasmídeos foram seqüenciados inteiramente como se fossem insertos de cosmídeos, sendo denominados então seqüências de preenchimento de “gaps”, ou simplesmente GFSs (“Gap-filling Sequences”).

A metodologia que se mostrou mais adequada em termos de rapidez e eficiência foi o preparo de biblioteca de “shotgun” a partir do próprio plasmídeo, cujo tamanho normalmente estava na faixa de 5 quilobases (2 quilobases do plasmídeo mais inserto de aproximadamente 3 quilobases). Devido à grande quantidade necessária para o sucesso de uma nebulização, utilizou-se como método de clivagem mecânica a sonicação.

Uma quantidade de aproximadamente 100 ng de plasmídeo suspenso em 10,0 µL de água em tubos eppendorf de 100,0 µL era submetida à sonicação em um sonicador “de copo” (“cup sonicator”), disponibilizado pelo Prof. Dr. Paulo Arruda (CBMEG –UNICAMP). Utilizaram-se pulsos de 45 segundos. Seguiram-se então os protocolos de subclonagem dos fragmentos em pUC18, seqüenciamento e montagem do inserto idênticos aos utilizados para as bibliotecas de “shotgun” de cosmídeos.

III I. Determinação de ORFs nos insertos de cosmídeos

Finalizado o seqüenciamento, as seqüências consenso do insertos do cosmídeos foram depositadas no banco de dados do laboratório de bioinformática, onde foram examinadas pelo programa Glimmer²⁹ para a determinação de ORFs, sendo então disponibilizada para que fosse realizada a anotação das ORFs encontradas.

O Glimmer faz uma busca em seqüências de DNA por regiões com alta probabilidade estatística de codificarem proteínas, baseado num treinamento prévio, determinando então fases abertas de leitura (“open reading frames”, ou simplesmente ORFs) que se iniciem pelos códons: ATG, GTG e TTG. Para a análise dos insertos de cosmídeos o programa foi treinado com seqüências de DNA da própria *Xylella fastidiosa*, para a obtenção do padrão estatístico de utilização de códons em regiões codificantes. Após a análise das seqüências dos insertos de cosmídeo pelo Glimmer, foram criados arquivos para o programa Sequin³⁰ contendo as posições das ORFs nos insertos dos cosmídeos e arquivos contendo a tradução em fase de leitura correta de cada ORF de um determinado cosmídeo. Esses arquivos foram disponibilizados pelo laboratório de bioinformática aos demais grupos.

III J. Anotação de ORFs

A anotação de uma ORF consiste na identificação da ORF como suposto gene que codifique certa proteína. Essa identificação baseia-se na similaridade que a seqüência protéica obtida da tradução da ORF tem como uma seqüência de uma proteína conhecida, presente em um banco de dados.

A busca de similaridades entre seqüências protéicas foi realizada através do serviço BLASTP (v. 2.0³¹), utilizando a matriz de comparação BLOSUM 62, no banco de dados público não redundante disponibilizado pelo National Center for Biotechnology Information (NCBI³²).

A anotação das ORFs presentes nas seqüências consenso dos insertos dos cosmídeos foi realizada através do programa Sequin, sendo possíveis as seguintes descrições de uma ORF, como determinado pela coordenação, de acordo com o resultado da busca em bancos de dados:

1. Putative Xase, ou Putative X protein: para as ORFs cujas seqüências eram similares a proteínas ou genes de função conhecida presentes nos bancos de dados, sendo a probabilidade da similaridade ser ao acaso fortuito (E-value do BLAST) menor do que 10^{-5} (menos que 1 em 10.000);
2. Conserved Hypothetical X kDa protein: para as ORFs cujas seqüências eram similares a proteínas ou genes de função desconhecida (hipotéticos) presentes nos bancos de dados, sendo a probabilidade da similaridade ser ao acaso fortuito (E-value do BLAST) menor do que 10^{-5} (menos que 1 em 10.000);
3. Hypothetical X kDa protein: para as ORFs cujas seqüências não eram similares a proteínas ou genes presentes nos bancos de dados, ou se a probabilidade da similaridade a proteínas e genes presentes nos bancos de dados ser ao acaso fortuito (E-value do BLAST) for maior ou igual a 10^{-3} (mais que 1 em 1.000).

4. Nos casos em que a probabilidade da similaridade entre a seqüência da ORF anotada e uma proteína, ou gene, dos bancos de dados fosse ao acaso fortuito (E-value do BLAST) estar na faixa de 10^{-5} a 10^{-3} , era realizada uma verificação utilizando-se o PSI-BLAST³³ e uma busca dentro do banco de dados de famílias de proteínas e de domínios em proteínas (Pfam e HMM-PROSITE). A ORF poderia então ser anotada de três formas diferentes de acordo com o resultado de tais buscas:

- a) se a busca no PSI-BLAST convergisse para uma determinada proteína ou gene até a quinta iteração do programa então ela era anotada como Putative X (X sendo o nome da enzima ou proteína a qual era mais similar), mas com a probabilidade de similaridade do BLAST e menção a busca no PSI-BLAST,
- b) se não houvesse convergência no PSI-BLAST, mas a proteína se enquadrasse numa família de proteínas por apresentar todos os domínios protéicos relacionados a essa família, então ela era anotada como Putative Y (Y sendo o nome da família protéica),
- c) se não houvesse convergência ou a seqüência traduzida da ORF não se enquadrasse diretamente em nenhuma família protéica, então a ORF era anotada como Conserved Hypothetical X kDa protein e faz-se menção às buscas realizadas.

Além de determinar com quais proteínas já conhecidas uma determinada ORF era similar, na anotação foram incluídas informações, quando existentes, a respeito da função da proteína similar, da reação catalisada, da necessidade de cofatores, etc. Também incluiu-se o código da proteína mais similar em um banco de dados e qual a probabilidade da semelhança entre as duas proteínas ser ao acaso fortuito (E-value do BLAST).

Cada ORF anotada foi inicialmente classificada de acordo com a classificação existente para os genes de *Escherichia coli*³⁴. Essa classificação se baseia na rota metabólica em que a proteína, codificada pelo gene em questão, participa. O metabolismo de *Escherichia coli* é dividido em sete categorias principais: I. Metabolismo Intermediário; II. Biossíntese de Pequenas Moléculas; III. Metabolismo de Macromoléculas; IV. Estrutura celular; V. Processos Celulares; VI. Outras funções; VII. Genes hipotéticos ou não classificados.

Durante o processo de anotação das ORFs dos insertos dos cosmídeos também foram realizadas buscas por tRNAs. Utilizou-se o programa tRNAscan-SE (v. 1.1³⁵), que analisa a seqüência do inserto do cosmídeo e procura por seqüências similares a tRNAs bacterianos. No caso de encontrá-la, o programa realiza a verificação da possibilidade da seqüência formar a estrutura secundária em trevo característica dos tRNAs. Somente no caso em que ambas as condições (seqüência similar e provável estrutura secundária) fossem satisfeitas é que a região determinada foi anotada como uma seqüência codificando um tRNA.

Após a anotação das ORFs e tRNAs de um inserto de um cosmídeo no programa Sequin, o arquivo gerado foi submetido ao laboratório de bioinformática.

III K. Busca de vias metabólicas e anotação de categorias de genes

Como parte da anotação do genoma de *Xylella fastidiosa*, foi realizada uma distribuição de categorias de genes a diversos pesquisadores, baseadas inicialmente nas categorias existentes para *Escherichia coli*.

Essa distribuição teve como objetivo verificar quais as rotas metabólicas tinham sido identificadas no genoma e fazer uma busca mais minuciosa por ORFs que codificassem proteínas que tivessem papel importante no metabolismo e que ainda não haviam sido anotadas.

Como houve muitas regiões do genoma de *Xylella fastidiosa* que acabaram sendo seqüenciadas através de bibliotecas de “shotgun” do DNA genômico da bactéria ou a partir de seqüenciamento de insertos de uma biblioteca de *Xylella fastidiosa* em fagos lambda, não puderam ser disponibilizados arquivos Sequin que contivessem as ORFs dessas regiões. Para realizar a busca de ORFs não anotadas utilizou-se então uma metodologia diferente da descrita para a anotação de cosmídeos.

Sem ORFs definidas e, portanto, sem seqüências protéicas nas regiões de “shotgun” genômico a metodologia adotada foi a de comparar seqüências das proteínas a serem encontradas diretamente com seqüências nucleotídicas, utilizando para isso o programa TBLASTN³¹, rotina derivada do BLASTP.

O TBLASTN realiza uma tradução reversa de uma dada seqüência protéica utilizando um código genético preferencial e compara, nos dois sentidos, o resultado com uma seqüência nucleotídica, e sua complementar, de um banco. No caso em questão, o código genético foi o bacteriano e os bancos de dados utilizados foram os de contíguos genômicos (formados por vários insertos de cosmídeos com sobreposição e/ou insertos de cosmídeos ligados por seqüências de “shotgun” genômico), de somente seqüências de “shotgun” genômico e de todas as seqüências do projeto Genoma *Xylella fastidiosa*, nessa ordem de preferência. Esperou-se assim cobrir todo o possível universo de seqüências produzidas tentando-se identificar as proteínas faltantes.

Para cada caso iniciou-se a análise através do TBLASTN a partir da proteína em questão presente em *Escherichia coli*. Não sendo detectada nenhuma semelhança nos três bancos, utilizou-se então as proteínas de outros organismos. No caso de similaridade encontrada no banco de contíguos genômicos, a região foi anotada. No caso de similaridade nos bancos de seqüências de “shotgun” genômico e de todas as seqüências do projeto foi realizada uma verificação para assegurar de que não se tratava de contaminação.

Essa precaução deveu-se ao fato de o inserto dos cosmídeos ser varrido por um programa de busca de seqüências de *Escherichia coli* no momento de submissão da montagem ao laboratório de informática, assegurando assim a inexistência de contaminação devido aos processos de subclonagem.

Entretanto, tal varredura não foi feita nas submissões dos cromatogramas dos plasmídeos, das bibliotecas de “shotgun” genômico ou de cosmídeos, submetidos ao laboratório de bioinformática. Assim, esses bancos de dados continham certa quantidade de seqüências contaminantes provenientes de *Escherichia coli* devido aos processos de clonagem e subclonagem anteriores ao seqüenciamento dos plasmídeos. Portanto, no caso da similaridade entre a proteína de *Escherichia coli* e uma seqüência de tais bancos ser maior do que 98% o resultado foi descartado por ser considerado como contaminação. No caso de não haver similaridade da proteína de *Escherichia coli* mas sim da proteína de outro organismo com tais bancos, o resultado foi considerado positivo.

Perto do fim do seqüenciamento do genoma de *Xylella fastidiosa*, quando os contíguos genômicos já contavam com várias centenas de quilobases, o laboratório de bioinformática realizou análises com o Glimmer sobre tais contíguos, gerando um banco de dados de ORFs de *Xylella fastidiosa*. Foram então anotadas as ORFs encontradas pelas buscas utilizando o TBLASTN.

A busca das vias metabólicas se baseou principalmente nas vias descritas para *Escherichia coli* e outras eubactérias, presentes em diversos bancos de dados públicos. Foram usados principalmente os bancos de vias metabólicas Kyoto Encyclopedia of Genes and Genomes (KEGG³⁶) e EcoCyc/MetaCyc³⁷, além do Genbank³⁸ e as ferramentas presentes nele, principalmente o “Clusters of Orthologous Groups of proteins” (COG³⁹).

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo

IV. Resultados

Foram seqüenciados três cosmídeos em nosso laboratório (QH): X0QH-07A10, X0QH-02E06 e X0QH-06A08; e outros dois em colaboração com o grupo coordenado pela Profa. Dra. Suely Gomes (IQ-USP São Paulo, QS): X0QH/QS-02G12 e X0QH/QS-05D08). Também foram seqüenciados clones das bibliotecas de “shotgun” genômico 07, 08, 10 e 15. Por fim foi realizado o seqüenciamento completo de um inserto em plasmídeo da biblioteca 10 de “shotgun” genômico (X0QH1013G02), para o fechamento de um “gap” entre contíguos genômicos, denominado GFS (“gap-filling sequence”) X0QH-00J42. No mapa de disposição de clones seqüenciados (Anexo, página 71), os clones seqüenciados em nosso laboratório estão marcados em vermelho.

Foi realizada a anotação dos cosmídeos seqüenciados em nosso laboratório, com exceção ao 05D08, e de regiões de cosmídeos seqüenciados por outros grupos. Foram realizadas verificações de anotações na Categoria I, Metabolismo Intermediário, assim como buscas às ORFs que pudessem codificar proteínas importantes dentro dessa categoria que não estavam presentes no banco de dados de ORFs de *Xylella fastidiosa*.

IV A. Seqüenciamento e anotação de insertos de cosmídeos

IV A 1. Cosmídeo X0QH-07A10

O primeiro cosmídeos seqüenciado foi o 07A10, entre a metade de julho de 1998 e março de 1999, em um período de aproximadamente 9 meses. Foram realizadas duas bibliotecas de plasmídeos do 07A10, denominadas K e L. A biblioteca L do 07A10 foi a primeira a ser realizada e, ao contrário das demais, foi a única a utilizar digestão parcial do cosmídeo com a nuclease Mug Bean para a obtenção de fragmentos a serem clonados o sítio *Sma*I do vetor pUC18. A partir da biblioteca K foi utilizada a nebulização para a subclonagem dos cosmídeos.

A montagem final do inserto compreendeu 939 seqüências, numa extensão de 42.406 pb, contendo 593 seqüências provindas de plasmídeos das bibliotecas K e L do cosmídeo 07A10, 197 seqüências de bibliotecas do cosmídeo 02E06 e 149 seqüências provenientes de bibliotecas de “shotgun” genômico. A taxa de erro final do inserto do cosmídeo foi determinada como 0,05 / 10.000 pb. Os relatórios de estágio e de término de seqüenciamento estão presentes no Anexo (página 101).

Uma vez que a digestão inicial de verificação de ausência de contaminação na placa do cosmídeo foi realizada com a enzima de restrição *EcoRI*, foi realizada a mesma digestão eletrônica do inserto no vetor Lawrist4. Foram obtidas bandas de tamanho similar, confirmando que o seqüenciamento e a montagem do inserto foram corretos (Anexo, página 102).

A análise da seqüência do inserto do cosmídeo 07A10, realizada pelo programa GLIMMER no laboratório de bioinformática, determinou 63 ORFs das quais foram analisadas para anotação 38, na região do inserto entre as bases 10.186 e 42.405, de acordo com a metodologia descrita na seção de Materiais e Métodos. As ORFs do cosmídeo 07A10 são mostradas na Figura 6 e a anotação correspondente é mostrada na Tabela II.

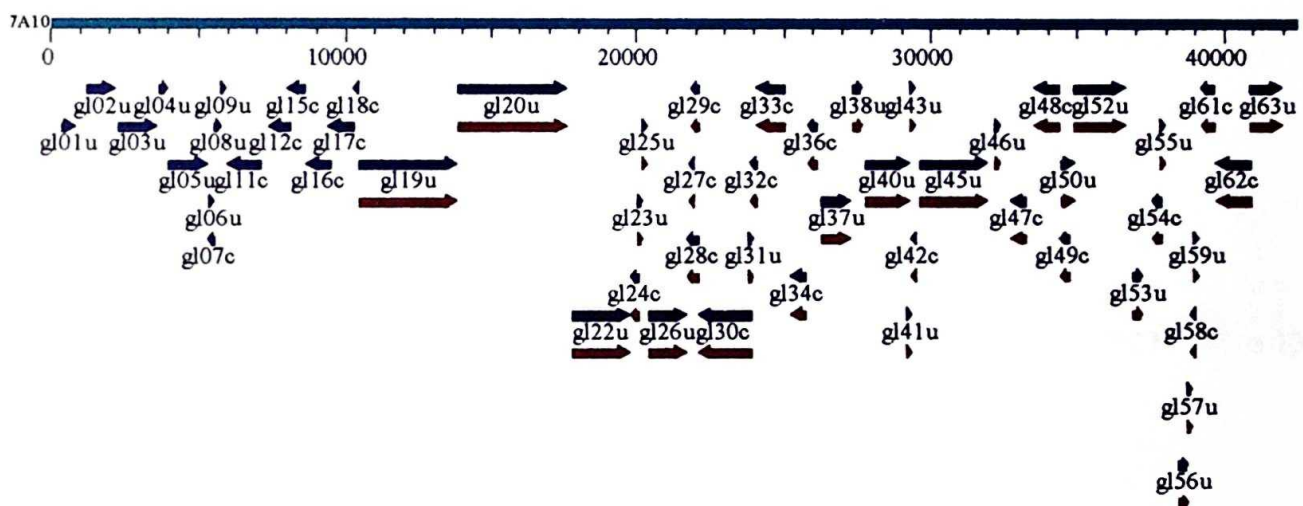


Figura 6: Anotação das ORFs do cosmídeo X0QH-07A10. São mostradas em azul as ORFs encontradas pelo programa Glimmer na seqüência do cosmídeo 07A10. Estão em vermelho as ORFs anotadas para esse cosmídeo, lembrando que as demais ORFs não foram anotadas uma vez que fazem parte de região de sobreposição a outro cosmídeo já anotado, no caso o 02E06. As ORFs estão denominadas pelo código gl#c(u), onde gl significa uma ORF Glimmer, seguido do número da ORF no cosmídeo e da letra u, no caso do sentido da ORF ser o mesmo da seqüência do cosmídeo (uncomplemented), ou da letra c, no caso da ORF estar no sentido inverso ao da seqüência dos cosmídeo, ou seja na seqüência complementar (complemented).

Tabela II: ORFs anotadas no cosmídeo X0QH-07A10. São mostrados os código da ORF e as informações referentes ao resultado de maior similaridade da busca realizada através do BLASTP: o nome da proteína, o organismo a que pertence, seu código no GenBank (gi) ou no SwissProt (sp) e a probabilidade esperada para a similaridade ser ao acaso (E-value).

ORF	Proteína	Organismo	Código # (Accession #)	E-value (BLASTP)
XF-07A10-gl19	Exodesoxirribonuclease V, cadeia gama	<i>Escherichia coli</i>	sp P07648	6e-89
XF-07A10-gl20	Exodesoxirribonuclease V, cadeia beta	<i>Escherichia coli</i>	sp P08394	2e-95
XF-07A10-gl22	Exodesoxirribonuclease V, cadeia alfa	<i>Escherichia coli</i>	sp P04993	3e-58
XF-07A10-gl23	Proteína hipotética de 6.50 kDa			
XF-07A10-gl24	Proteína hipotética de 6.10 kDa			
XF-07A10-gl25	Proteína hipotética de 5.22 kDa			
XF-07A10-gl26	Tryptophanyl-tRNA sintetase	<i>Thermotoga maritima</i>	gi 4981003	7e-41
XF-07A10-gl27	Proteína hipotética de 5.09 kDa			
XF-07A10-gl28	Proteína hipotética de 10.67 kDa			
XF-07A10-gl29	Proteína hipotética de 6.96 kDa			
XF-07A10-gl30	DNA Primase	<i>Escherichia coli</i>	sp P02923	e-131
XF-07A10-gl31	Proteína hipotética de 6.24 kDa			
XF-07A10-gl32	Proteína hipotética de 7.69 kDa			
XF-07A10-gl33	Proteína BrkB	<i>Deinococcus radiodurans</i>	gi 6460437	3e-27
XF-07A10-gl34	Proteína hipotética conservada de 16.9 kDa	<i>Legionella pneumophila</i>	gi 1575483	2e-26
XF-07A10-gl36	Proteína S21 da subunidade ribossomal 30S	<i>Escherichia coli</i>	sp P02379	9e-17
XF-07A10-gl37	O-sialoglicoproteína endopeptidase	<i>Escherichia coli</i>	gi 882587	1e-102
XF-07A10-gl38	Diidroneopterina aldolase	<i>Escherichia coli</i>	sp P31055	4e-25
XF-07A10-gl40	Proteína hipotética conservada de 3.6 kDa	<i>Escherichia coli</i>	sp P39455	2e-84
XF-07A10-gl41	Proteína hipotética de 3.84 kDa			
XF-07A10-gl42	Proteína hipotética de 3.76 kDa			
XF-07A10-gl43	Proteína hipotética de 5.96 kDa			
XF-07A10-gl45	Beta-glicosidase	<i>Bacteroides fragilis</i>	gi 2253101	1e-137
XF-07A10-gl46	Proteína hipotética 9.21 kDa			
XF-07A10-gl47	Proteína-ribossomal—alanina acetiltransferase	<i>Escherichia coli</i>	sp P09453	2e-22
XF-07A10-gl48	Fosfatidiltransferase	<i>Helicobacter pylori</i>	gi 2199520	6e-27

ORF	Proteína	Organismo	Acession #	E-value
XF-07A10-gl49	Proteína hipotética de 6.77 kDa			
XF-07A10-gl50	Proteína hipotética de 15.36 kDa			
XF-07A10-gl52	Prolyl-tRNA sintetase	<i>Escherichia coli</i>	sp P16659	1e-165
XF-07A10-gl53	Proteína DNA-ligante	<i>Bordetella bronchiseptica</i>	gi 4107150	4e-08
XF-07A10-gl54	Proteína hipotética de 7.82 kDa			
XF-07A10-gl55	Proteína hipotética de 4.20 kDa			
XF-07A10-gl56	Proteína hipotética conservada de 13.29 kDa	<i>Deinococcus radiodurans</i>	gi 6458527	9e-22
XF-07A10-gl57	Proteína hipotética de 6.16 kDa			
XF-07A10-gl58	Proteína hipotética de 3.30 kDa			
XF-07A10-gl59	Proteína hipotética de 4.24 kDa			
XF-07A10-gl61	Proteína regulatória de sistema de tradução de sinal de dois componentes	<i>Pseudomonas aeruginosa</i>	sp P43501	6e-32
XF-07A10-gl62	Proteína hipotética conservada de 40.0 kDa	<i>Bacillus subtilis</i>	sp P26646	3e-54
XF-07A10-gl63	Protease integral a membrana	<i>Escherichia coli</i>	sp P25662	2e-45.

Obs: As ORFs marcadas em azul não fazem mais parte do banco de dados de ORFs de *Xylella fastidiosa*. A explicação para tal expurgo provavelmente é a existência de sobreposição a outras ORFs, sendo nesses casos eliminadas as hipotéticas ou as menores.

IV A 2. Cosmídeo X0QH-02E06

O segundo cosmídeo seqüenciado foi o 02E06, entre setembro de 1998 e início de março de 1999, num período de 6 meses. O 02E06 foi seqüenciado para ligar o cosmídeo 07A10 ao 07A04 (seqüenciado pelo grupo da Profa. Dra. Mayana Zatz, IB-USP – São Paulo, UI), uma vez que o mapeamento por hibridização eletrônica com seqüências das regiões das extremidades do inserto do 02E06 indicava que seu inserto tinha sobreposições com os insertos desses dois outros cosmídeos. Duas bibliotecas de plasmídeos do 02E06 foram feitas, ambas usando nebulização, denominadas L e M.

A montagem final do inserto compreendeu 1212 seqüências, numa extensão de 43.113 pb, contendo 611 seqüências das bibliotecas do 02E06, 389 seqüências da biblioteca do cosmídeo 07A04, 200 seqüências do 07A10 e 12 seqüências de bibliotecas de “shotgun”. A taxa de erro final do inserto do cosmídeo foi determinada como 0,08 / 10.000 pb. Os relatórios de estágios e término de seqüenciamento estão presentes no Anexo (página 73). A digestão eletrônica com *EcoRI* e *HindIII* mostraram bandas iguais à observada em fotografia do gel de agarose das digestões utilizadas para verificação de unicidade de cosmídeos na placa recebida da coordenação (Anexo, 92).

A análise da seqüência do inserto do cosmídeo 02E06, realizada pelo programa GLIMMER no laboratório de bioinformática, determinou a presença de 65 ORFs, das quais foram analisadas 26 para anotação (as outras estavam em sobreposição com os outros dois cosmídeos, já anotados), na região do inserto compreendida entre as bases 18.094 e 33.047, de acordo com a metodologia descrita na seção de Materiais e Métodos. A anotação dessa região ficou a cargo do estudante de doutorado do Prof. El-Dorry, Eric Bonaccorsi. As ORFs do cosmídeo 02E06 anotadas são mostradas na Figura 7 e a anotação correspondente é mostrada na Tabela III.

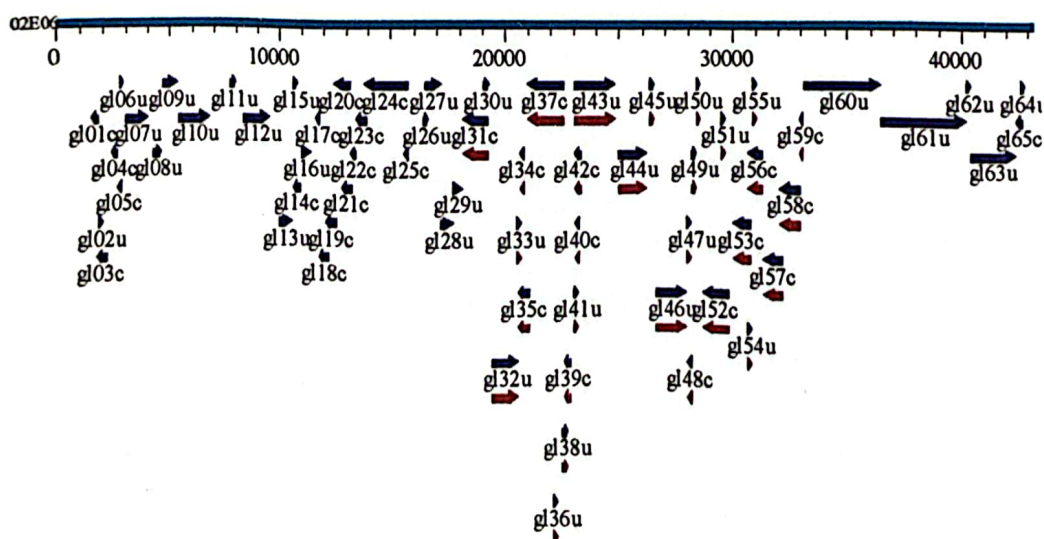


Figura 7: Anotação das ORFs do cosmídeo X0QH-02E06. São mostradas em azul as ORFs encontradas pelo programa Glimmer na seqüência do cosmídeo 02E06. Estão em vermelho as ORFs anotadas para esse cosmídeo, lembrando que as demais ORFs não foram anotadas uma vez que fazem parte de região de sobreposição a outros cosmídeos já anotados, no caso o 07A04 e o 07A10. As ORFs estão denominadas pelo código gl#(u), onde gl significa uma ORF Glimmer, seguido do número da ORF no cosmídeo e da letra u, no caso do sentido da ORF ser o mesmo da seqüência do cosmídeo (uncomplemented), ou da letra c, no caso da ORF estar no sentido inverso ao da seqüência dos cosmídeo, ou seja na seqüência complementar (complemented).

Tabela III: ORFs anotadas no cosmídeo X0QH-02E06. São mostrados os código da ORF e as informações referentes ao resultado de maior similaridade da busca realizada através do BLASTP: o nome da proteína, o organismo a que pertence, seu código no GenBank (gi) ou no SwissProt (sp) e a probabilidade esperada para a similaridade ser ao acaso (E-value).

ORF	Proteína	Organismo	Código # (Accession #)	E-value (BLASTP)
XF-02E06-gl31	Proteína integral da membrana	<i>Escherichia coli</i>	sp P42601	2e-69
XF-02E06-gl32	Proteína hipotética conservada de 42.8 kDa	<i>Escherichia coli</i>	sp P75876	2e-50
XF-02E06-gl33	Proteína hipotética de 3.88 kDa			
XF-02E06-gl34	Proteína hipotética de 3.82 kDa			
XF-02E06-gl35	Proteína hipotética de 13.36 kDa			
XF-02E06-gl36	Proteína hipotética de 4.59 kDa			
XF-02E06-gl37	Transportador de aminoácidos	<i>Bacillus subtilis</i>	gi 2633281	2e-87
XF-02E06-gl38	Proteína hipotética de 9.77 kDa			
XF-02E06-gl39	Proteína hipotética de 5.71 kDa			
XF-02E06-gl40	Proteína hipotética de 4.37 kDa			
XF-02E06-gl41	Proteína hipotética de 9.12 kDa			
XF-02E06-gl42	Proteína hipotética de 6.41 kDa			
XF-02E06-gl43	Transportador ABC de nitrato, permease	<i>Phormidium laminosum</i>	sp Q51881	3e-11
XF-02E06-gl44	Transportador ABC de nitrato, proteína de ligação a ATP	<i>Synechocystis sp.</i>	sp P73265	3e-46
XF-02E06-gl45	Proteína hipotética de 10.15 kDa			
XF-02E06-gl46	Proteína hipotética conservada de 50.2 kDa	<i>Chlamydia trachomatis</i>	sp P27850	3e-11
XF-02E06-gl47	Proteína hipotética de 3.64 kDa			
XF-02E06-gl48	Proteína hipotética de 5.00 kDa			
XF-02E06-gl49	Proteína hipotética de 7.39 kDa			
XF-02E06-gl50	Proteína hipotética de 7.58 kDa			
XF-02E06-gl51	Proteína hipotética de 4.43 kDa			
XF-02E06-gl52	Precursor de lipoproteína	<i>Shigella flexneri</i>	sp P43262	1e-25
XF-02E06-gl53	Proteína de tolerância a tolueno	<i>Escherichia coli</i>	gi 4336801	1e-11
XF-02E06-gl54	Proteína hipotética de 4.10 kDa			

ORF	Proteína	Organismo	Acession #	E-value
XF-02E06-g155	Proteína hipotética de 4.44 kDa			
XF-02E06-g156	Proteína de tolerância a tolueno	<i>Pseudomonas putida</i>	gi 4336800	4e-19
XF-02E06-g157	Proteína de tolerância a tolueno	<i>Escherichia coli</i>	gi 4336799	3e-52
XF-02E06-g158	Proteína de tolerância a tolueno	<i>Pseudomonas putida</i>	gi 4336798	4e-58
XF-02E06-g159	Proteína hipotética de 3.33 kDa			

Obs: As ORFs marcadas em azul não fazem mais parte do banco de dados de ORFs de *Xylella fastidiosa*. A explicação para tal expurgo provavelmente é a existência de sobreposição a outras ORFs, sendo nesses casos eliminadas as hipotéticas ou as menores.

IV A 3. Cosmídeo X0QH-06A08

O terceiro cosmídeo seqüenciado foi o 06A08, entre fim de dezembro de 1998 e final de março de 1999, num período de 3 meses. Do mesmo modo que o 02E06, o 06A08 foi seqüenciado para fazer a ligação entre dois cosmídeos, no caso o 07A01 (seqüência pelo grupo do Prof. Dr. Paulo Arruda, CBMEG-UNICAMP, Campinas, CC) e o 07C04 (seqüenciado pelo grupo da Profa. Dra. Eliana Lemos, UNESP, Jaboticabal, JE). Foi feita um única biblioteca, utilizando nebulização, denominada A.

A montagem final do inserto compreendeu seqüências, numa extensão de 39.491 pb, contendo 545 seqüências do 06A08, 395 seqüências do 07A01, 90 seqüências de bibliotecas de “shotgun” e 26 seqüências do 07C04. A taxa de erro final do inserto do cosmídeo foi determinada como 0,00 / 10.000 pb. Os relatórios de estágios e término de seqüenciamento estão presentes no Anexo (página 98). As digestões eletrônicas utilizando tanto *EcoRI*, *HindIII* como as duas conjuntamente, mostraram as mesmas bandas obtidas pelas digestões do cosmídeo em si com tais enzimas (Anexo, página 99).

A análise da seqüência do inserto do cosmídeo 06A08, realizada pelo programa GLIMMER no laboratório de bioinformática, determinou 52 ORFs, das quais 47 foram analisadas para anotação, de acordo com a metodologia descrita na seção de Materiais e Métodos. As ORFs do cosmídeo 06A08 anotadas são mostradas na Figura 8 e a anotação correspondente é mostrada na Tabela IV.

