INSTITUTO DE QUÍMICA
Universidade de São Paulo
Universidade de São Paulo

Instituto de Química

**Seqüenciamento e anotação de parte do genoma de *Xylella fastidiosa* e
análise das vias de biossíntese de pequenas moléculas e cofatores**

Cássia Docena

Dissertação de Mestrado

Orientador: Prof. Dr. Ronaldo Bento Quaggio

São Paulo

26 de outubro de 2000

Ficha Catalográfica
Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Docena, Cássia
D636s Sequenciamento e anotação de parte do genoma de *Xylella fastidiosa* e análise das vias de biossíntese de pequenas moléculas e cofatores / Cássia Docena. -- São Paulo, 2000.
156p.

Dissertação (mestrado) – Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica.
Orientador : Quaggio, Ronaldo Bento

1. Bioquímica molecular 2. Biologia molecular I. T. II.
Quaggio, Ronaldo Bento, orientador.

574.88 CDD

**“Seqüenciamento e Anotação de Parte do Genoma de
Xylella fastidiosa e Análise das Vias de Biossíntese de
Pequenas Moléculas e Cofatores”.**

CÁSSIA DOCENA

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

Aprovada por:

Prof. Dr. RONALDO BENTO QUAGGIO
IQ - USP
(Orientador e Presidente)

Prof. Dr. PEDRO SOARES DE ARAUJO
IQ - USP

Prof. Dr. ARTHUR GRUBER
FMVZ - USP

SÃO PAULO
26 DE OUTUBRO DE 2000.

AGRADECIMENTOS

Ao Prof. Dr. Ronaldo Bento Quaggio, agradeço não somente a orientação neste projeto, mas também toda a sua dedicação para meu aprendizado.

A atenção que recebi dos professores Dra. Ana Cláudia Rasera da Silva e Dr. Fernando de Castro Reinach foi essencial para a realização do meu trabalho.

Agradeço toda a ajuda e companherismo aos amigos do projeto: Ari José Scattone Ferreira, Adriana Matsukuba e Cássio Baptista que sempre estiveram ao meu lado. Gostaria de agradecer também aos colegas de laboratório Frank Jefferson Alarcon Barrientos, Elaine Cristina Favaro, Eric Kors Vidsiunas, Ilda Souza Costa e Daniela Costa. Ao amigo que infelizmente não está mais entre nós Dellano Marcatti de Freitas.

Foi também indispensável a ajuda que recebi de todos os que colaboraram indiretamente com o projeto.

Aos meus pais e às minhas irmãs, agradeço todo o apoio que me deram para a realização de mais esta etapa profissional.

Agradeço à FAPESP pela bolsa de mestrado concedida.

1. INTRODUÇÃO	5
I.1. ORGANIZAÇÃO GERAL DO PROJETO	7
I.2. ESTRATÉGIAS DE CLONAGEM E MONTAGEM DE GENOMAS BACTERIANOS	8
I.3. A CLOROSE VARIEGADA DOS CITRUS.....	11
I.4. A BACTÉRIA XYLELLA FASTIDIOSA WELLS.....	13
I.5. TRANSMISSÃO E CONTROLE.....	15
I.6. COLONIZAÇÃO	17
II. OBJETIVOS.....	20
III. MATERIAL E MÉTODOS	21
III.1. SEQUENCIAMENTO DO GENOMA DE XYLELLA FASTIDIOSA	21
III.2. SEQUENCIAMENTO DE COSMÍDEOS PELO LABORATÓRIO QR.....	22
1- MEIOS DE CULTURA.....	22
2- ANTIBIÓTICOS	23
3- ENZIMAS.....	23
4- CEPAS DE <i>ESCHERICHIA COLI</i> UTILIZADAS NAS BIBLIOTECAS.....	23
5- VETORES UTILIZADOS NAS BIBLIOTECAS	23
6- OBTENÇÃO DE COSMÍDEOS EM LARGA ESCALA	23
7- BIBLIOTECA.....	24
8- OBTENÇÃO DE PLASMÍDEOS PARA SEQUENCIAMENTO (“MINIPREPS”).....	25
9- SEQUENCIAMENTO	25
10- MONTAGEM (“FINISHING”) DOS COSMÍDEOS	26
11- CONSTRUÇÃO DE “PRIMERS”	27
12- PROCURA E ANOTAÇÃO DE ORFS.....	27
13- BUSCA DE TRNAs.....	28
14- ANOTAÇÃO DOS COSMÍDEOS.....	28
15- REVISÃO DA CATEGORIA II	30
IV. RESULTADOS OBTIDOS.....	32
IV.1 BIBLIOTECAS DE NEBULIZAÇÃO E RESTRIÇÃO	32
IV.2. SEQUENCIAMENTO DE COSMÍDEOS.....	33
IV.3. ANOTAÇÃO	34
3.1 ANOTAÇÃO DO COSMÍDEO 02A11	34
3.2 ANOTAÇÃO DO COSMÍDEO 05B03	36
3.3 ANOTAÇÃO DO COSMÍDEO 02D09	38
3.4 ANOTAÇÃO DO COSMÍDEO 09A12	42
3.5 ANOTAÇÃO DO COSMÍDEO 07H03.....	44
IV.4. ANÁLISE DA CATEGORIA DE BIOSÍNTESE DE PEQUENAS MOLÉCULAS.....	47
IV.4.1. AMINOÁCIDOS	48
4.1. FAMÍLIA DO GLUTAMATO	49
4.2. FAMÍLIA DO ASPARTATO/ PIRUVATO	52
4.3. FAMÍLIA GLICINA/ SERINA	56
4.4. FAMÍLIA DOS AMINOÁCIDOS AROMÁTICOS.....	58
4.5. HISTIDINA	59
IV.4.2. NUCLEOTÍDEOS.....	60
4.2.1. RIBONUCLEOTÍDEOS DE PURINA	60
4.2.2. RIBONUCLEOTÍDEOS DE PIRIMIDINA.....	60

4.2.3. 2'-DESOXIRRIBONUCLEOTÍDEOS.....	61
4.2.4. SALVAMENTO DE NUCLEOSÍDEOS E NUCLEOTÍDEOS	61
IV.4.3. NUCLEOTÍDEOS E NUCLEOTÍDEOS DE AÇÚCARES	62
IV.4.4. COFATORES, GRUPOS PROSTÉTICOS E TRANSPORTADORES	62
4.4.1. BIOTINA	62
4.4.2. ÁCIDO FÓLICO.....	62
4.4.3. LIPOATO	63
4.4.4. MOLIBDOPTERINA	64
4.4.5. PANTOTENATO E COENZIMA A.....	64
4.4.6. PIRIDOXINA.....	65
4.4.7. NUCLEOTÍDEOS DE PIRIDINA.....	65
4.4.8. TIAMINA.....	66
4.4.9. RIBOFLAVINA.....	66
4.4.10. TIORREDOXINA, GLUTARREDOXINA, GLUTATIÃO.....	67
4.4.11. MENAQUINONA (VITAMINA K2) E UBIQUINONA (COENZIMA Q)	67
4.4.12. PROTO E SIROHEME	68
4.4.13. PROTEÍNA TRANSPORTADORA DE CARBOXIL BIOTINA (BCCP)	69
4.4.14. COBALAMINA	69
4.4.15. ENTEROBACTINA.....	69
4.4.16. BIOPTERINA	70
IV.4.5. BIOSÍNTESE DE ÁCIDOS GRAXOS E FOSFATÍDICOS.....	70
IV.4.6. POLIAMINAS	70
<u>V. DISCUSSÃO</u>	<u>71</u>
V.I. ESTRATÉGIAS DE SEQUENCIAMENTO.....	71
V.II ANÁLISE DA CATEGORIA DE BIOSÍNTESE DE PEQUENAS MOLÉCULAS.....	72
V. III. CONCLUSÃO.....	77
<u>V.I BIBLIOGRAFIA</u>	<u>79</u>
<u>V.II. ANEXOS</u>	<u>85</u>

ABREVIATURAS

ATP: adenosina trifosfato;

dATP, dCTP, dGTP, dTTP: 2'desoxinucleosídeos 5'trifosfato de adenina, citosina, guanina e timina respectivamente;

ddNTP: didesoxinucleosídeos 5'trifosfato de quaisquer bases nitrogenadas;

dNTP: 2'desoxinucleosídeos 5'trifosfato de quaisquer bases nitrogenadas;

EDTA: ácido etilenodiaminotetracético;

EPS: Polissacarídeos extracelulares, "extracellular polysaccharide";

IPTG: isopropil-tio- β -galactopiranosídeo;

kDa: quilodaltons;

kb: quilobases;

kV: quilovolts;

L: litro

M: molar;

mM: milimolar;

mL: mililitro;

mg: miligrama;

ng: nanograma;

nm: nanômetro;

nt: nucleotídeo;

pb: pares de bases;

PCR: reação de polimerização em cadeia;

PD: "Pierce's disease";

p.: página;

RNAse: ribonuclease A;

rpm: rotações por minuto;

V: volts;

μ L: microlitro;

μ g: micrograma;

X-gal: 5-bromo-4-cloro-3-indolil-B-D-galactopiranosídeo;

Ω : ohms

RESUMO

Agente da clorose variegada dos *Citrus* (CVC) ou amarelinho, *Xylella fastidiosa*, uma bactéria gram-negativa limitada ao xilema, tem grande importância na agricultura do Estado de São Paulo. Identificada em 1987, na região de Bebedouro, a CVC já está presente em quase todas as áreas citrícolas do país, afetando a produção de frutos e seus derivados.

O projeto genoma, criado pela FAPESP em 1997 com o apoio do FUNDECITRUS, habilitou laboratórios da capital e interior paulistas a seqüenciar o genoma deste fitopatógeno, contribuindo internacionalmente para a elucidação de sua completa informação genética.

Nosso laboratório (QR) obteve aproximadamente 1.900 leituras de bibliotecas aleatórias de “shotgun”. Seqüenciamos 13 cosmídeos perfazendo um total linear de 533.571 pares de bases. Anotamos 125 genes prováveis de cinco cosmídeos e realizamos a análise de uma categoria dos genes anotados, referente à biossíntese de pequenas moléculas, onde incluem-se a biossíntese de aminoácidos, purinas e pirimidinas, ácidos graxos, poliaminas, cofatores, grupos prostéticos e transportadores de elétrons.

X. fastidiosa possui, mesmo que incompletas, as vias metabólicas para a produção de todos os aminoácidos, sendo provavelmente a glutamina, o principal meio para a obtenção de nitrogênio. Estão presentes as vias para a biossíntese de ribonucleotídeos de purina e pirimidina. *X. fastidiosa* parece conseguir alongar ácidos graxos a partir de acetil-CoA e é capaz de sintetizar uma grande variedade de cofatores e grupos prostéticos, entretanto, estão ausentes os genes envolvidos na biossíntese de biotina, cobalamina e enterobactina.

1. INTRODUÇÃO

O seqüenciamento de DNA é a determinação de toda ou parte da seqüência nucleotídica de sua molécula. Nos últimos 20 anos, desde que os métodos de seqüenciamento foram primeiramente introduzidos, o seqüenciamento de DNA e conseqüentemente de genomas inteiros revolucionaram as bases da Biologia Molecular.

Técnicas para o seqüenciamento foram desenvolvidas recentemente. A introdução do seqüenciamento automático com marcadores fluorescentes em 1986 (SMITH *et al.*, 1986) possibilitou a expansão da ciência dos genomas. O primeiro artigo que descreve o seqüenciamento completo de um microrganismo foi publicado em 1977 com o seqüenciamento do Bacteriófago phi-X174 (5.386 nt) realizado por Sanger (SANGER *et al.*, 1977a). O genoma da varíola (poxvírus) com 186.000 pares de bases foi o primeiro a ser completamente seqüenciado pela tecnologia de seqüenciamento automático (MASSUNG *et al.*, 1993). Deste modo, o seqüenciamento genômico de vários organismos como bactérias, arqueobactérias e eucariotos tem crescido exponencialmente a uma extensão tal, que o seqüenciamento completo do genoma humano já está praticamente concluído. Recentemente foram publicadas a seqüências dos cromossomos 22 e 21 (DUNHAM *et al.*, 1999 e HATTORI *et al.*, 2000), há dois anos as seqüências dos cromossomos II e III de *Plasmodium falciparum* (GARDNER *et al.*, 1998 e BOWMAN *et al.*, 1999) e há um ano a seqüência do cromossomo I de *Leishmania major*. Atualmente, por volta de vinte e oito genomas bacterianos já se encontram totalmente seqüenciados (tabela 1); os genomas eucarióticos de *Saccharomyces cerevisiae* (GOFFEAU *et al.*, 1996 e 1997), *Caenorhabditis elegans* (Sequencing Consortium, 1998) e *Drosophila melanogaster* (MYERS *et al.*, 2000) foram completados e o genoma de *Arabidopsis thaliana* (LIN & Genome Consortium, 1999) com dois cromossomos já seqüenciados deve estar finalizado em poucos meses, assim como a seqüência de outros organismos de importância médica, econômica ou científica (atualizações constam no endereço eletrônico: www.ncbi.nlm.nih.gov/Entrez/Genome/main_genomes.html).

As razões fundamentais para o conhecimento da seqüência nucleotídica de uma molécula de DNA são as predições a respeito de sua função. No curso do seqüenciamento de diversos genomas demonstrou-se que muitos genes são descobertos somente quando suas seqüências totais tornam-se conhecidas. Entretanto, o conhecimento da estrutura de um gene é somente parte da resposta que precisamos; a necessidade vai mais além: está na elucidação da função desses genes através de uma nova abordagem em biologia freqüentemente chamada de genoma funcional — a interpretação da função dos genes (JORDAN & PASSOS, 2000).

Como todos os organismos pertencem a uma árvore evolutiva comum, o estudo do genoma de um organismo pode propiciar informações a respeito de organismos filogeneticamente relacionados. Assim, comparações entre genomas distantemente relacionados propiciam previsões da universalidade de mecanismos biológicos e identificam modelos experimentais para se estudar processos complexos. Comparações entre genomas fortemente relacionados podem proporcionar previsões detalhadas da estrutura de um gene e sua função (COLLINS *et al.*, 1998).

Tabela 1: Genomas seqüenciados. Fonte: COG Database. Julho 2000. <http://www.ncbi.nlm.nih.gov/COG> (exceto *X. fastidiosa*, *C. elegans*, *D. melanogaster*, *P. aeruginosa* e *Buchnera sp*, fonte Entrez).

Microorganismo	Genoma (pares de bases)	Conclusão	Genes
<i>Haemophilus influenzae</i>	1.830.138	1995	1.709
<i>Mycoplasma genitalium</i> G37	580.074	1995	480
<i>Mycoplasma pneumoniae</i> M129	816.394	1996	677
<i>Synechocystis</i> PCC6803	3.573.470	1996	3.169
<i>Borrelia burgdorferi</i>	1.519.856	1997	1.256
<i>Aquifex aeolicus</i>	1.551.335	1997	1.522
<i>Helicobacter pylori</i> 26695	1.667.867	1997	1.553
<i>Methanobacterium thermoautotrophicum</i> Δ4	1.751.377	1997	1.869
<i>Bacillus subtilis</i>	4.214.814	1997	4.100
<i>Archaeoglobus fulgidus</i>	2.178.400	1997	2.407
<i>Methanococcus janashii</i>	1.664.970	1998	1.715
<i>Treponema pallidum</i>	1.138.011	1998	1.031
<i>Chlamydia trachomatis</i> serovar D	1.042.519	1998	894
<i>Pyrococcus horikoshii</i> OT3	1.738.505	1998	2.064
<i>Mycobacterium tuberculosis</i> H37 Rv	4.411.529	1998	3.918
<i>Escherichia coli</i> K-12 MG1655	4.639.221	1998	4.289
<i>Rickettsia prowazekii</i>	1.111.523	1998	834
<i>Chlamydia pneumoniae</i>	1.230.230	1998	1.052
<i>Caenorhabditis elegans</i>	Aproximadamente 97 milhões	1998	19.000
<i>Aeropyrum pernix</i> K1	1.669.695	1999	2.694
<i>Helicobacter pylori</i> J99	1.643.831	1999	1.491
<i>Thermotoga maritima</i>	1.860.725	1999	1.846
<i>Saccharomyces cerevisiae</i>	12.068.327 (16 cromossomos)	1999	5.932
<i>Drosophila melanogaster</i>	Aproximadamente 137 milhões	2000	13.600
<i>Campilobacter jejuni</i>	1.641.481	2000	1.654
<i>Xylella fastidiosa</i> 9a5c	2.731.573	2000	2.782
<i>Pseudomonas aeruginosa</i> PA01	6.3 milhões	2000	5.570
<i>Buchnera sp</i> APS	648.467	2000	590

Os seqüenciamentos de genomas completos levam à identificação de genes inerentes a determinados organismos que conferem a diversidade taxonômica existente em nosso meio ambiente. A anotação dos genes tem um importante papel inicial na identificação de processos metabólicos e sua importância para o organismo. Pressupondo que para cada patógeno existe uma ou mais vias metabólicas envolvidas com mecanismos específicos de patogenicidade, fica clara a importância da

identificação de genes dessas vias. Em combinação com experimentos para a determinação de sua função, estratégias baseadas em genoma devem favorecer estudos relacionados à patogenicidade bacteriana.

O projeto de seqüenciamento do genoma da bactéria *Xylella fastidiosa* foi desenvolvido pela FAPESP em 1997 conjuntamente com o FUNDECITRUS (Fundo Paulista de Defesa da Citricultura) com o intuito de criar-se um consórcio de laboratórios que, juntos, fariam parte de uma organização para seqüenciamento: a rede ONSA (“Organization for Nucleotide Sequencing and Analysis”).

I.1. ORGANIZAÇÃO GERAL DO PROJETO

A supervisão geral do projeto foi realizada por um Comitê, denominado “Steering Committee”, nomeado pela FAPESP e composto por pesquisadores experientes em seqüenciamento. A estratégia de seqüenciamento do genoma foi definida por um Laboratório Coordenador de DNA responsável pela geração e distribuição de fragmentos do genoma para seqüenciamento com a colaboração de mais 35 laboratórios, dois dos quais, considerados centrais de seqüenciamento e de apoio técnico aos demais.

Um Centro de Bioinformática na Unicamp (Universidade Estadual de Campinas) responsabilizou-se pelo estabelecimento de ferramentas para a coleta e centralização dos dados das seqüências geradas pelos laboratórios de seqüenciamento, assim como a determinação e controle de sua qualidade. Também esteve envolvido na análise da seqüência completa do genoma e sua anotação.

O clone de *Xylella fastidiosa* (conhecido como 8.1.b clone 9.a.5.c) utilizado para seqüenciamento foi fornecido pelo Laboratório de Biologia Celular e Molecular dirigido pelo Prof. J.M. Bové e Dr. M. Garnier no Centro de Pesquisas de Bordeaux no Institut National de la Recherche Agronomique, França. Esse clone derivou-se da cultura original de *X. fastidiosa* (conhecido como 8.1.b) o qual foi submetido ao postulado de Koch (CHANG *et al.*, 1993a).

As estratégias para seqüenciamento envolveram a construção de bibliotecas de plasmídeos, contendo fragmentos do genoma gerados de forma aleatória por nebulização (“shotgun”) ou por enzimas de restrição, e de bibliotecas de cosmídeos, contendo fragmentos de aproximadamente 40kb, que cobriam o genoma total de forma redundante. A elaboração de sub-bibliotecas destes cosmídeos, bem como as técnicas utilizadas para seu seqüenciamento foi decidida individualmente pelos laboratórios envolvidos no projeto.

1.2. ESTRATÉGIAS DE CLONAGEM E MONTAGEM DE GENOMAS BACTERIANOS

A primeira precaução a ser tomada antes do início da construção de uma biblioteca genômica é certificar-se de que o DNA do organismo é estável nas condições de cultura. Em bactérias com tempo de geração muito curtos as regiões flanqueadoras a origem de replicação podem estar super-representadas de 4 a 8 vezes quando comparadas a região de término da replicação, o que pode ser minimizado com a adição de cloranfenicol à cultura bacteriana. É aconselhável a extração do DNA no estágio inicial da fase estacionária, embora um argumento contra a sua utilização seja a difícil lise da célula nesta etapa quando comparada à fase exponencial (FRANGEUL *et al.*, 1999). Adicionalmente, regiões com seqüências repetidas são difíceis de ter seu seqüenciamento montado e podem ocasionar a perda de um segmento.

Muitos projetos de seqüenciamento genômico de organismos modelos como *S. cerevisiae* (GOFFEAU *et al.*, 1996 e 1997), *E. coli* (BLATTNER *et al.*, 1997) e *B. subtilis* (KUNST *et al.*, 1997), foram completados com base numa gama de informações preliminares, incluindo o mapa físico desses organismos, assim como uma biblioteca ordenada de cosmídeos, Lambda, YACs (“yeast artificial chromosome”) ou BACs (“bacterial artificial chromosome”). Contudo, numa nova abordagem de seqüenciamento, a qual envolve a fragmentação aleatória do genoma total ou “shotgun”, esses dados preliminares não são mais indispensáveis para o início do projeto. Todavia, para obter-se sucesso, características como o tamanho do genoma, o conteúdo G+C e a presença de plasmídeos devem ser levadas em consideração. Mesmo não sendo indispensável, é recomendável a obtenção de um mapa físico do genoma pois, dessa maneira, obtém-se uma maior confiabilidade no seqüenciamento. Em particular, a escolha das estratégias de clonagem tornou-se crucial, uma vez que a construção de bancos genômicos de baixa qualidade pode levar em estágios posteriores à dificuldades na montagem.

Duas estratégias de seqüenciamento são freqüentemente utilizadas. A primeira apóia-se numa biblioteca contendo grandes fragmentos de inserto para a construção de um mapa físico, com clones que se sobrepõem, cobrindo a totalidade do genoma. A segunda utiliza o seqüenciamento direto de clones obtidos por “shotgun”. Outras técnicas como a utilização de transposons e “primer walking” são menos freqüentemente utilizadas como estratégia de seqüenciamento completo.

Muitos métodos têm sido usados para ordenar clones, incluindo análise de restrição de fragmentos por “fingerprinting” e mapas de hibridizações. Este último, identifica segmentos de DNA homólogos onde um grande número de clones podem ser analisados e ordenados simultaneamente.

Atualmente, a estratégia de seqüenciamento mais utilizada é a abordagem por seqüenciamento de bibliotecas geradas aleatoriamente. Este método foi inicialmente aplicado no seqüenciamento de genomas de alguns vírus e da bactéria *H. influenzae* (FLEICHMANN *et al.*, 1995), até que em março de 2000, anunciou-se o seqüenciamento do primeiro eucarioto, *Drosophila melanogaster* (MYERS *et al.*, 2000), pela empresa americana CELERA. Por esse método, um grande número de clones gerados por nebulização aleatória do genoma total do organismo é seqüenciado. Estatisticamente, exige-se uma alta redundância na cobertura das regiões, sendo necessário seqüenciar-se até 10 vezes o tamanho do genoma para cobrir completamente a molécula de DNA. Isso se deve ao fato de que algumas regiões do genoma possuem baixa taxa de eficiência de clonagem, uma vez que podem tratar-se de segmentos que contêm genes cujos produtos são tóxicos à célula hospedeira, o que pode ser resolvido com a clonagem de insertos grandes e pequenos de DNA. Deste modo, ganha-se na qualidade final do seqüenciamento mas ainda assim, apesar da redundância, algumas regiões são representadas por uma única ou nenhuma seqüência.

O seqüenciamento orientado pela técnica de “primer walking” foi primeiramente utilizado na montagem do genoma de *Saccharomyces cerevisiae* (WIEMAN *et al.*, 1993). Uma biblioteca deve ser gerada, freqüentemente de clones grandes com sobreposição, para os quais são desenhados iniciadores com base nas suas extremidades; esses clones servirão, posteriormente, de molde para o seqüenciamento de uma nova região a sua juzante. Neste caso, a redundância pode ser menor que 2,4 vezes o genoma completo (BAPTISTA, 2000) mas perde-se muito tempo na construção de novos iniciadores, além de, no caso de regiões repetitivas, ocorrer alinhamento duplo dos iniciadores na fita de DNA gerando seqüências duplicadas.

Outra metodologia inclui a abordagem por transposons (elementos genéticos móveis capazes de se inserir em outra região do genoma). Por esses métodos (BAPTISTA, 2000), insertos de 3 a 5 kb são clonados e transposons são inseridos, sobre os quais, são desenhados iniciadores. A vantagem está justamente no fato de que somente um par de iniciadores é utilizado, contudo, nenhum genoma foi completamente seqüenciado desta maneira. Comumente, mais de uma estratégia citada é usada para o fechamento de um genoma completo, incluindo também outras metodologias.

Durante a montagem e fechamento de um genoma, é de grande importância a qualidade das seqüências geradas. Muitos programas podem ser utilizados, mas deve-se ter em mente que a escolha da ferramenta correta deve ser feita no início do projeto. A fase de montagem é composta de três etapas principais: a primeira quando é feita a conversão dos dados gerados pelos seqüenciadores automáticos em seqüências nucleotídicas, a segunda, quando essas seqüências são usadas na montagem de

contíguos genômicos e por último, o acesso contínuo desses dados para a montagem de uma única sequência consenso, pois a montagem de mais de 10.000 clones em poucos contíguos, requer tipicamente, programas baseados em algoritmos complexos.

As primeiras ferramentas computacionais utilizadas na montagem de seqüências foram desenvolvidas por STADEN (1979, 1982) e vêm se modernizando nos últimos anos. Os algoritmos mais recentes permitem a decisão automática se duas seqüências se sobrepõem ou não para formar agrupamentos e constroem seqüências consenso. Os programas mais utilizados nesta etapa são o “TIGR Assembler” (SUTTON *et al.*, 1995), CAP2 (HUANG, 1996) e Phrap (P. GREEN, Universidade de Washington, Seattle, EUA). Outras ferramentas como o programa Consed (GORDON *et al.*, 1998) dedicam-se a edição das montagens, mostrando sua estrutura graficamente.

Após a montagem dos contíguos, isto é, de clones que se sobrepõem para formar uma fita consenso contínua, algumas regiões não cobertas (“gaps”) podem ainda permanecer remanescentes. Contudo, a ordem dos contíguos pode ser deduzida através da comparação de regiões codificantes (CDs) com proteínas de um banco de dados; se o final de um contíguo codifica parte de uma proteína e a extremidade de outro contíguo codifica o restante da mesma, provavelmente, esses contíguos são vizinhos. Ainda, a ordem dos genes em operons pode ser conservada entre o genoma que está sendo seqüenciado com outro genoma de um organismo correlato, referido como genoma controle. Neste caso, o genoma controle poderia servir de base na organização dos contíguos do genoma em andamento. Para tanto, é fundamental a identificação de ORFs (“Open Reading Frame” ou fase aberta de leitura) no genoma. Uma ORF corresponde a uma seqüência de DNA com um códon iniciador na extremidade 5’ seguida da região codificadora do gene e de uma região de término (“stop codon”). Todas as regiões codificadoras de proteínas se localizam dentro de uma ORF mas nem todas as ORFs representam um gene. Entretanto, a detecção de ORFs hipotéticas é uma etapa inicial para a busca de genes na ausência de outras informações.

Como citado anteriormente, no caso de *Xylella fastidiosa*, a coordenação do projeto optou por duas estratégias de seqüenciamento: bibliotecas aleatórias geradas por nebulização do genoma total ou por digestão com enzimas de restrição e uma biblioteca de cosmídeos. Para o fechamento do genoma clones de sub-bibliotecas de cosmídeos foram re-seqüenciados. Utilizou-se também o método de “primer walking” além do seqüenciamento de uma biblioteca de Fago Lambda, gerada pelo Laboratório do Prof. Dr. Paulo Arruda na UNICAMP.

1.3. A CLOROSE VARIEGADA DOS CITRUS

A clorose variegada dos *Citrus* (CVC) foi primeiramente observada em 1987, em plantações de laranjas localizadas na região de Bebedouro, centro-oeste do Estado de São Paulo. Inicialmente, foi chamada de “amarelinho”, em decorrência dos sintomas foliares apresentados.

A distribuição da *Xylella fastidiosa* no Brasil tem aumentado nos últimos anos. Em uma década a maioria dos estados cultivadores de *Citrus* relataram a ocorrência de CVC (LARANJEIRA *et al.*, 1996). No cinturão de São Paulo a doença é endêmica mas sua ocorrência e severidade não são uniformes entre as regiões (SALVA *et al.*, 1995).

Em cinco anos (de 1987 a 1993), a doença tornou-se presente em 134 municípios do Estado de São Paulo, 6 em Minas Gerais, 4 no Rio de Janeiro, 3 no Paraná e em outros Estados do país. Uma amostragem feita em Barretos, São Paulo, nessa ocasião mostrou que 12% das árvores estavam infectadas (GONZALES, 1992).

Atualmente, é a doença mais severa da citricultura brasileira, afetando toda a região produtora de citrus. Levantamentos realizados pelo FUNDECITRUS em 1994 e 1996 mostram o crescimento progressivo da doença nos pomares paulistas. Em 1994, 16,6% das plantas apresentavam os sintomas da doença e dois anos depois, 23,7% das plantas inspecionadas estavam contaminadas. Com base nesses dados, a projeção indica que 66 milhões de árvores tinham a doença, independentemente do grau de seus sintomas (Boletim FUNDECITRUS, 1998).

A clorose variegada dos *Citrus* afeta todos os cultivos comerciais de laranjas doces (Pêra, Hamlin, Natal, Valência, Valência Folha Murcha, Baianinha, Barão e outras).

Figura 1: Folhas apresentando os sintomas da CVC. À esquerda, face superior com pontuações circulares de cor amarela, à direita, face inferior com pontos de necrose. Fonte: Boletim FUNDECITRUS, 1998, 2ª edição.

Tangerinas, limões e alguns híbridos comerciais são considerados tolerantes (MACHADO, 1987).

Os principais sintomas da doença são o aparecimento de pontos de clorose espalhados pelas folhas e caule normalmente nos espaços inter vasos que se mostram como pontos marrons em ambos os lados do limbo (figura 1). Esses pontos podem coalescer até as margens das folhas podendo aparecer nos ramos ou em todo o dossel. Novos brotos podem ser afetados causando uma diminuição do tamanho da planta. As folhas mostram sinais de desordem nutricional, principalmente devido a

deficiência de zinco e boro e ausência de potássio. Todavia, o principal sintoma é a ocorrência de pequenos frutos (figura 2), até três vezes menores que o normal, de constituição dura, com sabor ácido, coloração amarela intensa com manchas marrons e amadurecimento precoce, que se tornam rejeitados para consumo, inclusive pela indústria de processamento de sucos (MENEGUCCI *et al.*, 1995). Em estágios avançados da doença, nota-se a presença de ramos secos, desfolhados e com pencas de frutos (figura 3). A produção é reduzida de 70 quilos em uma planta normal para apenas 17 quilos de frutos por árvore afetada.

Figura 2: Aspecto de um fruto normal (à esquerda) e de um fruto proveniente de uma laranjeira infectada com CVC (à direita). Observar a redução no tamanho do fruto. Fonte: Boletim FUNDECITRUS, 1998, 2ª edição. Foto: Henrique Santos

Figura 3: Presença de pencas de frutos em plantas severamente afetadas. Fonte: Boletim FUNDECITRUS, 1998, 2ª edição. Foto: Henrique Santos

No último século, muitos pomares têm sido atingidos pela bactéria *Xylella fastidiosa* e muitas perdas foram relatadas. Sendo o Estado de São Paulo, o segundo produtor mundial de suco concentrado— juntamente com o Estado da Flórida nos Estados Unidos são responsáveis por 90% da produção mundial— a CVC tornou-se, sem dúvida, a doença de maior impacto na economia citrícola da região. Nos primeiros estudos, os prejuízos foram endereçados a qualquer microrganismo que pudesse ser viável durante a fase de propagação vegetativa da planta (HEWITT, 1939).

Vírus (COCHRAN *et al.*, 1976) e bactérias da família das Rickettsiaceae (GOHEEN *et al.*, 1973) eram considerados os possíveis candidatos a agentes causadores da doença.

Além dos *Citrus* a bactéria já foi encontrada em diferentes espécies vegetais como videiras (doença de Pierce), pessegueiros, alfafa, amoreiras, carvalhos e café (PARADELA *et al.*, 1995).

Durante a década de 80, alguns estudos mostraram que a bactéria não fazia parte da família das Rickettsiaceae e inicialmente propuseram o nome de *Xylemella fastidiosum* como agente causal da doença de Pierce. No final do ano de 1980 foi relatado o isolamento de uma bactéria limitada ao xilema de videiras (*Vitis vinifera* L.) afetadas pela doença. Logo depois, o nome *Xylella fastidiosa* foi

usado num estudo comparativo e uma revisão sobre este novo gênero foi relatada (PURCELL & HOPKINS, 1996).

Desde o início, algumas hipóteses para a causa da CVC foram investigadas por cientistas brasileiros. Em 1987-88, alguns experimentos investigaram uma possível infecção por vírus mas nenhum resultado foi considerado conclusivo pois as plantas não estavam devidamente protegidas contra insetos. Em 1989, VITTI e colaboradores publicaram um dos primeiros experimentos

Figura 4: *X. fastidiosa* infectando folhas de laranjeira. Primeira observação. Fonte: *Citrus Variegated Chlorosis*, 1998. Foto: M. Garnier

considerando efeitos nutricionais como causa da CVC.

Estas desordens nutricionais poderiam criar condições favoráveis ao estabelecimento de um patógeno oportunista. Muitos sintomas mostravam também aspectos fisiológicos como por exemplo a toxicidade produzida por excesso de herbicidas (CASTRO, 1992).

Em dezembro de 1989, laboratórios do INRA (Institut National de Recherches Agronomiques) em Bordeaux, França, mostraram por microscopia eletrônica vasos xilemáticos entupidos com uma bactéria similar à *Xylella* em tecidos de plantas afetadas com CVC (figura 4).

Com a colaboração do Instituto Biológico de São Paulo, o INRA publicou seu sucesso na inoculação e isolamento da bactéria no hospedeiro, completando assim, o postulado de Koch. Ainda em 1993, informações da Universidade da Flórida diziam que a bactéria havia sido re-isolada de plantas inoculadas (BERETTA *et al.*, 1993). Foi então demonstrado que as desordens nutricionais observadas na sintomatologia da CVC eram uma consequência e não a causa da doença (CHANG *et al.*, 1993b).

I.4. A BACTÉRIA *Xylella fastidiosa* Wells

O termo “xylem-limited bactéria” (bactérias-limitadas ao xilema; XLM) define bactérias endoparasitas que vivem exclusivamente limitadas ao xilema de vegetais. Entretanto, são fastidiosas aquelas que crescem lentamente e exigem um meio complexo de crescimento “in vitro”. Três espécies são consideradas fastidiosas: *X. fastidiosa*, *Pseudomonas syzygii*, ambas transmitidas por insetos e

Clavibacter xyli (PURCELL & HOPKINS, 1996). Todas colonizam um nicho especial: o sistema de condução de água de uma variedade de hospedeiros vegetais.

Xylella fastidiosa Wells é uma bactéria gram-negativa, possui forma de bastão não flagelado, medindo de 0.3 a 0.5 µm de diâmetro e de 1 a 5 µm de comprimento (figura 5). Não forma esporos. Em temperaturas em torno de 20 a 25 °C, pH de 6.7 a 7.0 exibe crescimento “*in vitro*” somente após 5 a 7 dias de incubação.

Estudos iniciais (KAMPER *et al.*, 1985)

indicaram que o conteúdo de guanina + citosina das XLM varia de 50.1 a 54.0 mols%. Atualmente sabemos que é de 52.4 mols% na linhagem sequenciada, enquanto que, em *Rochalimaea quintana* (Rickettsiae) este valor é de 38.5 mols% (DAVIS *et al.*, 1980). Da mesma maneira, os genomas estão em torno de 1.4 ± 0.2 , ou 1.92 a 2.42 Mb,

com a cepa sequenciada apresentando 2,679,305 pares de bases mais dois plasmídeos: um de 51,158 pb (mega plasmídeo) e outro de 1,285 pb (mini plasmídeo), totalizando 2.731.573 pares de bases.

Estudos de bioquímica, sorologia, enzimas extra-celulares, ácidos graxos, DNA e RNA de 25 isolados de bactérias limitadas ao xilema, de videira, pessegueiro, amendoeira, olmo, ameixeira, plátano, carvalho e amoreira concluíram que o grupo era homogêneo e diferenças foram observadas na razão de crescimento “*in vitro*” (WELLS *et al.*, 1987). As hibridizações de DNA e as sequências dos 16S RNAs foram usadas para confirmar que as bactérias eram de fato uma única espécie associada ao subgrupo gamma das Eubacteria e à *Xanthomonas*. Ainda, foi proposta a criação de um novo gênero, *Xylella*, com somente uma espécie, *Xylella fastidiosa* (WELLS *et al.*, 1987). Atualmente, *X. fastidiosa* é considerada membro da subdivisão gamma, ordem Lysobacterales, família Lysobacteriaceae, grupo *Xanthomonas*, gênero *Xylella* com somente uma espécie (DONADIO & MOREIRA, 1998).

Embora toda bactéria gram negativa com crescimento limitado ao xilema de diversas plantas hospedeiras esteja incluída na espécie *X. fastidiosa*, existe variabilidade suficiente para a separação taxonômica em subespécies ou patovares (MACHADO, 1987). Diferentes cepas de *X. fastidiosa* requerem um meio de cultura complexo para crescimento e podem ser separadas em dois grupos distintos (HOPKINS, 1989) com base nas exigências nutricionais: o grupo da cepa causadora da doença de Pierce que cresce nos meios PD3-Pierce's disease agar (DAVIS *et al.*, 1978), PW-Perrywinkle Wilt Medium (CHANG *et al.*, 1993), BCYE e CS-20 e o grupo da cepa causadora da má formação em pêssegos que cresce em todos os meios citados, exceto PD3. Cepas de videira,

Figura 5: Microscopia eletrônica mostrando estrutura enrugada da parede de *X. fastidiosa*. Fonte: *Citrus Variegated Chlorosis*, 1998. Foto: C.M. Chagas

amendoeira, olmo e amoreira têm crescimento mais rápido em meio PW (em torno de 0.5 a 1.6 dias), enquanto que, cepas que ocorrem em pessegueiros e ameixeiras têm crescimento mais lento (em torno de 2 dias). A análise do DNA usando a técnica de RAPD (“Randomly Amplified Polymorphic DNA”) sugere a existência de cinco grupos de *X. fastidiosa*: o grupo dos *Citrus*, o grupo ameixa-olmo, o grupo videira-*Ambrosia* sp, o grupo das amendoeiras e o grupo das amoreiras (POOLER & HARTUNG, 1995).

1.5. TRANSMISSÃO E CONTROLE

A transmissão de *X. fastidiosa* depende necessariamente de insetos vetores para sua propagação natural e penetração nos tecidos vegetais susceptíveis.

Estudos comparativos utilizando laranjeiras protegidas e não protegidas contra insetos e térmitas demonstraram que as plantas não protegidas foram afetadas pela CVC. Em 1993 mostrou-se que havia um envolvimento de insetos como possíveis vetores na transmissão da doença (ROSSETI *et al.*, 1993). Como a bactéria permanece nos elementos traqueais, traqueídes, vasos (figura 6) e nos espaços intercelulares do xilema de seus hospedeiros, insetos que se alimentavam da seiva bruta foram identificados.

Figura 6: Observação de *X. fastidiosa* infectando um pecíolo de folhas de laranjeira. Fonte: Citrus Variegated Chlorosis, 1998. Foto: J.E.O. Lima

Muitas espécies de insetos sugadores são potenciais vetores na propagação da CVC, uma vez que se alimentam nos vasos xilemáticos. Alguns desses vetores vivem em plantas cítricas enquanto que outros, hospedam-se em arbustos utilizados para a proteção dos pomares. Estudos realizados pela ESALQ (Escola Superior de Agricultura Luiz de Queiroz) e pelo FUNDECITRUS identificaram cigarrinhas da família Ciccadellidae como os principais vetores na disseminação da doença (figura 7). Mais de 19 espécies da subfamília Ciccadellidae foram observadas sendo algumas consideradas os mais importantes vetores da propagação da CVC como *Dilobopterus costalimai*, *Oncometopia facialis* e *Acrogonia terminalis* (LOPES & BERETTA, 1996).

Estes insetos podem transmitir a bactéria durante todos os estágios de sua vida, entretanto, os adultos são mais importantes. As ninfas transmitem a bactéria, mas perdem essa capacidade após a

ecdise, contudo, a infectividade do adulto dura 60 dias e pode perdurar até a sua morte (PURCELL & HOPKINS, 1996).

Figura 7: Principais vetores de *X. fastidiosa*. Da esquerda para direita: *D. costalimai*, *O. fascialis*, *A. terminalis*. Fonte: <http://www.lbi.dcc.unicamp.br/services>

A coloração amarela das plantas afetadas atraem as cigarrinhas que se alimentam da seiva bruta. É possível que os mecanismos de transmissão das XLB sejam similares entre as cepas, entretanto, existem variações no tempo requerido de aquisição e inoculação do patógeno pelo vetor, o período de latência, eficiência na transmissão, persistência da bactéria no vetor e período de incubação na planta após a inoculação. Por exemplo, o principal vetor de *X. fastidiosa* em videiras, *Graphocephala atropunctata*, pode adquirir a bactéria em uma hora de alimentação em uma planta infectada e transmiti-la imediatamente após a sua aquisição (PURCELL & FINLAY, 1979).

Como as ninfas perdem a capacidade de transmissão da bactéria, foi observado por microscopia eletrônica o alojamento da bactéria na parte anterior do tubo digestivo do vetor. Em *G. atropunctata*, a bactéria foi encontrada aderida à cutícula que cobre a câmara sugadora (cibário) e na parte anterior do esôfago. Há evidências de que a bactéria colonize o canal alimentar formado pela junção do estilete maxilar e que se multiplique no esôfago mantendo a capacidade infectiva do vetor ao longo de sua vida (PURCELL & FINLAY, 1979).

O mecanismo de inoculação da bactéria depende do deslocamento das células no hospedeiro, que pode ocorrer por tensão negativa do xilema acarretando no movimento em direção ao vegetal. Alternativamente, a inoculação pode ocorrer devido ao processo de degustação da planta: líquidos que não são aceitos pelo inseto podem ser expelidos de volta à planta, transportando consigo células da bactéria descoladas do esôfago (LOPES, 1996).

Métodos para a identificação da doença foram desenvolvidos pelo FUNDECITRUS. Um método rápido é a verificação por microscopia ótica da presença da bactéria em extratos de pecíolos afetados. O método de ELISA (“Enzyme-Linked Immunosorbent Assay”) está sendo utilizado com

anticorpos produzidos pelo INRA. Outros métodos como “Dot-Immunobinding Assay” e “Western Blotting” também estão sendo usados (DONADIO & MOREIRA, 1998).

Propõe-se que a cada três meses por ano, as árvores de *Citrus* deverão ser submetidas a um diagnóstico de CVC e quando plantas infectadas forem diagnosticadas, a poda deve ser feita na forquilha do galho contaminado ou 30 cm abaixo da posição de observação dos sintomas. A prática da poda é realizada pelo citricultor durante sete meses do ano, de janeiro a julho, quando os sintomas são mais evidentes (Boletim FUNDECITRUS, 1998).

1.6. COLONIZAÇÃO

Os elementos traqueários das plantas hospedeiras proporcionam um ambiente físico e nutricional único para crescimento da bactéria *X. fastidiosa*. Os vasos xilemáticos propiciam um ambiente nutricional diluído; entretanto, estes contêm uma vasta gama de nutrientes essenciais, que variam de acordo com a espécie vegetal, sua saúde, sua idade, localização, e com os ciclos circadiano e sazonal (HOPKINS, 1989).

Dados sobre a seiva bruta de *Vitis rotundifolia*, espécie hospedeira de XLB, indicam que o xilema contém uma variedade de aminoácidos, ácidos orgânicos e íons inorgânicos, predominantemente em suas formas monoméricas, e concentrações muito baixas de reguladores de crescimento como citocininas e giberilinas. As concentrações totais de aminoácidos e ácidos orgânicos determinadas foram praticamente iguais (2 a 7 mmol/L) e os açúcares somaram a 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 xilema, dado este que é consistente com o requerimento deste composto em grandes quantidades para o crescimento da bactéria em meio de cultura.

Todos os micro e macro nutrientes essenciais analisados (NH_4^+ , NO_3^- , PO_4^{+3} , K^+ , Ca^{++} , Mg^{++} , Mn^{++} , Fe^{+3} , Cu^{++} , Zn^{++} , Na^+ , e Cl^-) foram detectados, sendo cálcio e potássio os macroelementos mais concentrados e cobre, o microelemento menos concentrado. Todos os aminoácidos, exceto triptofano foram encontrados, sendo a glutamina (2.3 a 5.0 mmol/L) cerca de 80% do total. Histidina, tirosina, cisteína, metionina e lisina ocorreram em níveis muito baixos. As concentrações de poliamina e de proteínas foram consideradas desprezíveis. Foram identificados sete outros ácidos orgânicos, entre os quais, os de maior concentração foram o ácido cítrico (0.3 a 2.0 mmol/L), tartárico (0.6 a 1.7 mmol/L) e málico (0.4 a 2.8 mmol/L). Na pequena fração referente aos açúcares foram detectados glicose,

frutose e sorbitol, ou manitol (o método de análise não distinguia entre ambos), sendo glicose e frutose em maior parte (FERREIRA, 2000).

Existem diversas considerações a respeito dos mecanismos de patogenicidade da bactéria. Os principais são a disfunção na condução do sistema de água ou produção de fitotoxinas. Evidências apoiam a hipótese de que o mecanismo de patogenicidade seja a oclusão dos vasos xilemáticos causando estresse hídrico à planta, caracterizado pela presença de agregados de bactérias presos às paredes vasculares.

Estudos de ultraestrutura demonstraram que a bactéria forma freqüentemente agregados nos vasos xilemáticos de pecíolos, nervuras foliares e tronco que aparentemente se mantêm unidos através de fibras produzidas pela bactéria. Estas fibras polissacarídicas e extracelulares foram identificadas como microfibrilas, fibras elétron-densas e fimbrias, e são mais abundantes nas extremidades do bastonete (HOPKINS, 1989). Também se parecem à massa entrelaçada de fibras de polissacarídeos que se estende da superfície de diversas bactérias, denominada, glicocálix. Esta matriz é composta por polissacarídeos extracelulares (EPSs) sintetizados por enzimas semelhantes às de *Xanthomonas campestris* pv *campestris*, que produzem a goma xantana. Vinte e seis genes que codificam para a biossíntese de fimbrias Tipo-4 foram identificados. Estas fimbrias são também encontradas nos pólos de uma gama de bactérias patogênicas onde atuam na adesão da superfície epitelial e na interação bactéria-bactéria para a formação de agregados celulares.

Estes agregados de fibras polissacarídicas que estão carregadas negativamente podem funcionar como uma resina de troca iônica, num ambiente diluído, servindo como um reservatório nutricional, ligando íons ao agregado e disponibilizando-os às células. Também podem funcionar contra defesas do hospedeiro. Evidências indiretas mostram que cepas de PD que possuem glicocálix formam agregados em meio de cultura. Estas cepas, são altamente virulentas, mas a partir de várias subculturas, perdem a capacidade de agregação e sua característica de virulência “in vivo” (HOPKINS, 1989).

