Bioinformática e anotação de genes em *Xanthomonas axonopodis* pv. *citri* e *Xylella fastidiosa*: metabolismo de ferro e biossíntese de pequenas moléculas

Eduardo Fernandes Formighieri

Dissertação apresentada ao Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, para obtenção do título de Mestre em Ciências, Área de Concentração: Energia Nuclear na Agricultura.

P I R A C I C A B A
Estado de São Paulo – Brasil
Março – 2002

Bioinformática e anotação de genes em *Xanthomonas axonopodis* pv. *citri* e *Xylella fastidiosa*: metabolismo de ferro e biossíntese de pequenas moléculas

EDUARDO FERNANDES FORMIGHIERI
Engenheiro Agrônomo

Orientador: Profa. Dra. **SIU MUI TSAI**

Dissertação apresentada ao Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, para obtenção do título de Mestre em Ciências, Área de Concentração: Energia Nuclear na Agricultura.

P I R A C I C A B A
Estado de São Paulo – Brasil
Março - 2002

Dados Internacionais de Catalogação na Publicação (CIP)
Seção Técnica de Biblioteca - CENA/USP

Formighieri, Eduardo Fernandes
Bioinformática e anotação de genes em *Xanthomonas axonopodis* pv. *citri* e *Xylella fastidiosa*: metabolismo de ferro e biossíntese de pequenas moléculas / Eduardo Fernandes Formighieri. - - Piracicaba, 2002.
177p. : il.

Dissertação (mestrado) - - Centro de Energia Nuclear na Agricultura, 2002.

1. Biologia molecular 2. Genomas 3. Sequência de aminoácidos 4. Sequenciamento I. Título

CDU 577.21

Dedico

A meus pais, Gentil e Márcia,
exemplos de tudo de bom que poderia desejar
e do que nem sabia que existia.
Pela luta de quem do nada criou o próprio sucesso.
Pela coragem de quem deixou tudo pelos filhos, e soube recuperar.
Principalmente, pelo amor.

A meus irmãos, Érica e Paulo, sempre presentes e brilhantes.
Imprescindíveis.

À minha namorada, Luciana, pela enorme paciência,
pelo carinho e amor. Por existir e estar ao meu lado.

Ao meu "irmão" Paulo César, sua avó Conceição e à "filha" Rúbia,
pelo exemplo de força e determinação.

Aos amigos e amigas,
Que são tudo.

E especialmente em memória de

Luiza e Pedro Valdir Formighieri,
eternos exemplos de que pode-se lutar uma vida toda
contra tudo e para todos, de cabeça erguida,
e um sorriso no rosto.

E de Rafael Palmero,
pela eterna alegria ☺

Dedico à saudade.

Agradecimentos

Agradeço

A Deus, e a todos os meus outros amigos, que tornaram minha vida a 1.100 km de casa possível.

À minha família, a mais especial do mundo. E à Lu.

À Dra. Siu Mui Tsai, pela amizade, confiança e orientação. Também pelo carinho materno.

Ao Dr. David Henry Moon, pela amizade, orientação, apoio e disponibilidade no que fosse preciso.

Às amigas especiais (em ordem alfabética) Aline Souza, Elena Perez, Lin Saito, Linda Lin Lee, Mariana Crepaldi de Paula e Rúbia Soares, por me ajudarem tanto, e sempre.

À Dra. Marília Caixeta Franco, pela orientação, apoio e paciência.

Ao amigo Luis Fernando Manesco, pelo apoio dentro e fora do expediente.

Aos velhos amigos daqui: Edenilson, Eduardo, Elaine, Fabiana, Matheus e Renata. Pela amizade e diversão.

Aos pesquisadores, funcionários e alunos do CENA, por tudo. Ao CENA, pela acolhida e paciência.

Ao Dr. Luiz Carlos E. Rodriguez, e ao colega Leandro A. V. Pinheiro, por me levarem ao mundo da pesquisa.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela bolsa concedida.

Enfim, a todos os que me ajudaram, ou complicaram minha vida, nesta época de profundo crescimento.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Siglas, Abreviações e Símbolos

AMP – Adenosina-5'-fosfato (*Adenosine-monophosphate*);

ATP – Adenosina tri-fosfato (*Adenosine-5'-triphosphate*);

BLAST – Ferramenta de Procura de Alinhamento Local Básica (*Basic Local Alignment Search Tool*);

CMP – Citidina-5'-fosfato (*Cytidine-monophosphate*);

COGs – Agrupamentos de Grupos Ortólogos de proteínas (*Clusters of Orthologous Groups of proteins*);

CVC – Clorose Variegada dos Citros;

DNA – Ácido Desoxirribonucléico (Deoxyribonucleic Acid);

EC# – Número EC (*Enzyme Classification number*);

GMP – Guanosina-5'-fosfato (*Guanosine-monophosphate*);

Hpa – Associado a hrp (*Hrp Associated*);

Hrc – Hrp conservado (Hrp Conserved);

Hrp – Resposta de Hipersensibilidade e Patogenicidade (*Hypersensitive Reaction and Pathogenicity*);

IMP – Inosina 5'-monofosfato (*Inosine-monophosphate*);

KEGG – Enciclopédia de Kyoto para Genes e Genomas (*Kyoto Encyclopedia of Genes and Genomes*);

MacOS – Sistema Operacional dos Computadores Macintosh (*Macintosh Operational System*);

NCBI – Centro Nacional para Informação em Biotecnologia (*National Center for Biotechnology Information*);

ONSA – Organização para Sequenciamento e Análise de Nucleotídeos (*Organization for Nucleotide Sequencing and Analysis*);

ORF – Quadro Aberto de Leitura (*Open Reading Frame*);

OSS – Sequenciamento de shotguns ordenado (*Ordered Shotgun Sequencing*);

PAUP – Análise Filogenética Utilizando Parcimônia (*Phylogenetic Analysis Using Parsimony*);

PC – Computador Pessoal (Personal Computer);

PRPP – Alfa-5-fosforribosil-1-fosfato (*5-phosphoribosyl-1-pyrophosphate*);

rDNA – Ácido Desoxirribunucléico Ribossômico (*Ribosomal Dexoryribonucleic acid*);

rRNA – Ácido Ribonucléico Ribossômico (*Ribosomal Ribonucleic Acid*);

SMART – Ferramenta de Pesquisa de Arquitetura Modular Simples (*Simple Modular Architecture Research Tool*);

TC# – Número TC (*Transport Classification number*);

TE – Elemento Transponível (*Transponible Element*);

tRNA – Ácido Ribonucléico transportador (*Transfer Ribonucleic Acid)*;

UMP – Uridina-5'-fosfato (*Uridine-monophosphate*);

Xcamp – *Xanthomonas campestris* pv. *campestris*;

Xcitri – *Xanthomonas axonopodis* pv. *citri*;

XCVC – *Xylella fastidiosa*, relacionada à Clorose Variegada dos Citros;

XPD – *Xylella fastidiosa*, relacionada à doença da uva (*Pierce's Disease*);

Bioinformática e anotação de genes associados ao ferro e biossíntese de pequenas moléculas em *Xylella fastidiosa* e *Xanthomonas axonopodis* pv. *citri*

Autor: Eduardo Fernandes Formighieri
Orientador: Profa. Dra. Siu Mui Tsai

RESUMO

Nos últimos anos, o sequenciamento de diversos organismos tem gerado uma grande quantidade de dados biológicos. Para possibilitar o armazenamento, gerenciamento e disponibilização destes dados de forma a potencializar ao máximo sua utilização, a bioinformática se desenvolveu conjuntamente, e numa velocidade também impressionante. O resultado desta parceria entre a biologia molecular e a informática pode ser visto na quantidade crescente de genomas completos sendo seqüenciados, e também no sucesso do Programa de Sequenciamento de Genomas (ONSA-FAPESP), que levou o Brasil ao pequeno grupo de países com esta capacidade científica. O desenvolvimento e adaptação de ferramentas computacionais e a formação de recursos humanos têm apresentado crescimento constante no Brasil e no mundo.

Este estudo envolve o desenvolvimento de bioinformática através da adaptação desta ao Laboratório de Biologia Celular e Molecular do CENA/USP, de anotação de genes nas bactérias *Xylela fastidiosa* (XCVC), *Xanthomonas axonopodis* pv. *citri* (Xcitri), *Xanthomonas campestris* pv. *campestris* (Xcamp) e

Xylella fastidiosa / *Pierce's Disease* (XPD) e de estudo de genes relacionados ao metabolismo do ferro, que inclui estudo filogenético do gene fur e de alguns receptores dependentes do sistema TonB. O gene fur mostrou-se suficientemente conservado para ser útil na comparação de grupos próximo de bactérias. No caso dos receptores, uma maior divergência foi encontrada, indicando uma grande variação entre receptores da mesma classe, e algumas possíveis falhas no processo de anotação ou novas classes de receptores fecA em *Xanthomonas*.

A anotação em Xcitri envolveu parte da categoria Processos Celulares, e parte da categoria Patogenicidade, Virulência e Adaptação. A anotação em XPD incluiu a categoria Biossíntese de Pequenas Moléculas, sendo feita uma comparação entre as quatro bactérias citadas. Entre XCVC e XPD não existem diferenças significativas nesta classe de genes, o mesmo ocorrendo entre Xcitri e Xcamp. O mesmo foi refletido na totalidade dos genomas. Na classe estudada, as espécies de *Xanthomonas* apresentaram alguns genes a mais que os isolados de *Xylella*.

Bioinformatics and annotation of the genes associated with Iron and the biosynthesis of small molecules in Xylella fastidiosa and Xanthomonas axonopodis pv. citri.

Author: Eduardo Fernandes Formighieri
Advisor: Profa. Dra. Siu Mui Tsai

SUMMARY

In recent years the sequencing of diverse organisms has generated a large quantity of biological data. To facilitate the storage, management and availability of this data in a form that maximizes its utilization bioinformatics has had to develop at an equally astonishing rate. The result of this partnership between molecular biology and informatics can be clearly seen by the increasing number of complete genomes being sequenced and also the success of the Genome Sequencing Program (ONSA-FAPESP) uniting Brazil with the small group of elite countries with this capacity. The development and adaptation of computational tools and the formation of human resources has presented constant growth in Brazil and worldwide.

This study involves the development of bioinformatics through the adaptation of this laboratory, Cellular and Molecular Biology at CENA/USP, for the annotation of genes from the bacteria Xylella fastidiosa from citrus (XCVC) and grape vines (XPD) and Xanthomonas axonopodis pv, citri (Xcitri) and Xanthomonas campestris pv. campestris (Xcamp). Genes associated with iron metabolism were also studied, including phylogentic analysis of the fur gene

and some of the TonB dependent receptors. The level of conservation of the fur gene was shown to be adequate for the comparison of related groups of bacteria. In the case of the receptors a greater divergence was observed indicating a large variation between receptors of the same class and a few possible annotation errors or a new class of FecA receptor found in Xanthomonas.

Annotation of Xcitri involved the categories for cellular processes and pathogenicity, virulence and adaptation. The annotation of XPD included the biosynthesis of small molecules and this category was used to compare all four of the bacterial genomes. Between XCVC and XPD no significant differences within this gene category were observed, the same was found between Xcitri and Xcamp. This situation was reflected when the entire genomes were compared. More genes were observed in this class in Xanthomonas when compared with Xylella.

Bioinformatique et annotation de gènes associés au fer et biosynthèse de petites molécules en *Xylella fastidiosa* e *Xanthomonas axonopodis* pv. *citri*

Auteur: Eduardo Fernandes Formighieri
Directeur: Profa. Dra. Siu Mui Tsai

RÉSUMÉE

Dans les dernières années, le séquençage de diverses organismes a généré une grande quantité de donnés biologiques. Pour rendre possible la stockage, manipulation et disponibilité de ces donnés de façon a potentialiser au maximum son utilisation, la bio informatique s'a développée en même temps et dans une vélocité aussi impressionnante. Le résultat de ce partenariat entre la biologie moléculaire et l'informatique peut être vu dans la quantité grandissante de génomes complets en train d'être sequénces, et aussi dans le succès du Programme de Séquençage de Génomes (ONSA-FAPESP) qui a emmené le Brésil au petit nombre de pays avec cette capacité scientifique . Le développement et adaptation des outils informatiques et la formation de ressources humains s'agrandisse constamment au Brésil comme dans le monde.

Cet étude comprends le développement de bio informatique à travers de son adaptation au laboratoire de biologie cellulaire et moléculaire du CENA/USP, de la annotation des gènes dans les bactéries *Xylela fastidiosa* (XCVC), *Xanthomonas axonopodis* pv. *citri* (Xcitri), *Xanthomonas campestris* pv. *campestris* (Xcamp) et *Xylella fastidiosa* / *Pierce's Disease* (XPD) et de l'étude

des gènes relatione au métabolisme du fer, qui incluse l'étude phylogénétique du gène fur et de quelques récepteurs dépendants du système TonB. Le gène fur s'a montré suffisamment conservée pour être utile dans la comparaison de groupes proches de bactéries. Dans le cas des récepteurs, une plus grande divergence a été rencontrée, ce qui indique une grande variation entre les récepteurs d'une même classe, et quelques possibles failles dans le processus de notation ou des nouvelles classes de récepteurs fecA en *Xanthomonas*.

La notation en Xcitri a inpliqué une partie de la catégorie Processus Cellulaires, et une partie de la catégorie Pathogènicité, Virulence et Adaptation. La notation en XPD a inclus la catégorie Biosynthèse de Petites Molécules, une comparaison ayant faite entre les quatre bactéries citées. Entre XCVC et XPD il n'existent pas des différences significatives dans cette classe de gènes, le même occurrent entre Xcitri et Xcamp. Le même a été reflété dans la totalité des génomes. Dans la classe étudiée, les espèces de *Xanthomonas* ont présentée quelques gènes de plus que ceux isolées de *Xylella*.

1. Introdução

Uma área da biologia molecular que tem evoluído espetacularmente nesses últimos anos tem sido a bioinformática, que permite a aplicação da tecnologia de informação ao gerenciamento de dados biológicos (Gibas & Jambeck, 2001). Em outros termos, a bioinformática é a aplicação da informática na biologia molecular, e consiste na utilização e no desenvolvimento de ferramentas computacionais para estudo e resolução de problemas biológicos. Ela permite o desenvolvimento de pesquisa de ponta aumentando velocidade e qualidade de informações, através da análise de grande quantidade de dados com grande precisão e tempo reduzido.

Essa forma de integração entre a pesquisa prática e a máquina tem gerado diversas ferramentas poderosas que têm sido disponibilizadas para gerenciar, acessar e apresentar os dados biológicos de forma inteligível e funcional.

O estudo de genomas, ou genômica, tem sido um dos campos de maior desenvolvimento da bioinformática no Brasil e no mundo. Houve um grande salto neste campo com o projeto de sequenciamento do genoma humano, tanto em equipamentos quanto em programas e recursos humanos. O Brasil, com o Projeto Genoma da *Xylella fastidiosa* (Simpson et al., 2000) e posteriormente, de outras bactérias – *Xanthomonas axonopodis* pv. *citri*, *Xanthomonas campestris* pv. *campestris* (Silva et al., 2002), *Xylella fastidiosa / Pierce's Disease* e *Leifsonia xyli*, ingressou no seleto grupo dos países que dominam esta tecnologia com a proposta inovadora do trabalho interligado através da

internet (Rede ONSA[1]). Esta opção, compatível com a estrutura brasileira, foi viabilizada principalmente pela Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP e pela bioinformática adaptada e desenvolvida no Instituto de Computação da Universidade de Campinas – SP (UNICAMP).

A parte final e mais importante do sequenciamento de um genoma é a anotação. Anotar é postular uma função ao produto de um ORF, o que define genes potenciais, com a maior chance de acerto possível. Para tanto, os ORFs passam por inúmeras revisões, em diferentes fases: anotação automática, verificação inicial, verificação mais detalhada, e anotação baseada na categoria. A última envolve uma verificação dos genes presentes e posterior anotação por vias metabólicas.

O estudo filogenético, ou estudo de relações evolutivas, é uma das formas de utilização de dados de sequenciamento, já que estas informações permitem um estudo mais preciso do que características morfológicas e fisiológicas.

Neste estudo, a análise dos dados incluiu desde a submissão de clones seqüenciados até a anotação de ORFs em diferentes genomas. Nas fases finais, genes associados à Biossíntese de Pequenas Moléculas foram comparados entre as espécies anotadas, e foi realizado estudo filogenético de alguns genes relacionados ao metabolismo do ferro.

[1] Página da rede ONSA no site da FAPESP – http://watson.fapesp.br/onsa/Genoma3.htm

Objetivos

- ✓ Instalar e adequar sistemas de bioinformática para análise e montagem de seqüências de DNA dos *genomas de Xylella fastidiosa, Xanthomonas axonopodis* pv. *citri*, *Xanthomonas campestris* pv. *campestris*, *Xylella fastidiosa* / *Pierce's Disease* e *Leifsonia*.

- ✓ Utilização de ferramentas de bioinformática para anotação de ORFs nos genomas das bactérias *Xylella fastidiosa, Xanthomonas axonopodis* pv. *citri*, *Xanthomonas campestris* pv. *campestris* e *Xylella fastidiosa* / *Pierce's Disease*.

- ✓ Estudo filogenético de genes associados ao metabolismo do ferro em *Xanthomonas*.

- ✓ Análise comparativa da Biossíntese de Pequenas Moléculas das quatro bactérias anotadas.

2. Revisão de Literatura

A bioinformática é a aplicação da informática na biologia molecular, e consiste no uso e desenvolvimento de ferramentas computacionais para estudo e resolução de problemas biológicos. Ela permite o desenvolvimento de pesquisa de ponta aumentando velocidade e qualidade de informações, através da análise de grande quantidade de dados com grande precisão e tempo reduzido.

Com o aumento do número e ritmo no seqüenciamento de genomas bacterianos, há maior necessidade de desenvolvimento de técnicas acuradas para comparação de genomas e de banco de dados para facilitar a derivação da funcionalidade dos genomas, a identificação de enzimas, de operons putativos e caminhos metabólicos e derivar a classificação filogenética dos microrganismos (Bansal, 1999).

Para documentar a funcionalidade de regiões genômicas associadas à utilização do íon férrico e ao metabolismo de pequenas moléculas em quatro bactérias (*Xylella fastidiosa* – CVC = XCVC, *Xylella fastidiosa* – Pierce´s Disease = XPD, *Xanthomonas axonopodis* pv. *citri* – Xcitri, *Xanthomonas campestris* pv. *campestris* – Xcamp), alguns aspectos relevantes devem ser abordados para permitir a análise comparativa entre os genomas bacterianos.

Bioinformática e Genômica

Segundo Waterman (2000), o acúmulo de dados genéticos gerou a necessidade da criação de bancos de dados de acesso internacional para nucleotídeos e proteínas, e conseqüentemente surgiu uma área de

especialidade para esta nova realidade, na interface entre as ciências biológicas e de informação. Baxevanis & Ouellette (1998) afirmam que mais do que uma intersecção entre as biologias molecular e a computacional, a bioinformática é um novo caminho de trabalho exaustivo e significativo. Segundo Gibas & Jambeck (2001), a bioinformática é um subconjunto da biologia computacional, a aplicação de técnicas analíticas quantitativas à modelagem de sistemas biológicos. Setubal & Meidanis (1997) definem a biologia computacional como o desenvolvimento e uso de matemática e técnicas computacionais para ajudar a resolver os problemas da biologia molecular.

No Brasil os termos bioinformática e biologia computacional ainda se confundem, e ambas são tratadas como a aplicação da informática na biologia molecular, consistindo no uso e desenvolvimento de ferramentas computacionais para estudo e resolução de problemas biológicos. Ela permite o desenvolvimento de pesquisa de ponta aumentando velocidade e qualidade de informações, através da análise de grande quantidade de dados com grande precisão e tempo reduzido.

O estudo de genomas, ou genômica, tem sido um dos campos de maior desenvolvimento da bioinformática no Brasil e no mundo. Houve um grande salto com o projeto de sequenciamento do genoma humano, tanto em equipamentos quanto em programas e recursos humanos. Segundo Ferreira (2000), “a determinação de seqüência do genoma de organismos patógenos tem se destacado pela necessidade do entendimento e controle dos mesmos”.

O interesse na determinação da seqüência do genoma está na possibilidade de identificação de genes de importância econômica, científica ou social (Pallen, 1999). Mas o conhecimento dos genomas é a parte inicial das respostas, que embasará a parte mais importante da pesquisa, conhecida atualmente como genoma funcional (Jordan & Passos, 2000). Este tipo de pesquisa busca aplicações práticas das informações obtidas. Existem, por

exemplo, diferentes projetos funcionais estudando os dados gerados da *Xylella fastidiosa* da CVC[2].

Os diversos genomas seqüenciados, de bactérias, arqueobactérias e eucariotos[3] podem abrir novos caminhos em pesquisa, na busca de respostas e de questões atualmente obscuras nessa área da ciência (Strauss & Falkow, 1997).

O Brasil, com o Projeto Genoma – *Xylella fastidiosa*, ingressou no seleto grupo dos países que dominam esta tecnologia com a proposta inovadora do trabalho interligado através da internet (Rede ONSA[4]). Esta opção, compatível com a estrutura brasileira, foi viabilizada principalmente pela Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP e pela bioinformática adaptada e desenvolvida no Instituto de Computação da Universidade de Campinas – SP (UNICAMP).

A rede de serviços desenvolvida foi incorporada a outros projetos, como os o sequenciamento de genomas de *Xanthomonas axonopodis* pv. *citri*, *Xanthomonas axonopodis* pv. *campestris* e *Xylella fastidiosa* da uva, e de ESTs, como do câncer e da cana-de-açúcar. A cada novo projeto os recursos computacionais e humanos são melhorados, cumprindo um dos objetivos do programa ONSA, que é a qualificação dos recursos humanos no Brasil.

Os serviços disponibilizados nestes projetos incluem desde a submissão de clones seqüenciados até uma estrutura que facilite a anotação dos ORFs.

Sequenciamento de DNA e Montagem

Sequenciamento de DNA é a determinação de sua seqüência nucleotídica, em parte ou por completo (Docena, 2000). O seu principal objetivo é a predição das funções dos produtos dos genes, uma vez que as

[2] Página do Genoma Funcional da *Xylella* CVC – http://watson.fapesp.br/funcional/main.htm
[3] Lista do NCBI de genomas seqüenciados – http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/allorg.html
[4] Página da rede ONSA no site da FAPESP – http://watson.fapesp.br/onsa/Genoma3.htm

características biológicas dos organismos dependem diretamente destes genes. A determinação das seqüências oferece informações detalhadas que auxiliam a compreensão dos processos químicos, bioquímicos e biológicos (Ferreira, 2000). Muitos genes só são descobertos quando o genoma é seqüenciado (Docena, 2000).

O seqüenciamento de um genoma provê informações experimentais sobre as quais pode-se inferir sobre a importância de determinados processos metabólicos para o mesmo (Ferreira, 2000).

Na década de 60, quando os primeiros seqüenciamentos (de tRNA) foram realizados, não se considerava possível o conhecimento de um genoma inteiro. O DNA só começou a ser seqüenciado nos anos 70, mas o seqüenciamento só começou a se desenvolver realmente a partir do Método de Terminação da Cadeia (*Chain Termination Method*), desenvolvido pelo Dr. Fred Sanger e equipe (Hausmann, 1997).

Em 1986 surgiu o seqüenciamento automático com marcadores fluorescentes (Smith et al., 1986). O primeiro organismo a ser inteiramente seqüenciado por este método foi o do poxvírus (causador da varíola), com 186.000 pares de bases (Massung et al., 1993).

Estratégias de seqüenciamento de genomas

Segundo Lin (2001), os seqüenciadores atuais conseguem ler no máximo 1000 pares de bases, o que exige que o DNA genômico, para ser seqüenciado, seja quebrado em pequenos pedaços, chamados doravante de clones. Posteriormente, estes clones são comparados e alinhados, e o DNA original remontado.

Existem diferentes estratégias para clonagem e montagem de genomas, e algumas delas podem ser associadas no mesmo seqüenciamento. Segundo Docena (2000), alguns organismos foram seqüenciados com base em diversas informações preliminares, como mapa físico, biblioteca ordenada de cosmídeos

e Lambda. No método de "*shotgun*", que envolve a fragmentação aleatória do genoma total, estes dados preliminares não são indispensáveis. Apesar disso, é recomendável a utilização do mapa físico para maior confiabilidade da montagem.

As duas estratégias mais freqüentemente utilizadas atualmente são: a construção de bibliotecas de grandes fragmentos de inserto para construção de mapa físico, com clones sobrepostos cobrindo todo o genoma; e o seqüenciamento direto de clones de *shotgun*, a mais utilizada (Docena, 2000).

Um método complementar ao *shotgun* é o OSS (*Ordered Shotgun Sequencing*), que foi proposto em 1993 para amenizar os problemas de montagem. Sua inovação é a utilização de clones grandes, de dois mil até centenas de milhares de pares de bases (pb). Inicialmente somente suas extremidades são seqüenciadas (Lin, 2001). Estas pontas seqüenciadas permitem que os clones sejam localizados na montagem de *shotguns*. São utilizados para corrigir regiões específicas do genoma que apresentam problemas de montagem. Como exemplo, pode-se escolher um cosmídeo (clone de cerca de 40.000 pb) para resolver um ou mais complicadores de montagem.

A montagem dos fragmentos de DNA e o fechamento do genoma são o próximo passo no seqüenciamento. Ocorrem desde o início da obtenção dos cromatogramas seqüenciados, e permitem um acompanhamento da qualidade destes e da evolução do progresso e da qualidade do trabalho. Docena (2000) indica três fases para a montagem: (1) a conversão de cromatogramas em seqüências nucleotídicas; (2) a montagem de contíguos genômicos; e (3) a montagem destes contíguos numa seqüência consenso.

Podemos dividir didaticamente a montagem em duas fases distintas: micro e macro-montagem. Na micro-montagem pode-se incluir as montagens realizadas com cromatogramas, que possuem valores de qualidade para cada posição, e que vão determinar trechos de seqüência consenso. É o caso das montagens iniciais de *shotgun*, e de bibliotecas de cosmídeos e de plasmídeos (clones pequenos).

A macro-montagem, por trabalhar com grandes trechos de seqüências, utiliza os consensos gerados pelas micro-montagens ao invés dos inúmeros cromatogramas que cada uma possui. Esta opção agiliza e viabiliza estas montagens, além de evitar uma perda de tempo e de poder de processamento ao não refazer um trabalho já realizado.

As macro-montagens também atuam na resolução de alguns problemas de fechamento, ao indicar regiões que devem ser melhoradas, seja pelo reseqüenciamento de clones ou pela construção de bibliotecas de cosmídeos ou plasmídeos. Lin (2001) cita três complicadores do processo de montagem: (1) erros nos clones, como fragmentos quiméricos ou contaminação com vetores; (2) regiões repetidas; e (3) falta de cobertura. Cabe ainda ressaltar, como problemas a serem resolvidos: regiões de transposons, que podem confundir a montagem; e trechos de queda de qualidade abrupta no seqüenciamento ("compressão"), causados normalmente por estruturas secundárias ou trechos muito ricos em GC.

Anotação

Foi realizada anotação nas quatro bactérias citadas. Na *Xylella* CVC iniciou-se o aprendizado, com participação da fase inicial da anotação que foi feita por cosmídeo seqüenciado. Posteriormente, nas *Xanthomonas*, em duas categorias: Processos Celulares e Patogenicidade, Virulência e Adaptação. Na *Xylella* PD auxiliando a coordenação da categoria Biossíntese de Pequenas Moléculas.

Anotação é o processo de interpretação de novas seqüências genômicas em informações biológicas úteis, através da integração de análises computacionais com dados biológicos (Lewis et al., 2000). Também pode ser vista como determinação e caracterização de genes, elementos ativos ou funcionais, ou de espaços intergênicos ou genes não funcionais, elementos inativos (Lee, 2001).

O passo inicial da anotação de genes é a localização de ORFs, que segundo Docena (2000) são seqüências de DNA com códon iniciador na extremidade 5´ seguida da região codificadora do gene e de uma região de término. O termo ORF significa "*open reading frame*", e pode ser traduzido como "quadro aberto de leitura". A tradução dos nucleotídeos em aminoácidos é possível em três fases de leitura para cada lado da fita de DNA. Quando estão na mesma fase de leitura um códon iniciador, um trecho de nucleotídeos e um códon terminador, nesta ordem, temos um ORF. Docena (2000) afirma que todos os genes são ORFs, mas nem todo ORF é um gene.

Anotar é postular uma função ao produto de um ORF, um gene em potencial. Este trabalho é realizado através da comparação dos ORFs com seqüências de genes conhecidas. É a parte final e mais importante de um seqüenciamento de genoma. Segundo Lee (2001), a anotação é muito importante em genomas seqüenciados por fornecer informações preliminares sobre a presença ou ausência de vias metabólicas. Exige diferentes fases para filtrar as inúmeras informações encontradas no meio científico e obter uma lista de ORFs que represente, com a maior chance de acerto possível, a seqüência de genes reais do organismo.

Processos Celulares

Durante a anotação da categoria V (Processos Celulares) da *Xanthomonas*, houve uma divisão das famílias existentes entre os integrantes do grupo, e foram estudadas as famílias de transporte 2.C.1, 9.A.1, 1.B.18, 8.A.3 e 1.B.14, segundo classificação de Milton Saier.

A família 2.C.1 é a família de proteínas auxiliares para energização de transporte ativo mediado por receptores da membrana externa. Possui dois sistemas parálogos: TonB-ExbB-ExbD e TolA-TolQ-TolR. A família 9.A.1 é de transportadores de polissacarídeos, incluindo a exportação de exopolissacarídeos. Possui proteínas auxiliares, pertencentes às famílias 1.B.18 – proteínas auxiliares da membrana externa, e 8.A.3 – auxiliares da membrana periplásmica. A famílias 1.B.14 inclui diversos receptores da

membrana externa, normalmente dependentes do sistema de energização da família 2.C.1[5].

Biossíntese de Pequenas Moléculas

Segundo Docena (2000), a Biossíntese de pequenas moléculas inclui a biossíntese de aminoácidos; nucleotídeos; açúcares; cofatores, grupos prostéticos e transportadores; ácidos graxos; e poliaminas.

Os aminoácidos incluem as famílias:

- ✓ glutamato – arginina, glutamato, glutamina e prolina;
- ✓ aspartato/piruvato – alanina, valina, leucina, aspartato, asparagina, metionina, treonina, lisina e isoleucina;
- ✓ glicina – serina, glicina e cisteína;
- ✓ aminoácidos aromáticos – histidina, corismato, fenilalanina, triptofano e tirosina.

Os nucleotídeos :

- ✓ Ribonucleotídeos de purina;
- ✓ Ribonucleotídeos de pirimidina;
- ✓ 2'-Deoxiribonucleotídeos;
- ✓ Salvamento de Nucleotídeos e Nucleosídeos;

Cofatores, Grupos Prostéticos e Transportadores:

- ✓ Biotina;
- ✓ Ácido fólico;
- ✓ Lipoato;
- ✓ Molibdopterina;
- ✓ Pantotenato;
- ✓ Piridoxina;
- ✓ Nucleotídeos de pirimidina;

[5] Informações obtidas nas páginas organizadas pelo Dr. Milton Saier – http://www-biology.ucsd.edu/~msaier/transport/2_C_1.html para a família 2.C.1, final 9_A_1.html para a família 9.A.1 e assim por diante.

- ✓ Tiamina;
- ✓ Riboflavina;
- ✓ Tioredoxina, glutarredoxina, glutationa;
- ✓ Menaquinona, ubiquinona;
- ✓ Protoheme, siroheme;
- ✓ Proteína transportadora de carboxil biotina;
- ✓ Cobalamina;
- ✓ Enterobactina; e
- ✓ Biopterina.

Mecanismos Associados ao Ferro

Ferro é considerado um fator limitante para o crescimento da bactéria tanto em seu ambiente natural quanto em associação com seu hospedeiro infectado. Para manter a homeostase do ferro, muitas bactérias sintetizam um regulador global, a proteína Fur (Ferric Uptake Regulator), codificada pelo gene regulador da absorção do ferro *fur* (Guerinot, 1994). De um modo geral, Fur atua como um repressor transcricional (Litwin & Calderwood, 1993). Por outro lado, a regulação transcricional negativa mediada pelo Fur tem sido descrita pelos mesmos autores. Em adição ao controle da absorção do ferro, Fur também regula os genes envolvidos em virulência (Litwin & Calderwood, 1993; Ochsner et al., 1995), tolerância à acidez (Foster & Spector, 1995; Hall & Foster, 1996), os genes protetores ao stress oxidativo (Hassettet al., 1996) e outros fatores codificados por genes regulados por ferro tais como os sistemas de aquisição de ferro (ex. sideróforos), hemolisina e toxinas (Litwin & Calderwood, 1993). Fur é, portanto, um importante regulador global no metabolismo geral (Hantke, 1987; Tsolis et al., 1995).

Apesar de abundante na natureza, a biodisponibilidade do ferro é extremamente baixa. Ao longo da evolução bacteriana, complexos sistemas para absorção do ferro foram gerados para permitir a sua sobrevivência. Como exemplo, citamos o processo da aquisição de ferro pela bactéria através da secreção de sideróforos e a subseqüente absorção de complexos sideróforos de ferro, governada em grande parte pela proteína Fur (Bagg & Neilands, 1987b).

Na presença de ferro, Fur inativo é ligada a íons ferrosos disponíveis no citoplasma e se torna um repressor ativo de transcrição que se liga a uma região de DNA localizada nas vizinhanças da seqüência do promotor conhecido como caixa Fur (Loprasert et al., 1999). A ligação ao repressor Fur na caixa Fur bloqueia a transcrição do gene (Bagg & Neilands, 1987a; Escolar et al., 1998).

O gene *fur* tem sido caracterizado em muitas bactérias Gram-negativas (Staggs & Perry, 1991; Beall & Sanden, 1995; Hassett et al., 1996; Achenbach & Yang, 1997) e recentemente em bactérias Gram-positivas (Bsat et al., 1998). Fur é uma proteína pequena (15-18 kDa) contendo muitas regiões altamente conservadas, importantes para a sua função (Braun et al., 1990; Coy et al., 1994). A análise de seqüências em genomas bacterianos tem indicado que algumas bactérias têm múltiplas proteínas Fur (Bsat et al., 1998). *B. subtilis* tem pelo menos três Fur homólogos estruturais. Essas proteínas têm diversas funções, sugerindo que muitas funções do Fur e proteínas semelhantes a Fur ainda estão para serem identificadas (Bsat et al., 1998; Vasil & Ochsner, 1999).

Na caracterização, organização genômica e análise filogenética do gene *fur* em *Klebsiella pneumoniae*, demonstrou-se a capacidade da seqüência desta proteína em refletir uma relação filogenética entre espécies, sugerindo que o gene *fur* de modo semelhante ao 16S rRNA, pode não estar sujeito à transferência horizontal entre as diversas bactérias avaliadas (Achenback & Yang, 1997).

Relações Filogenéticas e Evolução

O sequenciamento de genes e, em especial, de genomas inteiros, pode ajudar no entendimento da evolução das bactérias em seu ambiente. A evolução biológica tem como pré-requisito a variação genética, dada principalmente por mutações, que aparentam ocorrer em maior freqüência em situações de estresse ambiental (Schloter et al., 2000).