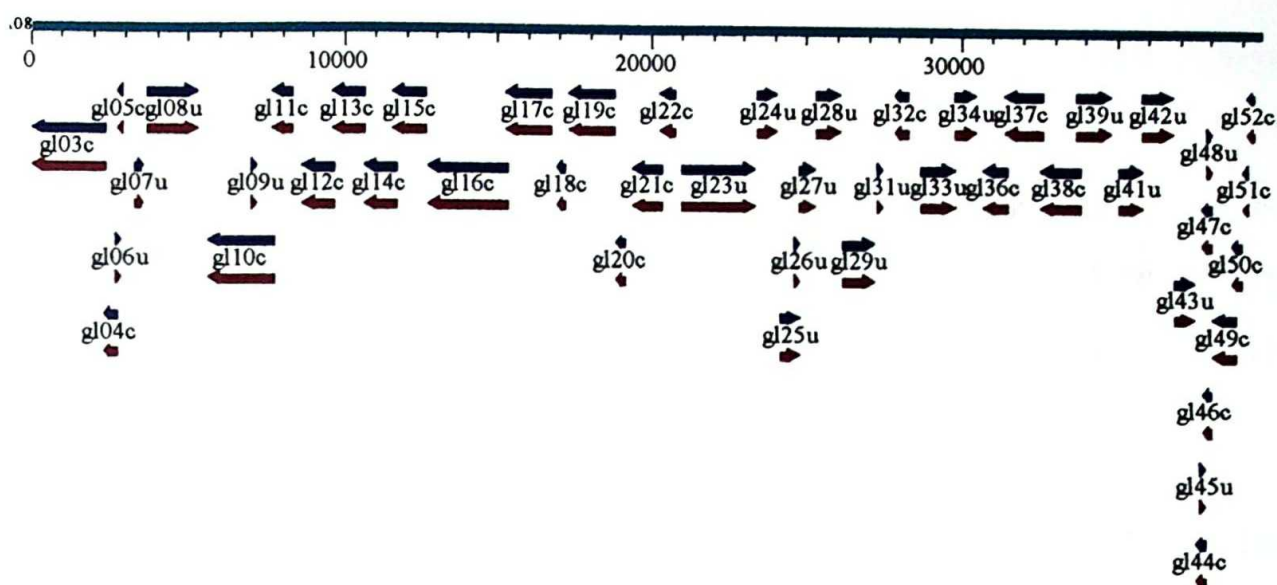


Figura 8: Anotação das ORFs do cosmídeo X0QH-06A08. São mostradas em azul as ORFs encontradas pelo programa Glimmer na seqüência do cosmídeo 06A08. Todas as ORFs desse cosmídeo foram anotadas. As ORFs estão denominadas pelo código gl#c(u), onde gl significa uma ORF Glimmer, seguido do número da ORF no cosmídeo e da letra u, no caso do sentido da ORF ser o mesmo da seqüência do cosmídeo (uncomplemented), ou da letra c, no caso da ORF estar no sentido inverso ao da seqüência dos cosmídeo, ou seja na seqüência complementar (complemented).

Tabela IV: ORFs anotadas no cosmídeo X0QH-06A08. São mostrados os código da ORF e as informações referentes ao resultado de maior similaridade da busca realizada através do BLASTP: o nome da proteína, o organismo a que pertence, seu código no GenBank (gi) ou no SwissProt (sp) e a probabilidade esperada para a similaridade ser ao acaso (E-value).

ORF	Proteína	Organismo	Código # (Accession #)	E-value (BLASTP)
XF-06A08-gl03	Proteína de biogênese do citocromo tipo C (proteína de tolerância a cobre)	<i>Chlamydia trachomatis</i>	gi 3329039	9e-26
XF-06A08-gl04	Proteína de biogênese do citocromo tipo C (proteína periplásmica de tolerância a cátions divalentes)	<i>Escherichia coli</i>	sp P36654	2e-17
XF-06A08-gl05	Proteína hipotética de 3.05 kDa			
XF-06A08-gl06	Proteína hipotética de 4.22 kDa			
XF-06A08-gl07	Chaperonina de 10 kDa	<i>Coxiella burnetii</i>	sp P19422	9e-28
XF-06A08-gl08	Chaperonina de 60 kDa	<i>Allochromatium vinosum</i>	sp P31293	0.0
XF-06A08-gl09	Proteína hipotética de 5.18 kDa			
XF-06A08-gl10	Proteína hipotética de 78.92 kDa			
XF-06A08-gl11	Proteína hipotética de 23.43 kDa			
XF-06A08-gl12	Dolicol-fosfato manosiltransferase	<i>Bacillus subtilis</i>	gi 2984166	5e-06

ORF	Proteína	Organismo	Acession #	E-value
XF-06A08-gl13	dTDP-glicose 4,6-desidratase	<i>Synechocystis sp.</i>	gi 1653195	1e-120
XF-06A08-gl14	UDP-glicose 4-epimerase	<i>Azospirillum brasilense</i>	gi 604222	1e-21
XF-06A08-gl15	GDP-manose 4,6-desidratase	<i>Aquifex aeolicus</i>	gi 2983560	e-117
XF-06A08-gl16	Manosiltransferase	<i>Escherichia coli</i>	gi 2125945	e-144
XF-06A08-gl17	Proteína hipotética de 51.45 kDa			
XF-06A08-gl18	Proteína hipotética de 5.34 kDa			
XF-06A08-gl19	Cistationina beta-sintase	<i>Dictyostelium discoideum</i>	sp P46794	1e-62
XF-06A08-gl20	Proteína hipotética de 8.66 kDa			
XF-06A08-gl21	Proteína hipotética de 32.43 kDa			
XF-06A08-gl22	Proteína hipotética de 15.23 kDa			
XF-06A08-gl23	Receptor para transporte de ferro dependente de TonB	<i>Escherichia coli</i>	sp P75780	0.0
XF-06A08-gl24	Proteína hipotética conservada de 25.4 kDa	<i>Escherichia coli</i>	gi 2981041	1e-64
XF-06A08-gl25	Proteína hipotética conservada de 24.20 kDa	<i>Deinococcus radiodurans</i>	gi 6459570	9e-12
XF-06A08-gl26	Proteína hipotética de 5.10 kDa			
XF-06A08-gl27	Proteína hipotética conservada de 19.48 kDa	<i>Deinococcus radiodurans</i>	gi 6459571	2e-09
XF-06A08-gl28	Proteína hipotética de 29.63 kDa			
XF-06A08-gl29	Proteína hipotética conservada de 35.31 kDa	<i>Escherichia coli</i>	sp P33944	6e-40
XF-06A08-gl31	Proteína hipotética de 4.83 kDa			
XF-06A08-gl32	Proteína hipotética conservada de 12.1 kDa	<i>Haemophilus influenzae</i>	sp Q57134	4e-11
XF-06A08-gl33	Proteína hipotética conservada de 41.91 kDa	<i>Pseudomonas aeruginosa</i>	gi 4545243	3e-63
XF-06A08-gl34	Fator de virulência	<i>Pseudomonas aeruginosa</i>	gi 4545244	1e-48
XF-06A08-gl36	Proteína hipotética conservada de kDa	<i>Escherichia coli</i>	sp P76570	3e-26
XF-06A08-gl37	Permease	<i>Rickettsia prowazekii</i>	gi 3861171	7e-23
XF-06A08-gl38	Proteína hipotética de 42.02 kDa			
XF-06A08-gl39	5'-fosforribosil-5-aminoimidazol sintetase	<i>Escherichia coli</i>	sp P08178	1e-106
XF-06A08-gl41	5'-fosforribosilglicinamida transformilase	<i>Escherichia coli</i>	sp P08179	1e-38
XF-06A08-gl42	Proteína hipotética de 34.47 kDa			

ORF	Proteína	Organismo	Acession #	E-value
XF-06A08-gl43	Proteína hipotética de 25.31 kDa			
XF-06A08-gl44	Proteína hipotética de 7.24 kDa			
XF-06A08-gl45	Proteína hipotética de 4.09 kDa			
XF-06A08-gl46	Proteína hipotética de 5.19 kDa			
XF-06A08-gl47	Proteína hipotética de 7.52 kDa			
XF-06A08-gl48	Proteína hipotética de 6.38 kDa			
XF-06A08-gl49	Timidilato quinase	<i>Pyrococcus horikoshii</i>	sp O59366	2e-22
XF-06A08-gl50	Proteína hipotética de 8.17 kDa			
XF-06A08-gl51	Proteína hipotética de 4.83 kDa			
XF-06A08-gl52	Proteína hipotética de 6.07 kDa			

Obs: As ORFs marcadas em azul não fazem mais parte do banco de dados de ORFs de *Xylella fastidiosa*. A explicação para tal expurgo provavelmente é a existência de sobreposição a outras ORFs, sendo nesses casos eliminadas as hipotéticas ou as menores.

IV A 4. Cosmídeo X0QH/QS-02G12

O cosmídeo 02G12 foi recebido em abril e foi terminado no início de agosto de 1999, período de 5 meses. Ele foi seqüenciado por seu inserto ter sobreposição com o do 01A03 (seqüenciado pelo grupo do Prof. Dr. Arthur Gruber, FMVZ-USP, São Paulo, UV) numa de suas extremidades e, na época em que foi distribuído, não haver sobreposição de sua outra extremidade, estendendo assim a região do genoma já seqüenciada. Cada grupo fez uma biblioteca independentemente, ambas por nebulização. A biblioteca feita por nosso grupo foi denominada G, enquanto a do grupo QS foi denominada N.

A montagem final do inserto do cosmídeo compreendeu seqüências, numa extensão de 41.105 pb, contendo 396 seqüências da biblioteca G do 02G12 (QH), 394 seqüências da biblioteca N do 02G12 (QS), 88 seqüências de bibliotecas de "shotgun", 26 seqüências do 01A03 e 2 seqüências das extremidades do inserto realizadas pela coordenação (grupo do Dr. Andrew Simpson, LICR, São Paulo, IL) para experimentos de hibridização. A taxa de erro da montagem final do inserto do cosmídeo foi determinada como 0,00 / 10.000 pb.

Os relatórios de estágios e término de seqüenciamento estão presentes no Anexo (página 93). Também são apresentadas no Anexo (página 94) as digestões eletrônicas da seqüência do inserto do cosmídeo no vetor Lawrist4, que fornecem padrões bastante semelhantes aos obtidos pelas digestões do cosmídeo com as enzimas *EcoRI*, *HindIII* e ambas conjuntamente.

A análise da seqüência do inserto do cosmídeo 02G12, realizada pelo programa GLIMMER no laboratório de bioinformática, determinou 63 ORFs das quais foram analisadas para a anotação 15, na região do inserto entre as bases 28.176 e 41.104, de acordo com a metodologia descrita na seção de Materiais e Métodos. As ORFs do cosmídeo 02G12 anotadas são mostradas na Figura 9 e a anotação correspondente é mostrada na Tabela V.

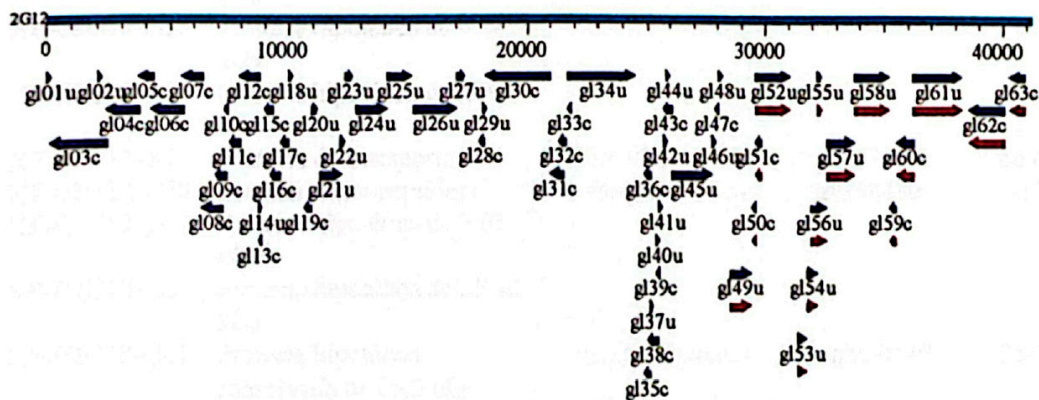


Figura 9: Anotação das ORFs do cosmídeo X0QH/QS-02G12. São mostradas em azul as ORFs encontradas pelo programa Glimmer na seqüência do cosmídeo 02G12. Estão em vermelho as ORFs anotadas para esse cosmídeo, lembrando que as demais ORFs não foram anotadas uma vez que fazem parte de região de sobreposição a outro cosmídeo já anotado, no caso o 01A03. As ORFs estão denominadas pelo código gl#(u), onde gl significa uma ORF Glimmer, seguido do número da ORF no cosmídeo e da letra u, no caso do sentido da ORF ser o mesmo da seqüência do cosmídeo (uncomplemented), ou da letra c, no caso da ORF estar no sentido inverso ao da seqüência dos cosmídeo, ou seja na seqüência complementar (complemented).

Tabela V: ORFs anotadas no cosmídeo X0QH/QS-02G12. São mostrados os código da ORF e as informações referentes ao resultado de maior similaridade da busca realizada através do BLASTP: o nome da proteína, o organismo a que pertence, seu código no GenBank (gi) ou no SwissProt (sp) e a probabilidade esperada para a similaridade ser ao acaso (E-value).

ORF	Proteína	Organismo	Código # (Acessão #)	E-value (BLASTP)
XF-02G12-gl49	Epimerase de Açúcares-nucleotidados	<i>Escherichia coli</i>	gi 1788642	5e-51
XF-02G12-gl50	Proteína hipotética de 4.04 kDa			
XF-02G12-gl51	Proteína hipotética de 6.48 kDa			
XF-02G12-gl52	Serina protease periplásmica, proteína de heat shock	<i>Escherichia coli</i>	pir S45229	4e-92
XF-02G12-gl53	Proteína hipotética de kDa			
XF-02G12-gl54	Proteína hipotética de 17.40 kDa			
XF-02G12-gl55	Proteína hipotética de 4.78 kDa			
XF-02G12-gl56	Proteína hipotética de 25.43 kDa			
XF-02G12-gl57	Proteína de transporte	<i>Bacillus subtilis</i>	gi 2635675	6e-07
XF-02G12-gl58	Leucina aminopeptidase	<i>Aquifex aeolicus</i>	gi 2984306	e-131
XF-02G12-gl59	Proteína hipotética de 7.65 kDa			
XF-02G12-gl60	Proteína hipotética de 25.42 kDa			
XF-02G12-gl61	Proteína hipotética conservada de 26.3 kDa	<i>Escherichia coli</i>	sp P30149	3e-19
XF-02G12-gl62	UDP-N-acetilmuramato—L-alanina sintetase	<i>Escherichia coli</i>	sp P37773	e-114
XF-02G12-gl63	Adenilato quinase	<i>Synechococcus sp.</i>	sp O24706	1e-41

Obs: As ORFs marcadas em azul não fazem mais parte do banco de dados de ORFs de *Xylella fastidiosa*. A explicação para tal expurgo provavelmente é a existência de sobreposição a outras ORFs, sendo nesses casos eliminadas as hipotéticas ou as menores.

IV A 5. Cosmídeo X0QH/QS-05D08

O 05D08, também seqüenciado em conjunto com o grupo QS, foi recebido em junho e terminado em setembro de 1999, período de 4 meses. O 05D08 foi escolhido para ser seqüenciado pois parecia estender o seqüenciamento a partir do cosmídeo 07C06 (seqüenciado pelo grupo do Prof. Dr. Celso Marinho, UNESP, Botucatu, BG), que tinha sido desmembrado pois descobrira-se ter um inserto aparentemente quimérico, ou seja, o inserto compreendia dois fragmentos genômicos de *Xylella fastidiosa* que não eram contíguos.

Como no seqüenciamento do X0QH/QS-02G12, cada grupo fez sua biblioteca, sendo que o grupo QS fez 4 bibliotecas, denominadas L, M, N e P, sendo utilizadas maior número de seqüências da última biblioteca com a qual obteve-se melhor qualidade (P). Nosso grupo preparou um única biblioteca denominada D. Todas as bibliotecas foram realizadas por nebulização.

A medida que o inserto do cosmídeo foi sendo montado, descobriu-se que ele era quimérico tal como o 07C06 (a região de junção dos fragmentos não contíguos nos dois cosmídeos é a mesma), indicando que o 05D08 talvez fosse uma réplica do 07C06. A montagem final do inserto quimérico do 05D08 compreendeu 1095 seqüências, numa extensão de 44.837 pb, contendo 336 da biblioteca D do 05D08 (QH), 195 seqüências das bibliotecas L, M, N e P do 05D08 (QS), 137 seqüências de bibliotecas de “shotgun”, 191 seqüências do 06H03 (UV, QV), 115 seqüências do 07C06 (BG), 105 seqüências do 07H02 (QS), 6 seqüências do 11G06 (seqüenciado pelos grupos do Prof. Dr. Fernando Ferreira Costa e do Prof. Dr. Luis Eduardo Aranha Camargo, ESALQ-USP, Piracicaba, EC), 2 seqüências das extremidades do inserto feitas pelo grupo do Prof. Dr. Sérgio Verjovski-Almeida (IQ-USP – São Paulo, QV) e outras 2 seqüências das extremidades do inserto feitos pelo grupo do Dr. Luiz R. Nunes (UMG, Mogi das Cruzes, MC). A grande quantidade de seqüências de outros cosmídeos se deve ao fato de os dois fragmentos não contíguos presentes no inserto do cosmídeo se sobrepunham em regiões totalmente diferentes da montagem geral do genoma de *Xylella fastidiosa*.

Quando a montagem do inserto foi inicialmente submetida ao laboratório de bioinformática, ele tinha a extensão de 44.837 pb. Visto ser um inserto quimérico, o inserto foi então dividido em dois fragmentos: um de 34.237 pb denominado e submetido como 05D08 e outro de 10.600 pb denominado e submetido como 15D08. A montagem e desmembramento do inserto do cosmídeo foram realizados pela Profa. Dra. Aline Silva (QS). A taxa de erro tanto do 05D08 como do 15D08 foi determinada como 0,00 / 10.000 pb. Os relatórios de estágio e término de cada fragmento estão no Anexo (05D08, página 96 e 15D08, página 97) Uma vez que o cosmídeo teve de ser desmembrado para a submissão, já que era quimérico, não foi realizada a digestão computacional do inserto para comparação com o padrão de digestão com enzimas de restrição.

A análise da seqüência do fragmento 05D08, realizada pelo programa GLIMMER no laboratório de bioinformática, determinou 49 ORFs, mas nenhum dos dois fragmentos foi anotado por se sobreporem a cosmídeos previamente anotados.

IV B. Fechamento de “gaps”

IV B 1. GFS X0QH-00J42

Uma das últimas etapas do seqüenciamento do genoma da bactéria *Xylella fastidiosa* foi o fechamento dos “gaps”. Para tal, a coordenação investiu em várias estratégias para que houvessem várias opções para os “gaps” existentes.

Uma das estratégias que se provou bastante eficaz foi o seqüenciamento completo de insertos de “shotgun” genômico como se fossem insertos de cosmídeos. Dessa forma obtinha-se um seqüência sobre o “gap” de grande qualidade, já que o procedimento envolvia a subclonagem do inserto do plasmídeo em uma biblioteca, o que por sua vez proporcionava muitas seqüências em diversos pontos, reforçando a qualidade da seqüência consenso da região do “gap” em si (tais seqüências estão presentes no mapa de disposição dos clones, Anexo página 71, em amarelo).

Foi seqüenciado em nosso laboratório, o inserto do plasmídeo G02 da placa 13 da biblioteca 10 de “shotgun” genômico, que tinha aproximadamente 3,0 quilobases de extensão. Tal plasmídeo foi escolhido, uma vez que as seqüências das extremidades de seu inserto estarem em contíguos genômicos diferentes e, portanto, o inserto formava uma ponte entre os dois contíguos. Após o resseqüenciamento das extremidades do inserto e a confirmação de que o plasmídeo realmente poderia ajudar a fechar um “gap” foi realizado o seqüenciamento completo de seu inserto como descrito na parte de Materiais e Métodos.

O inserto foi denominado então GFS X0QH-00J42, sendo obtido o término de seu seqüenciamento com 73 seqüências, sendo 67 da biblioteca do plasmídeo e 6 de seqüências de “shotgun” genômico, num total de 2622 bases no inserto. Os relatórios de estágios e término do seqüenciamento são mostrados no Anexo (página 103).

IV C. Seqüenciamento de clones de bibliotecas de “shotgun” genômico

No decurso do final do seqüenciamento de cosmídeos a coordenação decidiu voltar a seqüenciar bibliotecas de “shotgun”, o que se mostrou de grande valia para o preenchimento dos “gaps” (mapa de cobertura do seqüenciamento, Anexo página 72).

Assim, nosso grupo consentiu em receber placas de clones de bibliotecas de “shotgun”, mais precisamente das bibliotecas 07, 08, 10 e 15. Todas essas quatro bibliotecas foram realizadas a partir da nebulização de DNA total de *Xylella fastidiosa* e clonagem de fragmentos no sítio de *Sma*I do vetor pUC18 (para cada uma das bibliotecas o tamanho do inserto difere de acordo com a tabela mostrada na introdução). O número de seqüências obtidas de cada biblioteca e o número de bases correspondentes são mostrados na Tabela VI.

Tabela VI: Seqüenciamento de bibliotecas de “shotgun” genômico

Biblioteca	# seqüências	# bases
07	103	38.904
08	277	110,291
10	265	102,618
15	366	126,342

IV D. Anotação de regiões de cosmídeos seqüenciados por outros laboratórios

Além da anotação realizada no cosmídeos seqüenciados por nosso laboratório, também foram realizadas anotações em regiões de outros cosmídeos, com o objetivo de acelerar o processo de anotação do genoma como um todo. As regiões anotadas, o número de ORFs e o número de quilobases verificado são mostradas na Tabela VII.

Tabela VII: Regiões anotadas de cosmídeos seqüenciados por outros grupos

Cosmídeo	Intervalo anotado / # pb	# ORFs anotadas *
01A01	21.159-26.764 / 5.605	6
05G03	18.064-36.496 / 18.432	27
06D06	23.471-26.305 / 2.834	3
06H03	23.055-37.321 / 14.266	25
07B01	18.401-24.072 / 5.671	8
09C12	8.374-22.062 / 13.688	19
09D10	36.199-39.260 / 3.061	6
10B10	18.513-26.879 / 8.366	15
11A03	26.059-39.260 / 13.201	11
11A10	15.883-22.064 / 6.181	1
Total	91.305	121

* estão incluídas as ORFs hipotéticas.

IV E. Anotação da Categoria I: Metabolismo Intermediário

Dentro desse objetivo foram realizadas buscas pelas enzimas envolvidas nas principais rotas metabólicas do Metabolismo Intermediário que foi subdividido em diversas subcategorias. A hierarquia da categoria I e o número de enzimas anotadas em cada subcategoria são mostrados na Figura 10. Uma tabela detalhada contendo o nome das enzimas distribuídas nas subdivisões da categoria I está presente no Anexo (página 104).

Em princípio considerando-se somente a baixa similaridade de seqüência entre enzimas de uma via metabólica não se pode descartar a presença de uma rota. São conhecidas em vários organismos enzimas cujas seqüências diferem das mais conservadas, mas cujas estruturas são semelhantes, fornecendo o ambiente necessário para que ocorra a reação catalisada⁴⁰. Há casos em que ocorrem enzimas que evoluíram de forma diferente mas acabam por catalisar a mesma reação^{41,42}. Entretanto numa primeira instância, a busca por rotas metabólicas através de similaridade entre seqüências já pode evidenciar pelo menos as diferenças existentes entre *Xylella fastidiosa* e outras bactérias em relação à presença de proteínas similares.

A busca das enzimas e das vias metabólicas basearam-se principalmente em *Escherichia coli* devido à vasta quantidade de informações existente sobre o metabolismo desse organismo, além de pertencer também ao grupo Gama das Eubactérias. Tal estratégia identificou grande parte das supostas enzimas de *Xylella fastidiosa*, entretanto, houve exceções como a frutose-bisfosfato aldolase, que se assemelha muito mais às enzimas presentes em plantas do que às presentes em bactérias (BLASTP e alinhamento de seqüências, Anexo, páginas 112 e 113 respectivamente).

Considerando as fontes de energia, temos que a utilização tanto de glicose como de manose, e provavelmente frutose e galactose, é certa. Lactose, lactato e glicerol também podem ser candidatos a fonte de energia. Os Profs. Drs. Sérgio Verjovski-Almeida (IQ-USP, São Paulo) e João Meidanis (DCC-UNICAMP, Campinas), responsáveis pela categoria de transporte na anotação de *Xylella fastidiosa*, identificaram uma série de transportadores de carboidratos, inclusive enzimas envolvidas no sistema de transporte e fosforilação (“phosphotransferase system”: PTS, não foram identificadas proteínas II referentes a especificidade do sacarídeo transportado, entretanto) (Anexo, página 111), assegurando o transporte de glicose, frutose, galactose, glicerol distintamente e provavelmente de outros sacarídeos.

Categoria / Subcategoria	# de enzimas	
I. Metabolismo Intermediário	TOTAL	252
<u>I.A. Degradação</u>	<u>Subtotal</u>	<u>28</u>
I.A.1. Degradação de polissacarídeos		1*
I.A.2. Degradação de pequenas moléculas		27
• Aminas		
• Aminoácidos		
• Compostos carbônicos		
• Ácidos Graxos		
<u>I.B. Metabolismo Intermediário Central</u>	<u>Subtotal</u>	<u>62</u>
I.B.1. Aminoaçúcares		4
I.B.2. Entner-Doudoroff		2
I.B.3. Gliconeogênese		2
I.B.4. Via do Glioxilato		0
I.B.5. Metabolismo da glicose, miscelânea		1
I.B.6. Via das Pentoses, ramo não-oxidativo		3
I.B.7. Hidrólise de Nucleotídeos		0
I.B.8. Interconversões de Nucleotídeos		2
I.B.9. Compostos contendo fósforo		4
I.B.10. Conversões com múltiplos objetivos		15
I.B.11. Biossíntese/conversões de açúcares nucleotídicos		5
I.B.12. Metabolismo do enxofre		6
<u>I.C. Metabolismo Energético, Carbono</u>	<u>Subtotal</u>	<u>86</u>
I.C.1. Respiração Aeróbica		17
I.C.2. Respiração Anaeróbica e Fermentação		4
I.C.3. Transporte de elétrons		25
I.C.4. Glicólise		9
I.C.5. Via das Pentoses, ramo oxidativo		2
I.C.6. Piruvato desidrogenase		3
I.C.7. Ciclo dos Ácidos Tricarboxílicos		17
I.C.8. Interconversões para ATP por força próton-motriz		9
<u>I.D. Funções Regulatórias</u>	<u>Subtotal</u>	<u>75</u>

Figura 10: Hierarquia e número de enzimas anotadas na Categoria I: Metabolismo intermediário. *
A maioria das enzimas digestivas de polissacarídeos foi classificada dentro da categoria de virulência, visto a possibilidade de estarem relacionadas à movimentação da bactéria nos vasos do xilema.

A utilização de glicose pode se dar pela fosforilação a glicose-6-fosfato por glicocinase ou pelo sistema PTS. A frutose, sendo assimilada pelo sistema PTS, torna-se frutose-1-fosfato que pela ação da frutose-1,6-bisfosfato aldolase presente (similar a adolase C de plantas) poderia ser dissociada em diidroxiacetona-fosfato e gliceraldeído. Manose, se transportada pelo sistema PTS, torna-se manose-6-fosfato, podendo ser metabolizada a manose-1-fosfato ou frutose-6-fosfato por fosfomanomutase e fosfomanoisomerase respectivamente. A utilização de lactose é assegurada pela presença de beta-galactosidase. A fosforilação de galactose é provável, embora a galactocinase não tenha sido identificada, visto que as demais enzimas da via de utilização de galactose (UTP-galactose-1-fosfato uridiltransferase e UDP-glicose-4-epimerase) foram identificadas. A utilização de glicerol seria realizada por fosforilação a glicerol-3-fosfato, pela glicerol quinase, e posterior oxidação a gliceraldeído-3-fosfato por uma desidrogenase. O lactato por sua vez seria metabolizado a piruvato diretamente pela ação da lactato desidrogenase.

Foi detectado um sistema celulolítico completo, contendo beta-glicosidase, exo-beta-1,4-celobiosidase e endo-glucanase extracelulares, além de xilanases. Tais enzimas podem estar envolvidas na degradação das paredes do xilema, constituída principalmente de celulose e hemicelulose, objetivando a obtenção de fonte de carbono e energia ou como um mecanismo de dissipação dentro dos vasos da plantas (daí estarem classificadas primariamente na categoria de virulência e patogenicidade).

Foram encontradas várias lipases que degradam triacilgliceróis em glicerol e ácidos graxos, mas o complexo de beta-oxidação dos ácidos graxos está aparentemente ausente. A bactéria entretanto tem as enzimas necessárias para a utilização do glicerol como fonte de energia. A degradação de aminoácidos se restringe àqueles que aparentemente são mais abundantes no xilema: glutamina, glutamato, asparagina, aspartato, alanina, glicina, serina e treonina.

Em termos de metabolismo energético, foram identificadas todas as enzimas da glicólise, as subunidades do complexo piruvato desidrogenase, as enzimas e subunidades de todas as etapas do ciclo dos ácidos tricarboxílicos e as envolvidas na cadeia de transporte de elétrons, assim como a ATP sintase (transportadora de prótons). Tanto as subunidades da ATP sintase como as do complexo NDH1 da cadeia de transporte de elétrons (NADH:ubiquinona desidrogenase) estão agrupadas em dois operons idênticos aos de *Escherichia coli*.

Não foram identificadas enzimas relacionadas à utilização tanto de compostos nitrogenados ou sulfurados como aceptores de elétrons, podendo-se dizer então que *Xylella fastidiosa* é um organismo dependente da respiração aeróbica.

Entre as vias aparentemente ausentes estão a via do glioxilato (não foram identificadas a isocitrato liase nem a malato sintase) e a gliconeogênese (não foram identificadas a piruvato carboxilase, fosfoenolpiruvato carboxiquinase nem frutose-bisfosfatase). A via das pentoses está incompleta, sendo que duas enzimas-chaves não foram identificadas: fosfogluconato desidrogenase (descarboxilante) e transaldolase.

As duas principais surpresas com relação à anotação de vias metabólicas foram a não identificação de várias das enzimas da gliconeogênese e a aparente ausência de duas enzimas da via das pentoses que, se confirmada, indicaria que *Xylella fastidiosa* teria de obter ribose-5-fosfato para a síntese de nucleotídeos usando uma via alternativa.

A principal importância da gliconeogênese seria a obtenção de glicose, ou derivados fosforilados nas posições 1 ou 6, para serem utilizados em processos biossintéticos, como a síntese da parede celular, a partir de compostos não-glicídicos como: intermediários do ciclo dos ácidos tricarboxílicos, glicerol, lactato e aminoácidos gliconeogênicos. Assim, na falta de glicose, a degradação de tais compostos pode supri-la ao metabolismo²⁵.

A gliconeogênese descrita em vários organismos utiliza a maior parte das reações envolvidas na glicólise e suas enzimas, com exceção das três etapas altamente exergônicas: fosforilação de glicose a glicose-6-fosfato, fosforilação de frutose-6-fosfato a frutose-1,6-bisfosfato e desfosforilação e oxidação de fosfoenolpiruvato a piruvato. A desfosforilação de glicose-6-fosfato é realizada pela glicose-6-fosfatase, a desfosforilação de frutose-1,6-bisfosfato é realizada pela frutose-1,6-bisfosfatase e a formação de fosfoenolpiruvato é realizada a partir de oxaloacetato pela fosfoenolpiruvato carboxiquinase, sendo possível a utilização de piruvato a partir de sua carboxilação a oxaloacetato pela piruvato carboxilase (Figura 11).

A busca pelas últimas quatro enzimas em *Xylella fastidiosa* não obteve sucesso, entretanto foi identificada uma via alternativa para a obtenção de fosfoenolpiruvato, presente em vários procariotos, que lança a dúvida sobre a real ausência da gliconeogênese. Foram identificadas a enzima málica (malato oxidoreductase) responsável pela descarboxilação oxidativa de malato a piruvato e a enzima fosfoenolpiruvato sintase responsável pela formação de fosfoenolpiruvato a partir de piruvato (Figura 12).