Um aspecto importante na colonização de hospedeiros por *X. fastidiosa* está no movimento sistemático da bactéria dentro dos vasos xilemáticos. A taxa de dispersão do patógeno no xilema é limitada pelo comprimento do vaso e o tempo requerido para que sejam rompidas as membranas pontuadas que delimitam os mesmos e que impedem as células bacterianas de migrarem de um vaso ao outro. O movimento da bactéria se dá, portanto, através do rompimento dessas membranas, sugerindo que a bactéria possa degradá-las através da produção de enzimas digestivas concentradas pelo glicocálix. Tais membranas são constituídas em princípio de parede celular contendo celulose, hemicelulose, pectinas, ligninas e algumas proteínas. Das enzimas pectolíticas capazes dessa função,

foram encontradas no genoma seqüenciado um precursor de poligalacturonase e uma celulase. Outra possibilidade é a de que o hospedeiro possa produzir enzimas degradativas que seriam induzidas em resposta à presença de patógenos e despejá-las nos vasos xilemáticos. Alternativamente, a multiplicação da bactéria nos vasos poderia produzir pressão suficiente para romper as membranas pontuadas fisicamente (HOPKINS, 1989).

As relações entre a população bacteriana e os sintomas da doença como as desordens nutricionais devido ao baixo conteúdo de zinco e potássio sugerem que a produção de toxinas pela bactéria não está envolvida nos sintomas da doença. Sintomas de estresse hídrico em videiras com Doença de Pierce, como engruvinhamento marginal da folha, parecem estar co-relacionados ao bloqueio do sistema de transporte hídrico pela oclusão dos elementos traqueários devido à produção de tiloses, produzidas pelo hospedeiro, gomas ou grupamentos de bactérias (HOPKINS, 1989).

Com exceção a poucos hospedeiros como videiras e pessegueiros, *X. fastidiosa* pode ser considerada um patógeno oportunista. Cepas de *X. fastidiosa* podem residir em vasos do xilema de muitos hospedeiros sem causar a doença, mas acumulam-se e desencadeiam os sintomas se o hospedeiro sofrer algum tipo de estresse. Fatores estressantes que podem interagir sinergeticamente à bactéria incluem seca, outras doenças, poda, superprodução de frutos, maturação normal de frutos e senescência do vegetal (HOPKINS, 1989).

II. OBJETIVOS

Este projeto teve como objetivos:

- 1- Seqüenciar em maior número possível clones de bibliotecas de “shotgun” e restrição e apoiar tecnicamente os demais Laboratórios envolvidos no projeto;
- 2- Dar continuidade à elaboração de bibliotecas de cosmídeos bem como seu completo seqüenciamento durante o andamento do Projeto Genoma;
- 3- Realizar posterior análise, através da anotação dos genes de cinco cosmídeos, predizendo a respeito da estrutura e função biológica dos genes anotados e classificando-os em categorias, utilizando como ferramentas, bancos de dados disponíveis na Internet;
- 4- Verificar as ORFs anotadas na categoria de Biossíntese de Pequenas Moléculas, organizando-as nas vias metabólicas referente aos seus produtos e comparar essas vias com as presentes em outros microrganismos.

III. MATERIAL E MÉTODOS

III.1. SEQUÊNCIAMENTO DO GENOMA DE *Xylella fastidiosa*

A estratégia da coordenação do projeto foi elaborada para se ter o máximo de flexibilidade na conclusão do fechamento do genoma, incorporando extensiva redundância de dados provenientes dos Laboratórios de sequenciamento.

O DNA utilizado foi isolado do clone 9.a.5.c (CHANG *et al.*, 1993a) a partir do qual foram feitas: uma biblioteca de cosmídeos; algumas bibliotecas de plasmídeos, obtidas através da fragmentação aleatória do DNA genômico por enzimas de restrição ou nebulização (“shotgun”); e uma biblioteca de fago Lambda.

A biblioteca de cosmídeos foi construída pela Dra. Anete Pereira Souza no laboratório do Dr. Hoheisel em Heidelberg, Alemanha, cobrindo o genoma total. Cada cosmídeo possuía em torno de 40 kb. A clonagem do DNA foi feita em vetor Lawrist4 através da digestão parcial do mesmo com a enzima *MboI* e ligação ao sítio *BamHI* do vetor segundo o protocolo de Jens Hanke & Jörg D. Hoheisel, 1998 (Figura 8).

A validação da biblioteca realizou-se através de hibridização de sondas de cosmídeos ao DNA total da bactéria *X. fastidiosa*, através de “random primers”, utilizando a técnica de “Southern Blot”.

Um cosmídeo era considerado terminado (“finished”), se a montagem estivesse de acordo com normas estabelecidas pelo Laboratório de Bioinformática. No caso de haver muitas ou poucas discrepâncias, de acordo com os critérios de qualidade, o cosmídeo era considerado fechado (“one contig”) ou quase fechado (“close to finished”), respectivamente.

Dois tipos de bibliotecas de plasmídeos foram construídas pelo Laboratório Coordenador de DNA e distribuídas aos demais laboratórios de seqüenciamento; as primeiras, foram construídas através da digestão do DNA total com enzimas de restrição como *EcoRI/MunI* e *NsiI/PstI*, as demais, foram construídas com DNA total nebulizado ou sonicado (“shotgun”) e clonado nos vetores m13 mp19, pUC 18, pUC19 ou Janus (Tabela 1. Resultados).

A biblioteca de fago Lambda foi construída e seqüenciada pelo Laboratório do Prof. Dr. Paulo Arruda, em Campinas, com o intuito de preencher regiões do genoma que não haviam sido clonadas em cosmídeos. Os “gaps” foram resolvidos com o seqüenciamento de clones das bibliotecas de nebulização ou restrição cujas extremidades se voltavam “para fora” das extremidades de contíguos de cosmídeos. Estes clones foram seqüenciados inteiramente através de diferentes técnicas como “primer-walking”, nebulização ou sonicação e digestão por enzimas de restrição. Estas seqüências foram denominadas de GFSs (“Gap-filling Sequences”).

III.2. SEQÜENCIAMENTO DE COSMÍDEOS PELO LABORATÓRIO QR

1- Meios de cultura

1.1 2xTY

16 g/L triptona; 10 g/L extrato de levedura; 5 g/L NaCl; pH 7,4.

1.2 Terrific Broth

12 g/L triptona; 24 g/L extrato de levedura; 4,8 g/L glicerol.

A seguinte solução de sais é colocada na proporção de 1:10 no momento de uso:

2,89 g/100 mL KH_2PO_4 ; 15,68 g/100mL K_2HPO_4

1.3 SOB

20 g/L triptona; 5 g/L extrato de levedura; 0,58 g/L NaCl; 0,2 g/L KCl; pH 7,5.

1.4 Placas TYE

10 g/L triptona; 5 g/L extrato de levedura; 8 g/L NaCl; 15 g/L agar bacteriológico.

2- Antibióticos

2.1 Ampicilina sódica ou carbenicilina

Solução estoque 200 mg/mL (Royton), esterilizada por filtração e estocada a – 20°C.

Solução de uso: 100 µg/mL.

2.2 Kanamicina

Solução estoque 200 mg/mL, esterilizada por filtração e estocada a – 20°C.

Solução de uso: 50 µg/mL.

3- Enzimas

New England Biolabs

Hind III (20.000 U/mL), T4 DNA polimerase (3.000 U/mL), T4 DNA ligase (400.000 U/mL), T4 polinucleotídeo quinase (10.000 U/mL), Fragmento Klenow da DNA polimerase I (5.000 U/mL)

4- Cepas de *Escherichia coli* utilizadas nas bibliotecas

DH10B: *endA1 recA1 relA1 deoR thi-1 supE44 λgyrA96 merA F⁻ Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 Δ(lacZYA-argF)U169;*

JM109: *endA1 recA1 thi-1 gyrA96 hsdR17(r_K-m_K⁺) e14-(McrA-) Δ(lac-proAB)[F⁻ traD36proABlacI^qΔM15].*

5- Vetores utilizados nas bibliotecas

pUC18 *Sma*I/CIP

pUC19 *Sma*I/CIP

Amersham Pharmacia Biotech

Cat 27-4860-01

6- Obtenção de Cosmídeos em larga escala

O Laboratório Coordenador de DNA distribuiu placas de Petri com um repique de bactérias contendo o cosmídeo em meio sólido TYE (kanamicina 50 µg/mL). Desta cultura, eram repicadas cinco colônias isoladas em meio líquido 2xYT com o antibiótico para a verificação da ocorrência de

contaminação entre cosmídeos; o DNA do cosmídeo era extraído através de lise alcalina (SAMBROOK *et al.*, 1989) e digerido com a enzima de restrição *HindIII* (5 µg DNA, 0.5 µL enzima, 1 µL tampão Neb2 10x concentrado, 10 µL água MilliQ q.s.p, incubação a 37°C por 3 horas). Todos os cosmídeos foram digeridos com *HindIII* para posterior verificação do padrão de digestão com o cosmídeo já sequenciado (digestão eletrônica). Na ausência de contaminação, uma das culturas era escolhida para se proceder ao repique em larga escala (500mL de 2xYT) para posterior procedimento de isolamento do cosmídeo. Uma alíquota de cada cultura foi mantida estocada em glicerol (40% v/v) a – 70°C.

A obtenção dos cosmídeos purificados em larga escala se deu através da lise alcalina de uma cultura de bactérias contendo o cosmídeo e posterior ultracentrifugação da amostra em Cloreto de cério (1g/mL) e Brometo de etídio (0,5 mg/mL) (SAMBROOK *et al.*, 1989), quando o bandejamento referente à massa de cosmídeos pôde ser identificado na amostra de DNA centrifugado. Para a retirada do Brometo de etídio foram feitas extrações com Butanol saturado com 1M NaCl e para a posterior retirada do sal, procedeu-se a uma diálise em tampão TE (10 mM TRIS; 1 mM EDTA) segundo os protocolos modificados descritos no item anexos (página 86).

7- Biblioteca

Optamos pela nebulização dos cosmídeos obtidos em larga escala. Para tal, utilizamos um nebulizador de inalação onde a amostra era forçada a passar por um pequeno orifício. Posteriormente a quebra do DNA dos cosmídeos, este foi tratado com as enzimas T4 DNA polimerase, Klenow e Polinucleotídeo kinase para o reparo das pontas (“blunt end”) e para que fosse possível a ligação com o vetor (vide anexos, página 88).

Após este procedimento isolamos fragmentos de DNA com tamanho em torno de 1.500 a 3.000 pares de bases através de um gel de agarose (0,8%) e procedemos à ligação com o vetor. Para a construção das bibliotecas utilizamos os vetores pUC 18 ou pUC 19, abertos no sítio de *SmaI*, e tratados com fosfatase alcalina por 15 minutos (<http://circuit.neb.com/cgi-bin/onlineorder.pl/neb/products/nucleic/304-1.html>). Após a ligação dos insertos com os vetores, bactérias da linhagem de *E. coli* DH10B foram transformadas por eletroporação e plaqueadas em meio de cultura sólido TYE contendo 100 µg/mL de carbenicilina, X-gal (0,032 mg/mL) e IPTG (0,032 mg/mL). As colônias foram selecionadas através da expressão de *lacZ*; bactérias contendo o plasmídeo, ou seja, vetor ligado a inserto, produziram colônias brancas ao serem plaqueadas.

8- Obtenção de plasmídeos para seqüenciamento (“Minipreps”)

No início do projeto obtivemos material a partir de mini preparações de DNA plasmidial baseado no método de lise alcalina (BIRBOIN & DOLY, 1979) com modificações realizadas pelo Dr. Marcelo Bento Soares (Universidade de Iowa, EUA). Outro método, conhecido como “boiling prep” ou “micro wave protocol” foi introduzido para a obtenção simultânea de 96 clones em placas, adaptado do método de Holmes e Quigley (1981). Através desse método, 96 colônias brancas são repicadas em meio Terrific Broth com adição de carbenicilina e incubadas em agitador, 300 rpm, a 37 °C por 24 horas. O meio é descartado após centrifugação e a cultura é ressuspensa em tampão MW-TWEEN 20 e fervida por dois minutos em forno de microondas. Após nova centrifugação, é coletada a fase superior que contém os plasmídeos (consultar anexos à página 89).

No caso de seqüências de fechamento de cosmídeos (“gap-filling sequences”) os clones das placas contendo colônias transfectadas com fragmentos de DNA tiveram seus plasmídeos isolados para seqüenciamento através de lise alcalina descrita em Sambrook *et al.*, 1989.

9- Seqüenciamento

A reação de seqüenciamento do DNA baseou-se numa adaptação do método Dideoxi de Sanger (SANGER *et al.*, 1977) com a utilização de um kit de seqüenciamento fornecido pela Perkin Elmer Biosystems (item anexos, p.90). O método utiliza uma polimerase termo estável (*Ampli Taq* polimerase) com “primers” de iniciação universal (M13-21) ou reverso (M13 Reverse) específicos para os vetores utilizados. Os ddNTPs não radioativos são ligados covalentemente à substâncias fluorescentes específicas para cada ddNTP, que após excitação pelo feixe de laser do seqüenciador (ABI PRISM 377 – PE), emitem fluorescência em comprimentos de onda diferentes identificando cada uma das quatro bases (A, T, C, G).

Após a precipitação das amostras (vide p.90), estas são aplicadas no seqüenciador onde separam-se por eletroforese em gel de poliacrilamida e uréia utilizando placas de gel de 48 cm em módulo de programação “Seq Run 48E-1200” por sete horas. A reação de seqüenciamento cíclico é linear e foi adaptada do protocolo original da Perkin Elmer para otimização do processo. As seqüências são armazenadas em um microcomputador Power Macintosh através de um programa de coleta de dados (ABI Prism Data Collection) onde os dados brutos são convertidos em cromatogramas através do programa Sequencing Analysis versão 3.3. Os dados obtidos foram enviados ao Laboratório de Bioinformática e para um computador, plataforma UNIX, no Laboratório QR.

O cosmídeo 07H04 apresentou uma região onde se formava uma alça (“hairpin”) bloqueando a passagem da *Taq*DNA polimerase proveniente do kit de seqüenciamento da Perkin Elmer. O problema foi resolvido com seqüenciamento manual, utilizando o kit da Perkin Elmer “AmpliCycle Sequencing Kit” com isótopos radioativos a uma temperatura de amplificação de 68 °C .

10- Montagem (“finishing”) dos cosmídeos

A montagem de um cosmídeo consiste na comparação e junção entre seqüências de vários insertos que, unidas umas as outras, formam uma seqüência consenso para gerar apenas um contíguo (“contig”) de seqüências.

Os cosmídeos foram montados, utilizando as leituras das bibliotecas de “shotgun”, através do programa phred+Phrap+crossmatch e consed (<http://bozeman.mbt.washington.edu>). Para cada cromatograma considerado de boa qualidade e enviado à Campinas, o programa phred atribuía valores de qualidade (“quality values ou QV”) para cada base seqüenciada. Este valor de qualidade das bases seqüenciadas, também denominado de “phred values” está relacionado com a probabilidade de uma base (A, T, C ou G) estar errada. É calculado pela seguinte fórmula:

$$QV = -10 \log (P_E) \text{ onde,}$$

QV= valor de qualidade,

P_E = probabilidade da base seqüenciada estar errada.

Somente valores (“phred-value”) maiores que 20 foram considerados satisfatórios, ou seja, o erro aceitável no seqüenciamento de cada base deveria ser menor que 1 base em cada 100 seqüenciadas. O programa Phrap realizava, então, a comparação e montagem de contíguos das seqüências analisadas pelo programa phred. O programa consed é responsável pela visualização gráfica dos arquivos de montagem dos contíguos e determinação de regiões de baixa qualidade, (atribuindo uma escala de cores para cada base), seqüenciamento em uma fita única ou discrepâncias de alta qualidade.

Após o fechamento virtual dos cosmídeos, as respectivas seqüências foram enviadas pela rede (Internet) ao Laboratório de Bioinformática na Unicamp. Um relatório final mandado pela Unicamp diagnosticava problemas de qualidade na montagem final (“assembly”) do cosmídeo para que fossem feitos os refinamentos necessários na seqüência gerada. Os critérios estabelecidos para que um cosmídeo fosse considerado terminado (“finished”) foram:

- 1) todos os cromatogramas utilizados na montagem do cosmídeo deveriam estar depositados no banco de dados em Campinas;

- 2) a montagem do cosmídeo deveria ser realizada utilizando somente as seqüências depositadas no Laboratório de Bioinformática;
- 3) a montagem dos cosmídeos poderia ser realizada com ou sem “screening” contra Lawrist4. Quando realizada sem “screening” contra Lawrist4, as seqüências contendo as regiões do vetor deveriam estar presentes nas extremidades da montagem.
- 4) a taxa de erro aceitável na montagem deveria ser igual ou menor que 1 em cada 10.000 bases (0.01%);
- 5) cada base do consenso deveria ser confirmada por ambas as fitas de DNA, pelo menos uma vez;
- 6) a qualidade de cada base da seqüência consenso não deveria ser menor que “phred value” 20 nem poderiam ocorrer bases, em leituras, com qualidade igual ou superior a “phred value” 40 que discordassem da seqüência consenso, de acordo com o programa phred.

11- Construção de “primers”

Com o intuito de se atingir a taxa de erro de 1 a cada 10.000 pares de bases seqüenciados, “primers” foram desenhados para cobrir regiões dos cosmídeos onde as seqüências não apresentavam a qualidade desejada. Para a construção desses “primers” utilizamos o programa “Gene Runner” e a síntese dos mesmos foi realizada em sua maioria pela empresa Bio Synthesis. Os “primers” possuíam em torno de 18 pares de bases. As reações de seqüenciamento com os “primers” desenhados foram feitas utilizando como “template” (molde) o próprio DNA do cosmídeo não nebulizado ou a leitura (“read”) correspondente.

12- Procura e anotação de ORFs

Após serem considerados “finished” as seqüências dos cosmídeos tornaram-se disponíveis para a anotação dos genes. A determinação das ORFs (fases abertas de leitura ou “open reading frames”) realizou-se através do programa GLIMMER (SALSBERG *et al.*, 1998), que faz uma busca das possíveis regiões de DNA que codificariam uma dada proteína. Esta busca, realizada pelo Laboratório de bioinformática, baseia-se num treinamento prévio do programa por códons de iniciação e término de síntese de proteínas com alta probabilidade de flanquearem uma região codificante. O treinamento do programa foi realizado com seqüências da própria *Xylella fastidiosa* para a obtenção do padrão estatístico de utilização de códons em regiões codificantes. Os códons com maior ocorrência foram ATG, GTG, TTG e CTG para iniciação e TAA, TAG e TGA para término. A utilização do

GLIMMER, exige que a sequência do cosmídeo esteja em formato FASTA. Este programa pode ser instalado em máquinas UNIX/ LINUX.

A análise das ORFs pelo programa GLIMMER tornou-se disponível ao programa Sequin (BENSON *et al.*, 1999), que contém arquivos com as posições relativas de cada ORF ao cosmídeo e sua tradução. O Sequin gera uma lista de ORFs, representadas por setas que indicam a fita codificante de DNA (C ou U). O laboratório de bioinformática adotou como U a fita não complementar (“uncomplemented strand”) onde a sequência é igual ao mRNA e está representada na sequência do formato FASTA. Já, C significa fita complementar (“complemented strand”) e a sequência igual a do mRNA está representada na fita complementar a da sequência do FASTA.

O programa faz o reconhecimento do codon de iniciação, por exemplo, o ATG é procurado na sequência FASTA da seguinte maneira:

5'-ATG-3' e 3'-TAC-5' nesse caso a transcrição ocorre no sentido da sequência FASTA (U ou +)

5'-CAT-3' e 3'-GTA-5' nesse caso a transcrição ocorre no sentido inverso (C ou -).

13- Busca de tRNAs

Foi utilizado o programa tRNA scan-SE, versão 1.1 (LOWE & EDDY,1997), disponível na página <http://www.genetics.wustl.edu/eddy/>. Para executar o programa é necessário apresentar a sequência FASTA do cosmídeo, utilizando como parâmetro de busca organismos procariontes. No caso da similaridade ser encontrada, o programa verifica se é formada a estrutura secundária típica dos tRNAs.

14- Anotação dos cosmídeos

Utilizamos o programa BLAST (“Basic Local Alignment Search Tool”) do NCBI (“National Center for Biotechnology Information”) através do programa Sequin. O Laboratório de Bioinformática providenciou que as ORFs geradas pelo GLIMMER fossem uma a uma comparadas às sequências depositadas nos bancos de dados, utilizando o programa BLASTp versão 2.0 (ALTSCHUL & GISH, 1996), para a verificação de similaridades. Entretanto, cada laboratório ficou responsável por conferir este processo e anotar os comentários específicos às proteínas encontradas, classificando-as em categorias de acordo com uma tabela de classificação de *Escherichia coli* (RILEY, 1993).

Para a conferência de similaridades duvidosas ou fracas, programas como o PSI-BLAST (“Position Specific Interated BLAST”; ALTSCHUL *et al.*, 1997) e P-FAM (“Protein families database of alignments and HMMs”; <http://www.sanger.ac.uk/Software/Pfam>) foram adotados. Utilizamos

como parâmetro de busca de similaridade, a matriz Blosom 62, com filtro e não redundância (“non redundant protein database, nr”). Após a sequência de aminoácidos FASTA da ORF ser submetida ao programa BLASTp, uma janela lista todas as possíveis proteínas com as quais a ORF tem similaridade. Nessa janela, verificamos o valor de E (“expected value”) que indica a chance do alinhamento ter ocorrido ao acaso fortuito (ALTSCHUL & KOONIN, 1998). Um $E=0$ indica o valor máximo de confiança na similaridade entre a ORF e a proteína indicada. Outras informações como o tamanho da proteína e função foram indicadas.

Uma ORF poderia ser descrita com base na sua similaridade ou não com proteínas dos bancos de dados da seguinte maneira:

1. “Putative protein”, quando as seqüências das ORFs eram similares a proteínas ou genes de função conhecida, depositados nos bancos de dados, com “E-value” menor a 10^{-5} ;
2. “Conserved Hypothetical Protein” (kDa), quando as seqüências das ORFs eram similares à proteínas com função desconhecida (também determinadas por programas e consideradas hipotéticas), com “E-value” menor a 10^{-5} ;
3. “Hypothetical Protein” (kDa), quando as seqüências da ORF não eram similares à proteínas ou genes presentes nos Bancos de Dados, com “E-value” maior ou igual a 10^{-3} .
4. Quando $10^{-5} < \text{“E-value”} < 10^{-3}$, então a seqüência da ORF era analisada pelos programas PSI-BLAST, para a verificação de similaridades fracas, P-FAM e HMM-PROSITE; estes dois últimos com referência aos domínios protéicos.

Assim, as ORFs poderiam ser classificadas da seguinte maneira:

- a. se a busca no PSI-BLAST convergisse para uma determinada proteína ou gene até a quinta interação, então a ORF era anotada como “Putative Protein” mas com “E-value” da submissão ao BLASTp e menção a busca no PSI-BLAST;
- b. se após cinco interações do programa PSI-BLAST não houvesse a convergência a nenhuma proteína ou gene mas a busca no P-FAM ou HMM-PROSITE identificasse um domínio de uma família de proteínas, então a ORF era identificada como “Putative Protein” pertencente à família identificada;
- c. se não houvesse convergência nem similaridade a alguma família protéica, a ORF era considerada como “Hypothetical Protein” (kDa).

Para todas as ORFs anotadas havia um campo de descrições a respeito da função do gene ou proteína, reação catalisada, classificação em uma categoria funcional, acesso aos bancos de dados do Gene Bank ou Swiss Prot e “E-value”. As categorias podem ser encontradas no “site” <http://onsona.lbi.ic.unicamp.br/xf/> [outubro, 2000]. Como parte do projeto, o Laboratório QR ficou responsável pela verificação de cada ORF classificada na categoria de biossíntese de pequenas moléculas (categoria II). As principais categorias são identificadas como: I-metabolismo intermediário, II-biossíntese de pequenas moléculas, III-metabolismo de macromoléculas, IV-estrutura celular, V-processos celulares, VI-elementos genéticos móveis, VII-patogenicidade, virulência e adaptação, VIII-proteínas hipotéticas, IX-ORFs com categoria indefinida.

15- Revisão da categoria II

Pesquisamos as possíveis funções dos genes relacionados a essa categoria em uma série de páginas da Internet que possibilitaram a busca de referências bibliográficas acerca desses genes. As páginas mais utilizadas foram [atualização em setembro, 2000]:

<http://ncbi.nlm.nih.gov/PubMed/>

<http://www.kegg.com/>

<http://ecocyc.pangeasystems.com>

Esta última foi de importância fundamental na compreensão das vias de biossíntese de pequenas moléculas em *E. coli* e sua comparação às proteínas encontradas em *Xylella*. Assim, com base nas vias de *E. coli*, pudemos decidir quais proteínas necessárias para a biossíntese de determinada molécula estavam presentes em *Xylella* e qual sua importância na via biossintética (se a proteína é essencial ou se existe uma rota alternativa na produção de determinada molécula).

Como muitas regiões do genoma não haviam sido anotadas por estarem cobertas somente com leituras de “shotgun” e fago Lambda estas ainda não estavam disponíveis nos arquivos do Sequin. Por consequência, muitos genes pertencentes às vias não foram identificados e perderam-se nestas seqüências. Para suprir essa falha procedeu-se a dois caminhos: um realizado pelo Laboratório de Bioinformática onde a medida que as regiões eram cobertas o programa GLIMMER identificava as ORFs dessas regiões quando era feito uma busca por similaridade no BLASTp; e outro caminho inverso, no qual a procura de cada gene das vias que não havia sido anotado era feito através da comparação das seqüências das proteínas ausentes em *Xylella* mas presentes em outros organismos contra o genoma da *Xylella* utilizando para isto o programa tBLASTn.

Quando os genes não eram encontrados em *Xylella fastidiosa*, procedeu-se à busca dos mesmos em diferentes microrganismos obedecendo ao critério de busca em genomas seqüenciados de microrganismos Gram-positivos ou negativos com tamanhos de genoma maiores ou menores que o genoma de *Xylella*. Assim, pudemos obter os genes conservados numa determinada via biossintética para diferentes microrganismos.

Utilizamos também para essa análise o banco de dados do COG (“Cluster of Orthologues Groups”) do NCBI (TATSOV *et al.*, 2000), o qual consiste numa tentativa de classificar filogeneticamente as proteínas codificadas em 21 genomas de bactérias, arqueobactérias e eucariotos.

Os COGs são identificados com base na comparação de seqüências protéicas entre genomas completados utilizando o programa BLAST após mascarar regiões “coiled-coil” e de baixa complexidade. A construção do banco de dados do COG é baseada na noção de que qualquer grupo com mais de 3 proteínas, provenientes de genomas distantes, que sejam mais similares umas às outras do que à outras proteínas do mesmo genoma, são preferencialmente classificadas como pertencentes ao mesmo grupamento ortólogo. Esta predição é também verdadeira nos casos em que as similaridades entre proteínas é relativamente baixa. Sendo assim, as proteínas ortólogas têm tipicamente o mesmo domínio conformacional e provavelmente as mesmas funções.

IV. RESULTADOS OBTIDOS

IV.1 BIBLIOTECAS DE NEBULIZAÇÃO E RESTRIÇÃO

As bibliotecas de nebulização e restrição foram geradas pelo Laboratório Coordenador de DNA e distribuídas aos laboratórios de sequenciamento. O Laboratório QR sequenciou apenas bibliotecas do isolado zero de *X. fastidiosa* provenientes do clone do INRA. Foram sequenciados aproximadamente 1.900 leituras (“reads”) com 1.435.823 pares de bases. Deste total, não foi estimada a quantidade de pares de bases com “phred value” menor que 20 (Tabela 1).

Biblioteca	Descrição da biblioteca	Total de leituras	Total de pares de bases	Total de vetor
X0-0001	DNA total, digerido com <i>NsiI</i> em pBS(+) digerido com <i>PstI</i>	96	77.751	13.524
X0-0002	DNA total, digerido com <i>NsiI</i> em pBS(+) digerido com <i>PstI</i>	163	127.431	22.262
X0-0003	DNA total, digerido com <i>NsiI</i> em pBS(+) digerido com <i>PstI</i>	143	117.500	22.469
X0-0301	DNA total, digerido com <i>MunI</i> em pBS(+) digerido com <i>EcoRI</i>	142	115.148	46.630
X0-0302	DNA total, digerido com <i>MunI</i> em pBS(+) digerido com <i>EcoRI</i>	109	80.904	26.429
X0-0303	DNA total, digerido com <i>MunI</i> em pBS(+) digerido com <i>EcoRI</i>	61	47.246	18.980
X0-0304	DNA total, digerido com <i>MunI</i> em pBS(+) digerido com <i>EcoRI</i>	46	35.723	18.640
X0-1003	Não descrita	115	77.250	6.331
X0-1004	Não descrita	45	34.889	2.841
X0-1005	Não descrita	113	80.229	5.545
X0-1006	Não descrita	99	62.185	4.660
X0-1021	Não descrita	51	42.859	2.918
X0-1022	Não descrita	76	60.226	5.265
X0-1023	Não descrita	95	78.706	7.201
X0-1038	Não descrita	105	80.621	6.728
X0-1040	Não descrita	109	107.887	7.367
X0-1041	Não descrita	51	28.387	1.712
X0-1042	Não descrita	106	66.950	6.494
X0-1101	DNA total nebulizado, vetor m13 mp19, sítio <i>SmaI</i> . Fragmentos de 1.500 a 3.000 pares de bases	79	49.085	1.073
X0-1103	DNA total nebulizado, vetor m13 mp19, sítio <i>SmaI</i> . Fragmentos de 1.500 a 3.000 pares de bases	96	64.846	972
Total		1.900	1.435.823	225.414

Tabela 1: Estimativa de seqüências geradas a partir de placas de bibliotecas de nebulização e restrição.

IV.2. SEQUENCIAMENTO DE COSMÍDEOS

Foram totalmente sequenciados 13 cosmídeos utilizando-se sempre a estratégia de subclonagem por nebulização. A tabela 2 indica o total de leituras obtidas para cada cosmídeo e o total de clones obtidos pelo laboratório QR para um cosmídeo ser considerado fechado. Observamos que, como os cosmídeos sobrepunham-se, a quantidade de leituras que cobre um cosmídeo inteiro é sempre muito maior que o total necessário para seu fechamento. A taxa de erro indicada refere-se ao total de leituras com sobreposição.

Os primeiros cosmídeos a serem sequenciados foram o 07H04 e 07H03. Todos os clones para o sequenciamento destes cosmídeos foram obtidos através de minipreps por lise alcalina. Os demais clones dos outros cosmídeos foram obtidos através do protocolo de “boiling prep”.

O mapa das regiões de sobreposição entre os cosmídeos encontra-se disponível no item anexos, página 94.

Cosmídeo	Total de leituras	Total de leituras QR	Tamanho total do cosmídeo (pb)	Taxa de erro
01H09	1.415	693	37.445	0.00 / 10000 bp
02D09	1.078	859	42.240	0.00 / 10000 bp
02H01	2.047	573	40.885	0.00 / 10000 bp
04A04	1.915	847	46.887	0.00 / 10000 bp
05E12	792	599	37.060	0.00 / 10000 bp
06D10	1.207	812	41.346	0.03 / 10000 bp
07H03	640	540	48.325	0.00 / 10000 bp
07H04	1.058	689	39.395	0.00 / 10000 bp
08D10	978	705	44.268	0.00 / 10000 bp
08G12	1.117	610	36.829	0.00 / 10000 bp
09A10	1.164	907	44.492	0.00 / 10000 bp
09A12	833	733	36.483	0.00 / 10000 bp
09H12	1.377	849	37.916	0.00 / 10000 bp

Tabela 2: Sequenciamento dos cosmídeos. O total de leituras indica a quantidade de leituras que cobrem o cosmídeo, uma vez que ocorre sobreposição entre os mesmos. O total de leituras obtidas pelo laboratório QR indica a quantidade necessária de leituras para o fechamento dos cosmídeos.

IV.3. ANOTAÇÃO

Através da anotação dos cosmídeos, observamos que a maioria das ORFs encontradas pelo GLIMMER foram classificadas na categoria de proteínas hipotéticas (44,30%), sendo estas com baixa similaridade com qualquer proteína conhecida nos bancos de dados. As proteínas hipotéticas porém conservadas, totalizaram 10,4% das proteínas anotadas. Entretanto, observamos que a maioria das classes de categorias foi coberta, encontramos na categoria I de metabolismo intermediário proteínas referentes ao transporte de elétrons, proteínas participantes do ciclo do ácido tricarboxílico, degradação de pequenas moléculas e funções regulatórias globais, num total de aproximadamente 7,7%. Observamos ainda, proteínas envolvidas no processo de patogenicidade, adaptação e produção de toxinas totalizando 6,5% das proteínas encontradas. E mais, 8,45% das ORFs encontradas pertencem à classe de transportadores celulares e 13,9% pertencem à categoria de metabolismo de macromoléculas e biossíntese de pequenas moléculas.

Em alguns cosmídeos constatamos várias pequenas ORFs seguidas, em regiões intergênicas, com pouca ou nenhuma homologia às proteínas dos bancos de dados. Estas ORFs foram descartadas e colocadas em observação pelo Laboratório de Bioinformática. Dois problemas poderiam estar ocorrendo nestes casos: ou essas pequenas ORFs são reais, mas ainda não têm similaridade a nenhuma proteína dos bancos de dados, ou são artefatos aleatórios (BAPTISTA, 2000). Atualmente, algumas dessas ORFs podem ter se juntado às ORFs vizinhas formando uma ORF maior. Além disso, encontramos algumas ORFs que se sobrepunham às outras de maior tamanho com similaridade nos bancos de dados. Essas ORFs também foram expurgadas.

3.1 Anotação do cosmídeo 02A11

A anotação do cosmídeo 02A11 se deu apenas parcialmente (da posição 18.718 à 38.616) devido à sobreposição ao cosmídeo 09A12, anotado anteriormente por outro laboratório.

Neste cosmídeo não foi possível identificar a presença de tRNAs através do programa tRNA scan. As ORFs anotadas encontram-se na figura 1 e sua classificação na tabela 3.

Pertencem à categoria II, de biossíntese de pequenas moléculas, a ORF 26, envolvida na biossíntese de alanina e a ORF 22, com função na biossíntese de ribonucleotídeos de pirimidina.

Figura 1: ORFs anotadas em nosso laboratório do cosmídeo 02A11. As ORFs hipotéticas anotadas na Tabela não foram representadas. As setas em preto indicam o nome da ORF e sua posição relativa ao cosmídeo (por exemplo, XF⁽¹⁾-02A11⁽²⁾-gl⁽³⁾38⁽⁴⁾u⁽⁵⁾ quer dizer: (1) *X.fastidiosa*, (2) cosmídeo 02A11, (3) ORF encontrada pelo programa GLIMMER e (4) seu número correspondente à posição no cosmídeo, (5) c= complemented e u= uncomplemented). As setas rosas indicam a proteína ou gene anotado.

Tabela 3: Anotação do cosmídeo 02A11. A sequência examinada possui 19.898 pb. As ORFs são indicadas pelo nome do cosmídeo seguido da posição encontrada pelo GLIMMER. As colunas “de” e “até” indicam a posição relativa da ORF ao cosmídeo. A coluna denominada de “fita” indica a direção relativa da ORF à sequência FASTA do cosmídeo; (-) = transcrição ocorrendo no sentido da sequência FASTA, (+) = transcrição ocorrendo no sentido inverso. A coluna “acesso” indica o código do banco de dados em que se encontra a proteína mais similar, onde Sp= Swiss Prot e Gi= GenBank. Algumas ORFs encontradas pelo GLIMMER já não são mais descritas na nova versão de ORFs do genoma de *Xylella fastidiosa* sendo substituídas manualmente por terem tamanho muito pequeno (menor que 100 aminoácidos). Estas ORFs (indicadas em vermelho) podem ter se juntado às ORFs vizinhas ou quando situadas seguidamente, podem ter formado uma nova ORF com maior tamanho.

ORF	de	até	Fita	Proteína	E-value	Categoria	Acesso	Organismo
20	20427	18718	-	Proteína ribossomal 30S	0.0	III.B.2 III.C.1	Sp P02349	<i>E.coli</i>
21	20356	20604	+	Hipotética 9.54 kDa		VIII.B		
22	21294	20623	-	Citidilato quinase	3e-48	II.B.2	Sp P43829	<i>Haemophilus influenzae</i>
23	21840	22169	+	Hipotética 11.98 kDa		VIII.B		
24	22169	23260	+	Conservada hipotética 40.0 kDa	8e-64	VIII.A	Gi 2313129	<i>Helicobacter pylori</i>
25	23296	23135	-	Hipotética 6.11 kDa		VIII.B		
26	23349	24236	+	Beta alanina sintetase	5e-87	II.A.2	Gi 2313883	<i>H.pylori</i>

Continuação da tabela 3								
ORF	de	até	Fita	Proteína	E-value	Categoria	Acesso	Organismo
27	24323	25540	+	“Pherormone shutdown protein”	6e-39	VIII.A	Pir C53309	<i>Enterobacter faecalis</i>
28	28873	25703	-	Hipotética 118.48 kDa		VIII.B		
29	29706	28870	-	Proteína de membrana interna, transportador de açúcar ABC	1e-41	V.A.3 IV.A.1	Gi 6137042	<i>Strptomyces coelicolor</i>
30	30584	29703	-	Permease, transportador de açúcar ABC	8e-45	V.A.3 IV.A.1	PID d1029300	<i>Thermus sp</i>
31	31873	30581	-	Transportador de açúcar ABC	2e-21	V.A.3	Gi 777760	<i>Synechococcus sp</i>
32	33511	31877	-	Hipotética 62.88 kDa		VIII.B		
33	33391	33639	+	Hipotética 9.33 kDa		VIII.B		
34	34397	34107	-	Conservada hipotética 10.63 kDa	8e-8	VIII.A	Sp P29943	<i>Pseudomonas denitrificans</i>
35	34842	34967	+	Hipotética 4.64 kDa		VIII.B		
36	34903	35937	+	Conservada hipotética 37.91 kDa	2e-34	VIII.A	Sp P35154	<i>Bacillus subtilis</i>
37	35924	36907	+	Conservada hipotética 36.04 kDa	4e-21	VIII.A	Sp P35155	<i>B.subtilis</i>
38	36846	38330	+	Conservada hipotética	3e-62	VIII.A	Sp P37765	<i>E.coli</i>
39	38431	38616	+	Hipotética 6.69 kDa		VIII.B		

3.2 Anotação do cosmídeo 05B03

A anotação do cosmídeo 05B03 também se fez apenas parcialmente (da posição 503 à 25.588) pois este cosmídeo fazia sobreposição aos cosmídeos 07A11 e 09A12, anotados totalmente. A sequência anotada possui 25.085 pares de bases onde foram identificadas originalmente 30 ORFs pelo programa GLIMMER (Figura 2). Observamos ORFs anotadas na categoria de patogenicidade, virulência e adaptação (categoria VII), sendo estas ORFs pertencentes a um grupo de toxinas (ORFs 01, 15, 16 e 29) encontradas em *E. coli* e em um eucarioto. Observamos também, uma grande quantidade de ORFs hipotéticas (ORFs 17 à 28) com função totalmente desconhecida e que possuem baixa similaridade com outras proteínas existentes atualmente nos bancos de dados.

Figura 2: ORFs anotadas em nosso laboratório do cosmídeo 05B03. As ORFs hipotéticas anotadas na Tabela não foram representadas. As setas em preto indicam o nome da ORF e sua posição relativa ao cosmídeo; c= complemented e u= uncomplemented. As setas rosas indicam a proteína ou gene anotado.

Tabela 4: Anotação do cosmídeo 05B03. As ORFs são indicadas pelo nome do cosmídeo seguido da posição encontrada pelo GLIMMER. As colunas “de” e “até” indicam a posição relativa da ORF ao cosmídeo. A coluna denominada de “fita” indica a direção relativa da ORF à sequência FASTA do cosmídeo; (-) = transcrição ocorrendo no sentido da sequência FASTA, (+) = transcrição ocorrendo no sentido inverso. A coluna “acesso” indica em qual banco de dados se encontra a proteína mais similar, onde Sp= Swiss Prot e Gi= Gene Bank. Proteínas conservadas hipotéticas significam que são similares à proteínas hipotéticas descritas em outros organismos e proteínas descritas apenas como hipotéticas são possíveis proteínas existentes apenas em *Xylella* pois possuem muito baixa similaridade com qualquer proteína do banco de dados.

ORF	de	até	Fita	Proteína	E-value	Categoria	Acesso	Organismo
01	503	3652	+	Acriflavina resistente	0.0	VII.C V.A.7	Sp P76398	<i>E.coli</i>
02	4497	3805	-	Hipotética 25.76 kDa		VIII.B		
03	4553	4687	+	Hipotética 5.2 kDa		VIII.B		
04	5792	4743	-	Zinco-tipo álcool desidrogenase	1e-148	I.C.2	Sp P75691	<i>E.coli</i>

Continuação da tabela 4								
ORF	de	até	Fita	Proteína	E-value	Categoria	Acesso	Organismo
05	6400	6083	-	Conservada hipotética 11.62 kDa	1e-13	VIII.A	Sp P34251	<i>Saccharomyces cerevisiae</i>
06	6270	6587	+	Hipotética 11.99 kDa		VIII.B		
07	6556	6401	-	Hipotética 5.55 kDa		VIII.B		
08	7225	6608	-	Lisozima autolítica	2e-18	IV.A.2 III.B.3	Sp P34020	<i>Clostridium acetobutyricum</i>
09	7367	7642	+	Hipotética 10.42 kDa		VIII.B		
10	7650	7811	+	Hipotética 6.03 kDa		VIII.B		
11	8764	7838	-	Glutarredoxina	2e-26	II.D.10 I.C.3	Gi 3650389	<i>Schizosaccharomyces pombe</i>
12	8746	9606	+	Acetil-xileno esterase	8e-21	I.A.2	Gi 6469276	<i>S.coelicolor</i>
13	11156	9879	-	Aminotransferase	8e-85	I.A.2	Sp 033267	<i>Mycobacterium tuberculosis</i>
14	11378	11232	-	Hipotética 5.42 kDa		VIII.B		
15	11547	13730	+	Secretora de toxina, transportador ABC	0.0	VII.C V.A.6	Sp P10089	<i>E.coli</i>
16	13727	15148	+	Secretora de hemolisina, proteína D	5e-91	VII.C V.A.6	Sp P06739	<i>E.coli</i>
17	15164	15262	+	Hipotética 3.53 kDa		VIII.B		
18	15198	15443	+	Hipotética 8.67 kDa		VIII.B		
19	15558	15430	-	Hipotética 4.67 kDa		VIII.B		
20	15658	15762	+	Hipotética 3.72 kDa		VIII.B		
21	15759	16388	+	Hipotética 23.90 kDa		VIII.B		
22	16385	16657	+	Hipotética 9.72 kDa		VIII.B		
23	16660	17277	+	Hipotética 23.25 kDa		VIII.B		
24	17274	17546	+	Hipotética 9.97 kDa		VIII.B		
25	17958	18116	+	Hipotética 6.13 kDa		VIII.B		
26	17549	18172	+	Hipotética 23.49 kDa		VIII.B		
27	18169	18441	+	Hipotética 9.86 kDa		VIII.B		
28	18444	19061	+	Hipotética 23.34 kDa		VIII.B		
29	19058	25252	+	Bacteriocina	3e-61	VII.C IV.A	Gi 2961326	<i>Rhizobium leguminosarum</i>
30	25271	25588	+	Hipotética 11.49 kDa		VIII.B		

3.3 Anotação do cosmídeo 02D09

Este cosmídeo foi totalmente anotado (41.845 pb) e originalmente possuía 53 ORFs. Note que embora existam nomenclaturas de ORFs seguidas até o número 53, as ORFs 21 e 29 foram desconsideradas pois estavam contidas dentro das ORFs 20 e 28. As demais ORFs (em vermelho) já não existem mais na versão nova de ORFs de *Xylella fastidiosa*. Notamos que no início deste cosmídeo ocorre grande concentração deste tipo de ORF, assim como entre as posições 11.794 à 13.881. As

ORFs 14 e 15 pertencem a um sistema de regulação de dois componentes e as ORFs 28 e 30 pertencem à categoria de transportadores; ambos os pares possuem a mesma orientação. As ORFs 36, 37 e 38 também com mesma orientação, podem fazer parte de um operon e são constituintes de subunidades ribossomais. Em seguida (ORFs 39 e 40), encontramos duas ORFs que fazem parte do processo de divisão celular (Tabela 5).

Encontramos através do programa tRNA scan um tRNA para arginina na posição 3155 à 3232 (-). Este tRNA possui 78 pb com anticódon AGC localizado nas posições (relativas ao tRNA) 36-38.

Figura 3: ORFs encontradas pelo programa GLIMMER no cosmídeo 02D09. As ORFs hipotéticas anotadas na tabela 5 não foram representadas. As setas em preto indicam o nome da ORF e sua posição relativa ao cosmídeo; c= complemented e u= uncomplemented. As setas rosas indicam a proteína ou gene anotado.

Tabela 5: Anotação do cosmídeo 02D09. As ORFs são indicadas pelo nome do cosmídeo seguido da posição encontrada pelo GLIMMER. As colunas “de” e “até” indicam a posição relativa da ORF ao cosmídeo. A coluna denominada de “fita” indica a direção relativa da ORF à sequência FASTA do cosmídeo; (-) = transcrição ocorrendo no sentido da sequência FASTA, (+) = transcrição ocorrendo no sentido inverso. A coluna “acesso” indica em qual banco de dados se encontra a proteína mais similar, onde Sp= Swiss Prot e Gi= Gene Bank. Em vermelho, ORFs que atualmente não se encontram na versão nova de ORFs da *Xylella*. Estas, podem ter se juntado à ORFs vizinhas ou formado novas ORFs de maior tamanho. Neste último caso, a nomenclatura mudou.