Li (1997) cita três principais motivos para explicar a grande importância da utilização de seqüências de DNA para estudos evolutivos. Inicialmente, o DNA (e as proteínas) evoluem de maneira mais regular do que caracteres morfológicos e fisiológicos. Em segundo lugar, estes dados permitem melhor tratamento quantitativo do que dados morfológicos, e ainda são mais abundantes. Esta abundância é um fator muito útil principalmente no caso de microrganismos, que têm poucas informações morfológicas e fisiológicas disponíveis para estudo. Li (1997) lembra que embora os dados moleculares tenham revolucionado o estudo da taxonomia bacteriana, não se deve esquecer que as demais características são complementares e, portanto, não se deve abandonar estudos morfológicos e fisiológicos.

Diversos genes relacionados à patogenicidade estão associados a elementos transponíveis, como ilhas genômicas, plasmídeos e transposons. Os elementos transponíveis (TEs) são trechos de DNA com capacidade de alterar a sua localização no genoma, produzindo mutações e/ou rearranjos só possíveis por este meio, o que delega aos TEs um importante papel evolutivo, principalmente nas interações entre microrganismos e seus hospedeiros (Hentschel et al., 2000; Li, 1997).

Estes eventos de transferência horizontal de genes são passos chave em processos de especiação (Schloter et al., 2000). E o DNA também pode ser passado entre diferentes espécies. Bactérias podem “absorver” DNA do meio e integrá-lo ao seu genoma, mantendo sua funcionalidade. Este processo é conhecido como transformação (Li, 1997). De La Cruz & Davies (2000) afirmam que a virulência, ou capacidade de colonização de um novo nicho, acontece mais pela aquisição de genes patogênicos devido à transferência horizontal do que por mutação.

Segundo Dröge (1998), pode-se classificar os eventos de transferência de genes na natureza em três caminhos distintos: a detecção de genes homólogos em diferentes espécies; demonstrar a transferência experimentalmente em laboratório; e em estudo de campo. O estudo de genes homólogos, através da análise de seqüências de nucleotídeos ou aminoácidos, permite verificar se os

genes seguem o padrão evolutivo dos organismos, ou se diverge deste. Doolittle (2000) afirma que quanto mais recente a transferência, mais parecidas serão as seqüências homólogas.

A análise filogenética é o processo de desenvolver hipóteses sobre a relação evolutiva de organismos utilizando as suas características observáveis. Sua utilização foi iniciada com anatomia macroscópica, na árvore de vida de Lineu. Atualmente, a natureza quantitativa das seqüências permitiu o desenvolvimento de regras mais rigorosas para o desenho da árvore (Gibas & Jambeck, 2001).

A análise de filogenia inicia com a comparação das seqüências dos genes. Segundo Li (1997), o primeiro passo é o alinhamento de seqüências, que sobrepõe as mesmas e identificam locais de possíveis deleções e inserções. Lee (2001) afirma que estas comparações são importantes para melhor visualizar as diferenças entre as seqüências estudadas.

A árvore filogenética derivada de dados de seqüências pode ter ou não uma raiz, pois embora se siga o pressuposto da existência do ancestral comum, os métodos empregados permitem que se calcule a semelhança entre as seqüências e que se determine onde colocar pontos de ramificação (Gibas & Jambeck, 2001).

Segundo Li (1997), os objetivos dos estudos filogenéticos são a reconstrução das ligações genealógicas entre os organismos e estimar o tempo de divergência entre os organismos e seu ancestral comum. Pode-se estender este conceito para a busca de vínculos genealógicos de genes, e sua comparação com a evolução dos organismos.

Uma árvore pode representar uma filogenia, mas é preciso mais do que uma única análise evolutiva para tirar conclusões sobre a filogenia de um organismo completo. Somente quando as filogenias são construídas com base em quantidades suficientes de dados estas podem dar evidências para se inferir sobre a história evolutiva do organismo (Gibas & Jambeck, 2001).

As árvores filogenéticas são a forma de ilustrar a relação evolucionária entre um grupo de organismos (Li, 1997) ou genes. Segundo Li (1997) a árvore filogenética é composta por nós e ramos, onde somente um ramo conecta dois nós adjacentes. Os nós representam as unidades taxonômicas, e os ramos as relações entre estas.

Estas árvores podem ser classificadas em árvores de espécies, quando representam caminhos evolutivos de um grupo de espécies, ou árvores de genes, quando comparam genes que não seguem necessariamente a escala evolutiva da espécie (Lee, 2001).

Gibas & Jambeck (2001) descrevem quatro métodos para a construção das árvores filogenéticas. A baseada na distância entre pares produz árvores com raiz. É definida uma matriz de distância entre cada par de seqüências e estas são agrupadas de acordo com estas distâncias. O tamanho dos ramos teoricamente reflete o tempo evolutivo.

A árvore baseada na junção de vizinhos também utiliza uma matriz de distâncias. Seqüências mais próximas são consideradas vizinhas e o algoritmo pesquisa as menores distâncias e também conjuntos de vizinhos que diminuam o tamanho total da árvore.

O método da parcimônia máxima pesquisa o conjunto de árvores possíveis para localizar a que exige o menor número de substituições para explicar as diferenças entre as seqüências. Analisa apenas os locais que fornecem evidências evolutivas.

Além deste, o método baseado na estimativa de probabilidade máxima, que também avalia todas as topologias de árvores possíveis, mas utiliza métodos probabilísticos. Atribuindo probabilidade a cada alteração possível, calcula a melhor escolha.

3. Material e Métodos

Tanto o material quanto os métodos utilizados sofreram modificações ao longo do trabalho. Programas foram atualizados e novos equipamentos adquiridos, permitindo ainda a utilização de novos programas.

Muito tempo se gastou nos primeiros cosmídeos com verificação manual da montagem, retirada de trechos de vetores ou de baixa qualidade mal cortados, já que o programa *sequencher* é limitado neste aspecto. O processo de montagem no laboratório, inicialmente restrito ao sistema Macintosh e ao programa *sequencer*, ganhou muito em qualidade, velocidade e facilidade com a chegada dos PCs I e II e com a instalação dos programas phred, phrap e consed no sistema linux. A própria experiência com estes programas permitiu melhor aproveitamento de seus recursos e o desenvolvimento de uma estrutura para evitar erros.

Na anotação, com o decorrer dos projetos, a experiência foi permitindo melhorias tanto nas ferramentas de bioinformática disponibilizadas quanto nas estratégias de anotação dos grupos. Embora em cada projeto existam diversos anotadores iniciantes, a evolução dos recursos humanos é facilmente perceptível, principalmente se comparada ao projeto inicial (*X. fastidiosa* CVC). Além de uma melhor utilização de ferramentas já existentes, como PFAM e COGnitor, os seqüenciamentos dos projetos anteriores serviram como base de referência para a anotação. A comparação de genes das quatro bactérias (XCVC, XPD, Xcitri e Xcamp) permitiu, em muitos casos, maior precisão nas conclusões e até correções em anotações anteriores. O meio científico mundial é muito dinâmico, e a cada momento podem surgir novas informações que permitem a definição da anotação de um ORF.

O estudo da filogenia de alguns genes foi realizado após o fim da anotação por categorias da XPD. Foram estudados alguns programas, e foi escolhido o PAUP (Swofford, 2002), utilizando o alinhamento do CLUSTAL (Thompson et al., 1994). Dentre os programas acessíveis, o PAUP apresenta o recurso *bootstrap*, utilizado no trabalho.

Encontram-se em anexo os programas utilizados, endereços de páginas dos projetos e de páginas de busca e consulta (anexo 1).

3.1. Computadores

Estrutura pertencente ao Laboratório de Biologia Celular e Molecular do CENA, de uso compartilhado com outros pesquisadores.

- ✓ **Mac G3** – Power Macintosh G3 (160 Mb RAM, 6 Gb Disco Rígido, Mac OS 8.1, zip drive interno, gravador de CDs externo) – análise de géis, montagem, cópias de segurança;
- ✓ **Mac 7300** – Power Macintosh 7300/200 (160 Mb RAM, 2 Gb Disco Rígido, zip drive externo) – sequenciamento (acoplado ao seqüenciador ABI Prism™ 377 DNA Sequencer);
- ✓ **PC I** (Dual Pentium III 550 MHz, 256 Mb RAM, 20 Gb disco rígido) – Sistemas Operacionais Linux e Windows 98, prioritário para montagem em linux. Também utilizado para anotação e outros;
- ✓ **PC II** (Processador Pentium III 550 MHz, 256 Mb RAM, 20 Gb disco rígido) – Sistemas Operacionais Linux e Windows 98, reserva para casos de atualizações e problemas técnicos no PC I. Montagem em linux, anotação e demais demandas;
- ✓ **PC III** (Processador Pentium III 700 MHz, 256 Mb RAM, 30 Gb disco rígido) – Sistema Operacional Windows 98, utilizado em anotação.

3.2. Sequenciamento

Os métodos em bioinformática incluem a manutenção básica de computadores porque a exigência de prazos e qualidade dos projetos muitas vezes impede que sejam seguidos os caminhos normais da instituição.

Também serviços relacionados diretamente aos projetos genoma, como no sequenciamento e na montagem, e de forma indireta nestes e em outros projetos através do auxílio a pesquisadores.

O trabalho envolve desde o primeiro projeto genoma do Brasil, da *Xylella fastidiosa*, e demais citados no item 4.1, sendo seis programas de sequenciamento no total. Há uma evolução natural das páginas de serviços com o passar do tempo, mas as diferenças não justificam uma descrição para cada projeto. Serão descritos os processos mais importantes e os principais itens a se considerar em cada caso.

3.2.1. Análise de géis

Seqüenciador ABI Prism 377; sistema operacional MACOS 8; sequenciamento realizado com Mac 7300, acoplado ao seqüenciador; passagem para o Mac G3 com o auxílio de um disco zip; programa *Sequencing Analysis.*

- ✓ Abre-se arquivo de gel;
- ✓ Traqueamento – centralizar as linhas de extração nas colunas referentes a cada clone;
- ✓ Verificação do traqueamento – cada linha deve corresponder ao clone descrito, ou seja, deve-se ter certeza de que possíveis falhas no sequenciamento não levem a um erro de nomenclatura dos clones;
- ✓ Ajuste fino – aumentar o zoom para escolher a melhor região de leitura da coluna;
- ✓ Extrair os cromatogramas dos clones;
- ✓ Verificar clones que não podem ser aproveitados e apagá-los;
- ✓ Verificar nomenclatura.

3.2.2. Submissão de clones seqüenciados

- ✓ Selecionar os cromatogramas a serem submetidos e colocá-los em uma pasta;
- ✓ Compactá-los, utilizando o programa pkzip;
- ✓ Acessar a página de serviços do organismo;
- ✓ Completar os campos com código do laboratório, e-mail e arquivo a ser enviado;
- ✓ Submeter;
- ✓ No caso de submissão aceita, verificar e arquivar relatório;
- ✓ No caso de problema na submissão, resolver problema (podem ocorrer erros na nomenclatura, problemas na compactação, problemas no script de submissão etc. Se necessário, buscar auxílio no laboratório central).

3.2.3. Submissão de montagens

- ✓ Compactar o arquivo "*nome*.ace" da montagem:
  - o No linux: gzip <nome do arquivo>
- ✓ Acessar a página de serviço do organismo;
- ✓ Completar os campos com código do laboratório, e-mail e arquivo a ser enviado;
- ✓ Submeter;
- ✓ No caso de submissão aceita, verificar e arquivar relatório;
- ✓ No caso de problema na submissão, resolver problema (podem ocorrer erros na nomenclatura, problemas na compactação, problemas no script de submissão etc. Se necessário, buscar auxílio no laboratório central).

3.2.4. Download de dados

- ✓ Acessar a página do serviço de download;
- ✓ Digitar a lista de nomes de clones que se deseja, sendo um nome por linha;
- ✓ Submeter e busca;
- ✓ Salvar o arquivo da busca;
- ✓ Descompactá-lo.

3.2.5. Comparação de cromatogramas com *Sequencer – MacOS*

- ✓ Iniciar programa;
- ✓ Adicionar pasta de cromatogramas;
- ✓ Selecionar cromatogramas;
- ✓ Corte seletivo das pontas de baixa qualidade (*trim ends*);
- ✓ Corte seletivo de fragmentos de vetor (*trim vector*);
- ✓ Regulagem de parâmetros de montagem:
  - Tamanho mínimo – 40;
  - Identidade mínima – 90;
- ✓ Montagem automática;
- ✓ Regulagem de parâmetros de montagem:
  - Tamanho mínimo – 30;
  - Identidade mínima – 80;
- ✓ Montagem automática;
- ✓ Se restarem cromatogramas fora dos contíguos, ou se não se alcançou o número desejado de contíguos;
  - Regulagem de parâmetros de montagem:
    - Tamanho mínimo – 10;
    - Identidade mínima – 60;
  - Montagem Interativa. Neste caso, estuda-se individualmente cada montagem possível.

Sequencer - Depois de gerados os contíguos, pode-se visualizar e editar as montagens. Utilizado para visualização inicial de montagens, e comparação de seqüências. Substituído para montagens de cosmídeos e afins pelo programa phred, phrap e consed.

3.2.6. Desenho de primers

A escolha de oligonucleotídeos (primers) é um passo muito importante. O primeiro motivo é sua utilização para cobrir regiões não representadas, seja pela ausência de clones ou por problemas de sequenciamento. Esta demanda

ocorre principalmente no fechamento de cosmídeos, época de muita pressa. Em segundo lugar, sintetizar um primer é um trabalho caro, e normalmente se utilizam serviços internacionais.

Estas características tornam a escolha de primers um trabalho criterioso e importante, pois um primer mal feito custa dinheiro e tempo preciosos. O programa utilizado para o desenho de primers foi o Consed, com informações geradas nos programas Phred e Phrap. Este possui parâmetros padrão para a escolha de primers, que normalmente não são alterados. Além destes, existem informações relevantes utilizadas para escolher as melhores opções dentre os primers escolhidos pelo programa consed.

- ✓ Tamanho – o considerado ideal é 18 bases. O mínimo utilizado foi 16, e o máximo 20. Pode ser um pouco maior se necessário, mas não é recomendável que seja menor do que 16;
- ✓ Porcentagem GCs/ATs – busca-se entre 50 e 60% de GCs, sendo 50% o ideal. Estes valores dizem respeito às bactérias seqüenciadas, podendo este parâmetro mudar de acordo com o organismo;
- ✓ Distribuição das bases – quanto mais distribuídas, melhor. Seqüências maiores do que 3 repetições da mesma base são evitadas. Ex.: boa distribuição – ACCTGTAGGATCACTAGC, má distribuição - AAGTTCTTAAAGGGGCGAG;
- ✓ Distância em relação ao trecho alvo – no caso de trechos sem problemas, a distância vai depender do tamanho do "buraco" a ser coberto. Pode ser necessário o desenho de mais de um primer. No caso de problemas de sequenciamento, o primer deve estar o mais próximo possível de uma distância ideal, cerca de 80 a 100 bases. Se estiver mais próximo, pode não estar estabilizado quando chegar ao trecho problemático, e não apresentar qualidade neste, ou até mesmo não cobrir a região. Se estiver mais longe, a capacidade de ultrapassar o problema diminui com a distância;
- ✓ Verificar anelamento múltiplo – o programa realiza a busca de outros anelamentos na região montada no projeto utilizado. Se for utilizado o DNA do organismo inteiro, ou de regiões diferentes da montada, ou

mesmo se o vetor a ser utilizado não constar no arquivo específico do programa, será necessário comparar os primers escolhidos com a seqüência conhecida do DNA em questão. Pode-se utilizar, por exemplo, o programa cross-match.

- ✓ Dentro de clones – no caso de regiões problema, é recomendável que os primers estejam dentro dos limites de um clone seqüenciado. Isto facilita o processo de sequenciamento;
- ✓ Ponta 3' – existem opiniões divergentes sobre a que bases se deva dar preferência no final da seqüência. Alguns afirmam que seja melhor ATs, outros CGs. Foi utilizado preferencialmente o final com um ou dois CGs, com bom funcionamento dos primers.

3.3. Montagem de Fragmentos de DNA

Programa

A montagem é realizada utilizando-se o pacote de programas phred, phrap e consed (Ewing et al, 1998; ewing & Green, 1998; Gordon et al., 1998), em ambiente operacional linux. O pacote inclui outros programas, como o Repeat Master e Cross-match, e diversos scripts para facilitar a interação entre os mesmos. Após devidamente instalados, estes programas executam grande parte do serviço, restando ao operador alguns trabalhos mais específicos e a resolução de problemas.

Estrutura mínima

É exigida uma estrutura específica de diretórios para que os programas funcionem corretamente. Para cada montagem a se realizar, devem ser criados quatro diretórios: (1) <nome da montagem>, (2) "*edit_dir*", (3) "*phd_dir*", e (4) "*chromat_dir*", sendo os três últimos dentro do primeiro. Por exemplo, para a montagem de um cosmídeo de nome 00C10, cria-se um diretório com este nome e dentro dele os três outros diretórios. O nome dos diretórios internos não pode ser alterado, por uma determinação do próprio programa.

No diretório *chromat_dir* serão colocados os cromatogramas que se quer

montar inicialmente, assim como futuros novos cromatogramas. A partir do diretório *edit_dir* serão executados todos os comandos, e todos os arquivos gerados para visualização de dados e saída de informações úteis estarão neste diretório.

Estrutura utilizada

Além da estrutura mínima para o programa, foi adotado um modelo de estrutura de diretórios para facilitar as montagens de cosmídeos e plasmídeos, e evitar ao máximo a possibilidade de erro.

No diretório do cosmídeo, 00B01, p. ex., criam-se 5 diretórios: "00B01bkce", "00B01fim", "externos", "blast" e "pontas". Em cada um dos dois primeiros, cria-se a estrutura básica dos três diretórios *edit_dir*, *phd_dir* e *chromat_dir*. Em *externos*, cria-se o diretório "seleção", e em *blast* o diretório "resultados".

Na pasta 00B01bkce estará a montagem somente com os cromatogramas do próprio cosmídeo, que servirá como prova na comparação com montagens posteriores feitas na pasta 00B01fim, que conterá cromatogramas de *shotgun* (aleatórios), utilizados para a finalização da montagem. Este acompanhamento, auxiliado pelo uso do *Blast 2 sequences*, permite que se detectem diferenças nas montagens, que podem ser corretas ou não, mas que devem ser examinadas com cuidado.

A pasta "externos" serve como um depósito de cromatogramas de *shotgun* buscados para serem introduzidos na montagem. Facilita o controle. A pasta pontas é apenas um depositório dos cromatogramas que marcam as pontas de cada cosmídeo. Eles auxiliam a localização das pontas do cosmídeo que se está montando.

Na pasta *blast* são colocados arquivos fasta de contíguos e cromatogramas que se queira comparar com bancos de dados gerais ou específicos do organismo. Na pasta interna, "*resultados*", podem ser guardadas

comparações para futuras análises.

Funcionamento

O programa utiliza scripts, que são arquivos com seqüências de comandos e parâmetros. O script principal do pacote é o "phredPhrap.perl", que aciona os programas phred, repeat master, cross-match e phrap.

Inicialmente, o programa phred cria para cada cromatograma presente na pasta *chromat_dir* um arquivo correspondente na pasta *phd_dir*, com valores de qualidade atribuídos aos picos do sequenciamento de cada base. A seguir, estes arquivos são comparados, pelo programa cross-match, e depois se constroem contíguos, que são agrupamentos de pedaços de DNA com partes sobrepostas, que formam pedaços maiores de fita através do consenso dos cromatogramas reunidos - phrap. Também é gerado um arquivo em formato "ace", que permite a visualização e edição desta montagem através do programa consed.

Os principais scripts utilizados são:

- ✓ phredPhrap.perl – para casos comuns de montagem – descrito acima;
- ✓ fasta2Phd.perl – cria arquivos "phd" a partir de arquivos em formato fasta, sem a necessidade de cromatograma. Atribui qualidade uniforme;
- ✓ phd2Ace.perl – cria arquivos "ace" a partir de arquivo "phd". Permite o uso do consed em seqüências que não participam de alinhamento;
- ✓ ace2Oligos.perl – gera arquivo com primers presentes em arquivo de montagem (.ace).

O programa Consed permite a visualização de montagens e edição de dados para facilitar a finalização dos fragmentos de DNA. Inclui diversos recursos, como:

- ✓ Mostra valores de probabilidade de erro nos contíguos;
- ✓ Escolha de primers, com diversos parâmetros editáveis;
- ✓ Busca automática de problemas específicos (menu *navigate*);
- ✓ Escolha automática de clones para fechamento;

- ✓ A visualização de montagens ainda permite:
  - escolha manual de clones a serem feitos ou refeitos;
  - determinação de problemas;
  - auxílio na resolução destes;
- ✓ Visualização de cromatogramas alinhados;
- ✓ Busca de trechos de DNA (*search for string*);
- ✓ Busca de clones, com utilização de caracteres coringa;
- ✓ Marcações diversas (*tags*).

3.4. Anotação

A anotação de ORFs é realizada através da comparação das seqüências de nucleotídeos e aminoácidos de cada ORF com diversos bancos de dados genéticos (já descritos) em diferentes programas. A partir do cruzamento das inúmeras informações de alinhamento e homologia encontradas, determina-se a homologia do ORF.

3.4.1. Ferramentas para Anotação

O processo de anotação de genes é dinâmico, e utiliza as informações atuais dos bancos de dados genéticos públicos internacionais. Diante do crescente número de informações adicionadas diariamente, muitos casos de anotação podem ser alterados com novas descobertas. Mas as ferramentas utilizadas para buscar estas informações, embora sofram também atualizações, podem ser descritas de forma mais precisa para que o método empregado possa ser repetido. Novas ferramentas computacionais podem (e irão) surgir.

Neste item será feita uma descrição das principais ferramentas utilizadas no processo de anotação, assim como forma de entrada e saída de dados. Nos itens seguintes será descrita a forma como cada projeto utilizou estas ferramentas, sem a preocupação de descrever detalhes de utilização.

3.4.1.1. A Página de Anotação

Cada projeto de seqüenciamento do programa ONSA contou com facilidades bioinformáticas, e a maioria delas está reunida nas páginas de anotação. Estas páginas apresentam facilidades para agilizar o processo de anotação, através de informações organizadas, consultas automáticas e uma interface amigável.

Como base, será descrita a página de anotação da bactéria *Xanthomonas axonopodis* pv. *citri*. A página da *Xylella* CVC, anotada anteriormente, possuía menos recursos, e foi descrita em detalhes em Docena (2000) e Ferreira (2000). Os recursos adicionais encontrados no projeto da *Xylella* PD serão descritos posteriormente.

Os ORFs foram inicialmente localizados com o auxílio dos programas Glimmer 2.0 e Genemark 2.4, e foram pré-categorizados e anotados através de uma busca com o programa blastp. Este trabalho foi realizado pelo laboratório central de bioinformática do projeto – LBI/UNICAMP. Várias informações do primeiro hit foram colocadas automaticamente numa página gerada automaticamente para cada ORF. Os principais itens fornecidos na última versão desta página são:

- ✓ Pr*oduct*, com nome da proteína codificada pelo gene;
- ✓ *Gene name*, com nome do gene;
- ✓ *Chunk*, com informações sobre a localização do ORF, e *links* para a verificação de códon inicial, as seqüências em bases e aminoácidos, e ainda a qualidade da seqüência;
- ✓ *TC number*, código para genes de transporte;
- ✓ COG (Tatusov et al., 2001) – *Cluster of Orthologous Groups*, com informações geradas pelo programa *cognitor*, quando existentes;
- ✓ *Links* para resultados de: BLASTP/NOFILTER, COG, PSORT e PFAM;
- ✓ *Accession number*, número da entrada do gene homólogo nos bancos de dados genéticos GenBank (Benson et al., 2000) ou Swiss-Prot (Bairoch & Apweiler, 2000);
- ✓ *Organism,* com organismo do primeiro hit;

- ✓ *Identity,* que apresenta a porcentagem de semelhança do ORF com o gene homólogo, ou seja, a porcentagem de acertos;
- ✓ *Coverage Query,* porcentagem que demonstra relação entre tamanho do alinhamento e tamanho do ORF (neste caso no papel de *query*);
- ✓ *Coverage Subject,* porcentagem da relação entre o tamanho do alinhamento e o tamanho do gene homólogo (no papel de *subject*);
- ✓ *E-value,* apresenta a possibilidade do alinhamento ter sido feito de forma aleatória. Quanto menor, mais provável é a veracidade do alinhamento.

Além dos campos descritos, somente para visualização de dados, a página apresenta campos editáveis, onde o anotador que tem permissão pode alterar ou adicionar informações: *Product, Gene name, TC number, Primary Category, Secondary Category, Remarks e Notepad.* Ainda apresenta os campos que permitem marcação (sim/não), que apontam a presença de problemas no ORF *(Frameshift, Poin Mutation, Both, None, Orf with problem),* o término da anotação por parte de cada fase da mesma (*by category member, by category resposible, by member of scan team),* ou ainda mudanças no códon inicial ou localização do ORF em um intron (*Change start codon* e *Intron*, respectivamente*).*

A busca avançada (*advanced search*) do *gene editor* apresenta opções de busca bastante variadas, permitindo busca simultânea nos campos: C*hunk, Primary Category, Secondary Category, Product, Gene name, Remarks, Notepad, by category member, by category responsible, by member of scan team, Orf with problem e tRNA.* A busca básica procura a palavra nos campos *gene ID, gene name, remarks* e *product.*

A página da *Xylella* PD apresentou algumas mudanças em relação à da *Xanthomonas*, embora a totalidade de recursos tenha sido pouco modificada. Uma diferença funcional importante é que os *links* levam à abertura de novas janelas, ao contrário do que acontecia em *Xanthomonas*. Foram introduzidos também campos de números EC (Bairoch, 2000) e TC, *links* para resultados da página de busca SMART (Schultz et al, 1998), para BLAST (Altschul et al., 1997) contra a própria *Xylella* PD, e para mapas de vias metabólicas da página

do KEGG (Kanehisa et al., 2002), descritos abaixo:

- ✓ Número EC – número de categorização de enzimas segundo sua função. Cada EC representa uma função;
- ✓ Número TC – categorização de funções relacionadas somente ao transporte;
- ✓ SMART – programa de comparação que utiliza diversos processos de busca, incluindo domínios;
- ✓ BLASTP – um blast local que permite a comparação com os dados do organismo ainda não submetidos. Permite comparações com montagens prévias, contra seqüências, e contra ORFs (blast descrito no item 3.4.1.2);
- ✓ Mapas KEGG – a partir do número EC encontrado para o ORF, são colocados automaticamente *links* para cada uma das vias metabólicas em que este ORF pode estar presente (KEGG descrito no item 3.4.1.6);

3.4.1.2. **BLAST (NCBI)**

O blast é uma ferramenta de acesso via internet ou de instalação local que permite a comparação de uma seqüência com um banco de dados genéticos. Sua utilização no acesso via internet consiste em se colocar a seqüência que se quer comparar no campo determinado e rodar a procura. Dos parâmetros colocados como padrão, o único alterado normalmente é a retirada do filtro para regiões de baixa complexidade, como caudas poli-A e regiões com repetições de poucas letras (aminoácidos ou nucleotídeos). As seqüências devem estar em formato FASTA, ou seja, texto sem formatação, mas com quebras de linha e aproximadamente cinqüenta caracteres por linha. A primeira linha de cada seqüência é a do nome da seqüência, precedida de um “>”. Por exemplo:

```
>exemplo
ATGAAACTCTACAATCTTAAAGATCACAATGAGCAGGTCAGCTTTGCGC
GCAAAAATCAGGGGCTGTTTTTTCCGCACGACCTGCCGGAATTCAGCC
TTTACCGAATGTGAAAGTGGTTATCCTTTATACTGA
```

Programas para diferentes tipos de comparações segundo a tabela 3.1:

Programa	**Seqüência para busca (*query*)**	**Banco de Dados (*subject*)**
BLASTP	Aminoácidos	Aminoácidos
BLASTN	Nucleotídeos	Nucleotídeos
BLASTX	Nucleotídeos (traduzidos)	Aminoácidos
TBLASTN	Aminoácidos	Nucleotídeos
TBLASTX	Nucleotídeos (traduzidos)	Nucleotídeos (traduzidos)

Tabela 3.1 – Tipos de programas blast.

BLAST P

O blastp permite a comparação de seqüências de aminoácidos contra proteínas. Como mostrado na fig. 3.1, coloca-se a seqüência de aminoácidos no campo "*Search*" e realiza-se a procura contra proteínas (clicando em "BLAST!"). Pode-se regular o alvo da procura, através dos diferentes bancos de dados disponíveis para comparação (em "*Choose database*"). A opção "*Do CD-Search*" aciona a busca de domínios (exemplo de resultado na fig. 3.2).

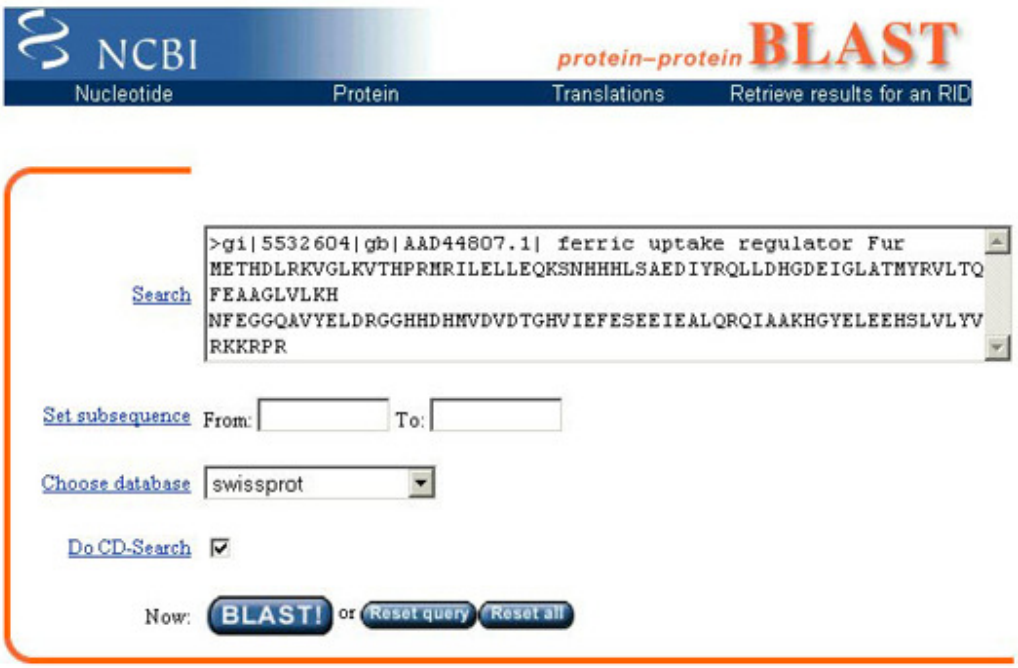


Figura 3.1 – Parte superior da página de entrada de dados no blastp.

Após o processamento da busca, é reportada uma página (fig. 3.2) contendo o resultado da busca de domínios, se a opção foi acionada; o número de identificação da procura – para facilitar a utilização do programa sem precisar esperar cada uma das buscas; e opções de formatação dos resultados. Após o tempo estimado, clicar em "FORMAT!".

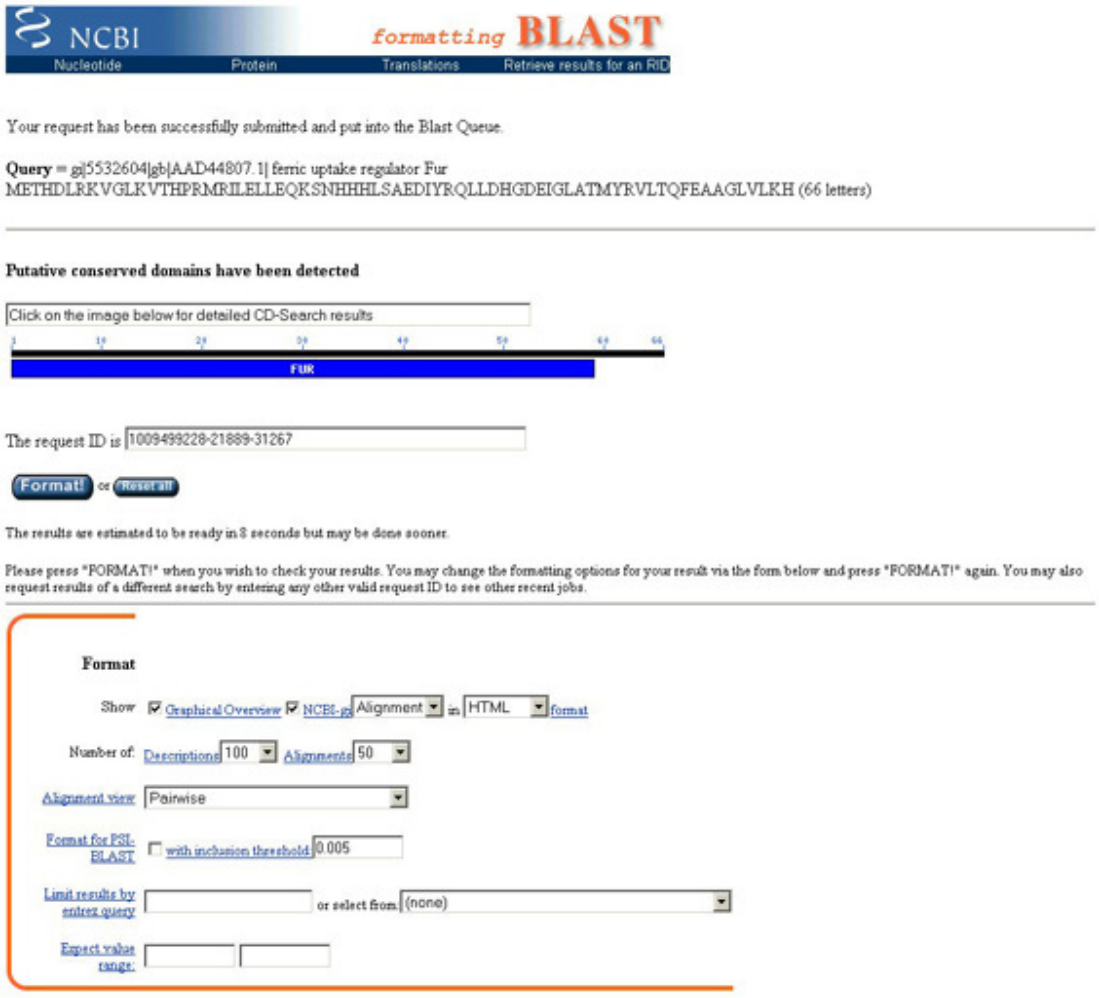


Figura 3.2 – Página de resultado intermediário do blastp, apresentando domínio encontrado, identificação da requisição e opções de formatação do resultado.

O resultado da busca é apresentado em uma página que está dividida aqui em três figuras para facilitar seu entendimento. Na figura 3.3 o resultado gráfico, permitindo visualização do tamanho dos genes homólogos em relação à seqüência inicial, além de diferentes cores que facilitam a análise.