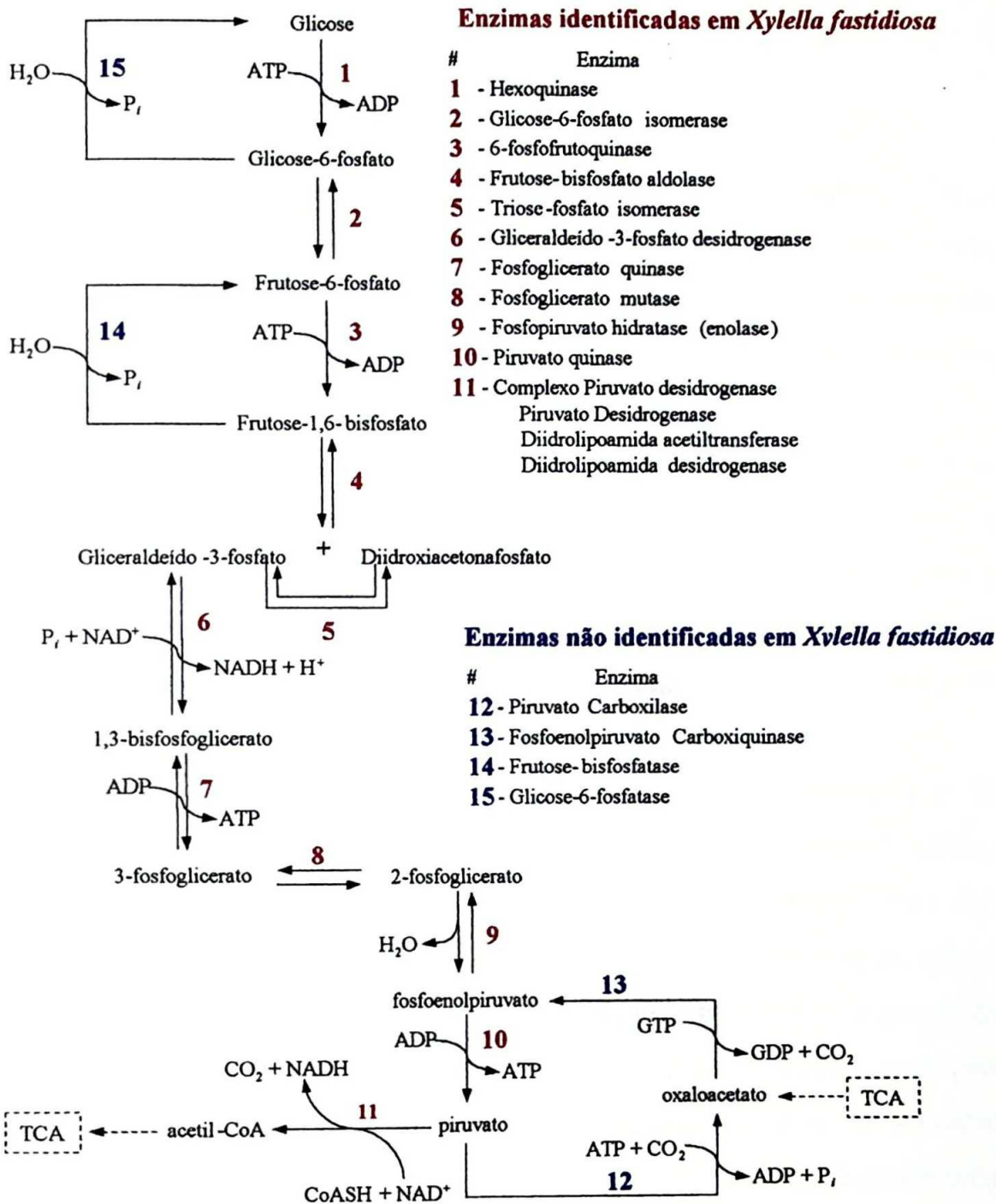


Figura 11: Reações envolvidas na Glicólise e na Gliconeogênese e formação de acetil-CoA a partir de piruvato. São indicadas quais enzimas foram identificadas e quais aparentemente não estão presentes em *Xylella fastidiosa*.

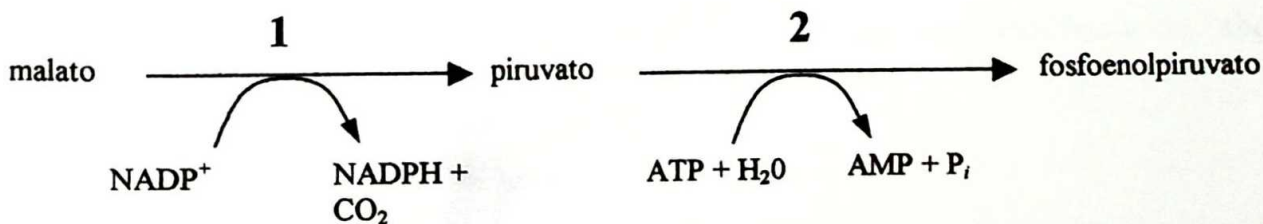


Figura 12: Via de obtenção de fosfoenolpiruvato a partir de malato. Foram identificadas em *Xylella fastidiosa* ambas as enzimas: 1 enzima málica; 2 fosfoenolpiruvato sintase.

V. Discussão

V A. Estratégias de Seqüenciamento

Com o grande número de projetos de seqüenciamento em andamento, há sempre uma busca pela determinação de uma estratégia que seja a mais eficaz e eficiente, fazendo com que a etapa de determinação da seqüência do genoma de um organismo se torne cada vez menos importante em vista das descobertas realizadas pela anotação e pelos projetos funcionais decorrentes.

Nesse âmbito, ao fim do seqüenciamento dos 2.679.305 pares de bases do genoma de *Xylella fastidiosa* (maiores dados sobre o genoma completo estarão disponíveis após publicação na página oficial do projeto na rede⁴³), tinha-se uma grande quantidade de informações com relação ao fato de terem sido utilizadas as mais diversas estratégias de seqüenciamento. Tais informações puderam servir como uma base de comparação entre as estratégias utilizadas.

A utilização de cosmídeos teve como principal ponto favorável o fato de estabelecer rapidamente uma base para um banco de dados de anotação de ORFs, visto que as anotações foram realizadas inicialmente em cada cosmídeo. O fato de terem insertos de tamanhos aproximados a 40 quilobases também foi importante para o seqüenciamento e a montagem correta de regiões contendo repetições de grande número de nucleotídeos. Entretanto, em algumas regiões do genoma repetições com pequenos números de nucleotídeos em cosmídeos diferentes, mas cujos insertos se sobrepunham, mostraram pequenas diferenças no número de repetições. Um desses casos ocorreu entre o cosmídeo X0QH-02E06 e os cosmídeo X0UI-07A04, numa região de repetição em tandem da seqüência CACCT. O X0QH-02E06 aparentemente possuía uma repetição a mais da seqüência em relação ao X0UI-07A04, embora o laboratório de bioinformática não tenha identificado problema de montagem em nenhum dos dois cosmídeos. A seqüência do X0UI-07A04 acabou sendo confirmada por uma seqüência de “shotgun” genômico.

Outro problema relacionado à utilização de cosmídeos, como de qualquer outro vetor que aceite insertos de grandes proporções, é a clonagem de insertos quiméricos formados por fragmentos de diferentes regiões do genoma que se ligam, formando um fragmento de DNA de tamanho compatível ao ser inserido no vetor. No projeto ocorreram casos de insertos quiméricos em alguns cosmídeos. Um exemplo ocorreu durante a montagem dos cosmídeos X0QH-07A10 e X0QH-02E06, indentificando-se com um fragmento quimérico numa das extremidades do X0QH-07A10.

O X0QH-02E06 foi escolhido para seqüenciamento baseado no fato das seqüências das extremidades de seu inserto hibridizarem eletronicamente com seqüências do X0QH-07A10 e do X0UI-07A04. Assim, sabendo-se da sobreposição entre os dois e nosso laboratório resolveu tentar montá-los conjuntamente para que o seqüenciamento de ambos fosse realizado mais rapidamente. Entretanto passadas algumas semanas de montagem conjunta, com ambos os cosmídeos com mais de 500 seqüências (número maior que o necessário normalmente para a obtenção da seqüência do inserto inteira, ainda que sem a qualidade desejada), não conseguia-se fazer a montagem da região contida nos insertos dos dois cosmídeos. Existiam um contíguo contendo seqüências somente do X0QH-02E06, outro com seqüências de ambos e um terceiro com seqüências somente do X0QH-07A10. Entretanto a somatória do tamanho do contíguo de seqüências do X0QH-02E06 e a região de sobreposição no contíguo formado por seqüências de ambos os cosmídeos fornecia o valor esperado para o inserto do cosmídeo X0QH-02E06.

Em uma visita ao laboratório de bioinformática, decidiu-se realizar a montagem dos cosmídeos em separado e obteve-se um único contíguo com as seqüências do X0QH-02E06. Contudo, se fossem adicionadas todas as seqüências já geradas no seqüenciamento do X0QH-07A10, que deveriam somente melhorar a qualidade do seqüenciamento do X0QH-02E06, o contíguo único era desmembrado. Seguiu-se então com a montagem dos cosmídeos em separado utilizando somente seqüências de um no outro cuidadosamente escolhidas para a melhoria da qualidade final. Ao final do seqüenciamento de ambos descobriu-se que o cosmídeo X0QH-07A10 continha na extremidade do inserto que se sobrepunha ao inserto do X0QH-02E06 uma região de aproximadamente 1200 pares de bases provinda de outra parte do genoma presente no cosmídeo X0QR-08D10.

Para se chegar a tal conclusão, foram realizadas hibridizações eletrônicas, buscas de similaridade de seqüência no genoma (BLASTN) e amplificação de região específica de diversos cosmídeos cujos insertos se sobrepunham à região de junção entre o X0QH-07A10 e o X0QH-02E06.

Para a amplificação foram desenhados dois primers: o primeiro na região de sobreposição comprovada entre os cosmídeos 07A10 e 02E06 (região que na montagem do 02E06 continha cromatogramas desse cosmídeo e do 07A10) e o segundo na região do 02E06 sem sobreposição entre cosmídeos de outros cosmídeos (região que na montagem do 02E06 só continha cromatogramas desse cosmídeo). Realizando a amplificação da região compreendida entre os primers, utilizando como molde tanto o 02E06, quanto os cosmídeos 06H08 e 06B12, que também faziam a ligação entre o 07A10 e o 07A04, confirmar se havia uma deleção interna ao 02E06, ou se havia uma inserção na extremidade do 07A10. Visto que a amplificação dos diversos moldes resultou em um amplificado de mesmo tamanho, concluiu-se que o 07A10 deveria conter uma inserção em sua extremidade que não estava presente nos outros três cosmídeos (Figura 13).

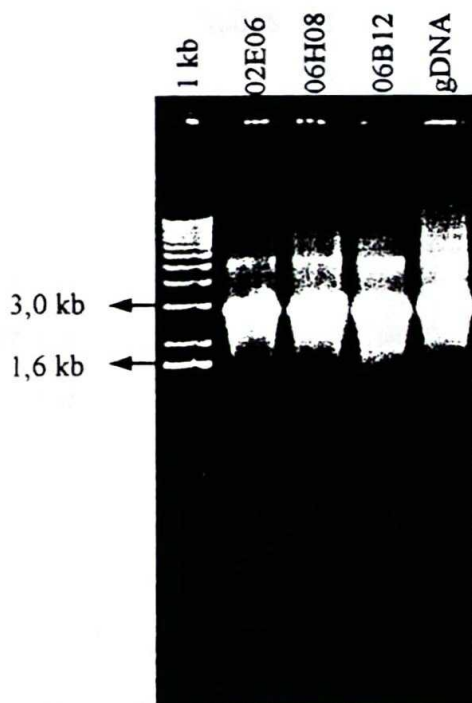


Figura 13: Amplificação de região de sobreposição entre os cosmídeos X0QH-02E06 e X0QH-07A10 utilizando como moldes os cosmídeos 02E06, 06H08, 06B12 e DNA genômico. A banda preferencialmente amplificada de aproximadamente 2,8 kb comprovou que não havia deleção no cosmídeo X0QH-02E06 e que a região de aproximadamente 1200pb na extremidade do inserto do cosmídeo X0QH-07A10 era quimérica.

Em contraposição aos problemas ocorridos no seqüenciamento de cosmídeos, a técnica de “shotgun” de DNA total, atualmente mais difundida, se mostrou realmente mais promissora. Mesmo que também ocorram os mesmos problemas de formação de insertos quiméricos, com uma probabilidade menor entretanto, o fato dessa técnica estar baseada na produção de seqüências em larga escala para a montagem direta da seqüência do genoma, e não da seqüência dos insertos de cosmídeos que então serviram para a montagem do genoma, fornece opções mais que suficientes para o descarte de seqüências duvidosas sem que tenha sido desperdiçado muito tempo na sua obtenção.

Além disso, o fato dos insertos em plasmídeos serem pequenos favorece o seqüenciamento de regiões que codifiquem proteínas que sejam nocivas às bactérias hospedeiras devido à clonagem de fragmentos dessas regiões. Tais regiões clonadas em cosmídeos normalmente levariam à perda do clone uma vez que as proteínas codificadas no inserto inviabilizariam o desenvolvimento da bactéria hospedeira, significando em termos de montagem de genoma uma região que se tornaria um “gap” (como pode ser observado no mapa de cobertura do seqüenciamento do cromossomo de *Xylella fastidiosa*, Anexo página 72).

O ponto negativo do seqüenciamento somente por “shotgun” é a montagem de regiões com grande número de repetições de um motivo de nucleotídeos ou de disposição correta de repetições de regiões grandes do genoma, com diferenças sutis. Entretanto, com o desenvolvimento cada vez maior dos programas de montagem de seqüências voltados para grandes volumes de informações como projetos genoma a cada dia tais problemas tem se tornado menores.

Ainda foram utilizadas outras técnicas como seqüenciamento a partir de biblioteca de *Xylella fastidiosa* em fagos lambda. Embora nosso laboratório não tenha se envolvido com tal técnica, as informações recebidas de outros grupos do projeto mostraram que essa técnica parece sofrer dos mesmos problemas que as bibliotecas de cosmídeos, com uma dificuldade adicional que é a preparação de DNA em grande escala a partir de fagos lambda. Mesmo assim, tal estratégia foi de extrema valia para a finalização do genoma devido ao fechamento de “gaps” por diversos clones de lambda seqüenciados pelo grupo do Prof. Dr. Paulo Arruda (UNICAMP – Campinas, CC).

Concluindo-se, os problemas das técnicas de clonagem para seqüenciamento genômico estão diretamente relacionados ao tamanho dos insertos nos diversos vetores. Enquanto que em cosmídeos e fagos lambda a possibilidade de clonagem de insertos grandes pode favorecer o quimerismo, os insertos pequenos das bibliotecas de “shotgun” podem causar problemas de montagem devido a repetições. A melhor resposta à questão de qual seria a mais apropriada é a utilização complementar de técnicas que utilizem insertos pequenos e grandes. Assim, a obtenção de grande quantidade de seqüências de bibliotecas de “shotgun” seguida do seqüenciamento de alguns cosmídeos especificamente selecionados para a confirmação da montagem obtida parece ser uma das melhores opções.

No caso específico do Projeto Genoma *Xylella fastidiosa*, a principal razão de se realizar o seqüenciamento de cosmídeos, que se mostrou como seu principal ponto favorável a curto prazo, foi o desenvolvimento dos pesquisadores envolvidos nas técnicas relacionadas à confecção de bibliotecas de “shotgun” de cosmídeos e na utilização dos programas de bioinformática relacionados a montagem de insertos, assim como determinação de ORFs e anotação das mesmas. Tal aprendizado refletiu-se nos outros projetos de seqüenciamento financiados pela FAPESP em andamento, onde o material humano decorrente do Projeto Genoma *Xylella fastidiosa* está sendo plenamente utilizado.

V B. Considerações sobre o Metabolismo de Carboidratos de *Xylella fastidiosa*

Muitas das questões relacionadas a como *Xylella fastidiosa* vive num ambiente cuja concentração de nutrientes é baixa e como e porquê ela forma agregados nos vasos do xilema devem estar intimamente relacionados com seu metabolismo.

Dentre os transportadores identificados há permeases para glicose e galactose, as proteínas do sistema PTS, proteínas de transporte ativo de açúcares do tipo ABC (ATP-binding cassette) e proteínas de transporte de glicerol. Há também enzimas de transporte para α -cetoglutarato e ácidos orgânicos de 4 carbonos, o que evidencia a provável utilização dos ácidos málico e cítrico que estão em grande quantidade na seiva bruta¹⁴, como mencionado na introdução. A presença de um transportador específico para glutamato, além de outras proteína de transporte de aminoácidos, também sugere a utilização do combinado glutamato/glutamina, cuja concentração na seiva bruta é a maior dentre os compostos orgânicos, como provável fonte de carbono, nitrogênio e energia como discutido a seguir.

Vamos aqui fazer uma análise das vias metabólicas relacionadas a utilização dos carboidratos como fonte de carbono e energia, fazendo uma comparação entre as vias centrais do metabolismo de carboidratos (Glicólise, Gliconeogênese, Ciclo dos Ácidos Tricarboxílicos, Ciclo do Glioxilato e Via das Pentoses) de *Xylella fastidiosa* e de alguns outros organismos cujos genomas foram seqüenciados e anotados. Para isso, utilizando o serviço de busca de vias funcionais do banco de dados COG³⁹, foram desenhados mapas de vias do metabolismo de carboidratos (Glicólise⁴⁴, Gliconeogênese⁴⁵ e Via de Entner-Doudoroff⁴⁶, Anexo, página 75; Ciclo dos Ácidos Tricarboxílicos⁴⁷ e Ciclo do Glioxilato⁴⁸, Anexo, página 82; e Via das Pentoses⁴⁹, Anexo, página 88).

A utilização de carboidratos, principalmente glicose, como fonte de carbono e energia é praticamente universal entre as bactérias, arqueobactérias e eucariotos. A forma como tais carboidratos são metabolizados pode entretanto diferir consideravelmente⁵⁰.

A rota metabólica presente na quase totalidade das bactérias modernas e eucariotos para o metabolismo de monossacarídeos, provindos diretamente do meio ambiente ou da degradação dos diversos tipos de carboidratos, é a glicólise, ou via de Embden-Meyerhoff-Parnas (Anexo, página 75). Os substratos iniciais de tal rota são geralmente glicose-6-fosfato ou frutose-6-fosfato provenientes do sistema PTS de transporte, da conversão desses próprios monossacarídeos aos derivados fosforilados por ação de quinases, ou pela conversão de outros monossacarídeos fosforilados (galactose a glicose-6-fosfato, manose a frutose-6-fosfato). O produto final da rota é o piruvato que constitui um dos principais intermediários do metabolismo central devido a sua utilização em diversas vias, tanto catabólicas como anabólicas.

Essa onipresença levou inicialmente a pensar-se que seria uma via metabólica proveniente dos primeiros microorganismos. Entretanto, em vista da falta de identificação de genes codificando enzimas dessa via nos genomas de arqueobactérias (fica evidente a inviabilidade da via para o metabolismo de carboidratos nos mapas da Glicólise, Gliconeogênese e Entner-Doudoroff referentes às arqueobactérias, Anexo, página 79), tal visão não é mais aceita e já há trabalhos sugerindo que a evolução da glicólise a partir da gliconeogênese, cujo papel anabólico seria mais antigo do que o catabólico da glicólise⁵⁰.

Nos organismos em que a glicólise não está presente, e mesmo em grande parte dos organismos que a utilizam como principal via de degradação de monossacarídeos, existe uma via metabólica, ou variações da mesma, denominada via de Entner-Doudoroff. Essa via é mostrada nos mapas de Glicólise, Gliconeogênese e Entner-Doudoroff, com glicose-6-fosfato como substrato inicial, seguindo para glucono-1,5-lactona-6-fosfato, 6-fosfogluconato e 2-desidro-3-desoxifosfogluconato. Normalmente é utilizada em bactérias para o metabolismo de glucoronato, com um balanço energético líquido de um mol de ATP por mol de glucoronato (ou glicose, dependendo do substrato inicial)⁵⁰.

Em arqueobactérias sacarolíticas, variações da via de Entner-Doudoroff que têm balanço energético nulo são as rotas de metabolismo de monossacarídeos mais presentes. As variações consistem em vias com alguns ou todos os intermediários desfosforilados⁵⁰. Assim, explica-se o fato de nos mapas baseados no COG algumas arqueobactérias aparentemente não terem como metabolizar monossacarídeos, pois tanto a glicólise como a via de Entner-Doudoroff tradicional estão incompletas (Anexo, página 79).

Foram descobertas, entretanto, na arqueobactéria hipertemofílica *Pyrococcus furiosus* tanto uma hexoquinase como uma fosfofrutoquinase dependentes de ADP (formadoras de AMP), o que indica a utilização da glicólise como a principal via de metabolismo de glicose, como comprovado por estudos de atividade enzimática e marcação com carbono ^{13}C . Tal descoberta, reforça a idéia de que a simples comparação de seqüências de proteínas homólogas é um mecanismo preliminar para a identificação de vias metabólicas, mas que não deve ser tomado como totalmente certo, visto a possibilidade da existência de enzimas análogas que desempenhem funções conhecidas de forma completamente diferente da já descrita.

Além de vias alternativas à glicólise, dentro dessa própria via há pelo menos dois casos já descritos de reações realizadas por enzimas não-similares: a conversão de frutose-1,6-bisfosfato em gliceraldeído-3-fosfato e diidroxiacetonafosfato; e a mudança da posição fosforilada entre 3- e 2-fosfoglicerato. A primeira reação pode ser realizada por dois tipos diferentes de frutose-1,6-bisfosfato aldolase, enquanto que a segunda reação pode ser realizada por três enzimas diferentes (Anexo, página 75).

Foram identificadas no genoma de *Xylella fastidiosa* ORFs codificando para as enzimas mais comuns na glicólise e também para as enzimas da versão fosforilada da via de Entner-Doudoroff, o que está de acordo com sua similaridade com *Escherichia coli* (Anexo, página 76). Assim, havendo carboidratos disponíveis, a obtenção de piruvato está de certa forma garantida por tais vias.

A partir do piruvato, pode ocorrer então sua redução a lactato (fermentação láctea); descarboxilação a acetaldeído, seguida de redução a etanol (fermentação alcoólica); ou formação de acetil-CoA. Foram identificadas em *Xylella fastidiosa* tanto ORFs codificando para lactato desidrogenase quanto várias para álcool desidrogenase, mas não foi identificada ORF codificando para a piruvato descarboxilase. Assim, visto que estão presentes as proteínas da cadeia de transporte de elétrons, com oxigênio como acceptor final, a importância da lactato desidrogenase pode estar mais relacionada à utilização de lactato como fonte de carbono, já que ela também pode realizar a reação de oxidação do lactato a piruvato. O fato de não ter sido identificada uma ORF codificando a piruvato descarboxilase inviabilizaria a fermentação alcoólica.

A obtenção de acetil-CoA no mundo microbiano também pode ser realizada por duas formas distintas: através do complexo da piruvato desidrogenase ou do complexo piruvato:ferredoxina oxidoreductase. O primeiro está ausente nas arqueobactérias, sendo considerado uma evolução do segundo quando as eubactérias mais modernas passaram a utilizar como aceptores de elétrons cofatores como NAD^+ , nesse caso, e FAD^{43} . Muitas das bactérias modernas entretanto ainda têm os genes codificantes para o segundo complexo. O complexo da piruvato desidrogenase foi identificado em *Xylella fastidiosa*, completando assim a via catabólica dos carboidratos até o ciclo dos ácidos tricarboxílicos.

O ciclo dos ácidos tricarboxílicos (TCA) é dentre as vias comparadas a que tem um maior número de etapas realizadas por grupos de enzimas não-similares. A descarboxilação oxidativa do α -cetoglutarato (oxoglutarato), a subsequente hidrólise do tioéster succinil-CoA com obtenção de ligação fosfato rica em energia, a hidratação do fumarato e a oxidação do malato a oxaloacetato são, cada uma, realizadas por duas enzimas que não possuem similaridade de seqüência (Anexo, página 82).

No TCA há uma distinção clara evolutiva entre as enzimas utilizadas em cada etapa. Enquanto que as arqueobactérias geralmente têm as etapas acima realizadas pelas enzimas menos comuns: oxoglutarato:ferredoxina oxidoreductase, succinil-CoA--acetoacetato-CoA transferase, fumarase classe I e malato desidrogenase arqueobacteriana; tais enzimas são menos usadas entre as eubactérias e não foram identificadas no genoma de levedura (eucarioto) (Anexo, página 87)⁴¹. Além dessas diferenças já identificadas, na maioria dos organismos comparados o TCA é a via com maior número de enzimas cujos genes não foram identificados nos genomas. Visto a importância anaplerótica do TCA (além dos elétrons obtidos na forma de NADH e FADH_2 serem necessários para a fosforilação oxidativa, os intermediários do ciclo são ponto de partida para uma série de reações de biossíntese) também é de supor que possam haver vários “non-homologous gene displacements”⁴¹ entre as enzimas dessa via.

Foram identificadas em *Xylella fastidiosa* todas as enzimas do TCA relacionadas às eubactérias mais modernas e aos eucariotos (Anexo, página 83). Tal fato, de certa forma é um pouco surpreendente, visto que fora *Escherichia coli*, *Mycobacterium tuberculosis* e *Bacillus subtilis*, as demais bactérias não tiveram todos os genes codificando para as enzimas do TCA identificados.

Também pode-se considerar que a presença de todas as enzimas do TCA talvez reflita a necessidade que *Xylella fastidiosa* tenha de sintetizar todos os intermediários, uma vez que está em meio pobre nesses compostos, como exceção do ácido cítrico e de α -cetoglutarato, devido a quantidade desse ácido e de glutamato/glutamina na seiva bruta¹⁴ (o metabolismo de tais aminoácidos fornece α -cetoglutarato).

A via do glioxilato, presente principalmente nas plantas e cujas enzimas, isocitrato liase e malatosintase, tiveram seus genes identificados em *Escherichia coli*, *Mycobacterium tuberculosis* (Anexo, página 83) e *Saccharomyces cerevisiae* (Anexo, página 87), tem um papel importante como uma rota que ultrapassa as etapas de descarboxilação do TCA e fornece grande quantidade de substratos gliconeogênicos (malato/oxaloacetato) a partir de acetil-CoA e conseqüentemente da beta-oxidação dos ácidos graxos. Nos eucariotos tal reação se desenvolve numa organela específica denominada glioxissomo²⁵.

Não foram identificadas ORFs para nenhuma das duas enzimas do ciclo do glioxilato em *Xylella fastidiosa* (Anexo, página 83), o que está de acordo com os fatos de também não terem sido identificados genes que codifiquem para proteínas envolvidas na beta-oxidação e da seiva bruta não conter ácidos graxos¹⁵. Entretanto foram identificados várias lipases e esterases, que estariam relacionadas à degradação de triacilgliceróis (Anexo, página 104). Talvez tais enzimas sejam secretadas pela bactéria quando esta habita os canais alimentícios dos insetos vetores, dessa forma obtendo glicerol que pode ser absorvido e metabolizado.

Completando as vias catabólicas de carboidratos há a via das pentoses, cujos principais objetivos são a produção de potencial redutor biossintético (NADPH) e ribose-5-fosfato, utilizada na síntese dos ácidos nucléicos (Anexo, página 88). Com base em comparações de seqüências das enzimas da via, foi sugerido que as arqueobactérias provavelmente possuem uma via alternativa para a obtenção de ribose-5-fosfato devido à quase ausência total de enzimas similares às conhecidas (Anexo, página 90)⁴².

Por ser uma via parcialmente integrada à glicólise (tem como substrato e alguns produtos, intermediários da glicólise) e as reações da transaldolase e transcetolase serem reversíveis, a ausência de enzimas na parte oxidativa da via (de glicose-6-fosfato a ribulose-5-fosfato) não prejudicaria a obtenção de ribose-5-fosfato a partir de frutose-6-fosfato e gliceraldeído-3-fosfato, o que poderia estar acontecendo em vários organismos se confirmadas as ausências principalmente da fosfogluconato desidrogenase em *Aquifex aeolicus* e *Helicobacter pylori* (Anexo, página 89).

A via das pentoses de *Xylella fastidiosa* (Anexo, página 89) está incompleta do ponto de vista de enzimas com seqüências similares às mais comuns. Não foram identificadas ORFs cujos produtos fossem similares à fosfogluconato desidrogenase e a nenhuma transaldolase. Dessa forma seria impossível utilizar tal rota para a obtenção de ribose-5-fosfato. Uma vez que tal composto é indispensável para a vida da bactéria, poderiam ser levantadas duas questões: ou uma das duas enzimas está presente devido a um “non-homologous gene displacement”⁴¹, ou *Xylella fastidiosa* teria uma via alternativa diferente para a obtenção de ribose-5-fosfato, como hipotetizado para as arqueobactérias⁴². Se a segunda hipótese se mostrasse verdadeira, poderia estar relacionada ao tempo elevado de duplicação da bactéria, que seria explicado por uma deficiência na síntese de ácidos nucleicos.

A utilização de glicose-6-fosfato para outros fins biossintéticos, ao invés de catabólicos, em condições em que a glicose é escassa, só é conseguido através da gliconeogênese. A síntese da parede celular das bactérias é uma das rotas biossintéticas que necessita de glicose-6-fosfato, que pela ação de uma fosfoglicomutase passa a glicose-1-fosfato e segue na via de síntese. A enzima frutose-1,6-bisfosfatase tem um papel chave na gliconeogênese visto que para suplantar a desfosforilação oxidativa do fosfoenolpiruvato a piruvato as bactérias possuem em sua maioria a fosfoenolpiruvato sintase, além de algumas terem as enzimas presentes em eucariotos: piruvato carboxilase e fosfoenolpiruvato carboxiquinase. Entretanto, dentre as enzimas da gliconeogênese, frutose-1,6-bisfosfatase é a que tem COG com menor número de enzimas⁴⁵.

Formam o COG de frutose-1,6-bisfosfatase somente as enzimas de *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori* e *Synechocystis* sp. Esse pequeno número de enzimas num COG e o fato de todas as demais bactérias com genoma seqüenciado terem a fosfoenolpiruvato sintase leva a crer que talvez haja uma classe de bisfosfatase ainda não identificada cuja similaridade de seqüência às já descritas seja muito pequena.

Como nas demais bactérias, não foi identificada uma ORF de *Xylella fastidiosa* que codificasse uma frutose-1,6-bisfosfatase similar às presentes no COG referente a essa enzima, embora estejam presentes ORFs para a enzima málica e a fosfoenolpiruvato sintase.

A tendência inicial é considerar que seja um caso de “non-homologous gene displacement”⁴¹. Entretanto, se confirmada a real ausência da enzima em *Xylella fastidiosa*, esse também poderia ser um outro possível fato explicativo do elevado tempo de duplicação da bactéria. Visto que, sem uma gliconeogênese funcional, *Xylella fastidiosa* não poderia dispor de outros substratos para a síntese da parede celular que não os próprios carboidratos que absorvesse, haveria uma constante competição para a utilização dos carboidratos, diminuindo a taxa de síntese da parede celular e aumentando o tempo de duplicação.

Considerando-se o metabolismo dos carboidratos como um todo, temos que a identificação de ORFs codificando todas as enzimas da glicólise e do TCA difere bastante do que é normalmente observado nos patógenos de animais, nos quais normalmente não são identificadas algumas enzimas e várias vezes chega-se a sugerir que a ausência dessas enzimas possa ser explicada pela obtenção dos intermediários metabólicos a partir do próprio hospedeiro.

Considerando uma linha comparativa, visto que já foram sequenciadas e anotadas duas bactérias próximas a *Xylella fastidiosa*: *Escherichia coli* e *Haemophilus influenzae*, podemos ter que o fato de terem sido identificadas as ORFs das enzimas da maioria das vias metabólicas pode somente estar relacionado a uma maior similaridade entre as proteínas de *Xylella fastidiosa* e as já presentes no banco de dados, salvo as exceções discutidas anteriormente.