ORF	de	até	Fita	Proteína	E-value	Categoria	Acesso	Organismo
01	119	6	-	Hipotética 4.19 kDa		VIII.B		
02	96	239	+	Hipotética 5.02 kDa		VIII.B		
03	305	174	-	Hipotética 4.87 kDa		VIII.B		
04	314	460	+	Hipotética 5.31 kDa		VIII.B		
05	488	619	+	Hipotética 4.74 kDa		VIII.B		
06	519	641	+	Hipotética 4.94 kDa		VIII.B		
07	667	515	-	Hipotética 6.13 kDa		VIII.B		
08	798	664	-	Hipotética 4.97 kDa		VIII.B		
09	885	1037	+	Hipotética 5.12 kDa		VIII.B		
10	2780	1020	-	Transportador ABC	e-102	V.A.7 IV.A	Gi 6460730	<i>Deinococcus radiodurans</i>
11	2758	2877	+	Hipotética 4.18 kDa		VIII.B		
12	3511	4533	+	Subunidade ribossomal 30S, proteína S2	2e-85	III.B.2	Sp P34831	<i>Spirulina platensis</i>
13	4701	5600	+	Fator de elongação EF-TS	2e-66	III.C.1	Sp P02997	<i>E.coli</i>
14	6438	5860	-	Sistema de 2 componentes, proteína regulatória	5e-29	I.E VII.G	Gi 5524743	<i>Sinorhizobium melioli</i>
15	7652	6435	-	Sistema de 2 componentes, proteína sensora	4e-17	I.E	Gi 703269	<i>S.melioli</i>
16	7647	7844	+	Hipotética 7.06 kDa		VIII.B		
17	7874	9139	+	GcpE	1e-39	VIII.A	Gi 6458062	<i>M.tuberculosis</i>
18	9120	9803	+	Conservada hipotética 23.45 kDa	2e-09	VIII.A	Gi 4155367	<i>H.pylori</i>
19	10748	9813	-	Conservada hipotética 32.10 kDa	2e-19	VIII.A	Sp O25832	<i>Pseudomonas sp</i>
20	11797	10742	-	UDP-N-acetilpiruvilglucosam ina redutase	3e-59	IV.B	Gi 290456	<i>E.coli</i>
22	12873	11794	-	Dihidroorotato oxidase	1e-81	II.B.2	Sp P05021	<i>E.coli</i>
23	12760	12903	+	Hipotética 5.05 kDa		VIII.B		
24	13032	13130	+	Hipotética 3.62 kDa		VIII.B		
25	13582	13139	-	Hipotética 16.30 kDa		VIII.B		
26	13524	13751	+	Hipotética 7.84 kDa		VIII.B		
27	13726	13881	+	Hipotética 5.59 kDa		VIII.B		
28	14529	13789	-	Transportador ABC	9e-61	V.A.7 IV.A	Sp Q56903	<i>Yersinia enterocolitica</i>

Continuação da tabela 5								
ORF	de	até	Fita	Proteína	E-value	Categoria	Acesso	Organismo
30	15339	14536	-	Transportador ABC	5e-41	V.A.7 IV.A	Sp Q56902	<i>Y. enterocolitica</i>
31	15864	16034	+	Hipotética 6.43 kDa		VIII.B		
32	16035	16130	+	Hipotética 3.68 kDa		VIII.B		
33	17538	16138	-	Asparaginil-tRNA sintetase	0.0	III.B.4	Sp P17242	<i>E.coli</i>
34	17662	17519	-	Hipotética 5.62 kDa		VIII.B		
35	17749	18087	+	Conservada hipotética 11.60 kDa	6e-24	VIII.A	Sp P36539	<i>E.coli</i>
36	18327	18758	+	Subunidade ribossomal 30S, proteína S6	2e-33	III.B.2	Sp O07813	<i>Neisseria gonorrhoeae</i>
37	18770	19000	+	Subunidade ribossomal 30S, proteína S18	7e-22	III.B.2	Sp P02374	<i>E.coli</i>
38	19056	19607	+	Subunidade ribossomal 50S, proteína L9	1e-33	III.B.2	Sp P02418	<i>E.coli</i>
39	19825	23328	+	Proteína segregadora do cromossomo, SMC	1e-118	V.B	Gi 2633966	<i>B.subtilis</i>
40	23387	24118	+	Proteína de divisão celular	1e-14	V.B	Sp P77173	<i>E.coli</i>
41	24733	24846	+	Hipotética 3.93 kDa		VIII.B		
42	24750	24854	+	Hipotética 3.96 kDa		VIII.B		
43	24847	27360	+	DNA ligase	1e-155	III.A.1	Sp P15042	<i>E.coli</i>
44	27350	27844	+	Hipotética 18.67 kDa		VIII.B		
45	27841	28791	+	Lisil-tRNA sintetase	4e-69	III.B.4	Sp P03812	<i>E.coli</i>
46	28943	29686	+	Hipotética 27.52 kDa		VIII.B		
47	29843	30907	+	Fator iniciador de tradução, subunidade IF-2B alfa	2e-86	III.C.1	Gi 4981447	<i>Thermotoga maritima</i>
48	31058	33781	+	DNA girase, subunidade A	0.0	III.A.1	Sp P48372	<i>P.aeruginosa</i>
49	33850	36525	+	Conservada hipotética 91.84 kDa	1e-109	VIII.A	Gi 2275219	<i>Arabidopsis thaliana</i>
50	37011	38804	+	Proteína ativadora de hemolisina	9e-44	VII.C	Gi 1772622	<i>Erwinia chrysanthemi</i>
51	39518	38916	-	Hipotética 20.83 kDa		VIII.B		
52	40659	39787	-	Succinil-CoA sintetase, cadeia alfa	1e-98	I.C.7	Sp P07459	<i>E.coli</i>
53	41851	40688	-	Succinil-CoA sintetase, cadeia beta	1e-121	I.C.7	Sp P07460	<i>E.coli</i>

3.4 Anotação do cosmídeo 09A12

A região anotada vai da posição 251 à 36.028 totalizando 36 ORFs. Note que as ORFs 13, 15 e 16 não estão presentes devido ao fato de estarem sobrepostas à outras ORFs vizinhas de maior tamanho. Neste caso, foram removidas manualmente.

No início do cosmídeo observamos a ocorrência de várias proteínas hipotéticas (em vermelho) não encontradas na nova versão de ORFs do genoma. As ORFs 20, 23 e 24 fazem parte de subunidades ribossomais. A ORF 22 (“GTP binding protein”) foi classificada como categoria VII.A (conservadas) pois sua função não foi estabelecida.

Utilizando o programa tRNA scan encontramos na posição 9910 à 9835 um tRNA para treonina (num total de 76 pb) com anticodon CGT (posição 34-36).

Figura 4: ORFs encontradas pelo programa GLIMMER no cosmídeo 09A12. As ORFs hipotéticas anotadas na Tabela não foram representadas. As setas em preto indicam o nome da ORF e sua posição relativa ao cosmídeo; c= complemented e u= uncomplemented. As setas rosas indicam a proteína ou gene anotado.

Tabela 6: Anotação do cosmídeo 09A12. As ORFs são indicadas pelo nome do cosmídeo seguido da posição encontrada pelo GLIMMER. As colunas “de” e “até” indicam a posição relativa da ORF ao cosmídeo. A coluna denominada de “fita” indica a direção relativa da ORF à sequência FASTA do cosmídeo; (-) = transcrição ocorrendo no sentido da sequência FASTA, (+) = transcrição ocorrendo no sentido inverso. A coluna “acesso” indica em qual banco de dados se encontra a proteína mais similar, onde Sp= Swiss Prot e Gi= Gene Bank. Em vermelho, ORFs que atualmente não se encontram na versão nova de ORFs da *Xylella*. Estas ORFs podem ter se juntado às ORFs vizinhas ou formado novas ORFs de maior tamanho. Neste último caso, a nomenclatura mudou.

ORF	de	até	fita	Proteína	E-value	Categoria	Acesso	Organismo
01	251	3817	+	Conservada hipotética 130.75 kDa	3e-80	VIII.A	Sp P39369	<i>E.coli</i>
02	3789	3962	+	Hipotética 6.19 kDa		VIII.B		
03	4514	3783	-	Conservada hipotética 26.80 kDa	7e-23	VIII.A	Pir S36438	<i>Podocoryne carnea</i>
04	5869	5961	+	Hipotética 3.42 kDa		VIII.B		
05	5968	6345	+	Hipotética 14.20 kDa		VIII.B		
06	6598	6756	+	Hipotética 5.78 kDa		VIII.B		
07	6984	4651	-	Conservada hipotética 85.54 kDa	3e-06	VIII.A	Gi 2314298	<i>H.pylori</i>
08	7270	7085	-	Hipotética 6.73 kDa		VIII.B		
09	7211	7387	+	Hipotética 6.26 kDa		VIII.B		
10	7747	8037	+	Hipotética 10.40 kDa		VIII.B		
11	9784	7979	-	Hipotética 67.88 kDa		VIII.B		
12	11341	10391	-	“lyzB protein”	1e-108	VII.C I.E	Gi 5915671	<i>Acinetobacter sp</i>
14	12085	11399	-	“Lipoprotein signal peptidase”	5e-30	III.C.1 IV.A	Sp P17942	<i>Pseudomonas fluorescens</i>
17	14882	12051	-	Isoleucina-tRNA sintetase	0.0	III.B.4	Sp P00956	<i>E.coli</i>
18	15826	14879	-	riboflavina quinase/FAD sintetase	6e-52	II.D.9	Sp P08391	<i>E.coli</i>
19	17597	15987	-	Fator de virulência MVIN	3e-85	VII.B	Sp P44958	<i>H.influenzae</i>
20	17637	17966	+	Subunidade ribossomal 30S, proteína S20	8e-19	III.B.2	Sp P02378	<i>E.coli</i>
21	18083	17658	-	Hipotética 15.91 kDa		VIII.B		
22	19222	18080	-	“GTP binding protein”	6e-94	VII.A	Sp P42641	<i>E.coli</i>
23	19582	19325	-	Subunidade ribossomal 50S, proteína L27	4e-26	III.B.2	Sp P02427	<i>E.coli</i>
24	19998	19606	-	Subunidade ribossomal 50S, proteína L21	3e-16	III.B.2	Sp P02422	<i>E.coli</i>
25	19639	20244	+	Hipotética 22.91 kDa		VIII.B		
26	20267	23164	+	excinuclease ABC, subunidade A	0.0	III.A.4	Sp P37434	<i>Salmonella typhimurium</i>
27	24238	23288	-	Hipotética 35.21 kDa		VIII.B		
28	24311	24204	-	Hipotética 3.91 kDa		VIII.B		
29	25796	24732	-	Hipotética 40.94 kDa		VIII.B		
30	27416	25848	-	GMP sintase	0.0	II.B.1	Sp P44335	<i>H.influenzae</i>

Continuação da tabela 6								
ORF	de	até	fita	Proteína	E-value	Categoria	Acesso	Organismo
31	27613	27491	-	Hipotética 4.55 kDa		VIII.B		
32	29109	27652	-	Dehidrogenase inosina-5'-monofosfato	1e-170	II.B.1	Sp P31002	<i>Acinetobacter calcoaceticus</i>
33	30113	29226	-	Metilenotetrahidrofolato desidrogenase/metilenotetrahidrofolato ciclohidrolase	6e-86	II.D.2	Sp P51696	<i>E.coli</i>
34	31222	30332	-	Hipotética 32.59 kDa		VIII.B		
35	33129	31222	-	Epimerase/desidratase	1e-111	IV.A	Gi 3135688	<i>Burkholderia pseudomallei</i>
36	34161	33142	-	"Lipopolysaccharide core biosynthesis protein"	2e-12	IV.A	Gi 836904	<i>Y. enterocolitica</i>
37	35351	34119	-	Conservada hipotética 45.17 kDa	4e-62	VIII.A	Sp P45576	<i>E.coli</i>
38	35625	35302	-	Hipotética 11.64 kDa		VIII.B		
39	36028	35648	-	Fator de integração ao hospedeiro, subunidade beta	8e-32	III.A.3	Sp Q51473	<i>P.aeruginosa</i>

3.5 Anotação do cosmídeo 07H03

Este cosmídeo foi anotado totalmente da região 2355 à 47470 totalizando 53 ORFs. As ORFs 17 e 18 fazem parte de um sistema de dois componentes de regulação de fosfato, que juntamente com as proteínas 15 e 16 podem fazer parte de um operon.

Neste cosmídeo, encontramos ORFs categorizadas nas diversas classes como metabolismo intermediário, biossíntese de pequenas moléculas, metabolismo de macromoléculas e adaptação.

Nenhum tRNA foi encontrado através do programa tRNA scan.

Continuação da tabela 7								
ORF	de	até	fita	Proteína	E-value	Categoria	Acesso	Organismo
09	3648	5006	+	“Outer membrane exporter factor”	3e-57	VII.C V.A.6	Sp P02930	<i>E.coli</i>
10	5285	5187	-	Hipotética 3.45 kDa		VIII.B		
11	5078	5299	+	Hipotética 8.17 kDa		VIII.B		
12	6397	5645	-	Conservada hipotética 27.5 kDa	3e-48	VIII.A	Sp P43341	<i>E.coli</i>
13	6462	6749	+	Hipotética 10.69 kDa		VIII.B		
14	6848	7027	+	Hipotética 7.04 kDa		VIII.B		
15	86556	7130	-	Exopolifosfatase	1e-102	I.B.9	Gi 4200042	<i>P.aeruginosa</i>
16	10821	8725	-	Polifosfato quinase	0.0	I.B.9	Gi 2463580	<i>P.aeruginosa</i>
17	12266	10926	-	Proteína sensora, reguladora de fosfato, sistema de dois componentes	1e-48	I.E	Sp P45608	<i>Klebsiella pneumoniae</i>
18	13084	12284	-	Regulador de transcrição, regulador de fosfato, sistema de dois componentes	7e-70	I.E	Sp P08402	<i>E.coli</i>
19	14678	13062	-	Conservada hipotética 62.5 kDa	1e-38	VIII.A	Gi 1788840	<i>E.coli</i>
20	14758	15114	+	Glutarredoxina 3	2e-17	II.D.10 I.C.3	Sp P37687	<i>E.coli</i>
21	15240	15076	-	Hipotética 5.88 kDa		VIII.B		
22	15596	15312	-	Hipotética 11.03 kDa		VIII.B		
23	15800	16807	+	Isocitrato desidrogenase	4e-75	I.C.7	Gi 6539562	<i>Oryza sativa</i>
24	17137	17024	-	Hipotética 4.48 kDa		VIII.B		
25	17756	17109	-	Hipotética 23.97 kDa		VIII.B		
26	17699	17809	+	Hipotética 4.33 kDa		VIII.B		
27	18091	18294	+	Hipotética 7.71 kDa		VIII.B		
28	18391	18912	+	Proteína-cisteína metiltransferase	4e-29	III.A.4	Sp P11742	<i>B.subtilis</i>
29	18941	20242	+	Precursor de fosfodiesterase-nucleotídeo pirofosfatase	3e-56	II.B.4	Gi 3925755	<i>S.pombe</i>
30	20335	20234	-	Hipotética 3.59 kDa		VIII.B		
31	20229	20339	+	Hipotética 3.93 kDa		VIII.B		
32	20797	21120	+	Ferridoxina	2e-43	I.C.3	Sp P00213	<i>Pseudomonas putida</i>
33	22659	21166	-	Mucina de intestino de inseto	3e-05	IX	Gi 2224921	<i>Trichoplusia ni</i>
34	23375	22776	-	Conservada hipotética 21.09 kDa	9e-34	VIII.A	Sp P46847	<i>E.coli</i>
35	23423	23815	+	Pterina-4-alpha-carbinolamina desidratase	6e-13	II.D.16	Sp P43335	<i>P.aeruginosa</i>

Continuação da tabela 7								
ORF	de	até	fito	Proteína	E-value	Categoria	Acesso	Organismo
36	23763	24236	+	Proteína <i>tonB</i>	2e-09	V.A.7 IV.D	Sp O34261	<i>Xanthomonas campestris</i>
37	25640	24648	-	Pseudouridilato sintase	4e-78	III.B.5	Sp P23851	<i>E.coli</i>
38	26037	25789	-	Hipotética 9.28 kDa		VIII.B		
39	26041	29439	+	Ribonuclease E	0.0	III.B.3	Sp P21513	<i>E.coli</i>
40	30532	29870	-	Regulador de transcrição, <i>GacA</i>	1e-47	I.E	Gi 4680222	<i>Pseudomonas aureofaciens</i>
41	31681	30782	-	Conservada hipotética 32.90 kDa	4e-34	VIII.A	Gi 1001499	<i>Synechocystis sp</i>
42	31981	31874	-	Hipotética 3.77 kDa		VIII.B		
43	33207	33094	-	Hipotética 4.31 kDa		VIII.B		
44	33320	33057	-	Hipotética 9.46 kDa		VIII.B		
45	33448	33299	-	Hipotética 5.51 kDa		VIII.B		
46	34336	33725	-	Superóxido dismutase (MN)	6e-94	VII.C	Sp P53654	<i>X.campestris</i>
47	34497	35018	+	Ribonuclease guanil-específica	3e-20	III.B.6	Sp P30289	<i>S.aureofaciens</i>
48	35015	35431	+	Hipotética 14.98 kDa		VIII.B		
49	37471	35486	-	Transportador ABC	1e-147	V.A.7	Sp P43672	<i>E.coli</i>
50	39250	37637	-	Conservada hipotética 59.10 kDa	6e-95	VIII.A	Sp P77649	<i>E.coli</i>
51	39359	39484	+	Hipotética 4.79 kDa		VIII.B		
52	39856	39599	-	Proteína B de aclimatação à temperatura	4e-25	VII.G	Sp P72192	<i>Pseudomonas fragi</i>
53	39880	39996	+	Hipotética 4.19 kDa		VIII.B		
54	40466	40371	-	Hipotética 3.77 kDa		VIII.B		
55	41336	43408	+	Diguanilato ciclase/ fosfodiesterase, domínio 2	2e-27		Gi 6458153	<i>D.radiodurans</i>
56	41519	43650	-	Proteína de choque térmico, HTPX	3e-81	VII.G III.C.3	Sp P23894	<i>E.coli</i>
57	45253	45146	-	Hipotética 4.05 kDa		VIII.B		
58	45566	45712	+	Hipotética 5.52 kDa		VIII.B		
59	47470	46232	-	Fator de elongação EF-Tu	1e-176	III.C.1	Sp P42481	<i>Thiobacillus cuprimus</i>

IV.4. ANÁLISE DA CATEGORIA DE BIOSÍNTESE DE PEQUENAS MOLÉCULAS

Tivemos como objetivos a busca pelas enzimas envolvidas nas principais rotas biossintéticas de pequenas moléculas. A organização e reclassificação das diversas ORFs anotadas nesta categoria baseou-se na descrição das vias de biossíntese de *Escherichia coli* por ser o microrganismo mais

estudado nessas condições e por pertencer ao mesmo grupo de classificação taxonômica de *Xylella fastidiosa* (Grupo Gama das Eubactérias).

Tendo como base a similaridade entre proteínas encontradas no banco de dados do BLAST, verificamos que, em *X. fastidiosa*, muitas vias biossintéticas estão incompletas. Entretanto, em princípio, considerando-se somente a similaridade de seqüências entre enzimas nos diferentes microrganismos não podemos inferir se uma rota está ou não ausente. São conhecidos vários microrganismos cujas seqüências de enzimas diferem das mais conservadas, mas que fornecem o ambiente necessário para que ocorra a reação catalisada.

Os primeiros complexos enzimáticos envolvidos na degradação da glicose em compostos alifáticos de carbono foram classificados na categoria I, de Metabolismo Intermediário. Os processos são exergônicos, isto é, produzem energia e suprem com esqueletos carbônicos as reações de síntese de pequenas moléculas, que são os componentes básicos na biossíntese de macromoléculas. Muitos destes intermediários (aminoácidos e nucleotídeos) contêm nitrogênio derivado do NH_4^+ . Outros tantos possuem enxofre, proveniente de íons SO_4^{2-} . Ao mesmo tempo, outras moléculas pertencentes à categoria de pequenas moléculas, como vitaminas e cofatores, são requeridas no funcionamento de muitas enzimas.

IV.4.1. Aminoácidos

A lista completa das vias de biossíntese de aminoácidos e dos genes ausentes e anotados em *Xylella fastidiosa*, encontra-se presente no item anexos a partir da página 99.

A oxidação da glicose e outros açúcares pela glicólise e ciclo do ácido tricarboxílico libera energia à célula bacteriana mas também provê unidades de seis à dois carbonos para a biossíntese de aminoácidos. A relação entre essas unidades e a formação dos aminoácidos, os blocos constitutivos das proteínas, está ilustrada na figura 6.

Xylella fastidiosa possui as vias metabólicas para a biossíntese de todos os aminoácidos. Algumas enzimas são hipoteticamente bifuncionais e outras não foram encontradas. Não foi possível identificar a enzima nitrato redutase para a assimilação direta de nitrogênio, contudo, sabemos que *Xylella fastidiosa* possui transportadores para amônia e para alguns aminoácidos dentre os quais, o glutamato e a glutamina, esta última apresentando-se em grande concentração na seiva bruta e sendo testada em meio de cultura mínimo como fonte única de nitrogênio, quando observou-se o crescimento da bactéria (Profª. Dra. Eliana Lemos- UNESP- Jaboticabal, comunicação pessoal).

ornitina é deacetilada por uma acetilornitina deacetilase codificada pelo gene *argE* presente em *X. fastidiosa*, *E. coli*, *B. subtilis*, *M. tuberculosis* e *H. influenzae*. Entretanto, em outras bactérias e fungos, o grupo acetil é reciclado em glutamato por uma ornitina acetiltransferase, igualmente presente em *B. subtilis*, sendo codificada neste microrganismo pelo gene *argJ*, o qual exerce também função de N-acetilglutamato sintase.

O controle da biossíntese de arginina foi estudado mais intensamente em *E. coli*. Os genes que controlam a biossíntese não constituem um operon neste microrganismo (MANDELSTAM *et al.*, 1982). Entretanto, todos estão sob controle regulatório do gene *argR*, não encontrado em *X. fastidiosa*. Em *E. coli*, quatro deles formam um grupamento (*argECBH*) e outros quatro ocorrem separadamente em pontos do genoma (MILLER, 1992). Todavia, em *X. fastidiosa*, o “cluster” é formado pelos genes *argFGEBCH*.

Uma peculiaridade encontrada no genoma da cepa K-12 de *E. coli* é o fato de existirem duas cópias do gene que codifica para as isozimas de ornitina carbamoiltransferase (*argF* e *argI*). O gene *argI* é o único encontrado em outras cepas de *E. coli* ou outras *Enterobacteriaceae*. Contudo, em *X. fastidiosa* observamos que somente uma cópia está presente, mas anotada como sendo o gene *argF*. Entretanto, quando comparamos sua sequência ao genoma de *E. coli*, observamos que sua semelhança é maior com o gene *argI* (BLASTp, E-value= 2e-22).

Xylella fastidiosa não possui a primeira enzima (N-acetilglutamato sintase) para a biossíntese de arginina. Contudo, esta enzima está também ausente em *H. influenzae* e *B. subtilis*, este último, apresentando um gene (*argJ*) com função dupla de N-acetilglutamato sintase e acetilornitina deacetilase. Uma consideração a ser feita aqui, é o fato da enzima N-acetilglutamato quinase (*argB*), encontrada em *X. fastidiosa* e em *Xanthomonas axonopodis* pv *citri* (Prof. Ronaldo Bento Quaggio, IQ-USP, comunicação pessoal) apresentar um domínio na região carboxiterminal (E-value= 0,082, P-fam), ainda que fraco, de N-acetilglutamato sintase (*argA*), fato não observado para a mesma enzima encontrada em *B. subtilis* (vide item anexos, página 155).

A via de biossíntese de arginina encontra-se também incompleta em *M. tuberculosis* e *H. influenzae*, estando totalmente ausente em *B. burgdorferi* e *C. pneumoniae*.

4.1.2. Glutamato

Todo nitrogênio para a síntese de macromoléculas nas bactérias é derivado do aminogruppo da glutamina e glutamato ou diretamente da incorporação de amônia.

O glutamato proporciona nitrogênio para a síntese da maioria dos aminoácidos, enquanto que a glutamina doa nitrogênio para a síntese de purinas, pirimidinas, histidina, triptofano, asparagina, NAD e *p*-aminobenzoato (NEIDHARDT *et al.*, 1987).

Quando a concentração de íons amônio no meio é suficientemente alta, amônia é incorporada diretamente em glutamato, glutamina e asparagina. Entretanto, quando sua concentração é baixa (menos de 0,1 mM) a amônia é incorporada somente em glutamina.

As reações responsáveis pela assimilação da amônia e síntese de glutamato e glutamina estão esquematizadas abaixo:

Estas três reações são catalisadas pela glutamina sintetase (*glnA*), glutamato sintase (*gltB* e *gltD*) e glutamato desidrogenase (*gdhA*), respectivamente. A reação catalisada pela glutamina sintetase é a única via biossintética para a síntese de glutamina. Quando cepas de *E. coli*, *Klebsiella* sp e *S. typhimurium* crescem num ambiente limitado em amônia, a glutamato desidrogenase não se envolve na sua assimilação e formação de glutamato. É papel da glutamina sintetase a assimilação de amônia e formação de glutamina e da glutamato sintase, a formação de glutamato.

Observamos que *Xylella fastidiosa* possui somente a via de conversão de glutamina em glutamato (genes *gltB* e *gltD*). Associado ao fato de que a glutamina está em grande concentração na seiva bruta e que a bactéria possui receptores para este composto, intui-se que esta via já é suficiente na obtenção de nitrogênio.

4.1.3. Glutamina

A glutamina sintetase (*glnA*) é essencial na formação de glutamina para a síntese de proteínas, purinas, pirimidinas e assimilação de amônia quando as células crescem num ambiente limitado em nitrogênio. Por ter um papel importante no metabolismo, sua regulação tanto em nível de síntese como de atividade catalítica é altamente controlada.

Como observado por Cássio Baptista em sua dissertação de mestrado (2000), *X. fastidiosa* possui um grupamento de genes no cosmídeo 07F02 potencialmente envolvidos na regulação da glutamina sintetase. Neste cosmídeo, a ORF gl28 (XF1842), que codifica para a glutamina sintetase (*glnA*) é regulada pela proteína codificada pela ORF gl29 (XF1843), proteína reguladora de nitrogênio,

PII (*glnB*). Ainda neste cosmídeo, a ORF gl30 (XF1844) está envolvida no transporte de amônia para o meio intracelular.

Altos níveis intracelulares de α -cetoglutarato induzem a formação de PII-UMP (vide anexo, página 104). PII-UMP interage com uma adenililtransferase (*glnE*) que catalisa a deadenilação da glutamina sintetase, ativando a enzima. Um alto nível intracelular de glutamina ativa a remoção de UMP de PII. PII interage com a adenililtransferase que catalisa a adenilação da glutamina sintetase, inibindo a síntese de glutamina.

PII também previne a conversão de NRI (ORF gl36 ou XF1848) a NRI-P catalisada pela proteína NRII (ORF gl37 ou XF1849). NRI-P é o ativador transcricional de *glnA*. Em níveis baixos de nitrogênio, NRII ativa a fosforilação de NRI a NRI-P que interage com sigma 54 ativando a transcrição de *glnA*.

4.1.4. Prolina

A biossíntese de prolina requer 3 enzimas codificadas pelos genes *proA*, *proB* e *proC*, presentes em *E. coli*, *B. subtilis*, *M. tuberculosis*, *N. meningitidis*, *H. influenzae* e *X. fastidiosa*. Outros microrganismos como *Pseudomonas aeruginosa*, *S. typhimurium* e *Brevibacterium flavum* também possuem a via de biossíntese completa (NEIDHARDT *et al.*, 1987).

Em *E. coli* os genes *proBA* formam um grupamento e *proC* tem localização pontual no genoma. O mesmo ocorre em *X. fastidiosa*. Contudo, não se sabe se em *E. coli* estes genes formam um operon transcrito por um mesmo promotor. Em *S. typhimurium*, *proBA* organizam-se em um operon transcrito na direção de *proB* para *proA*.

Estudos “in vitro” da enzima γ -glutamil quinase sugeriram que a atividade desta enzima é modulada por prolina e ADP (NEIDHARDT *et al.*, 1987).

4.2. Família do aspartato/ piruvato

4.2.1. Alanina

A L-alanina é formada a partir do piruvato por transaminação com glutamato catalisada pela alanina aminotransferase (*ilvE*), presente em *X. fastidiosa*, ou pela glutamato-piruvato transaminase (*gpt*). Em algumas bactérias a alanina pode ser formada diretamente pela aminação, dependente de NADH, do piruvato pela alanina desidrogenase. O piruvato também é o precursor de valina e leucina.

Em *E.coli*, observamos uma via alternativa na formação de alanina através da enzima valina piruvato aminotransferase (*avtA*), a qual é responsável pela transformação de L-valina e piruvato em L-alanina e α -cetoglutarato. Em *X. fastidiosa* esta enzima está ausente.

A única via para a síntese de D-alanina, um constituinte da parede celular, é a mudança de configuração de L-alanina. A alanina racemase catabólica de *E.coli* e *S. typhimurium* é induzida por L-alanina e por glicose. Em *X. fastidiosa* somente a alanina racemase biossintética está presente (gene *alr*). Mutações em ambos os genes (*alr* e *dadX*) são necessárias para torná-la D-alanina auxotrófica.

4.2.2. Asparagina

O aspartato dá origem a cinco aminoácidos, os constituintes das proteínas e ao diaminopimelato, que é encontrado em peptídeoglicanos da parede celular.

Em *E. coli* ocorre a presença de duas enzimas que transformam aspartato em asparagina: (i) *asnA* caracteriza-se por ser uma asparagina sintetase dependente de amônia e (ii) *asnB* utiliza a glutamina como doadora de nitrogênio. *X. fastidiosa* apresenta somente *asnB*, assim como *B. subtilis* (página 107, item anexos).

A presença de asparagina reprime tanto *asnA* como *asnB* em *E. coli*. Recentemente, o mecanismo de repressão por asparagina de *asnA* em *E. coli*, foi elucidado (MANDELSTAM *et al.*, 1982). Um gene designado de *asnC* (regulador de *asnA*) foi encontrado. Seu produto ativa a transcrição de *asnA*. Este gene também não está presente em *X. fastidiosa*, o que enfatizaria a ausência de uma proteína semelhante à *asnA* neste microrganismo.

Observamos ausência da via de biossíntese de asparagina em *B. burgdorferi*, *C. pneumoniae* e *N. meningitidis*.

4.2.3. Aspartato

O aspartato é sintetizado a partir do oxaloacetato por transaminação, com o glutamato sendo o doador do aminogruppo. As transaminases foram primeiramente estudadas por Rudman e Meister (NEIDHARDT *et al.*, 1987). Eles classificaram-nas em três grupos distintos pela variedade do aminoácido que doaria o aminogruppo ao α -cetoglutarato para formar glutamato (esta é a reação reversa das indicadas nos esquemas do item anexos). As transaminases A, têm atividade com aspartato, triptofano, tirosina e fenilalanina; as transaminases B (gene *ilvE*) atuam com leucina, isoleucina e valina e a transaminase C com a seguinte reação: L-valina + piruvato $\rightarrow$ 2-cetoisovalerato + L-alanina.

Somente em *B. burgdorferi*, observamos a ausência desta rota de biossíntese. *X. fastidiosa* apresenta uma ORF classificada nesta via, ainda que não fazendo parte da mesma, localizada no cosmídeo 06C10. Trata-se de uma aspartil/asparaginil beta hidrolase (XF2100) semelhante à mesma proteína em *Bos taurus* cuja função é a adição de um grupo hidróxido ao aspartato ou asparagina em domínios epidermais de alguns fatores de crescimento (informação retirada do “site” <http://www.dcc.ic.unicamp.br/xf/>).

4.2.4. Isoleucina, Valina

As vias de síntese de isoleucina e valina são, até certo ponto, paralelas. Ambas envolvem a condensação de um α -cetoácido com um grupo acetaldeído derivado do piruvato. Na via que leva à obtenção de valina, o aceptor do α -cetoácido é o próprio piruvato formando assim, acetolactato. Na via de isoleucina, o aceptor é o α -cetobutirato (2-oxobutanoato).

Deste maneira, na biossíntese de valina, o passo da condensação desloca um metabólito central especificamente para esta via. No caso de isoleucina, outro passo específico é necessário desde que o α -cetobutirato não é um intermediário das vias metabólicas centrais. Para esta função, *E. coli* e *S. typhimurium* utilizam uma treonina deaminase, produto do gene *ilvA*, que catalisa a formação do α -cetobutirato a partir de L-treonina. Este gene está ausente em *X. fastidiosa*. Contudo, observamos a presença de uma treonina deaminase catabólica (gene *tdcB*) que pertence ao mesmo COG da treonina deaminase biossintética, podendo servir como opção à ausência de *ilvA*.

As acetohidroxiácido sintases que catalisam a formação de α -acetolactato e acetohidroxibutirato estão na forma de três isozimas diferentes em *E. coli* e *S. typhimurium*. Nestes microrganismos, acetohidroxiácido sintase I e III são inibidas por valina. A atividade da acetohidroxiácido sintase II não é sensível a este composto, provavelmente levando em conta o fato de que muitas bactérias excretam valina no meio extracelular. Em *X. fastidiosa* somente a subunidade maior da acetolactato sintase II foi detectada (*ilvG*). *B. burgdorferi* e *C. pneumoniae* não possuem nenhuma das enzimas atuantes nesta via.

O próximo passo em cada uma das rotas que levam à valina e isoleucina é a isomerização e a redução dependente de NADPH dos acetohidroxiácidos por uma isômero redutase (*ilvC*) e a subsequente desidratação (*ilvD*) do intermediário formado para a obtenção dos precursores de isoleucina e valina. O último passo é a transaminação entre os 2 α -cetoácidos formados e o glutamato a

partir do produto de *ilvE*. Em *E. coli*, os genes *ilvGMEDAYC* formam um “cluster” e curiosamente, *X. fastidiosa* possui um agrupamento das enzimas *ilvGC* mas *ilvD* e *ilvE* são pontuais.

4.2.5. Leucina

Os genes *leuABCD*, em *E. coli*, organizam-se num operon. Observamos, em *X. fastidiosa*, que o gene *leuB* (posição XF2372 no mapa das ORFs, item anexos) é seguido de uma ORF hipotética e de *leuD* e *leuC* (anotado erroneamente no mapa como *leuA*, posição XF2375). A localização de *leuA* é pontual.

Nesta via biossintética, o α -cetoisovalerato, precursor imediato de valina, condensa-se com acetil-CoA sofrendo subseqüentes isomerização e descarboxilação oxidativa para formar α -cetoisocaproato. Este último, é o precursor imediato de leucina o qual é convertido tanto pela transaminase aromática (*tyrB*) quanto pela transaminase B (*ilvE*) em leucina. *X. fastidiosa* possui ambas as enzimas.

4.2.6. Lisina

Bactérias sintetizam lisina por uma via na qual o último passo é a descarboxilação do diaminopimelato. Este composto, assim como lisina em alguns organismos, faz parte do peptídeoglicano da parede celular.

Em *X. fastidiosa*, a ORF XF1116 codifica para uma possível enzima bifuncional com função de diaminopimelato descarboxilase e aspartoquinase III. Estas enzimas são codificadas pelos genes *lysA* e *lysC* em *E. coli*, que situam-se em regiões pontuais no genoma.

4.2.7. Metionina

Em *E. coli*, diaminopimelato, lisina, metionina, treonina e isoleucina, derivam parte ou todo seu esqueleto carbônico do aspartato (figura 6, p. 49). A primeira reação dessa via comum é catalisada, em *E. coli*, por três aspartato quinases, presentes em todas as *Enterobacteriaceae*, as quais diferem na maneira como são sintetizadas e reguladas.

X. fastidiosa não possui o gene *metL* (aspartoquinase II/ homoserina desidrogenase II), que em *E. coli* leva à produção de homoserina a partir do aspartato. Entretanto, é capaz de obter este composto através do produto do gene *thrA* (asparto quinase I/ homoserina desidrogenase I), o qual em *E. coli*

tem sua atividade regulada por treonina (asparto quinase II/ homoserina desidrogenase II não são reguladas por metionina).

O primeiro passo específico na biossíntese de metionina é catalisado pela homoserina succinil transferase, produto do gene *metA*, ausente em *X. fastidiosa*, o qual transforma homoserina em O-succinil homoserina. Todavia, foi verificado que *X. fastidiosa* e *M. tuberculosis* possuem uma O-acetiltransferase (*metA* em *M. tuberculosis*).

X. fastidiosa também não possui *metC* que leva à produção de homocisteína e nem *metZ*, que em *Pseudomonas aeruginosa*, faz o papel de enzima bifuncional (*metB* e *metC*) na conversão de homoserina à obtenção de homocisteína. Entretanto, *S. typhimurium* possui a enzima cistationina- γ -sintase que catalisa diretamente a reação de O-succinil-homoserina e sulfeto à homocisteína, pulando o intermediário cistationina (vide anexos, páginas 117 e 118).

X. fastidiosa possui a enzima que transforma homocisteína em L- metionina, independentemente de vitamina B₁₂ (cobalamina), entretanto, não possui a forma dependente de cobalamina mas possui a enzima (*metF*) que transforma 5,10 metileno THF em 5-metil-THF, que juntamente com homocisteína e através da catálise da transmetilase dependente de cobalamina, transformam-se em L-metionina.

4.2.8. Treonina

X. fastidiosa, *B. subtilis*, *M. tuberculosis*, *H. influenzae* e *N. meningitidis* possuem as mesmas enzimas de *E. coli* para a biossíntese de treonina, a partir de homoserina. *ThrA*, *thrB* e *thrC*, constituem um operon em *E. coli* e em *X. fastidiosa* estão localizadas em um “cluster” formando, provavelmente, um operon.

4.3. Família glicina/ serina

4.3.1. Cisteína

A biossíntese de cisteína em *S. typhimurium* e *E. coli* fornece um aminoácido necessário à síntese de proteínas, glutatião (GSH), metionina e outros metabólitos que contém enxofre. Também representa o meio pelo qual o enxofre é reduzido e fixado em moléculas orgânicas, um processo análogo à redução do nitrato e fixação de amônia. A redução de sulfato, nesses dois microorganismos é “assimilatória”, isto é, reduz somente a quantidade necessária para a biossíntese de aminoácidos que

contém enxofre e difere da redução de sulfato, realizada por organismos anaeróbios, na cadeia respiratória, como acceptor final de elétrons.

Em um ambiente contendo sulfeto, a biossíntese de cisteína é relativamente simples e requer a conversão de serina à *O*-acetilserina que então reage com sulfeto para formar cisteína. O crescimento bacteriano em formas oxidadas de enxôfre inorgânico, como o sulfato, requer um grupo mais complicado de reações que envolve sua mobilização e redução à sulfeto (esquema anexos, p. 120).

Outra via de incorporação direta do sulfeto em cisteína foi descrita em *S. typhimurium*. O tiosulfato pode reagir com *O*-acetilserina para produzir S-sulfocisteína, subseqüentemente reduzida à cisteína pela ação da *O*-acetilserina sulfidrilase, anotada em *X. fastidiosa*.

X. fastidiosa, não possui a enzima serina acetil transferase requerida no primeiro passo da conversão de L-serina à *O*-acetilserina. Nenhuma das enzimas para a incorporação direta de sulfato extracelular foi anotada.

4.3.2. e 4.3.3. Glicina/Serina

A biossíntese de serina, glicina e unidades de um Carbono (C1) constitui uma das vias metabólicas mais importantes aos microrganismos. Serina é usada na síntese de cisteína e triptofano e 3-fosfoserina, um intermediário na biossíntese de serina, é um precursor de piridoxina. A glicina é incorporada em purinas e compostos contendo o grupo heme. Unidades de C1 participam da síntese de purinas, timina, histidina e metionina.

Existem variações nas vias pelas quais a serina é formada a partir do 3-fosfoglicerato, um intermediário na degradação da glicose. Em alguns microrganismos (MANDELSTAM *et al.*, 1982), uma via não fosforilativa (que envolve glicerato e hidroxipiruvato) pode predominar, enquanto que, em outros microrganismos como *E. coli* e *S. typhimurium*, a via fosforilativa predomina (vide anexos, p. 121). A síntese de serina, nesses microrganismos, exige a presença de três enzimas: 3-fosfoglicerato desidrogenase (*serA*), 3-fosfoserinaminotransferase (*serC*) e 3-fosfoserina fosfatase (*serB*). Em *X. fastidiosa* não foi encontrada a ORF que codifica para a última enzima mas observou-se que esta também está ausente em *B. subtilis*.

A interconversão de serina em glicina ocorre pela ação da serina hidroximetiltransferase (*glyA*), presente em todos os microrganismos analisados. Esta última reação ainda produz 5,10-metilenotetrahidrofolato (5,10-mTHF), um importante contribuinte em unidades de C1 para o metabolismo celular.

4.4. Família dos aminoácidos aromáticos

4.4.1. Corismato

As vias de biossíntese dos aminoácidos aromáticos, fenilalanina, triptofano e tirosina possuem uma rota em comum, a partir do fosfoenolpiruvato e eritrose 4-fosfato, que leva à formação de corismato a partir do qual, as vias se ramificam. Essa via comum, anteriormente chamada de via do shiquimato, ainda possui quatro braços adicionais, que levam à formação de folato, ubiquinona, menaquinona e enterobactina, em *E. coli*, a partir do corismato.

O composto corismato, foi classificado nesta categoria, por ser o precursor dos aminoácidos citados.

Em *E. coli* e *S. typhimurium*, a conversão do fosfoenolpiruvato e de eritrose 4-fosfato em 3-deoxi-D-arabino-heptolosonato 7-fosfato é catalisada por três 3-deoxi-D-arabino-heptolosonato 7-fosfato sintases, isofuncionais separadas. Uma é inibida por tirosina, outra por fenilalanina e a terceira por triptofano (MANDELSTAM, *et al.*, 1982). Em *B. subtilis*, os produtos finais inibem as primeiras enzimas do ramo terminal de cada via. *X. fastidiosa* possui somente a sintase que é sensível a triptofano, no ramo comum que leva à biossíntese de corismato.

4.4.2. Fenilalanina e 4.4.4. Tirosina

A primeira reação, tanto para fenilalanina quanto para a biossíntese de tirosina, envolve a conversão do corismato ao prefenato. Os genes que codificam para (*tyrA*) a enzima bifuncional corismato mutase (COG 1605)/ prefenato desidrogenase (COG 0287) e (*pheA*) corismato mutase (COG 1605)/ prefenato desidratase (COG 0077) são inibidos por tirosina e fenilalanina, respectivamente. *X. fastidiosa* possui o gene *pheA* mas a ORF 08A04-gl33 (XF2338), que codifica para *tyrA*, está somente anotada como corismato mutase; entretanto, quando analisada pelo P-fam, verificamos que possui um domínio para prefenato desidrogenase (E-value= 2e-12).

A última reação das vias de síntese de fenilalanina e tirosina envolvem a transaminação dos α -cetoácidos respectivos, fenilpiruvato e P-hidroxifenilpiruvato, tendo o glutamato como doador do amino-grupo. Em *E. coli* e *S. typhimurium* essa reação pode ocorrer por conta de três enzimas em diferentes situações; a aminoácido aminotransferase (*ilvE*), a aminoácido aromático transaminase (*tyrB*), ou a aspartato aminotransferase (*aspC*). *X. fastidiosa* possui as três enzimas.

4.4.3. Triptofano

Como mostrado nas páginas 123 e 124 do item anexos, a primeira reação, a partir do corismato, na via de síntese do triptofano é catalisada pela antranilato sintase. A enzima ativa é composta por um agregado de duas moléculas; os peptídeos especificados pelos genes *trpE* e *trpD* e referidos como componentes I e II respectivamente. O componente I contém o sítio de ligação ao corismato. Na ausência do componente II, a enzima não catalisa a formação de antranilato com a glutamina sendo a fornecedora de nitrogênio (atividade de amidotransferase). Entretanto, essa reação pode ocorrer com eficiência reduzida na presença de amônia.

O componente II ainda está envolvido numa segunda atividade: a conversão do antranilato a antranilato 5-fosforibosil pirofosfato, atividade denominada de antranilato fosforibosil transferase.

Estudos mostraram (NEIDHARDT *et al.*, 1987) que em *S. typhimurium*, o sítio de ligação à glutamina pelo componente II se dá na região amino-terminal da molécula. Em *Serratia marcescens* as enzimas com função de amidotransferase e fosforibosiltransferase são codificadas por genes separados, *trpG* e *trpD*, respectivamente. Observamos, que *X. fastidiosa* possui também os dois genes com funções distintas. Os demais genes da via encontram-se presentes, contudo, o gene *trpR*, regulador da via em *E. coli*, não foi anotado.

4.5. Histidina

X. fastidiosa, *H. influenzae*, *M. tuberculosis*, *N. meningitidis* e *B. subtilis* possuem as mesmas enzimas para a biossíntese de histidina a partir de 5-fosforibosil-1-pirofosfato e adenosina tri-fosfato. As dez etapas desta via incluem reações complexas e não usuais. As oito enzimas que catalisam as reações são codificadas por genes que constituem um possível operon *his* em *X. fastidiosa* (ORFs XF2213 à 2222, genes *hisIFAHBCDG_S*). Duas proteínas, produtos dos genes *hisB* e *hisI* são bifuncionais e catalisam dois passos separados. Os substratos iniciais, fosforibosilpirofosfato (PRPP) e ATP, são centrais ao metabolismo intermediário e ligam a via de biossíntese de histidina à outras vias divergentes como a biossíntese de nucleotídeos de pirimidina, purina e piridina.

Em *E. coli* e *S. typhimurium* os produtos dos genes *hisR*, *hisS*, *hisT*, *hisU* e *hisW* estão relacionados ao controle do operon *his*. Nestes microrganismos estes genes são pontuais e os demais fazem parte do operon. Em *X. fastidiosa* estes genes reguladores não foram encontrados (exceto *hisS*, que encontra-se ao lado do possível operon. Vide mapa das ORFs no item anexos, página 98).

IV.4.2. Nucleotídeos

4.2.1. Ribonucleotídeos de purina

Nucleotídeos de purina podem ser derivados de fontes exógenas pela chamada via de salvamento ou podem ser sintetizados *de novo* a partir de precursores simples. Em ambos os casos o motivo de ribose 5'-fosfato é derivado do 5-fosforibosil- α -1-pirofosfato (PRPP).

O PRPP é requerido para a biossíntese, *de novo* e salvamento de purinas, pirimidinas, nucleotídeos de nicotinamida e para a biossíntese de histidina e triptofano. Sua síntese é catalisada pela fosforibosilpirofosfato sintetase.

A síntese de IMP (Inosina 5'-monofosfato) *de novo* contém 11 passos a partir de PRPP. São as enzimas codificadas pelos genes *purF*, *purD*, *purN*, *purT*, *purL*, *purM*, *purE*, *purC*, *purB*, *purH*, *purH*. Todos foram anotados em *Xylella* exceto o gene *purT*, glicinamida ribonucleotídeo transformilase 2, não requerente de folato, que pode ter sua atividade compensada pelo produto do gene *purN* (NEIDHARDT *et al.*, 1996).

A partir de IMP os genes *guaB*, *guaA* e *purA*, *purB* são responsáveis pela conversão à GMP e AMP respectivamente. Todos esses genes estão presentes em *X. fastidiosa*. Nucleotídeos de adenina e guanina podem ser interconvertidos pelo precursor IMP desde que as reações de síntese *de novo* são reversíveis. Todavia, outra via de interconversão entre esses compostos existe, onde são responsáveis os genes *add*, *amn*, *apt*, *deoD*, *gpt*, *gsk*, *guaA*, *guaB*, *guaC*, *hpt* e *purB*. Destes, somente *guaA*, *guaB*, *hpt* e *purB* estão presentes em *X. fastidiosa*. Ainda, AMP e GMP podem ser convertidos em ATP e GTP pelos genes *adk* e *gmk*, anotados em nossa categoria (NEIDHARDT *et al.*, 1996). Outros genes da tabela de conversão, foram retirados do “site” <http://ecocyc.pangeasystems.com> e estão ausentes em *X. fastidiosa*.

4.2.2. Ribonucleotídeos de pirimidina

Compostos intracelulares de pirimidina são quase que exclusivamente encontrados como nucleotídeos. Mais de 95% dos nucleotídeos são encontrados na fração ácida insolúvel das células como ácidos nucleicos. Esta fração consiste de nucleosídeos mono-, di-, e trifosfatos (NMPs, NDPs, e NTPs) e algumas coenzimas. As coenzimas que contêm nucleotídeos juntamente com mRNAs levam à uma reciclagem de NMPs e NDPs (NEIDHARDT *et al.*, 1996). A via de biossíntese *de novo* de

pirimidinas (NTPs) é uma sequência de reações na qual dTTP é o produto final e UTP, CTP e dCTP são compostos intermediários. O produto da primeira reação, carbamoilfosfato, é também um composto da biossíntese de arginina e constitui o único ramo desta via.