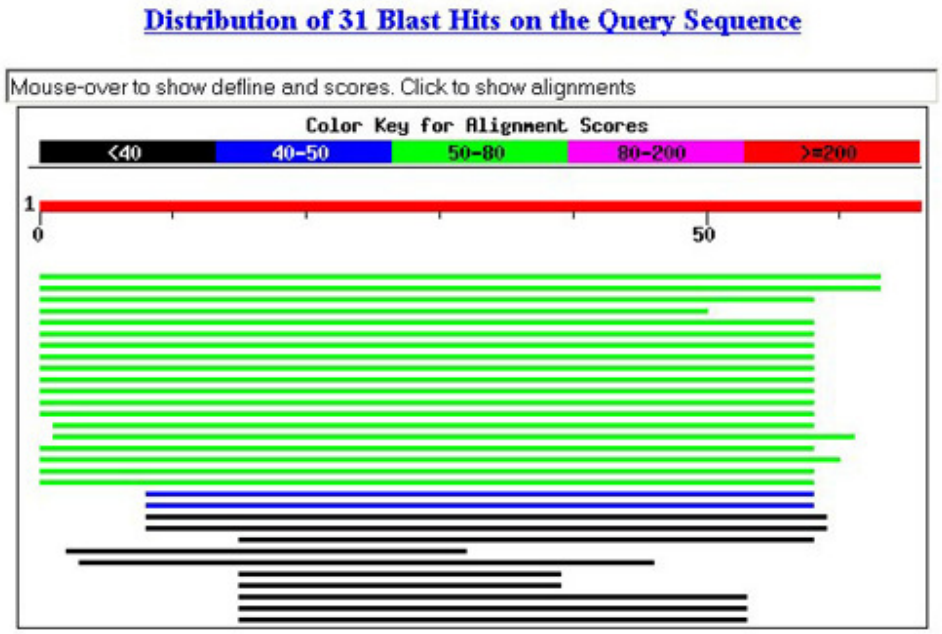


Figura 3.3 – Apresentação gráfica dos resultados, com variação de cor segundo o escore (*score)*, e graduação para comparação do tamanho da seqüência da busca (*query)* com os genes homólogos.

Na figura 3.4, encontra-se um exemplo da lista encontrada logo abaixo do gráfico, contendo os genes homólogos encontrados com alinhamento significante (aqui chamados de *hits*). Para cada um deles, apresentado em uma linha, temos: (1) Número de identificação em banco de dados genéticos, que é também um link para o registro deste gene; (2) Início do nome do produto do gene; (3) *Escore (Score)*, que também é um link para o alinhamento do query com este gene (figura 3.12); e (4) Valor E (*E value*), o principal parâmetro na escolha dos alinhamentos significativos.

Sequences producing significant alignments:		Score (bits)	E Value
gi|3913700|sp|Q52083|FUR_PSEPU	FERRIC UPTAKE REGULATION PRO...	80	9e-16
gi|417010|sp|Q03456|FUR_PSEAE	FERRIC UPTAKE REGULATION PROT...	75	3e-14
gi|1346056|sp|P33086|FUR_YERPE	Ferric uptake regulation pro...	70	6e-13
gi|3913693|sp|O68563|FUR_PSEFL	FERRIC UPTAKE REGULATION PRO...	68	2e-12
gi|2828519|sp|P45599|FUR_KLEPN	FERRIC UPTAKE REGULATION PRO...	68	3e-12
gi|120614|sp|P06975|FUR_ECOLI	FERRIC UPTAKE REGULATION PROT...	67	4e-12
gi|9911066|sp|Q57298|FUR_NEIMA	FERRIC UPTAKE REGULATION PRO...	64	4e-11
gi|3913699|sp|Q51008|FUR_NEIGO	FERRIC UPTAKE REGULATION PRO...	64	5e-11
gi|3913691|sp|O30330|FUR_ALCEU	FERRIC UPTAKE REGULATION PRO...	62	1e-10
gi|3913707|sp|O24755|FUR_VIBPA	FERRIC UPTAKE REGULATION PRO...	62	2e-10
gi|12644009|sp|P33087|FUR_VIBCH	FERRIC UPTAKE REGULATION PR...	62	2e-10
gi|585162|sp|P37736|FUR_VIBAN	FERRIC UPTAKE REGULATION PROT...	60	7e-10
gi|417012|sp|P33117|FUR_VIBVU	FERRIC UPTAKE REGULATION PROT...	60	9e-10
gi|3913696|sp|Q45765|FUR_BORPE	FERRIC UPTAKE REGULATION PRO...	60	9e-10
gi|3913692|sp|O30976|FUR_BRUAB	FERRIC UPTAKE REGULATION PRO...	58	2e-09
gi|3913688|sp|O07315|FUR_RHILV	FERRIC UPTAKE REGULATION PRO...	56	1e-08
gi|3913694|sp|P71333|FUR_HAEDU	FERRIC UPTAKE REGULATION PRO...	55	2e-08
gi|3913698|sp|Q48835|FUR_LEGPN	FERRIC UPTAKE REGULATION PRO...	54	6e-08
gi|1169780|sp|P44561|FUR_HAEIN	FERRIC UPTAKE REGULATION PRO...	54	6e-08
gi|3915681|sp|P48796|FUR_CAMJE	FERRIC UPTAKE REGULATION PRO...	47	6e-06
gi|3913697|sp|Q46463|FUR_CAMUP	FERRIC UPTAKE REGULATION PRO...	42	2e-04
gi|3913690|sp|O25671|FUR_HELPY	FERRIC UPTAKE REGULATION PRO...	39	0.002
gi|11386771|sp|Q9ZM26|FUR_HELPJ	FERRIC UPTAKE REGULATION PR...	38	0.004
gi|3913701|sp|Q55244|FUR_SYNP7	FERRIC UPTAKE REGULATION PRO...	30	0.52
gi|2493424|sp|Q28060|DSC3_BOVIN	DESMOCOLLIN 3A/3B PRECURSOR	27	5.4
gi|125960|sp|P23420|LAM4_XENLA	LAMIN L(III), ISOFORM 2 (LAM...	27	6.6
gi|125959|sp|P10999|LAM3_XENLA	LAMIN L(III), ISOFORM 1 (LAM...	27	6.7
gi|1706377|sp|P51110|DFRA_VITVI	DIHYDROFLAVONOL-4-REDUCTASE...	27	6.9
gi|125962|sp|P02545|LAMA_HUMAN	LAMIN A/C (70 KDA LAMIN)	27	7.4
gi|1346413|sp|P48679|LAMA_RAT	LAMIN A	27	8.5
gi|1346412|sp|P48678|LAMA_MOUSE	LAMIN A	26	9.3

Figura 3.4 – Lista de alinhamentos significativos com links para o registro de cada gene e para os alinhamentos.

Na figura 3.5 encontram-se exemplos de alinhamentos. Para cada alinhamento são descritos: Identificadores do gene; nome do produto; normalmente o nome do organismo; tamanho do gene; dados do alinhamento e o alinhamento. Os valores no início de cada linha do alinhamento indicam o número da primeira letra, e no final, da última. Estes valores permitem a verificação da cobertura do alinhamento, se este abrange início e fim de ambos os genes ou se apresenta homologia parcial.

```
                                                    Alignments

>gi|3913700|sp|Q52083|FUR_PSEPU FERRIC UPTAKE REGULATION PROTEIN (FERRIC UPTAKE REGULATOR)
          Length = 134

 Score = 79.7 bits (195), Expect = 9e-16
 Identities = 42/64 (65%), Positives = 52/64 (80%)

Query: 1   NFEGGQAVYELDRGGHHDHMVDVDTGHVIEFESEEIEALQRQIAAKHGYELEEHSLVLYV 60
           NF+GG AV+EL  GGHHDHMV+V+T  VIEF   EIE  QR+I A+HG+EL +H+LVLYV
Sbjct: 71  NFDGGHAVFELADGGHHDHMVNVETSEVIEFMDAEIEKRQREIVAEHGFELVDHNLVLYV 130

Query: 61  RKKR 64
           RKK+
Sbjct: 131 RKKK 134


>gi|417010|sp|Q03456|FUR_PSEAE FERRIC UPTAKE REGULATION PROTEIN (FERRIC UPTAKE REGULATOR)
          Length = 134

 Score = 74.7 bits (182), Expect = 3e-14
 Identities = 40/64 (62%), Positives = 49/64 (76%)

Query: 1   NFEGGQAVYELDRGGHHDHMVDVDTGHVIEFESEEIEALQRQIAAKHGYELEEHSLVLYV 60
           NF+GG AV+EL   GHHDHMV VDTG VIEF   EIE  Q++I  + G+EL +H+LVLYV
Sbjct: 71  NFDGGHAVFELADSGHHDHMVCVDTGEVIEFMDAEIEKRQKEIVRERGFELVDHNLVLYV 130

Query: 61  RKKR 64
           RKK+
Sbjct: 131 RKKK 134
```

Figura 3.5 – Exemplos de alinhamentos significativos entre o *query* e os genes (*subject*).

BLAST N

O blastn permite a comparação de nucleotídeos contra nucleotídeos. Coloca-se a seqüência no campo "*Search*" e realiza-se a procura ("BLAST!"). Da mesma forma, pode-se regular o alvo da procura através dos diferentes bancos de dados disponíveis para comparação. É indicado para altos valores de similaridade.

BLAST 2 Sequences

Permite o alinhamento entre duas seqüências quaisquer, apresentando resultados da comparação de forma gráfica e detalhada. Importante em diferentes momentos quando se busca uma definição precisa da homologia entre seqüências, ou mesmo na comparação de uma seqüência com ela mesma. Pode ser utilizada tanto com blastp ou blastn. A figura 3.6 apresenta a

entrada de dados de duas seqüências de aminoácidos, para um blastp sem filtro. Para rodar o programa, clica-se em "*Align*".

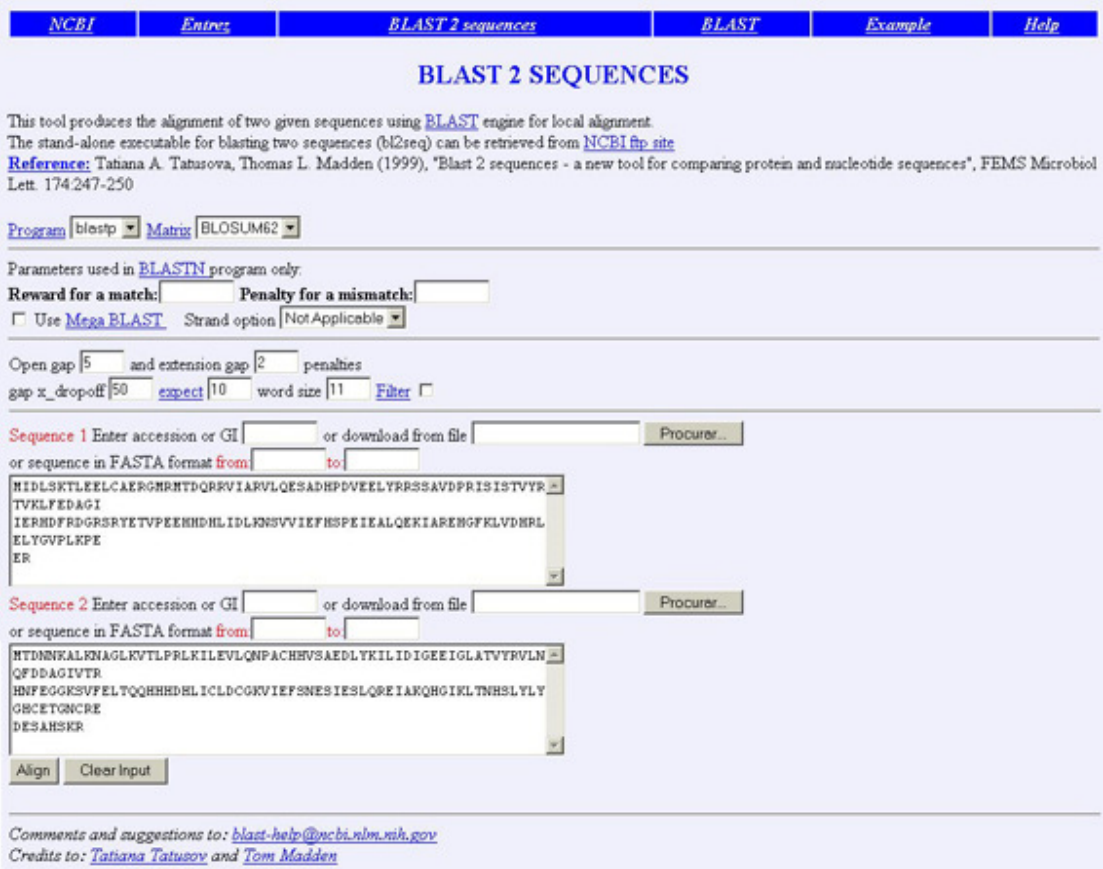


Figura 3.6 – Página de entrada de dados do Blast 2 Sequences

A figura 3.7 apresenta os resultados da comparação. Inicialmente, mostra os parâmetros utilizados na comparação. Abaixo destes, dados das seqüências alinhadas, e depois um gráfico mostrando os trechos homólogos. Na presença de trechos repetidos nas seqüências, seriam mostrados dois ou mais alinhamentos parciais. Quanto mais reta a linha do gráfico, mais perfeito é o alinhamento. A seguir, o alinhamento é mostrado com as barras horizontais, que facilitam a visualização de *gaps* e de possíveis desencontros no início ou final das seqüências, e com o alinhamento normal do blast, já descrito.

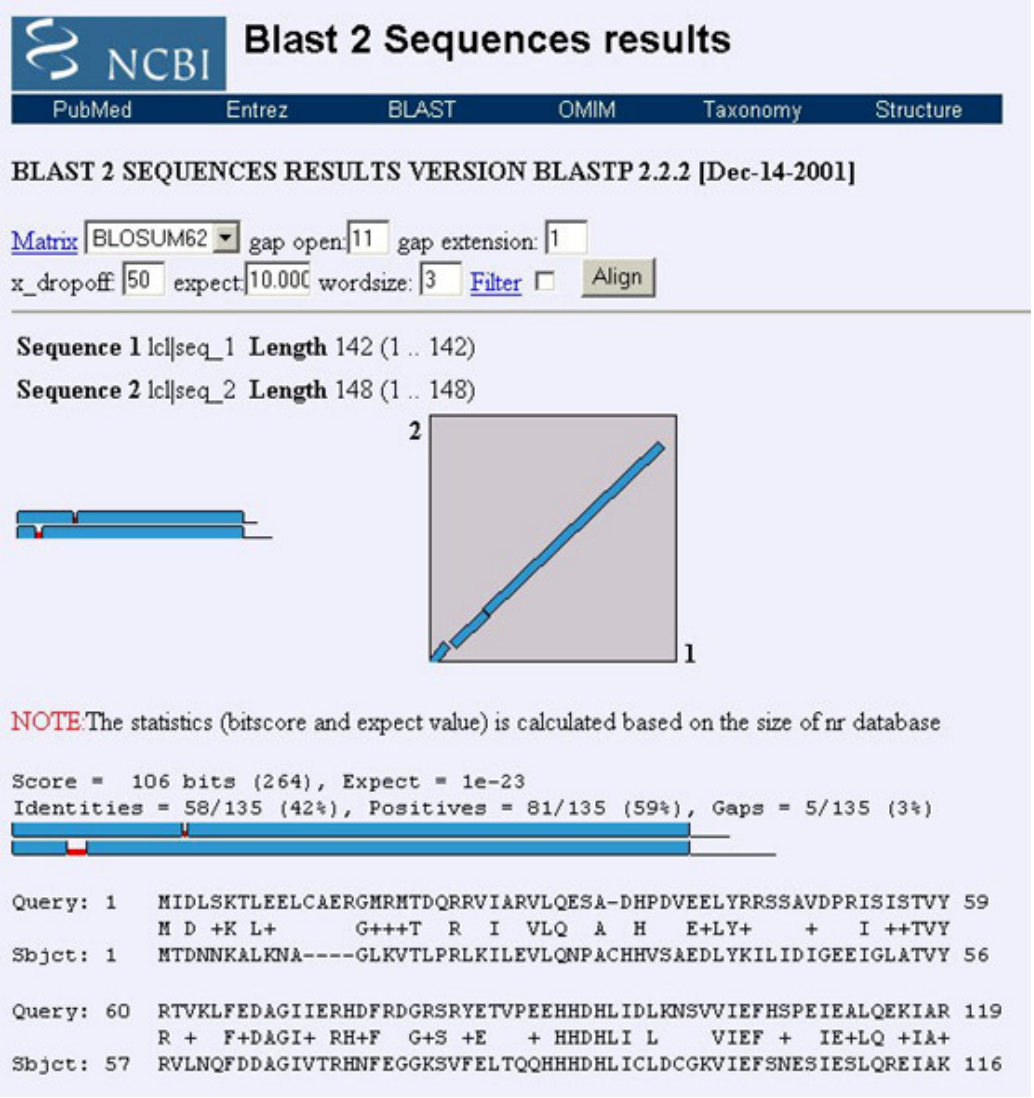
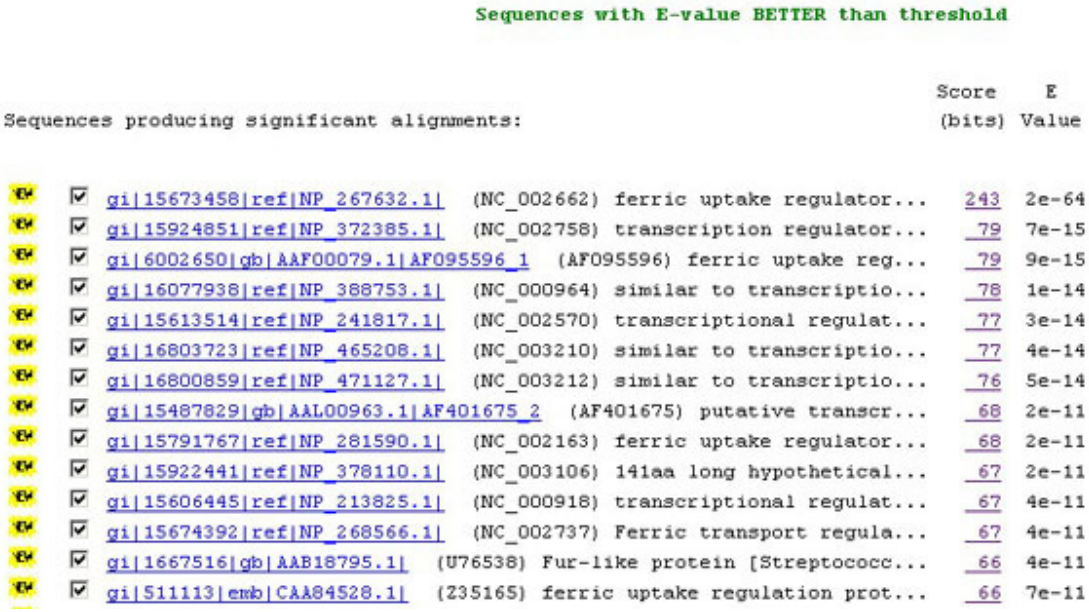


Figura 3.7 – Página de resultados do Blast 2 Sequences

PSI-BLAST

O psi-blast é uma ferramenta de alinhamentos múltiplos sucessivos. A partir do primeiro resultado de similaridade é construída uma matriz de pontuação de posicionamento específica, que irá direcionar as próximas comparações. A lista de seqüências homólogas estabiliza após alguns alinhamentos. O número destes depende da seqüência.

As figuras 3.8 e 3.9 mostram um exemplo de resultado obtido. Na primeira, os resultados da primeira interação, semelhantes a um blast comum. Na fig.3.9, os resultados da segunda interação. O banco de dados é restringido, os valores E do alinhamento do mesmo gene são alterados, e um marcador amarelo ou verde marca os genes que já faziam parte do grupo selecionado de melhores homologias e os que foram promovidos a este grupo.



Sequences with E-value BETTER than threshold

Sequences producing significant alignments:	Score (bits)	E Value
gi|15673458|ref|NP_267632.1| (NC_002662) ferric uptake regulator...	243	2e-64
gi|15924851|ref|NP_372385.1| (NC_002758) transcription regulator...	79	7e-15
gi|6002650|gb|AAF00079.1|AF095596_1 (AF095596) ferric uptake reg...	79	9e-15
gi|16077938|ref|NP_388753.1| (NC_000964) similar to transcriptio...	78	1e-14
gi|15613514|ref|NP_241817.1| (NC_002570) transcriptional regulat...	77	3e-14
gi|16803723|ref|NP_465208.1| (NC_003210) similar to transcriptio...	77	4e-14
gi|16800859|ref|NP_471127.1| (NC_003212) similar to transcriptio...	76	5e-14
gi|15487829|gb|AAL00963.1|AF401675_2 (AF401675) putative transcr...	68	2e-11
gi|15791767|ref|NP_281590.1| (NC_002163) ferric uptake regulator...	68	2e-11
gi|15922441|ref|NP_378110.1| (NC_003106) 141aa long hypothetical...	67	2e-11
gi|15606445|ref|NP_213825.1| (NC_000918) transcriptional regulat...	67	4e-11
gi|15674392|ref|NP_268566.1| (NC_002737) Ferric transport regula...	67	4e-11
gi|1667516|gb|AAB18795.1| (U76538) Fur-like protein [Streptococc...	66	4e-11
gi|511113|emb|CAA84528.1| (Z35165) ferric uptake regulation prot...	66	7e-11

Figura 3.8 – Parte da página de resultados de primeira interação do Psi-blast.

Figura 3.9 – Parte da página de resultados de segunda interação do Psi-blast.

BLAST X

O blastx compara uma seqüência de nucleotídeos traduzidos com proteínas, nas seis fases de leitura possíveis (três de cada lado). A forma de utilização é semelhante ao blastn. Seu uso é recomendado para seqüências de sequenciamento de qualidade não comprovada, ou para detecção de ORFs que possuam mutações pontuais que insiram ou deletem um número de bases diferente de um múltiplo de três.

Este tipo de mutação, nomeado nos projetos como *frameshift*, causa uma mudança na fase de leitura, que não pode ser detectada pelo blastp, utilizado normalmente. Busca-se blastx no caso de homologias parciais, para verificação da existência de *frameshift*.

3.4.1.3. COGnitor

COGnitor (Tatusov et al., 2001) – compara uma seqüência com banco de dados genéticos pré-classificados em COGs. Apresenta o resultado da comparação (fig. 3.11) e informações sobre os COGs que apresentaram homologia. O processo de comparação é semelhante ao do blast. Coloca-se a seqüência em formato fasta e roda-se a procura (compare to COGs) – fig.3.10. Também são considerados os valores E e de identidade para determinar até quais genes podem ter homologia significativa.

Figura 3.10 – Página de entrada de dados do programa COGnitor.

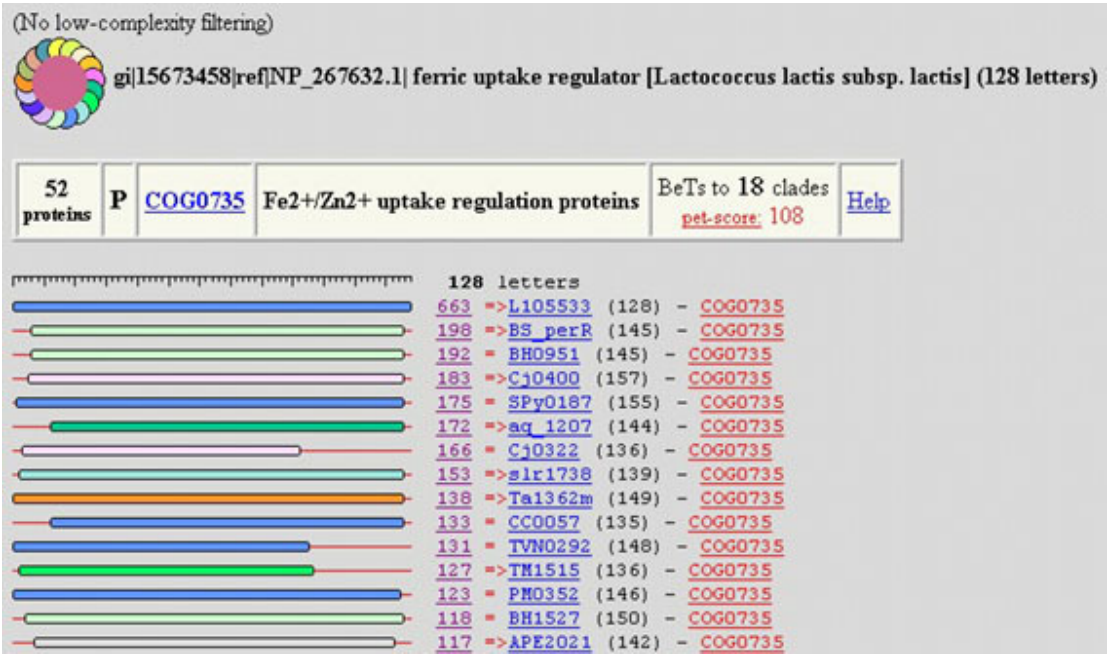


Figura 3.11 – Página de resultados do programa COGnitor – parte superior.

3.4.1.4. **PFAM**

PFAM (Bateman et al., 2002) – compara seqüência com banco de dados de domínios ou regiões de genes conservadas. Como mostrado na figura 3.12, coloca-se a seqüência no campo "*Protein sequence query*" e executa-se a procura (Enviar Consulta). Os resultados são apresentados com valor E e escore para cada domínio homólogo (fig. 3.13). Para cada domínio encontrado, tenha ele uma boa homologia ou não, existe um link para maiores detalhes do domínio, como no exemplo da fig. 3.14.

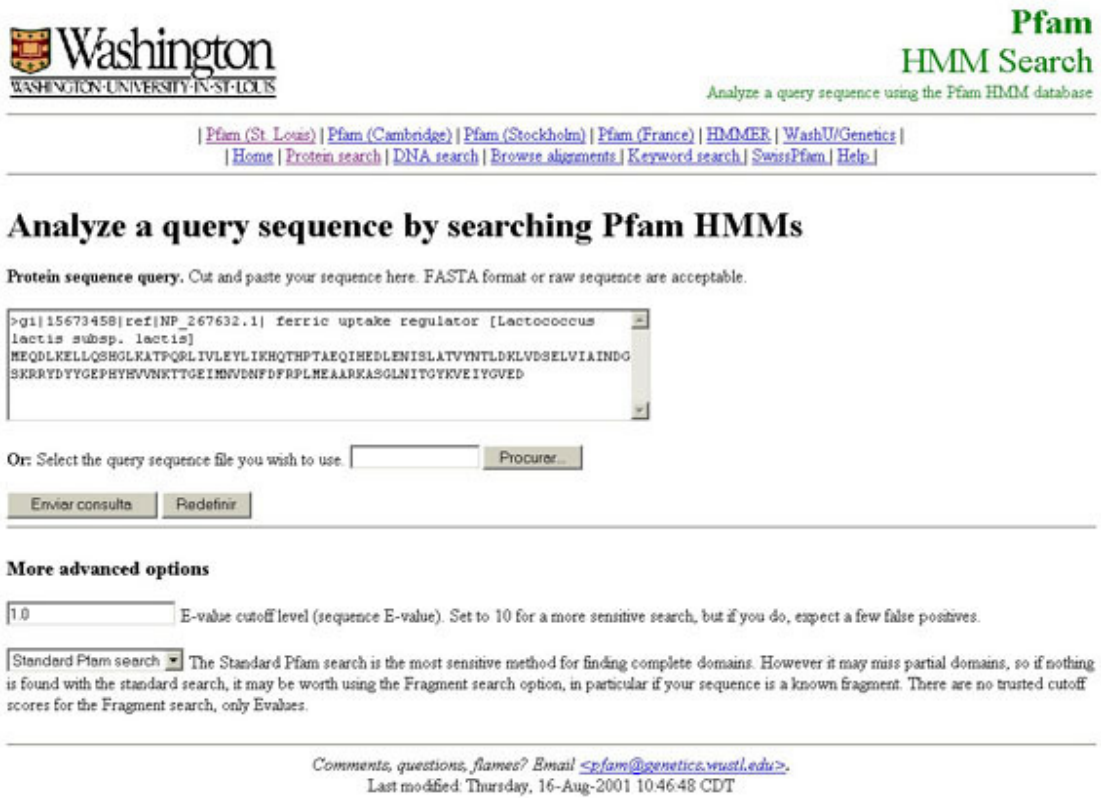


Figura 3.12 – Página de procura de proteínas do programa PFAM.

Pfam 6.6 (St. Louis) : Home | Analyze a sequence | Browse alignments | Text search | Swisspfam | Help |

Starting search. Estimated time: 16 seconds (assuming all Wulfpack nodes are running). Please wait...

Pfam HMM search results

[Go here for an explanation of the format of the results]

	Model	Seq-from	Seq-to	HMM-from	HMM-to	Score	E-value	Description
!!	FUR	11	124	1	124	84.4	1.2e-22	Ferric uptake regulator family
	HTH_5	12	96	1	82	-7.6	0.53	Bacterial regulatory protein, arsR family

overlap

FUR 11-124

```
Alignments of top-scoring domains:
FUR: domain 1 of 1, from 11 to 124: score 84.4, E = 1.2e-22
                  *->eaGlKiTpqRlkILevleksdeeHlsAEevYrellneedpniSlATV
                     + GlK+TpqRl +Le+l k+ + H++AE +++ l+     niSlATV
  gi|1567345    11    SHGLKATPQRLIVLEYLIKHQT-HPTAEQIHEDLE-----NISLATV 51

                  YRtLklleeaGivkrlefeggesrfElaqeaghHHdHlICekCGk...vi
                  Y+tL  l ++ +v  ++ +  ++r++     g+ H+H++   +G+  +v
  gi|1567345    52 YNTLDKLVDSELVIAINDG-SKRRYDYY---GEPHYHVVNKTTGEimnVD 97
```

Figura 3.13 – Página de resultados do programa PFAM.

Pfam entry: FUR

Accession number:	PF01475
Definition:	Ferric uptake regulator family
Author:	Bateman A
Alignment method of seed:	Clustalw
Source of seed members:	Prodom_2003 (release 99.1)
Gathering cutoffs:	0 0
Trusted cutoffs:	15.50 15.50
Noise cutoffs:	-14.70 -14.70
HMM build command line:	hmmbuild -F HMM SEED
HMM build command line:	hmmcalibrate --seed 0 HMM
Reference Number:	[1]
Reference Medline:	99003343
Reference Title:	Binding of the fur (ferric uptake regulator) repressor of
Reference Title:	Escherichia coli to arrays of the GATAAT sequence.
Reference Author:	Escolar L, Perez-Martin J, de Lorenzo V;
Reference Location:	J Mol Biol 1998;283:537-547.
Database reference:	PFAMB; PB055609;
Database Reference	INTERPRO; IPR002481;
Comment:	This family includes metal ion uptake regulator proteins,
Comment:	that bind to the operator DNA and controls transcription
Comment:	of metal ion-responsive genes. This family is also known
Comment:	as the FUR family.
Number of members:	71

Figura 3.14 – Página de detalhes do domínio encontrado – PFAM.

3.4.1.5. PSORT

PSORT (Nakai & Horton, 1999) – o processo de comparação é semelhante aos demais (fig. 3.15), sendo exibidos como resultado dados da provável localização na célula do produto do gene, em valores de zero a um (fig. 3.16).

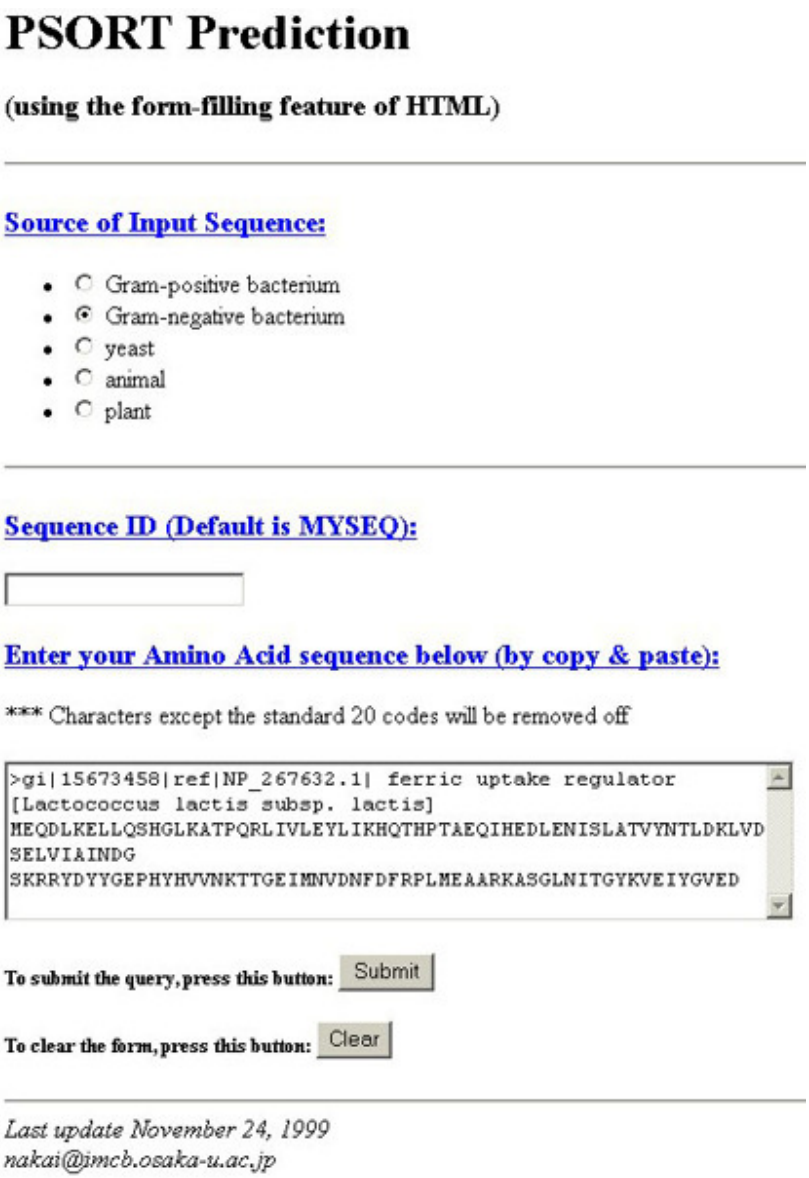


Figura 3.15 – Página de entrada de dados do programa PSORT.

Query Information

```
ORIGIN Gram-negative bacterium
BEGIN
>MYSEQ
 GIREFNPFER RICPTAKERE GLATRLACTC CCSLACTISS SPLACTISME
 QDLKELLQSH GLKATPQRLI VLEYLIKHQT HPTAEQIHED LENISLATVY
 NTLDKLVDSE LVIAINDGSK RRYDYYGEPH YHVVNKTTGE IMNVDNFDFR
 PLMEAARKAS GLNITGYKVE IYGVED
```

Result Information

```
PSORT --- Prediction of Protein Localization Sites
                                                version 6.4(WWW)
MYSEQ          176 Residues
*** Warning: The 1st amino acid is not methionine
Species classification: 2

*** Reasoning Step: 1

Lipop: Examining lipoprotein consensus (Klein et al.:modified)
     Possible modific. site: -1  CRend: 10
McG: Examining signal sequence (McGeoch)
     Length  of  UR:   5
     Peak Value of UR:   1.01
     Net Charge of CR: 1
     Discriminant Score:     -8.93
GvH: Examining signal sequence (von Heijne)
     Signal Score (-7.5): -2.56
     Possible cleavage site: 46
>>> Seems to have no N-terminal signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 1
ALOM: Finding transmembrane regions (Klein et al.)
     count: 0  value:   0.26 threshold:   0.0
     PERIPHERAL  Likelihood =  0.26
    modified ALOM score:  -0.55
Rule: cytoplasmic protein

*** Reasoning Step: 2


----- Final Results -----

               bacterial cytoplasm --- Certainty= 0.180(Affirmative
      bacterial periplasmic space --- Certainty= 0.000(Not Clear) -
         bacterial outer membrane --- Certainty= 0.000(Not Clear) -
         bacterial inner membrane --- Certainty= 0.000(Not Clear) -


----- The End -----
```

Figura 3.16 – Página de resultados do programa PSORT.