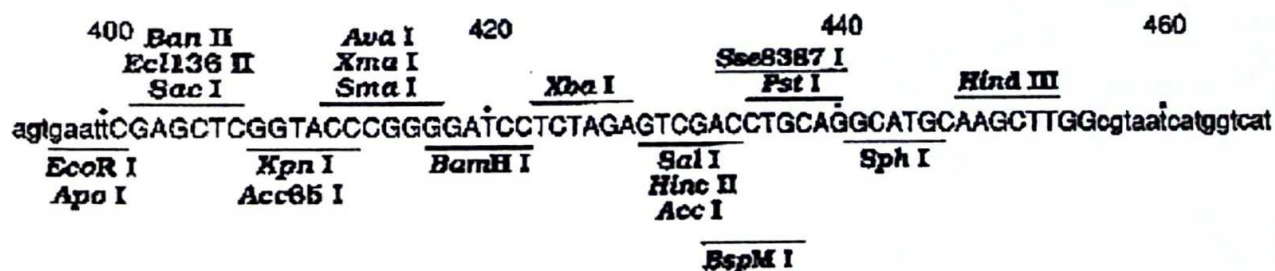
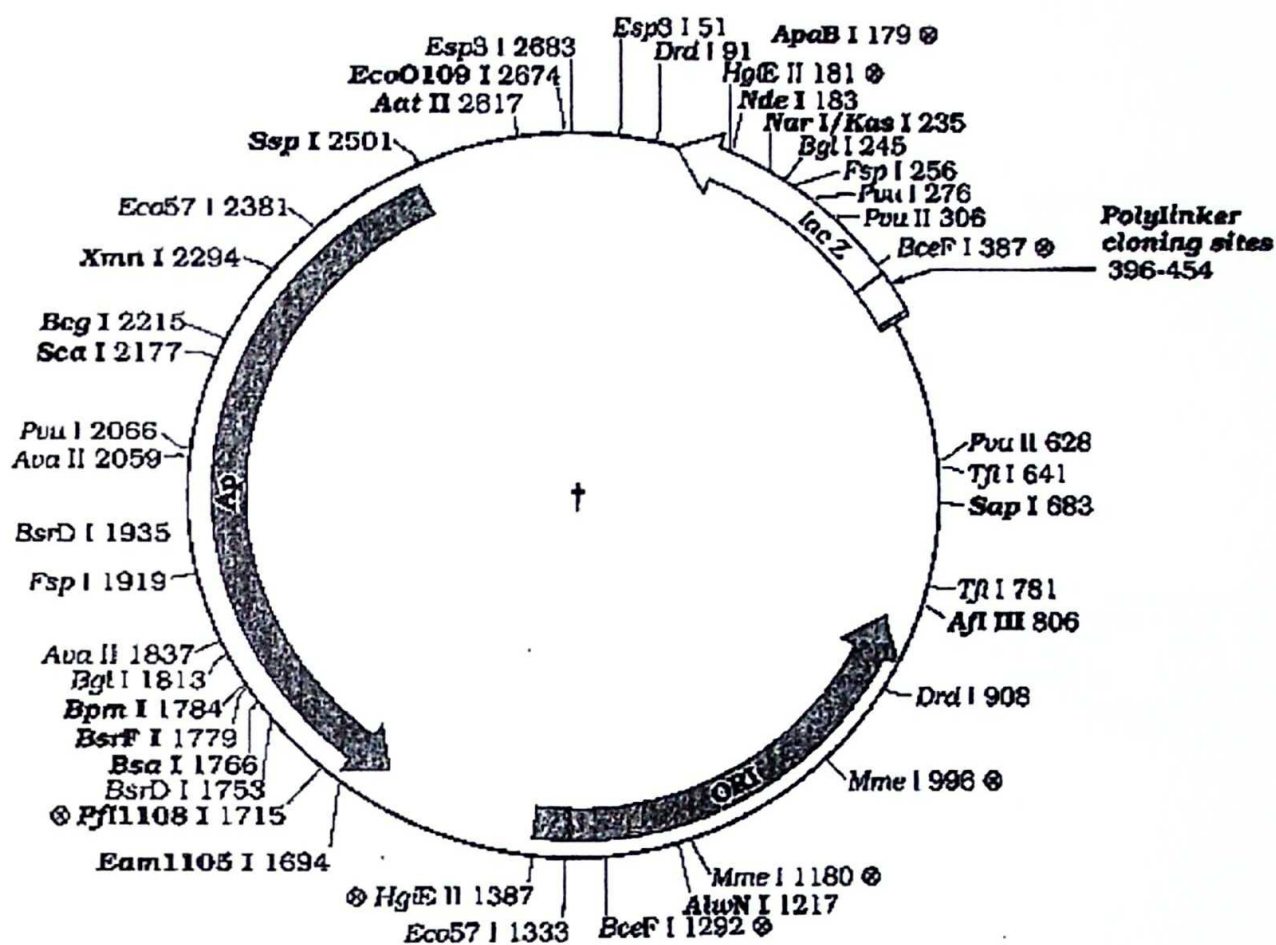
De outra forma poderíamos inferir que pelo xilema ser um meio pobre em comparação aos habitats de patógenos animais, *Xylella fastidiosa* necessita do metabolismo central completo pois precisa sintetizar senão todos, pelo menos a grande maioria dos intermediários. Tal visão pode ser reforçada em vista do fato de não terem sido identificadas enzimas de degradação para todos os aminoácidos, mas várias peptidases, de o mecanismo de “salvage” de nucleotídeos ser aparentemente bem apurado e de não terem sido identificadas proteínas do complexo de beta-oxidação de ácidos graxos.

VI. Anexo

Estão presentes nesse anexo: o mapa de restrição do plasmídeo pUC18, utilizado na clonagem de fragmentos das bibliotecas de “shotgun” genômico e de cosmídeos, mapas de disposição de clones seqüenciados e de cobertura do seqüenciamento do cromossomo de *Xylella fastidiosa*, mapas metabólicos comparativos entre diversos organismos com genomas seqüenciados e anotados e os relatórios de estágio e término do seqüenciamento de cada cosmídeo e da GFS (“gap-filling sequence”). Também são apresentadas digestões dos cosmídeos com enzimas de restrição e as correspondentes digestões eletrônicas do cosmídeo com a seqüência do inserto obtida, devidamente orientada no vetor. As informações referentes aos cosmídeos estão agrupadas em ordem alfabética do nome do cosmídeo, e não em ordem cronológica, como no texto da dissertação.

Ao final do anexo estão relacionadas, em uma lista subdividida hierarquicamente, as anotações da categoria I de Metabolismo Intermediário e uma lista das proteínas de transporte relacionadas a carboidratos, ácidos orgânicos e álcoois presentes no banco de dados de *Xylella fastidiosa* em Fevereiro de 2000. Por fim há o resultado de um BLASTp da ORF codificando a frutose-1,6-bisfosfato aldolase de *Xylella fastidiosa*.

VI A. Mapa do plasmídeo pUC18



Primers utilizados no seqüenciamento dos insertos clonados no sítio de *SmaI* do vetor pUC18 das bibliotecas de "shotgun" genômico e de cosmídeos de *Xylella fastidiosa*:

m13 Universal Sequence Primer:
5'-d[G T A A A A C G A C G G C C A G T]-3'

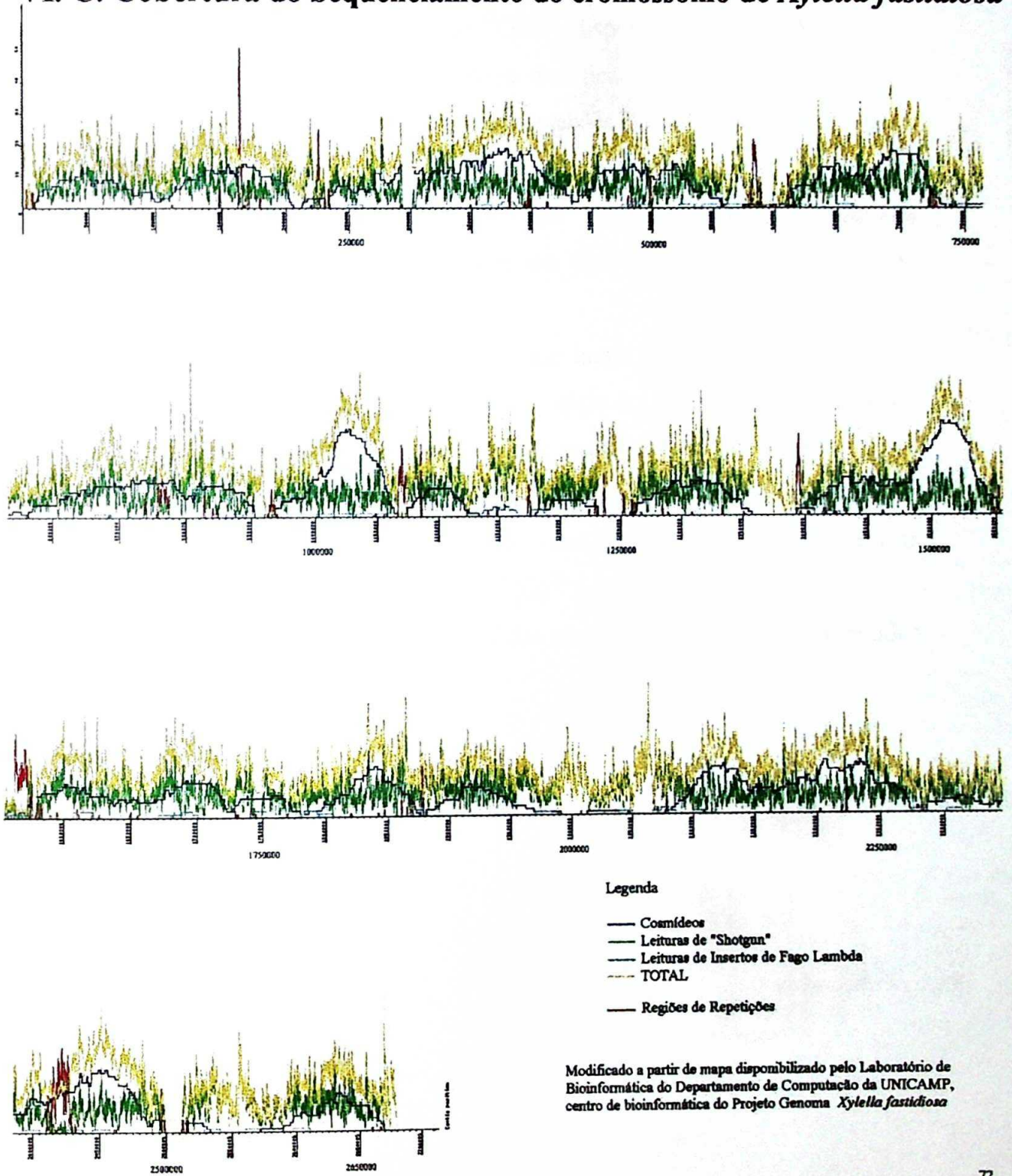
m13 Reverse Sequence Primer:
5'-d[C A G G A A A C A G C T A T G A C]-3'

VI.B Disposição dos clones sequenciados no cromossomo de *Xylella fastidiosa*

Escala — 15000 pb



VI. C. Cobertura do Seqüenciamento do cromossomo de *Xylella fastidiosa*



VI D. Mapas Metabólicos Comparativos

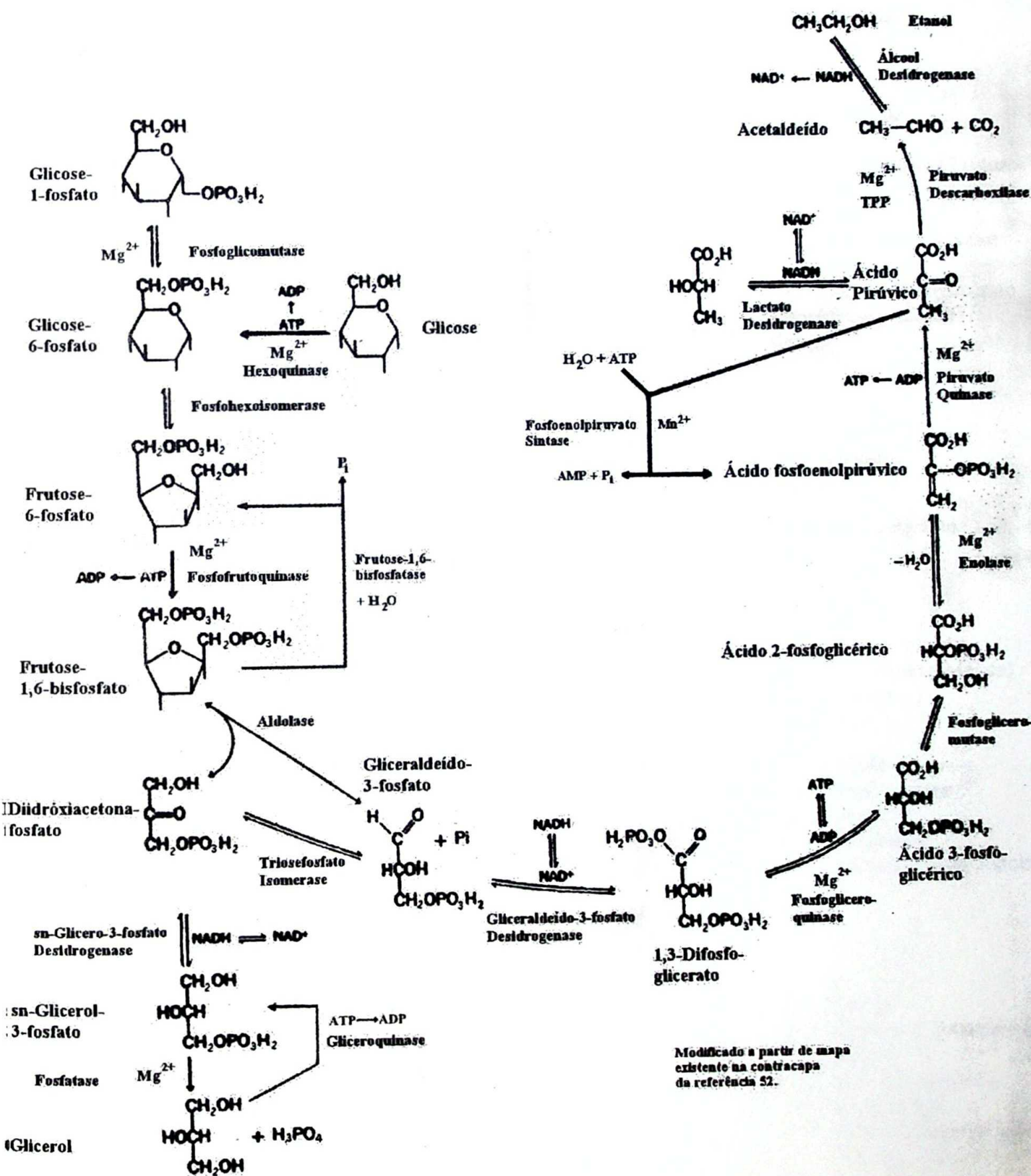
Seguem os esquemas de mapas metabólicos baseados no serviço de vias metabólicas e funcionais do banco de dados de COGs³⁹, disponibilizado pelo NCBI.

Primeiramente são apresentadas, em mapa modificado a partir do existente na referência 51, a glicólise, gliconeogênese e as fermentações láctea e alcoólica. A seguir é mostrado um esquema genérico da glicólise, gliconeogênese e via de Entner-Doudoroff fosforilada que é utilizado como base para os esquemas seguintes representando as enzimas ortólogas identificadas em diversos organismos (bactérias, arqueobactérias e levedura).

Da mesma forma que para a glicólise, há um mapa apresentando o ciclo dos ácidos tricarboxílicos (TCA), ou ciclo de Krebs, e o ciclo do glioxilato (modificado a partir do existente na referência 52). Seguindo-o há o esquema genérico desses ciclos e os esquemas derivados para diversos organismos (bactérias, arqueobactérias e levedura). Os esquemas para o TCA e ciclo do glioxilato foram baseados no serviço de vias metabólicas e funcionais do banco de dados de COGs³⁹ e na referência 41.

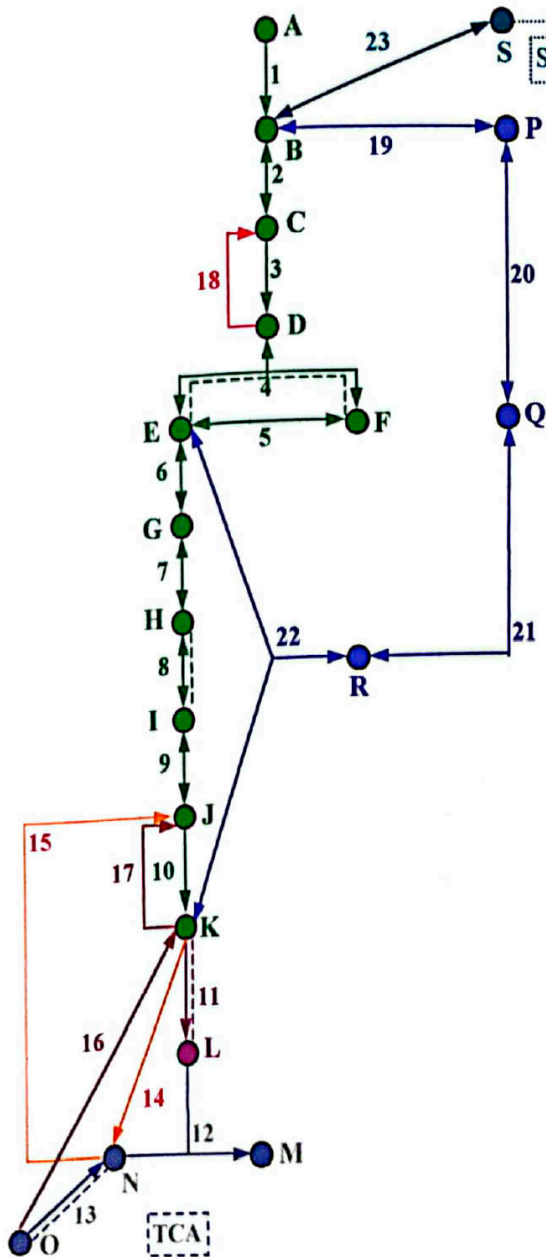
Por fim, há um esquema genérico da via das pentoses e os esquemas derivados para os diversos organismos.

Glicólise, Fermentação e Gliconeogênese



Modificado a partir de mapa existente na contracapa da referência 52.

Glicólise, Glioneogênese, Entner-Doudoroff



Enzimas (linha contínua / tracejada)

Glicólise

- 1 Glucoquinase
- 2 Glucose-6-fosfato Isomerase
- 3 Fosfofrutoquinase
- 4 Frutose-1,6-bisfosfato Aldolase / Frutose-1,6-bisfosfato Aldolase Dhna
- 5 Triosefosfato Isomerase
- 6 Gliceraldeído-3-fosfato Desidrogenase
- 7 3-Fosfoglicerato Quinase
- 8 Fosfoglicerato Mutase 1 / Fosfoglicerato Mutase (Fructose-2,6-bisfosfatase) / Fosfogliceromutase
- 9 Enolase
- 10 Piruvato Quinase
- 11 Piruvato Desidrogenase / Piruvato Ferredoxin Oxidoreductase
- 12 Citrato Sintase
- 13 Malato Desidrogenase Bacteriana / Archeobacteriana

Gliconeogênese

- 14 Piruvato Carboxilase
- 15 Fosfoenolpiruvato Carboxiquinase
- 16 Malato Oxidoreductase (Enzima Málica)
- 17 Fosfoenolpiruvato Sintase
- 18 Frutose-1,6-bisfosfatase

Entner-Doudoroff Fosforilada

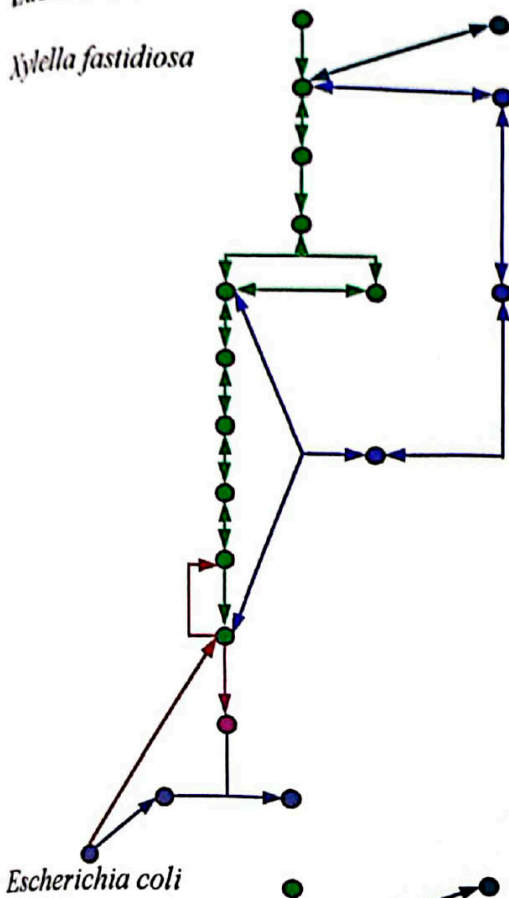
- 19 Glucose-6-fosfato Desidrogenase
- 20 6-Fosfogluconolactonase
- 21 6-Fosfogluconato Desidratase
- 22 2-Desidro-3-desoxifosfogluconato Aldolase
- 23 Fosfoglicomutase

Substâncias

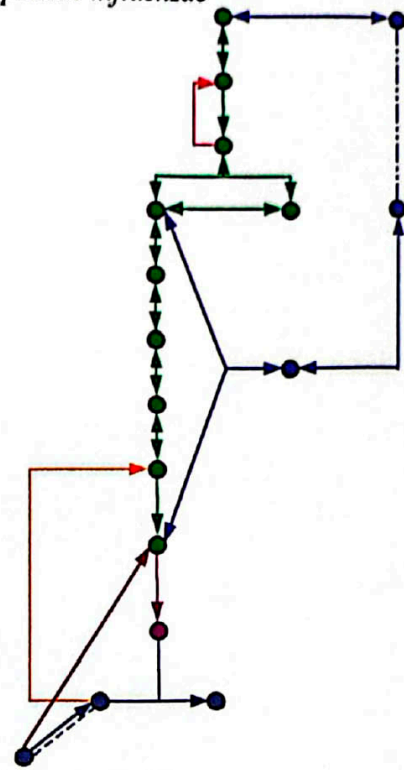
A	Glicose	H	3-Fosfoglicerato	O	Malato
B	Glicose-6-fosfato	I	2-Fosfoglicerato	P	Glucono-1,5-lactona-6-fosfato
C	Frutose-6-fosfato	J	Fosfoenolpiruvato	Q	6-Fosfogluconato
D	Frutose-1,6-bisfosfato	K	Piruvato	R	2-Desidro-3-desoxifosfogluconato
E	Gliceraldeído-3-fosfato	L	Acetil-CoA	S	Glucose-1-fosfato
F	Diidroxiacetonafosfato	M	Citrato		
G	1,3-Bisfosfoglicerato	N	Oxaloacetato		

Eubactérias

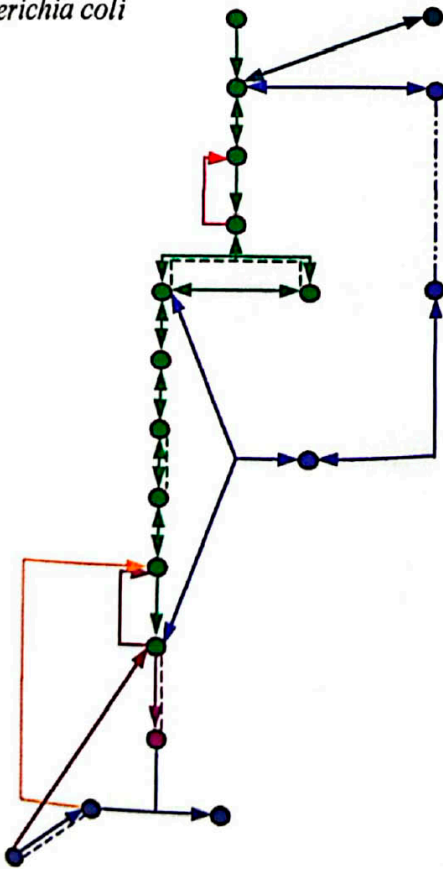
Xylella fastidiosa



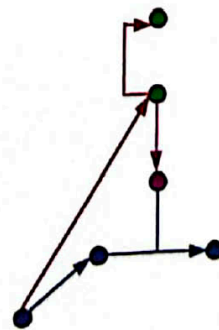
Haemophilus influenzae



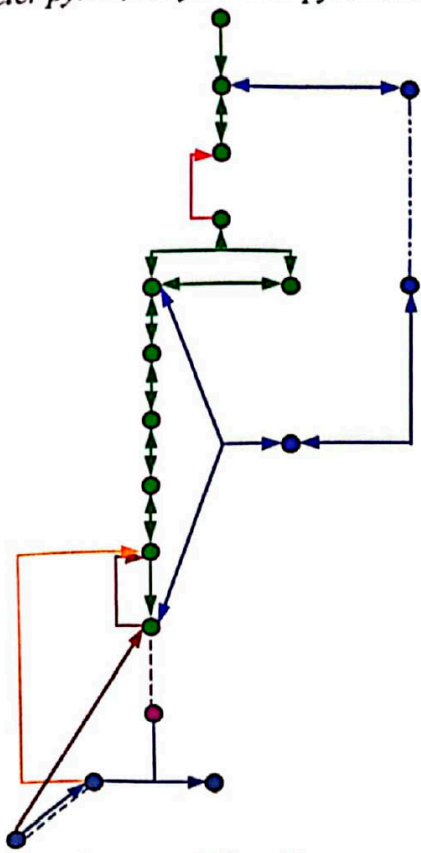
Escherichia coli



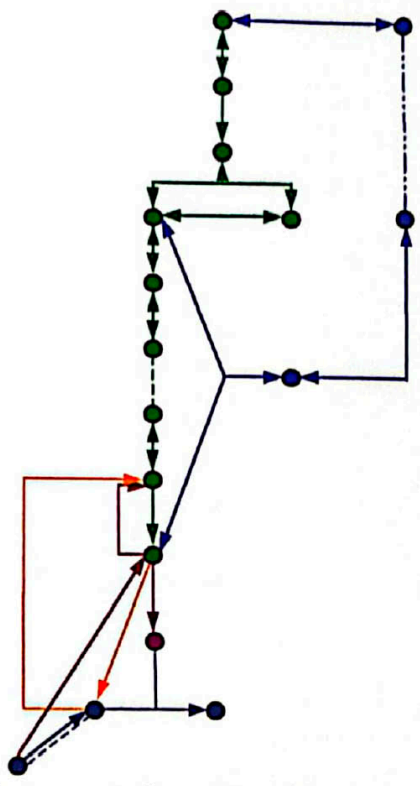
Rickettsia prowazekii



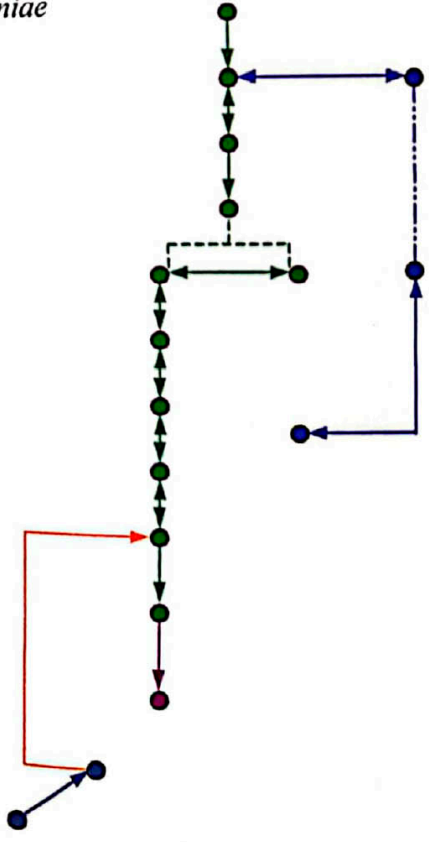
Helicobacter pylori, Helycobacter pylori J99



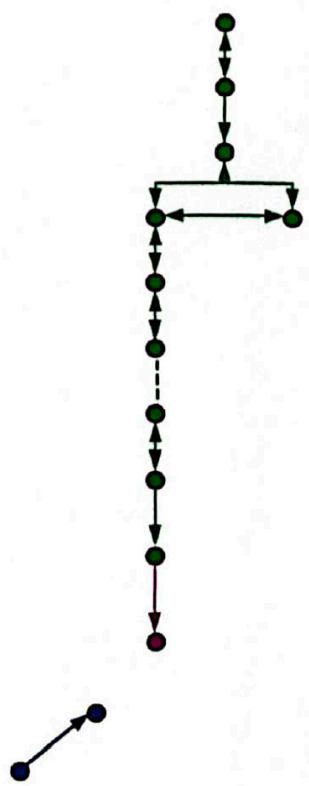
Bacillus subtilis



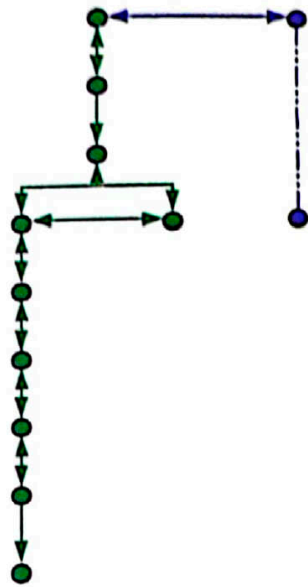
Chlamydia trachomatis, Chlamydia pneumoniae



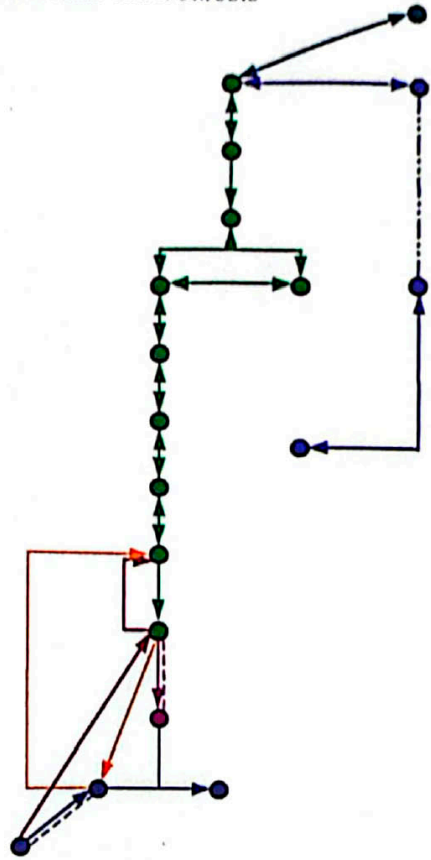
Mycoplasma genitalium, Mycoplasma pneumoniae