A síntese de UTP e CTP envolvem as enzimas *carAB*, *pyrBI*, *pyrC*, *pyrD*, *pyrE*, *pyrF*, *pyrH*, *ndk*, *pyrG* e *cmk*. *X. fastidiosa* não possui a subunidade regulatória da aspartato carbamoiltransferase (*pyrI*) mas a subunidade catalítica está presente. Os genes *pyrG* e *pyrH* foram classificados na categoria I, de metabolismo intermediário mas participam dessa via de síntese. A enzima que converte dCTP a dTTP não foi citada.

4.2.3. 2'-Desoxirribonucleotídeos

Deoxirribonucleosídeos são sintetizados a partir de seus ribonucleosídeos correspondentes por enzimas denominadas redutases. Nestas reações, a hidroxila C-2 da D-ribose é substituída por um hidrogênio.

Esta via está completa em *X. fastidiosa*, exceto pelo gene *nrdD*, presente em *E. coli* e *H. influenzae*, ativado sob condições de anaerobiose com papel substituível pela atividade da enzima ribonucleosídeo difosfato redutase (*nrdB*).

4.2.4. Salvamento de nucleosídeos e nucleotídeos

A via de salvamento de pirimidina em *E. coli* serve a três funções fisiológicas. A primeira é a assimilação de bases exógenas e nucleosídeos. A segunda é a disponibilização das pentoses provenientes de nucleosídeos exógenos como fonte de carbono e energia e os amino-grupos de compostos que contêm citosina como fonte de nitrogênio. A terceira função é de reutilização de bases livres e nucleosídeos produzidos intracelularmente a partir da reciclagem destes últimos (NEIDHARDT *et al.*, 1996).

Dos genes envolvidos no salvamento de pirimidinas (*cdd*, *codAB*, *tsx*, *nupCG*, *tdk*, *deoA*, *upp*, *uraA*, *udk* e *udp*), nenhum encontra-se anotado em *X. fastidiosa*. Somente *B. subtilis* possui esta via completa, com exceção do gene *tsx*, o qual não pertence ao “Cluster of Ortholog Groups”. *X. fastidiosa* também não possui os demais genes envolvidos no salvamento de purinas.

IV.4.3. Nucleotídeos e nucleotídeos de açúcares

Antiga classificação em *E. coli* para a síntese de nucleotídeos de açúcares; inexistente atualmente. Duas ORFs foram classificadas nesta categoria; um precursor de glucolactonase e uma manosil transferase.

IV.4.4. Cofatores, grupos prostéticos e transportadores

4.4.1. Biotina

O precursor na via de síntese de biotina é o ácido 7-ceto-8-aminopelargônico (7-KAP). Em *E. coli* pimeloil coenzima A (CoA) condensa-se com L-alanina para formar 7-KAP.

A sequência de reações para a síntese de pimeloil CoA, ainda não foi totalmente elucidada mas suspeita-se que os genes *bioC* e *bioH* participem da etapa anterior a síntese de 7-KAP. Esta aminocetona sofre transaminação para dar origem ao ácido 7,8 diaminopelargônico (DAPA), o qual incorpora CO₂ para formar detiobiotina. A introdução de enxofre com a formação do anel de tetrahidrotiopeno é o último passo na sequência de reações.

Dos microrganismos estudados, somente *X. fastidiosa*, *H. influenzae* e *N. meningitidis* possuem ORFs com possível função de *bioC* e *bioH*. Nesses e nos demais microrganismos (com exceção de *B. burgdorferi*) a síntese de biotina, a partir do ácido 7-ceto-8-aminopelargônico, parece estar completa.

4.4.2. Ácido fólico

Em contraste com humanos, que necessitam de folato na dieta, muitos microrganismos e plantas possuem a habilidade de síntese de ácido fólico. O produto final da biossíntese do ácido fólico é o dihidrofolato. Este derivado do ácido fólico não é ativo nas reações de transferência de 1 Carbono e deve ser reduzido à tetrahidrofolato. Todos os derivados do ácido fólico que atuam como reservatórios de unidades de 1 Carbono derivam-se do tetrahidrofolato, o qual é formado a partir do dihidrofolato por uma redução dependente de NADPH e catalisada pela dihidrofolato redutase.

A via de biossíntese do ácido fólico pode ser dividida em duas partes:

Primeiro, o corismato (precursor aromático) é convertido ao ácido *p*-aminobenzoico (PABA) pela ação de três enzimas (produtos dos genes *pabA*, *pabB* e *pabC*), ausentes em *X. fastidiosa*, *C. pneumoniae* e *H. influenzae*. Verificou-se que o gene *pabB* possui 26% de similaridade ao componente I da antranilato sintase, gene *trpE* (mesmos COGs 0147), contudo, apesar da similaridade, a atividade de *pabB* não é substituída por *trpE* e vice-versa (NEIDHARDT *et al*, 1996).

Segundo, o motivo de pteridina do folato é derivado do GTP numa série de reações que compreende cinco enzimas. A enzima GTP ciclohidrolase (*folE*), converte GTP em H₂-neopterina trifosfato numa reação composta de quatro passos. A conversão de dihidroneopterina trifosfato à dihidroneopterina ocorre pela ação de duas enzimas distintas. A primeira etapa é catalisada pela dihidroneopterina trifosfato pirofosfatase e a remoção do terceiro grupo fosfato é catalisada por uma fosfomonoesterase não específica (representada pela dihidroneopterina monofosfato desfosforilase, no esquema, página 134).

A dihidroneopterina aldolase (*folB*) catalisa a formação de 2-amino-4-hidroxi-6-hidroximetil-7,8-dihidropterina (H₂-neopterina) a partir de dihidroneopterina e a 6-hidroximetil-7,8-dihidropterina pirofosfoquinase (*folK*) converte H₂-neopterina em 2-amino-4-hidroxi-6-hidroximetil-7,8-dihidropterina difosfato (H₂-pteroato). *FolE*, *folB* e *folK*, estão presentes em *B. subtilis*, *M. tuberculosis*, *H. influenzae* e *X. fastidiosa*. As duas enzimas que levam à conversão de H₂-neopterina-trifosfato à dihidropterina não foram anotadas em *X. fastidiosa* e não foram analisadas para os demais microorganismos.

Após esta série de reações, o PABA e a 6-hidroximetil-7,8-dihidropterina pirofosfato são unidos para formar dihidropteroato pela dihidropteroato sintase. A enzima bifuncional *folC* adiciona então uma molécula de glutamato para formar dihidrofolato (dihidropteroilmonoglutamato). Este último é reduzido à tetrahydrofolato pela dihydrofolato redutase (*folA*) e a enzima bifuncional *folC* (função de folipoliglutamato sintase) adiciona 2 resíduos de glutamato para formar tetrahydropteroiltriglutamato. Este ramo da síntese de ácido fólico está completo em *X. fastidiosa* e demais microrganismos analisados (com exceção de *B. burgdorferi*).

4.4.3.Lipoato

O ácido lipóico é derivado do ácido octanóico, obtido diretamente da via de biossíntese de ácidos graxos. Desta maneira, a via de ácido lipóico consiste de dois passos nos quais átomos de

enxofre são introduzidos no esqueleto carbônico do ácido octanóico nas posições C-8 e C-6. WHITE (1986) mostrou que o enxofre do ácido lipóico é derivado da cisteína.

Dois genes estão envolvidos na síntese do ácido lipóico em *E. coli*: *lipA* e *lipB*, ambos presentes em *X. fastidiosa*, *M. tuberculosis*, *N. meningitidis* e *H. influenzae*. Mutantes *lipB*⁻ retém a habilidade parcial em sintetizar ácido lipóico mas produzem baixos níveis de atividade da piruvato desidrogenase e α -cetoglutarato desidrogenase (REED & CRONAN, 1993). Dados sugerem que o produto de *lipA* é necessário para a inserção do primeiro enxofre no esqueleto do ácido octanóico. A função exata de *lipB* ainda não foi estabelecida.

4.4.4. Molibdopterina

Estudos do cofator molibdênio na sulfito oxidase do fígado, levaram à caracterização estrutural da pterina denominada molibdopterina (MPT). MPT é um dos componentes de uma família de cofatores encontrados em todas as enzimas que contém molibdênio, com única exceção da nitrogenase.

Dentre os microrganismos analisados, possuem via de síntese de molibdopterina somente *E. coli*, *B. subtilis*, *M. tuberculosis* e *H. influenzae*. *X. fastidiosa* não possui capacidade de síntese de molibdopterina pela via em questão (como pode ser verificado na tabela da página 136, anexos). Duas proteínas, que aqui deveriam estar classificadas (XF0193 e XF2604), foram acomodadas numa nova categoria (4.4.16).

4.4.5. Pantotenato e coenzima A

O ácido pantotênico, uma das vitaminas B, é fornecido pela condensação dependente de ATP do ácido pantóico e ácido 2-aminopropiônico (β -alanina). O pantotenato é usado principalmente na biossíntese de coenzima A e proteína carregadora de acila. A biossíntese de CoA pode ser dividida em três partes: a formação de pantoato e β -alanina; a síntese de pantotenato; e a conversão deste último à coenzima A. As reações estão sumarizadas na página 137 do item anexos.

Em *E. coli*, a via principal de síntese de β -alanina se dá por α -descarboxilação do ácido aspártico (gene *panD*, também presente em *X. fastidiosa*). Baseados na observação de que mutantes *panD*⁻ cresciam com suplementação de dihidrouracil ou β -ureidopropionato, SLOTNICK e WEINFELD (1957) propuseram que a β -alanina para a biossíntese de pantotenato podia ser derivada do catabolismo de uracila, fato corroborado posteriormente. Finalmente, também parece possível que a β -alanina pode ser derivada por transaminação do semialdeído malônico.

O precursor do ácido pantóico é o α -cetoisovalerato, um intermediário na biossíntese de valina. Duas enzimas participam deste ramo na síntese de pantotenato: cetopantoato hidroximetiltransferase (*panB*) e 2-dehidropantoato redutase. Somente *panB* foi anotada em *X. fastidiosa*.

X. fastidiosa possui também a enzima pantotenato sintase (pantoato- β -alanina ligase, produto do gene *panC*) e é capaz de catalisar a síntese de pantotenato dependente de ATP. Entretanto, não possui *coaA*, primeira enzima na via de síntese de coenzima A. Outros microrganismos como *B. subtilis* e *N. meningitidis* também carecem desta e das demais enzimas que participam da síntese de CoA baseada na via presente em *E. coli*; levando a hipótese de que esta via não é conservada entre os microrganismos analisados.

4.4.6. Piridoxina (vitamina B₆)

Em *E. coli* e outros microrganismos, duas vias de obtenção de piridoxina podem ser definidas. A primeira é a via de biossíntese e a segunda, é a via de recuperação. A via de recuperação serve à conversão de outras formas de piridoxina, transportadas para o meio intracelular, em formas coenzimaticamente ativas. Piridoxina fosfato oxidase (*pdxH*) e piridoxal quinase participam desta via.

Na sua biossíntese, 1-Deoxy-D-xilulose serve como precursora de unidades de 5 carbonos, C-2', 2, 3, 4, 4', do piridoxol e 4-hidroxi-L-treonina, gerada pela condensação de glicina com glicolaldeído, é precursora de C-6, 5, 5', N-1 (NEIDHARDT, 1996).

Em *X. fastidiosa* foram anotados somente os genes *pdxA*, *dxs*, *pdxJ* e *pdxH* (vide anexos, p. 140). Contudo, o gene *serC* (EC 2.6.1.52), anotado na via de biossíntese de serina também pode estar envolvido na conversão de 3-hidroxi-4-fosfo-hidroxi- α -cetobutirato em fosfo-hidroxi-treonina. Mesmo assim, este ramo da via de síntese de piridoxal 5'-fosfato, ainda encontra-se incompleto pela ausência de genes similares aos *gapA*, *epd* e *pdxB* de *E. coli*.

4.4.7. Nucleotídeos de piridina

O NAD⁺ pode ser sintetizado a partir do aspartato através de uma série de cinco reações e é reciclado por vias conhecidas como ciclo dos nucleotídeos de piridina (PNCs).

A primeira enzima na via de biossíntese do NAD⁺, L-aspartato oxidase (produto de *nadB*), produz ácido iminoaspártico a partir do L-aspartato. Este intermediário é muito instável e se decompõe em oxaloacetato e amônia se a quinolinato sintase (produto do gene *nadA*) estiver ausente. Esta enzima

catalisa a condensação do iminoaspartato e dihidroxiacetona fosfato em quinolinato. A transcrição de *nadAB* é estritamente regulada pelo repressor *nadR* (ausente em *X. fastidiosa*).

O quinolinato é descarboxilado e convertido em um nucleotídeo pela QA fosforibosiltransferase (*nadC*). O produto desta reação, NAMN, é o ponto de convergência das vias de síntese e de reciclagem de NAD^+ .

O NAMN, produzido por reciclagem ou síntese de novo, é adenilado por *nadD* produzindo deamino-NAD, o qual sofre amidação por *nadE* (NAD sintase) para produzir NAD^+ . A principal diferença entre NAD sintases de bactérias e fungos é o doador do grupo amida. Bactérias utilizam NH_3 , enquanto fungos utilizam NH_3 ou glutamina.

X. fastidiosa possui esta via de síntese completa.

4.4.8. Tiamina

A tiamina, ou vitamina B_1 , é formada pela condensação de motivos de pirimidina e tiazol, os quais são sintetizados independentemente para convergir ao ramo de síntese de tiamina. A forma ativa da vitamina é a tiamina pirofosfato. Dessa maneira, a biossíntese de tiamina pode ser dividida em três partes: as vias individuais que levam à obtenção de pirimidina (4-amino-5-hidroximetil-2-metilpirimidina) e tiazol (5-(2-hidroxietil)-4-metiltiazol), e a conversão desses compostos na forma ativa da coenzima.

A via de síntese de tiamina, caracteriza-se pela sua diversidade bioquímica (NEIDHARDT, 1996). Diferenças entre microrganismos existem, não somente nas rotas que levam aos precursores de cada um dos motivos, mas também nos estágios finais dos processos de biossíntese.

Em *E. coli*, foi clonado um agrupamento de 5 genes (*thiCEFGH*) envolvidos nas vias de síntese dos motivos individuais (MILLER, 1992). Em *X. fastidiosa*, os genes *thiC*, *thiE/B*, *thiF* e *thiG*, são pontuais no genoma. Este microrganismo carece dos genes *thiJ/N*, *thiM* e *thiH* (o qual pode ter sua atividade substituída pelas proteínas *thiG* e *thiF*).

4.4.9. Riboflavina (vitamina B_2)

Verificou-se que a via de biossíntese de riboflavina é conservada em todos os microrganismos analisados (exceto *B. burgdorferi*, onde encontra-se ausente). Todavia, existem algumas diferenças

entre as enzimas analisadas: as proteínas GTP ciclohidrolase II e 3,4-dihidro-2-butanona-4-fosfato sintase são, em *E. coli*, codificadas pelos genes *ribA* (COG 0807) e pelo gene *ribB* (COG 0108), respectivamente, e localizam-se pontualmente no genoma. Entretanto, em *B. subtilis*, *M. tuberculosis* e *C. pneumoniae*, esse gene é bifuncional. *X. fastidiosa* possui as ORFs XF0953 e XF1992, anotadas como GTP ciclohidrolase II/ 3,4-dihidro-2-butanona-4-fosfato sintase e proteína de biossíntese de riboflavina. Quando procuramos verificar suas similaridades ao banco de dados do “COG” verificamos que a primeira ORF citada é semelhante à 3,4-dihidro-2-butanona-4-fosfato sintase e que a última é semelhante somente à GTP ciclohidrolase II.

Além disso, *X. fastidiosa* possui duas ORFs: XF0950 e XF1748 estão anotadas como riboflavina deaminase e 5-fosforibosilamino uracil redutase, respectivamente. Em *E. coli*, o gene *ribD* tem papel bifuncional equivalente à atividade dessas duas proteínas.

4.4.10. Tiorredoxina, glutarredoxina, glutatíão

A biossíntese de glutatíão (GSH), a partir de L-glutamato e L-cisteína, está completa em *X. fastidiosa*. Genes que codificam as proteínas tiorredoxina (*trxA*) e glutarredoxina (*grxC*) estão presentes.

4.4.11. Menaquinona (vitamina K₂) e ubiquinona (coenzimaQ)

Bactérias Gram-negativas, anaeróbias facultativas, incluindo *E. coli* e *S. typhimurium* possuem geralmente isopreno-quinonas da série benzênica e naftalênica. As benzoquinonas são chamadas de ubiquinonas (Q-n) e as naftoquinonas são chamadas menaquinonas (MK-n) ou demetilmenaquinonas (DMK-n). Onde n refere-se ao número de unidades isoprenil presentes na cadeia lateral (LEHNINGER, 1976).

Apesar do fato de que ambas as quinonas originam-se da via do shiquimato, existem diferenças importantes entre elas. Uma delas é a formação do núcleo quinonóide onde as vias divergem a partir do corismato. Na via de biossíntese de ubiquinona, o corismato perde um grupo piruvóil pela ação da corismato liase (*ubiC*), ausente em *X. fastidiosa* e demais microrganismos analisados; para formar 4-hidroxibenzoato. A enzima necessária ao passo seguinte (*ubiA*) está presente em *X. fastidiosa* e leva à formação de 3-octaprenil-4-hidroxi benzoato.

Mutantes de *E. coli*, *ubiD*⁻, conseguem sintetizar 20% do total de 2-octaprenilfenol produzido por uma cepa selvagem, provavelmente por uma via alternativa de descarboxilação do 4-hidroxibenzoato (NEIDHARDT *et al.*, 1987). *X. fastidiosa* também não possui essa enzima.

As reações de hidroxilação subseqüentes, levam a conversão do 2-octaprenilfenol à 2-octaprenil-3-metil-5-hidroxi-6-metoxi-1,4-benzoquinona. Observou-se que mutantes de *E. coli* nas monoxigenases aeróbicas; *ubiB*⁻, *ubiH*⁻ e *ubiF*⁻, crescidos em condições de anaerobiose, levavam à síntese de ubiquinona levantando a hipótese de que outras enzimas de hidroxilação alternativas devem estar presentes nestas condições (NEIDHARDT *et al.*, 1987). Destes, somente o gene *ubiH* foi identificado em *Xylella* e nos demais microrganismos analisados.

UbiG e *ubiE*, enzimas responsáveis pelas reações de metilação alternadas às reações de hidroxilação, também estão presentes em *X. fastidiosa* e *N. meningitidis*.

Nenhuma das enzimas participantes da via de biossíntese de menaquinona está presente em *X. fastidiosa* com exceção de *ubiE*, que participa de ambas as vias em reações de metilação.

4.4.12. Proto e siroheme

Hemes são membros da família de tetrapirróis, que incluem também as clorofilas, bilinas e corrinas. Eles têm participação no metabolismo energético, processos oxidativos; são componentes do aparato de transferência de elétrons e atuam como grupos prostéticos em várias situações.

Os passos na biossíntese de hemes, a partir de precursores comuns, constituem o ramo conservado da via e suas ramificações levam à grupos específicos de produtos como proto e siroheme. Eucariotos não vegetais e α -proteobactérias têm como precursores de uroporfirinogênio III, a glicina e succinil CoA. *E. coli* e outros microrganismos utilizam o glutamato (vide p. 143). Nos microrganismos analisados, verificamos que todos, exceto *B. burgdorferi* e *H. influenzae* possuem conservadas as enzimas participantes dessa etapa.

A partir daí, a via ramifica-se para dar origem à siroheme, o qual é um grupo prostético de muitas nitrito e sulfito redutases, e protoheme.

O siroheme é derivado do uroporfirinogênio III por metilação, oxidação e quelação catalisadas por uma enzima com tripla função em *E. coli*. *CysG* ou siroheme sintase, está também presente em *X. fastidiosa*, *N. meningitidis*, *B. subtilis* e *M. tuberculosis*.

X. fastidiosa também é capaz de sintetizar protoheme pela mesma via de *E. coli*. A função do gene *hemG* (protoporfirina oxidase, ausente em *X. fastidiosa*) pode ser substituída pelo produto do gene *hemK*.

4.4.13. Proteína transportadora de carboxil biotina (BCCP)

A conversão de acetil-CoA em malonil-CoA na biossíntese de ácidos graxos, é catalisada pela enzima acetil-CoA carboxilase, uma enzima polimérica constituída de filamentos longos de monômeros enzimáticos. Cada unidade monomérica contém quatro subunidades diferentes. Uma dessas subunidades é denominada de proteína-biotina-carboxil-transportadora (BCCP). As ORFs incluídas nesta categoria tanto em *E. coli* como em *X. fastidiosa* estão atualmente classificadas na Biossíntese de ácidos graxos (II.E).

4.4.14. Cobalamina (vitamina B₁₂)

Adenosil-cobalamina (Ado-CBL) é uma das moléculas biologicamente ativas mais importantes das encontradas na natureza. Pertence à classe dos tetrapirróis cíclicos e é sintetizada somente em procariotos, incluindo aeróbios e anaeróbios estritos e anaeróbios facultativos tanto em bactéria como em arqueobactérias.

X. fastidiosa não possui capacidade de síntese de vitamina B₁₂, entretanto, possui receptores para este composto (XF1604).

4.4.15. Enterobactina

Enteroquelina é um sideróforo que capacita *E. coli* à recrutar ferro do ambiente. É um composto de três unidades de 2,3-dihidroxibenzoilserina unidos numa estrutura cíclica por pontes de lactona.

Enterobactina é sintetizada sob condições de baixas concentrações de ferro e excretada ao meio ambiente onde liga-se a Fe(III) com alta afinidade e especificidade. Os complexos de ferrissideróforos formados são aceitos pela célula por transportadores específicos.

X. fastidiosa não possui nenhum dos genes presentes nesta via, contudo, genes para receptores de enterobactina (ORFs XF2134 e 2137) e transportadores de ferro (ORFs XF 0932 e 0933) foram identificados.

4.4.16. Biopterina

Esta categoria está ausente na classificação de *E. coli*, entretanto, foi criada para acomodar as ORFs 05G02-GL31 (XF0193), que corresponde à proteína 6-piruvil tetrahydrobiopterina sintase (*ptpS*) e 07H03-GL35 (XF2604), que corresponde à proteína pterina-4- α -carbinolamina desidratase. Contudo, estas proteínas estão envolvidas na síntese de tetrahydropterina, que em *E. coli* fazem parte da via de síntese do cofator molibdopterina.

IV.4.5. Biossíntese de ácidos graxos e fosfatídicos

A glicose nos microrganismos é convertida, pela glicólise, a piruvato e acetil-CoA, a partir da qual são sintetizados os ácidos graxos. Na reação global de síntese que é catalisada por um conjunto de sete proteínas: o complexo da ácido graxo-sintetase; a acetil-CoA é a precursora da cadeia de ácido graxo e a malonil-CoA contribui com outras unidades de carbono. Dessa maneira, um resíduo acetílico e outros malonílicos sofrem sucessivas etapas de condensação para formar a molécula de ácido graxo.

Os genes anotados nesta via de síntese encontram-se na tabela (item anexos, página 152). Além das enzimas necessárias a síntese: acetil-CoA carboxilase, ACP acetiltransferase, ACP malonil-transferase, 3-oxoacil-ACP sintase, 3-oxoacil-ACP redutase, uma família de 3-hidroxiacil ACP desidratases e enoil-AP redutase; outras enzimas foram anotadas em *X. fastidioasa*.

IV.4.6. Poliaminas

A biossíntese de putrescina, espermidina e cadaverina é bem conhecida em *E. coli* (anexos, p. 154). Em pHs baixos, estes compostos são sintetizados e provavelmente constituem um mecanismo de defesa contra a acidez, dado às suas características básicas. Mutantes onde a primeira enzima da via de síntese de putrescina está ausente não são capazes de crescer em pHs ácidos. As poliaminas ligam-se fortemente aos ácidos nucleicos e são abundantes em células em rápida proliferação.

A diamina putrescina pode ser feita diretamente por descarboxilação da ornitina (gene *speC*) ou indiretamente pela descarboxilação da arginina em agmatina (*speA*) seguida de sua hidrólise para formar putrescina (*speB*). Destas enzimas, somente *speA* está presente em *X. fastidiosa*.

Putrescina e S-adenosilmetionina descarboxilada geram espermidina pela catálise dos genes *speD* e *speE*, presentes em *X. fastidiosa*. A cadaverina é produto da descarboxilação da lisina pela lisina descarboxilase (*cadA*) que está ausente em *X. fastidiosa*.

V. DISCUSSÃO

V.I. Estratégias de seqüenciamento

A introdução de métodos de seqüenciamento de DNA automatizados, juntamente com os avanços no campo da bioinformática, introduziram a biologia e a medicina em uma nova era; a era da ciência genômica.

O desenvolvimento de métodos para a montagem e seqüenciamento de grandes segmentos de DNA é um ponto crítico nos projetos de seqüenciamento genômico. Um assunto muito abordado num projeto de seqüenciamento é a decisão de quais as estratégias que serão utilizadas para que este tenha um andamento rápido, mas preciso, de uma maneira fácil para chegar-se ao resultado final.

A estratégia de seqüenciamento utilizando “shotgun” é, atualmente, a mais difundida. Até 1995, o maior genoma seqüenciado por uma estratégia de fragmentação aleatória foi o do bacteriófago Lambda com um tamanho genômico de 48.502 pb (SANGER *et al.*, 1982), até que o microrganismo *Haemophilus influenzae* (FLEICHMANN *et al.*, 1995) de 1.9 Mpb teve seu genoma seqüenciado por este método, quando sua eficiência foi comprovada. Esta aproximação provou, até o momento, ser a de menor custo dentre as outras estratégias como bibliotecas cosmídeos e fago Lambda pois não requer um “set” de grandes segmentos ordenados ou outros subclones, enquanto proporciona alta redundância e exatidão minimizando o esforço requerido para o fechamento do genoma.

A questão mais crítica na utilização desta estratégia é a disponibilidade de programas e algoritmos capazes de interpretar e juntar o grande número de dados gerados. Como o número de regiões não amostradas cresce com o tamanho do genoma, isto é, quanto maior o genoma, mais “gaps” podem ocorrer e como o tamanho dessas regiões não é determinado, utilizam-se, comumente, cosmídeos-alvos que representam a região não amostrada no seqüenciamento pela abordagem de “shotgun”.

Outro ponto negativo da estratégia de “shotgun” é que pequenas seqüências repetitivas entremeadas, tendem a confundir a montagem. Contudo, um fator importante na escolha dessa estratégia é o fato de que sendo os plasmídeos pequenos, genes que codificam proteínas nocivas às bactérias hospedeiras devido à clonagem poderão ter sua expressão interrompida pelo fato de estarem incompletos. Tais regiões clonadas em um cosmídeo normalmente levariam à perda do clone, uma vez que a proteína tóxica poderia inviabilizar o crescimento da bactéria hospedeira. O seqüenciamento automático, com rendimento de 500 a 700 pb, reduz os problemas associados aos elementos repetitivos na montagem final e tem sido utilizado para minimizar rearranjos e outros artefatos de clonagem (FRASER & FLEICHMANN, 1997).

A utilização de cosmídeos teve como objetivo o aprendizado da técnica, uma vez que o projeto genoma de *Xylella fastidiosa* foi o pioneiro no país e seu intuito era além de determinar sua seqüência genômica treinar os laboratórios nas diversas técnicas de seqüenciamento. Um ponto favorável na utilização de cosmídeos foi o estabelecimento de uma base para um banco de dados de anotação do genoma uma vez que a anotação foi iniciada com base em cada cosmídeo. O tamanho dos cosmídeos também foi importante para a montagem correta de regiões que continham repetições de grande número de nucleotídeos. Entretanto, esta metodologia sofre com a clonagem de insertos quiméricos, formados por fragmentos do genoma de diferentes regiões que se ligam mas que não são vizinhas no genoma original.

O laboratório de seqüenciamento coordenado pelo Prof. Dr. Paulo Arruda, na Unicamp, apontou como principal problema na confecção das bibliotecas de Fago Lambda a dificuldade na preparação de DNA em larga escala para seqüenciamento e a presença de quimeras.

Assim, a melhor opção para se seqüenciar um microrganismo seria primeiramente a obtenção de grande quantidade de leituras obtidas a partir de bibliotecas de “shotgun” associada a utilização complementar de métodos que permitam a inserção de grandes regiões do genoma alvo, como bibliotecas de cosmídeo e de Fago Lambda.

V.II Análise da categoria de biossíntese de pequenas moléculas

Xylella fastidiosa cresce lentamente no xilema de seus hospedeiros, ambiente considerado hostil quando comparado ao habitat de outros microorganismos, alimentando-se da seiva bruta e formando agregados celulares.

Analizamos a biossíntese de pequenas moléculas fazendo uma comparação entre as vias centrais em *Escherichia coli* e de alguns outros organismos cujos genomas foram seqüenciados e anotados. Para isto, utilizamos o serviço de busca de vias funcionais do banco de dados do COG onde também procurava-se saber se enzimas com seqüências diferentes nos diversos microrganismos poderiam desempenhar o mesmo papel das proteínas consideradas ausentes em *Xylella fastidiosa*. Genes ortólogos são relacionados por descendência vertical e são responsáveis pelas mesmas funções em organismos diferentes, em contraste aos genes parálogos, os quais são homólogos relacionados por duplicação e têm funções similares mas não idênticas (MUSHEGIAN & KOONIN, 1996).

Procuramos comparar organismos com tamanhos de genoma diferentes como os menores *Borrelia burgdorferi* e *Chlamydia pneumoniae* e maiores como *Escherichia coli*, *Mycobacterium tuberculosis* e *Bacillus subtilis*. Entretanto, a análise das vias nesses microrganismos mostra-nos que seus habitats naturais suprem a ausência completa de muitas vias biossintéticas. Por exemplo podemos citar a ausência total das vias de biossíntese de aminoácidos em *Borrelia burgdorferi*, bactéria causadora de artrite reumatóide e doença de Lyme e da ausência completa de vias de biossíntese de arginina, glutamato, glutamina, prolina, asparagina, isoleucina, valina, lisina, cisteína, serina, triptofano e histidina em *C. pneumoniae*. Da mesma maneira, podemos inferir que pelo fato do xilema ser um meio pobre em nutrientes em comparação à outros habitats de patógenos animais, *X. fastidiosa* necessita de vias metabólicas completas para sintetizar a maioria das moléculas enquadradas nesta categoria.

X. fastidiosa possui mecanismos especiais para concentrar e absorver nutrientes do ambiente. De fato, foram identificados um total de 140 genes que codificam proteínas relacionadas ao transporte, representando um total de 4.8% de todas as ORFs anotadas. Existe um grande número de transportadores de íons, carboidratos, aminoácidos, peptídeos, sulfato, fosfato e vitamina B₁₂. Cinco receptores externos de membrana foram encontrados, incluindo sideróforos, ferro-cromo e heme-receptores envolvidos no transporte de ferro que levam à hipótese de que devido a mobilização deste e outros íons, como o manganês, por parte da bactéria, ocorra uma redução nos micronutrientes essenciais à planta, levando aos sintomas típicos de clorose observados nas folhas.

Dentre os transportadores identificados em *Xylella fastidiosa*, alguns são específicos para o transporte de glutamato e uma série de outros são membros da família dos transportadores ABC. Estes transportadores são sistemas que utilizam energia da hidrólise do ATP para importar nutrientes celulares como por exemplo aminoácidos, dentre os quais a glutamina, cuja concentração na seiva

bruta é a maior dentre os compostos orgânicos, sugerindo servir como fonte de nitrogênio e precursora dos demais aminoácidos cujas vias foram identificadas.

Observamos que embora o gene *gdhA* da biossíntese de glutamato não esteja presente como em *E. coli*, *Xylella* é capaz de produzir glutamato através dos genes *gltB* e *gltD* que codificam para glutamato sintase que tem também como função, além da biossíntese de glutamato, a remoção de glutamina, o primeiro produto de assimilação de amônia. Primeiramente a amônia é incorporada pela glutamina sintetase (*glnA*) em glutamina e então, a amida da glutamina é transferida ao 2-cetoglutarato para formar duas moléculas de glutamato. O glutamato proporciona nitrogênio para a síntese da maioria dos aminoácidos. Estas enzimas proporcionam uma rota de incorporação do nitrogênio em compostos orgânicos e formam um elo de ligação entre o metabolismo de carboidratos e aminoácidos.

Fato interessante é a disposição de alguns genes no cosmídeo 07F02, sequenciado e anotado no Laboratório IC por Cássio Baptista. As proteínas XF1842, 1843, 1844, 1848 e 184949 formam um grupamento de genes, interrompido por ORFs hipotéticas, potencialmente envolvidos no metabolismo de nitrogênio.

Observamos que as vias de biossíntese de aminoácidos como arginina, isoleucina, valina, metionina, cisteína e serina estão aparentemente incompletas em *X. fastidiosa*. As demais possuem suas rotas completas, mesmo que em determinados passos não catalisem as mesmas reações que *E. coli*.

Em princípio considerando-se somente a similaridade de seqüências entre enzimas de uma via metabólica não se pode descartar a presença de uma rota. A inspeção do conjunto de genes ortólogos identificou um número razoável de enzimas ausentes em vias essenciais. Entretanto, há casos em que as enzimas evoluíram de uma maneira diferente mas acabam por catalisar uma mesma reação. Como exemplo, podemos citar a enzima aminoácido acetiltransferase (gene *argJ*) de *B. subtilis* que catalisa o primeiro e quinto passos da via de biossíntese de arginina.

Na família do glutamato, a via que leva à síntese de arginina encontra-se incompleta quando comparada à rota utilizada por *E. coli*. A estrutura da enzima N-acetilglutamato quinase (*argB*) em *Xylella fastidiosa* (E-value=4.4e-22, P-fam) mostrou-nos que esta possui um domínio carboxi-terminal da enzima acetilglutamato sintase (E-value=0.082, P-fam), primeira enzima da via de biossíntese de arginina, e única ausente para esta via em *Xylella fastidiosa*. Em *Escherichia coli* observa-se que *argA* (E-value=1.1e-09) divide também um domínio teórico, ainda mais forte, com *argB* (E-value=1.8e-22). Neste caso porém, não sabemos se esse domínio de *argB* é funcional, embora a proteína esteja anotada somente como acetilglutamato sintase. Para a função de quinase, um outro gene em *E. coli* (*argB*)

exerce esse papel e esse gene não possui o domínio de acetiltransferase. Entretanto, observamos que em *B. Subtilis* a sintase também está ausente quando comparada à sintase de *E. coli*, mas que neste microrganismo, existe uma enzima bifuncional (*argJ*), citada anteriormente, que faz além desse papel, a função do gene *argE*. Esta enzima, não tem nenhuma semelhança no genoma de *X. fastidiosa* levando-nos a crer que neste microrganismo, N-acetilglutamato quinase pode ter um papel bifuncional, com função de *argB* e *argA*, pois o mesmo ocorre em *Xanthomonas citri*, outro fitopatógeno com genoma sendo seqüenciado atualmente. Se este fato for comprovado, podemos inferir uma duplicação parcial do gene original bifuncional, em *E. coli*, com a posterior perda da função do domínio de *argA* (vide item anexos, páginas 155 e 156).

Duas enzimas, em *E. coli* são capazes de catalisar o primeiro passo da via de biossíntese de isoleucina: a treonina desidratase biossintética, e a treonina desidratase catabólica; opção aparentemente utilizada por *X. fastidiosa* para compensar a ausência da primeira enzima. Somente a subunidade maior de uma das isozimas de acetolactato sintase II foi detectada em *X. fastidiosa* (*ilvG*). Esta última enzima, foi purificada de *S. typhimurium* e consiste em 2 subunidades; uma maior e outra menor. O gene que codifica para a subunidade maior está a montante do gene que codifica para a subunidade menor (*ilvM*). Na ausência de *ilvM* nenhuma atividade “in vitro” da proteína pôde ser quantificada mesmo quando um plasmídeo “high-copy” contendo *ilvG* era inserido na bactéria mutante *ilvM* (NEIDHARDT *et al.*, 1987). Deste modo, a via que leva à biossíntese de valina encontra-se incompleta quando comparada à via de *E. coli*.

A biossíntese de leucina e lisina encontram-se completas em *X. fastidiosa*. O composto diaminopimelato é o precursor da síntese de lisina e de compostos da parede celular. Uma conexão metabólica liga a síntese de lisina à síntese do material da parede celular. As sínteses de lisina, metionina e treonina dividem duas reações no começo de cada via biossintética; as reações correspondentes à síntese de L-aspartato semialdeído a partir do L-aspartato. A primeira reação é catalisada pela aspartato quinase na forma de três isozimas, cada uma participando num dos três produtos finais. Cada uma dessas isozimas de aspartato quinase são reguladas pelo produto final correspondente. Observamos a presença de uma enzima bifuncional em *X. fastidiosa* que supre a função dos genes *lysA* e *lysC*, últimos passos na biossíntese de lisina. *X. fastidiosa* também possui o gene regulador dessa via (*lysR*).

A via de cisteína está aparentemente incompleta em *X. fastidiosa* e *N. meningitidis*. Cisteína é usada na síntese de metionina, ambos contendo enxofre. Esta via, encontra-se totalmente ausente em *B. burgdorferi* e *C. pneumoniae*.

X. fastidiosa não possui o gene que em *E.coli* leva ao primeiro e terceiro passo da via de biossíntese de metionina (gene *metL*). *X. fastidiosa* e *M. tuberculosis*, também não possuem a enzima homoserina O-succiniltransferase, quarto passo na biossíntese de metionina. Entretanto, foi verificado que ambos microorganismos possuem uma homoserina O-acetiltransferase (*metA* em *M. tuberculosis*). *X. fastidiosa* também não possui *metC* que leva à produção de homocisteína e nem *metZ*, que em *Pseudomonas aeruginosa*, faz o papel de enzima bifuncional (*metB* e *metC*) na conversão de homoserina à obtenção de homocisteína. *X. fastidiosa* possui a enzima que transforma homocisteína em L- metionina, independentemente de vitamina B₁₂ (cobalamina), entretanto não possui a forma dependente de cobalamina mas possui a enzima (*metF*) que transforma 5,10 metileno THF em 5-metil-THF, que juntamente com homocisteína e através da catálise da transmetilase dependente de cobalamina, transformam-se em L-metionina. A cobalamina é levada à célula através de transportadores, pois *X. fastidiosa* não possui a via de síntese usual para esse composto.

X. fastidiosa não é capaz de transformar 3-fosfoserina em L-serina mas possui capacidade de transformar 3-fosfoglicerato em 3-fosfoserina através das enzimas produtos dos genes *serA* e *serC*. Como também possui o gene *glyA*, responsável pela biossíntese de glicina a partir de L-serina, este aminoácido deve ser produzido por alguma enzima desconhecida que substitui a atividade de *serB*.

O corismato é um intermediário na síntese de três aminoácidos: tirosina, fenilalanina e triptofano. A primeira reação na síntese de corismato é catalisada por três enzimas separadas em *E. coli*, cada uma sujeita à regulação do conteúdo celular dos três aminoácidos citados; *X. fastidiosa* possui somente a isozima responsiva ao triptofano. As vias de biossíntese de triptofano, fenilalanina e tirosina, a partir do corismato, estão completas.

Todos os genes para biossíntese de ribonucleotídeos de purina e pirimidina estão presentes com exceção de *purT* para purinas, que codifica a glicinamida ribonucleotídeo transformilase 2, não requerente de folato mas que pode ter sua atividade compensada pelo produto do gene *purN*. Os genes *nrdA* e *nrdD* também estão ausentes no metabolismo de deoxiribonucleotídeos.

Xylella fastidiosa parece conseguir sintetizar e alongar ácidos graxos a partir de acetato, embora alguns genes não tenham sido encontrados como é o caso de *acpD* e *acpS*, este último, também não sendo encontrado em *Synechocystis* sp, *Haemophilus* sp e *Mycoplasma genitalium*. Na etapa seguinte, existem duas ORFs para *fabF* e nenhuma para *fabB*. Estes genes, são similares e pertencem ao mesmo grupo comparativo ortólogo (COG).

Xylella fastidiosa é capaz de sintetizar uma grande variedade de cofatores e grupos prostéticos. Estão incompletas as vias de biossíntese de ácido fólico, pantotenato e coenzima A, além de

piridoxina, tiamina e ubiquinona. Contudo, este microrganismo não é capaz de sintetizar cobalamina, molibdopterina, menaquinona e enterobactina.

V. III. Conclusão

Nos últimos anos, o seqüenciamento genômico proporcionou a oportunidade de comparações entre organismos diferentes em nível molecular.

Pela primeira vez, tornou-se possível comparar o conjunto completo de proteínas presentes na célula de um determinado organismo e identificar as proteínas responsáveis para cada função celular. Nos casos em que nenhuma proteína conhecida é encontrada para a realização de uma função particular, os seus substitutos podem ser previstos a partir da carga de genes com função ainda não estabelecida.

A análise computacional de genomas procarióticos completos mostrou, até o momento, que as proteínas microbianas são, em geral, bem conservadas — 70% delas contém regiões conservadas (KOONIN *et al*, 1998). Este fato permitiu delinear famílias de ortólogos entre uma ampla extensão filogenética, e em muitos casos prever funções com considerável precisão.

O conhecimento de genes conservados responsáveis pela manutenção das funções celulares levou-nos à informação a respeito da carga mínima de genes necessária para a sobrevivência da célula. MUSHIGIAN & KOONING (1996) compararam os genomas de *H. influenzae* e *M. genitalium* e catalogaram os genes conservados em ambos os organismos. Em seu trabalho, adicionaram genes com função anteriormente desconhecida para completar as etapas críticas das vias metabólicas cruciais para a manutenção da vida celular. Esta lista, pode representar o genoma mínimo necessário para que se procedam essas reações celulares.

Deste modo, subtraindo-se o total de genes encontrados num dado microrganismo do total de genes do genoma mínimo, fatores que especificam as características fenotípicas que fazem de um organismo ser uma espécie distinta talvez estejam sendo revelados. Se esse exercício é realizado em uma espécie patogênica, os genes resultantes podem incluir traços necessários à sua virulência. Muitos genes de virulência são encontrados como ilhas de patogenicidade. Três genes de hemolisinas foram encontrados em *X. fastidiosa* sendo dois deles vizinhos e anotados no cosmídeo 05B03 (ORFs XF2397 e 2398). Juntamente com os genes de hemaglutininas, fazem parte de uma categoria de genes que até então não havia sido identificada em nenhum fitopatógeno. As hemolisinas não dividem em *Xylella*, extensivas homologias, exceto no domínio C-terminal onde as repetições ricas em glicina se localizam

(Prof. Márcio Lambais, comunicação pessoal). Essas repetições são caracteristicamente distribuídas entre patógenos Gram-negativos indicando que em *X. fastidiosa* possa ter ocorrido um mecanismo de transferência lateral de genes.

O principal objetivo da realização do seqüenciamento de *Xylella fastidiosa* foi a obtenção de dados a respeito dos possíveis mecanismos de patogenicidade da mesma e suas relações bioquímicas com o hospedeiro e vetor. Outra meta ainda a ser atingida é saber qual a diversidade genética entre as diferentes cepas dos diferentes patovares que compõe a espécie *Xylella fastidiosa*, (MACHADO, 1987) como vem sendo feito através da realização do novo projeto genoma de *Xylella*, causadora da doença de Pierce, em videiras.

Questões relacionadas ao tamanho real do genoma da cepa seqüenciada, a ocorrência de plasmídeos relacionados à patogenicidade em citros, e a presença de genes de virulência ou avirulência, que estariam relacionados aos sintomas da CVC ou aos vetores, puderam ser respondidas.

A comparação gênica entre microrganismos tem se tornado importante na busca de novas tecnologias para o combate à doenças dado o grande número de microrganismos seqüenciados. Assim, através da anotação dos genes de *Xylella fastidiosa* (SIMPSON *et al.*, 2000), foram identificadas uma variedade de mecanismos associados à virulência e patogenicidade da mesma como, por exemplo, os genes responsáveis pela produção de uma goma polissacarídica (EPS), provável causadora do agrupamento da bactéria no xilema que leva ao bloqueio dos vasos xilemáticos e conseqüentemente ao estresse hídrico e a produção de fímbrias, já anteriormente observadas por micrografia eletrônica, responsáveis pela adesão da bactéria ao aparelho bucal do hospedeiro (genes *pilS*, *pilR*, *pilG*, *pilH*, *pilI*, *pilJ* e *chpA*).

Concluimos que embora alguns genes não foram identificados em *Xylella*, estes podem não ser essenciais à via de síntese em questão ou podem ter sua função substituída por proteínas que ainda não foram identificadas, inerentes à *Xylella fastidiosa*. Enquanto que a obtenção de seqüências genômicas já caíram na rotina de muitos laboratórios, a análise funcional dos genes preditos aparece como um desafio. Através do projeto do genoma funcional de *X. fastidiosa* e a comparação com o genoma da cepa causadora da doença de Pierce, esperamos que genes atualmente considerados “putativos” e determinantes de patogenicidade tenham suas funções estabelecidas.

V.I. BIBLIOGRAFIA

- ALSBERG, S., DELSHER, A., KASIF, S., WHITE, O. Microbial gene identification using interpolated Markov Models. *Nucleic Acids Research*, v.26, p.544-548, 1998.
- ALTSCHUL, F.S., KOONIN, E.V. Iterative profile searches with PSI-BLAST: a tool for discovery in protein databases. *TIBS*, v.23, p.444-447, 1998.
- ALTSCHUL, F.S., GISH, W. Local alignment statistics. *Methods of Enzymology*, v.266, p.460-480, 1996.
- ALTSCHUL, F.S., 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 Res.*, v.25, p.3389-3402, 1997. Apud: NCBI [Online]. Available: <http://www.ncbi.nlm.nih.gov/blast/psiblst.cgi> [September, 2000].
- BAPTISTA, C. Seqüenciamento e anotação de quatro regiões (136.635 pb) do genoma de *Xylella fastidiosa*. São Paulo, 2000. 95p. (Dissertação de Mestrado- Instituto de Ciências Biomédicas II- Universidade de São Paulo).
- BENSON, D.A., BOGUSKI, M.S., LIPMAN, D.J., OSTELL, J., OULLETTE, B.F., RAPP, B., WHEELER, D.L. GenBank. *Nucleic Acids Research*, v.27, p.12-17, 1999.
- BERETTA, MJG., GARCIA JR, A., LEE, E.F., DERRICK, K.S., BARTLE, G., TOMÉ NETO, J. Observation on citrus variegated chlorosis in Brazil. Proc. Of the 12th In: CONFERENCE OF THE INTERNATIONAL. IOCV. 12, 1993, India. *Proceedings...* p.291-293. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.
- BIRBOIN, H.C., DOLY, J. A rapid alkaline extration procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, v.7, p.1513-1523, 1979.
- BLATTENER, F.R., *et al.* The complete genome sequence of Escherichia coli K-12. *Science*, v.277, p.1453-1474, 1997.
- BOLETIM FUNDECITRUS. Manual de convivência com a CVC. 2.ed. Araraquara, 1998. n.p.
- BOWMAN, S., *et al.* The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature*, v.400, p.532-538, 1999.
- CASTRO, P.R. de C. e C. Aspectos fisiológicos do amarelinho. In: ENCONTRO TÉCNICO SOBRE AMARELINHO DOS CITROS, 1992, Piracicaba. P.24-26. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.
- CHANG, C.J., *et al.* Culture and serological detection of the xylem limited bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. *Current Microbiology*, v.27, p.137-142, 1993a.