3.4.1.6. KEGG

KEGG (Kanehisa et al., 2002) – a Enciclopédia de Genes e Genomas de Kyoto (KEGG) apresenta um sistema de busca com diversas alternativas. Inclui vias metabólicas e regulatórias, através da procura por nome de vias, nomes de produtos ou códigos (número EC), ou ainda pela comparação de uma seqüência ao seu banco de dados.

O processo de busca por similaridade de seqüência consiste na colocação de uma seqüência no campo apropriado, seleção de aminoácidos ou nucleotídeos, e execução da busca (*Exec*). As figuras 3.17 e 3.18 mostram parte da página deste serviço no site do KEGG.

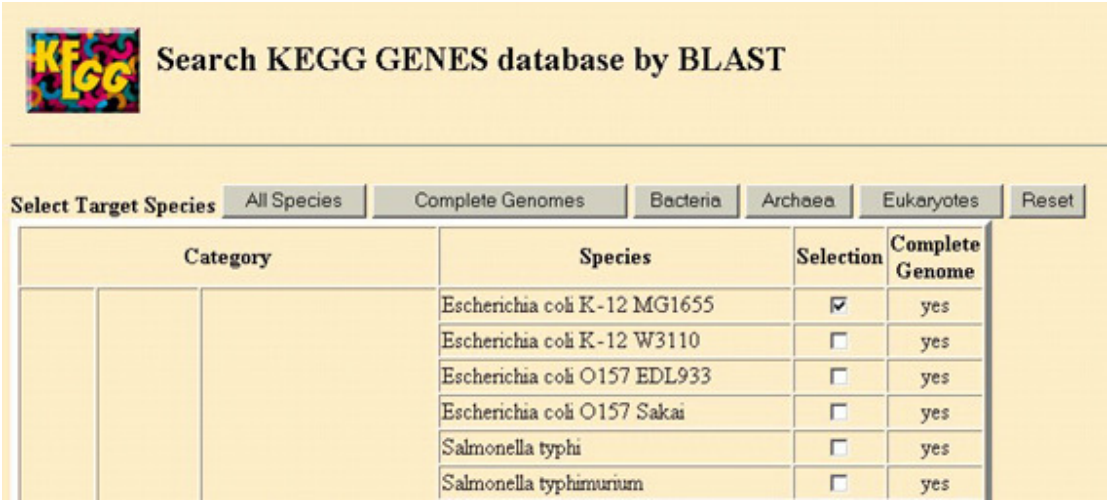


Figura 3.17 – Parte inicial da página de procura por genes do banco de dados do KEGG, utilizando o blast.

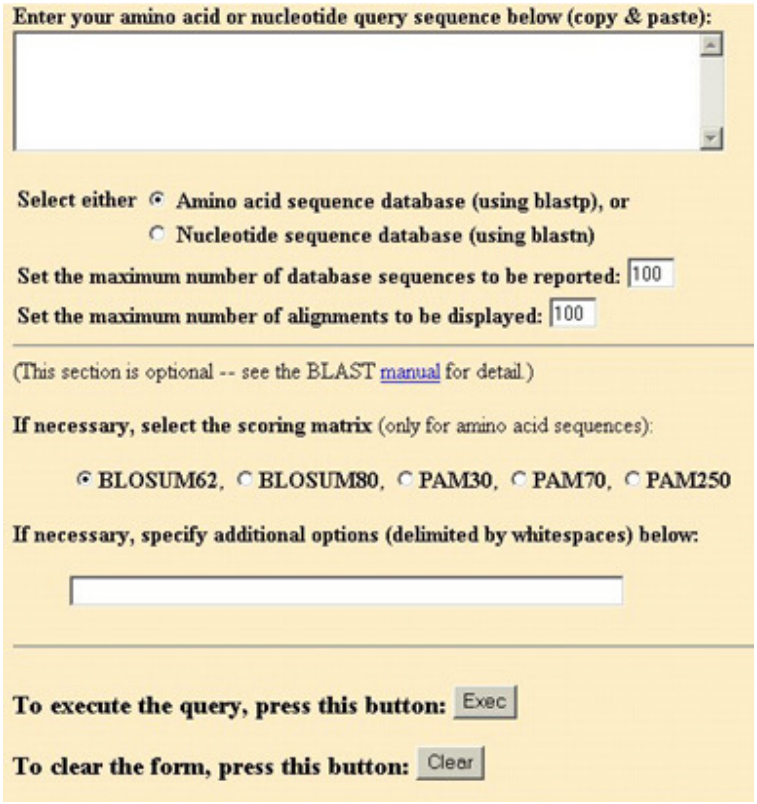


Figura 3.18 – Parte final da página de procura por genes do banco de dados do KEGG, utilizando o blast.

Pode-se utilizar o banco de dados para localizar genes a partir do número EC, como no exemplo da fig. 3.19. Na figura seguinte, 3.20, encontra-se o resultado da busca, e na fig. 3.21 um dos mapas citados.

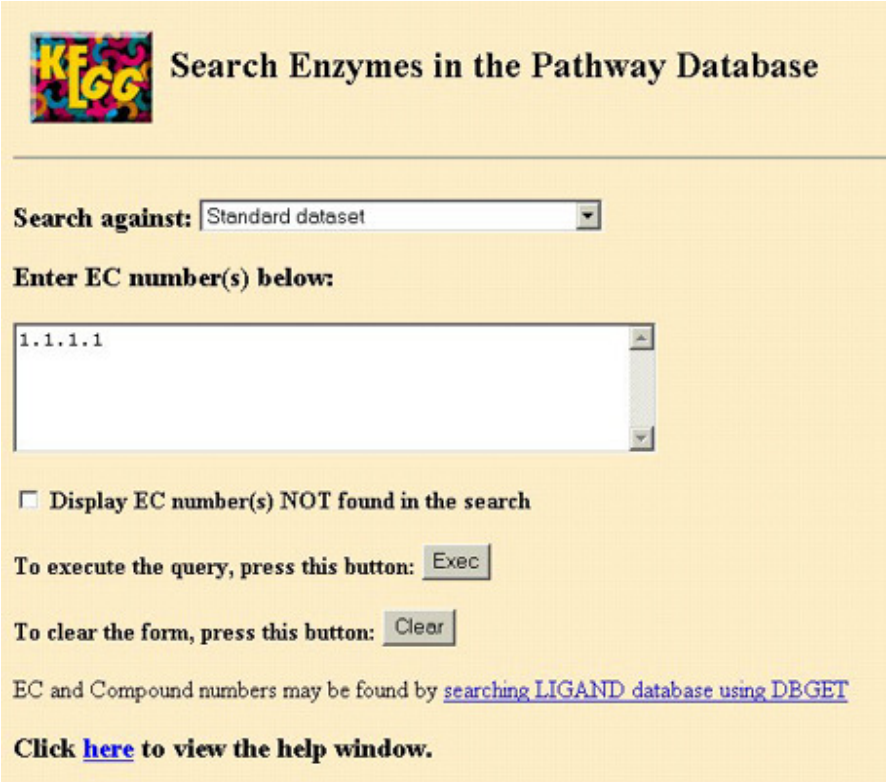


Figura 3.19 – Busca de informações no KEGG por número EC.

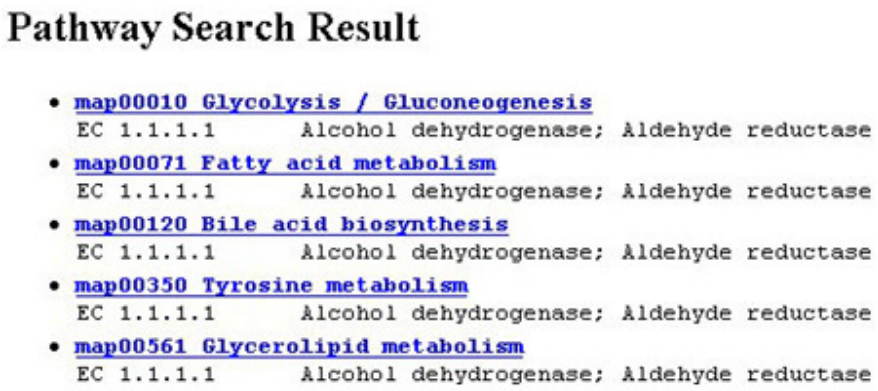


Figura 3.20 – Resultado de busca por número EC no KEGG – mapas onde o EC citado é um dos componentes da via.

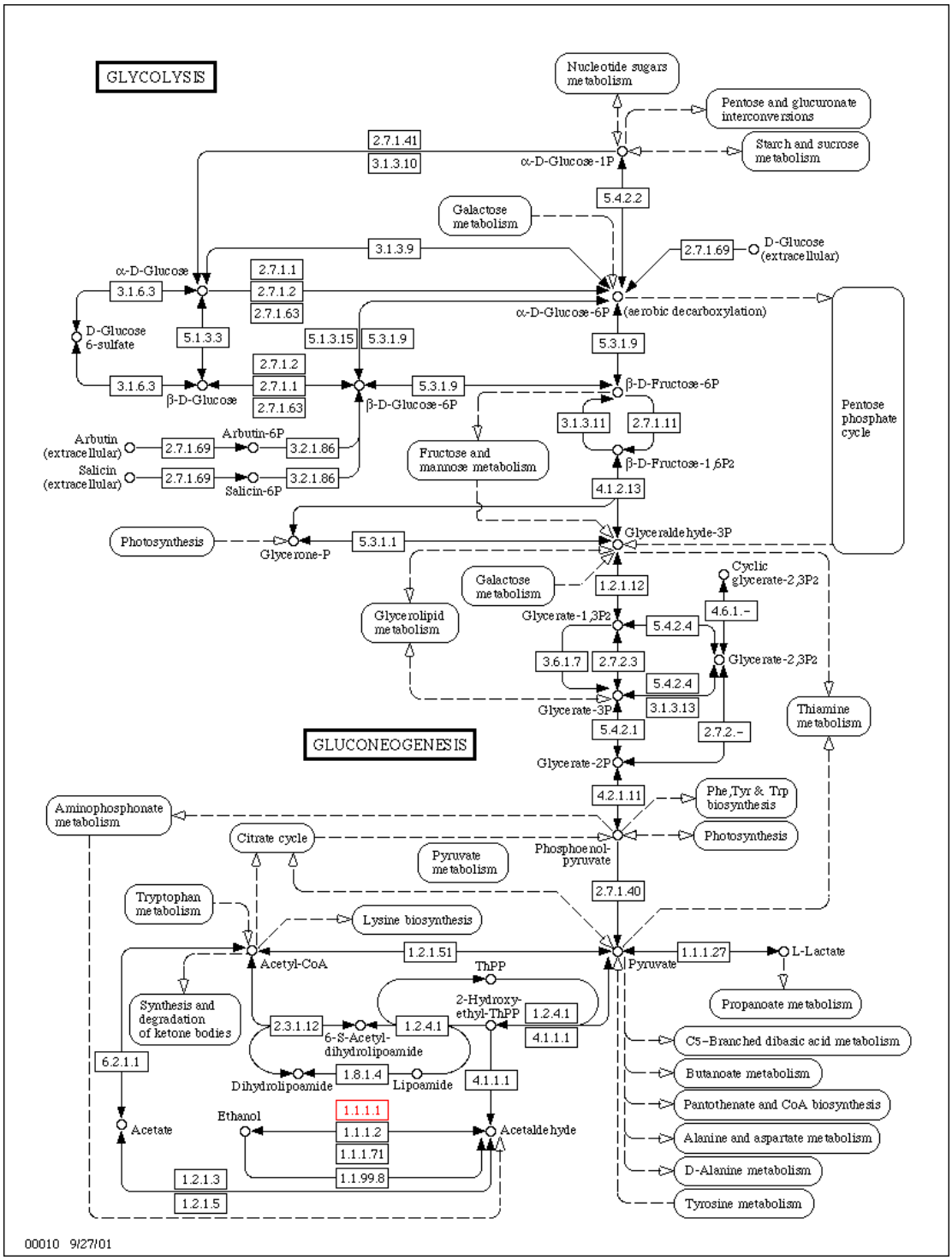


Figura 3.21 – Exemplo de mapa de via metabólica do KEGG – "map 00010 Glycolysis / Gluconeogenesis"

3.4.2. Xylella – CVC

Cada projeto apresentou diferentes fases de anotação. Neste item serão apresentadas diferentes fases de maneira a evitar sobreposição e demonstrar a evolução do método empregado.

Foi o primeiro processo de anotação no projeto Genoma. A anotação foi feita por trecho de DNA - cosmídeo, e após esta fase algumas pessoas realizaram a anotação por categorias. Não havia a estrutura atual de anotação, e os resultados eram formatados e enviados ao laboratório central.

Neste projeto, a anotação dos genes consiste na postulação de função para a proteína correspondente ao gene e preparação das informações para a submissão a um banco de dados genético. O programa utilizado para esta organização dos dados foi o Sequin, que formata os dados para submissão ao GenBank. Neste programa, para cada gene são introduzidas as informações (exemplo da anotação da *Xylella*):

- ✓ Descrição – função postulada; em region/protein/description;
- ✓ Comentários – justificativa; em properties/coments;
- ✓ Categorizar segundo classificação de *E. coli;*
- ✓ Submissão do cosmídeo anotado para o laboratório central através da página de serviços.

Para a comparação dos genes com os ORFs da *Xylella*, utilizam-se os resultados de comparação no blastp. Um exemplo de interpretação dos resultados é (segundo a análise do valor E):

- ✓ Maioria dos hits mais fortes que 1e-5 correspondem à mesma proteína: A função postulada é a dos hits de valor menor;
- ✓ Alguns hits fortes (mais do que 1e-5) correspondem a uma proteína e outros a outra: Auxílio do Psi-b*last* para decidir entre as opções, mas ambas devem constar no comentário;
- ✓ Hits mais fortes entre 1e-5 e 1e-3: Utilização do Psi-*blast* para confirmar se está clara a função ou não;

- ✓ Hits mais fortes tem valor inferior a 1e-3: Mesmo procedimento que a anterior caso haja interesse na orf.

3.4.3. *Xanthomonas*

Diferentemente do previsto no plano inicial, o sistema de anotação adotado no projeto não foi o de resolução por perguntas, mas um sistema composto por três partes, independente dos métodos adotados pelos chefes de cada categoria: na primeira etapa, os membros de cada categoria verificaram os ORFs a eles destinados e fizeram uma anotação inicial; num segundo momento, os chefes de categoria verificaram estes ORFs e deram seu aval; e finalmente, num terceiro estágio, um grupo chamado *scan team* realizou a verificação final dos ORFs.

Os chefes de categoria tiveram liberdade para definir as próprias regras internas aos grupos, já que ficaram responsáveis pelas informações obtidas. Houve diferentes sistemas de anotação, e a seguir serão descritos os métodos empregados nas duas categorias em que o autor participou (V e VII).

O Laboratório de Bioinformática (LBI) do Instituto de Computação (IC) da UNICAMP disponibilizou uma página com ferramentas de anotação e informações obtidas de forma automatizada, descrita anteriormente no item 3.4.1.1.

Além das ferramentas e informações apresentadas nas páginas de anotação, a estrutura criada permitiu a troca de informações através da disponibilização de relatórios parciais, da troca de e-mails para as diferentes listas criadas, dos links para outros serviços, das trocas de dados e também pela assessoria em casos de dúvidas, tanto entre a coordenação e os membros, como entre os membros da anotação.

É possível obter muita informação sobre cada ORF, mas existem informações mais importantes e uma ordem mais adequada para analisá-las:

- ✓ O primeiro passo é verificar o códon inicial (*start codon*), ou seja, o tamanho mais provável do ORF;
- ✓ Verifica-se a qualidade da seqüência, que é boa na maioria absoluta dos casos (boa em todos os casos após ter terminado a fase de sequenciamento);
- ✓ Verifica-se o resultado do blastp, onde serão analisados os *hits* mais fortes. A pré-categorização já separou os ORFs com hits muito fracos (*E-value* maior do que e-5), enviando-os para as categorias VIII e IX. Então se verificam os hits mais fortes e a cobertura dos mesmos (*coverage query e coverage subject*). Quanto menor o *E-value*, melhor o hit, e quanto mais próxima de 100% a cobertura, melhor.

Podem ocorrer casos de fácil solução, quando a maioria dos hits fortes bate com um mesmo gene, ou mais complicados, quando existem hits fortes com diferentes genes. No segundo caso, exige-se uma pesquisa mais aprofundada para verificação dos genes, se são nomes diferentes para o mesmo gene, se são genes naturalmente semelhantes, ou se é o caso de uma anotação anterior errada. Ainda pode ser necessária a busca da vizinhança do gene para definir a inclusão num *cluster* ou *operon*. De qualquer maneira, exige-se um estudo mais detalhado, e talvez até a consulta a um especialista na área para a resolução do caso.

Existem outras ferramentas para ajudar nesta decisão:

- ✓ O blastx apresenta uma visão diferente da mesma busca, podendo mostrar um resultado mais definido nos casos de inserções e deleções;
- ✓ O Psi-blast gera buscas sucessivas, e pode ajudar a definir o gene mais próximo;
- ✓ Dados adicionais como COG e PFAM podem auxiliar a escolha;
- ✓ Os organismos dos genes homólogos também são importantes, já que é dada preferência à própria *Xanthomonas*, e a organismos mais próximos.

Não é possível criar regras aplicáveis a todos os casos, mas existem várias ferramentas onde buscar informações, e o projeto definiu quais dados

seriam prioritários. O padrão principal de busca é o blastp sem filtro, de onde é retirado o valor E oficial de cada ORF. Os demais recursos do blast podem ser usados, mas para os casos de dúvida, e acompanhados das demais informações auxiliares.

A anotação por categorias verificará a anotação anterior, fazendo o caminho reverso e buscando os genes que deveriam estar no genoma. Neste caso, a maioria das anotações foi confirmada, mas alguns casos foram corrigidos em função do contexto mais completo da anotação.

O nome do gene seguiu sempre que possível o padrão *E. coli*, de três letras minúsculas, e se necessário uma letra maiúscula ou número. Os casos de genes novos foram definidos e padronizados no final do projeto. A seguir as particularidades do processo de anotação de cada categoria.

3.4.3.1. Anotação na Categoria V – Processos Celulares

Coordenador: João Meidanis
Membros: **Eduardo Fernandes Formighieri**; Jeny Rachid dos Santos; Nilce Maria Martinez Rossi

Há citações de dois tipos de categorias, sendo a primeira a categorização oficial do projeto (anexo 2) e a segunda utilizada apenas na anotação da categoria V do mesmo, que é a de famílias de proteínas de transporte organizada pelo Dr. Milton Saier (página inicial do site na fig. 3.22). Esta segunda define o número TC de cada gene associado a transporte.

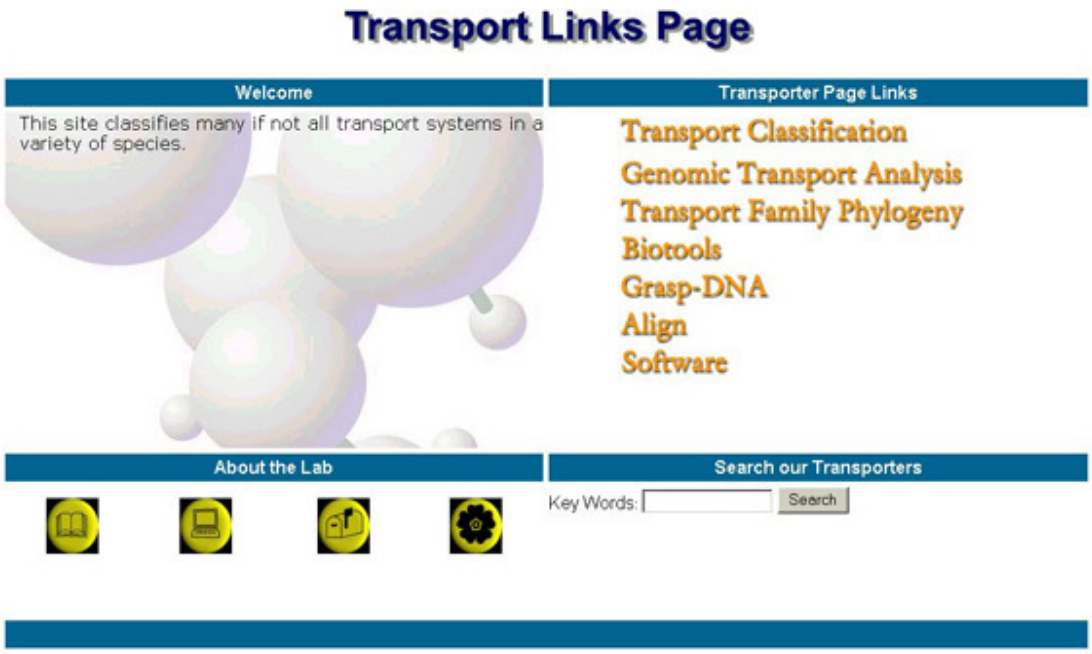


Figura 3.22 – Página do *site* de transporte do Dr. Milton Saier.

O sistema de anotação adotado nesta categoria foi composto por três fases principais. Num primeiro momento, os ORFs foram divididos por faixas da numeração entre os quatro integrantes do grupo (autor responsável pela faixa entre 6000 e 7999).

Os seguintes passos foram seguidos:

- ✓ Checar se o códon inicial está correto, e corrigir se for o caso;
- ✓ Verificar categorias primária e secundária;
- ✓ Verificar nome do gene e do produto;
- ✓ Verificar a presença de *frameshift*;
- ✓ Determinar o número TC;
- ✓ Anotar observações, quando for o caso.

A verificação do número TC é realizada utilizando-se o blastp contra banco de dados de proteínas relacionadas a transporte, na página de blast local da UNICAMP – figura 3.23.

Figura 3.23 – Parte da página de blast da UNICAMP – serviço de banco de dados de proteínas relacionadas a transporte.

A segunda fase ocorreu pela anotação por famílias de transporte, sendo o autor responsável pelas famílias:

- ✓ 1.B.14 – *The Outer Membrane Receptor (OMR) Family*;
- ✓ 2.C.1 - *The TonB-ExbB-ExbD / TolA-TolQ-TolR (TonB) Family of Auxiliary Proteins for Energization of Outer Membrane Receptor (OMR)-Mediated Active Transport*;
- ✓ 9.A.1 – *The Polysaccharide Transporter (PST) Family*;
- ✓ 1.B.18 – *The Outer Membrane Auxiliary (OMA) Protein Family*;
- ✓ 8.A.3 – *The Cytoplamic Membrane-Periplasmic Auxiliary-1 (MPA1) protein with Citoplasmic (C) Domain (MPA1-C or MPA1+C) Family).*

Inicialmente a busca foi baseada nos números TC definidos na fase 1, e em seguida nos sistemas destinados a cada anotador. Para cada sistema, foram procurados os genes correspondentes para a descrição de cada família de transporte envolvida. Nesta fase, ocorreu a descrição de ORFs pertencentes a outras categorias, principalmente à VII – Patogenicidade, Virulência e Adaptação, devido à preferência dada à mesma devido à sua importância. Ou seja, ORFs relacionadas a transporte, mas que têm envolvimento com a categoria VII tiveram a categoria primária colocada como VII, mas foram incluídas nos relatórios da categoria de transporte.

A diferença principal da anotação desta categoria em relação às demais é a utilização dos números TC, e da classificação por famílias de proteínas de transporte. Após a anotação por famílias de transporte e o envio de relatórios, restaram algumas dúvidas, e houve a inclusão de ORFs vindos de outras categorias que estavam terminando sua anotação. Nesta fase final, os casos foram discutidos entre os componentes do grupo e os ORFs novos foram anotados.

3.4.3.2. Anotação na Categoria VII – Patogenicidade, Virulência e Adaptação.

Coordenadores: Jesus A. Ferro; Luis Roberto Furlan; Rui P. Leite Jr.
Membros e subcategorias: segundo tabela 3.2.

Membro	Subcategoria
Alessandra de Souza	D, H
Alexandre do Amaral	C
Ana Cristina Dávila	C
Antonio Rossi Filho	H
Christian Greggio	
Daniela Truffi	A, C
Eduardo Formighieri	**A, B, E**
Eduardo Hilario	
Eliana Lemos	Pil
Haroldo Pereira Jr	C, F, H
Linda Lee	A, B, E, H
Luciano Oliveira	
Luis Antonio Peroni	C
Luis Eduardo Aranha Camargo	B, C
Manoel Victor Lemos	Pil
Marco Aurelio Takita	A, B, H
Marcos Antonio Machado	
Maria Teresa Novo	C
Marilia Franco	A, B
Regina Maria Barretto Cicarelli	G
Roberto Noda	C, G

Tabela 3.2 – Distribuição das subcategorias entre os integrantes da categoria VII de anotação.

Numa primeira fase, os ORFs destinados à categoria VII foram distribuídos aleatoriamente aos integrantes da categoria, cabendo oito ORFs por pessoa, para uma verificação inicial da anotação automática.

Os seguintes passos foram seguidos:

- ✓ Checar se o códon inicial está correto, e corrigir se for o caso;
- ✓ Verificar categorias primária e secundária;
- ✓ Verificar nome do gene e do produto;
- ✓ Conferir se o *accession code* confere com o melhor *hit*;
- ✓ Verificar no *medline* se existe alguma informação importante para colocar no campo *notepad*;

- ✓ No caso de informações muito seguras e evidentes, adicionar ao campo *remarks*.

Na segunda fase, a anotação foi realizada através de grupos responsáveis por subcategorias, segundo a tabela 3.2. Os resultados da anotação foram enviados na forma de relatórios para os coordenadores da categoria e posteriormente disponibilizados na página de anotação do projeto. Os chefes de categoria têm permissão para mudar a categoria primária dos ORFs. Foram gerados relatórios para as três categorias.

Além da verificação inicial nos casos de ORFs novas na categoria, os grupos responsáveis por subcategorias desenvolveram um relatório sobre cada subcategoria, incluindo:

- ✓ Genes presentes;
- ✓ Função dos mesmos;
- ✓ Possíveis envolvimentos em mecanismos de patogenicidade;
- ✓ Atualização do mesmo relatório após prazo dado;
- ✓ Inclusão de figuras e/ou esquemas, se necessário;
- ✓ Referências bibliográficas importantes.

Assim como no caso das famílias da categoria de transporte, houve uma revisão bibliográfica das subcategorias, a busca dos genes presentes nos principais bancos de dados contra o DNA da *Xanthomonas*, a verificação da falta ou sobra de genes nos clusters encontrados, e o relato dos processos celulares possivelmente presentes ou ausentes.

3.4.4. Xylella – PD

Este item descreve a estrutura adotada na anotação do organismo e informações sobre processos não descritos anteriormente.

Na fase 1-A cada anotador ficou responsável por 40 ORFs, segundo lista disponibilizada em página da internet. Esta página foi a mesma para as fases

seguintes, sendo a lista de cada anotador atualizada. Nesta fase foram verificados:

- ✓ Nome do produto do ORF;
- ✓ Nome do gene;
- ✓ Categorias primária e secundária.

Na fase 1-B, houve redistribuição dos ORFs não anotadas na fase anterior, sendo distribuídas 4 ou 5 ORFs para cada um. O objetivo era o mesmo da fase 1-A. Na fase 2 os ORFs foram redistribuídos para que cada anotador verificasse ORFs anotados por outra pessoa (checagem dupla). Quarenta e três para cada um. Os objetivos foram:

- ✓ Confirmar a anotação da fase 1;
- ✓ Indicação de *frameshift* ou *point mutation*;
- ✓ Ajuste de códon inicial.

ORFs que deveriam ser apagadas foram organizadas em listas e enviadas à central de bioinformática, que posteriormente fez a remoção.
Na fase 3 cada anotador recebeu um intervalo diferente de ORFs, e foram verificados:

- ✓ Confirmar a anotação da fase 2;
- ✓ Verificar o número EC ou TC;
- ✓ Verificar os mapas do KEGG e do Ecocyc (fig.3.24) em que o ORF seja um dos componentes – se via está completa ou não, e buscar genes faltantes;
    - o Comparação do gene faltante com genoma - blastn;
    - o Comparar similar do componente faltante com clones de shotgun - blastn;
    - o Comparar seqüência similar à do componente faltante com ORFs - blastp;
    - o No caso de encontrar ORF não anotada corretamente, relatar no notepad;
    - o Enviar relatório com vias incompletas.

Figura 3.24 – Página de busca do Ecocyc, que permite busca de genes e vias.

A fase 4 foi a anotação por categorias. A categoria II – Biossíntese de Pequenas Moléculas, foi coordenada por David Moon e Siu Mui Tsai. Os dois outros integrantes foram: **Eduardo Fernandes Formighieri** e Fabiana Cannavan.

Inicialmente, os ORFs atribuídos à categoria foram divididos entre os integrantes e a anotação foi verificada.

A segunda parte foi baseada nas vias metabólicas da bactéria *E. coli*. Foram conferidas todas as vias de biossíntese de pequenas moléculas deste organismo e buscados os elementos faltantes, como descrito na fase 3.

A seguir foram comparados todos os genes desta categoria das bactérias *Xylella fastidiosa* CVC, *Xanthomonas campestris* pv. *campestris* e *Xanthomonas axonopodis* pv. *citri* com os da *Xylella fastidiosa* PD (tabela resultante no anexo 3).

Antes da fase final, todos os procedimentos anteriores foram feitos para ORFs destinadas à categoria durante o processo de anotação.

No fim da anotação, foi escrito o relatório final da categoria (anexo 4), enviado para a equipe responsável pela redação do artigo a ser submetido. O rascunho deste artigo foi conferido. Posteriormente será gerado relatório de sugestão de atualizações de anotação nas bactérias anteriores do projeto Genoma (*X. fastidiosa* CVC, Xcamp e Xcitri).

3.5. Filogenia

3.5.1. CLUSTALW

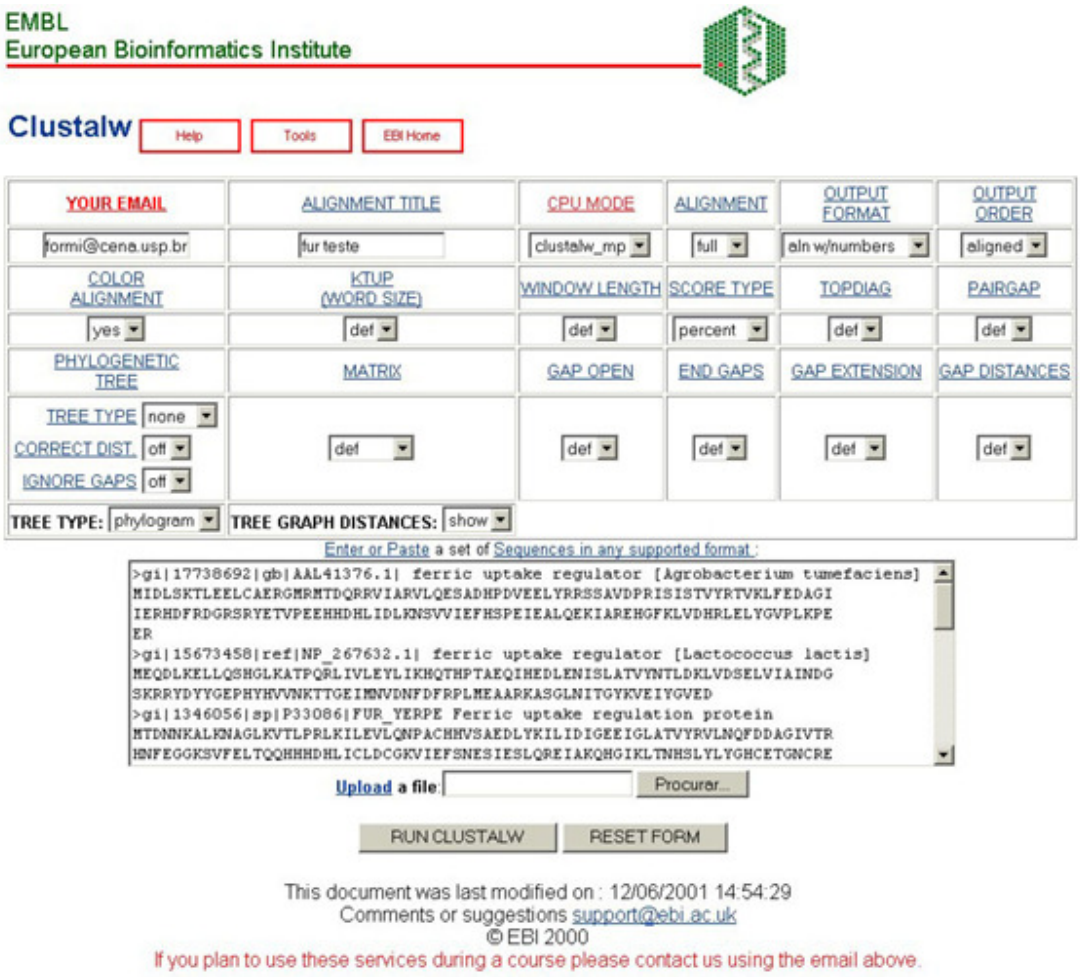


Figura 3.25 – Página de entrada de dados em ClustalW. Seqüências em formato fasta, e alinhamento colorido acionado.

Na figura 3.25 um exemplo de entrada de dados na página do ClustalW (Thompson et al., 1994) do *Europen Bioinformatics Institute*. São utilizadas seqüências em formato fasta. É inserido o e-mail (exigência do programa). A opção "*color alignment*" é acionada (*yes*) para facilitar a visualização do alinhamento múltiplo. Executa-se o programa (*Run CLUSTALW*).

Existem outros parâmetros que podem ser alterados conforme a necessidade. Serão gerados: um alinhamento colorido das seqüências inseridas e uma mapa filogenético das mesmas. A página de resultados é apresentada nas figuras 3.26, 3.27 e 3.28. Pode-se utilizar o Clustaw localmente (clustalx). Organizam-se as seqüências num arquivo em formato fasta, abre-se este arquivo e executa-se o alinhamento. O programa produz um arquivo com extensão "aln" onde se encontra o alinhamento gerado.