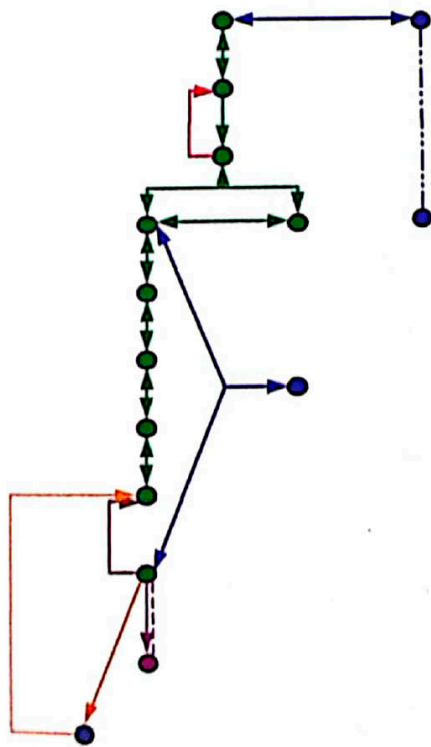
Borrelia burgdorferi



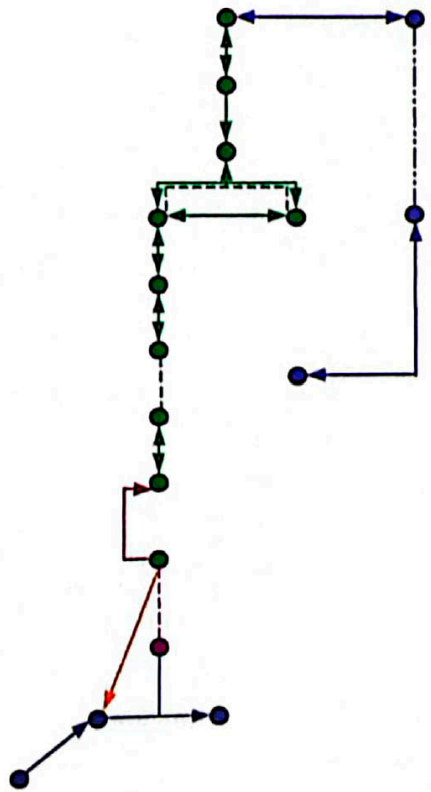
Mycobacterium tuberculosis



Treponema pallidum

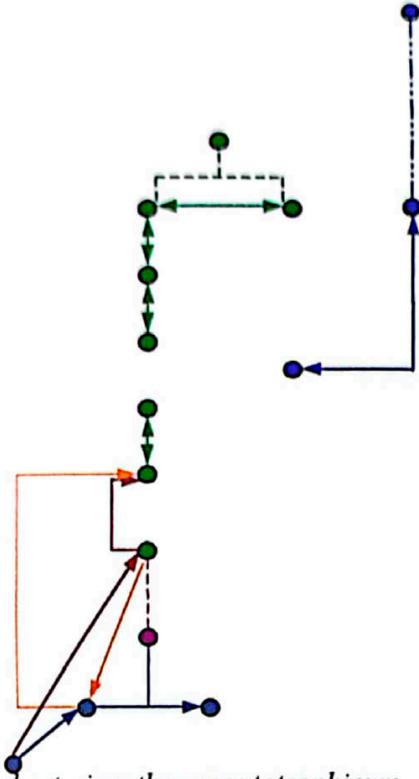


Aquifex aeolicus



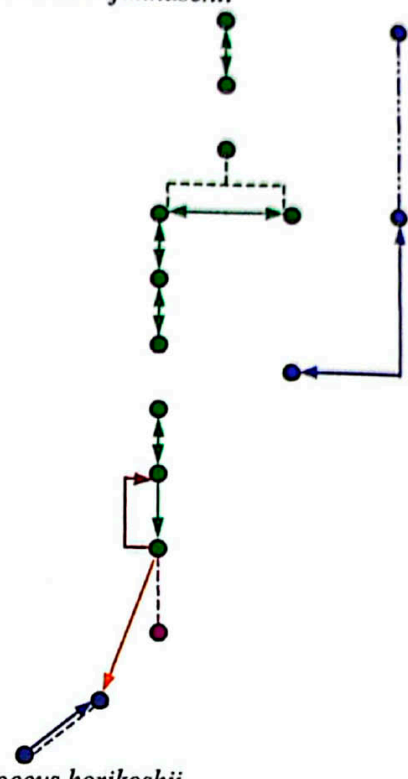
Arqueobactérias

Archeoglobus fulgidus

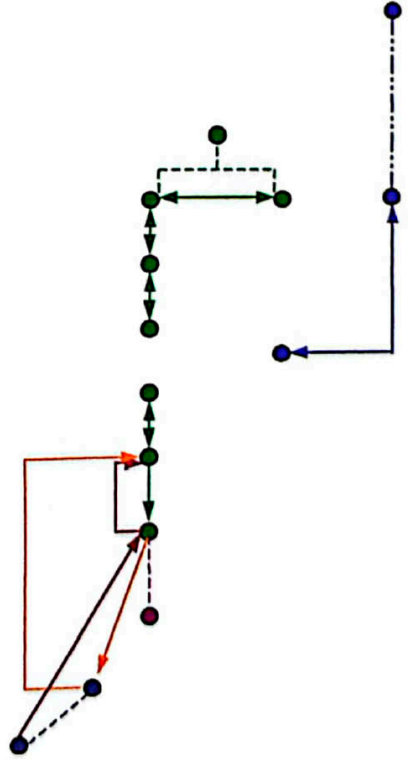
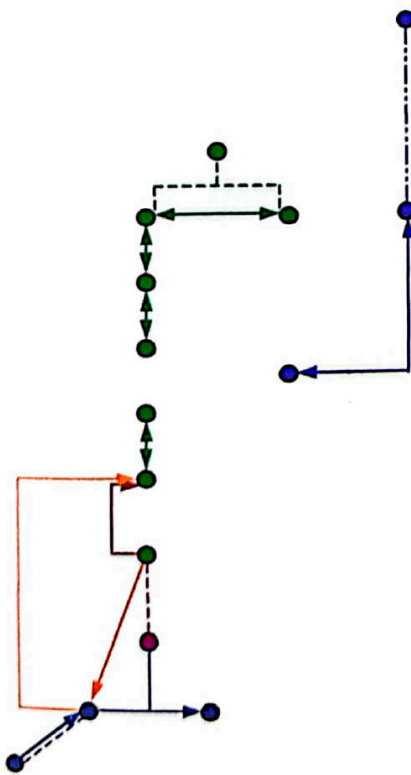


Methanobacterium thermoautotrophicum

Methanococcus jannaschii

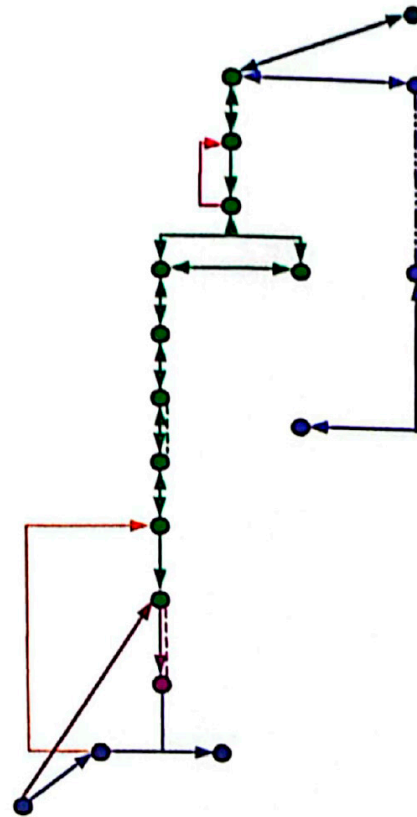


Pyrococcus horikoshii

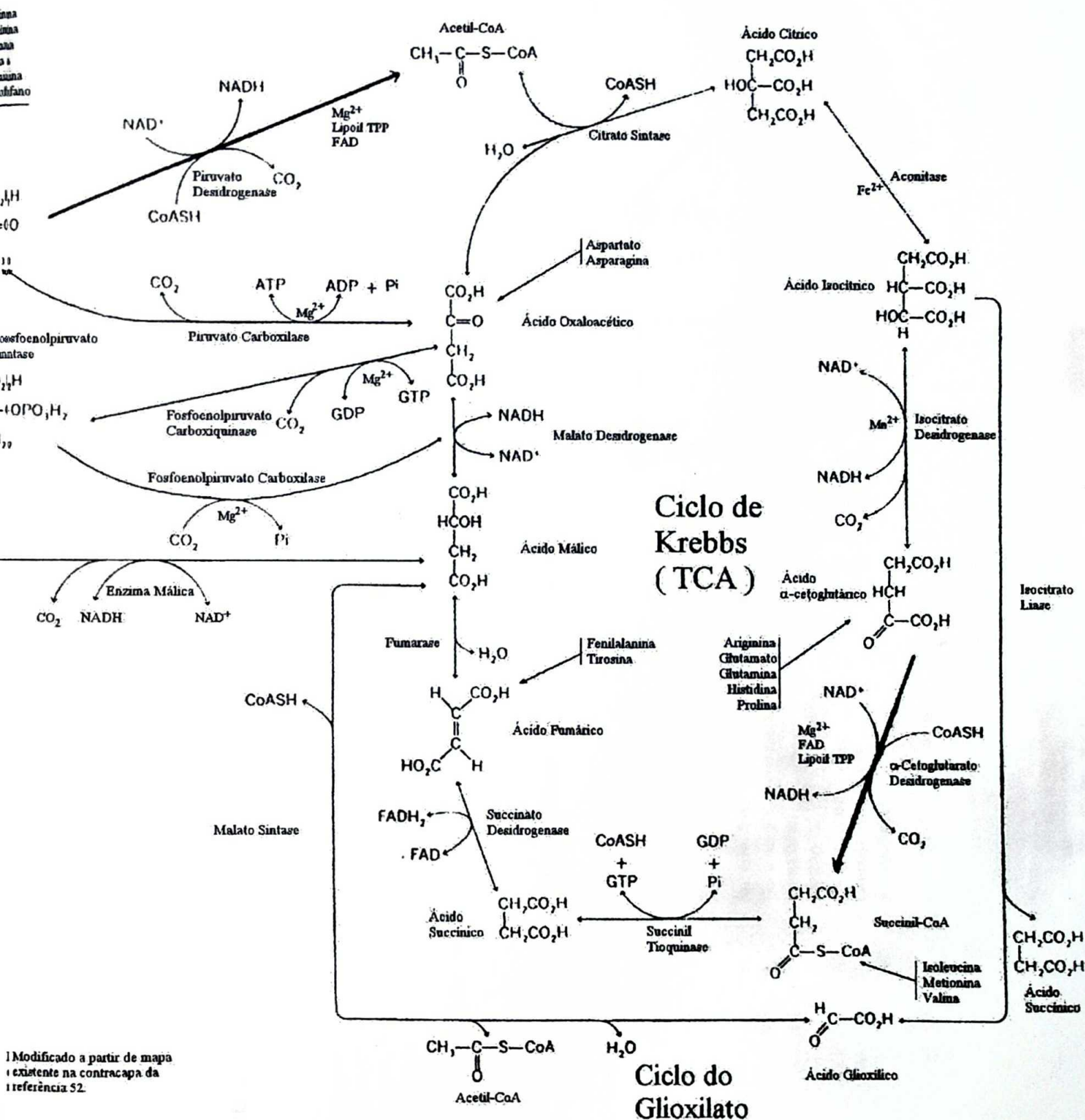


Eucarioto unicelular

Saccharomyces cerevisiae



Os mapas desenhados foram baseados no serviço de buscas de vias funcionais do banco de dados COG³⁹.



Modificado a partir de mapa existente na contracapa da referência 52.

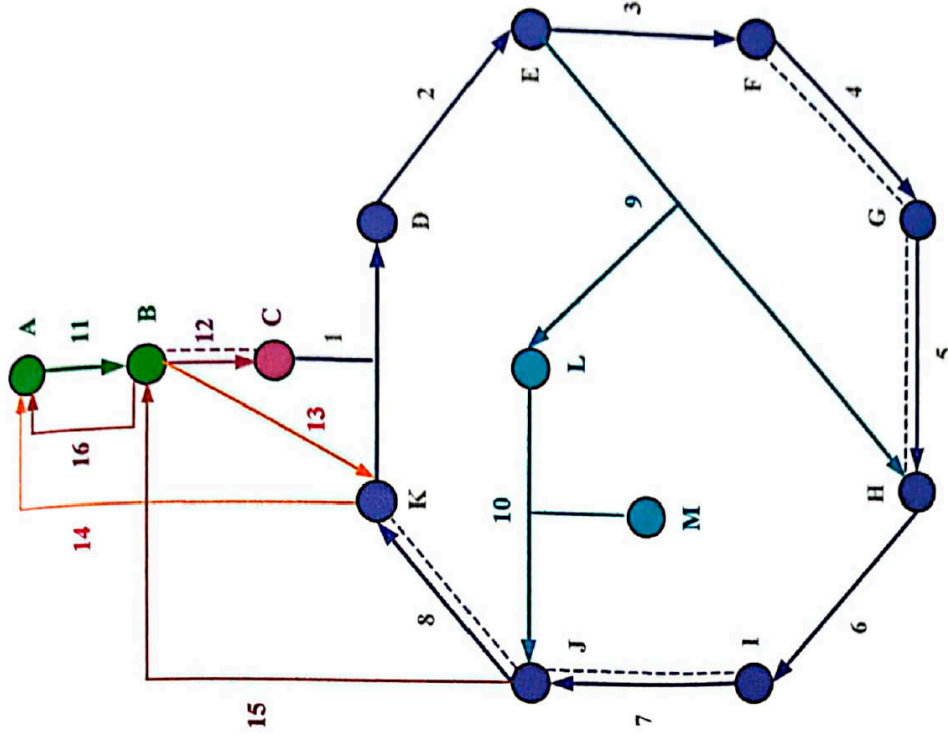
Ciclo dos Ácidos Tricarboxílicos, Via do Glioxilato

Enzimas (linha continua / tracejada)

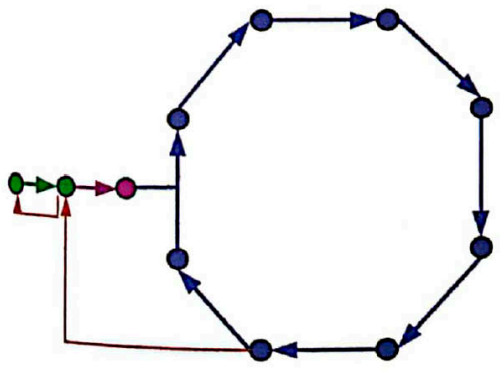
- 1 Citrato Sintase
- 2 Aconitase
- 3 Isocitrato Desidrogenase
- 4 Oxoglutarato Desidrogenase / Oxoglutarato Ferredoxin Oxidoreductase
- 5 Succinil-CoA Sintetase / Succinil-CoA-acetoacetato-CoA transferase
- 6 Succinato Desidrogenase
- 7 Fumarase Classe II / Classe I
- 8 Malato Desidrogenase Bacteriana / Arqueabacteriana
- 9 Isocitrato Liase
- 10 Malato Sintase
- 11 Piruvato Quinase
- 12 Piruvato Desidrogenase / Piruvato Ferredoxin Oxidoreductase
- 13 Piruvato Carboxilase
- 14 Fosfoenolpiruvato Carboxiquinase
- 15 Malato Oxidoreductase (Enzima Málica)
- 16 Fosfoenolpiruvato Sintase

Substâncias

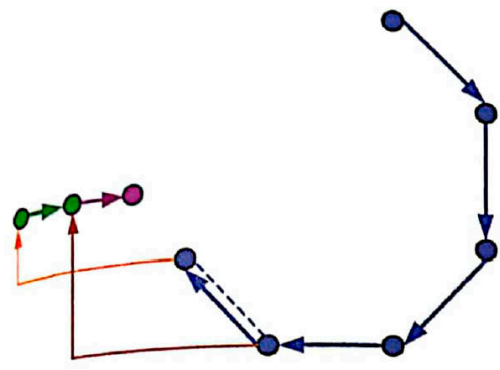
A	Fosfoenolpiruvato	G	Succinil-CoA
B	Piruvato	H	Succinato
C	Acetil-CoA	I	Fumarato
D	Citrato	J	Malato
E	Isocitrato	K	Oxaloacetato
F	Oxoglutarato	L	Glioxilato
		M	Acetil-CoA



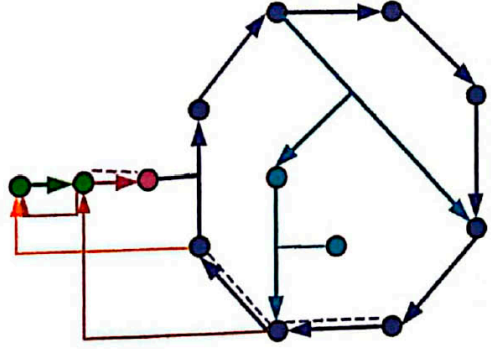
Eubactérias
Xylella fastidiosa



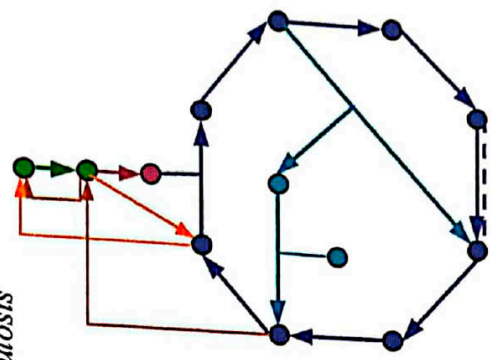
Haemophilus influenzae

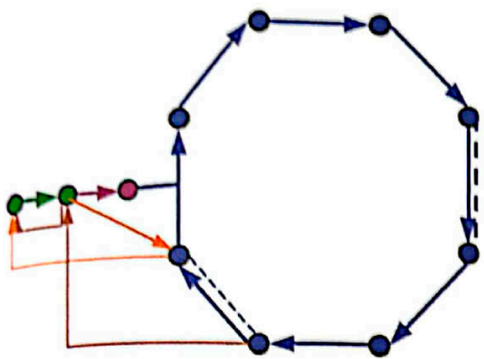


Escherichia coli

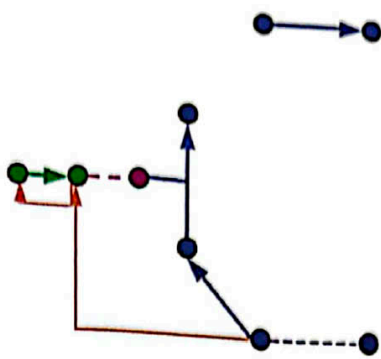


Mycobacterium tuberculosis

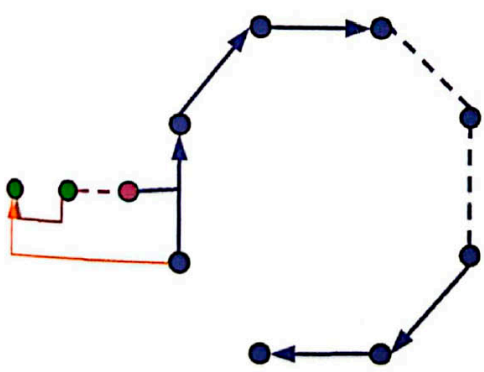




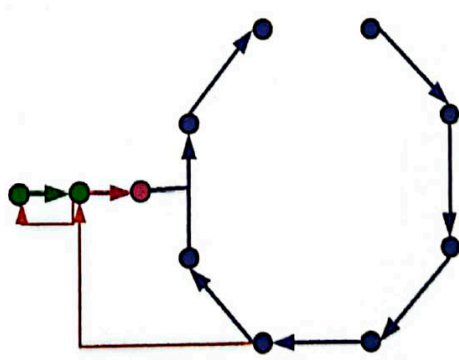
Bacillus subtilis



Thermotoga maritima

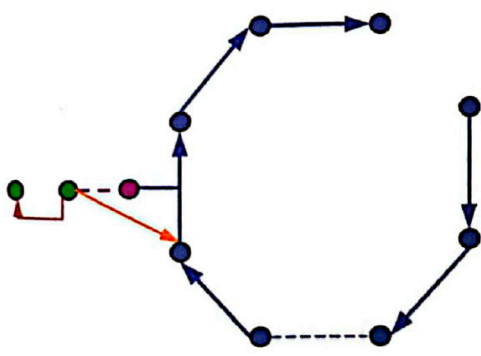


Helicobacter pylori,
Helicobacter pylori J99

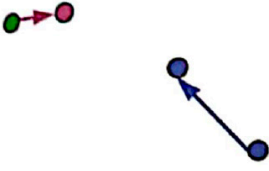


Rickettsia prowazekii

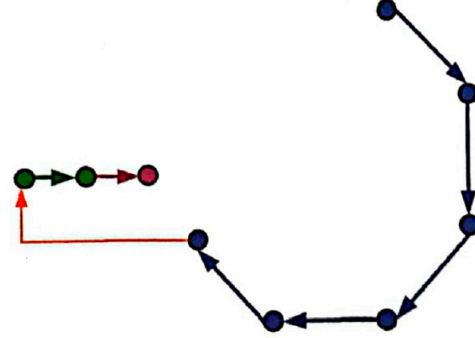
Aquifex aeolicus



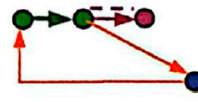
Mycoplasma genitalium,
Mycoplasma pneumoniae;



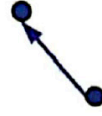
Chlamydia trachomatis,
Chlamydia pneumoniae



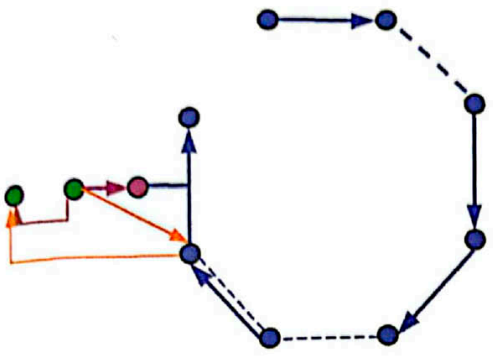
Treponema pallidum



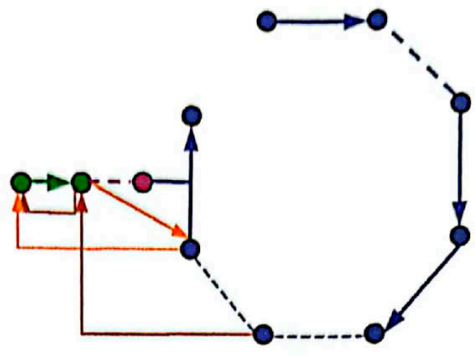
Borrelia burgdorferi



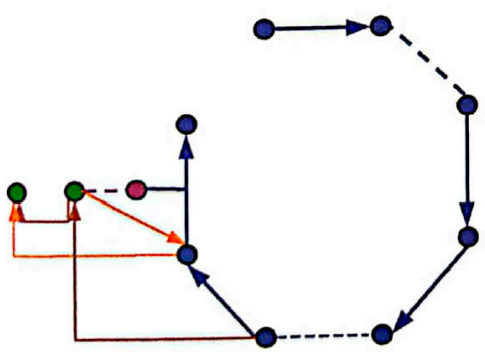
*Methanobacterium
thermoautotrophicum*



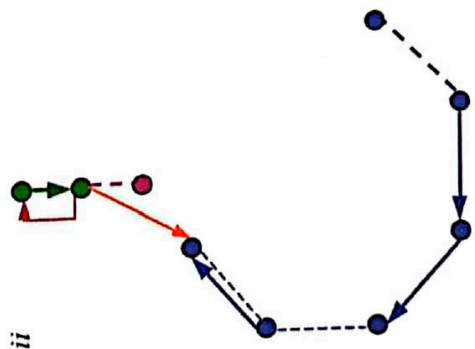
Pyrococcus horikoshii



Arqueabactérias
Archeoglobus fulgidus

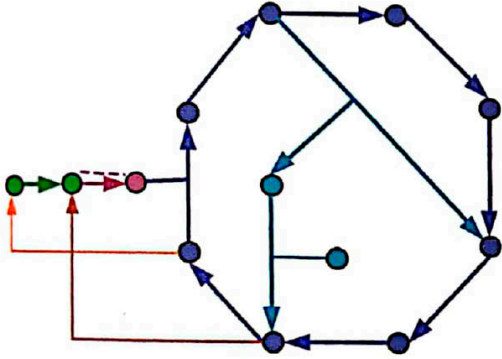


Methanococcus jannaschii



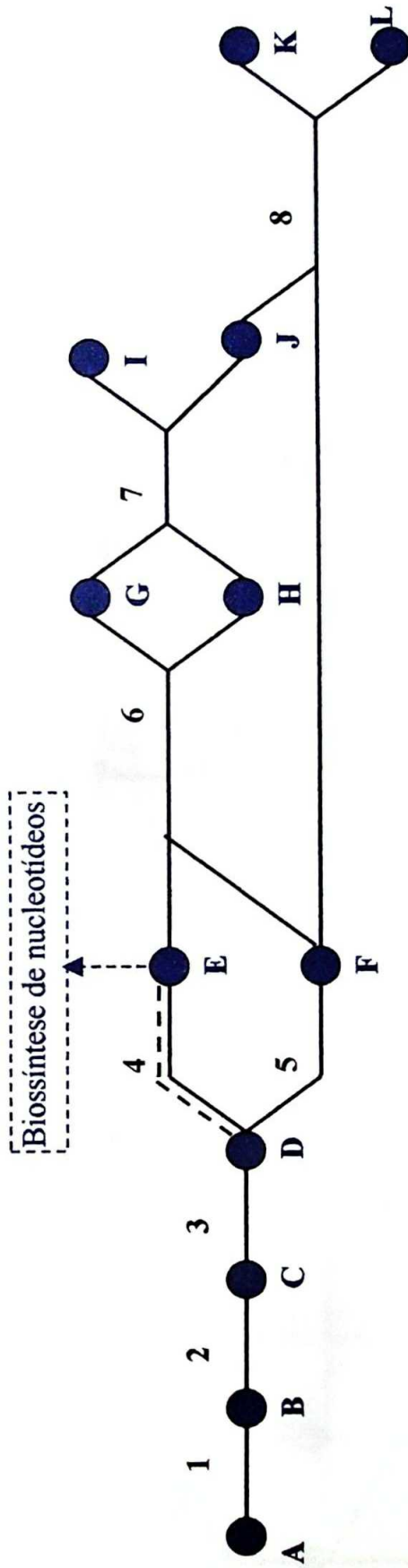
Eucarioto unicelular

Saccharomyces cerevisiae



Todos os mapas desenhado foram baseados nos dados da referência 41 e no serviço de busca de vias funcionais do banco de dados COG³⁹.

Via das Pentoses



Enzimas (linha contínua / tracejada)

- | | |
|---|---|
| 1 | Glicose-6-fosfato Desidrogenase |
| 2 | 6-Fosfogluconolactonase * |
| 3 | Fosfogluconato Desidrogenase |
| 4 | Ribulose-5-fosfato Isomerase / Isomerase RpiB |
| 5 | Ribulose-5-fosfato Epimerase |
| 6 | Transcetolase |
| 7 | Transaldolase |
| 8 | Transcetolase |

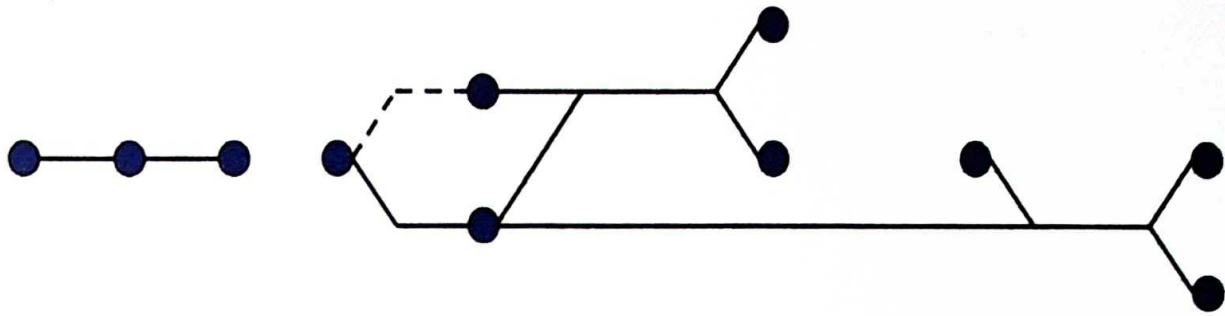
*Obs: não há COG para essa enzima, mas visto que a reação catalisada é espontânea e de velocidade relativamente rápida assumiu-se que estando presente a 6-fosfogluconato- δ -lactona, também estará o 6-fosfogluconato. Tal assumpção está marcada pela linha tracejada em vez da linha contínua da enzima.

Substâncias

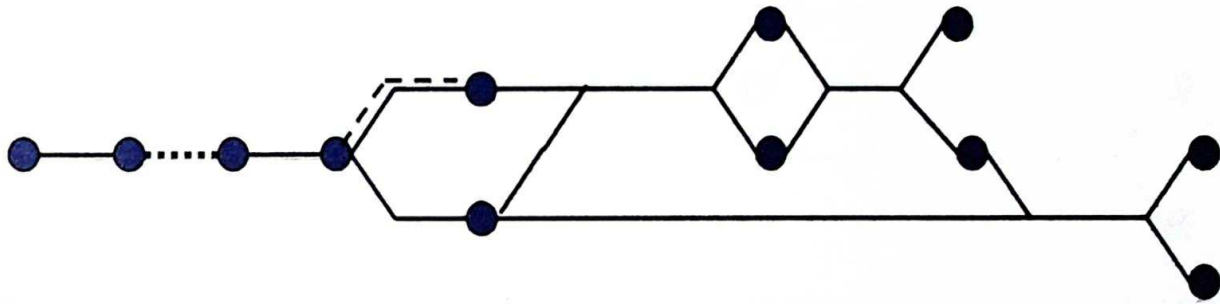
- | | |
|---|-------------------------------------|
| A | Glicose-6-fosfato |
| B | 6-Fosfogluconato- δ -lactona |
| C | 6-Fosfogluconato |
| D | Ribulose-5-fosfato |
| E | Ribose-5-fosfato |
| F | Xilulose-5-fosfato |
| G | Sedoheptulose-7-fosfato |
| H | Gliceraldeído-3-fosfato |
| I | Frutose-6-fosfato |
| J | Eritrose-4-fosfato |
| K | Frutose-6-fosfato |
| L | Gliceraldeído-3-fosfato |

Eubactérias

Xylella fastidiosa

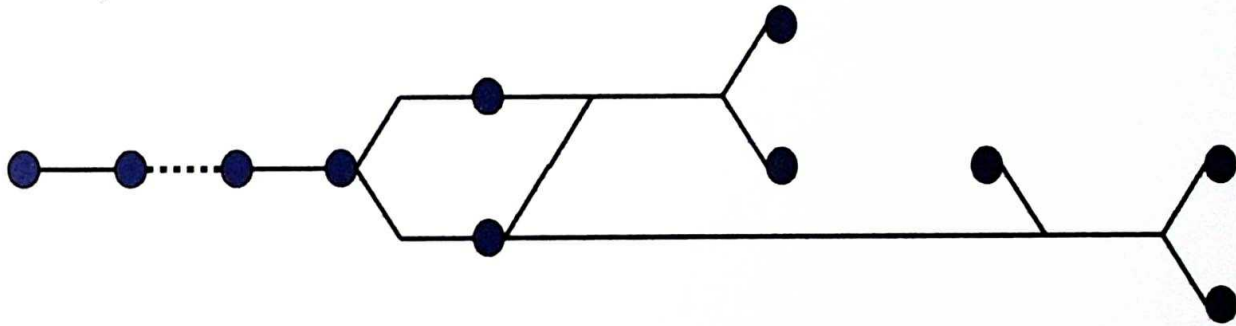


Escherichia coli, *Haemophilus influenzae**, *Bacillus subtilis*, *Mycobacterium tuberculosis*, *Synechocystis* sp., *Thermotoga maritima*, *Chlamydia trachomatis**, *Chlamydia pneumoniae**.

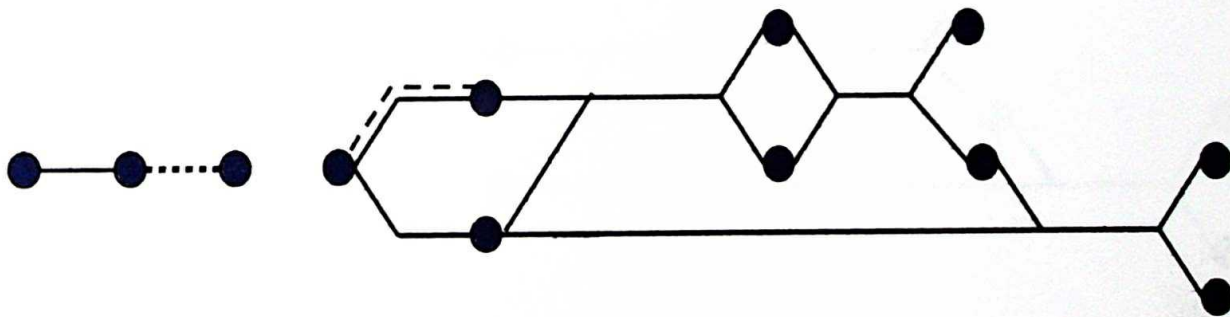


* não possuem a Ribulose-5-fosfato isomerase B

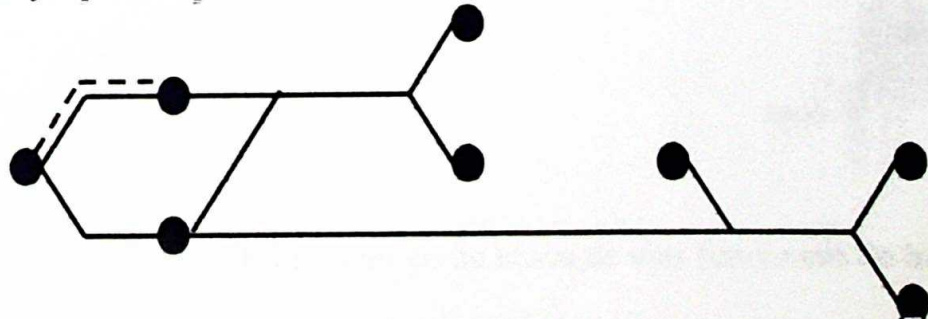
Treponema pallidum



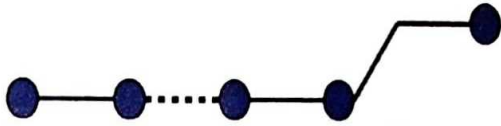
Aquifex aeolicus, *Helicobacter pylori*, *Helicobater pylori* J99



Mycoplasma genitalium, *Mycoplasma pneumoniae*

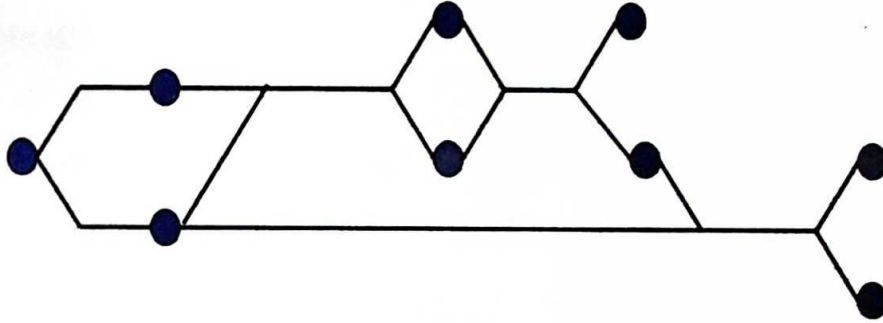


Borrelia burgdorferi

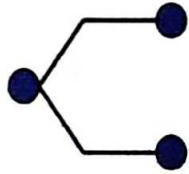


Arqueabactéria

Methanococcus jannaschii

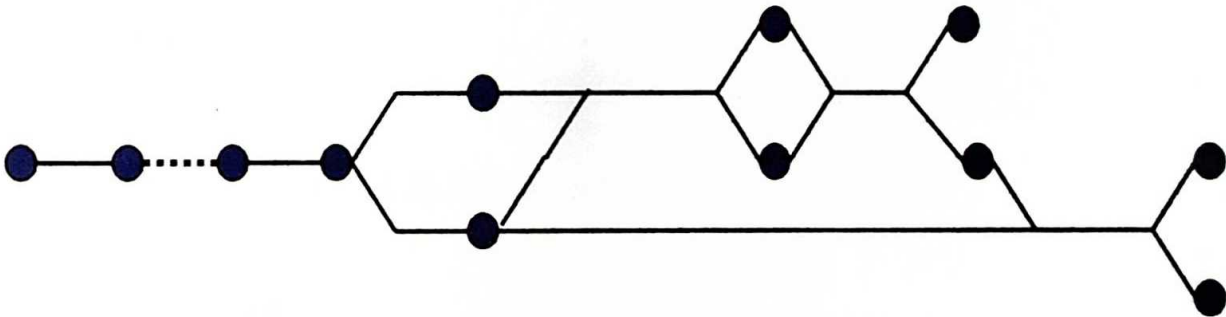


Archaeoglobus fulgidus, *Methanobacterium thermoautotrophicum*, *Pyrococcus horikoshii*



Eucarioto unicelular

Saccharomyces cerevisiae



Todos os mapas desenhados foram baseados no serviço de busca de vias funcionais do banco de dados COG³⁹.