- CHANG, C.J., *et al.* Citrus variegated chlorosis: cultivation of the causal bacterium and experimental reproduction of the disease. Conference of the International IOCV.12, India, Proceedings, p.294-300, 1993b. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.
- COCHRAN, L.C. , HUTCHINS, L.M. Virus disease of stone fruit. *Handbook*, v.437, p.96-102, 1976. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro:FAPESP/FUNDECITRUS, 1998. 166p.
- COLLINS, F.S., PATRINOS, A., JORDAN, E., CHAKRAVARTI, A., GESTELAND, R., WALTERS, L. *et al.* New Goals for the U.S. Human Genome Project: 1998-2003. *Science*, v.282, p.682-689, 1998.
- DAVIS, M.J., WHITECOMB, R.F., GILLASPIE, A.G.Jr. Fastidious bacteria of plant vascular tissue and invertebrates (including so-called rickettsia-like bacteria). In: **The Prokaryotes: A handbook on habitats, isolation, and identification of Bacteria**, ed. M.P. Starr, H.O. Stolp, H.G. Truper, A. Balows, A. Schlegel. pp.2172-88. New York/ Heidelberg/ Berlin: Springer-Verlag. Apud: HOPKINS, D.L. *Xylella fastidiosa*: Xylem-limited bacterial pathogen of plants. *Ann. Rev. Phytopathol.*, v.27, p.271-90, 1989.
- DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. cap.The bacterium *Xylella fastidiosa*, p.22-35.
- DUNHAM, I., *et al.* The DNA sequence of chromosome 22. *Nature*, v.402, p.489-495, 1999.
- FERREIRA, A.J.S. Sequenciamento de parte do genoma de *Xylella fastidiosa* e análise do metabolismo de carboidratos no genoma anotado. São Paulo, 2000. 116p. (Dissertação de Mestrado- Instituto de Química- Universidade de São Paulo).
- FLEICHMANN, R.D., *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, v.269, p.496-512, 1995.
- FRANGEUL, L., NELSON, K.E., BUCHRIESER, C., DANCHIN, A., GLASER, P., KUNST, F. Cloning and assembly strategies in microbial genome projects. *Microbiology*, v.145, p.2625-2634, 1999.
- FRASER, C.M., FLEISCHMANN, R.D. Strategies for whole microbial genome sequencing and analysis. *Electrophoresis*, v.18, n.8, p.1207-1216, 1997.
- GARDNER, M.J., *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science*, v.282, p.1126-32, 1998.
- GENOME CONSORTIUM. *Arabidopsis thaliana* chromosome IV complete sequence. *Nature*, v.402, p.769-777, 1999.
- GOFFEAU, A. *et al.* Life with 6.000 genes. *Science*, v.274, p.563-567, 1996.
- GOFFEAU, A. *et al.* The yeast genome directory. *Nature*, v.387, suppl., p. 5-105, 1997.
- GONZALES, M.A.Os cinco anos de convivência com a clorose variegada dos citros. Datilografado, 1992. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.

- GOHEEN, A.C., NYLAND, G., LOWE, S. K. Association of Rickettsia-like organism with Pierce's disease for grapevine and alfalfa dwarf and heat therapy of the disease in grapevines. *Phytopathology*, v.63, p.341-345, 1973.
- HANKE, J., HOHEISEL, D. Construction and Screening of Cosmid Libraries. 1998. Não publicado.
- HATTORI, *et al.* The DNA sequence of human chromosome 21. *Nature*, v.405, p.311-319, 2000.
- HEWITT, W.B. A transmissible disease of grapevines. *Phytopathology*, v.29, p.10, 1939. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.
- HOLMES, D.S., QUIGLEY, M. A rapid boiling method for the preparation of bacterial plasmids. *Annal. Biochemistry*, v.114, p.193-197, 1981.
- HOPKINS, D.L. *Xylella fastidiosa*: xylem-limited bacterial pathogen of plants. *Annu. Rev. Phytopathology*, v.27, p.271-290, 1989.
- HUANG, X. An improved sequence assembly program. *Genomics*, v.33, p.21-31, 1996.
- JORDAN, C.N.B., PASSOS, G.A.S. Projeto transcriptoma. *Biotechnologia*, v.12, p.34-37, 2000.
- KAMPER, S.M., FRENCH, W.J., DEKLOET, S.R. Genetic Relationships of some fastidious xylem-limited bacteria. *International Journal Systematic Bacteriology*, v.35, p.185-188, 1985.
- KOONIN, E.V., TATUSOV, R.L., GALPERIN, M.Y. Beyond complete genomes: from sequence to structure and function. *Current Opinion in Structural Biology*, v.8, p.355-363, 1998.
- KUNST, F., *et al.* The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature*, v.390, p.249-256, 1997.
- LARANJEIRA, F.F. *et al.* Avaliação serológica e sintomatologia da ocorrência de *Xylella fastidiosa* e clorose variegada dos citros (CVC) no banco do germoplasma ativo do IAC. *Fitopatologia Brasileira*, v.21 (suplemento), p.335, 1996. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.
- LEHNINGER, A.L. **Bioquímica**. São Paulo: Editora Edgard Blücher LTDA, 1976. v.2 Catabolismo e a produção da energia das ligações de fosfato, p.351.
- LIN, X., *et al.* Sequence and analysis of chromosome II of *Arabidopsis thaliana*. *Nature*, v.402, p.761-768, 1999.
- LOPES, J.R.S. Mecanismos de transmissão de *Xylella fastidiosa* por cigarrinhas. *Laranja*, Cordeirópolis, v.17, p.79-92, 1996. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.
- LOPES, J.R.S., BERETTA, M.J.G. A confirmação da transmissão por cigarrinhas do agente causal da clorose variegada dos citros, *Xylella fastidiosa*. *Fitopatologia Brasileira*, v.21, suplemento p.343, 1996. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.

- LOWE, T.M., EDDY, S.R. t-RNA scan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, v.25, p.955-964, 1997.
- MACHADO, M. Citrus Variegated Chlorosis (CVC), a new destructive citrus disease in Brazil, and the xylem limited bacteria, *Xylella fastidiosa*. 1987. 11p. Apud: [Online]. Available: <http://www.dcc.unicamp.br/genoma/xylella/html> [January, 2000].
- MANDELSTAM, J., Mc QUILLEN, K., DAWES, I. **Biochemistry of Bacterial Growth**. 3.ed. Boston: Blackwell Scientific Publications, 1982. 449p.
- MASSUNG, R.F., ESPOSITO, J.J., LIU, L.I., QI, J., UTTERBACK, T.R., KNIGHT, J.C., AUBIN, L., YURAN, T.E., PARSONS, J.M., LOPAREV, V.N., *et al.* *Nature* v.366, p.748-751, 1993.
- MENEGUCCI, J.L.P., PAIVA, L.V., SOUTO, R.F., CARVALHO, S.A. Alterações físico-químicas de frutos de laranheiras *Valência* com sintomas de clorose variegada dos citros. *Revista brasileira de citricultura*, v.17, p.153-155, 1995. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro:FAPESP/FUNDECITRUS, 1998. 166p.
- MILLER, J.H. **A short course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria**. Plainview, N.Y., U.S.A.: Cold Spring Harbor Laboratory Press, 1992. pt.1.
- MYERS, E. W., *et al.* A Whole Genome Assembly of Drosophila. *Science*, v.287, p.2196-2204, 2000.
- MUSHEGIAN, A.R., KOONIN, E.V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci.*, v.93, p.10268-10273, 1996.
- National Center For Biotechnology Information. [Online]. Available: <http://ncib.nlm.nih.gov> [September, 2000].
- NEIDHARDT, F.C. coord. **Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology**. Washington, D.C., U.S.A.: American Society for Microbiology, 1987. v.1.
- NEIDHARDT, F.C. coord. **Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology**. 2.ed. Versão 1.0 For windows. [CD-ROM]. Washington, D.C., U.S.A.: American Society for Microbiology, 1996. v.1.
- New England BioLabs. [Online]. Available: <http://circuit.neb.com/cgi-bin/onlineorder.pl/neb/products/nucleic/304-1.html> [September, 2000].
- PARADELA FILHO, O., SUGIMORI, M.H., RIBEIRO, I.J., *et al.* Primeira constatação em caféiro da *Xylella fastidiosa* causadora da clorose variegada dos citros, *Laranja*, 16. Cordeirópolis, 1995. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro:FAPESP/FUNDECITRUS, 1998. 166p.
- PERKIN ELMER BIOSYSTEMS. ABI PRISM Big Dye Terminator Cycle Sequencing- Ready Reaction Kits. Original and version 2.0. Protocol. U.S.A., p.1.1-5.7, 1999.
- POOLER, M.R., HARTUNG, J.S. Specific detection and identification of *Xylella fastidiosa* strains causing citrus variegated chlorosis. *Current Microbiology*, v.31, p.177-181, 1995.

- PURCELL, A.H., FINLAY, A.H. Evidence for noncirculative transmission of Pierce's disease bacterium by sharpshooter leafhoppers. *Phytopathology*, St Paul, v.69, p.393-395, 1979. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.
- PURCELL, A.H., HOPKINS, D. Fastidious xylem-limited bacteria plant pathogen. *Annu. Rev. Phytopathology*, v.34, p.131-151, 1996.
- REED, K.E., CRONAN, J.E. Lipoic acid metabolism in *Escherichia coli*: sequencing and functional characterization of the *lipA* and *lipB* genes. *J. Bacteriology*, v.175, p.1325-1336, 1993.
- ROSSETTI, V., CARVALHO, M.L.V., CHAGAS, C.M. Evidência indireta do envolvimento de insetos vetores na transmissão de clorose variegada dos citros. *Fitopatologia Brasileira*, v.18, p.226, 1993 (suplemento). Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro:FAPESP/FUNDECITRUS, 1998. 166p.
- ROSSETTI, V. Pontos básicos sobre a clorose variegada dos citros. *Laranja*, Cordeirópolis, v.15, 1994. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro:FAPESP/FUNDECITRUS, 1998. 166p.
- SALSBERG, S.L., DELCHER, A.L., KASIF, S., WHITE, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Research*, v.26, p.544-548, 1998.
- SALVA, R.A., ROBERTO, S.R., CARLOS, E.F. Situação da clorose variegada dos citros no Estado de São Paulo, *Laranja*, Cordeirópolis, v.16, p.155-164, 1995. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro:FAPESP/FUNDECITRUS, 1998. 166p.
- SAMBROOK, J., FRITSCH, E. F., MANIATIS, T. **Molecular Cloning: A Laboratory Manual**. 2.ed. Plainview, N.Y., U.S.A.: Cold Spring Harbor Laboratory Press, 1989. vol.1, cap.1, p.1.25-1.31.
- SANGER, F., AIR, G.M., BARRELL, B.G., BROWN, N.L., COUSON, A.R., FIDDES, C.A., HUTCHISON, C.A., SLOCOMBE, P.M., SMITH, M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, v.265, p.687-695, 1977a.
- SANGER, F., NIKLEN, S., COUSON, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, v.74, p.5463-5467, 1977b.
- The *C. elegans* SEQUENCING CONSORTIUM. Genome Sequence of the nematode *C. elegans*: A Platform for Investigating Biology. *Science*, v.282, p.2012-2018, 1998.
- SIMPSON, A.J.G. *et al.* The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, v. 406-13, p.151-159, 2000.
- SMITH, L.M., SANDERS, J.Z., KAISER, R.J., HUGHES, P., DODD, C., CONNEL, C.R., HEINER, C., KENT, S.B.H., HOOD, L.E. Fluorescence detection in automated DNA sequence analysis. *Nature*, v. 321, p.674-679, 1986.

- SLOTNICK, I.J., WEINFELD, H. Dihydrouracil as a growth factor for mutant strains of *Escherichia coli*. *J. Bacteriol.*, v.73, p.122–125, 1957.
- SUTTON, G., WHITE, O., ADAMS, M., KERLAVAGE, A. TIGR assembler: a new tool for assembling large shotgun sequencing projects. *Genome Sci Technol*, v.1, p.9-19, 1995.
- TATSOV, R.L., GALPERIN, M.Y., KOONIN, E.V. The COG database: a tool for genome analysis of protein function and evolution. *Nucleic Acids Research*, v.28, p.33-36, 2000.
- The Phred/Phrap/Consed System Home Page. [Online]. Available: <http://bozeman.mbt.washington.edu/> [September, 2000].
- tRNAScan-SE: tRNA detection in large-scale genome sequence. [Online]. Available: <http://www.genetics.wustl.edu/eddy/software/#trnascan> [September, 2000].
- VITTI, G.C., MARCHI, R.J., BORELLA, M.L., VAZ FILHO, D. Estudo de prováveis desequilíbrios nutricionais em pomares cítricos no município de Colina-SP. Jaboticabal: FCAV, UNESP, 1989, 12p. (mimeografado). Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro:FAPESP/FUNDECITRUS, 1998. 166p.
- WELLS, J.M., RAJU, B.C., JUNG, H.Y., WEISBURG, W.G., MANDELCO-PAUL, L., BRENNER, D.J. *Xylella fastidiosa* gen nov., sp. Nov. gramnegative, xylem limited fastidious plant bacteria related to *Xanthomonas* spp. *International Journal of Systematic Bacteriology*, Washington, v.37, p.136-143, 1987. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.
- WELLS, J.M., RAJU, B.C., MYLAND, G. Isolation, cultive and pathogeniticy of the bacterium causing phony disease of peach. *Phytopathology*, v.73, p.859-862, 1983. Apud: DONADIO, L.C., MOREIRA, C.S., coord. **Citrus Variegated Chlorosis**. Bebedouro: FAPESP/FUNDECITRUS, 1998. 166p.
- WIEMANN, S., SCHWAGER, C., RUPP, T., ZIMMERMANN, J., GROTHUES, D., SENSEN, C., ERFLE, H., BANVERI, A., ANSORGE, W. Sequencing and analysis of 51.5 kilobases on the left arm of chromosome XI from the *Saccharomyces cerevisiae* reveals 23 open reading frames including the *fas 1* gene. *Yeast*, v.9, p.1343-1348, 1993.
- WHITE, R. H. Stable isotope studies on the biosynthesis of lipoic acid. *Biochemistry*, v.19, p.15–19, 1980. Apud: NEIDHARDT, F.C. coord. ***Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology**. 2.ed. Versão 1.0 For windows. [CD-ROM]. Washington, D.C., U.S.A.: American Society for Microbiology, 1996. v.1.

V.II. ANEXOS

Neste item encontram-se:

- a) os métodos utilizados pelo laboratório para a realização do seqüenciamento a partir de cosmídeos, bem como a descrição sucinta da obtenção de plasmídeos para a confecção de bibliotecas e as reações de seqüenciamento feitas por PCR utilizando o kit Big Dye Terminator da Perkin –Elmer;
- b) o mapa físico dos cosmídeos;
- c) o mapa das ORFs anotadas e os genes anotados;
- d) as principais vias de biossíntese de aminoácidos encontradas em *Escherichia coli*, das quais basearam-se as conclusões de presença ou ausência das vias em outros microrganismos, juntamente com tabelas associadas a essas vias;
- e) o artigo do genoma de *Xylella fastidiosa* publicado na revista Nature.

Métodologia utilizada para o seqüenciamento

Os métodos descritos a seguir foram modificados com o intuito de otimizar o seqüenciamento do genoma em larga escala. Os métodos descritos são, em sua maioria, adaptações dos procedimentos descritos por Sambrook *et al.*, 1989.

Obtenção de Cosmídeos em larga escala

1) Repique de bactérias contendo os Cosmídeos Inocular em 500 mL de meio 2xTY, 1% de cultura saturada do respectivo cosmídeo. Incubar a 37 °C por 12 horas sob agitação. Centrifugar por 10 minutos a 6000 rpm. Descartar sobrenadante	5) Coletar o sobrenadante, filtrando-o com gaze.
2) Ressuspender pellet de bactérias em 50 mL da seguinte solução : 25 mM TRIS-Cl pH 8,0 50 mM glicose 10 mM EDTA	6) Precipitação Adicionar ao sobrenadante igual volume de Isoprapanol. Deixar por 60 minutos a -70 °C. Centrifugar a 5000 rpm por 15 minutos a 4 °C. Descartar sobrenadante. Adicionar ao pellet 300 mL de solução de etanol 70%. Centrifugar 5000 rpm por 15 minutos a 4 °C. Descartar sobrenadante.
3) Lise Adicionar ao pellet dissolvido 50 mL de: 0,5 % SDS 0,2 N NaOH Agitar por 1 minuto vigorosamente.	7) Secar sob vácuo por aproximadamente 1 hora.
4) Neutralização Adicionar 50 mL de Acetato de potássio pH 4,8 a 5,5 agitando vigorosamente. Centrifugar a 7000 rpm por 15 minutos	8) Ressuspender em 10 mL de TE. Solução de TE 10 mM TRIS-Cl 0,1 mM EDTA

Ultracentrifugação em gradiente de Cloreto de cézio

Dissolver 9 gramas de Cloreto de cézio em 9 mL de água Milli Q em tubo Corex. Adicionar 0,5 mL de solução de Brometo de etídio (10 mg/mL).
Adicionar 10 mL da solução de DNA (preparado por larga escala).
Centrifugar 10000 rpm por 10 minutos.
Transferir sobrenadante para tubo "quick-seal" completando o volume com óleo mineral. Selar em seladora.
Centrifugar a 60000 rpm por 16 horas em ultra-centrífuga.
Após a centrifugação, coletar a banda inferior em tubo Falcon.
Lavar por 6 vezes com igual volume de Butanol saturado com 1 M de NaCl.
Centrifugar a 4000 rpm por 20 minutos.
Dialisar "over-night" em TE trocando a solução de diálise.
Após a diálise coletar a solução de DNA e verificar sua concentração em espectrofotômetro através de sua densidade óptica em 600 nm.

Biblioteca	
<p>1)Nebulização do cosmídeo Em um nebulizador colocar a seguinte solução: 50 µg de DNA do cosmídeo 30% glicerol 1 mL qsp água Milli Q Nebulizar sob pressão de nitrogênio de 0.6 kgf/cm² por 45 segundos. Centrifugar a um pulso de 1000 rpm. Repetir o procedimento de Nebulização sob pressão por 3 vezes. Centrifugar 1000 rpm por 4 minutos . Verificar o volume coletado.</p>	<p>4)Reparo das pontas 5 µg DNA 4 µL Neb 2 Buffer (10x conc) 1,6 µL dNTP (2,5 mM) 2 µL T4 polimerase 1 µL Klenow 4 µL ATP (20 mM) 1 µL PNK 40 µL final com água MilliQ incubar 15°C 20 min</p>
<p>2)Precipitação Adicionar 10% do volume coletado de 3 M de Acetato de sódio pH 6,8. Adicionar 2 volumes de etanol. Deixar over-night a -20 °C. Centrifugar a 14000 rpm por 30 minutos. Descartar sobrenadante. Lavar com 1 ml de Etanol a 70%. Centrifugar a 14000 rpm por 15 minutos. Descartar sobrenadante.</p> <p>3)Secar sob vácuo. Ressuspender em 10 µL de 10 mM TRIS pH 8,5. Verificar a concentração do DNA através de sua densidade óptica em espectrofotômetro em 600 nm</p>	<p>5)Low Melting Gel 0,8% Agarose Low Melting 25 mL TAE 1x1,3 µL Brometo de etídio (10 mg/mL) Correr os 40 µL reparados nas seguintes condições: 150V 150 mA 400 W Após a corrida, sob UV longo cortar pedaço de gel correspondente ao peso molecular de 1000 a 3000. Pesar o bloco de gel e colocá-lo em tubo eppendorf de 1,5 mL.</p>

Extração do DNA do gel de agarose	
<p>1)Pesar o bloco de agarose Adicionar 30 µL de NaCl 1M Aquecer a 65 °C por 8 a 10 minutos . Verificar se ocorreu perfeita dissolução do gel.</p>	<p>5)Adicionar 1V de butanol saturado com 1M NaCl Centrifugar a 14000 rpm por 10 minutos. Coletar a fase de baixo verificando seu volume.</p>
<p>2)Adicionar 1V de fenol agitar em vortex 2 minutos. Centrifugar a 14000 rpm por 10 minutos. Coletar sobrenadante em novo eppendorf.</p>	<p>6)Precipitação Adicionar 1/10 do volume coletado de Acetato de sódio pH 6,8. Adicionar 2,5 volumes de etanol ou 1 volume de isopropanol. Incubar over-night a -20 °C . Após, centrifugar a 14000 rpm por 30 minutos a 4 °C. Lavar com 1 mL de Etanol 70%. Centrifugar a 14000 rpm por 15 minutos. Secar sob vácuo. Ressuspender em 20 µL TRIS 10 mM pH 8,5.</p>
<p>3)Adicionar ainda ao eppendorf original (reextração) 20 µL TE agitar em vortex 2 minutos. Centrifugar a 14000 rpm por 10 minutos. Juntar os sobrenadantes. Nova extração com 1V de fenol</p>	<p>7)Correr gel para quantificar a amostra</p>
<p>4)Coletar sobrenadante Adicionar 1 V Clorofórmio. Agitar em vortex 2 minutos. Centrifugar a 14000 rpm por 10 minutos. Coletar sobrenadante.</p>	

Ligação do DNA com vetor

1) Vetor utilizado: pUC 19 cortado com Sma I e tratado com Fosfatase alcalina por 15 minutos.

1 μ L DNA
2 μ L Buffer 10x LB
1 μ L de enzima T4 DNA Ligase
10 μ L qsp Água Milli Q
Incubar 16 horas a 16 °C.

2) Precipitação

1 μ L Acetato de sódio 3 M pH 6,8
20 μ L Etanol
Incubar por 1 hora a -70 °C
Centrifugar 30 minutos 14.000 rpm. Descartar sobrenadante.
Após, adicionar 1 mL de Etanol 80% e centrifugar 30 minutos em 14.000 rpm. Descartar sobrenadante.
Centrifugar 15 minutos a 14.000 rpm. Descartar sobrenadante.
Ressuspender em 5 μ L de água Milli Q estéril.

Eletroporação

1) Preparo de células competentes DH10B:

Distribuir em 2 erlenmeyers com capacidade de 2 litros, 500 mL de meio de cultura 2xTY.
Inocular 5 mL de cultura prévia de 24 horas (em meio 2xTY) das células bacterianas.
Incubar a 37 °C sob agitação até que sua OD_{600nm} esteja em 0,6.
Dividir 250 mL da cultura em 4 tubos de centrifuga.
Centrifugar por 30 minutos a 6000 rpm.
Ressuspender o pellet de cada tubo com 125 mL de Glicerol a 10% (esterilizado por filtração).
Juntar o conteúdo das 2 garrafas.
Centrifugar por 15 minutos a 6000 rpm.
Descartar sobrenadante.
Ressuspender em 250 mL de Glicerol 10%.
Centrifugar por 15 minutos a 6000 rpm.
Descartar sobrenadante.
Ressuspender em 2 mL de GYT (10% glicerol, 0,125% extrato de levedo e 0,25% triptona).
Distribuir em aliquotas de 100 μ L por tubo eppendorf estéril

2) Reações

Retirar células competentes DH10B do freezer -70 °C e deixar em isopor com gelo.
Preparar as seguintes reações em eppendorfs a 4 °C
40 μ L células DH10B
1 μ L da reação de ligação
Transferir as reações para cubetas de eletroporação identificadas. Manter a 4 °C.

3) Condições de da eletroporação para eletroporador Gene Pulser- Bio Rad, para cuvetas de 0,2 cm:

Voltagem de 2,5 kV
Capacitância de 50 μ F
Resistência de 100 Ω

Eletroporar.

4) Imediatamente após a eletroporação, adicionar 800 μ L de meio SOC agitando para cima e para baixo com pipeta.

Transferir para tubo de ensaio mantendo no gelo.
Incubar por uma hora a 37 °C 200 rpm.
Após, realizar plaqueamento conforme se segue:
plaquear 200 μ L por placa de Petri (100 x 150 mm)
Incubar as placas a 37 °C em estufa, "over night".

Meios de cultura para a eletroporação

Meio SOC

Adicionar 10 mM de Cloreto de Magnésio e 20 mM de Glicose.

Meio TYE

Dissolver 100 ml de TYE no microondas

Adicionar

16 μ L X-gal (200mg/ml)

16 μ L IPTG (200mg/ml)

100 μ L Carbecilina (200mg/ml)

Preparação de MINI PREPS	
Método de lise alcalina em tubos eppendorf	
<p>1) Repique de bactérias transformadas em tubos de ensaio</p> <p>Meio 2x TY</p> <p>Carbenicilina 100 µL/100mL (200mg/mL)</p> <p>Distribuir 1,5 mL de meio de cultura contendo o antibiótico por tubo de ensaio estéril.</p> <p>Picar uma colônia branca por tubo com palito de dente estéril. Deixar o palito no tubo.</p> <p>Incubar 16 horas 37 °C a 300 rpm.</p>	<p>2) Estoque em glicerol</p> <p>Retirar 100 µL de cultura e adicionar 100 µL de glicerol 40% estéril (esterilizado por filtração)</p> <p>Congelar com gelo seco e manter a - 70 °C</p>
<p>3) Transferir cultura para tubo eppendorf.</p> <p>Centrifugar 14.000 rpm por 10 minutos.</p> <p>Descartar sobrenadante.</p> <p>Adicionar 100 µL de GET-Rnase</p> <p>Ressuspender pellet em vortex</p>	<p>GET-RNase para 50 mL asp</p> <p>4,5 g dextrose</p> <p>0,625 mL TRIS-Cl 2M pH 8,0</p> <p>1,0 mL EDTA 0,5M</p> <p>0,75 mL ribonuclease A (10mg/mL)</p> <p>Adicionar primeiramente os três compostos.</p> <p>Uma vez dissolvidos, filtrar a solução.</p> <p>Adicionar a ribonuclease A. Manter a solução sob refrigeração.</p>
<p>4) Adicionar 200 µL de Tampão de lise celular.</p> <p>Inverter suavemente por 5 minutos.</p> <p>Incubar 5 minutos a temperatura ambiente.</p>	<p>Tampão de lise celular para 25 mL (feito na hora)</p> <p>1,25 mL SDS 20%</p> <p>0,5 mL NaOH 10N</p> <p>Dissolver.</p>
<p>5) Adicionar 200 µL de solução neutralizadora.</p> <p>Inverter os tubos por 3 minutos.</p> <p>Manter sob refrigeração a 4 °C por 30 minutos.</p>	<p>Solução neutralizadora para 250 mL.</p> <p>32,39 g Acetato de Potássio</p> <p>Dissolver o Acetato em pouca água</p> <p>Ajustar o pH para 4,8 com Ácido acético glacial</p> <p>Filtrar solução.</p>
<p>6) Centrifugar 20 minutos 14.000 rpm.</p> <p>Transferir 400 µL do sobrenadante para outro eppendorf identificado.</p> <p>Adicionar ao sobrenadante 300µL de Isopropanol.</p> <p>Centrifugar por 15 minutos a 14.000 rpm.</p> <p>Desprezar o sobrenadante.</p> <p>Adicionar 500 µL de Etanol 70% (feito a hora).</p> <p>Centrifugar por 15 minutos a 14.000 rpm.</p> <p>Secar sob vácuo.</p> <p>Ressuspender em 25 µL TRIS 10 mM pH 8,0</p> <p>Colocar em shaker por 5 minutos a 37 °C.</p> <p>Armazenar a 4°C ou em freezer a -20 °C.</p>	
Método em microplacas "Boiling prep protocol"	
<p>1) Meio de cultura</p> <p>100 mL de meio Terrific Broth</p> <p>100 µL de Carbenicilina (200 mg/mL)</p> <p>Distribuir 1 mL do meio com antibiótico por well da microplaca.</p> <p>Repique das bactérias</p> <p>Sob condições de esterilidade:</p> <p>Picar as colônias brancas com palitos de dentes estéreis, colocando-os nos wells.</p> <p>Deixar 10 minutos incubando a temperatura ambiente.</p> <p>Retirar os palitos de dente e lacerar a placa com selo Plate Sealer. Furar o selo em cada well.</p> <p>Incubar em shaker por 24 horas a 37 °C a 300 rpm.</p> <p>Após 24 horas de incubação, transferir 100 µL da cultura de bactéria para 100 µL de solução estéril de glicerol 40% em microplacas. Lacerar.</p> <p>Congelar as bactérias em glicerol.</p> <p>Utilizar gelo seco. Deixar a - 70°C.</p>	<p>2) Precipitação das bactérias:</p> <p>Centrifugar as placas por 8 minutos a 2700 rpm.</p> <p>Retirar o plástico.</p> <p>Após centrifugação, drenar o meio de cultura. Secar as placas invertidas em papel toalha por 5 minutos.</p> <p>O pellet de bactérias pode ser guardado a - 70 °C se não for utilizado no mesmo dia.</p> <p>Ressuspensão das bactérias:</p> <p>Adicionar ao pellet, 25 µL de água Milli Q estéril.</p> <p>Vortex até que os pellets fiquem bem dissolvidos.</p> <p>Adicionar 70 µL de solução de MW-Tween 20.</p> <p>Vortex por 30 segundos.</p> <p>Incubar a temperatura ambiente por 5 minutos.</p>

<p><u>Solução de STET-TWEEN 20</u> 5 mL Tween 20 2 mL NaCl 5 M 0,220 mL EDTA 0,5 M pH 8,0 1,6 mL Tris-Cl pH 8,0 100 mL água Milli Q qsp manter sob refrigeração (solução pode ser estocada na geladeira)</p>	<p><u>Solução de MW-TWEEN 20 (preparar no dia de sua utilização e mantê-la sob refrigeração)</u> 100 mL solução STET-TWEEN 20 50 mg lisozima (solubilizar à parte em um Eppendorf com 1 mL da solução acima) 2,4 mL RNase (10 mg/mL)</p>	<p><u>Solução RNase 10mg/mL</u> Dissolver quantidade suficiente em H2O Milli-Q e ferver por 15 min em banho-maria (deixar os tubos de vidro tampados em um béquer com água; começar a contar o tempo após o início da fervura). Aliquotar e congelar (manter a -20°C).</p>
<p>3)Aquecimento em microondas Retirar o plástico. Ferver o conteúdo dos wells no microondas (da Ana) em potência máxima. Repetir o procedimento com as placas invertidas (lados internos para fora). O tempo esperado de fervura é de 40 segundos para cada face da microplaca. Imediatamente após a fervura, retirar as placas do microondas e adicionar 300 uL de água deionizada. Pôr plástico. Vortex por 30 segundos. Deixar em banho de gelo e água por 10 minutos. Centrifugar por 30 minutos a 4000 rpm.</p>		<p>4)Coleta do DNA Transferir 50 µL do sobrenadante para uma microplaca (fundo em U). Verificar quantidade de DNA por meio de eletroforese em gel de agarose. Manter o DNA sob refrigeração (microplaca) ou a -20°C.</p>

PCR - PROTOCOLO SAVE MONEY Sistema de amplificação: Gene Amp PCR System- Perkin Elmer 2400 Gene Amp PCR System- Perkin Elmer 9600 Gene Amp PCR System- Perkin Elmer 9700 MJ Research Inc- Peltier Effect Cycling		
Ciclos de sequenciamento para plasmídeos Temperatura de 96 °C por 10 segundos; Temperatura de 50 °C por 5 segundos; Temperatura de 60 °C por 4 minutos. Total de 40 ciclos.	Ciclos de sequenciamento para cosmídeos Temperatura de 97 °C por 1 minuto 40 ciclos de: Temperatura de 96 °C por 10 segundos; Temperatura de 50 °C por 5 segundos; Temperatura de 60 °C por 4 minutos.	
1a)Preparando as reações em MicroAmp Trays para plasmídeos: 3 µL de DNA (200 a 400 ng) 1 µL primer (3,2 pmol) 2 µL Terminator Ready Reaction Mix 6 µL Buffer Save Money 20 µL qsp água Milli Q	1b)Preparando reações em Micro Amp Trays para cosmídeos: 2 µg DNA 15 pmol primer 2 µL Terminator Ready Reaction Mix 2 µL Buffer Save Money 10 µL qsp água MilliQ	Buffer Save Money (feito no dia) 200 mM Tris-Cl 1M pH 9,0 5 mM Cloreto de magnésio
2)Precipitação para 20 µL de reação: Após reação adicionar 80 µL de Isopropanol 75%. Agitar levemente. Incubar por 15 minutos a temperatura ambiente. Centrifugar por 45 minutos a 4000 rpm. Descartar sobrenadante.	3)Secagem Centrifugar os tubos invertidos em papel toalha para secagem do Isopropanol a 2 pulsos de 1000 rpm cada. Ressuspender em 2 µL de tampão Load buffer (5:1 de Formamida deionizada, 25 mM EDTA e 50 mg/mL de Blue Dextran). Aquecer a 94 °C por 2 minutos. Colocar em gelo. Aplicar no gel de sequenciamento 1 µL de cada amostra para um pente de 64 amostras.	

Gel para sequenciamento	
Uréia 18 g Acrilamida-bis-acrilamida (Page Plus) 6,6 mL Água Milli Q 24 mL	Dissolver e filtrar a solução em filtro Millipore de 0,22 micra. Colocar em uma proveta de 50 mL, 5 mL de TBE (10x). Juntar as soluções na proveta e completar o volume para 50 mL. Adicionar 250 µL de Persulfato de amônio (10%) e 35 µL de TEMED.

Classificação dos genes

adaptado segundo Riley, 1986

I. Metabolismo intermediário

A. Degradação

1. Degradação de polissacarídeos
2. Degradação de pequenas moléculas
Aminas, amino ácidos, compostos de carbono, ácidos graxos

B. Metabolismo Central Intermediário

3. Amino açúcares
4. Entner-Doudoroff
5. Gluconeogênese
6. Via do glioxalato
7. Miscelânea do metabolismo de glucose
8. Ramo não oxidativo, via das pentoses
9. Hidrólise de nucleotídeos
10. Interconversões de nucleotídeos
11. Compostos fosforados
12. Pool, conversões multipropósitos
13. Conversões de nucleotídeos
14. Metabolismo sulfúrico

C. Carbono, metabolismo energético

15. Respiração aeróbia
16. Respiração anaeróbia
17. Transporte de elétrons
18. Glicólise
19. Ramo oxidativo, via das pentoses
20. Piruvato dehidrogenase
21. Ciclo TCA (ácido tri-carboxílico)
22. Interconversão ATP-próton força motora
23. Fermentação

D. Funções regulatórias globais

II. Biossíntese de pequenas moléculas

A. Aminoácidos

1. Família do Glutamato | Assimilação de nitrogênio
Arginina, Glutamato, Glutamina, Prolina
2. Família do Aspartato e Piruvato
Alanina, Asparagina, Aspartato, Isoleucina, Valina, Leucina, Lisina, Metionina, Treonina
3. Família Glicina-Serina | Metabolismo sulfúrico
Cisteína, Glicina, Serina
4. Família dos Aminoácidos Aromáticos
Corismato, Fenilalanina, Tryptofano, Tirosina
5. Histidina

B. Nucleotídeos

1. ribonucleotídeos púricos
2. ribonucleotídeos pirimidínicos
3. 2'deoxiribonucleotídeos
4. Recuperação de nucleosídeos e nucleotídeos

- C. Açúcares e nucleotídeos
- D. Cofatores, grupos prostéticos, carregadores
 - 1. Biotina
 - 2. Ácido fólico
 - 3. Lipoato
 - 4. Molibdopterina
 - 5. Pantotenato
 - 6. Piridoxina
 - 7. Piridina nucleotídeos
 - 8. Tiamina
 - 9. Riboflavina
 - 10. Tioredoxina, Glutaredoxina, Glutathione
 - 11. Menaquinona, Ubiquinona
 - 12. Heme, Porfirina
 - 13. Proteína carreadora de carboxil biotina (BCCP)
 - 14. Cobalamina
 - 15. Enteroquelina
 - 16. Biopterina
- E. Biossíntese de ácidos graxos e fosfatídicos
- F. Poliaminas

III. Metabolismo de macromoléculas

- A. DNA
 - 1. Replicação
 - 2. Proteínas ligadoras à estrutura do DNA
 - 3. Recombinação
 - 4. Reparo
 - 5. Restrição, modificação
- B. RNA
 - 1. RNAs ribossomais e estáveis
 - 2. Proteínas ribossomais
 - 3. Ribossomos – maturação e modificação
 - 4. Aminoacil tRNA sintetases, modificação dos tRNAs
 - 5. Síntese de RNA, modificação, transcrição de DNA
 - 6. Degradação de RNA
- C. Proteínas
 - 1. Tradução e modificação
 - 2. Chaperones
 - 3. Degradação de proteínas
- D. Outras macromoléculas
 - 1. Polissacarídeos
 - 2. Fosfolipídeos

IV. Estrutura celular

- A. Componentes de membrana
- B. Saco de Mureína, peptídeoglicano
- C. Polissacarídeos de superfície, lipopolissacarídeos e antígenos
- D. Estruturas de superfície

V. Processos celulares

- A. Transporte
 - 1. Aminoácidos, aminas
 - 2. Ânions

- 3.Carboidratos, ácidos orgânicos, alcoóis
- 4.Cátions
- 5.Nucleosídeos, purinas, pirimidinas
- 6.Proteína, secreção peptídicas
- 7.Outros
- B. Divisão celular
- C. Quimotaxis e mobilidade
- D. Adaptação osmótica
- E. Morte celular

VI.Elementos genéticos móveis

- A. Funções relacionadas à fagos e prófagos
- B. Funções relacionadas à plasmídeos
- C. Funções relacionadas à transposons

VII.Patogenicidade, virulência e adaptação

- A. Avirulência
- B. Resposta hipersensiva e patogenicidade
- C. Produção de toxinas e detoxificação
- D. Degradação da parede celular do hospedeiro
- E. Exopolissacarídeos
 - 1.Proteínas de superfície
 - 2.Adaptação, condições atípicas
 - 3.Outros

VIII.Hipotéticas

- A. Proteínas conservadas hipotéticas
- B. Proteínas hipotéticas (inclui "no hits/only low score hits")

IX.ORFs com categoria indefinida

Mapa dos Cosmídeos – Cobertura total no cromossomo de *X. fastidiosa*

Legenda

- Cosmídeos
- PCRs/ plasmídeos de “shotgun”
- Fagos Lambda
- Cosmídeos sequenciados em nosso laboratório

Fonte: www.lbi.ic.unicamp.br/xf

Xylella fastidiosa

Main Chromosome Gene Map

Plasmid pXF51 Gene Map

Plasmid pXF1.3 Gene Map

- Transport
- Mobile genetic elements
- Pathogenicity and adaptation
- Conserved hypothetical
- Hypothetical
- X Frameshift/Point Mutation(FSPM)
- Intron
- tRNA (letter indicates aminoacid)

Table 3. Putative identification of *Xylella fastidiosa* genes. Gene numbers correspond to those in Fig. 1. Percentages represent per cent identities relative to the most relevant database hit. Conserved hypotheticals and per cent identities of genes with frameshift/stop codon in frame are not shown.

INTERMEDIARY METABOLISM/DEGRADATION/Degradation of small molecules		XF0187	sulfite synthesis pathway protein (cysJ/amtA)	45%	XF1535	citrinate synthase (glhA)	61%	XF443	two-component system, sensor protein (phs)	37%	
XF2395	acetyl-xylan esterase (axeA)	34%	INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/Aerobic respiration			XF1549	div, tri-olipoamide dehydrogenase (lpg)	58%	XF1473	two-component system, sensor protein (thiS)	35%
XF1250	arginine deaminase (rocF)	45%			XF1548	div, tri-olipoamide S-succinyltransferase (suR)	53%	XF625	two-component system, sensor protein (alg2)	40%	
XF1477	benzene 1,2-dioxygenase				XF1554	fumarate hydratase (fumC)	58%	XF849	two-component system, sensor protein (ntrB)	37%	
	ferredoxin protein (hcdB)	37%	XF0347	D-lactate dehydrogenase (ldhI)	38%	XF2586	iso, isolate dehydrogenase (icd)	49%	XF763	two-component system, sensor protein (colS)	34%
XF0840	beta-galactosidase (hga)	71%	XF1802	glycerol-3-phosphate dehydrogenase (lgsA)	46%	XF2700	isocitrate dehydrogenase (icd)	75%	XF546	two-component system, sensor protein (cpiS)	37%
XF0439	beta-glucosidase (bgJ)	41%	XF2766	glycerol-3-phosphate dehydrogenase (lpgJ)	54%	XF1211	malate dehydrogenase (mdh)	62%	XF747	two-component system, sensor protein (actS)	25%
XF0846	beta-mannosidase precursor	35%	XF0310	NADH-ubiquinone oxidoreductase, NQD1 subunit (nuoF)	53%	XF0942	malate quinone oxidoreductase (yoiH)	57%	XF597	two-component system, sensor protein (phoB)	36%
XF1234	carboxyphosphonate:pyruvate carboxyltransferase (ppbB)	62%	XF0314	NADH-ubiquinone oxidoreductase, NQD10 subunit (nuoJ)	45%	XF1550	succinate dehydrogenase iron-sulfur protein (sdhB)	54%	BIOSYNTHESIS OF SMALL MOLECULES/AMINO ACIDS BIOSYNTHESIS/Glutamate family/nitrogen assimilation		
XF7210	dioxygenase	48%	XF0315	NADH-ubiquinone oxidoreductase, NQD11 subunit (nuoK)	69%	XF1072	succinate dehydrogenase, membrane anchor subunit (sdhC)	38%	XF1001	acetylglutamate kinase (argB)	30%
XF1743	esterase (est)	49%	XF0317	NADH-ubiquinone oxidoreductase, NQD12 subunit (nuoE)	47%	XF1071	succinate dehydrogenase, membrane anchor subunit (sdhD)	38%	XF1003	argininosuccinate lyase (asl)	28%
XF1610	fructokinase	36%	XF0318	NADH-ubiquinone oxidoreductase, NQD13 subunit (nuoM)	50%	XF2548	succinyl-CoA synthetase, alpha subunit (sucI)	36%	XF099	argininosuccinate synthetase (argC)	36%
XF1740	glucose dehydrogenase B (yjiI)	44%	XF0319	NADH-ubiquinone oxidoreductase, NQD14 subunit (nuoN)	43%	XF2547	succinyl-CoA synthetase, beta subunit (sucC)	58%	XF1005	gamma-glutamyl phosphate reductase (proA)	51%
XF1064	glucose kinase (gik)	41%	XF0303	NADH-ubiquinone oxidoreductase, NQD17 subunit (nuoE)			INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/ATP-proton motive force interconversion		XF1004	glutamate 5-kinase	46%
XF1460	glucose kinase (gik)	35%	XF0311	NADH-ubiquinone oxidoreductase, NQD3 subunit (nuoG)	38%	XF1149	ATP synthase, A chain (atpB/atpC/paD)	46%	XF1710	glutamate synthase, alpha subunit (gluA/asbB)	53%
XF1965	haloalkane dehalogenase (lhaA)	30%	XF0308	NADH-ubiquinone oxidoreductase, NQD4 subunit (nuoD)	50%	XF1145	ATP synthase, alpha chain (atpA/uncA)	74%	XF709	glutamate synthase, beta subunit (gluD/aspB)	55%
XF2677	L-ascorbate oxidase (aao)	29%	XF0307	NADH-ubiquinone oxidoreductase, NQD5 subunit (nuoC)	63%	XF1147	ATP synthase, beta chain (atpF/uncF)	45%	XF1847	glutamine synthetase (gluA)	64%
XF1253	lipase (lipP)	46%	XF0306	NADH-ubiquinone oxidoreductase, NQD6 subunit (nuoB)	81%	XF1148	ATP synthase, delta chain (atpH/uncH)	37%	XF1002	N-acetyl-gamma-glutamyl-phosphate reductase	33%
XF0781	lipase/esterase (estA)	25%	XF0305	NADH-ubiquinone oxidoreductase, NQD7 subunit (nuoA)	53%	XF1146	ATP synthase, epsilon chain (atpE)	46%	XF918	ornithine carbamoyltransferase (argF)	36%
XF1825	NAD(P)H steroid dehydrogenase (rdh)	35%	XF0312	NADH-ubiquinone oxidoreductase, NQD8 subunit (nuoH)	56%	XF1144	ATP synthase, gamma chain (atpG/uncG/papC)	57%	XF792	pyruvate 5-carboxylate reductase (proC)	49%
XF1201	phosphoglycerate phosphatase (yfbT)	35%	XF0313	NADH-ubiquinone oxidoreductase, NQD9 subunit (nuoI)	60%		INTERMEDIARY METABOLISM/REGULATORY FUNCTIONS		XF1427	succinyl-homoserine aminotransferase (argM/astC/cstC)	48%
XF2470	phosphoglycerate phosphatase (cbzC)	37%				XF1241	aconitate hydratase 1 (aconA/acon)	...	BIOSYNTHESIS OF SMALL MOLECULES/AMINO ACIDS BIOSYNTHESIS/Aspartate family, pyruvate family		
XF2259	polyvinylalcohol dehydrogenase	22%				XF1143	ATP synthase, beta chain (atpD)	78%	XF0114	2,3,4,5-tetrahydropyridine-2-carboxylate	
XF0366	ribokinase (rbsK)	38%				XF1142	ATP synthase, C chain (atpE/uncE)	56%	XF1818	2-isopropylmalate synthase (leuA)	47%
XF0379	rubredoxin (rubA)	54%				XF1141	ATP synthase, delta chain (atpH/uncH)	37%	XF7375	2-isopropylmalate dehydratase large subunit (leuA)	63%
XF1819	threonine dehydratase catabolic (tdcB)	39%				XF1140	ATP synthase, epsilon chain (atpE)	46%	XF2374	3-isopropylmalate dehydratase small subunit (leuD)	79%
XF1181	triacylglycerol lipase precursor (lip)	56%				XF1144	ATP synthase, gamma chain (atpG/uncG/papC)	57%	XF7372	3-isopropylmalate dehydrogenase (leuB)	78%
XF0610	UDP-glucose 4-epimerase (galE)	27%					INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/Anaerobic respiration and fermentation		XF1171	10-methyltetrahydrofolate reductase (metF)	37%
XF7432	UDP-glucose-1-phosphate uridylyltransferase (gluB)	82%				XF1746	alcohol dehydrogenase (ydhK)	81%	XF7172	5-methyltetrahydrofolate:glutamate--homocysteine methyltransferase (metF)	56%
INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/Amino sugars					XF2389	alcohol dehydrogenase (ydhK)	78%	XF7621	acetolactate synthase isozyme II, large subunit (ilvG)	56%	
XF0141	glucosamine-1-fructose-6-phosphate aminotransferase (glmS)	54%	XF1746	alcohol dehydrogenase (ydhK)	81%	XF2387	alcohol dehydrogenase (ydhK)	78%	XF7473	acetyltransferase (ilvI)	56%
XF1464	glucosamine-1-fructose-6-phosphate aminotransferase	47%	XF1136	NADP-alcohol dehydrogenase	56%	XF1182	lipase modulator (act)	53%	XF7473	acetyltransferase (ilvI)	56%
XF2355	N-acetyl-beta-glucosaminidase (exo)	45%	XF1227	NADP-alcohol dehydrogenase (adh)	47%	XF1813	methanol dehydrogenase regulatory protein	53%	XF7473	acetyltransferase (ilvI)	56%
XF1465	N-acetyl-glucosamine-6-phosphate deacetylase (nagA)	39%	XF1734	NADP-alcohol dehydrogenase	56%	XF1830	nitrile hydratase activator	46%	XF7473	acetyltransferase (ilvI)	56%
INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/Enr/enr-Doudoroff					XF1843	nitrogen regulatory protein P-II (glnB)	87%	XF7473	acetyltransferase (ilvI)	56%	
XF1061	2-keto-3-deoxy-6-phosphogluconate aldolase (eda/iga/kdaA)	41%	INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/Carbon transport			XF0352	pentaphosphate guanosine-3'-pyrophosphatase (pppC)	47%	XF7473	acetyltransferase (ilvI)	56%
XF1062	6-phosphogluconate dehydratase (edd)	58%	XF2460	c-type cytochrome biogenesis membrane protein (cycK)	52%	XF2145	phosphate regulon transcriptional regulator (phoU)	43%	XF7473	acetyltransferase (ilvI)	56%
INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/Gluconogenesis					XF2459	c-type cytochrome biogenesis protein (cycJ)	39%	XF7473	acetyltransferase (ilvI)	56%	
XF0977	malate oxidoreductase (maoB)	57%	XF2462	c-type cytochrome biogenesis protein (cycL)	52%	XF0520	c-type cytochrome biogenesis protein (copper tolerance) (dsbD)	32%	XF7473	acetyltransferase (ilvI)	56%
XF1259	phosphoenolpyruvate synthase (ppsA/pps)	64%	XF2461	c-type cytochrome biogenesis protein (ironoxidation) (dsbE/csmG)	39%	XF1328	cytochrome B551 (yodB)	34%	XF7473	acetyltransferase (ilvI)	56%
INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/Non-oxidative branch, pentose pathway					XF1360	cytochrome C oxidase assembly factor (coxD)	41%	XF7473	acetyltransferase (ilvI)	56%	
XF0208	D-ribulose-5-phosphate 3-epimerase (rpe/dsd)	76%	XF1389	cytochrome C ubiquinol oxidase, subunit I (cyoB)	69%	XF1389	cytochrome C ubiquinol oxidase, subunit II (cyoA)	56%	XF7473	acetyltransferase (ilvI)	56%
XF2015	ribulose-5-phosphate isomerase A (rpiA)	57%	XF1390	cytochrome C ubiquinol oxidase, subunit II (cyoA)	56%	XF1389	cytochrome C ubiquinol oxidase, subunit III (cyoC)	63%	XF7473	acetyltransferase (ilvI)	56%
XF1936	transketolase 1 (tkiA/tki)	63%	XF1387	cytochrome C ubiquinol oxidase, subunit IV (cyoD)	41%	XF0253	electron transfer flavoprotein alpha subunit (etfA)	51%	XF7473	acetyltransferase (ilvI)	56%
INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/Nucleotide interconversions					XF0254	electron transfer flavoprotein beta subunit (etfB/etfS)	54%	XF7473	acetyltransferase (ilvI)	56%	
XF1288	CTP synthetase (pyrG)	65%	XF1298	electron transfer flavoprotein		XF1298	electron transfer flavoprotein ubiquinone oxidoreductase (etf-OU)	46%	XF7473	acetyltransferase (ilvI)	56%
XF1058	uridylyl kinase (pyrH)	54%	XF0557	electron transfer protein azurin I (azr1)	49%	XF0557	electron transfer protein azurin I (azr1)	49%	XF7473	acetyltransferase (ilvI)	56%
INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/Phosphorus compounds					XF0593	ferredoxin	56%	XF7473	acetyltransferase (ilvI)	56%	
XF0657	alkaline phosphatase (phoA)	32%	XF1964	ferredoxin [ykgJ]	47%	XF1964	ferredoxin [ykgJ]	47%	XF7473	acetyltransferase (ilvI)	56%
XF0904	ATP-binding protein (ybbZ)	55%	XF7601	ferredoxin	70%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF2590	exopolysphatase (ppxX)	42%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF2591	polyphosphate kinase (ppk)	33%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/Pool, multipurpose conversions					XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%	
XF2255	acetyl coenzyme A synthetase (acs)	65%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF1037	adenosylhomocystionase (ahcY)	65%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF0860	carbonic anhydrase (yafA)	46%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF2095	carbonic anhydrase	32%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF2249	deoxynucleoside-5-phosphate synthase (dxs)	59%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF1689	ferredoxin-NADP reductase (fpr)	55%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF2268	glycerol kinase (gikK)	58%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF0181	glycine cleavage H protein (gcvH)	48%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF0183	glycine cleavage T protein (gcvT)	55%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF1385	glycine carboxylase (gcvP)	57%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF2170	hydroxyacylglutathione hydrolase (glbB)	59%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF2171	inorganic pyrophosphatase (ppa)	45%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF1399	lactoylglutathione lyase (gluA)	63%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF0392	methionine adenosyltransferase (metK)	70%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
XF0913	topinone reductase	43%	XF0547	ferredoxin II (ydgM)	54%	XF0547	ferredoxin II (ydgM)	54%	XF7473	acetyltransferase (ilvI)	56%
INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/Sugar-nucleotide biosynthesis, conversions					XF0823	phosphoglycerate kinase (pgk)	68%	XF0823	phosphoglycerate kinase (pgk)	68%	
XF0609	GDP-mannose 4,6 dehydratase (gmd)	50%	XF1893	phosphoglycerate mutase (pgmA/pgm)	56%	XF1893	phosphoglycerate mutase (pgmA/pgm)	56%	XF0823	phosphoglycerate kinase (pgk)	68%
XF2775	nucleotide sugar epimerase (wnbF)	51%	XF0824	pyruvate kinase type II (pykA)	50%	XF0824	pyruvate kinase type II (pykA)	50%	XF0823	phosphoglycerate kinase (pgk)	68%
XF0260	phosphoglucomutase/phosphomannomutase (xanA)	84%	XF10303	triisophosphate isomerase (tppA/tpp)	48%	XF10303	triisophosphate isomerase (tppA/tpp)	48%	XF0823	phosphoglycerate kinase (pgk)	68%
XF1468	phosphomannomutase (mraA)	55%	INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/Oxidative branch, pentose pathway			XF1063	6-phosphogluconolactonase (pgl)	35%	XF0823	phosphoglycerate kinase (pgk)	68%
XF0259	phosphomannose isomerase-GDP-mannose pyrophosphorylase (xanB)	84%			XF1065	glucose-6-phosphate 1-dehydrogenase (zwf)	46%	XF0823	phosphoglycerate kinase (pgk)	68%	
XF1606	UDP-glucose dehydrogenase (ugd)	63%						XF0823	phosphoglycerate kinase (pgk)	68%	
INTERMEDIARY METABOLISM/CENTRAL INTERMEDIARY METABOLISM/Sulfur metabolism					XF0869	hydrolytic poamide acetyltransferase (pabB)	52%	XF0869	hydrolytic poamide acetyltransferase (pabB)	52%	
XF1497	3'-phosphoadenosine 5'-phosphotransferase (cysH)	57%	XF0869	hydrolytic poamide dehydrogenase (lpgA/lpg)	57%	XF0869	pyruvate dehydrogenase (aceE)	56%	XF0869	hydrolytic poamide acetyltransferase (pabB)	52%
XF1501	ATP sulfurylase, large subunit (nodJ)	55%	XF0869	pyruvate dehydrogenase (aceE)	56%	XF0869	pyruvate dehydrogenase (aceE)	56%	XF0869	hydrolytic poamide acetyltransferase (pabB)	52%
XF1500	ATP sulfurylase, small subunit (nodI)	67%						XF0869	hydrolytic poamide acetyltransferase (pabB)	52%	
XF1499	NADPH-sulfite reductase, flavoprotein subunit (cysJ)	40%	INTERMEDIARY METABOLISM/ENERGY METABOLISM, CARBON/TCA cycle			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	6-phosphofructokinase (pfkA)	31%
XF1498	NADPH-sulfite reductase, iron-sulfur protein (cysI)	52%	XF0274	6-phosphofructokinase (pfkA)	31%	XF1291	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofructokinase (pfkA)	31%	XF0274	enolase (eno)	63%	XF0274	6-phosphofructokinase (pfkA)	31%
			XF0274	6-phosphofruct							