Your ClustalW Results:

Use JalView: JalView SUBMIT ANOTHER JOB

Pairwise Scores:

```
CLUSTAL W (1.81) Multiple Sequence Alignments


Sequence format is Pearson
Sequence 1: gi|17738692|gb|AAL41376.1|          142 aa
Sequence 2: gi|15673458|ref|NP_267632.1|        128 aa
Sequence 3: gi|1346056|sp|P33086|FUR_YERPE      148 aa
Sequence 4: gi|15980619|emb|CAC92876.1|         148 aa
Sequence 5: gi|15895892|ref|NP_349241.1|Fe      138 aa
Sequence 6: gi|15793093|ref|NP_282915.1|fe      144 aa
Sequence 7: gi|15791767|ref|NP_281590.1|        157 aa
Sequence 8: gi|15281308|dbj|BAB63408.1|         147 aa
Sequence 9: gi|120614|sp|P06975|FUR_ECOLI       148 aa
Start of Pairwise alignments
Aligning...
```

Figura 3.26 – Início da página de resultados do CLUSTALW.

```
CLUSTAL W (1.81) multiple sequence alignment

gi|1346056|sp|P33086|FUR_YERPE      -MT---------DNNKALKNA-GLKVTLPRLKILEVLQNPACHHVSAEDL 39
gi|15980619|emb|CAC92876.1|         -MT---------DNNKALKNA-GLKVTLPRLKILEVLQNPACHHVSAEDL 39
gi|120614|sp|P06975|FUR_ECOLI       -MT---------DNNTALKKA-GLKVTLPRLKILEVLQEPDNHHVSAEDL 39
gi|15281308|dbj|BAB63408.1|         -MS---------DNNQALKKA-GLKVTLPRLKILEVLQEPECQHISAEDL 39
gi|15793093|ref|NP_282915.1|fe      -MEK-------FNNIAQLKDS-GLKVTGPRLKILDLFETHAEEHLSAEDV 41
gi|17738692|gb|AAL41376.1|          MIDL-------SKTLEELCAERGMRMTDQRRVIARVLQES-ADHPDVEEL 42
gi|15791767|ref|NP_281590.1|        MLIENVEYDVLLERFKKILRQGGLKYTKQREVLLKTLYHS-DTHYTPESL 49
gi|15673458|ref|NP_267632.1|        ----------MEQDLKELLQSHGLKATPQRLIVLEYLIKH-QTHPTAEQI 39
gi|15895892|ref|NP_349241.1|Fe      ----------MNDISTIFKEKKLKLTPQRIAVYKYLKST-HEHPSAETI 38
                                              .    :     :: *  *  :   :       *   * :

gi|1346056|sp|P33086|FUR_YERPE      YKIL--IDIGEEIGLATVYRVLNQFDDAGIVTRHNFEGGKSVFELTQQHH 87
gi|15980619|emb|CAC92876.1|         YKIL--IDIGEEIGLATVYRVLNQFDDAGIVTRHNFEGGKSVFELTQQHH 87
gi|120614|sp|P06975|FUR_ECOLI       YKRL--IDMGEEIGLATVYRVLNQFDDAGIVTRHNFEGGKSVFELTQQHH 87
gi|15281308|dbj|BAB63408.1|         YKKL--IDIGEEIGLATVYRVLNQFDDAGIVTRHHFEGGKSVFELATQHH 87
gi|15793093|ref|NP_282915.1|fe      YRIL--LEEGVEIGVATIYRVLTQFEQAGILQRHHFETGKAVYELDKGDH 89
gi|17738692|gb|AAL41376.1|          YRRS--SAVDPRISISTVYRTVKLFEDAGIIERHDFRDGRSRYETVPEEH 90
gi|15791767|ref|NP_281590.1|        YMEIKQAEPDLNVGIATVYRTLNLLEEAEMVTSISFGSAGKKYELANKPH 99
gi|15673458|ref|NP_267632.1|        HEDL------ENISLATVYNTLDKLVDSELVIAIN-DGSKRRYDYYG-EP 81
gi|15895892|ref|NP_349241.1|Fe      YKAI--QSDYPTMSLATVYKALKTLAEVHLIQELNVGEGNFRYDANS-SS 85
                                    :           :.::*:*..:  : :  ::       .   ::

gi|1346056|sp|P33086|FUR_YERPE      HDHLICLDCGKVIEFSNESIESLQREIAKQHGIKLTNHSLYLYGHCETGN 137
gi|15980619|emb|CAC92876.1|         HDHLICLDCGKVIEFSNESIESLQREIAKQHGIKLTNHSLYLYGHCETGN 137
gi|120614|sp|P06975|FUR_ECOLI       HDHLICLDCGKVIEFSDDSIEARQREIAAKHGIRLTNHSLYLYGHCAEGD 137
gi|15281308|dbj|BAB63408.1|         HDHLVCLDCGKVIEFSDETIEQRQKQVAEKYNVRLTNHSLYLYGHCITGD 137
gi|15793093|ref|NP_282915.1|fe      HDHIVCVKCGEVTEFHNPEIEALQDKIAEENGYRIVDHALYMYGVCS--D 137
gi|17738692|gb|AAL41376.1|          HDHLIDLKNSVVIEFHSPEIEALQEKIAREHGFKLVDHRLELYGVPLKPE 140
gi|15791767|ref|NP_281590.1|        HDHMICKNCGKIIEFENPIIERQQALIAKEHGFKLTGHLMQLYGVCGDCN 149
gi|15673458|ref|NP_267632.1|        HYHVVNKTTGEIMNVDNFDFRPLMEAARKASGLNITGYKVEIYGVED--- 128
gi|15895892|ref|NP_349241.1|Fe      HPHIQCLSCGKVDDIMGITFDNLNKDVSSHTDYDVISNKLYFYGICKDCK 135
                                    * *:     . : :. .  :           .  : .  : :**

gi|1346056|sp|P33086|FUR_YERPE      CREDESAHSKR 148
gi|15980619|emb|CAC92876.1|         CREDESAHSKR 148
gi|120614|sp|P06975|FUR_ECOLI       CREDEHAHEGK 148
gi|15281308|dbj|BAB63408.1|         CADNDDAHNA- 147
gi|15793093|ref|NP_282915.1|fe      CQAKGKR---- 144
gi|17738692|gb|AAL41376.1|          ER---------- 142
gi|15791767|ref|NP_281590.1|        NQKAKVKI--- 157
gi|15673458|ref|NP_267632.1|        ----------
gi|15895892|ref|NP_349241.1|Fe      DKA-------- 138
```

Figura 3.27 – Alinhamento gerado pelo CLUSTALW

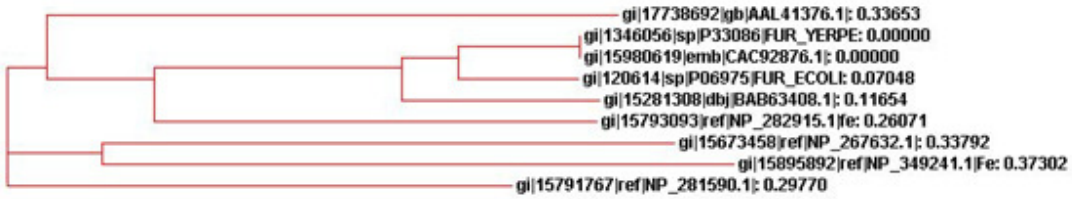


Figura 3.28 – Árvore filogenética gerada pelo CLUSTALW

3.5.2. PAUP

PAUP (Swofford, 2002) – programa de filogenia que cria a partir de arquivo com dados um arquivo de saída (*output.txt*) e alguns arquivos de árvores filogenéticas (extensão "tre"), visualizados pelo programa Treeview.

Forma de utilizar:

- ✓ Organizam-se as seqüências em formato fasta;
- ✓ Gera-se arquivo de alinhamento múltiplo utilizando o programa clustaw (conforme indicado acima – item 3.5.1);
- ✓ Cria-se arquivo *input.nex* (anexo 5, o nome pode ser alterado, mas deve ser um arquivo do tipo texto), segundo orientação do arquivo de ajuda do programa. Alguns dados importantes:
  - Tipo: DNA ou proteína (*datatype*);
  - Nomes das seqüências, na ordem do alinhamento (*tax labels*);
  - Matriz - alinhamento gerado no clustaw (*matrix*);
  - Número de seqüências (*dimensions n tax*);
  - Tamanho do alinhamento (*dimensions n char*);
  - Bootstrap 1000, ou pelo menos 500.
- ✓ Executa-se o programa PAUP no arquivo gerado (*input.nex*);
- ✓ Este gera diversos arquivos com os resultados do estudo filogenético:
  - Output.txt – resultados intermediários e finais;
  - Consenso.tre – resultado intermediário;
  - Trees.tre – filogenias geradas, resultado intermediário;
  - Treeconsenso.tre – filogenia final;
  - Bootstraping.tre – filogenia com valores de bootstraping na árvore definida em treeconsenso.

A importância da utilização do bootstrap está na maior confiabilidade na interpretação dos dados. Este executa inúmeras repetições da filogenia com diferentes ordenações, e com inclusão de seqüências aleatórias. O resultado é uma árvore com porcentagem de repetições nas divisões. Se uma divisão ocorreu em 100% das árvores formadas, o valor é 100. Se 34%, o valor é 34.

4. Resultados e Discussão

Este item apresenta resultados e discussões divididos em: (1) bioinformática – dados relacionados a sequenciamento, montagem e auxílio a pesquisadores; (2) anotação – apresentados em ordem cronológica de participação, nos organismos *Xylella fastidiosa* (CVC), *Xanthomonas axonopodis* pv. *citri* (e *X. campestris* pv. *campestris*); e *X. fastidiosa* (PD); e (3) Mecanismos associados ao ferro – com listagem de genes presentes e estudo de caso com filogenia.

4.1. Bioinformática

4.1.1. Projeto Xylella fastidiosa

Foram depositados 162 clones de shotgun e seqüenciados e montados dois cosmídeos (07B11 e 11A09).

4.1.2. Projeto *Xanthomonas axonopodis* pv. *citri*

Foram depositados 9247 clones *shotgun*, dos quais 8625 foram considerados válidos. O laboratório recebeu cinco cosmídeos para montar, além de duas regiões SGCs para finalizar. Os cosmídeos são selecionados para resolver regiões de difícil sequenciamento (compressão - C), que ocorrem principalmente por regiões ricas em GCs ou GTs, ou por estruturas secundárias. Podem também ser escolhidos para resolver regiões repetitivas (transposons - T), o que dificulta a montagem normal do trecho. A tabela 4.1 apresenta os dados finais destas montagens.

Cosmídeo	Contíguos de shotguns	Nº Placas	Nº Reads	Nº Oligos	Inserto
1HF02	490 T 510	7	721	14	39.465
1SC08	489 C 425 T 456 C 432	9	689	4	16.309
1HB01	516 C 442 T 455 T 467	14	453	19	36.534
1HC12	467 T 487	21	1688	3	41.351
0BB03	489 C 425 T 456 C 432	16	1232	4	40.151

Tabela 4.1 – Informações adicionais sobre cosmídeos montados.

A finalização dos SGCs (*shotgun contíguos*) foi realizada com o download das seqüências do trecho, a montagem do mesmo no programa phred/phrap/consed, e com a finalização das regiões problemáticas através do reseqüenciamento de clones e do sequenciamento utilizando oligonucleotídeos desenhados para as regiões mais difíceis. Em andamento a montagem de cinco clones para sub-bibliotecas.

4.1.3. Projeto *Xanthomonas campestris* pv. *campestris*

Foram depositados 5068 clones *shotgun*, sendo 4394 considerados válidos.

4.1.4. Projeto *Xylella fastidiosa* / Pierce's Disease

Foram depositados 5189 clones *shotgun*, sendo considerados válidos 4454. Foi montado um cosmídeo (07G02), com inserto de 42.205 bp.

4.1.5. Projeto *Leifsonia xyli* subsp. *xyli*

Depositados 8640 clones *shotgun*, sendo considerados válidos 7077. A montagem do cosmídeo 01C02 está em andamento, e o mesmo encontra-se com dois contíguos de 19.399 e 14.089 bp e um gap. Estão sendo seqüenciadas bibliotecas de clones das regiões com problemas (repetições e compressão). Já foram fechados três gaps entre os contíguos de shotgun.

4.1.6. Auxílio a Pesquisadores

Diversos pesquisadores do departamento ou de outros foram auxiliados através de discussão e definição de ferramentas a serem utilizadas, sobre a viabilidade de projetos, desenho de oligonucleotídeos, montagem de trechos de DNA seqüenciados, auxílio na utilização de ferramentas e nos processos de anotação.

Foi ministrada aula prática sobre "Bioinformática no Projeto Genoma", para os alunos da disciplina de pós-graduação Ecologia Experimental de Microrganismos, do CENA/USP. E também foi realizada palestra e aula prática sobre "Bioinformática e prospecção de genes em genomas bacterianos", para os alunos do curso de pós-graduação em Vigilância Sanitária de produtos, do Instituto Nacional de Controle de Qualidade em Saúde (INCQS, FIOCRUZ, RJ).

4.2. Anotação

São apresentados resultados de anotação nos quatro organismos. Inicialmente

4.2.1. *Xylella fastidiosa* CVC

Foi anotado preliminarmente o cosmídeo 11A05, que apresentou os seguintes genes: XF1575, XF1576, XF1577, XF1578, XF1579, XF1580, XF1582, XF1583, XF1584, XF1585, XF1589, XF1590, XF1591, XF1592, XF1593, XF1594, XF1595, XF1596, XF1597, XF1599, XF1600, XF1601, XF1602, XF1603, XF1604, XF1605, XF1606, XF1608, XF1609, XF1610, XF1611, XF1613, XF1614, XF1615, XF1617, XF1618, XF1620, XF1621, XF1622, XF1680, XF1681, XF1682, XF1683, XF1684, XF1685, XF1687, XF1688, XF1689, XF1690.

Nesta fase, não é significativo apresentar cada um destes genes, ou estatísticas sobre seu conteúdo, pois a anotação seqüencial objetiva uma anotação inicial dos genes, para verificar os genes que devem ser excluídos ou

inseridos, verificar os dados automáticos e limpar o genoma para facilitar a anotação posterior. Também serve para ajudar na montagem do genoma indicando cosmídeos contíguos.

Nesta época, existia uma grande demanda nas fases de sequenciamento e montagem, e a participação na anotação não se estendeu às fases seguintes, realizadas em sua maioria por anotadores mais experientes na época.

A necessidade de organização dos dados no programa Sequin, e a inexperiência da maioria dos anotadores e mesmo dos responsáveis pela estrutura de bioinformática tornaram o trabalho mais difícil. O organismo tomado como base foi a *E. coli*, que apresenta várias diferenças metabólicas. As anotações posteriores foram facilitadas pela presença do(s) genoma(s) seqüenciado(s) anteriormente.

4.2.2. *Xanthomonas*

Nesta bactéria, a participação na anotação se estendeu dos passos iniciais até a anotação por categorias, participando em duas distintas, incluindo o auxílio na redação dos relatórios da categoria para construção do artigo.

Na categoria Processos Celulares, o trabalho foi realizado em diversas fases por um número reduzido de pessoas (quatro). A comunicação foi realizada via internet e com reuniões esporádicas (integrantes pertencentes a diferentes cidades).

Na categoria Patogenicidade, Virulência e Adaptação, o trabalho foi dividido em grupos menores, sendo que três pesquisadores do CENA ficaram responsáveis por alguns subgrupos. Esta estrutura facilitou as discussões e o processo de anotação, pois as dúvidas são sanadas mais rapidamente e as conclusões das discussões mais freqüentes evitaram uma menor eficiência de trabalho, mais freqüente na categoria processos celulares, pela maior distância física.

4.2.2.1. Processos Celulares

Na primeira fase, foram anotados cerca de oitenta ORFs, entre os números 6000 e 7999.

Abaixo se encontra um resumo das informações obtidas na fase 2 da anotação desta categoria, divididas por famílias de transporte.

Família 2.C.1 – *The TonB-ExbB-ExbD / TolA-TolQ-TolR (TonB) Family of Auxiliary Proteins for Energization of Outer Membrane Receptor (OMR)-mediated Active Transport*

Esta família apresenta dois sistemas que energizam o transporte ativo mediado por OMR ("*outer membrane receptor*") através da pmf ("*proton motive force*"). Ambos sistemas parálogos estão presentes e completos. O primeiro contém os genes tonB, exbB, exbD1 e exbD2, respectivamente representados pelos ORFs 5419.1, 5417.1, 5414.1 e 5413.1. O gene exbD2 é provavelmente uma duplicação do exbD1. O segundo sistema contém os genes tolQ, tolR e tolA (ORFs 4507.1, 4509.1 e 4510.1, respectivamente), na mesma disposição encontrada em *Pseudomonas aeruginosa*. Existem vários outros ORFs homólogos ao gene tonB, mas fora do cluster e com hits mais fracos e/ou parciais.

Família 9.A.1 – *The Polysaccharide Transporter (PST) Family*

O sistema de secreção de EPS ("*exopolysaccharide*") está presente, através do gene gumJ (4276.1), e de dois genes correspondentes às suas proteínas auxiliares, gumB (4289.1) e gumC (4288.1). O gumB faz parte da família de transporte OMA (1.B.18 – *The Outer Membrane Auxiliary (OMA) Protein Family*), e o gumC da família MPA1-C (8.A.3 – *The Cytoplasmic Membrane-Periplasmic Auxiliary-1 (MPA-1) Protein with Cytoplasmic (C) Domain (MPA1-C or MPA1+C) Family*).

Família 1.B.14 – *The Outer Membrane Receptor (OMR) Family*

Foram encontrados vários ORFs desta família, principalmente *TonB-dependent receptors* e *ferric enterobactin receptors*.

Na categoria V.A.7 – *Cellular Processes/Transport/Other*, 44 ORFs:
66.1, 94.1, 124.1, 407.1, 630.1, 1361.1, 1567.1, 1715.1, 2113.1, 2119.1, 2146.1, 2290.1, 2437.1, 2851.1, 2874.1, 3077.1, 3162.1, 3367.1, 3446.1, 3448.1, 3671.1, 3893.1, 4192.1, 4203.1, 4314.1, 4403.1, 4480.1, 4551.1, 4621.1, 5074.1, 5079.1, 5100.1, 5118.1, 5303.1, 5547.1, 5603.1, 5697.1, 5701.1, 5739.1, 5806.1, 6107.1, 6219.1, 6314.1, 6861.1.

Na categoria V.A.4 – *Cellular Processes/Transport/Cations*, 7 ORFs:
1576.1; 4392.1; 4446.1; 4459.1; 4462.1; 4466.1; 5444.1.

Na fase três foram discutidos diferentes casos entre os integrantes do grupo, e entre estas trocas de informações o autor gerou a tabela encontrada no anexo 6 de genes relacionados ao TonB, das famílias de transporte 2.C.1 e 1.B.14. A maioria dos ORFs foi advinda da categoria VIII – *Hypothetical*, e a tabela foi sendo atualizada conforme a inclusão de novos ORFs na categoria.

4.2.2.2. Patogenicidade, Virulência e Adaptação

Subcategoria VII.A – Avirulência

Dos sete possíveis genes de avirulência encontrados no DNA cromossomal três foram confirmados na anotação final, quais sejam: avrBs2, avrXacA e avrXacB1. Nos plasmídeos (miniplasmídeo e megaplasmídeo) foram confirmados outros cinco (quatro pthAs e um avrXacB2). A tabela 4.2 apresenta mais detalhes destes genes. Apesar da menor homologia dos ORFs 106.2, 7474.1 e 10271.1, estes genes apresentaram semelhança suficiente para sua inclusão no grupo de genes potenciais da *X. axonopodis* pv. *citri*.

ORF	**Gene**	**Produto**	**Organismo Homólogo**	***E-value***	***Coverage Query* e *C. Subject***
35.1[1]	pthA	avirulence protein	*Xanthomonas a.* pv *citri*	0.0	103.3% / 100.0%
55.1[1]	pthA	avirulence protein	*Xanthomonas a.* pv *citri*	0.0	100.0% / 100.1%
100.2[1]	pthA	avirulence protein	*Xanthomonas a.* pv *citri*	0.0	100.0% / 100.1%
106.2[1]	avrXacB2	avirulence protein	*Pseudomonas s.* pv. *phaseolicola*	1e-10	96.3% / 96.3%
174.2[1]	pthA	avirulence protein	*Xanthomonas a.* pv *citri*	0.0	100.0% / 100.0%
5300.1[2]	avrBs2	avirulence protein	*Xanthomonas c.* pv *vesicatoria*	0.0	100.0% / 100.0%
7474.1[2]	avrXacB1	avirulence protein	*Pseudomonas s.* pv. *syringae*	2e-07	91.8% / 96.3%
10271.1[2]	avrXacA	avirulence protein	*Pseudomonas s.* pv. *phaseolicola*	2e-22	84.8% / 85.3%

Tabela 4.2 – Informações dos ORFs anotados na subcategoria VII.A.
1 – Encontrados no plasmídeo; 2 – Encontrados no cromossomo.

As proteínas dos genes citados nesta subcategoria têm ação no citoplasma bacteriano, de acordo com análise feita pelo programa PSORT. Ainda cabe ressaltar que estes genes não formam clusters, estando distribuídos distantes entre si, o que é normal devido à ação de avirulência ser dependente do produto de um único gene (genes monocistrônicos).

Subcategoria VII.B – Resposta de hipersensibilidade e patogenicidade

Foram encontrados vinte e seis ORFs no DNA cromossomal, detalhados na tabela 4.3. São hrps (*hypersensitive reaction and pathogenicity*), que elicitam reações de hipersensibilidade em plantas não hospedeiras e patogenicidade nas hospedeiras; hrcs (*hrp conserved*) e hpas (*hrp associated*). Os hrps e hrcs codificam componentes do Sistema de Secreção Tipo III, e os hpas codificam proteínas de função desconhecida, associada à patogenicidade.

ORF	**Gene**	**Produto**	**Organismo Homólogo**	***E-value***	***Coverage Query* e *C. Subject***
30000.1	hpa1	Hpa1 protein	*Xanthomonas o. oryzae*	4e-45	104.4% / 102.9%
7433.1	hpa2	Hpa2 protein	*Xanthomonas o. oryzae*	1e-72	98.6% / 84.0%
7406.1	hpaA	HpaA protein	*Xanthomonas c.* pv. *vesicatoria*	1e-118	101.5% / 100.0%
7399.1	hpaB	HpaB protein	*Xanthomonas c.* pv. *vesicatoria*	3e-83	100.0% / 96.3%
7410.1	hpaP	HpaP protein	*Xanthomonas c.* pv. *glycines*	1e-113	100.0% / 100.0%
7427.1	hrcC	HrcC protein	*Xanthomonas c.* pv. *vesicatoria*	0.0	100.0% / 100.0%
7416.1	hrcJ	HrcJ protein	*Xanthomonas c.* pv. *vesicatoria*	1e-141	100.0% / 100.0%
7420.1	hrcN	HrcN protein	*Xanthomonas o. oryzae*	0.0	100.0% / 100.0%
7409.1	hrcQ	HrcQ protein	*Xanthomonas c.* pv. *glycines*	1e-163	100.0% / 100.0%
7408.1	hrcR	HrcR protein	*Xanthomonas c.* pv. *vesicatoria*	1e-116	100.0% / 100.0%
7407.1	hrcS	HrcS protein	*Xanthomonas c.* pv. *glycines*	2e-40	100.0% / 100.0%
7423.1	hrcT	HrcT protein	*Xanthomonas o. oryzae*	1e-146	100.0% / 100.0%
7412.1	hrcU	HrcU protein	*Xanthomonas c.* pv. *glycines*	0.0	100.0% / 100.0%
7411.1	hrcV	HrcV protein	*Xanthomonas c.* pv. *glycines*	0.0	99.2% / 100.0%
7413.1	hrpB1	HrpB1 protein	*Xanthomonas o. oryzae*	8e-79	100.0% / 100.0%
7414.1	hrpB2	HrpB2 protein	*Xanthomonas o. oryzae*	8e-65	100.0% / 100.0%
7418.1	hrpB4	HrpB4 protein	*Xanthomonas c.* pv. *vesicatoria*	1e-111	100.0% / 100.0%

Tabela 4.3 – Informações dos ORFs anotados na subcategoria VII.B (continua).

ORF	Gene	Produto	Organismo Homólogo	*E-value*	*Coverage Query* e *C. Subject*
7419.1	hrpB5	HrpB5 protein	*Xanthomonas c.* pv. *vesicatoria*	1e-125	100.0% / 100.0%
7422.1	hrpB7	HrpB7 protein	*Xanthomonas o. oryzae*	8e-83	100.0% / 100.0%
7405.1	hrpD5	HrpD5 protein	*Xanthomonas c.* pv. *vesicatoria*	1e-162	100.0% / 100.0%
7404.1	hrpD6	HrpD6 protein	*Xanthomonas o. oryzae*	3e-31	98.8% / 98.8%
7402.1	hrpE	HrpE protein	*Xanthomonas o. oryzae*	7e-39	100.0% / 100.0%
7388.1	hrpF	HrpF protein	*Xanthomonas c.* pv. *vesicatoria*	0.0	98.3% / 100.5%
5823.1	hrpG	HrpG protein	*Xanthomonas c.* pv. *vesicatoria*	1e-147	100.0% / 100.0%
5819.1	hrpXct	HrpX protein	*Xanthomonas a.* pv. *Citri*	0.0	100.0% / 100.0%
6804.1	hrpX	HrpX protein	*Erwinia herbicola pv. gypsophilae*	2e-33	72.2% / 46.3%

Tabela 4.3 – Informações dos ORFs anotados na subcategoria VII.B. (continuação)

Há um *cluster* de 20.644 bp que inclui vinte e três destes ORFs (figura 4.1), e dois genes regulatórios (hrpG e hrpXct – figura 4.2) a 957.169 bp do mesmo. O outro ORF (que produz uma proteína relacionada ao HrpX) encontra-se mais distante do *cluster*, e apresenta menor similaridade. Este *cluster* é muito semelhante ao encontrado na *Xanthomonas axonopodis* pv. *vesicatoria*, tanto na orientação quanto no arranjo, sendo a única diferença entre os genes a ausência do hpa1 na *X.a.v.* Apenas o último ORF (6804.1 – hrpX) não apresentou alta homologia com genes descritos de *Xanthomonas*.

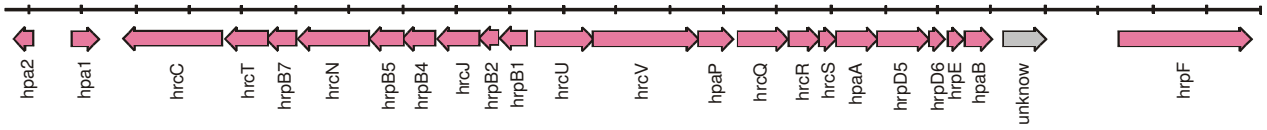


Figura 4.1 – Cluster do Sistema de Secreção Tipo III

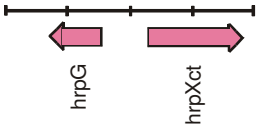


Figura 4.2 – Genes regulatórios

O local de ação dos produtos atribuído pelo PSORT está dividido entre o citoplasma bacteriano (10 casos) e a membrana interna da bactéria (15 casos), além de um caso no espaço periplasmático bacteriano.

Subcategoria VII.E - Exopolissacarídeos

Os exopolissacarídeos produzidos por bactérias patogênicas estão associados à habilidade em causar doença nas plantas. Na *Xanthomonas* é produzida a goma xanthana. Em *Xanthomonas campestris* pv. *campestris* este polímero é produzido pelo chamado *operon gum,* e todos os doze genes encontrados neste *operon* estão presentes na *X. a. citri*, com mesmo arranjo cromossomal e orientação. Estes ORFs estão descritos na tabela 4.4.

ORF	**Gene**	**Produto**	**Organismo Homólogo**	***E-value***	***Coverage Query* e *C. Subject***
4289.1	gumB	GumB protein	*Xanthomonas campestris*	1e-111	100.0% / 100.0%
4288.1	gumC	GumC protein	*Xanthomonas campestris*	0.0	99.8% / 100.0%
4286.1	gumD	GumD protein	*Xanthomonas campestris*	0.0	100.0% / 100.0%
4285.1	gumE	GumE protein	*Xanthomonas campestris*	0.0	100.0% / 100.2%
4284.1	gumF	GumF protein	*Xanthomonas campestris*	1e-178	99.7% / 99.5%
4282.1	gumG	GumG protein	*Xanthomonas campestris*	1e-130	100.0% / 100.0%
4279.1	gumH	GumH protein	*Xanthomonas campestris*	0.0	100.0% / 100.0%

Tabela 4.4 – Informações dos ORFs anotados na subcategoria VII.E (continua).

ORF	Gene	Produto	Organismo Homólogo	*E-value*	*Coverage Query* e *C. Subject*
4277.1	gumI	GumI protein	*Xanthomonas campestris*	1e-175	100.0% / 100.1%
4276.1	gumJ	GumJ protein	*Xanthomonas campestris*	0.0	100.0% / 101.0%
4275.1	gumK	GumK protein	*Xanthomonas o.* pv. *oryzae*	1e-167	100.0% / 100.0%
4274.1	gumL	GumL protein	*Xanthomonas o.* pv. *oryzae*	1e-145	100.0% / 100.0%
4272.1	gumM	GumM protein	*Xanthomonas campestris*	1e-135	99.6% / 99.6%
3111.1	xanA	phosphoglucomutase / phosphomannomutase	*Xanthomonas c.* pv. *campestris*	0.0	99.6% / 100.0%
3112.1	xanB	phosphomannose isomerase / GDP-mannose pyrophosphorylase	*Xanthomonas c.* pv. *campestris*	0.0	99.8% / 100.0%

Tabela 4.4 – Informações dos ORFs anotados na subcategoria VII.E (continuação).

O *operon* (figura 4.3) encontra-se entre as posições 3.033.000 e 3.048.000 do DNA cromossomal, abrangendo 12.560 bp. A maioria dos ORFs do *operon gum* tem homologia maior com *Xanthomonas campestris* pv. *campestris*, e apenas os ORFs gumK e gumL apresentaram maior homologia com *Xanthomonas oryzae* pv. *oryzae.*

Os dois ORFs que completam a categoria (xanA e xanB – figura 4) estão fora do *operon*, a 1.196.804 bp deste, e estão relacionados à biossíntese da goma xanthana.

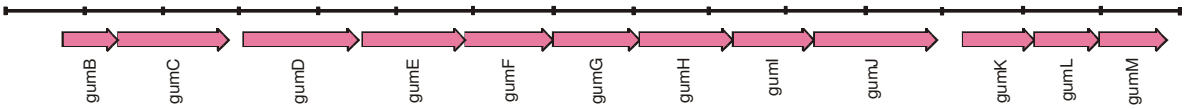


Figura 4.3 – Operon gum

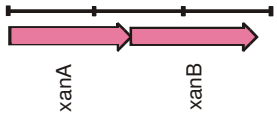


Figura 4.4 – Genes relacionados à biossíntese da goma xanthana.

O programa PSORT identificou a ação dos produtos destes genes na membrana interna da bactéria (10 casos), no citoplasma (três casos) e um caso no espaço periplasmático bacteriano.

4.2.3. *Xylella fastidiosa* PD

Nenhuma diferença significativa com a *Xylella* CVC foi identificada. A Xylella PD é capaz de sintetizar uma grande gama de pequenas moléculas necessárias à sua sobrevivência no xilema de vários hospedeiros.

Foram encontrados diversos genes bifuncionais, descritos no decorrer dos resultados.

No anexo 3 encontram-se as tabelas de comparação de genes presentes em ambos os isolados de *Xylella* e ambas as espécies de *Xanthomonas* estudadas. Também mapas de vias metabólicas tratadas neste item. Os mapas foram retirados dos sites do ECOCYC[6] (Karp et al., 2002) e do KEGG[7]. Foi utilizado o mapa considerado mais didático para a visualização da via biossintética. Passos diferentes dos descritos nos mapas foram descritos durante a discussão dos resultados.

As quatro bactérias apresentaram grande semelhança em praticamente todas as vias desta categoria, e preferiu-se apresentar os resultados para os quatro organismos de uma vez. Toda e qualquer diferença foi descrita e destacada na apresentação, e também pode ser examinada nas tabelas comparativas do anexo 3.

[6] http://ecocyc.pangeasystems.com/
[7] http://www.genome.ad.jp/kegg/kegg.html

Em diferentes vias metabólicas, uma ou mais enzimas presentes em *E. coli* está (estão) ausente(s), mas este fato também ocorre em diferentes bactérias gram-negativas, e aconteceu na maioria dos casos nas quatro bactérias comparadas. Infere-se que estas enzimas ausentes não sejam essenciais, ou tenham função substituída por novas estruturas.

II.A. Aminoácidos

A maioria dos genes encontrados em *E. coli* necessários à síntese de aminoácidos a partir de corismato, piruvato, 3-fosfoglicerato, glutamato e ácido oxaloacético foi identificada.

II.A.1 Família do Glutamato

Arginina

Na via da arginina, a partir do glutamato, apenas um gene não foi encontrado: argA (N-acetilglutamato sintase). Porém, o gene argB (N-cetilglutamato kinase) apresente homologia com o domínio deste gene faltante (argA), e infere-se que seja bifuncional. A ornitina carbamoiltransferase não está representada pelo gene argI, mas o argF executa a mesma função.

Genes presentes: argB, argC, argD, argE, argF, argG, argH, gltX, carA e carB.
Gene ausente: argA.
Tabela comparativa na p. 125, mapa na p. 142.

Glutamato e Glutamina

O glutamato pode ser obtido a partir de alfa-cetoglutarato através dos genes: gltB e gltD (glutamato sintase), tyrB (aspartato aminotransferase) e gdhA (NAD-glutamato desidrogenase). A glutamina pode ser obtida a partir do glutamato pelo gene glnA (glutamina sintetase).

Genes presentes: gdhA, gltD, gltB, tyrB (bifuncional), glnA e glnB.
Tabela comparativa na p. 125.

Prolina

Também a partir do glutamato, obtém-se a prolina, através dos genes proABC.

Genes presentes: proA, proB e proC.
Tabela comparativa na p. 125, mapa na p. 143.

II.A.2 Família do Aspartato / Piruvato
Alanina

A alanina é formada a partir do piruvato, por transaminação. A reação é catalisada pelo gene ilvE (alanina transferase) (Docena, 2000). A via descrita para *E. coli* apresenta também o gene avtA (valina piruvato aminotransferase), que não foi encontrado. A síntese de D-alanina ocorre com a mudança da L-alanina, pelo gene alr (alanina racemase).

Genes presentes: alr e ilvE. Gene ausente: avtA.
Tabela comparativa na p. 126.

Aspartato e Asparagina

O aspartato é sintetizado por transaminação do oxaloacetato (com glutamato como doador de aminogrupo). Está presente o gene envolvido: aspC (aspartato transaminase). Também pode ser obtido a partir do fumarato, através dos genes argH (EC# 4.3.2.1) e argG (EC# 6.3.4.5) ou dos genes purB (EC# 4.3.2.2) e purA (EC# 6.3.4.4). A asparagina é obtida a partir do aspartato por intermédio do gene asnB (asparagina sintase B).

Genes presentes: aspC, asnB, argG, argH, purA, purB e aspH.
Tabela comparativa na p. 126.

Isoleucina / Valina

Ambas as vias estão completas. A da isoleucina parte da treonina e utiliza os genes ilvAMGCD e E. A via da valina parte do piruvato e utiliza os genes ilvMGCDE. As *Xanthomonas* apresentam duas cópias do gene ilvA.

Genes presentes: ilvA, ilvC, ilvD, ilvE, ilvG, ilvM.
Tabela comparativa na p. 126, mapas nas p. 144 e 145.

Leucina

Via completa. O substrato inicial pode ser o piruvato ou o 2-oxoisovalerato (que exige a mais o gene leuA).

Genes presentes: leuA, leuB, icdA/leuB (bifuncional), leuC, leuD, tyrB (bifuncional) e ilvE.
Tabela comparativa na p. 127, mapa na p. 146.

Lisina

Não foi encontrado o gene dapC (N-sucinildiaminopimelato aminotransferase, EC# 2.6.1.17), nem uma via alternativa viável para completar a síntese da lisina a partir do aspartato. Existem outras aminotransferases com certa homologia com o gene ausente, como o gene aspC. Em *Xylella* PD, o resultado do blastp da comparação do gene dapC de *E. coli* apresenta maior homologia com os genes 13441 e 12911, sendo os valores E respectivamente 5e-30 e 1e-22.

Uma outra possibilidade seria a existência da diaminopimelato dehidrogenase (1.4.1.16), sem gene associado conhecido, que é uma via

alternativa apresentada pelo mapa 300 do KEGG (*Lysine Biosynthesis*[8]). Os genes de número EC mais próximo são os gltB e gltD (glutamato sintase, EC# 1.4.1.13).

Genes presentes: asd, dapA, dapB, dapD, dapE, dapE, dapF, lysA/lysC (bifuncional) e lysC/thrA (bifuncional).
Gene ausente: dapC.
Tabela comparativa na p. 127, mapa na p. 147.