VI E. Cosmídeo X0QH-02E06

VI E 1. Relatório de finalização de seqüenciamento

PRELIMINARY CERTIFICATION REPORT FOR COSMID 02E06

DATE: Thu 04-Mar-1999

READ COMPOSITION: 1212 reads;

QH-02E06 : 611 reads

QH-07A10 : 200 reads

UI-07A04 : 389 reads

plasmid : 12 reads

Contig size: 43567 bp

Insert size: 43113 bp (excludes lawrist and/or cosmid extensions)

LAWRIST LOCATIONS:

Left : first base after LAU1 = 1

Right : last base before LAU2 = 43113

EXPECTED ERROR RATE: 0.08 / 10000 bp

LOW CONSENSUS QUALITY (LCQ) REGIONS:

TOTAL: 0

HIGH QUALITY DISCREPANCIES (HQD):

X0QH-07A10-K081F 28378 - 28378 (1) --close to unaligned region:ok

X0QH-02E06-M109F 33711 - 33711 (1) --close to unaligned region:ok

TOTAL: 0

POSITIONS NOT CONFIRMED ON BOTH STRANDS (NCBS):

TOTAL: 0

CONCLUSION: Cosmid Finished

END OF ANALYSIS

VI E 2. Cronograma de estágios do seqüenciamento

02E06/Thu 04-Mar-1999
Finished

02E06/Thu 25-Feb-1999
Close to Finished

02E06/Thu 04-Mar-1999
Close to Finished

02E06/Wed 24-Feb-1999
One-contig

02E06/Thu 25-Feb-1999
Close to Finished

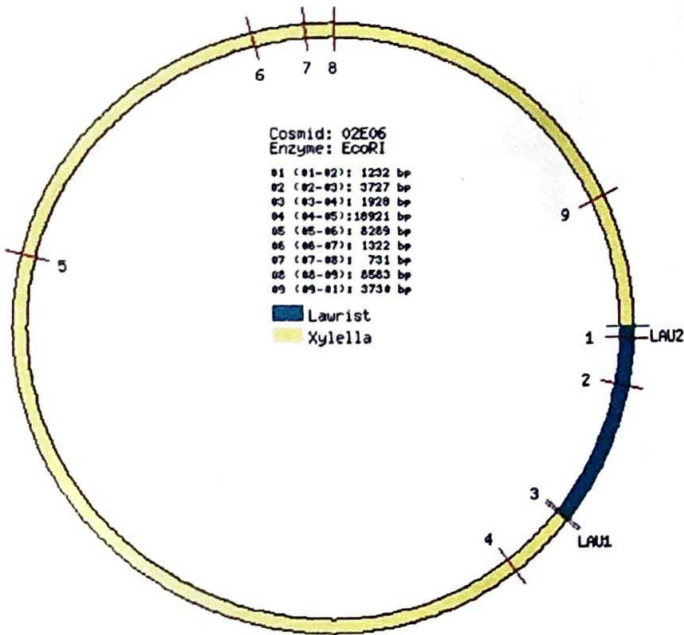
02E06/Wed 24-Feb-1999
One-contig

02E06/Thu 25-Feb-1999
Close to Finished

02E06/08-jan-1999
Submitted

VIE 3. Digestão do cosmídeo X0QH-02E06 com enzimas de restrição e digestão eletrônica

Digestão com *EcoRI* e *HindIII*



VI F. Cosmideo X0QH/QS-02G12

VI F 1. Relatório de finalização do seqüenciamento

PRELIMINARY CERTIFICATION REPORT FOR COSMID 02G12

DATE: Wed 11-Aug-1999 10:16:24

FILE SUBMITTED: X0QH-02G12-VI.zip

READ COMPOSITION: 906 reads;

IL-02G12 : 2 reads
QH-02G12 : 396 reads
QS-02G12 : 394 reads
UV-01A03 : 26 reads
plasmid : 88 reads

Contig size: 41606 bp

Insert size: 41105 bp (excludes lawrist and/or cosmid extensions)

LAWRIST LOCATIONS:

Left : first base after LAU1 = 69

Right : last base before LAU2 = 41173

EXPECTED ERROR RATE: 0.00 / 10000 bp

LOW CONSENSUS QUALITY (LCQ) REGIONS:

TOTAL: 0

HIGH QUALITY DISCREPANCIES (HQD):

TOTAL: 0

POSITIONS NOT CONFIRMED ON BOTH STRANDS (NCBS):

TOTAL: 0

CONCLUSION: Cosmid FINISHED

E. Coli contamination:

None detected.

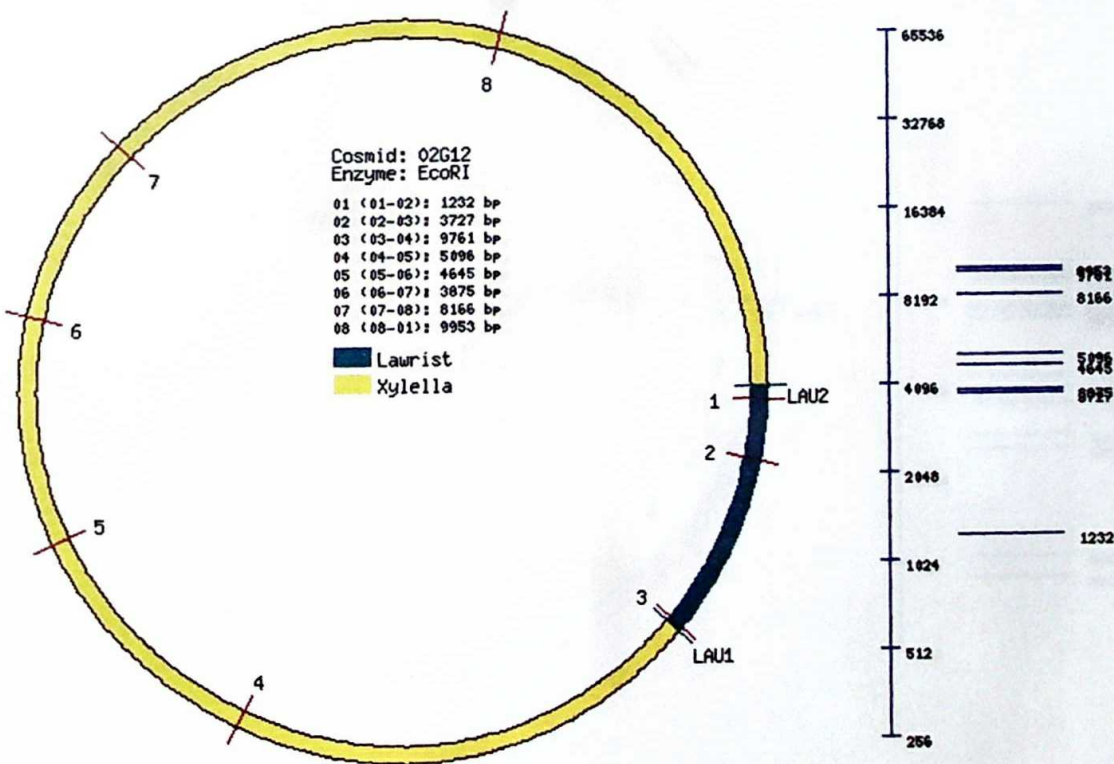
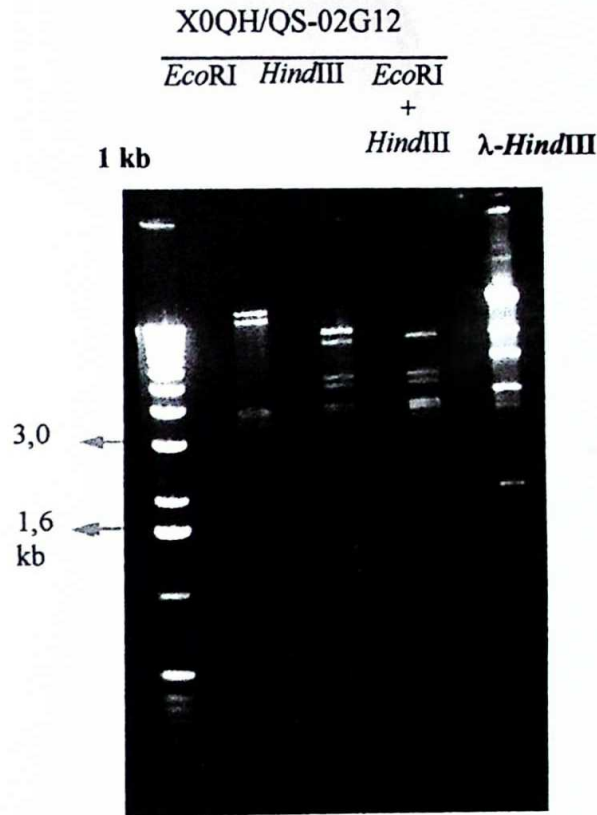
END OF ANALYSIS

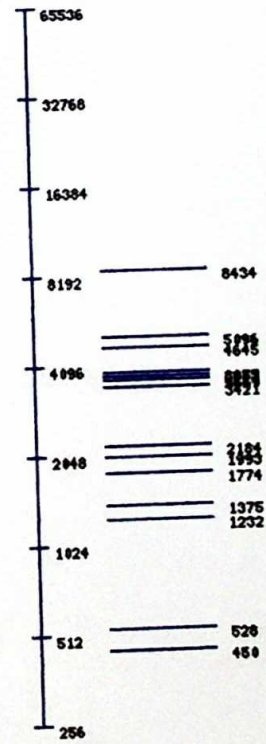
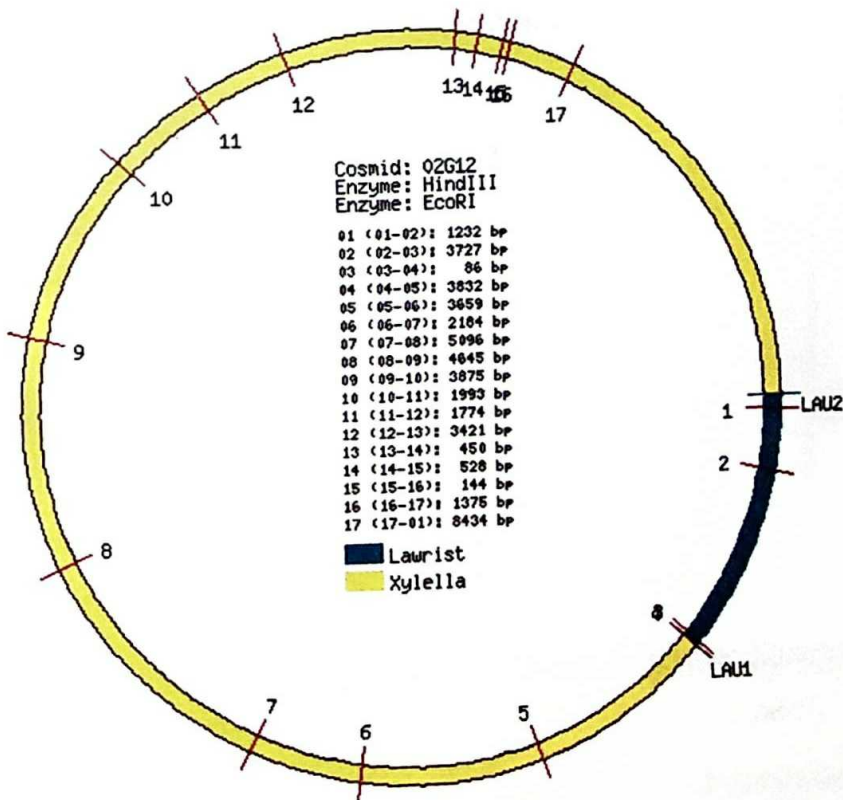
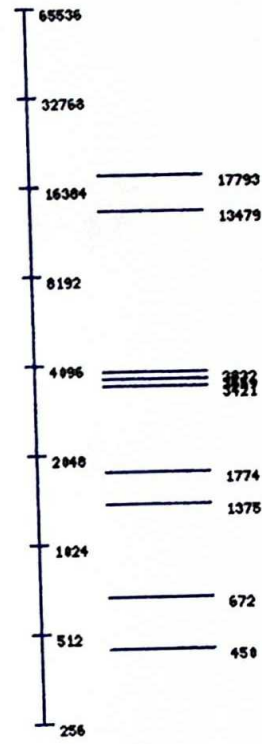
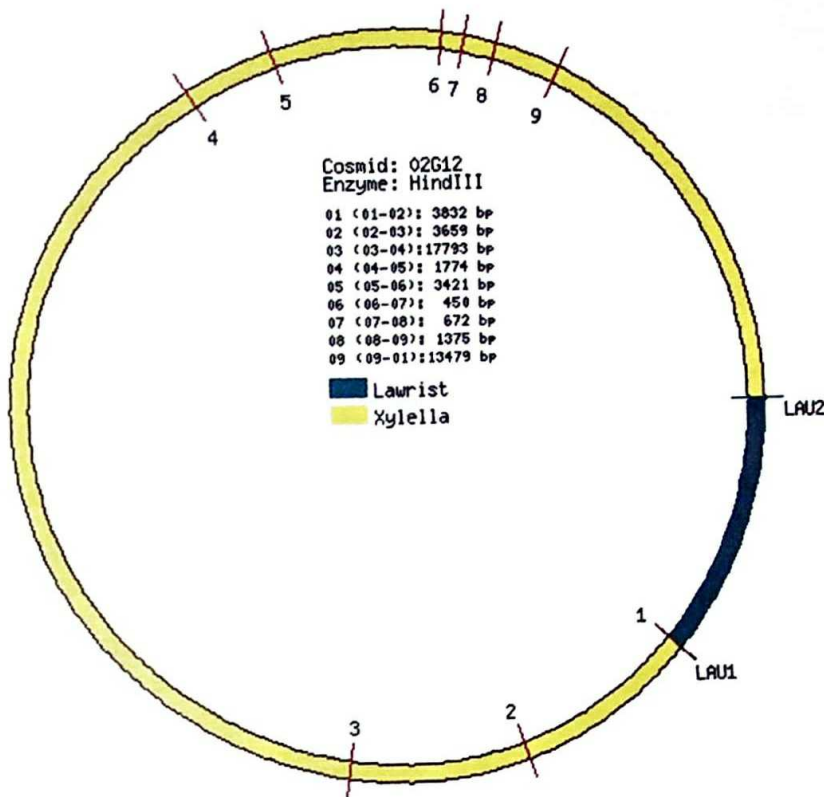
VI F 2. Cronograma de estágios de seqüenciamento

02G12/Wed 11-Aug-1999 Finished	Submitted
02G12/Mon 02-Aug-1999 Close to Finished	02G12/Wed 26-May-1999 Submitted
02G12/Mon 02-Aug-1999 Close to Finished	02G12/Wed 19-May-1999 Submitted
02G12/Thu 22-Jul-1999 Close to Finished	02G12/Mon 17-May-1999 Submitted
02G12/Fri 16-Jul-1999 Submitted	02G12/Sat 15-May-1999 Submitted
02G12/Mon 31-May-1999	

VI F 3. Digestão do cosmídeo X0QH/QS-02G12 com enzima de restrição e digestão eletrônica

Digestão com *EcoRI*, *HindIII* e *EcoRI* + *HindIII*





Cosmídeo X0QH/QS-05D08

VI G. Fragmento 05D08

VI G 1. Relatório de finalização do seqüenciamento

PRELIMINARY CERTIFICATION REPORT FOR GAP 05D08

DATE: Mon 25-Oct-1999

READ COMPOSITION: 1095 reads;

BG-07C06 : 115 reads
CM-11G06 : 6 reads
EP-11G06 : 6 reads
MC-05D08 : 2 reads
QH-05D08 : 336 reads
QS-05D08 : 195 reads
QS-07H02 : 105 reads
QV-06H03 : 61 reads
UV-05D08 : 2 reads
UV-06H03 : 130 reads
plasmid : 137 reads

Contig size: 45464 bp

Insert size: 34237 bp (excludes gap extensions)

EXTREMES LOCATIONS:

Left : first base after LOSTLAU (GGCGACCGGTCGTCCCATATTGCC) = 10707

Right : last base before LAUI (GATCCTAAGCTTGGGTCTCCCTA) = 44943

EXPECTED ERROR RATE: 0.00 / 10000 bp

LOW CONSENSUS QUALITY (LCQ) REGIONS:

TOTAL: 0

HIGH QUALITY DISCREPANCIES (HQD):

TOTAL: 0

POSITIONS NOT CONFIRMED ON BOTH STRANDS (NCBS):

TOTAL: 0

CONCLUSION: Gap FINISHED

E. Coli contamination:

None detected.

END OF ANALYSIS

VI G 2. Cronograma de estágios de seqüenciamento

05D08/Mon 25-Oct-1999

Finished

Finished

05D08/Wed 22-Sep-1999

05D08/Thu 23-Sep-1999

Finished

Close to Finished

05D08/Wed 22-Sep-1999

05D08/Wed 22-Sep-1999

One-conti

VI H. Fragmento 15D08

VI H 1. Relatório de finalização do seqüenciamento

PRELIMINARY CERTIFICATION REPORT FOR GAP 15D08

DATE: Wed 22-Sep-1999

READ COMPOSITION: 1095 reads;

BG-07C06 : 115 reads

CM-11G06 : 6 reads

EP-11G06 : 6 reads

MC-05D08 : 2 reads

QH-05D08 : 336 reads

QS-05D08 : 195 reads

QS-07H02 : 105 reads

QV-06H03 : 61 reads

UV-05D08 : 2 reads

UV-06H03 : 130 reads

plasmid : 137 reads

Contig size: 45464 bp

Insert size: 10600 bp (excludes gap extensions)

EXTREMES LOCATIONS:

Left : first base after LAU2 (CACTATAGAATACTAGGATC) = 1

Right : last base before LOSTLAU (TGATTCCCAGATGCCGCGTTGTGCATTT) = 10600

EXPECTED ERROR RATE: 0.00 / 10000 bp

LOW CONSENSUS QUALITY (LCQ) REGIONS:

TOTAL: 0

HIGH QUALITY DISCREPANCIES (HQD):

TOTAL: 0

POSITIONS NOT CONFIRMED ON BOTH STRANDS (NCBS):

TOTAL: 0

CONCLUSION: Gap FINISHED

END OF ANALYSIS

VI H 2. Cronograma de estágios de seqüenciamento

15D08/Wed 22-Sep-1999

Finished

15D08/Wed 22-Sep-1999

One-contig

VI I. Cosmídeo X0QH-06A08

VI I 1. Relatório de finalização do seqüenciamento

PRELIMINARY CERTIFICATION REPORT FOR COSMID 06A08
DATE: Thu 25-Mar-1999

READ COMPOSITION: 1056 reads;

CC-07A01 : 395 reads
JE-07C04 : 26 reads
QH-06A08 : 545 reads
plasmid : 90 reads

Contig size: 39987 bp
Insert size: 39491 bp (excludes lawrist and/or cosmid extensions)

LAWRIST LOCATIONS:

Left : first base after LAU2 = 1
Right : last base before LAU1 = 39491

EXPECTED ERROR RATE: 0.00 / 10000 bp

LOW CONSENSUS QUALITY (LCQ) REGIONS:

TOTAL: 0

HIGH QUALITY DISCREPANCIES (HQD):

TOTAL: 0

POSITIONS NOT CONFIRMED ON BOTH STRANDS (NCBS):

TOTAL: 0

CONCLUSION: Cosmid FINISHED

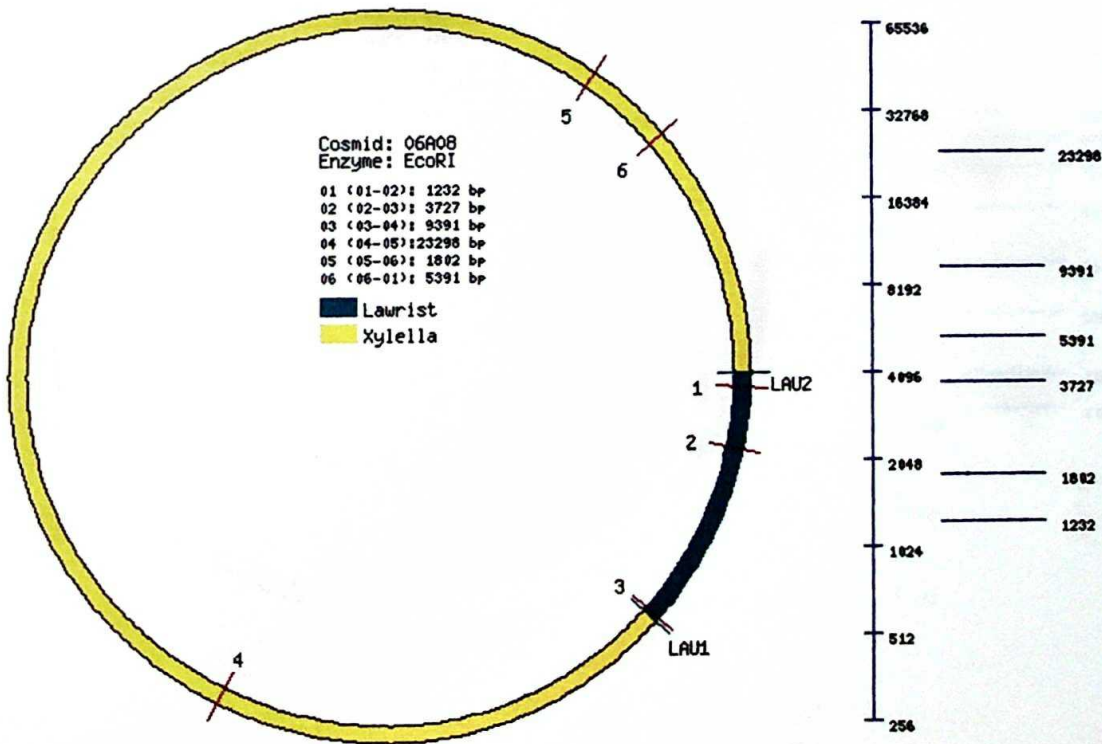
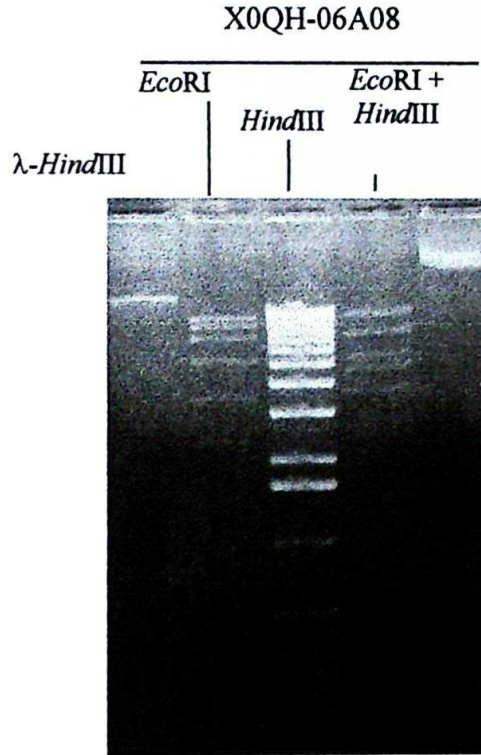
END OF ANALYSIS

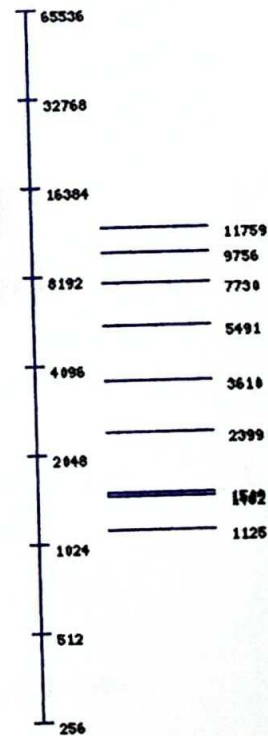
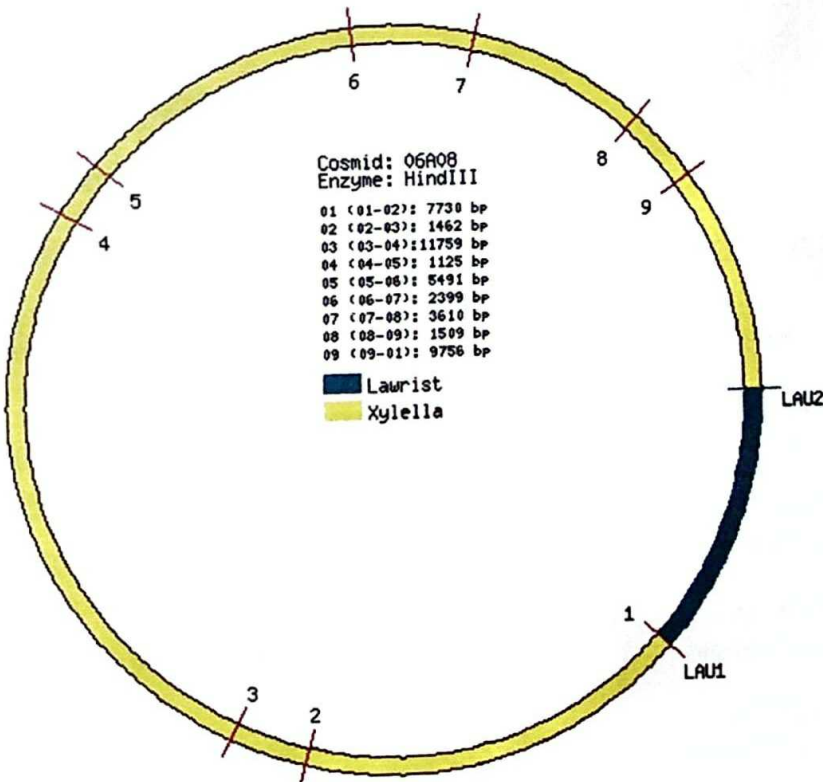
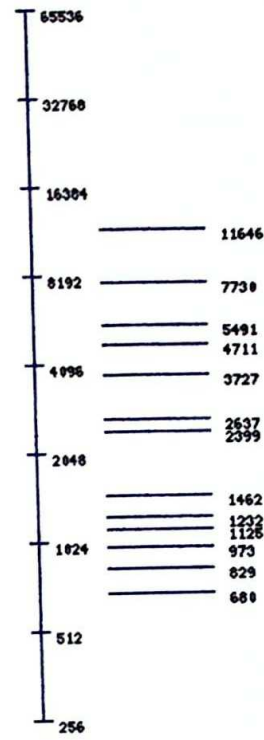
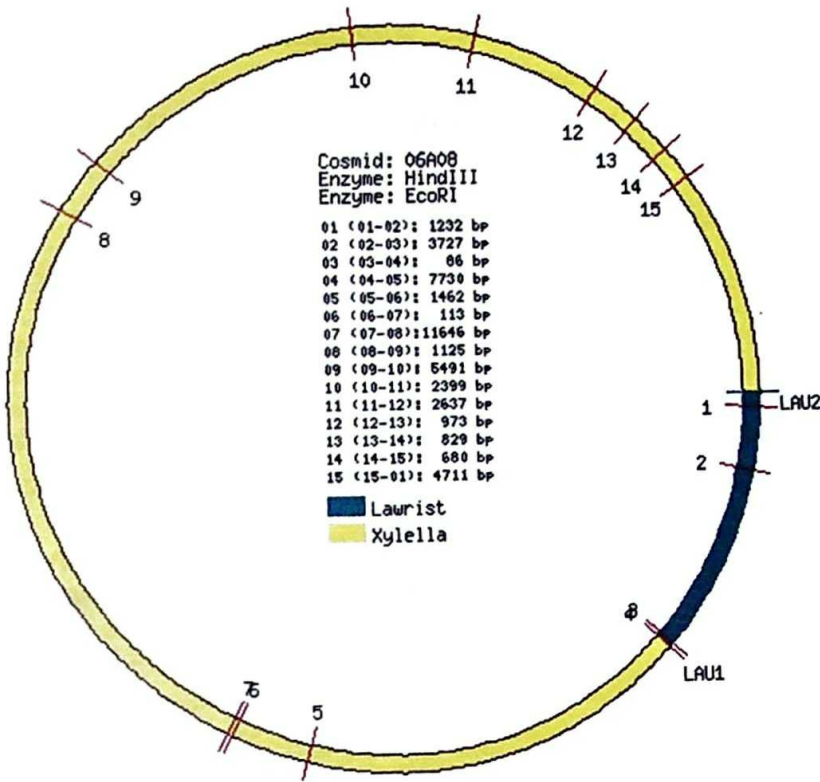
VI I 2. Cronograma de estágios de seqüenciamento

06A08/Thu 25-Mar-1999 Finished	One-contig 06A08/Mon 22-Feb-1999 One-contig
06A08/Fri 19-Mar-1999 Close to Finished	06A08/Mon 15-Feb-1999 One-contig
06A08/Mon 15-Mar-1999 Close to Finished	06A08/Sat 13-Feb-1999 One-contig
06A08/Sat 13-Mar-1999 One-contig	06A08/Sat 13-Feb-1999 One-contig
06A08/Fri 26-Feb-1999 One-contig	06A08/Fri 12-Feb-1999 Submitted
06A08/Thu 25-Feb-1999 One-contig	06A08/Fri 12-Feb-1999 Submitted
06A08/Wed 24-Feb-1999	

VII 3. Digestão do cosmídeo X0QH-06A08 com enzimas de restrição e digestão eletrônica

Digestões com *EcoRI*, *HindIII* e *EcoRI* + *HindIII*





VI J. Cosmideo X0QH-07A10

VI J 1. Relatório de finalização do seqüenciamento

PRELIMINARY CERTIFICATION REPORT FOR COSMID 07A10

DATE: Thu 25-Mar-1999

READ COMPOSITION: 939 reads;

QH-02E06 : 197 reads

QH-07A10 : 593 reads

plasmid : 149 reads

Contig size: 42922 bp

Insert size: 42406 bp (excludes lawrist and/or cosmid extensions)

LAWRIST LOCATIONS:

Left : first base after LAU2 = 517

Right : last base before LAU1 = 42922

EXPECTED ERROR RATE: 0.05 / 10000 bp

LOW CONSENSUS QUALITY (LCQ) REGIONS:

TOTAL: 0

HIGH QUALITY DISCREPANCIES (HQD):

TOTAL: 0

POSITIONS NOT CONFIRMED ON BOTH STRANDS (NCBS):

TOTAL: 0

CONCLUSION: Cosmid FINISHED

END OF ANALYSIS

VI J 2. Cronograma de estágios de seqüenciamento

07A10/Thu 25-Mar-1999
Finished

One-contig

07A10/Thu 25-Mar-1999
Close to Finished

07A10/Wed 10-Mar-1999
One-contig

07A10/Sat 20-Mar-1999
Close to Finished

07A10/Tue 09-Mar-1999
One-contig

07A10/Fri 19-Mar-1999
One-contig

07A10/Thu 11-Feb-1999
One-contig

07A10/Mon 15-Mar-1999
One-contig

07A10/Wed 10-Feb-1999
One-contig

07A10/Sat 13-Mar-1999
One-contig

07A10/Wed 03-Feb-1999
One-contig

07A10/Wed 10-Mar-1999

07A10/Fri 08-Jan-1999
One-contig

VI J 3. Observação com relação à extremidade quimérica do cosmídeo X0QH-07A10

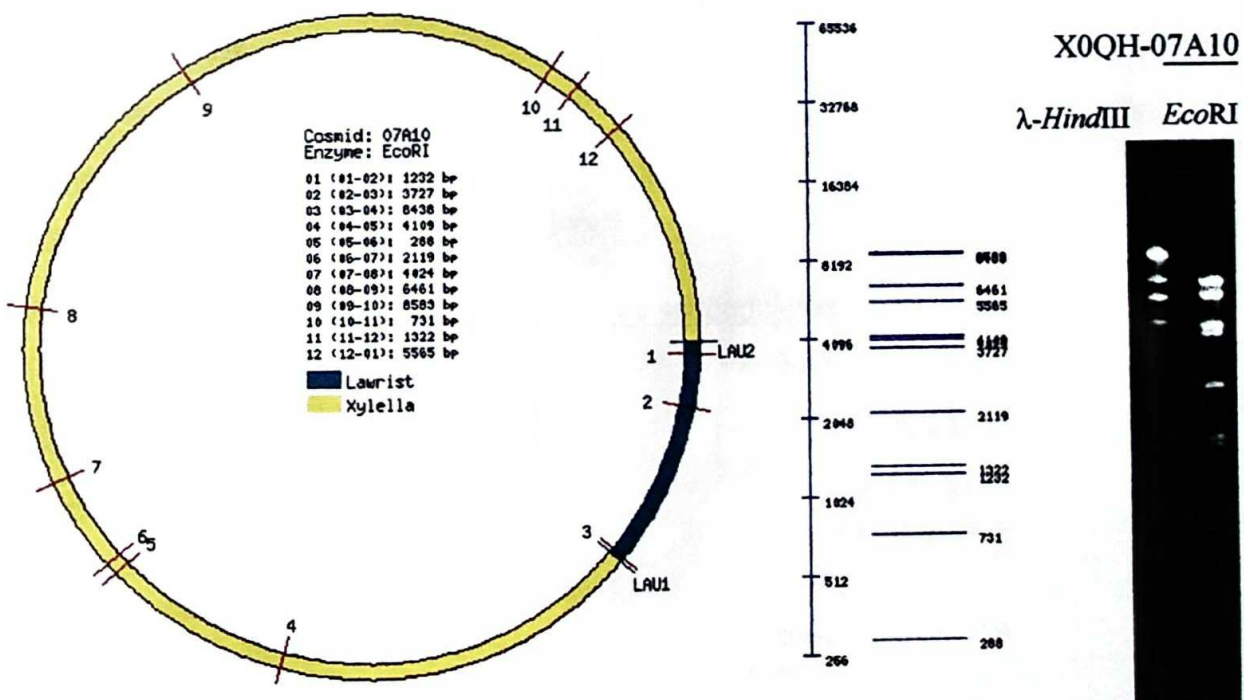
Devido à descoberta de que a região inicial do cosmídeo X0QH-07A10 era quimérica, através dos experimentos de amplificação e pela evidência de não sobreposição exata do cosmídeo 02E06 a tal região, foi adicionado o seguinte comentário à seqüência fasta do cosmídeo, indicando a remoção manual das bases problemáticas pelo laboratório de bioinformática:

Esse cosmídeo tem 1157 bases no seu início que aparentemente não são parte do genoma (evidência disto foi enviada pelo grupo QH). A seqüência fasta consenso original foi modificada em 5 de maio de 1999 e as primeiras 1157 bases foram manualmente removidas.