XF2213	phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase bifunctional enzyme (hisI/hisIE)	56%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Molybdopterin	XF0466	molybdopterin biosynthesis protein (moeB)	53%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Others	XF0760	DNA mismatch repair protein MutL (mutL)	42%		
XF2215	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (hisA)	44%	XF1545	molybdopterin biosynthesis protein	42%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Others	XF1916	coenzyme F390 synthetase	25%	XF0148	DNA repair protein (radC)	41%
BIOSYNTHESIS OF SMALL MOLECULES/NUCLEOTIDES BIOSYNTHESIS/Purine ribonucleotides			BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Pantothenate			BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Others			XF1299	DNA repair protein specific for alkylated DNA (alkB)	28%	
XF0587	5'-phosphoribosyl-5-aminoimidazole synthetase (purM/purG)	59%	XF0229	3-methyl-2-oxobutanoate hydroxymethyltransferase (panB)	55%	XF1044	[3r-hydroxymyristoyl] ACP dehydrase (fabZ)	49%	XF1326	DNA-3-methyladenine glycosidase (mag1)	35%	
XF0585	5'-phosphoribosylglycinamide transformylase (purN)	43%	XF0231	aspartate 1-decarboxylase precursor (panD)	53%	XF2269	3-alpha-hydroxyxysteroid dehydrogenase	34%	XF1806	DNA-damage-inducible protein (dinJ/pcca)	---	
XF0275	adenylate kinase (ack)	47%	XF0230	pantoate-beta-alanine ligase (panC)	46%	XF1073	3-oxoacyl-[ACP] reductase	42%	XF2081	DNA-damage-inducible protein (dinJ)	57%	
XF1553	adenylosuccinate synthetase (purB)	57%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Pyridoxine			XF0671	3-oxoacyl-[ACP] reductase (fabG)	66%	XF0647	endonuclease III (nhi)	67%	
XF0455	adenylosuccinate synthetase (purA)	58%	XF0060	pyridoxal phosphate biosynthetic protein (pdxJ)	49%	XF0673	3-oxoacyl-[ACP] synthase II (fabF)	60%	XF3967	exonuclease ABC subunit B (uvrB)	83%	
XF1949	amidophosphoribosyltransferase (purF)	61%		pyridoxal phosphate biosynthetic protein (pdxA)	50%	XF1970	3-oxoacyl-[ACP] synthase III (fabF)	24%	XF3111	exonuclease ABC subunit C (uvrC)	50%	
XF1975	bifunctional purine biosynthesis protein (purH)	59%		pyridoxamine 5'-phosphate oxidase (fprA)	52%	XF0203	acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (accD/debB/uspJ)	62%	XF3164	exodeoxyribonuclease	43%	
XF2429	glutamine amidotransferase (guaA)	68%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Pyridine nucleotides			XF0672	acyl carrier protein (acpP)	72%	XF2022	exodeoxyribonuclease I (sbcB/xonA/cpeA)	43%	
XF1976	glycinamide ribonucleotide synthetase (purD)	64%	XF0939	pyridoxal phosphate biosynthetic protein (pdxA)	50%	XF0771	acyl carrier protein (acpC)	40%	XF1933	exodeoxyribonuclease II (xthA)	46%	
XF0560	GMP synthase	37%	XF1337	pyridoxamine 5'-phosphate oxidase (fprA)	52%	XF0572	beta-hydroxydecanoyl-ACP dehydratase (fabA)	63%	XF3660	exodeoxyribonuclease small subunit (xseB)	39%	
XF1503	guanylate kinase (gmk/spoR)	52%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Pyridine nucleotides			XF1639	beta-ketoacyl-[ACP] synthase II (fabF/fabI)	34%	XF1755	exodeoxyribonuclease VII large subunit (xseA)	41%	
XF2430	incisome-5'-monophosphate dehydrogenase (guaB)	64%	XF0048	nucleoside diphosphate kinase (ndk)	68%	XF1817	beta-ketoacyl-[ACP] synthase III (fabH)	62%	XF0071	formamidopyrimidine DNA glycosylase (mutM/fpg)	51%	
XF0458	nucleoside diphosphate kinase (ndk)	68%		phosphoribosyl pyrophosphate synthetase (rskA/prs)	60%	XF0048	biotin carboxyl carrier protein of acetyl-CoA carboxylase (accC/fabG)	53%	XF3170	formamidopyrimidine DNA glycosylase (mutM/fpg)	51%	
XF2644	phosphoribosyl pyrophosphate synthetase (rskA/prs)	60%		phosphoribosyl-AMP cyclohydrolase (hisA)	56%	XF0049	biotin carboxylase subunit of acetyl CoA carboxylase (accC/fabG)	76%	XF1902	holliday junction binding protein.	69%	
XF2671	phosphoribosylaminoimidazole carboxylase, ATPase subunit (purK)	50%	XF1924	L-aspartate oxidase	40%	XF1087	cardiolipin synthase	33%	XF1904	holliday junction binding protein.	42%	
XF2672	phosphoribosylaminoimidazole carboxylase, catalytic subunit (purE)	66%	XF1961	NH3-dependent NAD synthetase (nadE/adgA)	47%	XF1209	cardiolipin synthase (cls/nov)	34%	XF1905	holliday junction resolvase, endonuclease (xseA)	52%	
XF0205	phosphoribosylaminoimidazole- succinocarboxamide synthase (purC)	53%	XF1097	nicotinate phosphoribosyltransferase (pncB)	43%	XF2716	ketoreductase	29%	XF0045	transcription-repair coupling factor (trfA)	49%	
XF1423	phosphoribosylformylglycinamide synthetase (purL/purI)	50%	XF1925	nicotinate-mono-nucleotide pyrophosphorylase (nadC)	54%	XF0670	malonyl CoA-ACP transacylase (fabD)	51%	XF2692	uracil-DNA glycosylase (ung)	55%	
BIOSYNTHESIS OF SMALL MOLECULES/NUCLEOTIDES BIOSYNTHESIS/Pyrimidine ribonucleotides			XF1923	quinolinate synthetase A (nadA)	42%	XF1365	phosphatidylserine decarboxylase (psd)	44%	MACROMOLECULE METABOLISM/DNA METABOLISM/Restriction, modification			
XF2226	aspartate carbamoyltransferase (pyrB)	52%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Thiamine			BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Thiamine			XF1368	adenine-specific methylase	46%	
XF1107	carbamoyl-phosphate synthase large chain (carB/purA)	68%	XF1048	1-deoxy-D-xylulose 5-phosphate reductoisomerase (dxr)	46%	XF0144	biosynthetic arginine decarboxylase (speA)	45%	XF2297	DNA methylase	40%	
XF1106	carbamoyl-phosphate synthase small chain (carA)	68%	XF0621	phosphomethylpyrimidine kinase	45%	XF1539	S-adenosyl methionine decarboxylase	66%	XF0641	DNA methyltransferase (sfllm)	31%	
XF2439	cytidylate kinase (cmkA)	48%	XF0378	thiamin-phosphate pyrophosphorylase (thiE)	41%	XF0143	spermidine synthase (speE)	41%	XF1774	DNA methyltransferase	---	
XF0980	dihydroorotate dehydrogenase (pyrC)	30%	XF0594	thiamine biosynthesis lipoprotein ApeB precursor (acpE)	33%	MACROMOLECULE METABOLISM/DNA METABOLISM/Replication			XF0924	DNA processing chain A (smf/dprA)	37%	
XF2571	dihydroorotate dehydrogenase (pyrD)	51%	XF0783	thiamine biosynthesis protein (thiG)	57%	XF2680	ATP-dependent DNA helicase (rep)	43%	XF1935	methyltransferase	37%	
XF0153	orate phosphate ribosyl transferase (pyrE)	51%	XF1888	thiamine biosynthesis protein (thiC/thiA)	67%	XF0682	ATP-dependent helicase (yaaA)	43%	XF1988	methyltransferase (kpnIM)	33%	
XF0034	oxotidine 5'-phosphate decarboxylase (pyrF)	41%	XF0956	thiamine-mono-phosphate kinase (thiL)	44%	XF1229	ATP-dependent helicase	57%	XF1604	site-specific DNA-methyltransferase (spHm)	39%	
BIOSYNTHESIS OF SMALL MOLECULES/NUCLEOTIDES BIOSYNTHESIS/2'-Deoxyribonucleotides			BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Riboflavin			XF0001	chromosomal replication initiator (dnaA)	53%	XF2313	site-specific DNA-methyltransferase	41%	
XF0762	deoxycytidine triphosphate deaminase (ddc)	65%	XF1748	5-amino-6-4S-phosphoribosylamino-uracil reductase	21%	XF2552	DNA gyrase subunit A (gyrA)	57%	XF2724	type I restriction-modification system (hsdM)	25%	
XF1441	phosphohydrolase (fouI/II)	45%	XF0954	6,7-dimethyl-8-ribityllumazine synthase (ribH)	45%	XF0005	DNA gyrase subunit B (gyrB)	60%	XF0297	type I restriction-modification system (hsdM)	---	
XF1196	ribonucleoside-diphosphate reductase alpha chain (rndA)	44%	XF0953	GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase (ribA)	52%	XF2558	DNA ligase (ligA/ligD/nal/pdcC/lop)	46%	XF0295	type I restriction-modification system endonuclease	---	
XF1197	ribonucleoside-diphosphate reductase beta chain (rndB)	34%	XF1992	riboflavin biosynthesis protein (ribA)	36%	XF1103	DNA polymerase I (polA/rsaA)	52%	XF2721	type I restriction-modification system endonuclease (hsdR1)	20%	
XF1448	thioredoxin reductase (trxB)	70%	XF2419	riboflavin biosynthesis protein (ribF)	43%	XF0136	DNA polymerase III holoenzyme chi subunit (holC)	34%	XF2725	type I restriction-modification system endonuclease	53%	
XF0580	thymidylate kinase	39%	XF0952	riboflavin synthase alpha chain (ribE)	45%	XF0204	DNA polymerase III, alpha chain (dnaE/polC)	48%	XF0296	type I restriction-modification system endonuclease (hsdR)	47%	
BIOSYNTHESIS OF SMALL MOLECULES/NUCLEOTIDES BIOSYNTHESIS/Salvage of nucleosides and nucleotides			XF0950	riboflavin-specific deaminase (ribD/ribG)	51%	XF0002	DNA polymerase III, beta chain (dnaN)	51%	XF2722	type I restriction-modification system specificity determinant	27%	
XF2089	5'-nucleotidase	42%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Thioredoxin, glutaredoxin, glutathione			XF0676	DNA polymerase III, delta subunit (holB)	35%	XF2726	type I restriction-modification system specificity determinant	21%	
XF2150	adenosine tetraphosphatase (apaH)	44%	XF0984	gamma-glutamyltranspeptidase (ggt)	39%	XF2178	DNA polymerase III, epsilon chain (dnaE)	45%	XF2727	type I restriction-modification system specificity determinant (hsdS)	24%	
XF0150	dUTPase (dut/dnaS/soi)	53%	XF1426	glutamate-cysteine ligase precursor (gshI)	57%	XF0430	DNA primase (dnaG/dnaP/parB)	45%	XF2741	type I restriction-modification system specificity determinant (hsdS)	24%	
XF1831	formyltetrahydrofolate deformylase (purU)	47%	XF2585	glutaredoxin (grx)	53%	XF0205	DNA primase (dnaG/dnaP/parB)	45%	MACROMOLECULE METABOLISM/RNA METABOLISM/Ribosomal proteins			
XF2354	hypoxanthine-guanine phosphoribosyltransferase (hpt)	27%	XF2394	glutaredoxin-like protein	54%	XF1776	DNA topoisomerase III (topA)	48%	XF2438	30S ribosomal protein S10 (rpsJ)	67%	
XF2599	phosphodiesterase-nucleotide pyrophosphate precursor	39%	XF1956	glutathione synthetase (gshB/gsh-III)	53%	XF0149	DNA topoisomerase III (topA)	48%	XF1151	30S ribosomal protein S10 (rpsJ)	72%	
XF2353	purine nucleoside phosphorylase	45%	XF1199	thioredoxin (trxA)	34%	XF1935	glucose inhibited division protein (gidB)	45%	XF1174	30S ribosomal protein S11 (rpsK)	86%	
XF2332	thymidylate synthase (thyA)	70%	XF2174	thioredoxin (tybN)	31%	XF2106	glucose inhibited division protein A (gidA)	65%	XF2631	30S ribosomal protein S12 (rpsL)	81%	
BIOSYNTHESIS OF SMALL MOLECULES/SUGARS AND SUGAR NUCLEOTIDES BIOSYNTHESIS			XF2698	thioredoxin (trxA/tsc/fipA)	49%	XF1383	helicase, ATP dependent (hrpA)	50%	XF1173	30S ribosomal protein S13 (rpsM)	85%	
XF1297	glucanotransferase precursor	37%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Thioredoxin, glutaredoxin, glutathione			XF2157	DNA polymerase III, epsilon chain (dnaE)	45%	XF1165	30S ribosomal protein S14 (rpsN)	62%	
XF0608	mannosyltransferase (mntA)	38%	XF0661	geranyltransferase (farnesyl-diphosphate synthase) (ispA)	54%	XF0430	DNA primase (dnaG/dnaP/parB)	45%	XF0238	30S ribosomal protein S15 (rpsO/secC)	62%	
BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Biotin			XF0068	hydroxybenzoate octaprenyltransferase (ubiA/cys)	49%	XF0205	DNA primase (dnaG/dnaP/parB)	45%	XF0107	30S ribosomal protein S16 (rpsB)	65%	
XF1357	8-amino-7-oxononanoate synthase (bioF)	42%	XF1391	octaprenyl-diphosphate synthase (ispB/cel)	42%	XF1776	DNA topoisomerase III (topA)	48%	XF1161	30S ribosomal protein S17 (rpsD)	51%	
XF0189	adenosylmethionine-8-amino-7-oxononanoate amino transferase (bioB)	47%	XF1538	ubiquinone biosynthesis protein (ucoG)	29%	XF1383	helicase, ATP dependent (hrpA)	50%	XF2580	30S ribosomal protein S18 (rpsR)	70%	
XF1356	biotin biosynthesis protein (bioH/bioB)	36%	XF1833	ubiquinone biosynthesis protein (aarF)	43%	XF2689	primosomal protein N' (priA)	44%	XF1156	30S ribosomal protein S19 (rpsS)	60%	
XF0064	biotin synthase (bioB)	60%	XF1487	ubiquinone/menaquinone transferase (ubiE)	58%	XF0361	replicative DNA helicase (dnaB/groP/grpA)	53%	XF2421	30S ribosomal protein S20 (rpsT)	54%	
XF2098	biotin synthesis protein (bioC)	36%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Heme, porphyrin			XF1127	TldD protein (tldD)	61%	XF0434	30S ribosomal protein S21 (rpsU)	67%	
XF0356	cytochrome P-450 hydroxylase	37%	XF0017	coproporphyrinogen III oxidase, aerobic (hemF)	63%	XF1392	single-stranded DNA binding protein (ssb)	56%	XF1157	30S ribosomal protein S22 (rpsV)	63%	
XF0377	cytochrome P450-like enzyme (bioI)	37%		hydroxylase (ubiH/visB)	34%	XF1558	single-stranded DNA binding protein (ssb)	35%	XF1169	30S ribosomal protein S23 (rpsW)	71%	
XF2477	dethiolbiotin synthetase (bioD)	44%		3-demethylubiquinone-9-3-methyltransferase (ubiG/purK)	49%	XF1644	single-stranded DNA binding protein (ssb)	54%	XF2630	30S ribosomal protein S24 (rpsX)	67%	
BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Folic acid			XF0635	2-octaprenyl-6-methoxyphenol hydroxylase (ubiH/visB)	34%	XF1778	single-stranded DNA binding protein (ssb)	25%	XF1167	30S ribosomal protein S25 (rpsY)	60%	
XF0278	2-amino-4-hydroxy-6- hydroxymethyl-dihydropteridine pyrophosphokinase (folK)	50%	XF2471	3-demethylubiquinone-9-3-methyltransferase (ubiG/purK)	49%	MACROMOLECULE METABOLISM/DNA METABOLISM/Recombination			XF1166	30S ribosomal protein S26 (rpsZ)	56%	
XF1458	2-amino-4-hydroxy-6- hydroxymethyl-dihydropteridine pyrophosphokinase (folK)	35%	XF0661	geranyltransferase (farnesyl-diphosphate synthase) (ispA)	54%	XF2558	chromosome segregation protein (smc)	27%	XF1536	30S ribosomal protein S27 (rpsL)	80%	
XF2431	bifunctional methylentetrahydrofolate dehydrogenase/methylenetetrahydrofolate cyclohydrolase (folD)	58%	XF0068	hydroxybenzoate octaprenyltransferase (ubiA/cys)	49%	XF0446	DNA-binding protein (tshB)	33%	XF1177	30S ribosomal protein S28 (rpsM)	69%	
XF2331	dihydrofolate reductase type III	43%	XF1391	octaprenyl-diphosphate synthase (ispB/cel)	42%	XF1939	DNA-binding protein (tsh)	58%	XF1168	30S ribosomal protein S29 (rpsN)	53%	
XF0436	dihydroorotidylase (folB)	50%	XF1538	ubiquinone biosynthesis protein (ucoG)	29%	XF1910	histone-like protein (tshB/tshU)	64%	XF0110	30S ribosomal protein S30 (rpsP)	53%	
XF0091	dihydroorotidylase (folB)	52%	XF1833	ubiquinone biosynthesis protein (aarF)	43%	XF1943	histone-like protein	54%	XF1155	30S ribosomal protein S31 (rpsQ)	65%	
XF1946	folylpolyglutamate synthase/dihydrofolate synthase (folC/dedC)	44%	XF1487	ubiquinone/menaquinone transferase (ubiE)	58%	XF0482	single-stranded DNA binding protein (ssb)	49%	XF1156	30S ribosomal protein S32 (rpsR)	63%	
XF1983	GTP cyclohydrolase I (folE)	52%	XF1332	uroporphyrinogen decarboxylase (hemE)	54%	XF1392	single-stranded DNA binding protein (ssb)	56%	XF1154	30S ribosomal protein S33 (rpsW)	52%	
BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Biotin			XF1799	uroporphyrinogen-III synthase (hemD)	27%	XF1558	single-stranded DNA binding protein (ssb)	35%	XF1163	30S ribosomal protein S34 (rpsX)	48%	
XF1270	lipote biosynthesis protein B (lipB)	55%	BIOSYNTHESIS OF SMALL MOLECULES/COFACTORS, PROSTHETIC GROUPS, CARRIERS BIOSYNTHESIS/Cobalamin			XF2437	integration host factor, beta subunit (himD)	72%	XF2423	30S ribosomal protein S35 (rpsY)	64%	
XF1269	lipic acid synthetase (lipA/lip)	60%	XF1886	phosphoglycerate mutase (pgmA)	32%	XF1809	recombination protein (recR)	5				

MACROMOLECULE METABOLISM/RNA METABOLISM/Ribosomes - maturation and modification			XF1940	peptide methionine sulfoxide reductase (msrA/msr)	59%	XF1045	outer membrane antigen (oma)	37%	XF2538	fimbrial assembly protein (pilC)	56%
			XF2642	peptidyl rRNA hydrolase (pth)	49%	XF0384	outer membrane hemin receptor (phuR)	22%	XF7539	fimbrial protein	48%
XF0108	16S rRNA processing protein (rimM)	36%	XF0644	peptidyl-prolyl cis-trans isomerase (mip)	38%	XF0343	outer membrane protein (nopB)	31%	XF7542	fimbrial protein	53%
XF2607	ribonuclease E (rne/msr/hmp1)	43%	XF0652	peptidyl-prolyl cis-trans isomerase (shyD)	43%	XF0873	outer membrane protein (ompW)	30%	XF0083	fimbrial subunit precursor	42%
XF1125	ribonuclease G (lca/mg)	53%	XF0838	peptidyl-prolyl cis-trans isomerase (lsuA)	32%	XF1053	outer membrane protein (ompP1)	28%	XF0487	fimbriin	37%
XF0939	ribosomal large subunit pseudouridine synthase D (hld/sfbH)	54%	XF1191	peptidyl-prolyl cis-trans isomerase (ppdH)	25%	XF1123	outer membrane protein	34%	XF0539	fimbriin (fimA)	38%
XF2201	ribosomal protein L11 methyltransferase (prmA)	46%	XF1212	peptidyl-prolyl cis-trans isomerase (ppdG/ppi)	60%	XF1739	outer membrane protein (fomA)	43%	XF1791	fimbriin	48%
XF2532	ribosomal protein S6 modification protein (rimK)	---	XF1605	peptidyl-prolyl cis-trans isomerase	36%	XF2345	outer membrane protein (smpA)	38%	XF0081	outer membrane usher protein precursor (fimO)	36%
XF1200	ribosomal small subunit pseudouridine synthase (rsuA)	42%	XF0442	phosphatidyltransferase (ptr)	37%	XF1024	outer membrane protein H.8 precursor	42%	XF1953	plus biogenesis protein (pilJ)	46%
XF0236	ribosomal-binding factor A (ribA)	35%	XF0926	polypeptide deformylase (dsf/fms)	55%	XF1811	outer membrane protein Slp precursor (slp)	31%	XF1954	plus biogenesis protein (pilI)	33%
XF0441	ribosomal-protein-alanine acetyltransferase (rimI)	38%	XF2156	protein phosphatase (lyoA)	35%	XF1547	peptidoglycan-associated outer membrane lipoprotein precursor (pcp/opp)	29%	XF2544	plus biogenesis protein (pilB)	55%
XF2358	rRNA/rRNA methyltransferase (lycA)	57%	XF1446	protein transferase (aet)	51%	XF0033	PilE protein (pilE)	37%	XF1955	plus protein (pilG)	82%
XF1985	rRNA/rRNA methyltransferase (yihH)	50%	XF2585	protein-L-isoaspartate O-methyltransferase (pcm)	34%	XF0975	polyphosphate-selective porin O (oprO)	23%	XF0031	PilX protein (pilX)	28%
MACROMOLECULE METABOLISM/RNA METABOLISM/Aminoacyl tRNA synthetases, tRNA modification			XF1051	ribosome recycling factor (rrf)	58%	XF0321	porin O precursor (oprO)	23%	XF0032	PilY1 gene product (pilY1)	32%
XF0124	alanine tRNA synthetase (alaS/fobB)	56%	XF2244	signal peptidase I (lepB)	40%	XF0028	pre-pilin like leader sequence (fimT)	37%	XF1274	PilY1 gene product (pilY1)	30%
XF0147	arginyl tRNA synthetase (argS)	40%	XF1437	thiol-disulfide interchange protein (dsbA)	39%	XF1363	soluble lytic murein transglycosylase precursor (sls/lytV)	25%	XF2537	pre-pilin leader peptidase (xpsO)	76%
XF2563	asparaginyl-tRNA synthetase (asnS/ass)	66%	XF0353	translation initiation inhibitor	54%	CELL STRUCTURE/MUREIN SACculus, PEPTIDOGLYCAN					
XF1856	aspartyl tRNA synthetase (aspS)	53%	MACROMOLECULE METABOLISM/PROTEIN METABOLISM/Chaperones			XF0852	alanine racemase (alr)	45%	XF1632	twitching motility protein (pilU)	60%
XF0995	cysteinyl tRNA synthetase (cysS)	46%	XF0616	10kDa chaperonin (groES/rtpA)	58%	XF0799	D-alanine-2-D-alanine ligase B (ddlB/ddl)	51%	XF1633	twitching motility protein (pilT)	74%
XF1338	glutamyl-tRNA synthetase (glnS)	54%	XF0515	60kDa chaperonin (nopA/groEL)	77%	XF0855	lipoprotein (nopD/ppb)	39%	XF6777	type 4 fimbriae assembly protein (pilZ)	62%
XF0822	glutamyl tRNA synthetase (glnX)	54%	XF1452	chaperone (loxA)	41%	XF0416	lipoprotein precursor (lacL)	36%	XF0479	type IV pilin (pilE)	35%
XF1960	glycyl tRNA synthetase alpha chain (glyY/glySIA)	74%	XF2233	DnaJ protein (dnaJ)	32%	XF1715	monofunctional biosynthetic peptidoglycan transglycosylase (mtgA)	50%	CELLULAR PROCESSES/TRANSPORT/Amino acids, amines		
XF1959	glycyl tRNA synthetase beta chain (glyS/glySIB)	39%	XF2340	DnaK protein (dnaK/grpP/grp/seq)	72%	XF0759	N-acetylmuramoyl-L-alanine amidase precursor (lamC)	45%	XF0408	amino acid transporter (yhdG)	43%
XF2222	histidyl tRNA synthetase (hisS)	36%	XF0991	DnaK suppressor	46%	XF2646	outer membrane lipoprotein precursor (loB/hemM)	29%	XF2730	amino acid transporter (rhtC)	28%
XF2418	isoleucyl tRNA synthetase (ileS/ivS)	54%	XF0978	heat shock protein G (hspG)	58%	CELLULAR PROCESSES/TRANSPORT/Anions			XF2707	cationic amino acid transporter	43%
XF2176	leucyl tRNA synthetase (leuS)	58%	XF2341	heat shock protein Gtp (gtpE)	38%	XF1896	outer membrane protein P6 precursor (pa/ompP6)	37%	XF2708	cationic amino acid transporter	41%
XF1112	lysyl tRNA synthetase (lysU)	56%	XF1186	peptidyl-prolyl cis-trans isomerase (tig)	32%	XF1814	penicillin binding protein (pbpA)	39%	XF2142	ABC transporter phosphate permease (pstC/phoW)	50%
XF2555	lysyl tRNA synthetase (yieA/genX)	47%	MACROMOLECULE METABOLISM/PROTEIN METABOLISM/Protein degradation			XF1614	penicillin-binding protein 1A (mrca/ponA)	39%	XF2143	ABC transporter phosphate permease (pstA/phoT)	53%
XF0927	methionyl tRNA formyltransferase (fmt)	56%	XF2260	alanine dipeptidyl peptidase	30%	XF0368	penicillin-binding protein 1B (ponB)	41%	XF1344	ABC transporter sulfate binding protein (sbp)	64%
XF0549	methionyl tRNA synthetase (metG)	54%	XF0138	aminopeptidase A (pepA/xerB/capD)	47%	XF0884	penicillin-binding protein 2 (mrda/pbp2)	38%	XF1345	ABC transporter sulfate permease (cysJ/lytJ)	50%
XF0742	phenylalanyl tRNA synthetase alpha chain (pheT)	42%	XF1488	aminopeptidase N	36%	XF0795	phospho-N-acetylmuramoyl-pentapeptide-transferase (mraY/murX)	59%	XF1346	ABC transporter sulfate permease (cysW)	49%
XF0741	phenylalanyl tRNA synthetase alpha chain (pheS)	59%	XF7009	aminopeptidase P (pepP)	46%	XF2185	rare lipoprotein A (rlpA)	30%	XF0412	nitrate ABC transporter ATP-binding protein (nrtA)	46%
XF0445	prolyl tRNA synthetase (proS/drpA)	55%	XF1188	ATP-dependent Ctp protease ATP binding subunit Ctpx (clpX/capC)	73%	XF1305	rod shape-determining protein (mrse/bvz/rodY)	77%	XF2144	phosphate ABC transporter ATP-binding protein (pstB/phoT)	65%
XF0751	ribonuclease D (rnd)	31%	XF1187	ATP-dependent Ctp protease proteolytic subunit (clpP/lofP)	73%	XF1311	rod shape-determining protein (mrseD)	32%	XF1347	sulfate ABC transporter ATP-binding protein (cysA)	61%
XF2781	ribonuclease P (rnp)	37%	XF0381	ATP-dependent Ctp protease subunit (clpB/rnpM)	69%	XF0907	soluble lytic murein transglycosylase precursor (lytJ)	42%	CELLULAR PROCESSES/TRANSPORT/Carbohydrates, organic acids, alcohols		
XF1314	S-adenosylmethionine: tRNA ribosyltransferase-isomerase (queA)	52%	XF1443	ATP-dependent Ctp protease subunit (clpA/lofD)	61%	XF0798	UDP-N-acetylmuramoyl-L-alanine ligase (murG)	47%	XF2446	ABC transporter sugar permease (lacF)	40%
XF2286	seryl tRNA synthetase (serS)	59%	XF1189	ATP-dependent serine proteinase La (lon/capP/dcg/muc/lofA)	64%	XF0276	UDP-N-acetylmuramoyl-L-alanine ligase (mpl)	50%	XF2447	ABC transporter sugar permease (lacF)	37%
XF0736	threonyl tRNA synthetase (thrS)	59%	XF2704	carboxyl-terminal protease (ctpaA)	41%	XF0278	UDP-N-acetylmuramoyl-L-alanine ligase (mpl)	50%	XF2448	ABC transporter sugar-binding protein (malE)	24%
XF0109	tRNA (guanine-N1)-methyltransferase (trmD)	51%	XF1282	carboxypeptidase related protein	33%	XF1118	UDP-N-acetylmuramoylalanine--D-glutamate ligase (murD)	29%	XF0887	alpha-ketoglutarate permease symporter (kgp/mxaA)	64%
XF1362	tRNA adenylyltransferase (lca)	52%	XF0156	cystine protease (cprB)	28%	XF0793	UDP-N-acetylmuramoylalanine--D-glutamate--2,6-diaminopimelate ligase (murE)	43%	XF0976	D-glucose/galactose transport protein (dctA)	66%
XF0090	tRNA delta(2)-isopentenylpyrophosphate transferase (miaA/trpX)	49%	XF0015	dipeptidyl-peptidase	28%	XF0794	UDP-N-acetylmuramoylalanine--D-glutanyl-L-2-diaminopimelate--D-alanyl-D-alanyl-L-glutamate ligase (murF/mra)	41%	XF1809	glucose/galactose transporter (gluP)	31%
XF1440	tRNA methyltransferase (trmU/assU)	55%	XF1484	heat shock protein (hsfV/hspD)	65%	XF2572	UDP-N-acetylpyruvoylglucosamine reductase (murB)	43%	XF2267	glycerol uptake facilitator protein (glpF)	39%
XF1373	tRNA pseudouridine synthase A (truA/hsfA/leuK)	52%	XF0452	integral membrane proteinase (hflK/hflA)	33%	XF1050	undecaprenyl pyrophosphate synthetase (lppS/rhl)	51%	XF1406	H ₂ O kinase/phosphatase (pskK)	36%
XF0237	tRNA pseudouridine synthase B (truB)	42%	XF0453	integral membrane proteinase (hflC)	36%	CELL STRUCTURE/SURFACE POLYSACCHARIDES, LIPOPOLYSACCHARIDES, AND ANTIGENS			XF0320	Mg ²⁺ /citrate complex transporter (citN/citH)	60%
XF2475	tRNA/rRNA methyltransferase	41%	XF0280	leucine aminopeptidase (leuA)	34%	XF1289	2-dehydro-3-deoxyphosphocitronate aldolase (kdsA)	44%	XF1402	phosphotransferase system enzyme I (phbI)	40%
XF0428	tryptophanyl tRNA synthetase	34%	XF0435	o-sialoglycoprotein endopeptidase (gcpI)	63%	XF0105	3-deoxy-D-manno-octulosonic acid transferase (kdsA/waaA)	44%	XF1403	phosphotransferase system HPr enzyme (phbH)	52%
XF0169	tyrosyl tRNA synthetase (tyrS)	53%	XF0127	oligopeptidase A (pilC/opa/opaA)	50%	XF2299	3-deoxy-manno-octulosonate cytidyltransferase (kdsB)	50%	XF1067	sugar ABC transporter ATP-binding protein	53%
XF0134	valyl tRNA synthetase (valS)	51%	XF1479	peptidase (ptrB/tip)	44%	XF1419	acetyltransferase (lpxD/tirA/omsA)	26%	CELLULAR PROCESSES/TRANSPORT/Cations		
MACROMOLECULE METABOLISM/RNA METABOLISM/RNA synthesis, modification, DNA transcription			XF1544	peptidyl-di-peptidase (lcp)	53%	XF0812	dolichol-phosphate mannosyltransferase (dmT)	24%	XF2328	ABC transporter sodium permease (natB)	27%
XF0192	ATP-dependent RNA helicase (hflE)	52%	XF0220	proline dipeptidase (pepD)	33%	XF1838	dolichyl-phosphate mannosase synthase related protein	34%	XF1844	ammonium transporter (amtB)	57%
XF0252	ATP4-dependent RNA helicase (deaD)	55%	XF1510	proline imino-peptidase (pip/kap)	22%	XF0257	dTPD-4-dehydrothiamine 3,5-epimerase (rtdD)	75%	XF0395	bacterioferritin (bfr)	61%
XF2696	ATP-dependent RNA helicase (hflB/nrmA)	52%	XF0667	serine protease (pspB)	28%	XF0258	dTPD-glucose 4,6-dehydratase (rftB)	76%	XF2140	cation:proton antiporter (ybaL)	57%
XF0234	N utilization substance protein A (nusA)	56%	XF1026	serine protease (pspB)	27%	XF0811	dTPD-glucose 4,6-dehydratase (rftB)	53%	XF2134	ferric enterobactin receptor (ftrA)	22%
XF0227	polynucleotide adenyltransferase (pcnB)	48%	XF1851	serine protease	25%	XF1637	glycosyl transferase (lgsD)	29%	XF0932	ferric iron transport protein	47%
XF2606	pseudouridylylase synthase (lucC)	48%	XF1823	tail-specific protease (prc/tpa)	39%	XF1082	lipid A 4'-kinase (lpxK)	38%	XF0933	ferric iron transport protein B (feoB)	42%
XF1176	RNA polymerase alpha subunit (rpoA)	87%	XF0816	zinc protease	40%	XF0104	lipid A biosynthesis lauroyl acyltransferase (htrB/waaM)	40%	XF1426	ion transporter	38%
XF2633	RNA polymerase beta subunit (rpoB/grnO/nitB/rifron)	67%	MACROMOLECULE METABOLISM/OTHER MACROMOLECULES METABOLISM/Polysaccharides			XF1348	lipid A biosynthesis lauroyl acyltransferase (htrB/waaM)	26%	XF0900	magnesium and cobalt transport protein (lcaA)	46%
XF2632	RNA polymerase beta' subunit (rpoC/tabB)	71%	XF2714	alpha-L-fucosidase (fucA1)	36%	XF1042	lipid A disaccharide synthase (lpxB/pgsB)	26%	XF1015	manganese transport protein (mntH2)	64%
XF1502	RNA polymerase omega subunit (rpoZ)	43%	XF2278	dolichol-phosphate mannosyltransferase (dpm1)	24%	XF0879	lipopolysaccharide biosynthesis protein (ribU)	26%	XF1401	Mg ²⁺ transporter (mgpE)	26%
XF2638	transcription antitermination factor (nusG)	60%	XF0887	mannosyltransferase (mttA)	31%	XF2434	lipopolysaccharide core biosynthesis protein	37%	XF1398	Na ⁺ /H ⁺ exchange protein	24%
XF0955	transcription termination factor (nusB/ssyB)	49%	MACROMOLECULE METABOLISM/OTHER MACROMOLECULES METABOLISM/Phospholipids			XF0980	lipopolysaccharide synthesis enzyme (kdsB)	56%	XF2019	Na ⁺ /H ⁺ antiporter (yjcE)	42%
XF2699	transcription termination factor Rho (rho/ntrA/psuA/mscTsu)	72%	XF2330	proteinase (slpB)	22%	XF0778	O-antigen acetylase (oatA)	26%	XF0324	periplasmic iron-binding protein (afuA)	24%
XF1108	transcriptional elongation factor (greA)	62%	XF1026	serine protease (pspB)	27%	XF1413	polysialic acid capsule expression protein (kpsF)	45%	XF0921	polar amino acid transporter (yxeX)	46%
MACROMOLECULE METABOLISM/RNA METABOLISM/RNA degradation			XF1851	serine protease	25%	XF2154	saccharide biosynthesis regulatory protein (lpsX)	73%	XF1903	potassium uptake protein (kup/trkD)	48%
XF1257	oligoribonuclease (orn)	58%	XF1823	tail-specific protease (prc/tpa)	39%	XF0176	sugar transferase	43%	XF2329	sodium ABC transporter ATP-binding protein (natA)	47%
XF0239	polynucleotide phosphorylase (ppn)	64%	XF0816	zinc protease	40%	XF1045	UDP-3-O-(4-hydroxymyristoyl)-glucosamine N-acyltransferase (lpxD/tirA/asc)	41%	XF0599	TonB-dependent receptor for iron transport (ybaL)	47%
XF2615	ribonuclease	44%	CELL STRUCTURE/MEMBRANE COMPONENTS/Inner membrane			XF1646	UDP-3-O-(4-hydroxymyristoyl)-glucosamine N-acyltransferase (lpxD/tirA)	28%	XF1496	TonB-dependent receptor for iron transport (ybaL)	47%
XF2158	ribonuclease H (rnhA/rnh)	60%	XF2780	60kDa inner-membrane protein	40%	XF0486	UDP-3-O-(3-hydroxymyristoyl)-glucosamine N-acyltransferase (lpxD)	31%	XF0367	voltage-gated potassium channel beta subunit	50%
XF1041	ribonuclease III (rnhB)	55%	XF1640	ankyrin-like protein (ankZ)	26%	XF0903	UDP-3-O-(3-hydroxymyristoyl)-N-acetylglucosamine deacetylase (lpxC)	58%	CELLULAR PROCESSES/TRANSPORT/Protein, peptide secretion		
XF2246	ribonuclease III (rnc)	50%	XF2235	functional penicillin-binding protein 1C precursor (pbpC)	47%	XF1042	lipid A disaccharide synthase (lpxB/pgsB)	26%	XF2455	heme ABC transporter ATP-binding protein (lcaA)	41%
XF1505	ribonuclease PH (rph)	63%	XF0082	chaperone protein precursor (ecpD)	40%	XF0879	lipopolysaccharide biosynthesis protein (ribU)	26%	XF2456	heme ABC transporter membrane protein (cmhB)	50%
XF2145	ribonuclease T (rnt)	47%	XF0951	O-amino acid dehydrogenase subunit (dadA/dadR)	54%	XF2434	lipopolysaccharide core biosynthesis protein	37%	XF2457	heme ABC transporter membrane protein (cmcC)	48%
MACROMOLECULE METABOLISM/PROTEIN METABOLISM/Translation and modification			XF2334	diacylglycerol kinase (dgkA)	54%	XF0778	O-antigen acetylase (oatA)	26%	XF2261	oligopeptide transporter	40%
XF1018	arginine-tRNA-protein transferase (ate1)	25%	XF0340	disulfide bond formation protein B (dsbB)	33%	XF1415	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (murA/murZ)	62%	XF0806	preprotein translocase SecA subunit (secA/prlD/azp/pea)	60%
XF1436	disulfide oxidoreductase (dsbA)	38%	XF2433	epimerase/dehydratase protein (capD)	37%	XF1043	UDP-N-acetylglucosamine acyltransferase (lpxA)	44%	XF1172	preprotein translocase SecY subunit (secY/prlA)	55%
XF1834	disulfide isomerase	30%	XF0256	glucose-1-phosphate thymidyltransferase (rftA)	78%	CELL STRUCTURE/SURFACE STRUCTURES			XF2639	preprotein translocase subunit (secE)	33%
XF2629	elongation factor G (ltaA)	70%	XF0375	inner membrane protein (yjdB)	36%	XF0077	fimbrial adhesin precursor (mrkD)	30%	XF0224	preprotein translocase YaeC subunit (yaeC)	36%
XF2203	elongation factor P (ltaP)	39%	XF1033	membrane protein	25%	XF0078	fimbrial adhesin precursor (mrkD)	32%	XF2685	protease IV (sspA)	40%
XF2473	elongation factor P (ltaP)	63%	XF0764	membrane protein	38%	XF0080	fimbrial adhesin precursor (fima/pilA)	29%	XF0225	protein-export membrane protein (secD)	44%
XF2579	elongation factor Ts (tsf)	50%	XF0777	membrane protein (actII-3)	22%	XF0368	fimbrial assembly membrane protein (pilM)	61%	XF0226	protein-export membrane protein (secF)	39%
XF2628	elongation factor Tu (tufA)	82%	XF2230	penicillin-binding protein 6 precursor (dacC)	47%	XF0486	UDP-3-O-(3-hydroxymyristoyl)-glucosamine N-acyltransferase (lpxD)	28%	XF0304	protein-export membrane protein (secG)	43%
XF2640	elongation factor Tu (tufB)	82%	XF0151	phosphomannomutase (algC)	45%	XF0803	UDP-3-O-(3-hydroxymyristoyl)-N-acetylglucosamine deacetylase (lpxC)	58%	XF1801	protein-export protein (secB)	43%
XF2553	initiation factor elf-2B, alpha subunit-related	50%	XF1183	polysaccharide biosynthetic protein (yjaA/yjaB)	52%	XF1415	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (murA/murZ)	62%	XF0562	sec-independent protein (secE)	47%
XF1445	initiation factor IF-1 (finA)	73%	XF2333	prolipoprotein diacylglycerol transferase (lgt/lmpA)	48%	XF1043	UDP-N-acetylglucosamine acyltransferase (lpxA)	44%	XF0073	signal recognition particle protein (fhl)	63%
XF0235	initiation factor IF-2 (finB)	49%	XF1310	rod shape-determining protein (mrseC)	37%	CELL STRUCTURE/MEMBRANE COMPONENTS/Outer membrane constituents					
XF0737	initiation factor IF-3 (finC)	60%	XF1313	rod shape-determining protein (mrseD)	48%	XF2292	autolytic lysozyme (lyc)	30%			
XF0857	L-isoaspartate O-methyltransferase (pcm)	47%	XF1579	transmembrane protein (amp							