Metionina

Inicialmente a ausência do gene metC parecia impedir a síntese de metionina, mas existe um caminho alternativo entre a cistationina e a homocisteína, através do gene metB, e tendo como produto intermediário a O-Acetil L-homoserina. A ausência do metH é compensada pela presença do metE, que apresenta caminho alternativo para a obtenção de metionina a partir de homocisteína. A ausência do gene metL é compensada pelo gene thrA, de mesmo número EC.

Genes presentes: metA, metB, metE, lysC/thrA (bifuncional), asd e masA.
Genes ausentes: metC, metH e metL.
Tabela comparativa na p. 127, mapa na p. 148.

Treonina

Via completa, a partir de aspartato e utilizando os genes thrA, thrB, thrC e asd.

Genes presentes: lysC/thrA (bifuncional), thrB, thrC e asd.
Tabela comparativa na p. 128, mapa na p. 149.

II.A.3 família Glicina / Serina

Cisteína

[8] http://www.genome.ad.jp/dbget-bin/get_pathway?org_name=map&mapno=00300

Embora no mapa de *E. coli* só seja descrita a via utilizando o gene ausente cysE, o mapa 260 do KEGG (*Glycine, Serine and Threonine Metabolism*[9]) apresenta duas possibilidades de obtenção da cisteína a partir da serina. No primeiro caso, diretamente, através do gene cistationa beta-sintase (cysM2, EC# 4.2.1.22). No segundo, o mesmo gene cysM2 leva à formação de cistationa, que através do gene metB (cistationa gama sintase, EC# 4.2.99.9) chega à cisteína.

Genes presentes: cysM2, cysB, cysK, metB e csdB. Gene ausente: cysE.
Tabela comparativa na p. 128.

Glicina / Serina

O gene serB é uma ausência importante no biossíntese de serina e glicina. Existem outras fosfatases, mas as comparações com blast não mostram nenhum alinhamento significativo. Embora o gene glyA, que permite a conversão entre serina e glicina esteja presente, não foi encontrada via completa de síntese de serina a partir de 3-fosfoglicerato.

Genes presentes: serA, serC, glyA, ilvA e ydfG. Gene ausente: serB.
Tabela comparativa na p. 128.

II.A.4 Família dos Aminoácidos Aromáticos
Corismato

Via completa, com gene aroQ substituindo o gene aroD, de mesmo número EC. São sete passos desde o eritrose-4-fosfato. No quinto passo, existe a possibilidade de ação de dois genes (aroK e L), mas apenas o aroK está presente.

Genes presentes: aroA, aroB, aroC, aroE, aroG, aroK e aroQ. Gene ausente: aroL.

[9] http://www.genome.ad.jp/dbget-bin/get_pathway?org_name=map&mapno=00260

Tabela comparativa na p. 129, mapa na p. 150.

Fenilalanina

Via completa, composta por três etapas a partir do corismato. Utiliza três genes que são bifuncionais. Foi encontrada nos quatro organismos um gene composto por uma parte do gene pheA. O gene corismato mutase não apresenta a segunda função do pheA (prepenato dehidratase), e ainda não existe um consenso quanto ao seu nome nos bancos de dados genéticos.

Genes presentes: pheA (bifuncional), tyrA (bifuncional) e tyrB (bifuncional).
Tabela comparativa na p. 129, mapa na p. 151.

Triptofano

Via completa, a partir do corismato. Os genomas apresentam mais de uma cópia dos genes trpD e trpE, sendo as consideradas funcionais descritas na tabela correspondente. O gene trpD, que é originalmente bifuncional, aparece em dois genes distintos, cada um com o domínio correspondente de uma função.

Genes presentes: trpA, trpB, trpC, trpD, trpE, trpF e wrbA.
Tabela comparativa na p. 129, mapa na p. 152.

Tirosina

Via completa e também descrita a partir do corismato. Não apresenta nenhuma particularidade.

Genes presentes: pheA (bifuncional), tyrA (bifuncional) e tyrB (bifuncional).
Tabela comparativa na p. 130, mapa na p. 153.

II.A.5 Histidina

A via de biossíntese de histidina está completa. A partir de alfa-5-fosforribosil-1-fosfato (PRPP) e adenosina tri-fosfato (ATP) são dez etapas até a histidina. O quinto passo é realizado por uma proteína composta por duas subunidades (produtos de hisF e hisH).

Genes presentes: hisA, hisB (bifuncional), hisC, hisD, hisF, hisG, hisH e hisI (bifuncional).
Tabela comparativa na p. 130, mapa na p. 154.

II.B Nucleotídeos

II.B.1 Ribonucleotídeos de Purina

Via completa, a partir de PRPP, até IMP (inosina 5'-monofosfato ou ácido inosínico), GMP (guanosina-5'-fosfato) e AMP (adenosina-5'-fosfato). O terceiro passo é possível através de dois genes: purN e purT. O purT (fosforibosilglicinamida formiltransferase 2) permite um caminho alternativo, mas está presente somente nas *Xanthomonas*. Em *Xylella*, foi encontrado apenas o fosforibosilglicinamida formiltransferase (purN). O sexto passo é realizado por um produto de dois genes: purE (subunidade catalítica) e purK (subunidade ATPase).

Os genes adk e ndk podem converter AMP em ATP, e o gene gmk e ndk, GMP em GTP.

Genes presentes: guaA (bifuncional), purA, purB, purC, purD, purE, purF, purH (bifuncional), purK, purL, purM ou G, purN, purU, ushA, adk, gmk, ndk e prsA. Gene ausente: purT.
Tabela comparativa na p. 131, mapa na p. 155.

II.B.2 Ribonucleotídeos de Pirimidina

Via completa, a partir de glutamina. A ausência do gene pyrI não impede o funcionamento do gene pyrB, que é a subunidade catalítica. O mapa mostra

a via até a uridina-5'-fosfato (UMP). A citidina-5'-fosfato (CMP) é obtida a partir deste após 5 passos. De UMP para UDP (cmk, EC# 2.7.4.14), deste para UTP (ndk, EC# 2.7.4.6), de UTP para CTP (pyrG, EC# 6.3.4.2), de CTP para CDP (ndk) e finalmente de CDP para CMP (cmk).

Genes presentes: carA, carB, pyrB, pyrC, pyrD, pyrE, pyrF, pyrG, ndk, cmk, dnaS, trxA, thyA e ushA. Gene ausente: pyrI.
Tabela comparativa na p. 132, mapa na p. 156.

II.B.3 2´-Deoxiribonucleotídeos

Via completa. O gene ausente (nrdD) tem função substituível pelo gene nrdB.

Genes presentes: dcd, mutT, mutT/thiE (bifuncional), nrdA, nrdB, glyA e tmk. Gene ausente: nrdD.
Tabela comparativa na p. 132.

II.B.4 Salvamento de Nucleosídeos e Nucleotídeos

Apresenta alguns genes associados ao salvamento de purinas e pirimidinas, mas as vias não estão completas. Nenhum dos organismos parece ter a capacidade de salvamento pelas vias conhecidas atualmente. As *Xanthomonas* apresentam o gene add (adenosina deaminase, EC# 3.5.4.4), ausente em *Xylella*.

Genes presentes: apaH, deoD, hpt, tdk, nadD e enpP. Gene ausente: add.
Tabela comparativa na p. 132.

II.C Açúcares e Nucleotídeos de Açúcares

Classificação adotada em *E. coli* e não utilizada neste trabalho. Citada por ter sido mantida na categorização dos projetos.

II.D Cofatores, Grupos Prostéticos e Transportadores

II.D.1 Biotina

Via completa a partir de 6-Carboxihexanoil-CoA (ou Pimeloil-CoA). No mapa 780 do KEGG (Biotin Metabolism[10]) existe a possibilidade de síntese a partir do pimelato, mas o gene responsável (bioW, EC# 6.2.1.14), com poucos exemplos no próprio KEGG[11], não foi encontrado. As *Xanthomonas* apresentam duas cópias de bioA e bioC.

Genes presentes: bioA, bioB, bioC, bioD, bioF, bioH, bioI e birA. Gene ausente: bioW.
Tabela comparativa na p. 133, mapa na p. 157.

II.D.2 Ácido Fólico

Esta via está aparentemente incompleta, pela ausência dos genes pabABC. Estes três genes participam da via a partir de corismato para a obtenção de ácido p-aminobenzóico (PABA). Embora o mapa 790 do KEGG (*Folate biosynthesis*[12]) cite apenas o EC# 4.1.3.- (pabAB) para a conversão, e estes genes tenham certa homologia com genes trpD e trpE encontrados, não se pode afirmar que os genes semelhantes executem a mesma função. O restante da via biossintética está completo.

Genes presentes: folB, folC, folD, folE, folK, folP, thyA, metF, mutT, mutt/thiE (bifuncional), glyA e lpxC. Genes ausentes: pabA, pabB e pabC.
Tabela comparativa na p. 134, mapa na p. 158.

II.D.3 Lipoato

[10] http://www.genome.ad.jp/kegg/pathway/map/map00780.html
[11] http://www.genome.ad.jp/dbget-bin/www_bget?enzyme+6.2.1.14
[12] http://www.genome.ad.jp/kegg/pathway/map/map00790.html

Via completa, de apenas dois passos a partir do ácido octanóico. Esta síntese ainda não foi bem esclarecida.

Genes presentes: lipA e lipB.
Tabela comparativa na p. 134.

II.D.4 Molibdopterina

As *Xylella* apresentam apenas um gene desta via (moeB). Os genes não possuem números EC. As *Xanthomonas* apresentam os mesmos genes que *E. coli*, que possui a capacidade de síntese da molibdopterina, e deve apresentar a mesma capacidade. Apresenta ainda duas cópias do gene moeA (molibdopterina guanina dinucleotídeo sintase).

Genes presentes em *Xylella*: moeB.
Genes presentes em *Xanthomonas*: moaA, moaB, moaC, moaD, mobA, moeA e moeB.
Tabela comparativa na p. 135.

II.D.5 Pantotenato

Via incompleta. O pantotenato é obtido a partir da condensação do pantoato com beta-alanina - pantoato beta-alanina ligase (panC, EC# 6.3.2.1). A beta-alanina pode ser obtida a partir do aspartato com o gene panD (aspartato descarboxilase, EC# 4.1.1.11), mas o pantoato não, pela ausência do gene panE (2-dehidropantoato redutase, EC# 1.1.1.169). Não foi identificada uma via alternativa de obtenção do pantoato.

A ausência do gene coaA (pantotenato kinase, EC# 2.7.1.33) e a baixa similaridade do gene dfp encontrado (Bifuncional: EC# 6.3.2.5 e 4.1.1.36) indica que o pantotenato também não pode ser obtido a partir da coenzima A, e que esta não pode ser obtida a partir do pantotenato.

Genes presentes: panB, panC, panD, dfp (bifuncional), coaD e coaE. Genes ausentes: coaA e panE.
Tabela comparativa na p. 135, mapa na p. 159.

II.D.6 Piridoxina

Via aparentemente incompleta. A ausência dos genes pdxB (eritronato-4-fosfato dehidrogenase, EC# 1.1.1.-) e pdxK (piridoxamina kinase, trifuncional) impede a síntese de piridoxina pelos meios conhecidos. Mesmo com a presença em *Xanthomonas* do gene pdxY (piridoxal kinase 2, EC# 2.7.1.35), que executa uma das funções do gene pdxK, a via encontra-se incompleta pela falta do gene essencial pdxK.

Existe a possibilidade do gene ptsK (HPr kinase/fosfatase, EC# 3.1.3.-) executar a função faltante do gene pdxK, de piridoxina fosfato para piridoxina, mas não se pôde confirmar esta hipótese apenas com as comparações das seqüências encontradas.

Genes presentes em *Xylella*: pdxA, pdxH, pdxJ e dxr.
Genes presentes em *Xanthomonas*: pdxA, pdxH, pdxJ, pdxY e dxr.
Genes ausentes: pdxB e pdxK (bifuncional).
Tabela comparativa na p. 135.

II.D.7 Nucleotídeos de Piridina

Via completa, a partir do aspartato, nos isolados de *Xylella*. As *Xanthomonas* não têm os genes nadAB e pncA, apresentando apenas parte da via, a partir quinolinato.

Genes presentes em *Xylella*: nadA, nadB, nadC, nadD, nadE, pncA e pncB.
Genes presentes em *Xanthomonas*: nadC, nadD, nadE e pncB.
Tabela comparativa na p. 136, mapa na p. 160.

II.D.8 Tiamina

Provavelmente a síntese seja possível, embora a via esteja incompleta. O gene thiG executa a função do thiH. Os genes thiM e thiK não têm substitutos, mas fazem parte de passos não essenciais à síntese. Caso o gene pstK, que apresenta o EC# 3.1.3.-, possa executar o passo de tiamina fosfato para tiamina, a síntese será possível.

Genes presentes: thiC, thiD (bifuncional), thiE, mutT/thiE (bifuncional), thiG, thiL, apbE e dxs.
Genes ausentes: thiH, thiK e thiM.
Tabela comparativa na p. 136, mapa na p. 161.

II.D.9 Riboflavina

Via completa, a partir de GTP.

Genes presentes: ribA, ribA/ribB (bifuncional), ribD ou ribG (bifuncional), ribE, ribH e ribF (bifuncional).
Tabela comparativa na p. 136, mapa na p. 161.

II.D.10 Tioredoxina, Glutarredoxina e Glutationa

Os genes necessários para a síntese dos três compostos estão presentes.

Genes presentes: ggt, grxC, gshA, gshB, yneN, gst, trxA, trxB e trx.
Tabela comparativa na p. 137.

II.D.11 Menaquinona e Ubiquinona

Só está presente um gene que participa da biossíntese da menaquinona, a ubiE (ubiquinona/menaquinona transferase, EC# 2.1.1.-), sendo esta via inexistente.

A via da biossíntese da ubiquinona está presente, mas incompleta. Dentre os passos iniciais, estão faltando o primeiro e o terceiro, sendo responsáveis os genes ubiC (corismato piruvato liase, EC# 4.-.-.-), e ubiX e ubiD (3-octaprenil-4-hidroxibenzoato descarboxilase 1 e 2, ambos com EC# 4.1.1.-. A partir deste passo, todos os genes estão presentes.

Genes presentes: ispA, ispB, ubiA, ubiB, ubiE, ubiG, ubiH, dxs, coq7 e aroQ. Gene ausente: ubiC, ubiD e ubiX.
Tabela comparativa na p. 137, mapa da ubiquinona na p. 162.

II.D.12 Proto e Siroheme

Via completa. O caminho de glutamato a siroheme está completo, mas a ausência do hemG (protoporfirinogen oxidase, EC# 1.3.3.4) impediria que fosse obtido o protoheme pela via conhecida. No entanto, o produto do gene hemK executa a função do produto do hemG.

A *Xanthomonas citri* apresenta duas cópias a mais do gene hemL.

Genes presentes: cysG (bifuncional), hemA, hemB, hemC, hemD, hemE, hemF, hemH, hemK, hemL, hemN, hemY, gltX e cyoE. Gene ausente: hemG.
Tabela comparativa na p. 138, mapa na p. 163.

II.D.13 Proteína Transportadora de Carboxil Biotina (BCCP)

Presente em todos os organismos estudados.

Genes presentes: accB.
Tabela comparativa na p. 138.

II.D.14 Cobalamina

Os isolados de *Xylella* só possuem um gene relacionado a esta via: pgmA (fosfoglicerato mutase, EC# 5.4.2.1).

A Xcitri apresenta a via completa, e na Xcamp falta apenas o gene cobC, e esta tem a mais o gene cobB, de função na via ainda não esclarecida. Infere-se que a Xcitri a via esteja funcional, e não se pode afirmar o mesmo da Xcamp.

Gene presente em *Xylella*: pgmA.
Genes presentes em *Xanthomonas*: btuR, cobB, cobC, cobS, cobT, cobU e pgmA. CobB só na *X. campestris*, e cobB só na *X. citri*.
Tabela comparativa na p. 138, mapa na p. 164.

II.D.15 Enterobactina

Via incompleta. Presença apenas do gene entF.

Gene presente em *Xanthomonas*: entF.
Tabela comparativa na p. 139.

II.D.16 Biopterina

Existem alguns genes relacionados a esta via, mas não são suficientes a biossíntese.

Genes presentes: ptr1, ispD, ispE, ispF, gcpE e ygcM.
Tabela comparativa na p. 139.

II.E Biossíntese de Ácidos Graxos e Fosfatídicos

Esta biossíntese foi dividida em três fases para uma melhor discussão, segundo gráficos do ECOCYC[13]. A primeira diz respeito aos passos iniciais, e está completa. Inclui os genes fabB, fabH, accA, accB, accD e fabD.

[13] http://ecocyc.pangeasystems.com/

A segunda se refere à elongação de ácidos graxos saturados, e está incompleta. Nesta via falta o gene enoil-ACP redutase (NADH - fabI, EC# 1.3.1.9; ou NADPH - EC# 1.3.1.10, sem nome de gene). Apresenta os genes: fabA, fabB e fabG.

A terceira é a elongação de insaturados, e está completa, com os genes fabB, fabG e fabA.
Genes presentes em *Xylella*: accA, accC, accD, accP, acs, cdsA, dgkA, fabA, fabB, fabd, fabG, fabH, fabZ, drb0080, tesB, accB, cls e psd.
Genes presentes em *Xanthomonas*: os citados acima mais tesA. A *X. citri* tem um acpD.
Genes ausentes: acpE, cdh e fabI.
Tabela comparativa na p. 140, mapas nas p. 164 e 165.

II.F Poliaminas

Via incompleta. Ausência importante dos genes agmatinase (speB, EC# 3.5.3.11) e ornitina descarboxilase biossintética (speC, EC# 4.1.1.17).

Genes presentes: speA, speD e speE.
Genes ausentes: speB, spec, speF e speG.
Tabela comparativa na p. 141, mapa na p. 166.

4.3. Mecanismos Associados ao Ferro

Os genes envolvidos no metabolismo de ferro podem ser divididos de acordo com sua função no processo global. Genes que são sintetizados dentro da célula e excretados no meio para capturar ferro e trazê-lo de volta em uma forma não disponível aos outros organismos são chamados sideróforos. Durante o processo de assimilação, necessitam-se receptores que reconhecem especificamente estas moléculas acopladas ao ferro para iniciar o processo de internalização, visando à extração de ferro do sideróforo para participação no metabolismo celular. Uma vez dentro da célula, algumas proteínas utilizam

diretamente o ferro em reações químicas e outras monitoram os níveis do ferro dentro da célula.

4.3.1. Genes Presentes

Inúmeros genes associados ao metabolismo do ferro foram encontrados em *Xanthomonas*, incluindo vários receptores de sideróforos. Porém os genes necessários para a síntese de sideróforos não ribossômicos não foram localizados. O sistema energizador TonB está presente e completo, permitindo a energização da passagem de produtos pela membrana celular.

Dos genes regulados pelo Fur (*ferric uptake regulator*) encontrados em *Pseudomonas aeruginosa* por VASIL & OCHSNER, 1999, estão presentes os ORFs correspondentes: fpvA (7667), pfeA (25309), phuR (1712, 2743), feoAB (2230, 2231), fumC (4651), nuoA (3826), tonB (45789, 5419), tolQRA (4507, 4509, 4510) e piuB (2830). O gene fur (4702) também está presente.

Ambas as espécies de *Xanthomonas* apresentam muitas cópias de diversos receptores, como: cirA (8 em Xcamp e 9 em Xcitri), fecA (9 em Xcamp e 8 em Xcitri) e fhuA (7 em cada uma). Porém, destas cópias, apenas algumas possuem homologia boa com o domínio TonB boxC. Nenhum cirA tem homologia (utilizando PFAM) menor que 0,017, sendo considerado aceitável apenas abaixo de e-5 (10^{-5}). Dos ORFs fecA, 4 têm baixa homologia, e dos fhuA, 7 apresentam hits abaixo de 10^{-5}. De modo geral, são poucas as cópias de receptores relacionados ao ferro com boa homologia em relação aos domínios funcionais dos genes originais.

As *Xanthomonas* também possuem bacterioferritinas com boa homologia (5710, 683, 11450, 4849), uma proteína comigratória com bacterioferritina (3465) e uma bacterioferritina associado a ferredoxina (15018).

Basicamente, Xcitri e Xcamp têm dois receptores para sideróforos (pioverdina e enterobactina), e são competentes para a aquisição de Fe(II). Possuem os dois sistemas parálogos (TonB-ExbBD e TolAQR) de energização do

transporte ativo através da membrana externa. Existem diversos receptores dependentes deste sistema, muitos deles relacionados ao transporte de ferro, como: fepA (66 e 4551), fecA (630, 3367, 4446, 5074, 5079, 5444, 5701, 6219 e 20068), fhuA (1715, 2146, 3448, 4480, 4855, 6107, 15194, 25860 e 20006), iroN (124, 2851, 3671, 5547, 25839 e 20761), bfeA (4392, 4459, 4462, 4466 e 20019), fpvA, pfeA, cirA (94, 407, 1361, 1567, 2437, 3446, 5303, 6861, 25327, 25837 e 25853) e fhuE (1576 e 25820).

Os receptores citados realizam o transporte através da membrana externa, para dentro da célula. Eles apresentam especificidade para determinados produtos. O fepA tem como produto um receptor de Ferri-enterobactina (sideróforo), e também pode atuar em colicinas B e D. O bfeA e iroN também estão relacionados à enterobactina. O pfeA é um receptor específico para a enterobactina Fe (III). O produto do fhuA é receptor de Fe (III) hidroxamato, de colicina M e dos fagos T1, T5 e phi80. O produto de fhuE é requerido para a captura de Fe (III). O fecA é receptor de Fe (III) dicitrato. Os ORFs feoA e feoA estão relacionados ao transporte de Fe (II). FyuA está relacionado ao Fe (III). O gene fpvA codifica um receptor de ferripioverdina, um sideróforo. O produto do gene cirA é um receptor de colicina I e de Fe (III) catecholate (informações obtidas no site do Swiss-Prot, Bairoch & Apweiler, 2000).

Embora não tenham sido encontrados genes codificadores de sideróforos não ribossomais, a hipótese de que exista a produção de sideróforos de outros tipos ainda não está totalmente descartada. Além disso, é possível a utilização de sideróforos externos, o que justifica a presença da variada gama de receptores, mesmo na ausência dos transportadores correspondentes.

Os isolados de *Xylella* apresentaram um número bem menor de receptores: bfeA (11011 e 11031), yncD (6531), fhuA (14071), cirA (18911) e outros dois receptores dependentes de TonB. Destes, apenas dois (yncD e fhuA) apresentaram uma homologia boa com o domínio do TonB boxC. Apresentam ainda bacterioferritinas (15251, 24891 – este ORF não foi reconhecida em XCVC, mas a seqüência está presente), os genes feoAB (16091

e 16081), ferredoxinas (2471, 6291, 7731, 14431 e 18171) e uma proteína comigratória com bacterioferritina (bcp, 15891).

4.3.2. Filogenia

Foram realizados dois estudos de caso, com um gene bem conservado (fur) e com genes menos conservados (receptores da membrana externa, dependentes do sistema TonB).

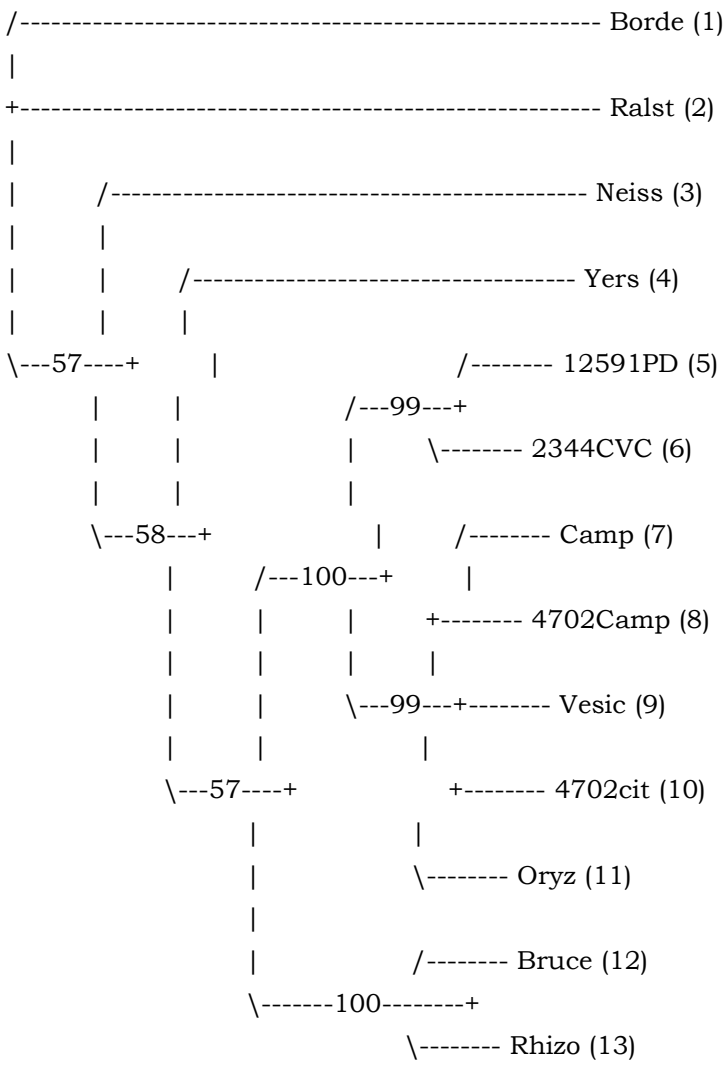


Figura 4.5 – Árvore filogenética do gene fur, utilizando o programa PAUP, usando parcimônia máxima. Valores de bootstrap (1000 réplicas). Retirado do arquivo "*output.txt*".

A figura 4.5 traz uma árvore filogenética de genes fur dos organismos seqüenciados, com genes fur encontrados de outras *Xanthomonas* e de alguns organismos mais distantes em termos de taxonomia para comparação (Wheeler et al., 2000). O gene apresenta alta homologia, mesmo em organismos mais distantes, indicando alto grau de conservação, o que pode ser confirmado pela árvore quase idêntica à classificação taxonômica (NCBI Taxonomy - Wheeler, 2000). A única diferença nas divisões desta é a colocação do gene de *Yersinia pestis*, que taxonomicamente deveria estar mais próximo do grupo *Xanthomonas* por fazer parte da subdivisão Gama. Na árvore apresentada, existe uma divisão entre o grupo *Xanthomonas* e a subdivisão Alfa, e depois a ligação com *Y. pestis*.

São pertencentes à subdivisão Alfa:

- ✓ Bruce (12) – *Brucella melitensis* biovar *abortus*, gi|3913692|sp|O30976; e
- ✓ Rhizo (13) – *Rhizobium leguminosarum* bv. *viciae*,gi|3913688|sp|O07315.

São pertencentes à subdivisão Beta:

- ✓ Borde (1) – *Bordetella pertussis*, gi|2120993|pir||I40326;
- ✓ Ralst (2) - *Ralstonia eutropha*, gi|3913691|sp|O30330;
- ✓ Neiss (3) – *Neisseria meningitides* Z2491, gi|15793093|ref|NP_282915.1.

São pertencentes à subdivisão Gama:

- ✓ 12591PD (5) – XPD, ORF 12591;
- ✓ 2344CVC (6) – XCVC, ORF 2344;
- ✓ Camp (7) – *Xanthomonas campestris* pv. *campestris*, gi|5532516|gb|AAD44765.1;
- ✓ 4702Camp (8) – Xcamp, ORF 4702;

- ✓ Vesic (9) – *Xanthomonas campestris* pv. *vesicatoria*, gi|5532520|gb|AAD44767.1;
- ✓ 4702cit (10) – Xcitri, ORF 4702;
- ✓ Oryz (11) – *Xanthomonas oryzae*, gi|5532604|gb|AAD44807.1; e
- ✓ Yers (4) – *Yersinia pestis*, gi|16122845|ref|NP_406158.1.

Os valores de bootstrap mostram que as divisões do grupo *Xanthomonas* e da subdivisão alfa são muito consistentes, ocorrendo em praticamente 100% dos casos. Da mesma forma, a divisão entre *Xylella* e *Xanthomonas*, e as divisões subseqüentes. As divisões anteriores apresentaram valores considerados baixos, mas sendo o objetivo do estudo a comparação dentro do grupo *Xanthomonas*, isto não afeta os resultados.

Por outro lado, os genes receptores dependentes do sistema TonB encontrados demonstraram grande variabilidade, a ponto de formarem grupamentos separados dos genes originais. O estudo filogenético auxiliou na detecção de genes mais próximos ou distantes dos originais, e principalmente permitiu uma visualização rápida de genes anotados de forma errada, na fase de escolha dos genes apropriados para o estudo.

A figura 4.6 apresenta a comparação filogenética, utilizando o PAUP, dos receptores relacionados com ferro de anotação mais consistente, em relação aos dados de comparação do blastp. Mesmo assim, alguns genes estão normalmente distantes dos originais, que formam em alguns casos braços separados, mas pertencendo a grupamentos maiores. Estes dados sugerem que tenha ocorrido transferência horizontal, já que além de os organismos não produzirem os sideróforos correspondentes aos receptores presentes, estes se encontram com várias cópias em diferentes graus de homologia. Existem ORFs classificados muito próximos dos genes originais, demonstrando alta probabilidade de serem verdadeiros, como o caso de: 25860 Xcitri fhuA, 25820 Xcitri fhuE e 4446 Xcitri e Xcamp fecA. Os ORFS de número maior que 25000 representam ORFs que não apresentam correspondente em Xcamp. Os da faixa de 20000 são presentes apenas em Xcamp.

Os ORFs bfeA, apesar de não apresentarem alta homologia com o original, formaram um grupo bem destacado na árvore, incluindo os genes de XCVC e XPD. O mesmo acontece para os ORFs de cirA e iroN. Os ORFs fecA 630 e 5074 de Xcitri e Xcamp, apesar de ter alta homologia com o fecA de *E. coli*, estão separados na árvore filogenética, provavelmente indicando uma subfamília deste tipo de receptor.

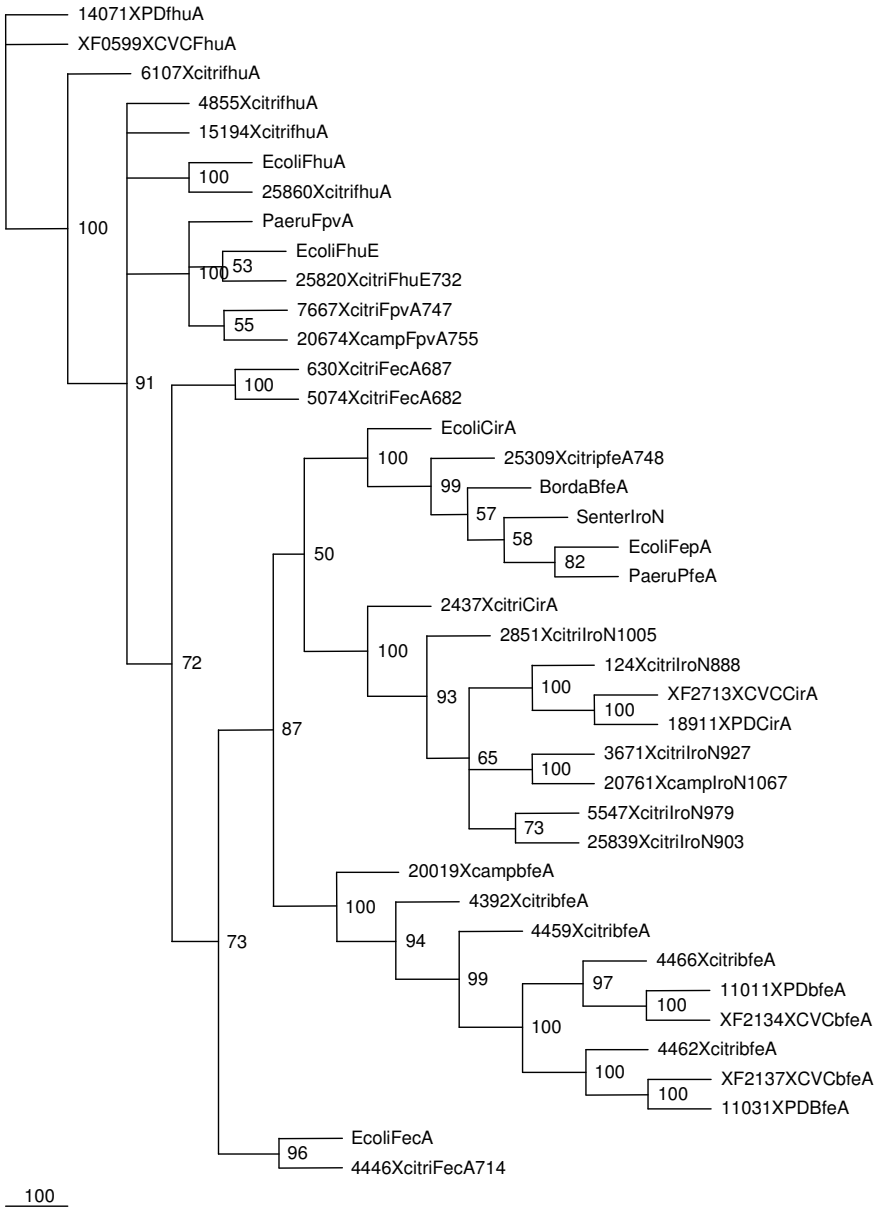


Figura 4.6 – Árvore filogenética, com valores de bootstraping, para alguns receptores de ferro dependentes do sistema tonB, gerada pelo PAUP (*bootstraping.tree*).

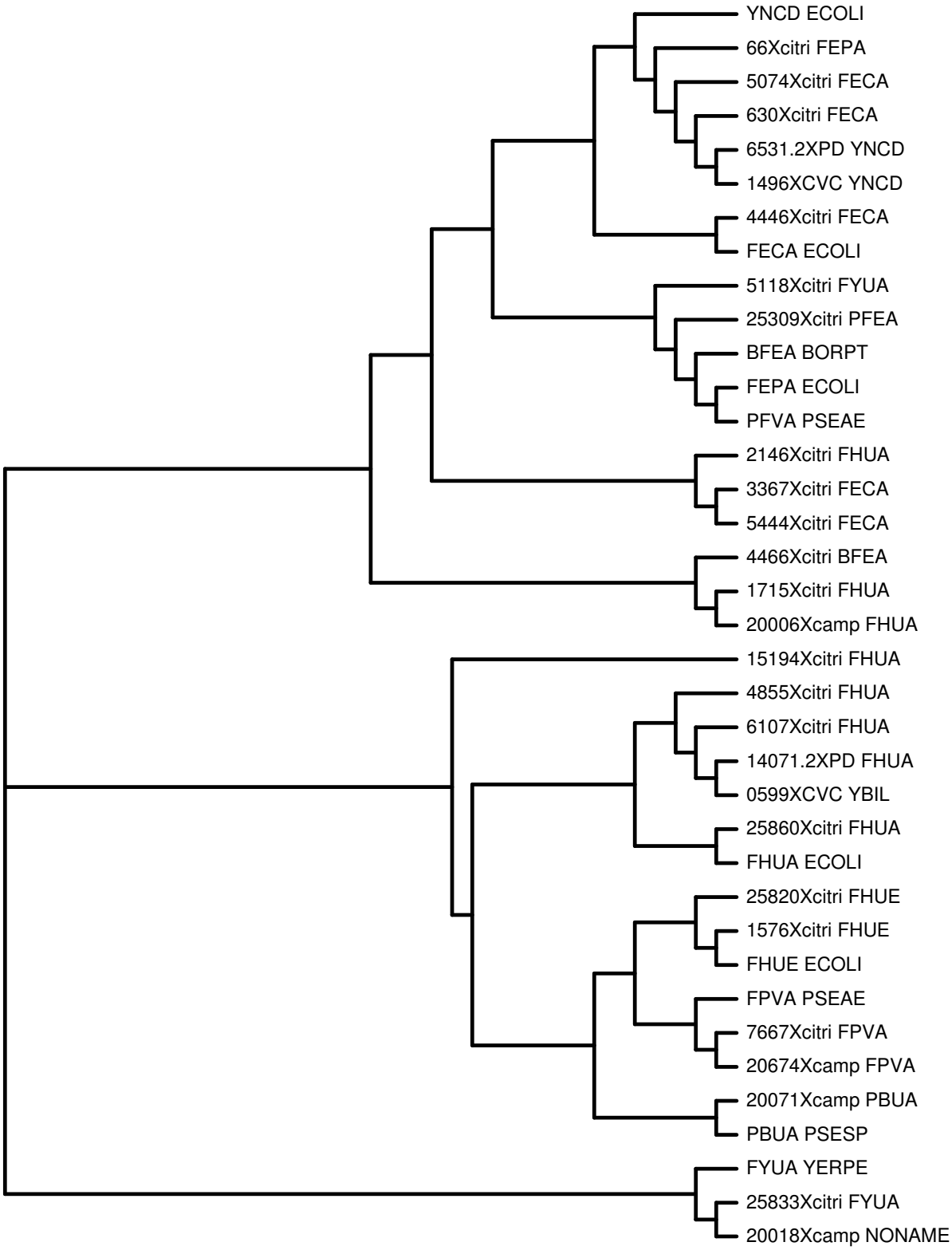


Figura 4.7 – Árvore filogenética de receptores da membrana externa relacionados ao Ferro, selecionados pela homologia com TonB boxC, gerada pelo ClustalW.