Originalmente: "This cosmid has 1157 bases at the beginning which apparently are not part of the genome (evidence about this related by QH group). The original fasta consensus sequence was modified on 5-may-1999 and the first 1157 bases were manually removed." (J.C. Setúbal)

VI J 4. Digestão do cosmídeo X0QH-07A10 com enzimas de restrição e digestão eletrônica

Digestão com *EcoRI*



VI K. GFS X0QH-00J42 (X0QH1013G02)

Clone da posição G02 da placa 13 da biblioteca 10 de "shotgun" genômico

VI K 1. Relatório de finalização do seqüenciamento

PRELIMINARY CERTIFICATION REPORT FOR GAP 00J42

DATE: Thu 16-Dec-1999

READ COMPOSITION: 73 reads;

QH-00J42 : 67 reads

plasmid : 6 reads

Contig size: 2840 bp

Insert size: 2622 bp (excludes gap extensions)

EXTREME LOCATIONS:

Left : first base after X0QH1013G02R (TGCCAAAGAAGAATTAATAT) = 21

Right : last base before X0QH1013G02F (CTCACCCAAAGATTGACGTA) = 2642

EXPECTED ERROR RATE: 0.00 / 10000 bp

LOW CONSENSUS QUALITY (LCQ) REGIONS:

TOTAL: 0

HIGH QUALITY DISCREPANCIES (HQD):

TOTAL: 0

POSITIONS NOT CONFIRMED ON BOTH STRANDS (NCBS):

1299-1299 (1): ABI base call confirms the base

TOTAL: 0

CONCLUSION: Gap Finished

E. Coli contamination:

None detected.

END OF ANALYSIS

VI K 2. Cronograma de estágios de seqüenciamento

	Submitted
00J42/Thu 16-Dec-1999 Finished	00J42/Mon 13-Dec-1999 Submitted
00J42/Thu 16-Dec-1999 Close to Finished	00J42/Mon 13-Dec-1999 Submitted
00J42/Tue 14-Dec-1999 one-contig (manually)	00J42/Wed 08-Dec-1999 Submitted
00J42/Tue 14-Dec-1999	

VI L. Lista de ORFs anotadas na Categoria I: Metabolismo Intermediário

INTERMEDIARY METABOLISM

INTERMEDIARY METABOLISM/DEGRADATION/ Degradation of polysaccharides

XF-07H04-GL14 extracellular endoglucanase precursor {*Xanthomonas campestris* (pv. *campestris*)}

INTERMEDIARY METABOLISM/DEGRADATION/ Degradation of small molecules

XF-05B03-GL12 acetylxylyl esterase (*axeA*) {*Streptomyces coelicolor* A3(2)}
XF-02D03-GL36 arginine deaminase (*rocF*) {*Bacillus halodurans*}
XF-04A09-GL48 benzene 1,2-dioxygenase, ferredoxin protein (*bedB*) {*Pseudomonas putida*}
XF-08A03-GL09 beta-galactosidase (*bga*) {*Xanthomonas manihotis*}
XF-07A10-GL45 beta-glucosidase (*bglX*) {*Escherichia coli*}
XF-09E09-GL33 beta-mannosidase precursor {*Thermotoga maritima*}
XF-02D03-GL06 carboxyphosphoenolpyruvate phosphonmutase (*prpB*) {*Escherichia coli*}
XF-11A09-GL46 dioxygenase {*Klebsiella oxytoca*}
XF-0.3-GL0214 esterase (*est*) {*Synechocystis* sp.}
XF-11A05-GL44 fructokinase {*Beta vulgaris*}
XF-0.3-GL0219 glucose dehydrogenase B (*ylil*) {*Escherichia coli*}
XF-01G04-GL46 glucose kinase (*glk*) {*Escherichia coli*}
XF-01A01-GL32 glucose kinase (*glk*) {*Zymomonas mobilis*}
XF-05F05-GL17 haloalkane dehalogenase (*dhaA*) {*Rhodococcus rhodochrous*}
XF-07C01-GL04 L-ascorbate oxidase (*aao*) {*Cucurbita maxima* (Pumpkin) (Winter squash)}
XF-02D03-GL41 lipase (*lipP*) {*Pseudomonas* sp. B11-1}
XF-03D03-GL19 lipase/esterase (*estA*) {*Pseudomonas aeruginosa*}
XF-07F02-GL04 NAD(P)H steroid dehydrogenase (*cdh*) {*Mus musculus*}
XF-07C06-GL21 phosphoglycolate phosphatase (*yfbT*) {*Escherichia coli*}
XF-0.3-GL3354 phosphoglycolate phosphatase (*cbbZC*) {*Alcaligenes eutrophus*}
XF-11A08-GL38 polyvinylalcohol dehydrogenase {*Pseudomonas* sp.}
XF-10F03-GL13 ribokinase (*rbsK*) {*Escherichia coli*}
XF-0.5-MN0006 rubredoxin (*rubA*) {*Acinetobacter calcoaceticus*}
XF-11B08-GL09 threonine dehydratase catabolic (*tdcB*) {*Escherichia coli*}
XF-11G06-GL08 triacylglycerol lipase precursor (*lip*) {*Pseudomonas* sp. (strain KWI-56)}
XF-06A08-GL14 UDP-glucose 4-epimerase (*galE*) {*Brevibacterium lactofermentum*}
XF-0.5-GL3353 UTP-glucose-1-phosphate uridylyltransferase (*gtaB*) {*Xanthomonas campestris*}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Amino sugars

XF-07A09-GL02 glucosamine--fructose-6-phosphate aminotransferase (*glmS*) {*Escherichia coli*}
XF-03C12-GL04 glucosamine--fructose-6-phosphate aminotransferase {*Deinococcus radiodurans*}
XF-07A02-GL13 N-acetyl-beta-glucosaminidase (*exo*) {*Vibrio furnissii*}
XF-03C12-GL03 N-acetylglucosamine-6-phosphate deacetylase (*nagA*) {*Escherichia coli*}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Entner-Doudoroff

- XF-01G04-GL43 2-keto-3-deoxy-6-phosphogluconate aldolase (eda/hga/kdgA) {*Escherichia coli*}
XF-01G04-GL44 6-phosphogluconate dehydratase (edd) {*Rhizobium meliloti*}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Gluconeogenesis

- XF-08D10-GL10 malate oxidoreductase (maeB) {*Haemophilus influenzae*}
XF-02D03-GL48 phosphoenolpyruvate synthase (ppsA/pps) {*Escherichia coli*}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Miscellaneous glucose metabolism

- XF-0.3-GL0623 phosphomannomutase (mrsA) {*Escherichia coli*}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Non-oxidative branch, pentose pathway

- XF-05G02-GL50 D-ribulose-5-phosphate 3-epimerase (rpe/dod) {*Serratia marcescens*}
XF-11A02-GL10 ribose-5-phosphate isomerase A (rpiA) {*Escherichia coli*}
XF-03H12-GL23 transketolase 1 (tktA/tkt) {*Escherichia coli*}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Nucleotide interconversions

- XF-02F10-GL09 CTP synthetase (pyrG) {*Escherichia coli*}
XF-01G04-GL39 uridylate kinase (pyrH) {*Synechocystis* sp. (strain PCC 6803)}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Phosphorus compounds

- XF-07C04-GL61 alkaline phosphatase (phoA) {*Escherichia coli*}
XF-02G04-GL16 ATP-binding protein (ybeZ) {*Escherichia coli*}
XF-07H03-GL15 exopolyphosphatase (ppx) {*Pseudomonas aeruginosa*}
XF-07H03-GL16 polyphosphate kinase (ppk) {*Escherichia coli*}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Pool multipurpose conversions

- XF-11A08-GL32 acetyl coenzyme A synthetase (acs) {*Escherichia coli*}
XF-01G04-GL14 adenosylhomocysteinase (ahcY) {*Rhodobacter capsulatus* (*Rhodopseudomonas capsulata*)}
XF-0.3-GL1498 carbonic anhydrase (yadF) {*Escherichia coli*}
XF-06D10-GL58 carbonic anhydrase {*Methanosarcina thermophila*}
XF-11A08-GL26 deoxyxylulose-5-phosphate synthase (dxs) {*Escherichia coli*}
XF-07C09-GL01 ferredoxin-NADP reductase (fpr) {*Azotobacter vinelandii*}
XF-11A07-GL55 glycerol kinase (glpK) {*Enterococcus faecalis* (*Streptococcus faecalis*)}
XF-05G02-GL17 glycine cleavage H protein (gcvH) {*Escherichia coli*}
XF-05G02-GL19 glycine cleavage T protein (gcvT) {*Escherichia coli*}
XF-08A02-GL33 glycine decarboxylase (gcvP) {*Escherichia coli*}
XF-07C07-GL33 hydroxyacylglutathione hydrolase (gloB) {*Escherichia coli*}
XF-07C07-GL48 inorganic pyrophosphatase (ppa) {*Escherichia coli*}
XF-07C12-GL16 lactoylglutathione lyase (gloA) {*Pseudomonas putida*}

- XF-07A04-GL26 methionine adenosyltransferase (metK) {Haemophilus influenzae}
XF-02G04-GL30 tropinone reductase {Arabidopsis thaliana}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Sugar-nucleotide biosynthesis, conversions

- XF-06A08-GL15 GDP-mannose 4,6 dehydratase (gmd) {Escherichia coli}
XF-11A07-GL41 nucleotide sugar epimerase (wbnF) {Escherichia coli}
XF-C13KB-GL12 phosphoglucomutase/phosphomannomutase (xanA) {Xanthomonas campestris (pv. campestris)}
XF-C13KB-GL13 phosphomannose isomerase-GDP-mannose pyrophosphorylase (xanB) {Xanthomonas campestris (pv. campestris)}
XF-11A05-GL37 UDP-glucose dehydrogenase (ugd) {Xanthomonas campestris pv. campestris}

INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/ Sulfur metabolism

- XF-01H09-GL04 3'-phosphoadenosine 5'-phosphosulfate reductase (cysH) {Salmonella typhimurium}
XF-01H09-GL09 ATP sulfurylase, large subunit (nodQ) {Rhizobium meliloti}
XF-01H09-GL08 ATP sulfurylase, small subunit (cysD) {Escherichia coli}
XF-01H09-GL07 NADPH-sulfite reductase, flavoprotein subunit (cysJ) {Salmonella typhimurium}
XF-01H09-GL06 NADPH-sulfite reductase, iron-sulfur protein (cysI) {Thiocapsa roseopersicina}
XF-05G02-GL23 sulfite synthesis pathway protein (cysQ/amtA) {Escherichia coli}

INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/ Aerobic respiration

- XF-0.3-GL2292 D-lactate dehydrogenase (dld1) {Aquifex aeolicus}
XF-09C12-GL23 glycerol-3-phosphate dehydrogenase (gpsA) {Escherichia coli}
XF-11A07-GL57 glycerol-3-phosphate dehydrogenase (glpD) {Pseudomonas aeruginosa}
XF-01A03-GL13 NADH-ubiquinone oxidoreductase, NQO1 subunit (nuoF) {Rickettsia prowazekii}
XF-01A03-GL07 NADH-ubiquinone oxidoreductase, NQO10 subunit (nuoJ) {Rhodobacter capsulatus (Rhodopseudomonas capsulata)}
XF-01A03-GL06 NADH-ubiquinone oxidoreductase, NQO11 subunit (nuoK) {Paracoccus denitrificans}
XF-01A03-GL05 NADH-ubiquinone oxidoreductase, NQO12 subunit (nuoL) {Paracoccus denitrificans}
XF-01A03-GL04 NADH-ubiquinone oxidoreductase, NQO13 subunit (nuoM) {Rhodobacter capsulatus (Rhodopseudomonas capsulata)}
XF-01A03-GL03 NADH-ubiquinone oxidoreductase, NQO14 subunit (nuoN) {Paracoccus denitrificans}
XF-01A03-GL14 NADH-ubiquinone oxidoreductase, NQO2 subunit (nuoE) {Rickettsia prowazekii}
XF-01A03-GL12 NADH-ubiquinone oxidoreductase, NQO3 subunit (nuoG) {Drosophila melanogaster (Fruit fly)}
XF-01A03-GL15 NADH-ubiquinone oxidoreductase, NQO4 subunit (nuoD) {Reclinomonas americana}
XF-01A03-GL16 NADH-ubiquinone oxidoreductase, NQO5 subunit (nuoC) {Rhizobium meliloti}
XF-01A03-GL17 NADH-ubiquinone oxidoreductase, NQO6 subunit (nuoB) {Solanum tuberosum (Potato)}

- XF-01A03-GL18 NADH-ubiquinone oxidoreductase, NQO7 subunit (nuoA) {*Rhodobacter capsulatus* (*Rhodospseudomonas capsulata*)}
- XF-01A03-GL09 NADH-ubiquinone oxidoreductase, NQO8 subunit (nuoH) {*Marchantia polymorpha* (Liverwort)}
- XF-01A03-GL08 NADH-ubiquinone oxidoreductase, NQO9 subunit (nuoI) {*Neurospora crassa*}

INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/ Anaerobic respiration and Fermentation

- XF-05B03-GL04 alcohol dehydrogenase (yahK) {*Escherichia coli*}
- XF-0.3-GL0240 NADP-alcohol dehydrogenase (adh) {*Thermoanaerobacter brockii* (*Thermoanaerobium brockii*)}
- XF-0.3-GL0230 NADP-alcohol dehydrogenase (yahK) {*Escherichia coli*}
- XF-0.3-GL0209 NADP-alcohol dehydrogenase (yahK) {*Escherichia coli*}
- XF-0.3-GL1124 NADP-dependent alcohol dehydrogenase {*Leishmania major*}

INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/ Electron transport

- XF-0.3-GL3369 c-type cytochrome biogenesis membrane protein (cycK) {*Pseudomonas fluorescens*}
- XF-0.3-GL3372 c-type cytochrome biogenesis protein (cycJ) {*Pseudomonas fluorescens*}
- XF-0.3-GL3367 c-type cytochrome biogenesis protein (cycL) {*Pseudomonas fluorescens*}
- XF-06A08-GL03 c-type cytochrome biogenesis protein (copper tolerance) (dsbD) {*Chlamydia trachomatis*}
- XF-0.3-GL3368 c-type cytochrome biogenesis protein/thioredoxin (dsbE/ccmG) {*Escherichia coli*}
- XF-06C10-GL31 cytochrome B561 (yodB) {*Escherichia coli*}
- XF-07B09-GL33 cytochrome C oxidase assembly factor (coxD) {*Synechocystis* sp.}
- XF-08A02-GL38 cytochrome O ubiquinol oxidase, subunit I (cyoB) {*Escherichia coli*}
- XF-07C12-GL01 cytochrome O ubiquinol oxidase, subunit II (cyoA) {*Escherichia coli*}
- XF-08A02-GL37 cytochrome O ubiquinol oxidase, subunit III (cyoC) {*Escherichia coli*}
- XF-08A02-GL36 cytochrome O ubiquinol oxidase, subunit IV (cyoD) {*Escherichia coli*}
- XF-C17KB-GL35 electron transfer flavoprotein alpha subunit (etfA) {*Paracoccus denitrificans*}
- XF-0.5-GL0382 electron transfer flavoprotein beta subunit (etfB/etfS) {*Bradyrhizobium japonicum*}
- XF-02F10-GL22 electron transfer flavoprotein ubiquinone oxidoreductase (etf-QO) {*Zymomonas mobilis*}
- XF-06F05-GL35 electron transfer protein azurin I (az1) {*Achromobacter xylosoxidans* subsp. *xylosoxidans*}
- XF-07H02-GL34 ferredoxin {*Haemophilus influenzae*}
- XF-05F05-GL15 ferredoxin (ykgJ) {*Escherichia coli*}
- XF-07H03-GL32 ferredoxin {*Pseudomonas putida*}
- XF-07A08-GL58 ferredoxin II (ydgM) {*Escherichia coli*}
- XF-03H11-GL17 flavohemoprotein (fhp) {*Alcaligenes eutrophus*}
- XF-11A04-GL59 oxidoreductase {*Schizosaccharomyces pombe*}
- XF-11A02-GL45 thioredoxin (yneN) {*Bacillus subtilis*}
- XF-02G04-GL24 ubiquinol cytochrome C oxidoreductase, cytochrome B subunit (petB) {*Chromatium vinosum*}
- XF-02G04-GL25 ubiquinol cytochrome C oxidoreductase, cytochrome C1 subunit (petC) {*Chromatium vinosum*}
- XF-02G04-GL23 ubiquinol cytochrome C oxidoreductase, iron-sulfur subunit (petA) {*Chromatium vinosum*}

INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/ Glycolysis

- XF-0.5-GL0408 6-phosphofructokinase (pfkA) {*Streptomyces coelicolor*}
- XF-02F10-GL13 enolase (eno) {*Bacillus subtilis*}
- XF-08A03-GL27 fructose-bisphosphate aldolase {*Pisum sativum* (Garden pea)}
- XF-0.3-GL2475 glucose-6-phosphate isomerase (pgi) {*Xanthomonas citri*}
- XF-0.3-GL2115 glyceraldehyde-3-phosphate dehydrogenase (gapA) {*Escherichia coli*}
- XF-08A03-GL31 phosphoglycerate kinase (pgk) {*Escherichia coli*}
- XF-07C09-GL06 phosphoglyceromutase (gpmA/gpm) {*Escherichia coli*}
- XF-08A03-GL29 pyruvate kinase type II (pykA) {*Escherichia coli*}
- XF-01A03-GL21 triosephosphate isomerase (tpiA/tpi) {*Escherichia coli*}

INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/ Oxidative branch, pentose pathway

- XF-01G04-GL45 6-phosphogluconolactonase (pgl) {*Pseudomonas aeruginosa*}
- XF-01G04-GL47 glucose-6-phosphate 1-dehydrogenase (zwf) {*Pseudomonas aeruginosa*}

INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/ Pyruvate dehydrogenase

- XF-0.3-GL1517 dihydrolipoamide acetyltransferase (pdhB) {*Alcaligenes eutrophus*}
- XF-0.3-GL1518 dihydrolipoamide dehydrogenase (lpdA/lpd) {*Ralstonia eutropha*}
- XF-09G12-GL15 pyruvate dehydrogenase (aceE) {*Escherichia coli*}

INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/ TCA cycle

- XF-01A03-GL41 aconitate hydratase 2 (acnB) {*Escherichia coli*}
- XF-07B05-GL07 citrate synthase (gltA) {*Pseudomonas aeruginosa*}
- XF-07B05-GL26 dihydrolipoamide dehydrogenase (lpd) {*Pseudomonas fluorescens*}
- XF-07B05-GL28 dihydrolipoamide S-succinyltransferase (sucB) {*Escherichia coli*}
- XF-07B05-GL37 fumarate hydratase (fumC) {*Pseudomonas aeruginosa*}
- XF-07F02-GL49 fumarate hydratase (fumB) {*Escherichia coli*}
- XF-07H03-GL23 isocitrate dehydrogenase (icd) {*Thermus aquaticus* (subsp. thermophilus)}
- XF-07B07-GL33 isocitrate dehydrogenase (icd) {*Azotobacter vinelandii*}
- XF-07C06E-GL08 malate dehydrogenase (mdh) {*Thermus aquaticus* (subsp. flavus)}
- XF-08D10-GL54 malate:quinone oxidoreductase (yojH) {*Escherichia coli*}
- XF-07B05-GL30 oxoglutarate dehydrogenase (odhA) {*Alcaligenes eutrophus*}
- XF-07B08-GL21 succinate dehydrogenase iron-sulfur protein (sdhB) {*Paracoccus denitrificans*}
- XF-07B08-GL20 succinate dehydrogenase, flavoprotein subunit (sdhA) {*Rickettsia prowazekii*}
- XF-07B08-GL18 succinate dehydrogenase, membrane anchor subunit (sdhC) {*Bradyrhizobium japonicum*}
- XF-0.5-GL1512 succinate dehydrogenase, membrane anchor subunit (sdhD) {*Bradyrhizobium japonicum*}
- XF-02D09-GL52 succinyl-CoA synthetase, alpha subunit (sucD) {*Escherichia coli*}
- XF-02D09-GL53 succinyl-CoA synthetase, beta subunit (sucC) {*Escherichia coli*}

INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/ ATP-proton
motive force interconversion

- XF-0.3-GL1105 ATP synthase, A chain (atpB/uncB/papD) {*Escherichia coli*}
XF-0.3-GL1112 ATP synthase, alpha chain (atpA/uncA) {*Vibrio alginolyticus*}
XF-0.3-GL1108 ATP synthase, B chain (atpF/uncF) {*Vibrio alginolyticus*}
XF-0.3-GL1115 ATP synthase, beta chain (atpD) {*Haemophilus influenzae*}
XF-0.3-GL1106 ATP synthase, C chain (atpE/uncE) {*Vibrio alginolyticus*}
XF-0.5-GL1614 ATP synthase, delta chain (atpH/uncH) {*Vibrio alginolyticus*}
XF-0.3-GL1116 ATP synthase, epsilon chain (atpC) {*Thiobacillus ferrooxidans*}
XF-0.3-GL1113 ATP synthase, gamma chain (atpG/uncG/papC) {*Escherichia coli*}

INTERMEDIARY METABOLISM/REGULATORY FUNCTIONS

- XF-0.5-FS0013 aconitate hydratase 1 (acnA/acn) {*Escherichia coli*}
XF-02F10-GL54 ATP:GTP 3'-pyrophosphotransferase (relA) {*Vibrio* sp. (strain S14 / CCUG 15956)}
XF-07A09-GL23 carbon storage regulator (csrA) {*Pseudomonas aeruginosa*}
XF-0.5-MN0004 cold shock protein (scoF) {*Streptomyces coelicolor*}
XF-0.3-GL3348 extragenic suppressor (suhB/ssyA) {*Escherichia coli*}
XF-08A04-GL38 heat-inducible transcriptional repressor (hrcA) {*Caulobacter crescentus*}
XF-07A09-GL26 LexA repressor (lexA) {*Xanthomonas citri*}
XF-11G06-GL09 lipase modulator (act) {*Pseudomonas* sp. (strain KWI-56)}
XF-0.5-GL2524 methanol dehydrogenase regulatory protein {*Deinococcus radiodurans*}
XF-07F02-GL10 nitrile hydratase activator {*Pseudomonas chlororaphis* (*Pseudomonas fluorescens* biotype D)}
XF-07F02-GL29 nitrogen regulatory protein P-II (glnB) {*Xanthomonas citri*}
XF-0.3-GL2286 pentaphosphate guanosine-3'-pyrophosphohydrolase (spoT) {*Escherichia coli*}
XF-07C07-GL16 phosphate regulon transcriptional regulator (phoU) {*Zymomonas mobilis*}
XF-02D03-GL72 poly(hydroxyalcanoate) granule associated protein (phaF) {*Pseudomonas oleovorans*}
XF-07B07-GL43 RNA polymerase sigma-32 factor (rpoH) {*Xanthomonas campestris* pv. *campestris*}
XF-07C12-GL27 RNA polymerase sigma-54 factor (rpoN) {*Xanthomonas campestris* (pv. *vesicatoria*)}
XF-07B09-GL51 RNA polymerase sigma-70 factor (rpoD) {*Pseudomonas putida*}
XF-11A08-GL15 RNA polymerase sigma-H factor (algU/algT) {*Pseudomonas aeruginosa*}
XF-07C12-GL26 sigma-54 modulation protein {*Azotobacter vinelandii*}
XF-02G04-GL28 stringent starvation protein A (sspA/ssp/pog) {*Escherichia coli*}
XF-02G04-GL29 stringent starvation protein B (sspB) {*Escherichia coli*}
XF-07C07-GL41 transcription-related protein (tex) {*Bordetella pertussis*}
XF-08D10-GL28 transcriptional regulator (gcvR) {*Escherichia coli*}
XF-0.3-GL0206 transcriptional regulator (opdE) {*Pseudomonas aeruginosa*}
XF-0.5-GL2587 transcriptional regulator (exsB) {*Synechocystis* sp.}
XF-11A04-GL04 transcriptional regulator (paiB) {*Bacillus subtilis*}
XF-11A09-GL27 transcriptional regulator (algH) {*Pseudomonas aeruginosa*}
XF-11A04-GL63 transcriptional regulator (AcrR family) {*Streptomyces coelicolor* A3(2)}
XF-02D03-GL43 transcriptional regulator (AraC family) (araL) {*Streptomyces lividans*}
XF-03D03-GL35 transcriptional regulator (ArsR family) (hlyU) {*Vibrio cholerae*}
XF-07B05-GL12 transcriptional regulator (Crp/Fnr family) (clp) {*Xanthomonas campestris* (pv. *campestris*)}
XF-07H04-GL41 transcriptional regulator (Fur family) (zur) {*Escherichia coli*}
XF-07A02-GL01 transcriptional regulator (Fur family) (fur) {*Xanthomonas campestris* pv. *vesicatoria*}
XF-03C12-GL05 transcriptional regulator (LacI family) {*Streptomyces coelicolor* A3(2)}

XF-08D10-GL16 transcriptional regulator (LuxR/UhpA family) (*agmR/glpR*) {*Pseudomonas aeruginosa*}
 XF-07H03-GL40 transcriptional regulator (LuxR/UhpA family) (*gacA*) {*Pseudomonas fluorescens*}
 XF-08A03-GL16 transcriptional regulator (LysR family) (*cysB*) {*Haemophilus influenzae*}
 XF-0.3-GL1129 transcriptional regulator (LysR family) (*act*) {*Pseudomonas aeruginosa*}
 XF-0.3-GL0235 transcriptional regulator (LysR family) (*yafC*) {*Escherichia coli*}
 XF-0.3-GL0200 transcriptional regulator (LysR family) {*Pseudomonas aeruginosa*}
 XF-11H05-GL09 transcriptional regulator (LysR family) (*ycjZ*) {*Escherichia coli*}
 XF-05G02-GL59 transcriptional regulator (MarR family) (*prsX*) {*Escherichia coli*}
 XF-07B09-GL44 transcriptional regulator (MarR family) (*yybA*) {*Bacillus subtilis*}
 XF-04A09-GL21 transcriptional regulator (MarR/EmrR family) {*Bacillus firmus*}
 XF-11A02-GL35 transcriptional regulator (PbsX family) {Bacteriophage P22}
 XF-03H11-GL06 transcriptional repressor (*korB*) {*Escherichia coli*}
 XF-11A04-GL34 transcriptional repressor (*korC*) {*Escherichia coli*}
 XF-0.5-GL2659 Trp operon transcriptional repressor (*trpR*) {*Enterobacter cloacae*}
 XF-03C11-GL30 tryptophan repressor binding protein (*wrbA*) {*Escherichia coli*}
 XF-0.3-GL1127 tryptophan repressor binding protein {*Archaeoglobus fulgidus*}
 XF-0.3-GL0231 tryptophan repressor binding protein {*Archaeoglobus fulgidus*}
 XF-03C12-GL13 two-component system, hybrid sensor/regulatory protein {*Deinococcus radiodurans*}
 XF-0.3-GL2327 two-component system, regulatory protein (*tctD*) {*Salmonella typhimurium*}
 XF-10F03-GL36 two-component system, regulatory protein (*popP/feuP/phoP*) {*Rhizobium leguminosarum*}
 XF-07A10-GL61 two-component system, regulatory protein (*pilH*) {*Pseudomonas aeruginosa*}
 XF-03C11-GL04 two-component system, regulatory protein {*Synechocystis sp.*}
 XF-07G02-GL14 two-component system, regulatory protein (*algR*) {*Pseudomonas aeruginosa*}
 XF-07F02-GL36 two-component system, regulatory protein (*glnG/ntrC/glnT*) {*Escherichia coli*}
 XF-08A04-GL30 two-component system, regulatory protein (*colR*) {*Pseudomonas fluorescens*}
 XF-05E12-GL26 two-component system, regulatory protein (*colR*) {*Pseudomonas fluorescens*}
 XF-05E12-GL08 two-component system, regulatory protein (*pilR*) {*Pseudomonas aeruginosa*}
 XF-02D09-GL14 two-component system, regulatory protein (*actR*) {*Sinorhizobium meliloti*}
 XF-07H03-GL18 two-component system, regulatory protein (*phoB*) {*Escherichia coli*}
 XF-0.3-GL2326 two-component system, sensor protein (*tctE*) {*Salmonella typhi*}
 XF-10F03-GL37 two-component system, sensor protein (*phoQ*) {*Escherichia coli*}
 XF-05G03-GL51 two-component system, sensor protein (*phs*) {*Calothrix viguieri*}
 XF-08D10-GL15 two-component system, sensor protein (*flhS*) {*Paracoccus denitrificans*}
 XF-07G02-GL13 two-component system, sensor protein (*algZ*) {*Pseudomonas aeruginosa*}
 XF-07F02-GL37 two-component system, sensor protein (*ntrB*) {*Thiobacillus ferrooxidans*}
 XF-05E12-GL25 two-component system, sensor protein (*colS*) {*Pseudomonas fluorescens*}
 XF-05E12-GL06 two-component system, sensor protein (*pilS*) {*Pseudomonas aeruginosa*}
 XF-02D09-GL15 two-component system, sensor protein (*actS*) {*Sinorhizobium meliloti*}
 XF-07H03-GL17 two-component system, sensor protein (*phoR*) {*Klebsiella pneumoniae*}

VI M. Lista de ORFs anotadas na Categoria de Processos Celulares: Transporte de carboidratos, ácidos orgânicos e álcoois

CELLULAR PROCESSES/TRANSPORT/Carbohydrates, organic acids, alcohols

- XF-02A11-GL29 ABC transporter sugar permease (SCF11.20) {*Streptomyces coelicolor* A3(2)}
- XF-02A11-GL30 ABC transporter sugar permease (LACF) {*Agrobacterium radiobacter*}
- XF-02A11-GL31 ABC transporter sugar-binding protein (malE) {*Thermococcus litoralis*}
- XF-07C10-GL48 alpha-ketoglutarate permease symporter (KGTP OR WITA) {*Escherichia coli*}
- XF-08D10-GL11 C4-dicarboxylate transport protein (DCTA) {*Rhizobium leguminosarum*}
- XF-03C12-GL06 glucose/galactose transporter (GLUP) {*Brucella abortus*}
- XF-11A05-GL43 glucose/galactose transporter (GLUP) {*Brucella abortus*}
- XF-11A07-GL56 glycerol uptake facilitator protein (GLPF) {*Escherichia coli*}
- XF-07C12-GL25 HPr kinase/phosphatase (PTSK) {*Bacillus subtilis*}
- XF-0.5-GL0475 Mg⁺⁺/citrate complex transporter (CITN OR CITH OR N15CR) {*Bacillus subtilis*}
- XF-07C12-GL19 phosphotransferase system enzyme I (PHBI) {*Alcaligenes eutrophus*}
- XF-07C12-GL20 phosphotransferase system HPr enzyme (PHBH) {*Alcaligenes eutrophus*}
- XF-0.4-GL1627 sugar ABC transporter ATP-binding protein (DR2153) {*Deinococcus radiodurans*}

VI N. BLASTP da Frutose-1,6-bisfosfato aldolase de *Xylella fastidiosa*

São mostrados na tabela a seguir os principais cabeçalhos dos melhores alinhamentos realizados através do serviço de BLASTP da página do NCBI³². As linhas pontilhadas indicam onde uma grande quantidade de alinhamentos foi pulado para se chegar a algum de maior interesse. O alinhamento de *Synechocystis* sp. é o primeiro referente a bactérias. A alta similaridade com a enzima de plantas pode indicar um evento de transferência horizontal, como "non-orthologous gene displacement"⁴¹.