XF1913	type V secretory pathway protein (mttC)	46%	XF2177	phage-related protein	24%	XF1827	organic hydroperoxide resistance protein (ohr)	70%	XF0595	virulence factor	48%
CELLULAR PROCESSES/TRANSPORT/Other			XF2132	phage-related protein	34%	XF2586	outer membrane export factor	33%	XF2420	virulence factor (mvvN)	39%
XF0875	ABC transporter ATP-binding protein (yusC)	49%	XF2291	phage-related protein	57%	XF2550	outer membrane hemolysin activator protein (hacB)	76%	XF0755	virulence protein (acvB)	39%
XF0944	ABC transporter ATP-binding protein (yjkK)	69%	XF2292	phage-related protein	47%	XF2131	polyketide synthase (PKS) (frnE)	35%	XF2679	virulence protein (acvB)	28%
XF1077	ABC transporter ATP-binding protein (yjcV)	55%	XF2295	phage-related protein	53%	XF1532	oxidative stress transcriptional regulator (oxyR)	81%	XF0745	virulence regulator (kvxA)	47%
XF1081	ABC transporter ATP-binding protein (msbA)	41%	XF2479	phage-related protein	32%	XF1038	peptide synthase	23%	XF0493	virulence regulator (kvxA)	46%
XF1223	ABC transporter ATP-binding protein (yadG)	54%	XF2482	phage-related protein	21%	XF2276	peptide synthase	26%	XF0506	virulence-associated protein E (vapE)	31%
XF1307	ABC transporter ATP-binding protein	47%	XF2484	phage-related protein	41%	XF0619	periplasmic divalent cation tolerance protein (cutA/cycY/cutA1)	37%	XF2121	virulence-associated protein E (vapE)	32%
XF1409	ABC transporter ATP-binding protein	60%	XF2501	phage-related protein	27%	XF1729	phenylacetaldolase dehydrogenase	61%	ORFS WITH UNDEFINED CATEGORY		
XF1475	ABC transporter ATP-binding protein (yabH)	63%	XF2504	phage-related protein	44%	XF1131	phenylacetaldolase dehydrogenase	61%	XF2149	ApA protein (apaG/corD)	45%
XF1602	ABC transporter ATP-binding protein	31%	XF2505	phage-related protein	34%	XF1215	phenylacetaldolase dehydrogenase	61%	XF1828	ATPase	28%
XF1733	ABC transporter ATP-binding protein (yheS)	45%	XF2511	phage-related protein	28%	XF1729	phenylacetaldolase dehydrogenase	61%	XF0961	bacterioferritin comigratory protein (bcf)	51%
XF2568	ABC transporter ATP-binding protein (rftE)	45%	XF2527	phage-related protein	41%	XF2093	precursor of drug resistance protein (acrA/mecA/lir)	42%	XF1120	bifunctional DGTP-pyrophosphatase/thiamine phosphate synthase (mutT/thiE1)	43%
XF2582	ABC transporter ATP-binding protein	42%	XF2528	phage-related protein	52%	XF0769	pteridine-dependent deoxygenase (rapK)	31%	XF1786	bifunctional transcriptional repressor of the biotin operon/biotin acetyl-CoA-carboxylase synthetase (bifA/bioR/dnbB)	40%
XF2617	ABC transporter ATP-binding protein (yup)	49%	XF2524	phage-related protein	43%	XF1530	subunit C of alkyl hydroperoxide reductase (ahpC)	70%	XF1874	carboxylesterase (estB)	37%
XF2695	ABC transporter ATP-binding protein (HseE)	47%	XF2526	phage-related protein	50%	XF1899	TolD protein (tolD)	48%	XF0063	competence protein F (comF)	37%
XF1474	ABC transporter membrane protein (ylnC)	33%	XF2527	phage-related protein	21%	XF1997	TonB protein (tonB)	72%	XF1129	competence related protein (comM)	35%
XF1476	ABC transporter membrane protein	69%	XF2528	phage-related protein	56%	XF1957	TonB protein (tonB)	26%	XF1078	DNA uptake protein (comA)	35%
XF0674	ABC transporter permease protein	48%	XF0500	phage-related repressor protein (racR)	50%	XF2267	TonB protein (tonB)	28%	XF2243	GTP binding protein (yhbZ)	69%
XF2567	ABC transporter permease protein (rftD)	35%	XF0696	phage-related repressor protein (pqi)	29%	XF2605	TonB protein (tonB)	36%	XF1213	GTP-binding elongation factor protein (tyrA)	69%
XF1604	ABC transporter vitamin B17 uptake permease (thdF)	44%	XF1658	phage-related repressor protein (rli)	24%	XF2397	toxin secretion ABC transporter ATP-binding protein (thyB)	61%	XF0088	GTP-binding protein (hflX)	54%
XF0406	export protein (yggI)	14%	XF1706	phage-related tail fiber protein (gp37)	25%	XF1841	undecaprenyl kinase (bacA)	38%	XF0465	GTP-binding protein (ylyK)	46%
XF0030	large-conductance mechanosensitive channel (mscL)	52%	XF0725	phage-related tail protein (gp1)	36%	XF1981	undecaprenyl kinase (bacA)	38%	XF1430	GTP-binding protein (yhaA)	52%
XF0680	permease	78%	XF0730	phage-related tail protein (gp1)	21%	XF1267	1,4-beta-cellulohydrolase (cbhA)	46%	XF2422	GTP-binding protein (yhbZ)	55%
XF2701	onlysaccharide export protein (mvp)	63%	XF0732	phage-related tail protein (gpK)	41%	XF0818	endo-1,4-beta-glucanase (engXCA)	64%	XF2641	GTP-binding protein	68%
XF1358	small-conductance mechanosensitive ion channel (yggB)	35%	XF0733	phage-related tail protein (gpD)	32%	XF2708	endo-1,4-beta-glucanase (egl)	47%	XF1668	HicB-related protein	38%
XF2751	solute Na+ symporter (pps)	58%	XF2490	phage-related tail protein (gpJ)	41%	XF0810	extracellular endoglucanase precursor (engXCA)	33%	XF1192	integral membrane protein	49%
XF0281	transport protein (yufF)	74%	XF2491	phage-related tail protein (gpJ)	40%	XF0445	family 3 glycoside hydrolase (xylA)	45%	XF1977	integral membrane protein	...
XF0651	transport protein (yxaH)	33%	XF2497	phage-related tail protein (gpJ)	35%	XF2466	polygalacturonase precursor (pgaIA)	...	XF2645	isopentenyl monophosphate kinase (ipk)	50%
XF1778	transport protein	27%	XF0508	phage-related terminase large subunit (gp2)	25%	XF2466	polygalacturonase precursor (pgaIA)	...	XF0145	oxidoreductase (ydfG)	50%
CELLULAR PROCESSES/CELL DIVISION			XF0711	phage-related terminase large subunit (gp2)	28%	PATHOGENICITY, VIRULENCE, AND ADAPTATION/HOST CELL WALL DEGRADATION			XF1744	oxidoreductase	71%
XF0659	cell cycle protein (mesJ)	42%	XF2045	conjugal transfer protein (trbN)	76%	XF1841	undecaprenyl kinase (bacA)	38%	XF1723	sugar-phosphate dehydrogenase (yprG)	55%
XF0286	cell division inhibitor (sulA)	39%	XF2046	conjugal transfer protein (trbL)	45%	PATHOGENICITY, VIRULENCE, AND ADAPTATION/HOST CELL WALL DEGRADATION			XF1724	sugar-phosphate dehydrogenase (yprG)	55%
XF1322	cell division inhibitor (minC)	35%	XF2048	conjugal transfer protein (trbL)	...	PATHOGENICITY, VIRULENCE, AND ADAPTATION/HOST CELL WALL DEGRADATION			pXf51 genes		
XF0093	cell division protein (Htb/Hts/Hrs/CtoZ)	64%	XF2049	conjugal transfer protein (trbL)	53%	XF1267	1,4-beta-cellulohydrolase (cbhA)	46%	INTERMEDIARY METABOLISM/REGULATORY FUNCTIONS		
XF0094	cell division protein (ftsJ/msrF)	52%	XF2050	conjugal transfer protein (trbH)	45%	XF0818	endo-1,4-beta-glucanase (engXCA)	64%	XF0001	transcriptional regulator	31%
XF0791	cell division protein (ftsL)	30%	XF2051	conjugal transfer protein (trbH)	65%	XF2708	endo-1,4-beta-glucanase (egl)	47%	XF0057	transcriptional regulator (korA)	40%
XF0795	cell division protein (ftsW)	44%	XF2052	conjugal transfer protein (trbG)	43%	XF0810	extracellular endoglucanase precursor (engXCA)	33%	MACROMOLECULE METABOLISM/DNA METABOLISM/Replication		
XF0800	cell division protein (ftsZ)	32%	XF2053	conjugal transfer protein (trbE)	49%	XF0445	family 3 glycoside hydrolase (xylA)	45%	XF0061	single-strand binding protein (ssb)	54%
XF0801	cell division protein (ftsA/divA)	45%	XF2054	conjugal transfer protein (trbD)	68%	XF2466	polygalacturonase precursor (pgaIA)	...	XF0003	topoisomerase I (topA/supX)	35%
XF0802	cell division protein (ftsZ)	59%	XF2055	conjugal transfer protein (trbC)	52%	PATHOGENICITY, VIRULENCE, AND ADAPTATION/EXOPLYACCHARIDES			MACROMOLECULE METABOLISM/DNA METABOLISM/Recombination		
XF1450	cell division protein (ftsK)	50%	XF2056	conjugal transfer protein (trbB)	69%	XF2370	GumB protein (gumB)	67%	XF0019	site-specific recombinase (rim)	91%
XF1910	cell division protein (ftsY)	45%	XF2058	conjugal transfer protein (traF)	59%	XF2369	GumC protein (gumC)	62%	MOBILE GENETIC ELEMENTS/PLASMID-RELATED FUNCTIONS		
XF2557	cell division protein (zipA)	27%	XF2059	conjugal transfer protein (traF)	73%	XF2367	GumD protein (gumD)	73%	XF0002	conjugal transfer protein (traA/virB1)	42%
XF2694	cell division protein (ftsX/HtsS)	31%	XF2060	conjugal transfer protein (traD)	43%	XF2366	GumE protein (gumE)	61%	XF0005	conjugal transfer protein (trbC/virB2)	36%
XF1320	cell division topological specificity factor (minE)	48%	XF1759	conserved plasmid protein	35%	XF2365	GumF protein (gumF)	44%	XF0006	conjugal transfer protein (traA/virB3)	39%
XF1651	cell filamentation protein (fici)	38%	XF2066	conserved plasmid protein (yacB)	35%	XF2364	GumH protein (gumH)	64%	XF0007	conjugal transfer protein (trbC/virB4)	43%
XF2281	chromosome partitioning protein	43%	XF2161	conserved plasmid protein	30%	XF2362	GumJ protein (gumJ)	65%	XF0011	conjugal transfer protein (traD/virB6)	23%
XF2287	chromosome partitioning protein (parA)	49%	XF2061	CNA primase (traC)	44%	XF2361	GumK protein (gumK)	69%	XF0012	conjugal transfer protein (traE/virB8)	33%
XF1785	chromosome partitioning related protein (soj)	27%	XF2444	pheromone shutdown protein (traB)	31%	XF2360	GumL protein (gumL)	65%	XF0013	conjugal transfer protein (traD/virB9)	34%
XF2247	GTP binding protein (era)	51%	XF1589	p asmid stabilization protein	50%	PATHOGENICITY, VIRULENCE, AND ADAPTATION/SURFACE PROTEINS			XF0014	conjugal transfer protein (traF/virB10)	40%
XF1084	partition protein (parA)	32%	XF1590	plasmid stabilization protein	62%	XF0889	hemagglutinin-like secreted protein (pspA)	25%	XF0015	conjugal transfer protein	...
XF1371	septum site-determining protein (minD)	68%	XF2031	plasmid stabilization protein (parD)	81%	XF2196	hemagglutinin-like secreted protein (pspA)	35%	XF0016	conjugal transfer protein (traG/virD4)	35%
CELLULAR PROCESSES/CHEMOTAXIS AND MOBILITY			XF2032	plasmid stabilization protein (parE)	75%	XF2775	hemagglutinin-like secreted protein (pspA)	35%	XF0036	conjugal transfer protein (trbL)	45%
XF1952	chemotaxis-related protein kinase (chpA)	42%	XF1574	plasmid-related protein (traN)	34%	XF1529	surface protein (hsf)	22%	XF0037	conjugal transfer protein (trbL)	54%
MOBILE GENETIC ELEMENTS/PHAGE-RELATED FUNCTIONS AND PROPHAGES			XF1679	plasmid-related protein (traN)	34%	XF1981	surface protein (hsf)	24%	XF0040	conjugal transfer protein (trbL)	52%
XF0089	host factor-I protein (hflg)	92%	XF1945	VirK protein (virK)	35%	XF1516	surface-exposed outer membrane protein (uspA1)	28%	XF0041	conjugal transfer protein (trbH)	44%
XF0719	phage-related baseplate assembly protein (gpV)	32%	MOBILE GENETIC ELEMENTS/TRANSPOSON- AND INTRON-RELATED FUNCTIONS			XF0432	Birk protein (birk)	33%	XF0042	conjugal transfer protein (trbG)	65%
XF0723	phage-related baseplate assembly protein (gpW)	40%	XF2063	DNA invertase (inv)	...	XF0665	cobalt-zinc-cadmium resistance protein (czrD)	53%	XF0043	conjugal transfer protein (trbG)	67%
XF0724	phage-related baseplate assembly protein (gpJ)	40%	XF1930	resolvase (trpR)	...	XF0785	heat shock protein (htrA/degP/pdi)	43%	XF0044	conjugal transfer protein (trbC)	...
XF2488	phage-related baseplate assembly protein (gpJ)	46%	XF1775	reverse transcriptase	...	XF2675	heat shock protein (htrX)	53%	XF0027	plasmid maintenance protein (pemK)	49%
XF2489	phage-related baseplate assembly protein (gpW)	40%	XF1931	transposase (trpA)	...	XF2234	low molecular weight heat shock protein (hspA)	38%	XF0060	plasmid replication protein (trnC)	34%
XF2492	phage-related baseplate assembly protein (gpV)	32%	XF2303	transposase (trpA)	71%	XF1379	lipase	31%	XF0059	plasmid replication/partition protein	32%
XF3727	phage-related contractile tail sheath protein (HfR2)	46%	XF0325	transposase DriA	64%	XF0637	organic solvent tolerance precursor (ompA/ostA)	29%	XF0029	plasmid stabilization protein (parE)	76%
XF2485	phage-related contractile tail sheath protein (HfR2)	46%	XF0535	transposase DriA	64%	XF1523	periplasmic glucan biosynthesis protein (mdoH)	37%	PATHOGENICITY, VIRULENCE, AND ADAPTATION/OTHER		
XF1728	phage-related contractile tail tube protein (HfR2)	41%	XF0326	transposase DriB	...	XF2682	periplasmic glucan biosynthesis protein (mdoG)	36%	XF0052	virulence-associated protein D (vapD)	54%
XF0683	phage-related [DNA polymerase (dpoI)]	50%	XF0536	transposase DriB	35%	XF0785	sulfur deprivation response regulator (sac1)	22%	pXf1.3 genes		
XF2525	phage-related [DNA polymerase (dpoI)]	49%	PATHOGENICITY, VIRULENCE, AND ADAPTATION/TOXIN PRODUCTION AND DETOXIFICATION			XF0858	survival protein (swrE)	60%	MOBILE GENETIC ELEMENTS/PLASMID-RELATED FUNCTIONS		
XF0513	phage-related endonuclease (hycV)	48%	XF1726	2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase (HfR2)	39%	XF2622	temperature acclimation protein B (tapB)	85%	XF0001	replication protein	41%
XF0631	phage-related integrase (hnt)	26%	XF0243	acetylflavin resistance protein (yepP)	23%	XF0418	toluene tolerance protein (ttg2C)	24%			
XF0678	phage-related integrase (hnt)	27%	XF2386	acetylflavin resistance protein (yepN)	42%	XF0419	toluene tolerance protein (ttg2C)	41%			
XF1647	phage-related integrase (hnt)	28%	XF2385	acetylflavin resistance protein D (yepN)	40%	XF0420	toluene tolerance protein (ttg2B)	42%			
XF1718	phage-related integrase (hnt)	91%	XF2407	bacteriocin	27%	XF0421	toluene tolerance protein (ttg2A)	50%			
XF2288	phage-related integrase (hnt)	26%	XF0165	beta-lactamase induction signal transducer protein (lampG)	38%	PATHOGENICITY, VIRULENCE, AND ADAPTATION/OTHER					
XF2530	phage-related integrase (hnt)	28%	XF1621	beta-lactamase-like protein (pdp)	24%	XF0290	aconitase (rpfA)	79%			
XF1554	phage-related lysozyme (hycV)	43%	XF0010	biopolymer transport Exb protein (exbB)	81%	XF1424	chitinase (chi)	43%			
XF1569	phage-related lysozyme (hycV)	43%	XF0012	biopolymer transport ExbD2 protein (exbD2)	78%	XF1527	general secretory pathway protein D precursor (xpsD/pefD)	68%			
XF2314	phage-related lysozyme (hycV)	42%	XF2232	catalase/peroxidase (cpeB)	66%	XF1517	general secretory pathway protein E precursor (xpsE/pefE)	81%			
XF0713	phage-related portal protein (gp4)	25%	XF2083	cation efflux system protein (czcA)	24%	XF1518	general secretory pathway protein F precursor (xpsF/pefF)	86%			
XF2498	phage-related portal protein (gp4)	25%	XF0262	colicin V precursor (cvaC)	36%	XF1519	general secretory pathway protein G precursor (xpsG/pefG)	78%			
XF0485	phage-related protein	34%	XF1948	colicin V production protein (cvpA)	31%	XF1520	general secretory pathway protein H precursor (xpsH)	67%			
XF0660	phage-related protein	56%	XF1220	colicin V secretion ABC transporter ATP-binding protein (cvaB)	40%	XF1521	general secretory pathway protein I precursor (xpsI/pefI)	55%			
XF0682	phage-related protein	50%	XF1216	colicin V secretion protein (cvaA)	25%	XF1522	general secretory pathway protein J precursor (xpsJ/pefJ)	56%			
XF0694	phage-related protein	49%	XF2064	component of multidrug efflux system (mexE)	28%	XF1523	general secretory pathway protein K precursor (xpsK/pefK)	60%			
XF0695	phage-related protein	53%	XF1341	copper homeostasis protein (cucC)	47%	XF1524	general secretory pathway protein L precursor (xpsL/pefL)	60%			
XF0704	phage-related protein	41%	XF0132	copper resistance protein A precursor (copA)	61%	XF0720	proteic killer active protein (higB)	33%			
XF0705	phage-related protein	32%	XF0133	copper resistance protein B precursor (copB)	38%	XF0721	proteic killer suppression protein (higA)	40%			
XF0707	phage-related protein	44%	XF1741	daunorubicin C-13 ketoreductase (dnrK)	31%	XF0723	queuine tRNA-ribosyltransferase (tgr/vacC)	61%			
XF0710	phage-related protein	27%	XF2416	drug tolerance protein (yrbB)	64%	XF0727	regulator of pathogenicity factors (rpfH)	72%			
XF1555	phage-related protein	34%	XF0993	drug-proton antiporter (htrV)	24%	XF1114	regulator of pathogenicity factors (rpfC)	59%			
XF1570	phage-related protein	29%	XF1765	drug-proton antiporter (htrV)	33%	XF11					

Mapa das ORFs

Análise da categoria

II. Biossíntese de Pequenas Moléculas

Os esquemas das vias de biossíntese de aminoácidos e cofatores foram baseados nas vias de biossíntese desses compostos em *Escherichia coli* encontrados no site,

<http://www.ecocyc.pangeasystems.com>

A legenda a seguir refere-se a todas as vias de biossíntese esquematizadas:

Em **vermelho**, representaram-se os microrganismos onde as enzimas estão presentes, na seguinte ordem: e= *Escherichia coli*, b= *Bacillus subtilis*, r= *Mycobacterium tuberculosis*, o= *Borrelia burgdorferi*, n= *Chlamydia pneumoniae*, s= *Neisseria meningitidis* (ausente no banco de dados do COG), h= *Haemophilus influenzae* e f= *Xylella fastidiosa* (ausente no banco de dados do COG). O traço (-) indica que no microrganismo representado nesta ordem a enzima está ausente. Em **negrito**, representou-se a sigla que determina o nome do gene. EC= enzyme code, COG= Cluster of Ortholog Groups. Os compostos estão representados nos quadrados em cinza e as enzimas, nos quadrados em branco. As setas pontilhadas indicam a regulação das vias pelos compostos indicados. Setas rígidas indicam o sentido das reações. Em *E. coli*, algumas enzimas são bifuncionais. Nestes casos, todas as subunidades foram representadas mas encontram-se em negrito e sublinhado as subunidades que atuam naquele passo específico da via.

A legenda abaixo representa todas as tabelas apresentadas:

Na coluna "Proteínas" encontram-se todas as enzimas envolvidas na biossíntese de pequenas moléculas segundo a via de *E. coli*, onde EC= enzyme code. Na coluna "COG", **em verde**, encontram-se os números das proteínas identificadas pelo "Cluster of Ortholog Groups". Notar que para *N. meningitidis* não existem COGs, pois a comparação dos genes presentes nesta bactéria aos outros microrganismos se deu com base somente aos nomes dos genes anotados nesse microrganismo. As ORFs anotadas em *X. fastidiosa* foram comparadas aos COGs do Banco de Dados do NCBI pelo Laboratório de Bioinformática, mas ainda não se encontram disponibilizados ao público neste banco de dados. As siglas em *itálico*, representam os nomes dos genes nos diferentes microrganismos; em *X. fastidiosa*, aparecem como o nome da ORF anotada.

II.A. AMINOÁCIDOS

II.A.1. Família do glutamato e assimilação de nitrogênio

1. ARGININA

Tabela 1: Lista dos genes anotados na via de biossíntese de Arginina

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
N-acetilglutamato sintase EC 2.3.1.1	1246	argA 1246		Rv2747 1246			argA		ORFs
N-acetilglutamato quinase EC 2.7.2.8	0548	argB 0548	argB 0548	argB 0548			argB		07C06-GL43 0548
N-acetil-(gamma)-glutamil- fosfato redutase EC 1.2.1.38	0002	argC 0002	argC 0002	argC 0002			argC		07C06-GL44 0002
Acetilornitina transaminase EC 2.6.1.11	0160	argD 0160	argD 0160	argD 0160			argD	H10949 0160	07C12-GL48 0160
Acetilornitina deacetilase EC 3.5.1.16	0624	argE 0624	argE 0624	Rv2141c 0624				H10102 0624	07C06-GL42 0624
Ornitina carbamoiltransferase, cadeia F EC 2.1.3.3	0078	argF 0078	argF 0078	argF 0078			argF	arcB 0078	07C06-GL40 0078
Argininosuccinato sintase EC 6.3.4.5	0137	argG 0137	argG 0137	argG 0137			argG	argG 0137	07C06-GL41 0137
arginino succinato liase EC 4.3.2.1	0165	argH 0165	argH 0165	argH 0165			argH	argH 0165	07C06-GL45 0165
Ornitina acetiltransferase / aminoácido acetiltransferase			argJ	argJ			argJ		
Ornitina carbamoiltransferase, cadeia I EC 2.1.3.3	0078	argI 0078						arcB	07C06-GL40 0078
Carbamoilfosfato sintase, cadeia pequena EC 6.3.5.5	0505	carA 0505	carA 0505	carA 0505					03C11-GL12 0505
Carbamoilfosfato sintase, cadeia grande EC 6.3.5.5	0458	carB 0458	carB 0458	carB 0458					03C11-GL10 0458

Legenda: proteínas envolvidas na biossíntese de Glutamato segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.1. Família do glutamato e assimilação de nitrogênio

2. GLUTAMATO

Tabela 2: Lista dos genes anotados na via de biossíntese de Glutamato

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>B. burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
Glutamato sintase, subunidade alpha EC 1.4.1.13	0069	<i>gltB</i> 0069	<i>gltB</i> 0069	<i>gltB</i> 0069					ORFs 07B07-GL16 0069
Glutamate sintase, subunidade beta EC 1.4.1.13	0493	<i>gltD</i> 0493	<i>gltA</i> 0493	<i>gltD</i> 0493					07B07-GL17 0493
Glutamato desidrogenase EC 1.4.1.4	0334	<i>gdhA</i> 0334	<i>gudB</i> <i>rocG</i> 0334	<i>Rv2476c</i> 0334			<i>gdhA</i> <i>gluD</i>	<i>gdhA</i> 0334	

Legenda: proteínas envolvidas na biossíntese de Glutamato segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.1. Família do glutamato e assimilação de nitrogênio

3. GLUTAMINA

Tabela 3: Lista dos genes anotados na via de biossíntese de Glutamina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
glutamina sintetase EC 6.3.1.2	0174	<i>glnA</i> 0174	<i>glnA</i> 0174	<i>glnA1</i> <i>glnA2</i> <i>glnA3</i> <i>glnA4</i> 0174			<i>glnA</i>	<i>glnA</i> 0174	07F02-GL28 0174

Legenda: proteínas envolvidas na biossíntese de Glutamina segundo a via de *E. coli*.

Regulação da biossíntese de glutamina

II.A. AMINOÁCIDOS

II.A.1. Família do glutamato e assimilação de nitrogênio

4. PROLINA

Tabela 4: Lista dos genes anotados na via de biossíntese de Prolina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
gamma-glutamyl fosfato redutase EC 1.2.1.41	0014	proA 0014	proA 0014	proA 0014			proA 0014	proA 0014	07C06-GL47 0014
gamma-glutamyl quinase EC 2.7.2.11	0263	proB 0263	proB 0263	proB 0263			proB 0263	proB 0263	07C06-GL46 0263
Pirrolina-5-carboxilato redutase EC 1.5.1.2	0345	proC 0345	proH 0345	proC 0345			proC 0345	proC 0345	07B07-GL15 0345

Legenda: proteínas envolvidas na biossíntese de Prolina segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.2. Família do aspartato e piruvato

1. ALANINA

Tabela 5: Lista dos genes anotados na via de biossíntese de Alanina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
alanina racemase, biossintética EC 5.1.1.1	0787	<i>alr</i> 0787	<i>dal</i> 0787	<i>alr</i> 0787	<i>alr</i> 0787			<i>alr</i> 0787	05G03-GL53 0787
transaminase B EC 2.6.1.42	0115	<i>ilvE</i> 0115	<i>pabC</i> 0115	<i>ilvE</i> 0115			<i>ilvE</i>	<i>ilvE</i> 0115	11A02-GL31 0115
transaminase C EC 2.6.1.66	0436	<i>avtA</i> 0436	<i>patA</i> 0436	Rv3565 0436		CP0259 0436	<i>avtA</i>	HI0286 0436	
Alanina racemase, catabólica EC 5.1.1.1	0787	<i>dadX</i> 0787							
N-carbamoyl-beta-alanina amidohidrolase									02A11-GL26 0388

Legenda proteínas envolvidas na biossíntese de Alanina segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.2. Família do aspartato e piruvato

2. ASPARAGINA

Tabela 6: Lista dos genes anotados nas vias de biossíntese de Asparagina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i>
Aspartato-amônia ligase EC 6.3.1.1	Não encontrado	<i>asnA</i>	<i>asnA</i>					<i>asnA</i>	
asparagina sintetase B EC 6.3.5.4	0367	<i>asnB</i> 0367	<i>asnB</i> <i>asnH</i> 0367	<i>asnB</i> 0367					07A09-GL34 0367

Legenda: proteínas envolvidas na biossíntese de Asparagina segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.2. Família do aspartato e piruvato

3. ASPARTATO

Tabela 7: Lista de genes anotados na via de biossíntese do Aspartato

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
aspartato aminotransferase EC 2.6.1.1	0436	<i>aspC</i> 0436	<i>aspB</i> 0436	<i>aspB</i> <i>aspC</i> 0436		CP0259 0436	<i>aspC</i>	<i>aspC</i> 0436	05803-GL13 0436
aspartil/ asparaginil beta-hidrolase (aspH)									06D10-GL51 sem COG

Legenda: proteínas envolvidas na biossíntese de Aspartato segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.2. Família do aspartato e pruvato

4. ISOLEUCINA

II.A. AMINOÁCIDOS

II.A.2. Família do aspartato e piruvato

5. VALINA

Tabela 8: Lista de genes anotados na via de biossíntese de Valina e Isoleucina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
Treonina deaminase EC 4.2.1.16	1171	<i>tdcB</i> 1171							11B08-GL09 1171
Treonina desidratase, proteína biossintética EC 4.2.1.16	1171	<i>ivA</i> 1171	<i>ivA</i> 1171	<i>ivA</i> 1171			<i>ivA</i>		
Acetolactato sintase I, subunidade maior EC 4.1.3.18	0028	<i>ivB</i> 0028	<i>ivB</i> 0028	<i>ivB</i> <i>ivB2</i> 0028				H10737 0028	
Ácido-ceto- reductoisomerase EC 1.1.1.86	0059	<i>ivC</i> 0059	<i>ivC</i> 0059	<i>ivC</i> 0059			<i>ivC</i>	<i>ivC</i> 0059	07F02-GL01 0059
Ácido desidróxi desidratase EC 4.2.1.9	0129	<i>ivD</i> 0129	<i>ivD</i> 0129	<i>ivD</i> 1029			<i>ivD1</i> <i>ivD2</i>	<i>ivD</i> 0129	07C10-GL36 1029
transaminase B EC 2.6.1.42	0115	<i>ivE</i> 0115	<i>ivE</i> 0115	<i>ivE</i> 0115			<i>ivE</i>	<i>ivE</i> 0115	11A02-GL31 0115
acetolactato sintase II, subunidade menor EC 4.1.3.18	0028	<i>ivG</i> 0028	<i>alsS</i> 0028	<i>ivG</i> 0028				<i>ivG</i> 0028	11B08-GL11 0028
acetolactato sintase III, subunidade menor EC 4.3.1.18	0440	<i>ivH</i> 0440					<i>ivH</i>	<i>ivH</i> 0440	
acetolactato sintase III, subunidade maior EC 4.3.1.18	0028	<i>ivI</i> 0028	<i>ivB</i> 0028	<i>ivB</i> 0028			<i>ivI</i>	<i>ivI</i> 0028	
acetolactato sintase II, subunidade maior EC 4.1.3.18	0440	<i>ivM</i> 0440		<i>ivX</i> 0440				<i>ivH</i> 0440	
acetolactato sintase I, subunidade menor EC 4.1.3.18	0440	<i>ivN</i> 0440	<i>ivN</i> 0440	<i>ivN</i> 0440					

Legenda: proteínas envolvidas na biossíntese de Isoleucina e Valina segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.2. Família do aspartato e piruvato

6. LEUCINA

Tabela 9: Lista de genes anotados na via de biossíntese de Leucina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
2-isopropilmalato sintase EC 4.1.3.12	0119	<i>leuA</i> 0119	<i>leuA</i> 0119	<i>leuA</i> 0119			<i>leuA</i>	<i>leuA</i> 0119	11B08-GL08 0119
3-isopropilmalato desidrogenase EC 1.1.1.85	0473	<i>leuB</i> 0473	<i>leuB</i> 0473	<i>leuB</i> 0473			<i>leuB</i>	<i>leuB</i> 0473	07A02-GL31 0473
3-isopropilmalato desidratase, subunidade maior EC 4.2.1.33	0065	<i>leuC</i> 0065	<i>leuC</i> 0065	<i>leuC</i> 0065			<i>leuC</i>	<i>leuC</i> 0065	07A02-GL33 0065
3-isopropilmalato desidratase, subunidade menor EC 4.2.1.33	0066	<i>leuD</i> 0066	<i>leuD</i> 0066	<i>leuD</i> 0066			<i>leuD</i>	<i>leuD</i> 0066	07A02-GL32 0066
Aminoácido-aromático aminotransferase EC 2.6.1.57	1448	<i>tyrB</i> 1448				CP0005 1448			03H11-GL44 1448
"Branched-chain" aminoácido aminotransferase EC 2.6.1.42	0115	<i>ilvE</i> 0115		<i>ilvE</i> 0115			<i>ilvE</i>	<i>ilvE</i> 0115	11A02-GL31 0115

Legenda: proteínas envolvidas na biossíntese de Leucina segundo a via de *E. coli*.

Via de Integração da biossíntese de Alanina, Valina e Leucina em bactérias

II.A. AMINOÁCIDOS

II.A.2. Família do aspartato e piruvato

7. LISINA

Tabela 10: Lista dos genes anotados na via de Biossíntese de Lisina

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
Aspartato-B-semialdeído desidrogenase EC 1.2.1.11	0136	<i>asd</i> 0136	<i>asd</i> 0136	<i>asd</i> 0136		<i>asd</i> 0136	<i>asd</i>	<i>asd</i> 0136	08A02-GL13 0136
dihidrodipicolinato sintase EC 4.2.1.52	0329	<i>dapA</i> 0329	<i>dapA</i> 0329	<i>dapA</i> 0329		<i>dapA</i> 0329	<i>dapA</i>	<i>dapA</i> 0329	08D10-GL27 0329
dihidrodipicolinato redutase EC 1.3.1.26	0289	<i>dapB</i> 0289	<i>dapB</i> 0289	<i>dapB</i> 0289		<i>dapB</i> 0289	<i>dapB</i>	<i>dapB</i> 0289	03C11-GL13 0289
2,3,4,5-tetrahidropiridina-2-carboxilato-N-succiniltransferase EC 2.3.1.117	2171	<i>dapD</i> 2171	<i>ykuQ</i> 2171				<i>dapD</i>	<i>dapD</i> 2171	07A09-GL39 2171
succinil-diaminopimelato EC 3.5.1.18	0624	<i>dapE</i> 0624	<i>argE</i> 0624	<i>dapE</i> <i>dapE2</i> 0624		CP0876 0624	<i>dapE</i>	<i>dapE</i> 0624	07A09-GL37 0624
diaminopimelato epimerase EC 5.1.1.7	0253	<i>dapF</i> 0253	<i>yutL</i> 0253	<i>dapF</i> 0253		<i>dapF</i> 0253	<i>dapF</i>	<i>dapF</i> 0253	04A09-GL35 0253
Diaminopimelato decarboxilase EC 4.1.1.20	0019	<i>lysA</i> 0019	<i>lysA</i> 0019	<i>lysA</i> 0019			<i>lysA</i>	<i>lysA</i> 0019	0.4-GL1696 * 0019 0527
aspartato kinase III EC 2.7.2.4	0527	<i>lysC</i> <i>apk</i> 0527	<i>dapG</i> <i>lysC</i> 0527	<i>ask</i> 0527		<i>lysC</i> 0527	<i>lysC</i>		0.4-GL1696 * 0019 0527

Legenda: proteínas envolvidas na biossíntese de Lisina segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.2. Família do aspartato e piruvato

8. METIONINA

Tabela 11: Lista dos genes anotados na via de Biossíntese de Metionina

Proteína	COG	E.coli	B. subtilis	M. tuberculosis	Borrelia burgdorferi	C. pneumoniae	N. meningitidis	H. influenzae	X.fastidiosa ORFs
Homoserina O-succiniltransferase EC 2.3.1.46	1897	metA 1897	metB 1897						
Aspartato-B-semialdeído desidrogenase EC 1.2.1.11	0136	asd 0136	asd 0136	asd 0136		asd 0136	asd	asd 0136	08A02-GL13 0136
homoserine O-acetiltransferase EC 2.3.1.31	2021			metA 2021			metA	met2 2021	0.3-GL3362 2021
cistationina gamma-sintase EC 4.2.99.9	0626	metB 0626	yrhB 0626	metB 0626			metB	metB 0626	05G03-GL32 0626
beta cistationase EC 4.4.1.8	0626	metC 0626	yjcJ 0626	metC 0626				metC 0626	
cobalamina independente metionina sintase EC 2.1.1.14	0620	metE 0620	metC 0620	metE 0620				metE 0620	11A07-GL50 0620
Metionina sintase, cobalimina dependente EC 2.1.1.13	0646	metH 0646 1410	yetJ 0646	metH 0646 1410			metH	metH 1410	
bifuncional aspartoquinase II/ homoserina desidrogenase II EC 2.7.2.4/EC 1.1.1.3	0460 0527 asp	hom 0460 0527	hom 0460 dapG 0527	thrA 0460 ask 0527			asp	thrA 0460 0527	
5,10-metilenotetrahidrofolato redutase Ec 1.7.99.5	0685	metF 0685	yitJ_2 0685				metF	metF 0685	0.3-GL1144 0685
o-succinil-homoserina sulfidrilase				metZ					

Legenda: proteínas envolvidas na biossíntese de Metionina segundo a via de E. coli.

II.A. AMINOÁCIDOS

II.A.2. Família do aspartato e piruvato

9. TREONINA

Tabela 12: Lista dos genes anotados na via de Biossíntese de Treonina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i>
Enzima bifuncional aspartoquinase/homoserina desidrogenase I EC 2.7.2.4/EC 1.1.1.3	0527 0460 hom	thrA 0460 0527	hom 0460 dapG 0527	thrA 0460 ask 0527					11A09-GL31 0460 0527
Aspartato-β-semialdeído desidrogenase EC 1.2.1.11	0136	asd 0136	asd 0136	asd 0136		asd 0136	asd	asd 0136	08A02-GL13 0136
homoserina quinase EC 2.7.1.39	0083	thrB 0083	thrB 0083	thrB 0083			thrB	thrB 0083	11A09-GL32 0083
treonina sintase EC 4.2.99.2	0498	thrC 0498	thrC 0498	thrC 0498			thrC	thrC 0498	11A09-GL33 0498

Legenda: proteínas envolvidas na biossíntese de Treonina segundo a via de *E. coli*.

II. BIOSÍNTESE DE PEQUENAS MOLÉCULAS

II.A. AMINOÁCIDOS

II.A.3. Família Glicina - Serina

1. CISTEÍNA

Tabela 13: Lista de genes anotados na via de biossíntese de Cisteína

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i> ORFs
cistationina beta-sintase ou serina sulfidrilase EC 4.2.1.22	0031			cysM2					06A08-GL19 0031
serina acetiltransferase	1045	cysE 1045	cysE 1045	cysE 1045			cysE	cysE 1045	
O-acetilserina tiol liase A EC 4.2.99.8	0031	cysK cysZ 0031	cysK 0031	cysK 0031			cysK	cysK cysZ 0031	08A03-GL18 0031
O-acetilserina tiol liase B EC 4.2.99.8	0031	cysM 0031		cysM cysM3 0031				cysK 0031	07A09-GL17 0031

Legenda: proteínas envolvidas na biossíntese de Cisteína segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.3. Família Glicina - Serina

2. GLICINA

3. SERINA

Tabela 14: Lista dos genes anotados na via de biossíntese de Glicina

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
fosfoserina fosfatase EC 3.1.3.3	0560	<i>serB</i> 0560		<i>serB</i> <i>serB2</i> 0560			<i>serB</i>	<i>serB</i> 0560	ORFs
Serina hidroximetiltransferase EC 2.1.2.1	0112	<i>glyA</i> 0112	<i>glyA</i> 0112	<i>glyA</i> <i>glyA2</i> 0112	<i>glyA</i> 0112	<i>glyA</i> 0112	<i>glyA</i>	<i>glyA</i> 0112	08D10-GL50 0112

Legenda: proteínas envolvidas na biossíntese de Glicina segundo a via de *E. coli*.

Tabela 15: Genes anotados na via de biossíntese de Serina

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
D-3-fosfoglicerato desidrogenase EC 1.1.1.95	0111	<i>serA</i> 0111	<i>serA</i> 0111	<i>serA</i> 0111				<i>serA</i> 0111	11A09-GL50 0111
fosfoserina fosfatase EC 3.1.3.3	0560	<i>serB</i> 0560		<i>serB</i> <i>serB2</i> 0560			<i>serB</i>	<i>serB</i> 0560	
fosfoserina aminotransferase	1932	<i>serC</i> 1932	<i>serC</i> 1932	<i>serC</i>			<i>serC</i>	<i>serC</i> 1932	08A04-GL20 1932

Legenda: proteínas envolvidas na biossíntese de Serina segundo a via de *E. coli*.

II. BIOSÍNTESE DE PEQUENAS MOLÉCULAS

II.A. AMINOÁCIDOS

II.A.3. Família Aminoácidos Aromáticos

1. FENILALANINA

Tabela 16: Lista de genes anotados na via de biossíntese de Fenilalanina

Proteína	COG	E.coli	B. subtilis	M. tuberculosis	Borrelia burgdorferi	C. pneumoniae	Neisseria meningitidis	H. influenzae	X.fastidiosa ORFs
fosfo-2-dehidro-3-deoksiheptonato aldolase	0722	aroG 0722	aroA 0722	aroG		aroG 0722	aroG	aroG 0722	03E01-GL39 0722
Corismato mutase EC 5.4.99.5/prefenato desidratase EC 4.2.1.51	1605 0077	pheA 1605 0077	pheA 1605 0077	pheA 1605 0077			pheA	pheA 1605 0077	08A04-GL19 0077 0.3-GL1118 1605
Aminoácido-aromático aminotransferase	1448	tyrB 1448				CP0005 1448	tyrB	aspC 1448	03H11-GL44 1448

Legenda: proteínas envolvidas na biossíntese de Fenilalanina segundo a via de *E. coli*.

II.A. AMINOÁCIDOS

II.A.3. Família dos Aminoácidos Aromáticos

2. TRIPTOFANO

Tabela 17: Lista dos genes anotados na via de biossíntese do Triptofano

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
triptofano sintase, cadeia alfa EC 4.2.1.20	0159	<i>trpA</i> 0159	<i>trpA</i> 0159	<i>trpA</i> 0159			<i>trpA</i>	<i>trpA</i> 0159	08A02-GL20 1059
triptofano sintase, cadeia beta EC 4.2.1.20	0133	<i>trpB</i> 0133	<i>trpB</i> 0133	<i>trpB</i> 0133			<i>trpB</i>	<i>trpB</i> 0133	08A02-GL19 0133
indol-3-glicerol fosfato sintase EC 4.1.1.48	0134 0135	<i>trpC</i> 0134 0135	<i>trpC</i> <i>trpF</i> 0134 0135	<i>trpC</i> 0134			<i>trpC</i> <i>trpF</i>	<i>trpC</i> 0134 0135	05G02-GL56 0134 08A02-GL17 0135
Antranilato sintase, componente II EC 4.1.3.27/ antranilato fosforibosiltransferase EC 2.4.2.18	0512 0547	<i>trpD</i> 0512 0547	<i>trpD</i> 0512 0547	<i>trpD</i> 0512 0547			<i>trpD</i>	<i>trpD</i> 0512 0547	05G02-GL55 0547
antranilato sintase, componente I EC 4.1.3.27	0147	<i>trpE</i> 0147	<i>trpE</i> 0147	<i>trpE</i> <i>trpE2</i> 0147			<i>trpE</i>	<i>trpE</i> 0147	05G02-GL53 09G12-GL09 0147
antranilato sintetase, componente II	0512		<i>pabA</i> 0512	<i>pabA</i> 0512				<i>trpG</i> 0512	05-GL2654 0512

Legenda: proteínas envolvidas na biossíntese de Triptofano segundo a via de *E. coli*.