Os genes presentes na figura 4.6 são:

- ✓ 14071XPDfhuA – XPD, ORF 14071;
- ✓ XF0599XCVCFhuA – XCVC, ORF 599;
- ✓ 6107XcitriFhuA – Xcitri e Xcamp, ORF 6107;
- ✓ 4855XcitriFhuA – Xcitri e Xcamp, ORF 4855;
- ✓ 15194XcitriFhuA – Xcitri e Xcamp, ORF 15194;
- ✓ EcoliFhuA – *E. coli*, gi|2507464|sp|P06971;
- ✓ 25860XcitriFhua – Xcitri, ORF 25860, exclusivo;
- ✓ PaeruFpva – *Pseudomonas aeruginosa*, gi|12230910|sp|P48632;
- ✓ EcoliFhuE – *E. coli*, gi|2507465|sp|P16869;
- ✓ 25820XcitriFhuE – Xcitri, ORF 25820, exclusivo;
- ✓ 7667XcitriFpvA – Xcitri e Xcamp, ORF 7667;
- ✓ 20674XcampFpvA – Xcamp, ORF 20674, exclusivo;
- ✓ 630XcitriFecA – Xcitri e Xcamp, ORF 630;
- ✓ 5074XcitriFecA – Xcitri e Xcamp, ORF 5074;
- ✓ EcoliCirA – *E. coli*, gi|2507462|sp|P17315;
- ✓ 253089XcitriPfeA – Xcitri, ORF 25309, exclusivo;
- ✓ BordaBfeA – *Bordetella pertussis*, gi|538279|gb|AAA98536.1;
- ✓ SenterIroN – *Salmonella enterica*, gi|2738252|gb|AAC46183.1;
- ✓ EcoliFepA – *E.coli*, gi|2507463|sp|P05825;
- ✓ PaeruPfeA – *Pseudomonas aeruginosa*, gi|548479|sp|Q05098;
- ✓ 2437XcitriCirA – Xcitri e Xcamp, ORF 2437;
- ✓ 2851XcitriIroN – Xcitri e Xcamp, ORF 2851;
- ✓ 124XcitriIroN – Xcitri e Xcamp, ORF 124;
- ✓ XF2713XCVCCirA – XCVC, ORF 2713;
- ✓ 18911XPDCirA – Xcitri e Xcamp, ORF 18911;
- ✓ 3671XcitriIroN – Xcitri e Xcamp, ORF 3671;
- ✓ 20761XcampIroN – Xcamp, ORF 20761, exclusivo;
- ✓ 5547XcitriIroN – Xcitri e Xcamp, ORF 5547;
- ✓ 25839XcitriIriN – Xcitri, ORF 25839, exclusivo;
- ✓ 20019XcampBfeA – Xcamp, ORF 20019, exclusivo;
- ✓ 4392XcitriBfeA – Xcitri e Xcamp, ORF 4392;
- ✓ 4459XcitriBfeA – Xcitri e Xcamp, ORF 4459;

- ✓ 4466XcitriBfeA – Xcitri e Xcamp, ORF 4466;
- ✓ 11011XPDBfeA – XPD, ORF 11011;
- ✓ XF2134XCVCBfeA – XCVC, ORF 2134;
- ✓ 4462XcitriBfeA – Xcitri e Xcamp, ORF 4462;
- ✓ XF2137XCVCBfeA – XCVC, ORF 2137;
- ✓ 11031XPDBfeA – XPD, ORF 11031;
- ✓ EcoliFecA – *E.coli*, gi|729471|sp|P13036;
- ✓ 4446XcitriFecA – Xcitri e Xcamp, ORF 4446;

As seqüências que compõem as figuras 4.7 e 4.8 foram selecionadas pela homologia do domínio necessário a todo receptor dependente do sistema TonB: o TonB boxC. Foram selecionados os ORFs com valor E na faixa de 10^{-5} ou menor. Não apresentaram os valores mínimos representantes dos genes: cirA e iroN, e o bfeA teve a representação reduzida. Após esta seleção, foram incluídos os genes originais.

Genes da filogenia das figuras 4.7 e 4.8:

- ✓ YNCD ECOLI – yncD, *E. coli*, gi|6137256|sp|P76115;
- ✓ FECA ECOLI – fecA, *E. coli*, gi|729471|sp|P13036;
- ✓ BFEA BORPT – bfeA, *Bordetella pertussis*, gi|538279|gb|AAA98536.1;
- ✓ FEPA ECOLI – fepA, *E. coli*, gi|2507463|sp|P05825;
- ✓ FHUA ECOLI – fhuA, *E. coli*, gi|2507464|sp|P06971;
- ✓ FHUE ECOLI – fhuE, *E. coli*, gi|2507465|sp|P16869;
- ✓ FPVA PSEAE – fpvA, *Pseudomonas aeruginosa*, gi|12230910|sp|P48632;
- ✓ PBUA PSEAE – pbuA, *Pseudomonas* sp. M114, gi|1172035|sp|Q08017;
- ✓ FYUA YERPE – fyuA, *Yersinia pestis*, gi|17380443|sp|P46359;
- ✓ 66Xcitri FEPA – Xcitri e Xcamp, ORF 66;
- ✓ 5074Xcitri FECA – Xcitri e Xcamp, ORF 5074;
- ✓ 630Xcitri FECA – Xcitri e Xcamp, ORF 630;
- ✓ 6531.1XPD YNCD – XPD, ORF 6531;
- ✓ 1496XCVC YNCD – XCVC, ORF 1496;
- ✓ 4446Xcitri FECA – Xcitri e Xcamp, ORF 4446;

- ✓ 5118Xcitri FYUA – Xcitri e Xcamp, ORF 5118;
- ✓ 25309Xcitri PFEA – Xcitri e Xcamp, ORF 25309;
- ✓ 2146Xcitri FHUA – Xcitri e Xcamp, ORF 2146;
- ✓ 3367Xcitri FECA – Xcitri e Xcamp, ORF 3367;
- ✓ 5444Xcitri FECA – Xcitri e Xcamp, ORF 5444;
- ✓ 4466Xcitri BFEA – Xcitri e Xcamp, ORF 4466;
- ✓ 1715Xcitri FHUA – Xcitri e Xcamp, ORF 1715;
- ✓ 20006Xcamp FHUA – Xcamp, ORF 20006, exclusiva;
- ✓ 15194Xcitri FHUA – Xcitri e Xcamp, ORF 15194;
- ✓ 4855Xcitri FHUA – Xcitri e Xcamp, ORF 4855;
- ✓ 6107Xcitri FHUA – Xcitri e Xcamp, ORF 6107;
- ✓ 14071.2XPD FHUA – XPD, ORF 14071;
- ✓ 0599XCVC YBIL – XCVC, ORF 599;
- ✓ 25860Xcitri FHUE – Xcitri, ORF 25860, exclusiva;
- ✓ 1576Xcitri FHUE – Xcitri e Xcamp, ORF 1576;
- ✓ 7667Xcitri FHUE – Xcitri e Xcamp, ORF 7667;
- ✓ 20674Xcamp FPVA – Xcamp, ORF 20674, exclusiva;
- ✓ 20071Xcamp PBUA – Xcamp, ORF 20071, exclusiva;
- ✓ 25833Xcitri FYUA – Xcitri, ORF 25833, exclusiva;
- ✓ 20018Xcamp NONAME – Xcamp, ORF 20018, exclusiva.

O gráfico da fig. 4.7 foi construído pelo programa clustalw, e da 4.8 pelo programa PAUP, incluindo o bootstraping. Na figura 4.7 pode-se observar que todos os receptores fizeram grupamento próximo aos respectivos genes “originais”. Entenda-se como originais as seqüências dos genes padrão, ou seja, dos genes que primeiro adotaram o nome descrito. Somente cinco ORFs ficaram mal colocados, não correspondendo ao esperado pela anotação: os ORFs fhuA de Xcitri e Xcamp: 2146, 1715 e 20006, e os ORFs fecA também de Xcitri e Xcamp 3367 e 5444. Pode-se destacar, por sua proximidade filogenética, os ORFs: 4464XcitriFecA, 25860XcitriFhuA, 25820XcitriFhuE, 1576XcitriFhuE, 7667XcitrFpvA, 20674XcitriFpvA, 20071XcampPbuA e 25833XcitriFyuA.

O gráfico apresentado na fig. 4.8 foi construído pelo programa PAUP, utilizando parsimônia máxima, e apresenta valores de bootstraping. Foi configurado para 500 réplicas. O dendograma mostrou alguns grupamentos semelhantes aos da fig. 4.7, mas formou um grupo de quatro ORFs distinto do grupo contendo o fecA original (inclui dois fecAs e dois yncDs). Um segundo grupo mais diversificado encontra-se na metade de baixo do gráfico contendo nove ORFs com ligação mais fraca aos originais correspondentes, e também genes originais formando ramificações isoladas: fyuA, yncD, bfeA, pfvA, fepA e fecA. Apresenta ainda quatro ORFs fhuA, muito distantes do fhuA original. O ORF 6107FhuAXcitri está também longe do original, apesar de ter boa homologia com o mesmo.

A árvore gerada com dados do bootstraping exige bastante tempo, principalmente no caso de muitas seqüências. Porém, o recurso é muito útil por acrescentar informações sobre a confiabilidade das ramificações. Além disso, pode-se confiar que as chances da ordem em que as seqüências foram colocadas não irá alterar os resultados.

Comparando-se as figuras 4.7 e 4.8, e considerando que, pelos motivos descritos, a fig. 4.8 represente melhor as relações evolutivas dos genes estudados, obtemos concordâncias e dicrepâncias. Nas duas árvores alguns genes estão agrupados de forma semelhante quanto à distância ao gene original, como os fhuA, fhuE, fpvA e pbuA próximos dos originais. Os ORFs fecA, na fig. 4.8, estão arranjados de forma diferente, que sugere que se revise a anotação do ORF 2146XcitriFHUA, por estar entre vários fecA com estrutura de divisões consistentes. Além de comparar o gene citado com fecA, pode-se comparar os ORFs yncD próximos aos fecA. As ramificações próximas também estão consistentes, o que sugere proximidade ou erro de anotação. A segunda possibilidade pode ser verificada mais facilmente que a primeira.

Apesar da degradação das seqüências destes receptores, presentes em organismos diferentes de sua origem, e da grande semelhança entre os receptores, o estudo filogenético mostrou-se uma ferramenta útil para o estudo específico do grupo, sem pretensões classificatórias para os organismos.

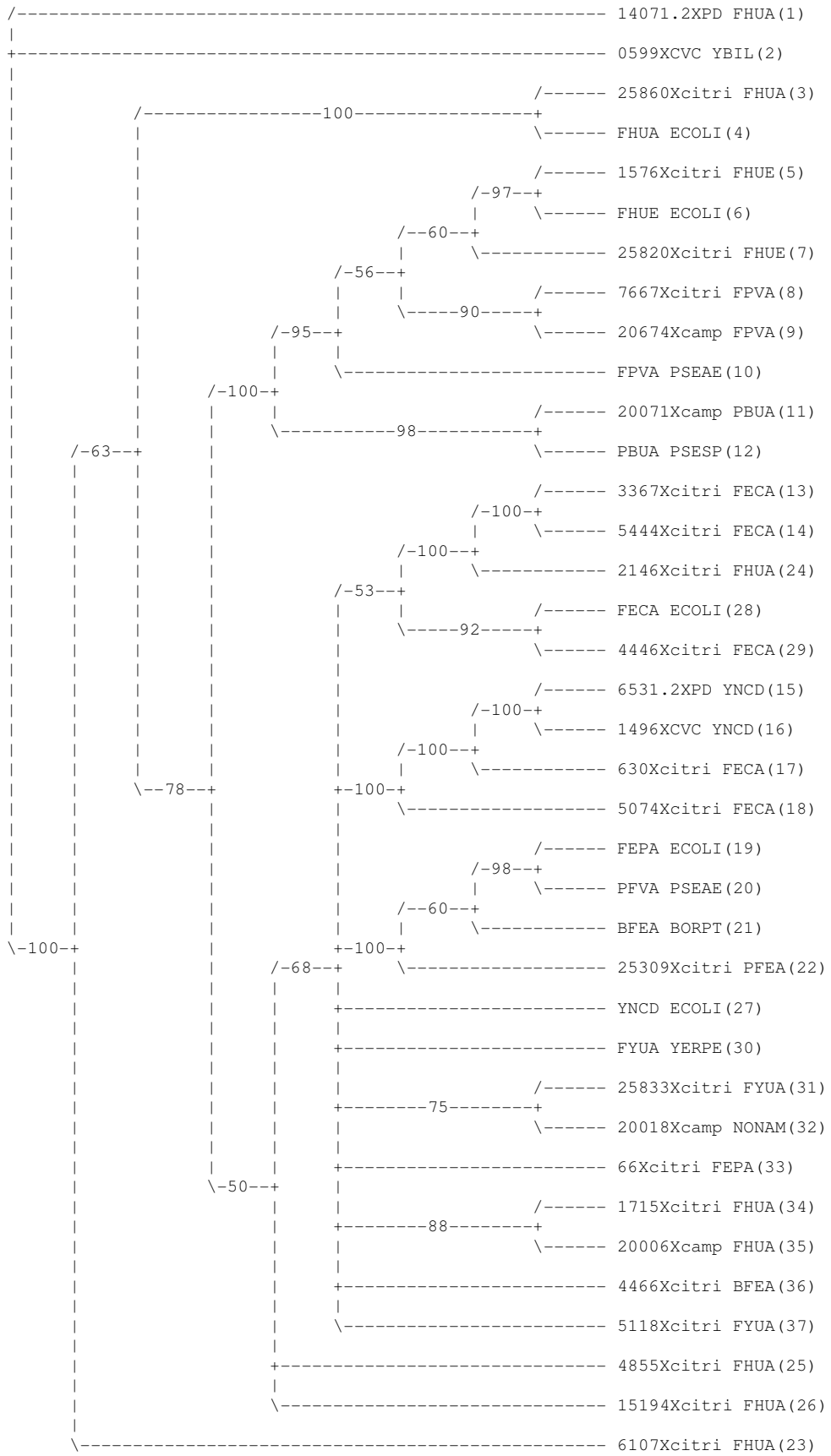


Figura 4.8 - Árvore filogenética de receptores relacionados ao Ferro, selecionados pela homologia com TonB boxC, gerada pelo PAUP, com valores de bootstraping.

5. Conclusões

- ✓ O gene fur, por sua alta conservação, é um ótimo candidato a estudos filogenéticos entre bactérias de grupos próximos.
- ✓ O programa Consed é capaz de desenhar bons primers para auxiliar o sequenciamento de regiões de baixa qualidade.
- ✓ A anotação dos ORFs é dependente da utilização conjunta de vários programas e bancos de dados genéticos, sendo considerados como principais programas: BLAST, PFAM e Cognitor.
- ✓ A filogenia pode ser utilizada como ferramenta auxiliar na anotação de ORFs de grupos de genes semelhantes e com baixa homologia.
- ✓ Em relação à biossíntese de pequenas moléculas, os dois isolados de *Xylella fastidiosa* estudados apresentam os mesmo genes. As espécies do gênero *Xanthomonas* estudadas apresentam poucas diferenças entre si. Xcitri tem a mais duas cópias do gene hemL e os genes cobC e acpD. Xcamp tem a mais o gene cobB.
- ✓ As diferenças efetivas entre os isolados de *Xylella fastidiosa* e as espécies do gênero *Xanthomonas* estudados na biossíntese de pequenas moléculas se resumem à biossíntese de: molibdopterina, piridoxina, nucleotídeos de piridina e cobalamina.
- ✓ O número de receptores dependentes do sistema TonB nas espécies do gênero *Xanthomonas* estudadas é muito maior do que nos isolados de *Xylella fastidiosa* estudados.

Referências Bibliográficas

ACHENBACH, L.A.; YANG, W. The *fur* gene from *Klebsiella pneumoniae*: characterization, genomic organization and phylogenetic analysis. **Gene**, v.185, p.201-207, 1997.

ALTSCHUL, S.F.; MADDEN, T.L.; SCHÄFFER, A.A. et al. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". **Nucleic Acids Research**, v.25, p.3389-3402, 1997.

BAGG, A.; NEILANDS, J.B. Ferric uptake regulation protein acts as a repressor employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. **Biochemistry**, v.26, p.5471-5477, 1987a.

BAGG, A.; NEILANDS, J.B. Molecular mechanism of regulation of siderophore mediated iron assimilation. **Microbiology Review**, v.51, p.509-518, 1987b.

BAIROCH, A. The ENZYME database in 2000. **Nucleic Acids Research**, v.28, p.304-305, 2000.

BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. **Nucleic Acids Research**, v.28, p.45-48, 2000.

BANSAL, A.K. An automated comparative analysis of 17 complete microbial genomes. **Bioinformatics**, v.15, p.900-908, 1999.

BATEMAN, A.; BIRNEY, E.; CERRUTI, R.D. et al. The Pfam protein families database. **Nucleic Acids Research**, v.30, p.276-280, 2002.

BEALL, B.W.; SANDEN, G.N. Cloning and initial characterization of the *Bordetella pertussis fur* gene. **Current Microbiology**, v.30, p.223-226, 1995.

BENSON, D.A.; KARSCH-MIZRACHI, I.; LIPMAN, D.J. et al. GenBank. **Nucleic Acids Research**, v.28, p.15-18, 2000.

BONAS, U.; ACKERVEKEN, G.V. Recognition of bacterial avirulence proteins occurs inside the plant cell: a general phenomenon in resistance to bacterial diseases? **The Plant Journal**, v.12, n.1, p.1-7, 1997.

BRAUN, V.; CHAFFER, S.; HANTKE, K. et al. Regulation of gene expression by iron. In: HAUSKA, G.; THAUER, R. (Ed.) **The molecular basis of bacterial metabolism.** New York: Springer, 1990. p.164-179.

BSAT, N.; HERBIG, A.; CASILLAS-MARTINEZ, L. et al. *Bacillus subtilis* contains multiple Fur homologues: identification of the iron uptake Fur and peroxide regulon PerR repressors. **Molecular Microbiology**, v.29, p.189-198, 1998.

CAMARGO, L.E.A. Análise genética da resistência e da patogenicidade. In: BERGAMIN FILHO, A.; KIMATI, H.; AMORIM, L. (Ed.) **Manual de fitopatologia**: princípios e conceitos. 3.ed. São Paulo: Agronômica Ceres, 1995. v.1, p.470-492.

CHAN, J.W.Y.F.; GOODWIN, P.H. The molecular genetics virulence of *Xanthomonas campestris*. **Biotechnology Advances**, v.17, p.489-508, 1999.

COLLAZO, C.M.; GALÁN, J.E. The invasion-associated type III secretion system in *Salmonella* - a review. **Gene**, v.192, p.51-59, 1997.

COY, M.; DOYLE, C.; BESSER, J. et al. Site directed mutagenesis of the ferric uptake regulation gene of *Escherichia coli*. **Biometals**, v.7, p.292-298, 1994.

DE LA CRUZ, F.; DAVIES, J. Horizontal gene transfer and the origin of species: lessons from bacteria. **Trends in Microbiology**, v.8, n.3, p.128-133, 2000.

DOCENA, C. Seqüenciamento e anotação de parte do genoma de *Xylella fastidiosa* e análise das vias de biossíntese de pequenas moléculas e cofatores. São Paulo, 2000. 156p. Dissertação (Mestrado) – Instituto de Química, Universidade de São Paulo.

DOOLITTLE, R.F. Searching for the common ancestor. **Research in Microbiology**, v.151, p.85-89, 2000.

DRÖGE, M.; PÜHLER, A.; SELBITSCHKA, W. Horizontal gene transfer as a biosafety issue: A natural phenomenon of public concern. **Journal of Biotechnology**, v.64, p.75-90, 1998.

ESCOLAR, L.; PERZ-MARTIN, J.; De LORENZO, V. Binding of the *fur* ferric uptake regulator repressor of *Escherichia coli* to arrays of the GATAAT sequence. **Journal of Molecular Biology**, v.283, p.537-547, 1998.

EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, v.8, p.186-194, 1998.

EWING, B.; HILLIER, L.; WENDL, M.C. et al. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**, v.8, p.175-185, 1998.

FASSBINDER, F.; VAN VLIET, A.H.M.; GIMMEL, V. et al. Identification of iron-regulated genes of *Helicobacter pylori* by a modified Fur titration assay (FURTA-Hp). **FEMS Microbiology Letters**, v.184, p.225-229, 2000.

FERREIRA, A.J.S. Seqüenciamento de parte do genoma de *Xylella fastidiosa* e análise do metabolismo de carboidratos no genoma anotado. São Paulo, 2000. 120p. Dissertação (Mestrado) – Instituto de Química, Universidade de São Paulo.

FOSTER, J.W.; SPECTOR, M.P. How Salmonella survive against the odds. **Annual Review of Microbiology**, v.49, p.145-174, 1995.

GIBAS, C.; JAMBECK, P. **Desenvolvendo bioinformática**: ferramentas de software para aplicações em biologia. Rio de Janeiro: Ed. Campus, 2001. 440p.

GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: a graphical tool for sequence finishing. **Genome Research**, v.8, p.195-202, 1998.

GUERINOT, M.L. Microbial iron transport. **Annual Review of Microbiology**, v.48, p.743-772, 1994.

HALL, H.K.; FOSTER, J.W. The role of Fur in the acid tolerance response of *Salmonella typhimurium* is physiologically and genetically separable from its role in iron acquisition. **Journal of Bacteriology**, v.178, p.5683-5691, 1996.

HANTKE, K. Selection procedure for deregulated iron transport mutants (fur) in *Escherichia coli* K12: *fur* not only affects iron metabolism. **Molecular & General Genetics**, v.210, p.135-139, 1987.

HASSET, D.J.; SOKOL, P.A.; HOWELL, M.L. et al. Ferric uptake regulator Fur mutants of *Pseudomonas aeruginosa* demonstrate defective siderophore-mediated iron uptake altered aerobic growth and decreased superoxide dismutase and catalase activities. **Journal of Bacteriology**, v.178, p.3996-4009, 1996.

HAUSMANN, R. **História da biologia molecular**. Ribeirão Preto: Sociedade Brasileira de Genética, 1997. 312p.

HENTSCHEL, U.; STEINERT, M.; HACKER, J. Common molecular mechanisms of symbiosis and pathogenesis. **Trends in Microbiology**, v.8, n.5, p.226-231, 2000.

HUECK, C.J. Type III protein secretion systems in bacterial pathogens of animals and plants. **Microbiology and Molecular Biology Reviews**, v.62, n.2, p.379-433, 1998.

JORDAN, C.N.B.; PASSOS, G.A.S. Projeto Transcriptona. **Biotecnologia**, v.12, p.21-31, 2000.

KANEHISA, M.; GOTO, S.; KAWASHIMA, S. et al. The KEGG databases at GenomeNet. **Nucleic Acids Research**, v.30, p.42-46, 2002.

KARP, P.D.; RILEY, M.; SAIER, M. et al. The Ecocyc database. **Nucleic Acids Research**, v.30, p.56-58, 2002.

KATZEN, F.; FERREIRO, D.U.; ODDO, C.G. et al. *Xanthomonas campestris* pv. *campestris* gum mutants: effects on xanthan biosynthesis and plant virulence. **Journal of Bacteriology**, v.180, n.7, p.1607-1617, 1998.

LEACH, J.E.; WHITE, F.F. Bacterial avirulence genes. **Annual Review of Phytopathology**, v.34, p.153-179, 1996.

LEE, L.L. Anotação de genes associados à patogenicidade e avirulência em *Xanthomonas axonopodis* pv. *citri*. Rio Claro, 2001. 129p. Dissertação (Mestrado) – Instituto de Biociências, Universidade Estadual Paulista "Júlio de Mesquita Filho".

LEWIS, S.; ASHBURNER, M.; REESE, M.G. Annotating eukaryote genomes. **Current Opinion in Structural Biology**, v.10, p.349-354, 2000.

LI, W.H. **Molecular evolution**. Massachusetts: Sinauer, 1997. 487p.

LIN, T.L. Montagem de fragmentos de DNA pelo método "Ordered Shotgun Sequencing" (OSS). Campinas, 2001. 135p. Dissertação (Mestrado) – Instituto de Computação, Universidade Estadual de Campinas.

LITWIN, C.M.; CALDERWOOD, S.B. Role of iron in regulation of virulence genes. **Clinical Microbiology Reviews**, v.6, n.2, p.137-149, 1993.

LOPRASERT, S.; SALLABHAN, R.; ATICHARTPONGKUL, S. et al. Characterization of a ferric uptake regulator (fur) gene from *Xanthomonas campestris* pv.*phaseoli* whit unusual primary structure, genome organization, and expression patterns. **Gene**, v.239, p.251-258, 1999.

MASSUNG, R.F.; ESPOSITO, J.J.; LIU, L.I.; et al. Potential virulence determinants in terminal regions of variola smallpox virus genome. **Nature**, v.366, p.748-751, 1993.

NAKAI, K.; HORTON, P. PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. **Trends in Biochemical Sciences**, v.24, p.34-35, 1999.

OCHSNER, U.A.; VASIL, A.I.; VASIL, M.L. Role of the ferric uptake regulator of *Pseudomonas aeruginosa* in the regulation of siderophores and exotoxin A expression: purification and activity on iron-regulated promoters. **Journal of Bacteriology**, v.177, p.7194-7201, 1995.

PALLEN, M.J. Microbial genomes. **Molecular Microbiology**, v.32, n.5, p.907-912, 1999.

SCHLOTER, M.; LEBUHN, M.; HEULIN, T. et al. Ecology and evolution of bacterial microdiversity. **Fems Microbiology Reviews**, v.24, p.647-660, 2000.

SCHULTZ, J.; MILPETZ, F.; BORK, P. et al. SMART, a simple modular architecture research tool: identification of signaling domains. **Proceedings of the National Academy of Sciences of the United States of America**, v.95, p.5857-5864, 1998.

SETUBAL, J.C.; MEIDANIS, J. **Introduction to computational molecular biology**. Boston: PWS Publishing, 1997.

SILVA, A.C.R. et al. Complete genome sequences of two *Xanthomonas* pathogens with similar genomes but different host specificities. **Nature** (Letters), aceito para publicação, 2002. http://cancer.lbi.ic.unicamp.br/xanthomonas/

SIMPSON, A.J.G. et al. (ONSA) The genome sequence of the plant pathogen *Xylella fastidiosa*. **Nature**, v.406, p.151-157, 2000. http://aeg.lbi.ic.unicamp.br/xf/

STAGGS, T.M.; PERRY, R.D. Identification and cloning of a *fur* regulatory gene in *Yersinia pestis*. **Journal of Bacteriology**, v.173, p.417-425, 1991.

STRAUSS. E.J.; FALKOW, S. Microbial pathogenesis: genomics and beyond. **Science**, v.276, p.707-712, 1997.

SWOFFORD, D.L. **PAUP***. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland: Sinauer Associates, 2002.

TATUSOV, R.L.; NATALE, D.A.; GARKAVTSEV, I.V. et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. **Nucleic Acids Research**, v.29, p.22-28, 2001.

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v.22, p.4673-4680, 1994.

TSOLIS, R.M.; BÄULMER, A.J.; STOJILJKOVIC, I. Et al. Fur regulon of *Salmonella typhimurium*: Identification of new iron-regulated genes. **Journal of Bacteriology**, v.177, p.4628-4637, 1995.

VASIL, M.L.; OCHSNER, U.A. The response of *Pseudomonas aeruginosa* to iron: genetics, biochemistry and virulence. **Molecular Microbiology**, v.34, n.3, p.399-413, 1999.

WATERMAN, M.S. **Introduction to computational biology**: Maps, sequences and genomes. London: Chapman & Hall, 1995. 431p.

WHEELER, D.L.; CHAPPEY, C.; LASH, A.E. et al. Database resources of the National Center for Biotechnology Information. **Nucleic Acids Research**, v.28, p.10-14, 2000.

WHITE, F.F.; YANG, B.; JOHNSON, L.B. Prospects for understanding avirulence gene function. **Current Opinion in Plant Biology**, v.3, n.4, p.291-298, 2000.

WILSON, T.J.G.; BERTRAND, N.; TANG, J.L. et al. The rpfA gene of *Xanthomonas campestris* pv.*campestris*, which is involved in the regulation of pathogenicity factor production, encodes an aconitase. **Molecular Microbiology**, v.28, n.5, p.961-970, 1998.

Bibliografia Recomendada

BRAUN, V.; HANTKE, K. Genetics of bacterial iron transport. In: WINKELMANN, G. (Ed.) **Handbook of microbial iron chelates**. Boca Raton: CRC Press, 1991. p.107-138.

DOERKS, T.; BAIROCH, A.; BORK, P. Protein annotation: detective work for function prediction. **Trends in Genetics**, v.14, n.6, p.248-250, 1998.

FERREIRA, L.P.; SALGADO, C.L. Bactérias. In: BERGAMIN FILHO, A.; KIMATI, H.; AMORIM, L. (Ed.) **Manual de fitopatologia**: princípios e conceitos. 3.ed. São Paulo: Agronômica Ceres, 1995. v.1, p.97-131.

FIELDS, C.; ADAMS, M.D.; WHITE, O. et al. Predicting the total number of human genes – Reply. **Nature Genetics**, v.8, n.2, p.114, 1994.

HOPKINS, D.L. Physiological and pathological characteristics of virulent and avirulent strains of the bacterium that causes Pierce's disease in grapevines. **Physiology and Biochemistry**, v.75, n.6, p.713-717, 1985.

LAMBAIS, M.R.; GOLDMAN, M.H.S.; CAMARGO, L.E.A. A genomic approach to the understanding of *Xylella fastidiosa* pathogenicity **Current Opinion in Microbiology**, v.3, p.459-462, 2000.

PAYNE, S.M. Iron acquisition in microbial pathogenesis. **Trends in Microbiology**, v.1, n.2, p.66-69, 1993.

PURCELL, A.H.; HOPKINS, D.L. Fastidious xylem-limited bacterial plant pathogens. **Annual Review of Phytopathology**, v.34, p.131-151, 1996.

RATLEDGE, C.; FOVER, L.G. Iron metabolism in pathogenic bacteria. **Annual Review of Microbiology**, v.54, p.881-941, 2000.

SETUBAL, J.; WERNECK, R.F. **A program for building contigs scaffolds in double-barreled shotgun genome sequencing**. Campinas: UNICAMP, Instituto de Computação, 2000. (Relatório Técnico 00-20).

STOJILJKOVIC, I.; BÄULMER, A.J.; HANTKE, K. Fur regulon in gram-negative bacteria: identification and characterization of new iron-regulated *Escherichia coli* by a fur tritation assay. **Journal of Molecular Biology**, v.236, p.531-545, 1994.

STOJILJKOVIC, I.; HANTKE, K. Functional domains of the *Escherichia coli* ferric uptake regulator protein. **Molecular and General Genetics**, v.247, p.199-205, 1995.

VATTANAVIBOON, P.; MONGKOLSUK, S. Evaluation of the role hydroxyl radicals and iron play in hydrogen peroxide killing of *Xanthomonas campestris* pv. *phaseoli*. **Fems Microbiology Letters**, v.169, p.225-260, 1998.