BLASTP 2.0.10

Query= XF-08A03-GL27 (334 letters) -Frutose-1,6-bisfosfato aldolase

Database: nr 449,381 sequences; 138,416,789 total letters

AC # (gi)	Descrição	Organismo	L	Score (bits)	E-value	Identidades	Positivos	Gaps
2982453	Fructose-bisphosphate aldolase-like protein	Arabidopsis thaliana	358	348	4e-95	179/332 (53%)	233/332 (69%)	3/332 (0%)
1168408	Fructose-bisphosphate aldolase, cytoplasmic isozyme 1	Pisum sativum	357	344	5e-94	175/332 (52%)	229/332 (68%)	3/332 (0%)
418644	Fructose-bisphosphate aldolase, cytosolic	Spinach	357	341	4e-93	175/332 (52%)	228/332 (67%)	3/332 (0%)
2213867	Fructose-bisphosphate aldolase	Mesembryanthemum crystallinum	357	340	1e-92	174/332 (52%)	228/332 (68%)	3/332 (0%)
113624	Fructose-bisphosphate aldolase, cytoplasmic isozyme	Spinacia oleracea	357	340	1e-92	174/332 (52%)	228/332 (68%)	3/332 (0%)
4581146	Putative fructose bisphosphate aldolase	Arabidopsis thaliana	358	333	2e-90	171/332 (51%)	225/332 (67%)	3/332 (0%)
3913008	Fructose-bisphosphate aldolase, cytoplasmic isozyme	Cicer arietinum	359	332	3e-90	172/333 (51%)	228/333 (67%)	4/333 (1%)
2130073	Fructose-bisphosphate aldolase isoenzyme C-1, cytosolic	Oryza sativa	358	331	6e-90	169/332 (50%)	227/332 (67%)	3/332 (0%)
1351913	Fructose-bisphosphate aldolase 2 (ALDO2)	Plasmodium berghei	358	325	4e-88	168/333 (50%)	223/333 (66%)	3/333 (0%)
5834626	Fructose-1,6-bisphosphate aldolase	Leishmania mexicana	371	289	3e-77	153/332 (46%)	202/332 (60%)	5/332 (1%)
3913016	Fructose-bisphosphate aldolase class I (fbp aldolase)	Synechocystis sp.	300	52.3	6e-06	73/284 (25%)	123/284 (42%)	38 / 284 (13%)

VI.O Alinhamento de frutose-1,6-bisfosfato aldolases com a suposta enzima de *Xylella fastidiosa*

Uma busca por domínios na seqüência protéica codificada pela suposta ORF de frutose-1,6-bisfosfato aldolase através do serviço Pfam[53] resultou na identificação do domínio de frutose-1,6-bisfosfato aldolase classe I. A partir desse resultado obtiveram-se as seqüências protéicas utilizadas no alinhamento base para busca de domínios. Tais seqüências e a seqüência protéica da suposta frutose-1,6-bisfosfato aldolase de *Xylella fastidiosa* (XF0826 U) foram comparadas utilizando o programa ClustalW, gerando o alinhamento a seguir.

		10	20	30	40	50	60
ALF1_PEA//	1	----	E I I N R Y Y A T P G K G I L A A D E S T G T I G K R L A S I N V E N I E A N R Q I L R E L L F T S P N - A				
ALF2_PEA//	1	----	E I I A N A Y G T P G K G I L A A D E S T G T I G K R L S S I N V E N V E S N R Q I L R E L L F T A S W L F				
ALF_MAIZE/	1	----	E I I N R A Y G T P G K G I L A A D E S T G T I G K R L S S I N V E N V E E N R R A L R E L L F C C P G - A				
ALF_ARATH/	1	----	E I I T A R Y A T P G R G I L A A D E S T E T I G K R F A G I N V E N T E S N R Q A Y R E L L F T S P G - S				
ALF2_CAEEL	1	----	E I H Q I A L K V Q D G K G I L A A D E S T G T I G K R L D A I N L E N N E T N R Q K Y R O L L F T T P N - L				
ALFA_HUMAN	1	----	E S D I A H R V A P G K G I L A A D E S T G S L A K R L Q S I G T E N T E E N R R F Y R O L L T A D D R V				
ALFB_CHICK	1	----	A S D I A Q R V A S G K G I L A A D E S V G T M G N R L Q R I N V E N T E E N R R A F R E I L F S S D A S I				
ALFR_DROME	1	----	E R E I A Q K V A P G K G I L A A D E S G P T M G K R L Q D I G V E N T E D N R R A Y R O L L F S T D P K L				
ALF_PLAFA/	1	----	E A T T A Q K V Q A G K G I L A A D E S T Q T T K K R F D V I K E N T I E N R A S Y R D L L E G T K G - L				
ALFC_PEA//	1	----	E V I T A T T A S P G R G I L A M D E S N A T C G K R L A S I G L E N T E V N R Q A W T L V T V P T - L				
ALF_TRYBB/	1	----	E I I T A K K T A P G K G I L A A D E S T G S C S K F F G G S Y T A E H R F Q R A L M L E C E G - F				
ALF_STACA/	1	----	N G K G F I A A D Q S G G S T P K L L D Y G V E E N E Y S - - N D E M F N L V H D M T R I I T S P A F N				
XF0826 U	1	----	M S I E Q A E T A G A V A S G K G I L I D E S A G T A K H F S S G E I I E E N R R A Y R E L L T T P K -				
		70	80	90	100	110	120
ALF1_PEA//	56	L Q Y S G V I L E E T L Y Q N S S E G K P V E L L Q E N N V I P G I K V D K G V V E L A G T D - G E T T T Q G F D					
ALF2_PEA//	57	L Q Y S G V I L E E T L Y Q N T A A G K P V D V L N E A G V L P G I K V D K G T V E L A G T D - G E T T T Q G L D					
ALF_MAIZE/	56	L Q Y S G V I L E E T L Y Q N T K D G K P V D V L K E G G V L P G I K V D K G T E V G T D - K E T T T Q G H D					
ALF_ARATH/	56	Y P C S G V I L E E T L Y Q N T S D G K P V D L M E N G V I P G I K V D K G L V D L A G T N - G E T T T Q G L D					
ALF2_CAEEL	56	N Q H S G V I L E E T F H Q S T D K G E K T D L I K Q G I V P G I K L D L G V V P L A G T I - G E G T T Q G L D					
ALFA_HUMAN	57	N P C S G G V I L E H E T L Y Q A D D G P P Q Y I K S K G G V G I K V D K G V P L A G T N - G E T T T Q G L D					
ALFB_CHICK	57	S I S G G V I L E H E T L Y Q D S S G P P A L I K E K G V V G K L D A G T A P L A G T N - G E T T I Q G L D					
ALFR_DROME	57	A E N S G V I L E H E T L Y Q A D D G T P A E I I K K G I L G K V D E G V V P L F G S E - D E V T T Q G L D					
ALF_PLAFA/	56	G K F I S G A L L E E T L R Q N E A G V P M V L I H N E N I I P G I K V D K G V N I P C T D - E E K S T Q G L D					
ALFC_PEA//	56	G E Y I S G A L L E E T L R Q S T T D G K I V D V L I E O N I I P G I K V D K G L V P L A G S N - D E S W C Q G L D					
ALF_TRYBB/	56	E Q Y S G V I L H E T L Y Q A K T G E T P Q Y I R R G V P G I K T D C G L E P I V E G A K G E Q M T A G E D					
ALF_STACA/	55	G E K I L G A L L E E T L D E V - E G Y T G S Y A D K G V P F I K V D K G L A E E D G V Q L M K P I P D I D					
XF0826 U	60	S D Y I S G A L L E E T L R Q S T K A G V P P K Y I A D H G I I P G I K V D K G A Y P L A G C P - G E L V T E G L D					
		130	140	150	160	170	180
ALF1_PEA//	115	S G A R C Q Q Y E A G A R F A K W R A V L K I G P N E - - P S E S I Q Q N A Q G L A R Y A I I C Q E N G L V L F V					
ALF2_PEA//	116	G G A R C R Q Y E A G A R F A K W R A V L K I G A N E - - P S E H S I H E N A Y G L A R Y A V I C Q E N G L V P T V					
ALF_MAIZE/	115	D G K P C A K Y E A G A R F A K W R A V L K I G P N E - - P S Q A I D L A Q G L A R Y A I I C Q E N G L V P T V					
ALF_ARATH/	115	S G A R C Q Q Y E A G G R F A K W R A F F K I G A T E - - P S V I S Q E D A R V L A R Y A I I C Q E N G L V P T V					
ALF2_CAEEL	115	K L A E P A A A K G G C G F A K W F C I N I G T H T - - P S H I G L E N A N V L A R Y A S I C Q A N G L V P T V					
ALFA_HUMAN	116	G S E R C A Q K D G D P A K W F C I L K I G E H T - - P S A A I M E N A N V L A R Y A S I C Q Q N G L V P T V					
ALFB_CHICK	116	K L A E R C A Q K D G D P G W R A V L K I S S T T - - P S Q A I Q E N A N T L A R Y A S I C Q Q N G L V P T V					
ALFR_DROME	116	D L A A R C A Q K D G C D P A K W F C I L K I G K N T - - P S Y Q S I L E N A N V L A R Y A S I C Q S Q R I V P T V					
ALF_PLAFA/	115	G L A E P C K E Y A G A R F A K W F T I V I D T A K G K P T D S I H E T W G L A R Y A S I C Q Q R A V P T V					
ALFC_PEA//	115	G L A S S A A Y Q Q G A R F A K W F T V S I P N G P - - S A L A V K E A W G L A R Y A A I S O D N G L V P T V					
ALF_TRYBB/	116	G Y I K A K Y I A M G C R I C Q F N Y K I Q N G T - - V S E A V R F N A E T L A R Y A I I S Q L C G L V P T V					
ALF_STACA/	114	K I L D A N E R - - G I F G T M S N I L E N N K E - - - - - A I E I V V K Q Q F E V I K E I I A A G L V P T I					
XF0826 U	119	G R A L E E Y I L G A R F A K W R A V I N I G D D I - - P S G C I D A V H A L A R Y A A L C Q E Q G L V P M V					

190 200 210 220 230 240

173 EPEITVDGSHDILKCAAEITETVLAACFKALNDHHVILEGTLTKLPNMVTPGSD-SPVVSPE
 ALF2_PEA// 174 EPEITVDGSHDILKCAAEITERVLAATFKALSDHNVILEGTLTKLPNMVTPGSD-APVVSPE
 ALF_MAIZE/ 173 EPEITVDGPHDIDRCAYVTEETVLAACFKALNEHHVILEGTLTKLPNMVTPGSD-SKVVTPE
 ALF_ARATH/ 173 EPEVITGGSHDILKCAAVTEETVLAAVFKALNYHHVILEGTLTKLPNMVTPGSD-SPVVSPE
 ALF2_CAEEL 173 EPEVTCDCGHDILAKAQKITEQLAFVYKALADHHVILEGTLTKLPNMVTPGQSSASASHE
 ALFA_HUMAN 174 EPEITPDGDHDLKRCQYVTEKFLAAVYKALSDHNVILEGTLTKLPNMVTPGHACTQKFSHE
 ALFB_CHICK 174 EPEVLPDGDHDLQRCQYVTEKFLAAVYKALNDHNVILEGTLTKLPNMVTPAGHSCPKQVYTPQ
 ALFR_DROME 174 EPEVLPDGDHDLGDAQKITEETVLAAVYKALSDHNVILEGTLTKLPNMVTPAGQSA-AKQNTPE
 ALF_PLAFA/ 175 EPEILLADGPHSEVCAVVTQKQLSCVFKALQENGVILEGALLKLPNMVTPAGYECTAATTTQ
 ALFC_PEA// 172 EPEITLDGEGHDIETFEVAQKWVEVYVLAENNFQFEGILKLPNMVTPGQESKDKASPT
 ALF_TRYBB/ 174 EPEVLDGTHDIEETQRVQHWSEVSLHRRGVWEGCLKLPNMVTPGQESGLQGHAE
 ALF_STACA/ 165 EPEVNIYAKDKAEATEANLAEAKKEDNLKQDQVWVK---TIPTKNAYSE-----
 XF0826 U 177 EPEVLDGDNHDLSAAYEATEATRSENLYEQNVILEGTLTKLPNMVTPGTDCEEASID

250 260 270 280 290 300

232 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALF2_PEA// 233 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALF_MAIZE/ 232 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALF_ARATH/ 232 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALF2_CAEEL 233 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALFA_HUMAN 234 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALFB_CHICK 234 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALFR_DROME 233 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALF_PLAFA/ 235 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALFC_PEA// 232 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALF_TRYBB/ 234 VLAEHTYRALQRTVPAAVPAVFLSGGQSEEEATNLNAINQKGGKPPKTSSTSGRALQ
 ALF_STACA/ 214 -----LIEHPQVIRVVALSGGYSRDEENKIHKOND-----GVIASVSRAL-
 XF0826 U 237 EVAEESTVLCVSTVPIIPGGVFLSGGQSDTAOSTAHNAINNOLD-PLPVPKTSSTSGRALQ

310 320 330 340 350

292 QSTKATVSGKKE-NVGAQDVFQARCANSEATLGTKGGGGS-GTGLVSESHKDKDKK
 ALF1_PEA// 293 QSTKAWGGKTE-NVGAQDAELTRAKANSEATLGTKGGASNLGAGVSESHKDKDKK
 ALF2_PEA// 293 QSTKAWGGKTE-NVGAQDAELTRAKANSEATLGTKGGASNLGAGVSESHKDKDKK
 ALF_MAIZE/ 292 QSTKAWAGKVE-NLEKAAAFQARCANSEATLGTKGGDAAADT---ESHKDKDKK
 ALF_ARATH/ 292 QSAKAWAGKPE-NVAKAAQKFTCKKANKDATLGTKGGWASGDSAIFENVVIGVRY
 ALF2_CAEEL 293 QSVKAWGGKDE-NLAAQKTLHRSKANGDASLGKYGEDAAG-AAESVFAKHSY
 ALFA_HUMAN 294 QSAKAWGGKKE-NLAAQEEYKRALNSLACQKQTPSGQAGAAVSESHKDKDKK
 ALFB_CHICK 294 QSAKAWGGKSE-NKAAQAEAFCKRAQINSLACRGQVVTSGKTDATQSFASVTV
 ALFR_DROME 293 QSVKAWAGKKE-NLAAQONEELKRAKANGACGNVTAGSVKGFAGKDTIHDHRP
 ALF_PLAFA/ 294 QSVNNTQGGKKE-NVAKAREVLLQRAEANSLATYGGKKGAGGEN-AGAVYKDKVY
 ALFC_PEA// 290 NTAKATVGGKPE-NVGAQEAALFRANSNLACQKQVYDGESEE-SKTEVFKGSSV
 ALF_TRYBB/ 294 QSAKRWGGKES-GVAGGRAVHRAKMNSLACQKQVNRADDDKD---SQSYVAGNTY
 ALF_STACA/ 255 VSDNMQQSDAE-FNEKIQEADT----IFDASVNA-----
 XF0826 U 296 QVAKMLVSDMGVFAKAAQVVYERAKEVGLVAGKVKG-----

A assinatura PROSITE[54] para a frutose-1,6-bisfosfato aldolase classe I: [LIVM]-E-x-E-[LIVM]-G-x(2)-[GM]-[GSTA]-x-E, está presente nesse alinhamento da posição 215 à posição 225 (em itálico). O sítio ativo das frutose-1,6-bisfosfato aldolases classe I está presente nessa assinatura (VILEGTLTKLPN). As enzimas presentes no alinhamento, além da *Xylella fastidiosa* (XF0826 U), estão relacionadas abaixo.

ALF1_PEA	Frutose-bisfosfato aldolase, isozima 1 citoplasmática	<i>Pisum sativum</i>
ALF2_PEA	Frutose-bisfosfato aldolase, isozima 2 citoplasmática	<i>Pisum sativum</i>
ALF_MAIZE	Frutose-bisfosfato aldolase, isozima citoplasmática	<i>Zea mays</i>
ALF_ARATH	Frutose-bisfosfato aldolase, isozima citoplasmática	<i>Arabidopsis thaliana</i>
ALF2_CAEEL	Frutose-bisfosfato aldolase 2	<i>Caenorhabditis elegans</i>
ALFA_HUMAN	Frutose-bisfosfato aldolase A	<i>Homo sapiens</i>
ALFB_CHICK	Frutose-bisfosfato aldolase B	<i>Gallus gallus</i>
ALFR_DROME	Proteína relacionada a aldolase	<i>Drosophila melanogaster</i>
ALF_PLAFA	Frutose-bisfosfato aldolase	<i>Plasmodium falciparum</i>
ALFC_PEA	Frutose-bisfosfato aldolase 1, cloroplástica	<i>Pisum sativum</i>
ALF_TRYBB	Frutose-bisfosfato aldolase, glicosomal	<i>Trypanosoma brucei brucei</i>
ALF_STACA	Frutose-bisfosfato aldolase classe I	<i>Staphylococcus carnosus</i>

VI P. Informações sobre o genoma de *Xylella fastidiosa*

Informações disponíveis na página oficial do Projeto Genoma *Xylella fastidiosa*, assim como um mapa dos genes codificados no genoma⁴³.

Cromossomo principal (NCBI AC: AE003849)

Número total de bases: 2.679.305.

G+C: 1.411.300 (52,67 %).

Número total de genes: 2.838;

2.782 genes codificantes;

49 tRNAs;

6 rRNAs;

1 tmRNA.

Orientação de genes U: 1.520

(primeira metade: 871 / segunda metade: 649).

Orientação de genes C: 1.318

(primeira metade: 550 / segunda metade: 768).

312 famílias de parálogos compreendendo 853 genes.

ORFs com designação funcional (genes codificando proteínas): 1.283.

ORFs com similaridade alta a proteínas hipotéticas: 310.

ORFs sem similaridade significativa em bancos de dados: 1.083.

Plasmídeo pXF51 (NCBI AC: AE003851)

Número total de bases: 51.158.

G+C: 25.358 (49,57 %).

Número total de genes: 64;

64 genes codificantes.

Orientação de genes U: 32 (primeira metade: 21 / segunda metade: 11).

Orientação de genes C: 32 (primeira metade: 11 / segunda metade: 21).

ORFs com designação funcional (genes codificantes): 30.

ORFs com similaridade alta a proteínas hipotéticas: 8.

ORFs sem similaridade significativa em bancos de dados: 24.

Plasmídeo pXF1.3 (NCBI AC: AE003850)

Número total de bases: 1.285.

G+C: 714 (55,56 %)

Número total de genes: 2;

2 genes codificantes.

ORFs com designação funcional (genes codificantes): 1.

ORFs com similaridade alta a proteínas hipotéticas: 0.

ORFs sem similaridade significativa em bancos de dados: 1.

rRNAs

Uma cópia localizada entre as bases 66.508 e 71.708 e uma segunda cópia entre as bases 172.224 e 177.424.

O operon de rRNA é formado pelos rRNAs 16S-23S-5S nessa ordem, sendo que Ala-TGC-tRNA e Ile-GAT-tRNA estão presentes entre os rRNAs 16S e 23S. As duas cópias são idênticas!

tRNAs

46 genes de tRNA distintos, em 49 ocorrências (três seqüências estão presentes duas vezes em cópias idênticas).

Todos os aminoácidos estão representados.

Lys-TTT, Ala-TGC, and Ile-GAT estão presentes duas vezes (cópias idênticas).

Gly-GCC está presente duas vezes (cópias quase idênticas).

Met-CAT está presente três vezes. As cópias são marcadamente diferentes.

De todos os anticódons, somente um é iniciado por A (Arg-ACG). Isso significa que somente um códon é terminado por T.

tmRNA

1 tmRNA encontrado, como notado por Kelly Williams (comunicação pessoal, kwilliam@bio.indiana.edu). A fase de leitura para a marca de proteólise é `gccaacgaagacaacttcgctgtagccgct`.

Esse tipo de RNA está descrito em Williams and Bartel, 1996, *RNA* 2:1396, and Keiler, Waller, Sauer, 1996, *Science* 27(1): 990-993. Maiores informações em <http://www.indiana.edu/~tmrna/>.

Genes com “frameshift” e mutação pontual (códon de terminação em fase)

11 genes com “frameshift”.

4 genes com mutação pontual (códon de terminação em fase).

2 genes com mutação pontual e “frameshift”.

Íntron do grupo II

1 íntron do grupo II localizado entre as posições 1691283 e 1693687.
Parte do gene XF1774 (DNA metiltransferase).

Origem de replicação

Essa região começa na primeira base após a proteína ribossomal L34 (gene XF2782) e termina uma base após o *gyrB* (gene XF0005).

Genes relacionados a fagos

Há 4 regiões de profagos (Xfp1, Xfp2, Xfp3, Xfp4).

Há 83 genes relacionados a fagos.

VII. Bibliografia

- 1 <http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/eub.html>
- 2 <http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/a.html>
- 3 <http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/euk.html>
- 4 <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome>
- 5 Pallen, M.J., "Microbial Genomes", *Molecular Microbiology*, **32** (5): 907-912 (1999).
- 6 Strauss, E.J. & Falkow, S., "Microbial pathogenesis: genomics and beyond", *Science*, **276**: 707-712 (1997).
- 7 Machado, M., "Citrus Variegated Chlorosis (CVC), a new destructive citrus disease in Brazil, and the xylem limited bacteria, *Xylella fastidiosa*", <http://www.dcc.unicamp.br/genoma/xylella.html> (1997).
- 8 Rossetti, V., Gonzales, M.A., Donadio, L.C., "History" in "Citrus Variegated Chlorosis", Donadio, L.C. & Moreira, C.S. (eds), Bebedouro, 1998
- 9 Wells, J., Raju, B., Jung, H., Weisburg, W., Mandelco-Paul, L., Brenner, D., "*Xylella fastidiosa* gen nov, sp nov gramnegative, xylem limited fastidious plant bacteria related to *Xanthomonas* ssp.", *International Journal of Systematic Bacteriology*, **37**, 136-143 (1987).
- 10 Purcell, A., <http://nature.berkeley.edu/xylella/portugese/brazil.html> (1999)
- 11 Carlos, E.F., Rodrigues Neto, J., Beretta, M.J.G., "*The bacterium Xylella fastidiosa*" in "Citrus Variegated Chlorosis", Donadio, L.C. & Moreira, C.S. (eds), Bebedouro, 1998
- 12 Hopkins, D.L., "*Xylella fastidiosa*: xylem limited bacteria pathogen of plants", *Annual Reviews of Phytopathology*, **27**: 271-290 (1989).
- 13 Purcell, A.H. & Hopkins, D.L., (1996), *Annual Reviews of Phytopathology*, **34**: 131-151
- 14 Andersen, P.C. & Brodbeck, B.V., "Diurnal and temporal changes in the chemical profile of xylem exudate from *Vitis rotundifolia*", *Physiologia Plantarum*, **75**: 63-70 (1989).
- 15 Hopkins, D.L., "Physiological and pathological characteristics of virulent and avirulent strains of the bacterium that causes Pierce's disease in grapevines", *Physiology and Biochemistry*, **75**(6): 713-717 (1985).

- 16 Fry, S.M., Huang, J.-S., Milholland, R.D., "Isolation and preliminary characterization of extracellular proteases produced by strains of *Xylella fastidiosa* from grapevines", *Biochemistry and Cell Biology*, **84**(4), 357-363 (1994).
- 17 Gravena, S., Lopes, J.R.S., Paiva, P.E.B., Yamamoto, P.T., Roberto, S.R., "The *Xylella fastidiosa* vectors" in "Citrus Variegated Chlorosis", Donadio, L.C. & Moreira, C.S. (eds), Bebedouro, 1998
- 18 <http://www.lbi.dcc.unicamp.br/xylella/strategy.html>
- 19 Sambrook, J., Fritsch, E. E., Maniatis, T., "Molecular Cloning: A Laboratory Guide", 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989
- 20 *Focus* (1986), **8**:2,9
- 21 Lorow, D & Jesse, J, (1990) *Focus* **12**,49
- 22 <http://fugu.hgmp.mrc.ac.uk/fugu/Lawrist4.html>
- 23 Hanahan D., "Studies on transformation of *Escherichia coli* with plasmids", *Journal of Molecular Biology*, **116**(5), 557-80 (1983).
- 24 Watson, J.D., Gilman, M., Winkowski, J., Zoller, M., "Recombinant DNA", 2nd edition, Scientific American Books, New York, 1996
- 25 Voet, D. & Voet, J.G., "Biochemistry", 2nd edition, John Wiley & Sons, New York, 1995
- 26 Ewing, B., Hillier, L., Wendl, M.C., Green, P., "Base-calling of automated sequence tracer using phred. I Accuracy assesment", *Genome Research*, **8**(3), 175-85 (1998).
- 27 Ewing, B., Green, P., "Base-calling of automated sequence tracer using phred. II Error probabilities", *Genome Research*, **8**(3), 186-94 (1998).
- 28 Gordon, D., Abajian, C., Green, P., "Consed: a graphical tool for sequence finishing", *Genome Research*, **8**(3), 195-202 (1998).
- 29 Salzberg, S., Delcher, A., Kasif, S. and White, O., "Microbial gene identification using interpolated Markov models", *Nucleic Acids Research*, **26** (2), 544-548, (1998).
- 30 Benson D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F., Rapp, B.^a, Wheeler, D.L., "Genbank", *Nucleic Acids Research*, **27**(1):12-17 (1999).
- 31 Altschul, S.F. & Gish, W., "Local alignment statistics", *Methods of Enzymology*, **266**:460-480 (1996).
- 32 <http://www.ncbi.nlm.nih.gov/BLAST/>

- 33 Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Research* **25**:3389-3402 (1997).
- 34 Riley, M., "Functions of the gene products of *Escherichia coli*", *Microbiological Reviews*, **57** (4), 862-952 (1993).
- 35 Lowe, T.M. & Eddy, S.R., "tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence", *Nucleic Acids Research*, **25**, 955-964 (1997).
- 36 <http://www.genome.ad.jp/kegg/kegg2.html>
- 37 <http://ecocyc.pangeasystems.com/>
- 38 <http://www.ncbi.nlm.nih.gov/ENTREZ/>
- 39 <http://www.ncbi.nlm.nih.gov/COG/>; Tatusov, R.L., Galperin, M.Y., Koonin, E.V., "The COG database: a tool for genome analysis of protein function and evolution", *Nucleic Acids Research*, **28**(1):33-36 (2000).
- 40 Galperin, M.Y., Walker, D.R., Koonin, E.V., "Analogous Enzymes: independent inventions in enzyme evolution", *Genome Research*, **8**: 779-790 (1998).
- 41 Huynen, M.A., Dandekar, T., Bork, P., "Variation and evolution of the citric-acid cycle: a genomic perspective", *Trends in Microbiology*, **7**(7): 281-291 (1999).
- 42 Makarova, K.S., Aravind, L., Galperin, M.V., Grishin, N.V., Tatusov, R.L., Wolf, Y.I., Koonin, E.V., "Comparative genomics of the Archea (Euryarchaea): evolution of conserved protein families, the stable core, and the variable shell", *Genome Research*, **9**: 608-628 (1999).
- 43 <http://onsona.lbi.dcc.unicamp.br/xf/>
- 44 <http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palog?sys=GLYCOLYSIS>
- 45 <http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palog?sys=GLUCONEOGENESIS>
- 46 <http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palog?sys=ENTNER-DOUDOROFF+PATHWAY>
- 47 <http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palog?sys=TCA+CYCLE>
- 48 <http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palog?sys=GLYOXYLATE+BYPASS>
- 49 <http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palog?sys=PENTOSE+PHOSPHATE+PATHWAY>
- 50 Romano, A.H. & Conway, T., "Evolution of carbohydrate metabolic pathways", 14th Forum in Microbiology, *Research in Microbiology* **147**: 448-455 (1996).

- 51 Kengen, S.W.M., de Bok, F.A.M., van Loo, N.-D., Dijkema, C., Stams, A.J.M., de Vos, W.M., "Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*", *The Journal of Biological Chemistry*, **269** (26): 17537-17541 (1994).
- 52 Conn, E.E. & Stumpf, P.K., "Outlines of Biochemistry" 3rd edition, Wiley International Edition, John Wiley & Sons, Inc., New York, 1972.
- 53 <http://pfam.wustl.edu/index.html>
- 54 <http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00143>

CURRICULUM VITAE

NOME: **ARI JOSÉ SCATTONE FERREIRA**

NASCIMENTO: São Paulo, 09 de Janeiro de 1976

EDUCAÇÃO:

Ensino Médio: Colégio Bandeirantes, São Paulo, 1993

Ensino Superior: Universidade de São Paulo, São Paulo, 1997
Bacharelado em Química com atribuições biotecnológicas.

OCUPAÇÃO:

Aluno de Pós-Graduação do Instituto de Química da
Universidade de São Paulo, 1998 até o presente.

Bolsista: Bolsa de Mestrado do Conselho Nacional de Pesquisa,
Fev/1998 a Jan/2000.

PRÊMIOS RECEBIDOS:

Prêmio de Mérito Científico e Tecnológico
Governo do Estado de São Paulo, 2000

Prêmio Lavoisier
Conselho Regional de Química – IV Região, 1998