A seguir, lista dos genes anotados na biossíntese de corismato:

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i>
3-fosfoshiquimato 1-carboxiviniltransferase EC 2.5.1.19	0128	<i>aroA</i> 0128	<i>aroE</i> 0128	<i>aroA</i> 0128		<i>aroA</i> 0128	<i>aroA</i>	<i>aroA</i> 0128	08A04-GL18 0128
3-hidroquinato sintase EC 4.6.1.3	0337	<i>aroB</i> 0337	<i>aroB</i> 0337	<i>aroB</i> 0337		<i>aroB</i> 0337	<i>aroB</i>	<i>aroB</i> 0337	06C10-GL38 0337
corismato sintase	0082	<i>aroC</i> 0882	<i>aroF</i> 0882	<i>aroF</i> 0882		<i>aroC</i> 0882	<i>aroC</i>	<i>aroC</i> 0882	08A02-GL09 0882
3-dehidroquinato desidratase EC 4.2.1.10	0710	<i>aroD</i> 0710	<i>aroC</i> 0710			bifuncional 0710	<i>aroD</i>		
Dehidroquinase, catabólica	0757		<i>yqhS</i> 0757	<i>aroD</i> 0757				<i>aroQ3</i> 0757	03H11-GL26 0757
shiquimato 5-desidrogenase	0169	<i>aroE</i> 0169	<i>aroD</i> 0169	<i>aroE</i> 0169		bifuncional 0169	<i>aroE</i>	<i>aroE</i> 0169	07C04-GL13 0169
3-deoxi-D-arabino-heptulosonato 7-fosfato sintase / corismato mutase	0722 1605 cho		<i>aroA</i> 0722 1605						
fosfo-2-dehidro-3-deoxiheptonato aldolase EC 4.1.2.15	0722	<i>aroG</i> <i>aroF</i> <i>aroH</i> 0722		<i>aroG</i> não encontrado		<i>aroG</i> 0722	<i>aroG</i>	<i>aroG</i> 0722	03E01-GL39 0722
Shiquimato quinase I EC 2.7.1.71	0703	<i>aroK</i> 0703	<i>arol</i> 0703	<i>aroK</i> 0703		<i>aroK</i> 0703	<i>aroK</i>	<i>aroK</i> 0703	06C10-GL39 0703
Shiquimato quinase II EC 2.7.1.71	0703	<i>arol</i> 0703	<i>arol</i> 0703						

II.A. AMINOÁCIDOS

II.A.4. Família dos Aminoácidos Aromáticos

3. TIROSINA

Tabela 18: Lista dos genes anotados na via de biossíntese de Tirosina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i>
Corismato mutase EC 5.4.99.5/prefenato desidrogenase EC 1.3.1.12	1605 0287	tyrA 1605 0287	aro H 1605 0287	tyrA 1605 0287				tyrA 1605 0287	08A04-GL33 1605
Aminoácido-aromático aminotransferase	1448	tyrB 1448				CP0005 1448	tyrB	aspC 1448	03H11-GL44 1448

Legenda: proteínas envolvidas na biossíntese de Tirosina segundo a via de *E. coli*

II. BIOSÍNTESIS DE PEQUEÑAS MOLÉCULAS

II.A. AMINOÁCIDOS

II.A.5. Histidina

Tabela 19: Lista dos genes anotados na via de biossíntese de Histidina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
fosforibosilformimino-5-amino-1-fosforibosil-4-imidazol carboxamida isomerase EC 5.3.1.16	0106	hisA 0106	hisA 0106	hisA 0106			hisA	hisA 0106	11A09-GL41 0106
imidazol glicerol fosfato desidratase EC 4.2.1.19/histidinol-fosfato fosfatase EC 3.1.3.15 enzima bifuncional	0131 0241	hisB 0131 0241	hisB 0131 hisJ 0241	hisB 0131 0241			hisB	hisB 0131 0141	11A09-GL40 0131 0241
histidinol-fosfato aminotransferase EC 2.6.1.9	0079	hisC 0079	hisC 0079	hisC hisC2 0079			hisC	hisC 0079	11A09-GL39 0079
histidinol desidrogenase EC 1.1.1.23	0141	hisD 0141	hisD 0141	hisD 0141			hisD	hisD 0141	11A09-GL38 0141
fosforibosil-ATP ciclohidrolase	Não encontrado						hisE	hisE	
Imidazol glicerol-fosfato sintase	0107	hisF 0107	hisF 0107	hisF 0107			hisF	hisF 0107	11A09-GL42 0107
ATP fosforibosiltransferase EC 2.4.2.17	0040	hisG 0040	hisG 0040	hisG 0040			hisG	hisG 0040	11A09-GL37 0040
imidazol glicerol fosfato sintase EC 2.4.2.1	0118	hisH 0118	hisH 0118	hisH 0118			hisH	hisH 0118	0.5-GL3071 0118
fosforibosil-AMP ciclohidrolase EC 3.5.4.19/fosforibosil-ATP pirofosfatase EC 3.6.1.61 enzima bifuncional	0139 0140	hisI 0139 0140	hisI 0139 0140	hisI 0139 0140			hisI	hisI 0139 0140	11A09-GL43 0139 0140

Legenda: proteínas envolvidas na biossíntese de Histidina segundo a via de *E. coli*.

II. BIOSÍNTESE DE PEQUENAS MOLÉCULAS

II. B. NUCLEOTÍDEOS

II. B. 1. Ribonucleotídeos de Purina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
adenilato quinase EC 2.7.4.3	0563	adk 0563	adk 0563	adk 0563	adk 0563		adk	adk 0563	02G12-GL63 0563
guanilato quinase EC 2.7.4.8	0194	gmK 0194	yioD 0194	gmK 0194		gmK 0194	gmK	gmK 0194	01H09-GL13 0194
GMP sintase EC 6.3.5.2	0518 0519	guaA1 0518 guaA2 0519	guaA 0518 0519	guaA 0518 0519	guaA 0518 0519	guaA 0518 0519	guaA	guaA 0518 0519	06F05-GL 30 0518 09A12-GL30 0519
inosina-5'-monofosfato desidrogenase EC 1.1.1.205	0516	guaB 0516	guaB 0516	guaB1 gua B2 guaB3 0516	guaB 0516	guaB, degenerado 0516	guaB	guaB 0516	09A12-GL32 0516
GMP redutase EC 1.6.6.8	0516	guaC 0516							
nucleosídeo difosfato quinase EC 2.7.4.6	0105	ndk 0105	ndk 0105	ndkA 0105	ndk 0105	ndk 0105		ndk 0105	0.3-GL2108 0105
fosforibosil pirofosfato sintetase EC 2.7.6.1	0462	prsA 0462	prs 0462	prsA 0462	prs 0462		prsA	prsA 0462	07C01-GL47 0462
adenilosuccinato sintetase EC 6.3.4.4	0104	purA 0104	purA 0104	purA 0104			purA	purA 0104	0.3-GL2118 0104
adenilosuccinato liase EC 4.3.2.2	0015	purB 0015	purB 0015	purB 0015			purB	purB 0015	07B05-GL35 0015
Fosforibosil aminoimidazol- succinocarboxamida sintetase EC 6.3.2.6	0152	purC 0152	purC 0152	purC 0152			purC	purC 0152	05G02-GL46 0152
glicinamida ribonucleotídeo EC 6.3.4.13	0151	purD 0151	purD 0151	purD 0151			purD	purD 0151	05F05-GL30 0151
Fosforibosil aminoimidazol carboxilase, subunidade catalítica EC 4.1.1.21	0041	purE 0041	purE 0041	purE 0041			purE	purE 0041	07C01-GL12 0041
Amido fosforibosil transferase EC 2.4.2.14	0034	purF 0034	purF 0034	purF 0034			purF	purF 0034	10H05-GL05 0034
aicar transformilase EC 2.1.2.3/ IMP ciclohidrolase EC 3.5.4.10	0138	purH 0138	purH 0138	purH 0138			purH	purH 0138	05F05-GL29 0138
Fosforibosil aminoimidazol carboxilase, subunidade ATPásica EC 4.1.1.2.1	0026	purK 0026	purK 0026	purK 0026			purK 0026	purK 0026	07C01-GL13 0026
Fosforibosil formil glicinamida sintetase EC 6.3.5.3	0046 0047	purL 0046 0047	purL 0047 purQ 0046	purL 0047 purQ 0046			purL	purL 0046 0047	07C12-GL43 0046 0047
5'-fosforibosil-5- aminoimidazol EC 6.3.3.1	0150	purM 0150	purM 0150	purM 0150			purM	purM 0150	06A08-GL39 0150

continuação									
Fosforibosil glicinamida formiltransferase EC 2.1.2.2	0299	purN 0299	purN 0299	purN 0299			purN	purN 0299	06A08-GL41 0299
Glicinamida ribonucleotídeo transformilase 2, não requer folato EC 2.1.2.2	0027	purT 0027	purT 0027	purT 0027					

Legenda: proteínas envolvidas na biossíntese de Ribonucleotídeos de Purina segundo a via de *E. coli*.

II.B.2. Ribonucleotídeos de Pirimidina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i>
Carbamoilfosfato sintase, cadeia pequena EC 6.3.5.5	0505	carA 0505	carA 0505	carA 0505					03C11-GL12 0505
Carbamoilfosfato sintase, cadeia grande EC 6.3.5.5	0458	carB 0458	carB 0458	carB 0458					03C11-GL10 0458
citidilato quinase EC 2.7.4.14	0283	cmk 0283	cmk 0283	cmk 0283	cmk 0283	cmk 0283	cmk	cmkA cmkB 0283	02A11-GL22 0283
aspartato carbamoiltransferase, cadeia catalítica EC 2.1.3.2	0540	pyrB 0540	pyrB 0540	pyrB 0540			pyrB		11A09-GL29 0540
dihidroorotase EC 3.5.2.3	0418	pyrC 0418	pyrC não encontrado	pyrC não encontrado			pyrC		07H02-GL27 0418
dihidroorotato dehidrogenase EC 1.3.3.1	0167	pyrD 0167	pyrD 0167	pyrD 0167			pyrD	pyrD 0167	02D09-GL22 0167
orotato fosforibosil transferase EC 2.4.2.10	0461	pyrE 0461	pyrE 0461	Rv0382c 0461		CP0139 0461	pyrE	pyrE 0461	10G06-GL36 0461
orotidina 5'-fosfato decarboxilase	0284	pyrF 0284	pyrF 0284	pyrF Not found			pyrF	pyrF 0284	03E01-GL54 0284
CTP sintase (UTP-amônia liase)	0504	PyrG 0504	CtrA 0504	Rv1699 0504	BB0575 0504	CPn0236 0504		HI1077 0504	XF1288 0524
Uridilato quinase	0528	PyrH 0528	SmbA 0528	Rv2883c 0528	BB0571 0528	CPn0698 0528		HI1065 0528	XF1058 0528
Aspartato carbamoiltransferase, subunidade regulatória	1781	pyrI							
uridina 5'-monofosfato sintase				umpA					

Legenda: proteínas envolvidas na biossíntese de Ribonucleotídeos de Pirimidina segundo a via de *E. coli*.

II.B.3. 2' Desoxirribonucleotídeos

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
deoxicitidina trifosfato deaminase EC 3.5.4.13	0717	dcd paxA 0717		dcd 0717		dcd 0717		dcd 0717	03D03-GL41 0717
ribonucleosídeo-difosfato redutase, cadeia alfa EC 1.17.4.1	0209	nrdA 0209	ndrE 0209	nrdZ 0209		nrdA 0209	nrdA	nrdA 0209	07C06-GL12 0209
Ribonucleosídeo-difosfato redutase, cadeia beta EC 1.17.4.1	0208	nrdB 0208	ndrF 0208	ndrB 0208		nrdB 0208	nrdB	nrdB 0208	07C06-GL13 0208
fosfohidrolase	0494	mufT 0494	mufT 0494	mufT 0494		mufT 0494		mufT 0494	03C12-gl33 0494
Anaeróbico ribonucleotídeo-trifosfato redutase, classe III EC 1.17.4.2	1328	nrdD 1328						nrdD nrdG 1328	
timidilato quinase EC 2.7.4.9	0125	trnk 0125	trnk 0125	trnk 0125	trnk 0125	trnk 0125	trnk	trnk 0125	06A08-GL49 0125
tioredoxina redutase EC 1.6.4.5	0492	trxB 0492	trxB 0492	trxB 0492	trxB 0492	trxB 0492		trxB 0492	03C12-GL26 0492

Legenda: proteínas envolvidas na biossíntese de Ribonucleotídeos de Pirimidina segundo a via de *E. coli*.

II.B.4. Salvamento de nucleosídeos e nucleotídeos

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
adenosina deaminase EC 3.5.4.4	1816	add 1816		add 1816					
AMP nucleosidase EC 3.2.2.4	0775	amn 0775	yrU 0775	Rv0091 0775		CP0972 0775			
diadenosina tetrafosfatase	0639	apaH 0639	yjbP 0639					apaH 0639	07C07-GL22 0639
Adenosina fosforibosil transferase EC 2.4.2.7	0503	apt 0503	apt 0503	apt 0503	apt 0503		apt	apt 0503	
citidina deaminase EC 3.5.4.5	0295	cdd 0295	cdd 0295	cdd 0295	cdd 0295			cdd 0295	
Citosina deaminase EC 3.5.4.1	0402	codA 0402							
Citosina/uracil/tiamina/alantoína permeases	1953	codB 1953	ywoE 1953						
2'3'-nucleotídeo-cíclico 2'-fosfodiesterase EC 3.1.4.16	0737	cpdB 0737	yfkN 0737					cpdB 0737	
Timidina fosforilase EC 2.4.2.4	0213	deoA 0213	pdp 0213	deoA 0213					
Deoxiribouratase, fosfopentomutase EC 5.4.2.7	1015	deoB 1015	drm 1015						
deoxiribose-fosfato aldolase EC 4.1.2.4	0274	deoC 0274	dra 0274	deoC 0274				deoC 0274	
Purina-nucleosídeo fosforilase	0813	deoD 0813	deoD 0813	deoD 0005				deoD 0813	

continuação									
Deoxiuridina trifosfatase EC 3.6.1.23	0756	dut 0756	yncF 0756	dut 0756		dut 0756	dut	dut 0756	10G06-GL33 0756
Guanina-xantina fosforibosiltransferase fosfotransferase EC 2.4.2.22	0503	gpt 0503	hprT 0503	apt 0503				gptB 0503	
Inosina-guanosina quinase EC 2.7.1.73	0524	gsk 0524	iolC 0524	rbsK 0524				rbsK 0524	
hipoxantina-guanina fosforibosiltransferase EC 2.4.2.8	0634	hpt 0634	hprT 0634	hpt 0634			hpt	hpt 0634	07A02-GL12 0634
Nucleosídeo permease	1972	nupC 1972	nupC 1972					HI0519 1972	
Permeases	0477	nupG 0477	presente 0477	Presente 0477	Presente 0477	Presente 0477		Presente 0477	
	Não encontrado	tsx							
timidina quinase EC 2.7.1.21	1435	tdk 1435	tdk 1435		tdk 1435			tdk 1435	
Uracil fosforibosiltransferase EC 2.4.2.9	0035	upp 0035	upp 0035	upp 0035			upp	upp 0035	
Uridina fosforilase EC 2.4.2.3	0813	udp 0813	deoD 0813					udp 0813	
Xantina/uracil permeases	2233	uraA 2233	Presente 2233					Presente 2233	
Uridina quinase	0572	udk 0572	udk 0572		BB0015 0572	CPn0735 0572		HI0132 0572	
Pirimidina-nucleosídeo fosforilase			pdp						
fosfodiesterase-nucleotídeo pirofosfatase									07H03-GL29 não encontrado
timidilato sintase EC 2.1.1.45	0207	thyA 0207	thyA 0207	thyA 0207			thyA	thyA 0207	08A04-GL27 0207
5'Metiltioadenosina /S-adenosil homocisteína nucleosidase	0775	pfs 0775	yrU 0775	Rv0091 0075	pfs_1 pfs_2 0075	CP0972 0075		pfs 0775	
Purina-nucleosídeo fosforilase	0005	xapA 0005	pnp 0005	pnp 0005					07A02-GL11 0005
5' nucleotidase									XF 2089 Não encontrado
formiltetrahidrofolato deformilase EC 3.5.1.10	0788	purU 0788	ykkE 0788	purU 0788				purU 0788	07F02-GL12 0788

Legenda: proteínas envolvidas no salvamento de nucleotídeos e nucleosídeos segundo as vias de *E. coli*.

II. BIOSÍNTESE DE PEQUENAS MOLÉCULAS

II. C. AÇÚCARES

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
Precursor de glucolactonase									02F10-GL20 não encontrado
Manosil transferase	0438	yefD 0438	YveN 0438	Rv1212c 0438				HI1698 0438	06A08-GL16 0438
UTP-glicose-1-fosfato uridil transferase							galU		

Legenda: proteínas envolvidas na biossíntese de açúcares segundo as vias de *E. coli*.

II. BIOSÍNTESE DE PEQUENAS MOLÉCULAS

II. D. COFADORES, GRUPOS PROSTÉTICOS, CARREADORES

II.D.1. Biotina

Tabela 20: Lista de genes envolvidos na via de biossíntese de Biotina

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i> ORFs
adenosilmetionina-8-amino-7-oxononanoato aminotransferase	0161	<i>bioA</i> 0161	<i>bioA</i> 0161	<i>bioA</i> 0161		<i>bioA</i> 0161	<i>bioA</i>	<i>bioA</i> 0161	05G02-GL24 0161
biotina sintetase	0502 0570	<i>bioB</i> 0502 <i>bioH</i> 0570	<i>bioB</i> 0502	<i>bioB</i> 0502		<i>bioB</i> 0502	<i>bioB1</i> <i>bioB2</i> <i>bioH</i>	<i>bioB</i> 0502	0.3-GL2743 0502 07B09-GL37 0570
Proteína de síntese de biotina	0500	<i>bioC</i> 0500		Rv0089 0500			<i>bioC</i>	<i>bioC</i> 0500	06D10-GL52 0500
detibiotina sintetase	0132	<i>bioD</i> 0132	<i>bioD</i> 0132	<i>bioD</i> 0132		<i>bioD</i> 0132	<i>bioD</i>	<i>bioD-1</i> <i>bioD-2</i> 0132	0.3-GL3346 0132
8-amino-7-oxononanoato sintase	0156	<i>bioF</i> 0156	<i>bioF</i> 0156	<i>bioF</i> <i>bioF2</i> 0156		<i>bioF</i> 0156	<i>bioF</i>	<i>bioF</i> 0156	07B09-GL36 0156
Citocromo P450	2124		<i>bioI</i> 2124	Rv0766c 2124					0.3-GL2279 10F03-GL23 2124

proteínas envolvidas na biossíntese de biotina segundo a via de *E. coli*,

II.D.2. Ácido Fólico

Tabela 21: Lista dos genes anotados na via de biossíntese de Ácido Fólico

Proteína	COG	E.coli	B. subtilis	M. tuberculosis	Borrelia burgdorferi	C. pneumoniae	Neisseria meningitidis	H. influenzae	X.fastidiosa
dihidrofolato redutase, tipo I EC 1.5.1.3	0262	folA 0262	dhfrA 0262	dhfrA 0262		folA 0262	folA	folA 0262	08A04-GL26 0262
dihidrofolato redutase, tipo III EC 1.5.1.3									
dihidroneopterin aldolase	1539	folB folX 1539	folA 1539	folX 1539		folB 1539	folB	folB 1539	07A10-GL38 1539
folilpoliglutamato sintase EC 6.3.2.17/ dihidrofolato sintase EC 6.3.2.12	0285	folC 0285	folC 0285	folC 0285			folC	folC 0285	10H05-GL02 0285
metilenotetrahidrofolato desidrogenase EC1.5.1.5/ metenil Ciclodrolase EC 3.5.4.9	0190	folD 0190	folD 0190	folD 0190	folD 0190	folD 0190	folD	folD 0190	09A12-GL33 0190
GTP ciclodrolase I EC 3.5.4.16	0302	folE 0302	mtrA 0302	folE 0302				folE 0302	05F05-GL40 0302
2-amino-4-hidroxi-6-hidroxi metil dihidropteridina pirofosfoquinase EC 2.7.6.3	0801	folK 0801	folK 0801	folK 0801		folK 0801	folK	folK 0801	01A01-GL36 11A03-GL52 0801
dihidropteroato sintase	0294	folP 0294	sul 0294	folP folP2 0294		folP 0294	folP	folP1 folP2 0294	07C10-GL44 0294
Proteína fol I							folI		
Para-aminobenzoato sintase glutamina amidotransferase, componente II EC 4.1.3.1	0512	pabA 0512	pabA 0512	pabA 0512			pabA	trpG	
Para-aminobenzoato sintase, componente I EC 4.1.3.1	0147	pabB 0147	pabB 0147	pabB 0147			pabB/pabC	trpE 0147	
4-amino-4-deoxicorismato liase EC 4.1.3.1	0115	pabC 0115	pabC 0115	pabC 0115			pabC/pabB	HI1193 0115	
Proteína de dihidrofolato biossintética			dhps						

Legenda: proteínas envolvidas na biossíntese de ácido fólico segundo a via de *E. coli*.

II.D.3. LIPOATO

Proteína	COG	E.coli	B. subtilis	M. tuberculosis	Borrelia burgdorferi	C. pneumoniae	Neisseria meningitidis	H. influenzae	X.fastidiosa
Sintetase de ácido lipóico	0320	lipA 0320	yufB 0320	lipA 0320		lipA 0320	lipA	lipA 0320	02D03-GL66 0320
Lipoato, proteína biossintética B	0321	lipB 0321		lipB 0321			lipB	lipB 0321	02D03-GL67 0321

Legenda: proteínas envolvidas na biossíntese de lipoato segundo a via de *E. coli*.

II.D.4. MOLIBDOPTERINA

Proteína	COG	E.coli	B. subtilis	M. tuberculosis	Borrelia burgdorferi	C. pneumoniae	Neisseria meningitidis	H. influenzae	X.fastidiosa
Proteína A de biossíntese do cofator molibdopterina	0535	moaA 0535	narA 0535	moaA moaA2 moaA3 0535				moaA 0535	
Proteína B de biossíntese do cofator molibdopterina	0521	moaB 0521	moaB 0521	moaB moaB2 0521				mog 0521	
Proteína de biossíntese do cofator molibdopterina	0315	moaC 0315	moaC 0315	moaC moaC2 moaC3 0315				moaC 0315	
Fator de conversão de molibdopterina, subunidade 1 [MPT]	1977	moaD 1977	moaD 1977	gphA moaD moaD2 1977				moaD 1977	
Fator de conversão de molibdopterina, subunidade 2 [MPT]	0314	moaE 0314	moaE 0314	gphA moaE moaE2 moaE3 0314				moaE 0314	
Molibdopterina-guanina dinucleotídeo, proteína biossintética A	Não encontrada	mobA	mobA mobB				mobA	mob mobB	
Proteína biossintética de molibdopterina, moeA	0303	moeA 0303	moeA 0303	moeA moeX 0303				moeA 0303	
Proteína biossintética de molibdopterina, moeB	0476 1179	moeB 0476 ygdL 1179	moeB 0476 yrvM 1179	moeB moeY 0476				moeB 0476 HI0118 1179	0.3-GL2099 0476 07B05-GL20 1179
Proteína biossintética de molibdopterina, mog	0521	mog 0521	moaB 0521	mog 0521				mog 0521	

Legenda: proteínas envolvidas na biossíntese de lipoato segundo a via de E. coli.

II.D.5. Pantotenato e Coenzima A

Tabela 23: Lista dos genes anotados na via de biossíntese de Pantotenato e CoA

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
2-dehidropantoato redutase		presente							
Panteteína fosfato adeniltransferase		presente							
Desfosfo-CoA quinase		presente							
P-pantotenocisteína decarboxilase		presente							
P-pantotenato-cisteína ligase		presente							
Pantotenato quinase	1072	coaA 1072		coaA 1071				coaA 1072	
Quetopantoato hidroximetiltransferase EC 2.1.2.11	0413	panB 0413	panB 0413	panB 0413			panB		11A03-GL53 0413
Pantoato-beta alanina ligase	0414	panC 0414	panC 0414	panC 0414			panC		0.3-GL2477 0414
aspartato 1-decarboxilase	0853	panD 0853	panD 0853	panD 0853			panD		0.3-GL2476 0853

Legenda: proteínas envolvidas na biossíntese de pantotenato e coenzimaA segundo a via de *E. coli*.

Tabela 24: Lista de genes anotados na via de biossíntese de Piridoxina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
Eritrose 4-fosfato desidrogenase		gapA							
Eritrose 4-fosfato desidrogenase		epd							
1-deoxixilulose-5-fosfato-sintase	1154	dxs 1154	yqiE 1154	dxs 1154		dxs 1154	dxs	dxs 1154	XF-11A08-GL26 1154
Fosfohidroxitreonina transaminase		serC							
Proteína biossintética de piridoxal fosfato	1995	pdxA 1995					pdxA		08A03-GL10 1995
Eritronato-4-fosfato desidrogenase	1052	pdxB 1052	yoaD 1052	Rv0728c 1052				HI1556 1052	
piridoxamina 5'-fosfato oxidase	0259	pdxH 0259		pdxH 0259			pdxH	pdxH 0259	06C10-GL41 0259
piridoxal fosfato, proteína biossintética	0854	pdxJ 0854					pdxJ		03H11-GL07 0854
Piridoxina quinase	2240	pdxK 2240			pdxK 2240			HI0405 2240	
Piridoxal quinase 2	2240	pdxY 2240						HI0405 2240	

Legenda: proteínas envolvidas na biossíntese de piridoxina segundo a via de *E. coli*.

II.D.6. PIRIDOXINA

II.D.7. Nucleotídeos de Piridina

Tabela 25: Lista dos genes anotados na via de biossíntese de Nucleotídeos de Piridina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
quinolinate sintetase A	0379	<i>nadA</i> 0379	<i>nadA</i> 0379	<i>nadA</i> 0379			<i>nadA</i>		07C09-GL33 0379
L-aspartato oxidase	0029	<i>nadB</i> 0029	<i>nadB</i> 0029	<i>nadB</i> 0029			<i>nadB</i>		07C09-GL34 0029
nicotinato mononucleotídeo pirofosforilase	0157	<i>nadC</i> 0157	<i>nadC</i> 0157	<i>nadC</i> 0157			<i>nadC</i>		03H12-GL01 0157
NH ₃ -dependente, NAD sintetase	0171	<i>nadE</i> 0171 0388	<i>nadE</i> 0171 0388	Rv2438c 0171 0388	<i>nadE</i> 0171		<i>nadE</i>		05F05-GL12 0171 0388
nicotinato fosforibosil transferase	1488	<i>pncB</i> 1488	<i>yueK</i> 1488	Rv0573c 1488			<i>pncB</i>		03C11-GL26 1488

Legenda: proteínas envolvidas na biossíntese de nucleotídeos de piridina segundo a via de *E. coli*.

Tabela 26: Lista dos genes anotados na via de biossíntese de Tiamina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
Tiamina, proteína biossintética	0422	<i>thiC</i> 0422	<i>thiA</i> 0422	<i>thiC</i> 0422			<i>thiC</i>	<i>thiC</i> No cog	0.4-GL2864 0422
fosfometilpirimidina quinase	0351	<i>thiD</i> 0351	<i>thiD</i> <i>thiK</i> 0351	<i>thiD</i> 0351			<i>thiD</i>	<i>thiD</i> 0351	07C04-GL03 0351
tiamina-fosfato pirofosforilase EC 2.5.1.3	0352	<i>thiE</i> 0352	<i>thiC</i> 0352	<i>thiE</i> 0352			<i>thiE</i>	<i>thiE</i> 0352	10F03-GL24 0352
Proteína <i>thiF</i>	1179 0476	<i>thiF</i> <i>moeB</i> 0476 <i>ygdL</i> 1179		<i>moeY</i> 0476			<i>thiF</i>	<i>moeB</i> 0476 HI0118 1179	07B05-GL20 1179
Proteína <i>thiG</i>	2022	<i>thiG</i> 2022	<i>yjbT</i> 2022	<i>thiG</i> 2022			<i>thiG</i>		03D03-GL16 2022
Proteína <i>thiH</i>	1060	<i>thiH</i> 1060		Rv1173 1060		CPn0513 CPn0911 1060			
Hidroximetil piridina quinase 2	0693	<i>thiJ</i> 0693	<i>yraA</i> 0693	Rv0052 0693	<i>thiJ</i> 0693				
tiamina quinase	Não encontrado	<i>thiK</i>							
tiamina-monofosfato quinase	0611	<i>thiL</i> 0611	<i>ydiA</i> 0611	<i>thiL</i> 0611			<i>thiL</i>	<i>thiL</i> 0611	08D10-GL36 0611
4-metil-5-beta-hidroxietiltiazol quinase	2145	<i>thiM</i> 2145	<i>thiK</i> 2145					HI04J5 2145	
1-deoxixilulose-5-fosfato-sintase	1154	<i>dxs</i> 1154	<i>yqiE</i> 1154	<i>dxs</i> 1154		<i>dxs</i> 1154	<i>dxs</i>	<i>dxs</i> 1154	11A08-GL26 1154
1-deoxi-D-xilulose 5-fosfato redutoisomerase	0743	<i>dxr</i> <i>yaeM</i> 0743	<i>yluB</i> 0743	Rv2870c 0743		<i>dxr</i> 0743	<i>dxr</i>	<i>dxr</i> 0743	01G04-GL27 0743
lipoproteína de biossíntese de tiamina	Não encontrado					<i>apbE</i>	<i>apbE</i>		

Legenda: proteínas envolvidas na biossíntese de Tiamina segundo a via de *E. coli*.

II.D.8. TIAMINA

II.D.9. Riboflavina

Tabela 28: Lista dos genes anotados na via de biossíntese de Riboflavina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
GTP ciclohidrolase II	0807	<i>ribA</i> 0807	<i>ribA</i> 0807 0108	<i>ribA</i> <i>ribA2</i> 0807 0108		<i>ribA+ ribB</i> 0807 0108	<i>ribA</i> <i>ribA+ribB</i>	<i>ribA</i> 0807	08D10-GL39 11A02-GL42 0108 0807
3,4-dihidroxi-2-butanona-4-fosfato	0108	<i>ribB</i> 0108	<i>ribA</i> 0108 0807	<i>ribA</i> <i>ribA2</i> 0108 0807		<i>ribB+ ribA</i> 0108 0807	<i>ribA+ribB</i>	<i>ribB</i> 0108	08D10-GL39 11A02-GL42 0108 0807
Deaminase riboflavina específica/ 5-fosforibosilamino uracil redutase	0117 1985	<i>ribD</i> 0117 1985	<i>ribG</i> <i>ribT</i> 0117 1985	<i>ribD</i> <i>ribG</i> 0117 1985		<i>ribG</i> CP0998 0117 1985	<i>ribD</i>	<i>ribD</i> 0117 1985	08D10-GL43 0117 0.3-GL0207 1985
riboflavina sintase, cadeia alfa	0307	<i>ribE</i> 0307	<i>ribB</i> 0307	<i>ribC</i> 0307		<i>ribE</i> 0307	<i>ribE</i>	<i>ribE</i> 0307	08D10-GL40 0307
Riboflavina sintase, cadeia beta	0054	<i>ribH</i> 0054	<i>ribH</i> 0054	<i>ribH</i> 0054		<i>ribH</i> 0054	<i>ribH</i>	<i>ribH</i> 0054	XF-08D10-GL38 0054
Riboflavina quinase/ FMN adeniltransferase	0196	<i>ribF</i> 0196	<i>ribC</i> <i>ribR</i> 0196	<i>ribF</i> 0196		<i>ribF</i> 0196	<i>ribF</i>	<i>ribF</i> 0196	09A12-GL18 0196

Legenda: proteínas envolvidas na biossíntese de Riboflavina segundo a via de *E. coli*.

II.D.10. Tiorredoxina, Glutaredoxina, Glutatião

a. Glutatião

Tabela 28: Lista dos genes anotados na via de biossíntese de Tiorredoxina, Glutaredoxina e Glutamina

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>N. meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i> ORFs
gamma-glutamyl transpeptidase	0405	<i>ggt</i> 0405	<i>ggt</i> 0405	<i>ggtA</i> 0405 <i>ggtB</i> 0405			<i>ggt</i>		07H02-GL33 0405
glutathione reductase, NADPH	1249	<i>gor</i> 1249	<i>pdhD</i> 1249	<i>gorA</i> 1249		CPn0833 1249		<i>gor</i> 1249	
glutaredoxina	0695	<i>grxA</i> 0695	<i>ytnI</i> 0695	<i>nrdH</i> 0695				<i>grx</i> 0695	
glutaredoxina 2	0695	<i>grxB</i> 0695						<i>grx</i> 0695	
glutaredoxina 3	0695	<i>grxC</i> 0695 <i>yibM</i> 0695							07H03-GL20 0695
"Glutaredoxin like protein" / "Rhodanese-related sulfurtransferases"	0278 0607	<i>ydhD</i> 0278 <i>glpE</i> 0607	<i>yqhl</i> 0607	<i>Rv3116</i> 0607	<i>glpE</i> 0607			<i>HI1165</i> 0278 <i>glpE</i> 0607	05B03-GL11 0278 0607
glutamato-cisteína ligase		<i>gshA</i>		<i>gshA</i>			<i>gshA</i>		07C12-GL50 não encontrado
glutatião sintetase	0189	<i>gshB</i> 0189					<i>gshB</i> 0189		05F05-GL-05 0189
Glutatião transferase	0625	<i>gst</i> 0625						<i>bphH</i> 0625	
"glutaredoxin-like protein", doadora de hidrogênio	Não encontrado	<i>nrdH</i>		<i>nrdH</i>					
tiorredoxina	0526	<i>trxA</i> 0526	<i>trxA</i> 0526	<i>trxC</i> 0526	<i>trxA</i> 0526	<i>trxA</i> 0526		<i>trxM</i> 0526	07C06-GL15 07B07-GL35 0526
Tiorredoxina 2	0526	<i>trxC</i> 0526 <i>yfiG</i> 0526		<i>trxA</i> 0526					
coenzima F390 sintetase	1541	1787664 1541							07C09-GL29 1541

Legenda: proteínas envolvidas na biossíntese de tiorredoxina, glutaredoxina e glutamina segundo as vias de *E. coli*.

II.D.11. Menaquinona e Ubiquinona

a. Menaquinona

b. Ubiquinona

Tabela 29: Lista de genes anotados na via de biossíntese de Ubiquinona e Menaquinona

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i> ORFs
geraniltransferase/farnesil-difosfato sintase	0142	ispA 0142	yqiD 0142	idsA idsA2 0142	ispB 0142	CPn0748 0142	ispA	ispA 0142	07C04-GL65 0142
octaprenil-difosfato sintase	0142	ispB 0142	0142	idsB 0142	ispB 0142	0142	ispB	ispB 0142	07C12-GL03 0142
4-dihidroxi-2-naftoato octapreniltransferase	1575	menA 1575	ywaB 1575	menA 1575				menA 1575	
naftoato sintase	0447	menB 0447	menB 0447	menB 0447				menB 0447	
O-succinilbenzoato-CoA sintase	1441	menC 1441	menC 1441	menC 1441				menC 1441	
2-succinil-6-hidroxi-2,4-ciclohexadieno-1-carboxilato sintase	1165	menD 1165	menD 1165	menD 1165				menD 1165	
O-succinilbenzóico ácido-CoA ligase	2228	menE 2228	menE 2228	menE 2228				menE 2228	
menaquinona-específica isocorismato sintase	1169	menF 1169	menF 1169	entC 1169				menF 1169	
S-adenosilmetionina 2-demetilmenaquinona metiltransferase	0684	menG 0684		menG 0684				menG 0684	
hidroxibenzoato octapreniltransferase	0382	ubiA 0382				ubiA 0382			0.3-GL2737 0382
corismato piruvato liase	Não encontrado	ubiC							
ubiquinona/menaquinona transferase	2226	ubiE 2226	gerCB 2226	ubiE 2226		ubiE 2226	ubiE		04A09-GL27 2226
2-octaprenil-3-metil-6-metoxi-1,4-benzoquinona hidrolase		ubiF							
3,4-dihidroxi-5-hexaprenilbenzoato metiltransferase	2227	ubiG 2227					ubiG		0.3-GL3353 2227
2-octaprenil-6-metoxifenol hidroxilase	0654	ubiH 0654	yhjG 0654	Rv1751 0654			ubiH		08A03-GL14 0654
octaprenil-4-hidroxibenzoato carboxilase		ubiD ubiX					ubiX 3		
Proteína de biossíntese de ubiquinona	0661	aarF não encontrado yigR 0661		Rv0647c 0661			aarF		07F02-GL14 0661 07B05-GL10 não encontrado
heptaprenil difosfato sintase, componente I			gerCA I						
heptaprenil difosfato sintase, componente II			gerCC 0142						

Legenda: proteínas envolvidas na biossíntese de tireodoxina, glutaredoxina e glutatona segundo as vias de *E. coli*.

II.D. 12. Heme e Siroheme

Tabela 30: Lista dos genes anotados nas vias de biossíntese de Heme e Siroheme

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i> ORFs
siroheme sintase	0007 1648	cysG 0007 1648	nasF 0007 1648	cysG 0007 1648			cysG1 cysG2		08A03-GL17 0007 1648
Heme O-sintase		cyoE							
glutamil-tRNA redutase	0373	hemA 0373	hemA 0373	hemA 0373		hemA 0373	hemA		07C01-GL43 0373
delta-aminolevulínico ácido desidratase	0113	hemB 0113	hemB 0113	hemB 0113		hemB 0113	hemB		11A07-GL02 0113
porfobilinogênio deaminase	0181	hemC 0181	hemC 0181	hemC 0181		hemC 0181	hemC		07G02-GL15 0181
uroporfirinogênio-III sintase	1587	hemD 1587	hemD 1587	Rv0511 1587		CPn0522 1587	hemD		09C12-GL27 1587
uroporfirinogênio III decarboxilase	0407	hemE 0407	hemE 0407	hemE 0407		hemE 0407	hemE		06C10-GL36 0407
coproporfirinogênio III oxidase, aeróbica	0408	hemF 0408							03E01-GL23 0408
protoporfirinogênio oxidase EC1.3.3.4	0716	hemG 0716	ykuN 0716					H10191 0716	
Protoheme ferro-liase	0276	hemH 0276	hemH 0276	hemI 0276		hemH 0276	hemH	hemH 0276	07A01-GL58 0276
protoporfirinogênio oxidase	0500	hemK 0500	ywkE 0500	hemK 0500	hemK 0500	CPn0114 0500		hemK 0500	01H09-GL23 0500
glutamato-1-semialdeído 2,1-aminomutase	0001	hemL 0001	gsaB 0001	hemL 0001		hemL 0001	hemL		11A07-GL08 0001
coproporfirinogênio III oxidase, anaeróbica	0635	hemN 0635	hemN 0635	hemN 0635	hemN 0635	hemN 0635	hemN	hemN 0635	
Protoporfirinogênio oxidase	1232	hemY Not found	hemY 1232	hemY 1232		hemY 1232			
glutamil-tRNA sintetase	0008	glfX 0008	glfX 0008	glfS 0008	glfX 0008	glfX 0008		glfX 0008	08A03-GL34 0008
Proteína de biossíntese de porfirina	0457	hemY não encontrado							11H05-GL49 0457

Legenda: proteínas envolvidas na biossíntese de heme e siroheme segundo as vias de *E. coli*.

II.D.14. Cobalamina

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i>
cobinamida adenosiltransferase/ cobalamina adenosiltransferase	2109	btuR 2109		cobA 2109					
cobalamina 5'-fosfato sintase/ cobalamina sintase	0368	cobS 0368		cobS 0368					
dimetilbenzimidazol fosforibosiltransferase	2038	cobT 2038		cobT 2038					
cobinamida quinase/cobinamida-P guanilil transferase	2087	cobU 2087		cobU 2087					

Legenda: proteínas envolvidas na biossíntese cobalamina segundo as vias de *E. coli*.

II.D.15. Enteroquelina

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
2,3-dihidro-2,3-dihidroxibenzoato desidrogenase	1028	entA 1028	dhbA 1028	Rv1941 1028		CPn0296 1028		H10155 1028	
Isocorismatase	1535	entB 1535	dhbB 1535						
Isocorismato sintase, específico de enterobactina	1169	entC 1169	dhbC 1169	entC 1169				menF 1169	
Fosfopanteteinil transferase, subunidade de enterobactina /enteroquelina sintase (complexo multienzimático)	Not found	entD							
Subunidade E de 2,3-dihidroxibenzoato-AMP ligase e enterobactina sintase (complexo multienzimático)	1021	entE 1021	dhbE 1021	RV2384 1021					
Enzima serina-ativadora, subunidade de enterobactina sintase (complexo multienzimático)	1020	entF 1020	dhbF 1020	RV2380c 1020					

Legenda: proteínas envolvidas na biossíntese enteroquelina segundo as vias de *E. coli*.

II.D.16. Biopterina

Proteína	COG	<i>E. coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X. fastidiosa</i>
6-piruvil tetrahydrobiopterina sintase	0720	YgcM 0720	YkvK 0720				ygcM	YgcM 0720	05G02-GL31 0720
pterina-4-alpha-carbinolamina desidratase	2154			moaB 2154					07H03-GL35 2154

Legenda: proteínas envolvidas na biossíntese biopterina.

II. BIOSÍNTESE DE PEQUENAS MOLÉCULAS

II.E. Ácidos Graxos e Fosfatídicos

Tabela : Lista dos genes anotados nas vias de biossíntese de ácidos graxos e fosfatídicos

Proteína	COG	E.coli	B. subtilis	M. tuberculosis	Borrelia burgdorferi	C. pneumoniae	Neisseria meningitidis	H. influenzae	X.fastidiosa
2-acilglicerofosfoetanola mina aciltransferase/acil-ACP sintetase	Não encon- trado	aac							
acetil-CoA carboxilase carboxiltransferase , subunidade alfa	0825	accA 0825	accA 0825	accD 0825 0777		aacA 0852	accA1 accA2	accA 0825	05G02-GL44 0825
BCCP-subunidade de biotina de "carboxil carrier protein" e acetil CoA carboxilase	0511	accB 0511	accB 0511	accA3 0511		aacB 0511	accB	accB 0511	03H11-GL24 0511
biotina carboxilase de acetil CoA carboxilase , cadeia alfa	0439	accC 0439	accC 0439	accA1 0439		aacC accC 0439	accC	accC 0439	03H11-GL23 0439
acetil-coenzima A carboxilase carboxil transferase , cadeia beta	0777	accD 0777		accD1 accD2 accD3 accD4 accD5 0777		aacD 0777	accD	accD 0777	0.3-GL0624 0777
"acyl carrier protein" fosfodiesterase	1182	acpD 1182						acpD 1182	
"acyl carrier protein"	0236	acpP 0236	acpA 0236	acpM 0236	BB0704 0236	acpP 0236	acp1 acp2	acpP 0236	09G12-GL11 0236
"acyl carrier protein"	0236								0.5-GL1098 0236
"holo-[acyl-carrier- protein] sintase"	0736	acpS dpj 0736		acpS 0736	BB0010 0736	acpS 0736	acpS		
Acil-coenzima A sintetase	0365	acs 0365	acsA 0365	acs 0365					
CDP-diacilglicerol pirofosfatase	2134	cdh 2134		cdh 2134					
CDP-diglicerídeo sintetase	0575	cdsA 0575	cdsA 0575	cdsA 0575	cdsA 0575	cdsA 0575	cdsA	cdsA 0575	01G04-GL28 0575
Ciclopropano acil fosfolípido sintase	2230	cfa 2230		Rv3392c 2230					
diacilglicerol quinase	0818	dgaA 0818	dgaA 0818				dgaA	dgaA 0818	
"beta-hidroxidecanoil- acyl carrier protein dehidratase"	0764	fabA 0764				CPn0651 0764		fabA 0764	07A01-GL46 0764
beta-quetoacil-ACP sintase I/ malonil-ACP decarboxilase	0304	fabB 0304		kasA kasB 0304		CPn0916 0304		fabB 0304	

continuação									
"malonil CoA-acil carrier protein transacilase"	0331	<i>fabD</i> 0331	<i>fabD</i> 0331	<i>fabD</i> 0331		<i>fabD</i> 0331	<i>fabD</i>	<i>fabD</i> 0331	09G12-GL14 0331
"3-oxoacil-[acil-carrier-protein] sintase II beta-ketoacyl-acp synthase II"	0304	<i>fabF</i> 0304		<i>kasA</i> 0304		<i>fabF</i> 0304	<i>fabF1</i> <i>fabF2</i>	<i>fabB</i> 0304	09G12-GL10 07G02-GL31 0304
"3-oxoacil[acyl-carrier-protein] redutase"	1028	<i>fabG</i> 1028	<i>fabG</i> 1028	<i>fabG1</i> 1028		<i>fabG</i> 1028	<i>fabG1</i> <i>fabG2</i>	<i>fabG</i> 1028	05G02-GL06 09G12-GL12 1028
"oxoacil-[acyl-carrier-protein] sintase III"	0332	<i>fabH</i> 0332		<i>fabH</i> 0332	BB0683 0332	<i>fabH</i> 0332	<i>fabH</i>	<i>fabH</i> 0332	11B08-GL06 05F05-GL21 0332
enol-ACP redutase (NADH)	0623	<i>fabI</i> 0623		<i>inhA</i> 0623		<i>fabI</i> 0623	<i>fabI</i>	<i>fabI</i> 0623	
beta-hidroxiacil-ACP desidratase	0764	<i>fabZ</i> 0764				<i>fabZ</i> 0764	<i>fabZ</i>	<i>fabZ</i> 0764	01G04-GL23 0764
acetoacetil-CoA redutase	1028	<i>fabG</i>							01A03-GL02 1028

Legenda: proteínas envolvidas na biossíntese de ácidos graxos e fosfatídicos segundo as vias de *E. coli*.

II. BIOSÍNTESE DE PEQUENAS MOLÉCULAS

II.F. Poliaminas

Tabela : Lista dos genes anotados na via de biossíntese de poliaminas

Proteína	COG	<i>E.coli</i>	<i>B. subtilis</i>	<i>M. tuberculosis</i>	<i>Borrelia burgdorferi</i>	<i>C. pneumoniae</i>	<i>Neisseria meningitidis</i>	<i>H. influenzae</i>	<i>X.fastidiosa</i>
Glutathionil espermidina sintetase/glutathionil espermidina amidase	0754	<i>gsp</i> 0754						H10929 0754	
Arginina descarboxilase, biossintética	1166	<i>speA</i> 1166	<i>speA</i>				<i>speA</i>		04-GL0249 1166
agmatinase	0010	<i>speB</i> 0010	<i>speB</i> não encontrado						
Ornitina descarboxilase	1982	<i>speC</i> 1982	<i>speA</i> 1982	<i>adi</i> 1982				<i>speF</i> 1982	
S-adenosil metionina descarboxilase	1585	<i>speD</i> 1585							07B05-GL11 1585
espermidina sintase	0421	<i>speE</i> 0421	<i>speE</i> 0421	<i>speE</i> 0421			<i>speE</i>		07A09-GL53 0421
Ornitina descarboxilase, indutível	1982	<i>speF</i> 1982	<i>speA</i> 1982	<i>adi</i> 1982				<i>speF</i> 1982	
Espermidina acetiltransferase		<i>speG</i>							
Lisina descarboxilase		<i>cadA</i>							

Legenda: proteínas envolvidas na biossíntese de poliaminas segundo as vias de *E. coli*.

Xylella fastidiosa ORF XF1001 (gene argB)

Modelo	Seq-de	Seq-até	HMM-de	HMM-até	Valor	E-value	Descrição
!! aaquinase	38	232	1	248	82.2	4.4e-22	Família Aminoácido quinase
Acetiltransf	334	404	1	71	5.8	0.082	Família Acetiltransferase (GNAT)

aaquinase 38-232
Acetiltransferase 334-404

Escherichia coli (gene argA)

Modelo	Seq-de	Seq-até	HMM-de	HMM-até	Valor	E-value	Descrição
!! aaquinase	26	237	1	248	83.5	1.8e-22	Família Aminoácido quinase
!! Acetiltransf	342	414	1	71	41.1	1.1e-09	Família Acetiltransferase (GNAT)

aaquinase 26-237
Acetiltransferase 342-414

Escherichia coli (gene argB)

Modelo	Seq-de	Seq-até	HMM-de	HMM-até	Score	E-value	Descrição
Smr	7	72	1	81	-22.2	0.52	Domínio Smr
!! aaquinase	3	202	1	248	106.2	2.6e-29	Família Aminoácido quinase

aaquinase 3-202

Bacillus subtilis (gene *argJ*)

	Modelo	Seq-de	Seq-até	HMM-de	HMM-até	Valor	E-value	Descrição
!!	ArgJ	22	406	1	405	866.4	3.8e-258	Família <i>argJ</i>

ArgJ 22-406

Bacillus subtilis (gene *argB*)

	Modelo	Seq-de	Seq-até	HMM-de	HMM-até	Valor	E-value	Descrição
!!	<i>aa</i> quinase	3	200	1	248	174.4	7.9e-50	Família Aminoácido quinase

aaquinase 3-200

OBSERVAÇÃO

NÃO FOI AUTORIZADA A INCLUSÃO DO(S)
ARTIGO(S) NESTE ARQUIVO