ANEXO 1

Programas Locais, Páginas de Projetos e
Páginas de Busca e Consulta

Programas locais

- ✓ Clustalw – alinhamentos múltiplos;
- ✓ *Excel* – Planilha de cálculos.
- ✓ *Netscape* ou *Internet Explorer* (navegador e correio eletrônico);
- ✓ *PAUP* – filogenia;
- ✓ *Phred, Phrap* & *Consed*, *Repeat Master*, *Cross Match* – montagem;
- ✓ *Sequencher* – montagem;
- ✓ *Sequencing Analysis* – sequenciamento;
- ✓ *Word* - Processador de textos;

Páginas dos Projetos – informações, serviços e suporte à anotação

- ✓ Projeto Genoma *Xylella*

http://aeg.lbi.ic.unicamp.br/xf/

- ✓ Projeto *Xylella* Funcional

http://watson.fapesp.br/funcional/main.htm

- ✓ Projeto Genoma *Xanthomonas*

http://genoma4.iq.usp.br/xanthomonas/

- ✓ Projeto Genoma *Leifsonia xyli* subsp. *xyli*

http://aeg.lbi.ic.unicamp.br/leifsonia/

- ✓ Projeto Genoma *Xylella fastidiosa / Pierce's Disease*

http://aeg.lbi.ic.unicamp.br/xf-grape/

Páginas de Programas, Busca e Consulta

- ✓ BLAST (NCBI)

http://www.ncbi.nlm.nih.gov/BLAST/

- ✓ CLUSTAL

http://www.ebi.ac.uk/clustalw/

- ✓ COGs - *Clusters of Orthologous Groups of proteins*

http://www.ncbi.nlm.nih.gov/COG/

- ✓ ECOCYC

http://ecocyc.pangeasystems.com/

- ✓ Entrez – Taxonomy (NCBI)

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Taxonomy

- ✓ ENZYME Search (ExPASy)

http://ca.expasy.org/enzyme/

- ✓ KEGG - *Kyoto Encyclopedia of Genes and Genomes*

http://www.genome.ad.jp/kegg/kegg2.html

- ✓ NCBI – *National Center of Biotechnology Information*

http://www.ncbi.nlm.nih.gov/

- ✓ PAUP

http://paup.csit.fsu.edu/index.html

- ✓ PFAM

http://pfam.wustl.edu/index.html

- ✓ Phred/Phrap/Consed

http://www.phrap.org/

- ✓ Proteínas de Transporte – Dr. Milton Saier Jr.

http://www-biology.ucsd.edu/~msaier/transport/index.html

- ✓ PSORT

http://psort.nibb.ac.jp/

- ✓ SMART – *Simple Modular Architecture Research Tool*

http://smart.embl-heidelberg.de/

- ✓ Swiss-Prot (ExPASy)

http://www.expasy.ch/sprot/

ANEXO 2

Categorização Oficial das bactérias:
Xanthomonas axonopodis pv. *citri*,
Xanthomonas campestris pv. *campestris* e
Xylella fastidiosa / *Pierce's Disease*

Categorização oficial adotada para a *X. axonopodis* pv. *Citri e X. campestris* pv. *campestris*

I. Intermediary metabolism
  A. Degradation
    1. Degradation of polysaccharides and oligosaccharides
    2. Degradation of small molecules
  B. Central intermediary metabolism
    1. Amino sugars
    2. Entner-Douderoff
    3. Gluconeogenesis
    4. Glyoxylate bypass
    5. Miscellaneous glucose metabolism
    6. Non-oxidative branch, pentose pathway
    7. Nucleotide hydrolysis
    8. Nucleotide interconversions
    9. Phosphorus compounds
    10. Pool, multipurpose conversions
    11. Sugar-nucleotide biosynthesis, conversions
    12. Sulfur metabolism
  C. Energy metabolism, carbon
    1. Aerobic respiration
    2. Anaerobic respiration and fermentation
    3. Electron transport
    4. Glycolysis
    5. Oxidative branch, pentose pathway
    6. Pyruvate dehydrogenase
    7. TCA cycle
    8. ATP-proton motive force interconversion
  D. Regulatory functions
    1. Two component systems
    2. Repressors
    3. Phosphatases
    4. Sigma factors and other regulatory components
    5. Not Used
II. Biosynthesis of small molecules
  A. Amino acids biosynthesis
    1. Glutamate family|nitrogen assimilation
    2. Aspartate family, pyruvate family
    3. Glycine-serine family|sulfur metabolism
    4. Aromatic amino acid family
    5. Histidine
  B. Nucleotides biosynthesis
    1. Purine ribonucleotides
    2. Pyrimidine ribonucleotides
    3. 2'-Deoxyribonucleotides

4. Salvage of nucleosides and nucleotides

C. Sugars and sugar nucleotides biosynthesis

D. Cofactors, prosthetic groups, carriers biosynthesis

1. Biotin
2. Folic acid
3. Lipoate
4. Molybdopterin
5. Pantothenate
6. Pyridoxine
7. Pyridine nucleotides
8. Thiamin
9. Riboflavin
10. Thioredoxin, glutaredoxin, glutathione
11. Menaquinone, ubiquinone
12. Heme, porphyrin
13. Biotin carboxyl carrier protein (BCCP)
14. Cobalamin
15. Enterochelin
16. Biopterin
17. Others

E. Fatty acid and phosphatidic acid biosynthesis

F. Polyamines biosynthesis

III. Macromolecule metabolism

A. DNA metabolism

1. Replication
2. Structural DNA binding proteins
3. Recombination
4. Repair
5. Restriction, modification

B. RNA metabolism

1. Ribosomal and stable RNAs
2. Ribosomal proteins
3. Ribosomes - maturation and modification
4. Aminoacyl tRNA synthetases, tRNA modification
5. RNA synthesis, modification, DNA transcription
6. RNA degradation

C. Protein metabolism

1. Translation and modification
2. Chaperones
3. Protein degradation

D. Other macromolecules metabolism

1. Polysaccharides
2. Phospholipids
3. Lipoprotein

IV. Cell structure

A. Membrane components

1. Inner membrane
2. Outer membrane constituents

- B. Murein sacculus, peptidoglycan
- C. Surface polysaccharides, lipopolysaccharides, and antigens
- D. Surface structures

V. Cellular processes
- A. Transport
  1. Amino acids, amines
  2. Anions
  3. Carbohydrates, organic acids, alcohols
  4. Cations
  5. Nucleosides, purines, pyrimidines
  6. Protein, peptide secretion
  7. Other
- B. Cell division
- C. Chemotaxis and mobility
- D. Osmotic adaptation
- E. Cell killing

VI. Mobile genetic elements
- A. Phage-related functions and prophages
- B. Plasmid-related functions
- C. Transposon- and intron-related functions

VII. Pathogenicity, virulence, and adaptation
- A. Avirulence
- B. Hypersensitive response and pathogenicity
- C. Toxin production and detoxification
- D. Host cell wall degradation
- E. Exopolysaccharides
- F. Surface proteins
- G. Adaptation, atypical conditions
- H. Other

VIII. Hypothetical
- A. Conserved hypothetical proteins
- B. Hypothetical proteins (includes no hits or only low score hits)

IX. ORFs with undefined category

Categorização oficial adotada para a *Xylella fastidiosa / Pierce's Disease*

I. Intermediary metabolism
  A. Degradation
    1. Degradation of polysaccharides and oligosaccharides
    2. Degradation of small molecules
    3. Degradation of lipids
  B. Central intermediary metabolism
    1. Amino sugars
    2. Entner-Douderoff
    3. Gluconeogenesis
    4. Glyoxylate bypass
    5. Miscellaneous glucose metabolism
    6. Non-oxidative branch, pentose pathway
    7. Nucleotide hydrolysis
    8. Nucleotide interconversions
    9. Phosphorus compounds
    10. Pool, multipurpose conversions
    11. Sugar-nucleotide biosynthesis, conversions
    12. Sulfur metabolism
  C. Energy metabolism, carbon
    1. Aerobic respiration
    2. Anaerobic respiration and fermentation
    3. Electron transport
    4. Glycolysis
    5. Oxidative branch, pentose pathway
    6. Pyruvate dehydrogenase
    7. TCA cycle
    8. ATP-proton motive force interconversion
  D. Regulatory functions
    1. Two component systems
    2. Repressors
    3. Phosphatases
    4. Sigma factors and other regulatory components
  E. Uncharacterized
II. Biosynthesis of small molecules
  A. Amino acids biosynthesis
    1. Glutamate family|nitrogen assimilation
    2. Aspartate family, pyruvate family
    3. Glycine-serine family|sulfur metabolism
    4. Aromatic amino acid family
    5. Histidine
  B. Nucleotides biosynthesis
    1. Purine ribonucleotides
    2. Pyrimidine ribonucleotides

3. 2'-Deoxyribonucleotides
4. Salvage of nucleosides and nucleotides

C. Sugars and sugar nucleotides biosynthesis
D. Cofactors, prosthetic groups, carriers biosynthesis
1. Biotin
2. Folic acid
3. Lipoate
4. Molybdopterin
5. Pantothenate
6. Pyridoxine
7. Pyridine nucleotides
8. Thiamin
9. Riboflavin
10. Thioredoxin, glutaredoxin, glutathione
11. Menaquinone, ubiquinone
12. Heme, porphyrin
13. Biotin carboxyl carrier protein (BCCP)
14. Cobalamin
15. Enterochelin
16. Biopterin
17. Others

E. Fatty acid and phosphatidic acid biosynthesis
F. Polyamines biosynthesis
G. Uncharacterized

III. Macromolecule metabolism
A. DNA metabolism
1. Replication
2. Structural DNA binding proteins
3. Recombination
4. Repair
5. Restriction, modification

B. RNA metabolism
1. Ribosomal and stable RNAs
2. Ribosomal proteins
3. Ribosomes - maturation and modification
4. Aminoacyl tRNA synthetases, tRNA modification
5. RNA synthesis, modification, DNA transcription
6. RNA degradation

C. Protein metabolism
1. Translation and modification
2. Chaperones
3. Protein degradation

D. Other macromolecules metabolism
1. Polysaccharides
2. Phospholipids
3. Lipoprotein

E. Uncharacterized

IV. Cell structure

- A. Membrane components
  1. Inner membrane
  2. Outer membrane
  3. Uncharacterized
- B. Murein sacculus, peptidoglycan
- C. Surface polysaccharides, lipopolysaccharides, and antigens
- D. Surface structures
- E. Uncharacterized

V. Cellular processes
- A. Transport
  1. Amino acids, amines
  2. Anions
  3. Carbohydrates, organic acids, alcohols
  4. Cations
  5. Nucleosides, purines, pyrimidines
  6. Protein, peptide secretion
  7. Other
- B. Cell division
- C. Chemotaxis and mobility
- D. Osmotic adaptation
- E. Cell killing
- F. Uncharacterized

VI. Mobile genetic elements
- A. Phage-related functions and prophages
- B. Plasmid-related functions
- C. Transposon- and intron-related functions
- D. Uncharacterized

VII. Pathogenicity, virulence, and adaptation
- A. Avirulence
- B. Hypersensitive response and pathogenicity
- C. Toxin production and detoxification
- D. Host cell wall degradation
- E. Exopolysaccharides
- F. Surface proteins
- G. Adaptation, atypical conditions
- H. Other

VIII. Hypothetical
- A. Conserved hypothetical proteins
- B. Hypothetical proteins (includes no hits or only low score hits)

IX. ORFs with undefined category

ANEXO 3

Tabela comparativa de genes da categoria
II – Biossíntese de Pequenas Moléculas das bactérias:
Xylella fastidiosa (CVC),
Xanthomonas axonopodis pv. *citri* (cancro),
Xanthomonas campestris pv. *campestris* e
Xylella fastidiosa / *Pierce's Disease*

&

Mapas metabólicos referentes às vias de
biossíntese de pequenas moléculas

Anotação Xylella PD - Categoria II - Biossíntese de Pequenas Moléculas

II.A. Aminoácidos

II.A.1. Família do Glutamato

Arginina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
argA		2.3.1.1				
argB	acetylglutamate kinase	2.7.2.8	2641	1001	6413	6413
argC	N-acetyl-gamma-glutamyl- phosphate reductase	1.2.1.38	2651	1002	6416	6416
argD	argininosuccinate lyase	2.6.1.11	5951	1427	2142	2142
argE	acetylornithine deacetylase	3.5.1.16	2631	1000	6412	6412
argF	ornithine carbamoyltransferase	2.1.3.3	2611	998	6408	6408
argG	argininosuccinate synthase	6.3.4.5	2621	999	6409	6409
argH	argininosuccinate lyase	4.3.2.1	2661	1003	6418	6418
carA	carbamoyl-phosphate synthase small chain	6.3.5.5	3571	1106	2345	2345
carB	carbamoyl-phosphate synthase large chain	6.3.5.5	3581	1107	2347	2347
gltX	glutamyl-tRNA synthetase	6.1.1.17	16951	822	1745	1745

Glutamato						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
gdhA	NAD-glutamate dehydrogenase	1.4.1.4	7231	2091	825	825
gltB aspB	glutamate synthase, alpha subunit	1.4.1.13	18891	2710	5371	5371
gltD aspB	glutamate synthase, beta subunit	1.4.1.13	18881	2709	5373	5373
tyrB	aspartate aminotranferase	2.6.1.57 2.6.1.1	231	36	3642	3642

Glutamina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
glnA	glutamine synthetase	6.3.1.2	9311	1842	7618	7618
glnB	nitrogen regulatory protein P-II	-	9301	1843	7617	7617

Prolina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
proA	gamma-glutamyl phosphate reductase	1.2.1.41	2681	1005	6423	6423
proB	glutamate 5-kinase	2.7.2.11	2671	1004	6421	6421
proC	pyrroline-5-carboxylate reductase	1.5.1.2	18901	2712	6088	6088

II.A.2. Família do Aspartato / Família do Piruvato

Alanina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
alr	alanine racemase	5.1.1.1	16691	852	957	957
avtA		2.6.1.66				
ilvE	branched-chain amino acid aminotransferase	2.6.1.42	7481	1999	3206	3206

Asparagina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
asnB	asparagine synthase B	6.3.5.4	741	118	4861	4861

Aspartato						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
argG	argininosuccinate synthase	6.3.4.5	2621	999	6409	6409
argH	argininosuccinate lyase	4.3.2.1	2661	1003	6418	6418
aspC	aminotransferase	2.6.1.1	12911	2396	2283	2283
aspC	aminotransferase	2.6.1.1	13441	2403	6452	6452
aspH	aspartyl/asparaginyl beta-hydroxylase	-	7171	2100	7363	7363
purA	adenylosuccinate synthetase	6.3.4.4	14801	455	6016	6016
purB	adenylosuccinate lyase	4.3.2.2	7021	1553	4655	4655

Isoleucina, Valina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
ilvA	threoninhe dehydratase (biosynthetic)	4.2.1.16	9511	1819	7359	7359
ilvC	ketol-acid reductoisomerase	1.1.1.86	9481	1822	2104	2104
ilvD	dihydroxy-acid dehydratase	4.2.1.9	601	99	7293	7293
ilvE	branched-chain amino acid aminotransferase	2.6.1.42	7481	1999	3206	3206
ilvG	acetolactate synthase isozyme II, large subunit	4.1.3.18	9491	1821	2102	2102
ilvM	acetolactate synthase isozyme II, small sbunit	4.1.3.18	9501	1820	2101	2101

Leucina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
icdA/ leuB	Isocitrate/isopropylmalate dehydrogenase	1.1.1.41 1.1.1.8	18131	2596	3404	3404
ilvE	branched-chain amino acid aminotransferase	2.6.1.42	7481	1999	3206	3206
leuA	2-isopropylmalate synthase	4.1.3.12	9521	1818	2097	2097
leuB	3-isopropylmalate dehydrogenase	1.1.1.85	12771	2372	2096	2096
leuC	3-isopropylmalate dehydratase large subunit	4.2.1.33	12791	2375	2092	2092
leuD	3-isopropylmalate dehydratase small subunit	4.2.1.33	12781	2374	2094	2094
tyrB	aspartate aminotranferase	2.6.1.57 2.6.1.1	231	36	3642	3642

Lisina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
asd	aspartate-semialdehyde dehydrogenase	1.2.1.11	5471	1371	3858	3858
dapA	dihydroxydipicolinate synthase	4.2.1.52	15871	963	3462	3462
dapB	dihydrodipicolinate reductase	1.3.1.26	3561	1105	2343	2343
dapC		2.6.1.17				
dapD	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase	2.3.1.117	721	114	4867	4867
dapE	succinyl-diaminopimelate desuccinylase	3.5.1.18	731	116	4862	4862
dapF	diaminopimelate epimerase	5.1.1.7	6391	1481	11368	11368
lysA/ lysC	bifunctional diaminopimelate decarboxylase/aspartate kinase	4.1.1.20 2.7.2.4	3671	1116	6063	6063
lysC/ thrA	bifunctional aspartokinase/homoserine dehydrogenase I	2.7.2.4 1.1.1.3	11621	2225	1067	1067

Metionina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
asd	aspartate-semialdehyde dehydrogenase	1.2.1.11	5471	1371	3858	3858
lysC/ thrA	bifunctional aspartokinase/homoserine dehydrogenase I	2.7.2.4 1.1.1.3	11621	2225	1067	1067
masA	enolase-phosphatase	3.-.-.-	11491	2211	2307	2307
metA	homoserine O-acetyltransferase	2.3.1.31	13541	2465	6437	6437
metA	homoserine O-acetyltransferase	2.3.1.31	16611	863	6301	6301
metB	cystathionine gamma-synthase	4.2.99.9	16601	864	6300	6300
metC		4.4.1.8				
metE	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	2.1.1.14	11951	2272	1837	1837
metH		2.1.1.13				
metL		1.1.1.3				

Treonina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
asd	aspartate-semialdehyde dehydrogenase	1.2.1.11	5471	1371	3858	3858
lysC/ thrA	bifunctional aspartokinase/homoserine dehydrogenase I	2.7.2.4 1.1.1.3	11621	2225	1067	1067
thrB	homoserine kinase	2.7.1.39	11611	2224	1066	1066
thrC	threonine synthase	4.2.99.2	11601	2223	1063	1063

II.A.3. Família Glicina/Serina

Cisteína						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
csdB	selenocysteine lyase	4.4.1.16	6301	1473	6103	6103
cysB	transcriptional regulator, LysR family	-	16871	833	23	23
cysE		2.3.1.30				
cysK	cysteine synthase	4.2.99.8	16891	831	26	26
cysK	cysteine synthase	4.2.99.8	821	128	223	223
cysM2	cystathionine beta-synthase	4.2.1.22	14041	603	3145	3145
metB	cystathionine gamma-synthase	4.2.99.9	16601	864	6300	6300

Glicina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
glyA	serine hydroxymethyltransferase	2.1.2.1	16011	946	5165	5165

Serina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
glyA	serine hydroxymethyltransferase	2.1.2.1	16011	946	5165	5165
ilvA	threoninhe dehydratase (biosynthetic)	4.2.1.16	9511	1819	7359	7359
serA	D-3-phosphoglycerate dehydrogenase	1.1.1.95	11431	2206	2316	2316
serB		3.1.3.3				
serC	phosphoserine aminotransferase	2.6.1.52	12431	2326	10912	10912
ydfG	oxireductase	1.1.1.-	951	145	1160	1160

II.A.4. Família de Aminoácidos Aromáticos

Corismato						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
aroA	3-phosphoshikimate 1-carboxyvinyltransferase	2.5.1.19	12411	2324	10908	10908
aroB	3-dehydroquinate synthase	4.6.1.3	5211	1334	6244	6244
aroC	chorismate synthase	4.6.1.4	5461	1369	3861	3861
aroE	shikimate 5-dehydrogenase	1.1.1.25	13921	624	2826	2826
aroG	phospho-2-dehydro-3-deoxyheptonate aldolase	4.1.2.15	161	26	7129	7129
aroK	shikimate kinase	2.7.1.71	5221	1335	6242	6242
aroL		2.7.1.71				
aroQ	catabolic dehydroquinase	4.2.1.10	291	47	7154	7154

Fenilalanina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
pheA	chorismate mutase	5.4.99.5	3861	1141	1024	1024
pheA	P-protein (chorismate mutase / prephenate dehydratase)	5.4.99.5 4.2.1.51	12421	2325	10910	10910
tyrA	chorismate mutase/prephenate dehydrogenase	5.4.99.5 1.3.1.12	12531	2338	4686	4686
tyrB	aspartate aminotranferase	2.6.1.57 2.6.1.1	231	36	3642	3642

Triptofano						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
trpA	tryptophan synthase alpha chain	4.2.1.20	5521	1376	3846	3846
trpB	tryptophan synthase beta chain	4.2.1.20	5511	1375	3849	3849
trpC	indole-3-glycerol phosphate synthase	4.1.1.48	1451	213	11458	11458
trpD	anthranilate phosphoribosyltransferase	2.4.2.18	1441	212	11459	11459
trpD	anthranilate synthase component II	4.1.3.27	1431	211	11463	11463
trpE	anthranilate synthase component I	4.1.3.27	1421	210	11466	11466
trpE	anthranilate synthase component I	4.1.3.27	13681	674	5678	5678
trpF	phosphoribosylanthranilate isomerase	5.3.1.24	5501	1374	3853	3853
wrbA	tryptophan repressor binding protein	-	3511	1094	3226	3226
wrbA	tryptophan repressor binding protein	-	3801	1133	1443	1443

Tirosina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
pheA	P-protein (chorismate mutase / prephenate dehydratase)	5.4.99.5 4.2.1.51	12421	2325	10910	10910
tyrA	chorismate mutase/prephenate dehydrogenase	5.4.99.5 1.3.1.12	12531	2338	4686	4686
tyrB	aspartate aminotranferase	2.6.1.57 2.6.1.1	231	36	3642	3642

II.A.5. Histidina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
hisA	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	5.3.1.16	11521	2215	1046	1046
hisB	imidazoleglycerolphosphate dehydratase/histidinol-phosphate phosphatase bifunctional enzyme	4.2.1.19 3.1.3.15	11541	2217	1049	1049
hisC	histidinol-phosphate aminotransferase	2.6.1.9	11551	2218	1050/ 2151/ 2486	1050/ 2151/ 2486
hisC	histidinol-phosphate aminotransferase	2.6.1.9	30261		1830	1830
hisD	histidinol dehydrogenase	1.1.1.23	11561	2219	1051	1051
hisF	imidazoleglycerol-phosphate synthase, cyclase subunit	2.4.2.-	11511	2214	1045	1045
hisG	ATP phosphoribosyltransferase	2.4.2.17	11571	2220	1052	1052
hisH	amidotransferase	2.4.2.-	11531	2216	1047	1047
hisI	phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase bifunctional enzyme	3.5.4.19 3.6.1.3	11501	2213	1044	1044

II.B. Nucleotídeos

II.B.1. Ribonucleotídeos de Purina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
adk	adenylate kinase	2.7.4.3	1891	275	2131	2131
gmk	guanylate kinase	2.7.4.8	6611	1503	1623	1623
guaA	glutamine amidotransferase	6.3.5.2 6.3.4.1	13191	2429	6505	6505
guaA	glutamine amidotransferase	6.3.5.2 6.3.4.1	14351	560	596	596
guaB	inosine-5'-monophosphate dehydrogenase	1.1.1.205	13201	2430	6504	6504
ndk	nucleoside diphosphate kinase	2.7.4.6	14781	458	6849	6849
prsA	phosphoribosyl pyrophosphate synthetase	2.7.6.1	18481	2644	3246	3246
purA	adenylosuccinate synthetase	6.3.4.4	14801	455	6016	6016
purB	adenylosuccinate lyase	4.3.2.2	7021	1553	4655	4655
purC	phosphoribosylaminoimidazole-succinocarboxamide synthase	6.3.2.6	1391	205	11474	11474
purD	Phosphoribosylamine-glycine ligase	6.3.4.13	7651	1976	7189	7189
purE	phosphoribosylaminoimidazole carboxylase, catalytic subunit	4.1.1.21	18631	2672	3792	3792
purF	amidophosphoribosyltransferase	2.4.2.14	7851	1949	3380	3380
purH	bifunctional purine biosynthesis protein	3.5.4.10 2.1.2.3	7661	1975	7186	7186
purK	phosphoribosylaminoimidazole carboxylase, ATPase subunit	4.1.1.21	18621	2671	3791	3791
purL	phosphoribosylformylglycinamidine synthetase	6.3.5.3	5901	1423	286	286
purM purG	5'-phosphoribosyl-5-aminoimidazole synthetase	6.3.3.1	14191	587	6141	6141
purN	5'-phosphoribosylglycinamide transformylase	2.1.2.2	14201	585	6144	6144
purT	phosphoribosylglycinamide formyltransferase 2	2.1.2.-			5870	5870
purU	formyltetrahydrofolate deformylase	3.5.1.10	9411	1831	1815	1815
ushA	5'-nucleotidase	3.1.3.5	7251	2089	2538	2538

II.B.2. Ribonucleotídeos de Pirimidina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
carA	carbamoyl-phosphate synthase small chain	6.3.5.5	3571	1106	2345	2345
carB	carbamoyl-phosphate synthase large chain	6.3.5.5	3581	1107	2347	2347
cmk	cytidylate kinase	2.7.4.14	13281	2439	6490	6490
dnaS	dUTPase	3.6.1.23	1011	150	1151	1151
ndk	nucleoside diphosphate kinase	2.7.4.6	14781	458	6849	6849
pyrB	aspartate carbamoyltransferase	2.1.3.2	11631	2226	6075	6075
pyrC	dihydroorotase	3.5.2.3	2511	988	6399	6399
pyrD	dihydroorotate dehydrogenase	1.3.3.1	17941	2571	1099	1099
pyrE	orotate phosphoribosyl transferase	2.4.2.10	1041	153	1132	1132
pyrF	orotidine 5'-phosphate decarboxylase	4.1.1.23	221	34	6701	6701
pyrG	CTP synthetase	6.3.4.2	4831	1288	3536	3536
pyrI		2.1.3.2				
thyA	thymidylate synthase	2.1.1.45	12491	2332	2712	2712
trxB	thioredoxin reductase	1.6.4.5	6081	1448	6826	6826
ushA	5'-nucleotidase	3.1.3.5	7251	2089	2538	2538

II.B.3. 2'-Deoxiribonucleotídeos						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
dcd	deoxycytidine triphosphate deaminase	3.5.4.13	17391	762	3888	3888
glyA	serine hydroxymethyltransferase	2.1.2.1	16011	946	5165	5165
mutT	dNTP pyrophosphohydrolase	3.6.1.-	6031	1441	6815	6815
mutT/ thiE	bifunctional DGTP-pyrophosphohydrolase/ thiamine phosphate synthase	3.6.1.- 2.5.1.3	3701	1120	5236	5236
nrdA	ribonucleoside-diphosphate reductase alpha chain	1.17.4.1	4231	1196	2897	2897
nrdB	ribonucleoside-diphosphate reductase beta chain	1.17.4.1	4241	1197	2892	2892
tmk	thymidylate kinase	2.7.4.9	14231	580	2789	2789

II.B.4. Salvamento de Nucleotídeos e Nucleosídeos						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
add	adenosine deaminase	3.5.4.4			1788	1788
apaH	diadenosine tetraphosphatase	3.6.1.41	11131	2150	2686	2686
deoD pnp	purine nucleoside phosphorylase	2.4.2.28	12641	2353	71	71
enpP	phosphodiesterase-nucleotide pyrophosphatase precursor	3.6.1.9	18161	2599	1705	1705
hpt	hypoxanthine-guanine phosphoribosyltransferase	2.4.2.8	12651	2354	70	70
nadD	nicotinate-nucleotide adenyltransferase	2.7.7.18	11311	2179	3946	3946
tdk	thymidine kinase	2.7.1.21			6692	6692

II.C. Açúcares e Nucleotídeos de Açúcares

II.D. Cofatores, Grupos Prostéticos e Transportadores

II.D.1. Biotina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
bioA	adenosylmethionine-8-amino-7-oxononanoate aminotransferase	2.6.1.62	1281	189	6355/ 789	6355/ 789
bioB	biotin synthase	2.8.1.6	371	64	7375	7375
bioC ubiE	biotin synthesis protein (Methyltransferase)	2.1.1.-	7181	2099	7365/ 10872	7365/ 10872
bioD	dethiobiotin synthetase	6.3.3.3	13641	2477	3167	3167
bioF	8-amino-7-oxononanoate synthase	2.3.1.47	5361	1357	7374	7374
bioH	biotin biosynthesis protein	3.1.1.1	5351	1356	7369	7369
bioI	cytochrome P450-like enzyme	1.14.-.-	15361	377	4456	4456
birA	bifunctional transcriptional repressor of the biotin operon/biotin acetyl-CoA-carboxylase synthetase	6.3.4.15	9751	1796	2795	2795
cypX bioI	cytochrome P-450 hydroxylase	1.14.-.-	15521	356	4456	4456

II.D.2. Ácido Fólico						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
dhfRIII	dihydrofolate reductase type III	1.5.1.3	12481	2331	2710	2710
folB	dihydroneopterin aldolase	4.1.2.25	14941	436	3595	3595
folC dedC	folylpolyglutamate synthase/dihydrofolate synthase	6.3.2.17	7871	1946	3376	3376
folD	bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase	3.5.4.9 1.5.1.5	13211	2431	6502	6502
folE	GTP cyclohydrolase I	3.5.4.16	7611	1983	6657	6657
folK	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	2.7.6.3	1561	228	5067	5067
folK	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	2.7.6.3	6151	1456	15208	15208
folP	dihydropteroate synthase	2.5.1.15	551	91	3509	3509
glyA	serine hydroxymethyltransferase	2.1.2.1	16011	946	5165	5165
lpxC	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	3.5.1.-	17061	803	5232	5232
metF	5,10-methylenetetrahydrofolate reductase	1.7.99.5	3711	1121	5239	5239
mutT	dNTP pyrophosphohydrolase	3.6.1.-	6031	1441	6815	6815
mutT/ thiE	bifunctional DGTP-pyrophosphohydrolase/thiamine phosphate synthase	3.6.1.- 2.5.1.3	3701	1120	5236	5236
pabA		4.1.3.-				
pabB		4.1.3.-				
pabC		4.-.-.-				
thyA	thymidylate synthase	2.1.1.45	12491	2332	2712	2712

II.D.3. Lipoato						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
lipA lip	lipoic acid synthetase	2.8.1.-	4741	1269	5040	5040
lipB	lipoate-protein ligase B	6.3.4.-	4751	1270	5039	5039

II.D.4. Molibdopterina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
moaA	molybdenum cofactor biosynthesis protein A	-			394	394
moaB	molybdopterin biosynthesis protein B	-			3234	3234
moaC	molybdenum cofactor biosynthesis protein C	-			398	398
moaD	molybdopterin-converting factor chain 1	-			399	399
moaE	molybdopterin-converting factor chain 2	-			400	400
mobA	molybdopterin guanine dinucleotide synthase	-			6875	6875
moeA	molybdopterin guanine dinucleotide synthase	-			6858/ 6874	6858/ 6874
moeB	molybdopterin biosynthesis protein	-	6961	1545	4792	4792
moeB	molybdopterin biosynthesis protein	-	14712	466	6859	6859

II.D.5. Pantotenato						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
	Pantothenase	3.5.1.22	13311	2443	6483	6483
coaA		2.7.1.33				
coaD kdtB	phosphopantetheine adenyltransferase	2.7.7.3	2451	980	4183	4183
coaE	dephospho-CoA kinase	2.7.1.24	17621	2536	2035	2035
dfp	DNA/pantothenate metabolism flavoprotein	6.3.2.5 4.1.1.36	1001	149	1152	1152
panB	3-methyl-2-oxobutanoate hydroxymethyltransferase	2.1.2.11	1571	229	3420	3420
panC	pantoate--beta-alanine ligase	6.3.2.1	1581	230	3417	3417
panD	aspartate 1-decarboxylase precursor	4.1.1.11	1591	231	3415	3415
panE		1.1.1.169				

II.D.6. Piridoxina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
dxr	1-deoxy-D-xylulose 5-phosphate reductoisomerase	1.1.1.-	2971	1048	4892	4892
pdxA	pyridoxal phosphate biosynthetic protein	-	16811	839	2682	2682
pdxB		1.1.1.-				
pdxH	pyridoxamine 5'-phosphate oxidase	1.4.3.5	5231	1337	6239	6239
pdxJ	pyridoxal phosphate biosynthetic protein	-	341	60	5411	5411
pdxK		2.7.1.49 2.7.1.- 2.7.1.35				
pdxY		2.7.1.35			4688	4688

II.D.7. Nucleotídeos de Piridina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
nadA	quinolinate synthetase A	-	8051	1923		
nadB	L-aspartate oxidase	1.4.3.16	8041	1924		
nadC	nicotinate-mononucleotide pyrophosphorylase	2.4.2.19	8031	1925	3795	3795
nadD	nicotinate-nucleotide adenyltransferase	2.7.7.18	11311	2179	3946	3946
nadE adgA	NH3-dependent NAD synthetase	6.3.5.1	7741	1961	10273	10273
pncA	pyrazinamidase/nicotinamidase	3.5.1.19	11971	2274		
pncB	nicotinate phosphoribosyltransferase	2.4.2.11	3521	1097	2229	2229

II.D.8. Tiamina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
apbE	thiamine biosynthesis lipoprotein ApbE precursor		14121	594	6117	6117
dxs	deoxyxylulose-5-phosphate synthase	2.2.1.1	11811	2249	4257	4257
mutT/ thiE	bifunctional DGTP-pyrophosphohydrolase/thiamine phosphate synthase	3.6.1.- 2.5.1.3	3701	1120	5236	5236
thiC	thiamine biosynthesis protein	-	8331	1888	2115	2115
thiD	phosphomethylpyrimidine kinase	2.7.4.7 2.7.1.49	13931	621	3473	3473
thiE	thiamin-phosphate pyrophosphorylase	2.5.1.3	15351	378	2167	2167
thiG	thiamine biosynthesis protein	-	17231	783	5527	5527
thiH		-				
thiK		2.7.1.89				
thiL	thiamine-monophosphate kinase	2.7.4.16	15941	956	5179	5179
thiM		2.7.1.50				

II.D.9. Riboflavina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
ribA	riboflavin biosynthesis protein	3.5.4.25	7551	1992	2938	2938
ribA/ ribB	GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase	3.5.4.25	15971	953	5174	5174
ribD ribG	riboflavin-specific deaminase	1.1.1.193 3.5.4.26	15991	950	5168	5168
ribE	riboflavin synthase alpha chain	2.5.1.9	15981	952	5173	5173
ribF	Riboflavin kinase / FAD synthetase	2.7.7.2 2.7.1.26	13101	2419	5844	5844
ribH	6,7-dimethyl-8-ribityllumazine synthase/riboflavin synthase beta chain	2.5.1.9	15961	954	5176	5176

II.D.10. Tiorredoxina, Glutarredoxin e Glutatião						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
ggt	gamma-glutamyltranspeptidase	2.3.2.2	2481	984	1784/ 4179	1784/ 4179
grx	glutaredoxin-like protein	-	12891	2394	2270	2270
grxC	glutaredoxin	-	18121	2595	3401	3401
gshA gsh1	glutamate-cysteine ligase precursor	6.3.2.2	5961	1428	5136	5136
gshB	glutathione synthetase	6.3.2.3	7781	1956	4580	4580
gst	glutathione S-transferase	2.5.1.18	4341	1210	3338	3338
trx	thioredoxin	-	11261	2174	3953	3953
trxA	thioredoxin	-	4261	1199	2890	2890
trxA	thioredoxin	-	18801	2698	743	743
trxB	thioredoxin reductase	1.6.4.5	6081	1448	6826	6826
yneN	thioredoxin	-	7561	1990	3227	3227

II.D.11. Menaquinona, Ubiquinona						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
aroQ	catabolic dehydroquinase	4.2.1.10	291	47	7154	7154
coq7	ubiquinone biosynthesis protein	-	6911	1538	11926	11926
dxs	deoxyxylulose-5-phosphate synthase	2.2.1.1	11811	2249	4257	4257
ispA	geranyltranstransferase (farnesyl-diphosphate synthase)	2.5.1.10	13751	661	3920	3920
ispB	octaprenyl-diphosphate synthase	2.5.1.-	5631	1391	6057	6057
ubiA	hydroxybenzoate octaprenyltransferase	2.5.1.-	391	68	7379	7379
ubiB aarF	ubiquinone biosynthesis protein	-	9391	1833	7547	7547
ubiC		4.-.-.-				
ubiD		4.1.1.-				
ubiE	ubiquinone/menaquinone transferase	2.1.1.-	6471	1487	11357	11357
ubiF	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinone hydroxylase	1.14.13.-	16861	834	2667	2667
ubiG	3-demethylubiquinine 3-methyltransferase / 2-octaprenyl-6-hydroxyphenol methylase	2.1.1.64	13581	2471	6373	6373
ubiG	3-demethylubiquinine 3-methyltransferase / 2-octaprenyl-6-hydroxyphenol methylase	2.1.1.64	5661	1397	10877	10877
ubiH	2-octaprenyl-6-methoxyphenol hydroxylase	1.14.13.-	16851	835	2669	2669
ubiX		4.1.1.-				

II.D.12. Proto e Siroheme						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
cyoE	cytochrome c oxidase assembly factor	2.5.1.-	5381	1360	3608	3608
cysG	siroheme synthase	2.1.1.107 4.99.1.-	16881	832	25/ 7265	25/ 7265
gltX	glutamyl-tRNA synthetase	6.1.1.17	16951	822	1745	1745
hemA	glutamyl-tRNA reductase	1.2.1.-	18521	2648	3239	3239
hemB	delta-aminolevulinic acid dehydratase	4.2.1.24	12181	2306	2836	2836
hemC	hydroxymethylbilane synthase	4.3.1.8	10501	1627	11380	11380
hemD	uroporphyrinogen-III synthase	4.2.1.75	9721	1799	7596	7596
hemE	uroporphyrinogen decarboxylase	4.1.1.37	5201	1332	6249	6249
hemF	coproporphyrinogen III oxidase, aerobic	1.3.3.3	131	17	4131	4131
hemG		1.3.3.4				
hemH	ferrochelatase	4.99.1.1	14291	566	1420	1420
hemK	protoporphyrinogen oxidase		6691	1512	2609	2609
hemL	glutamate-1-semialdehyde 2,1-aminomutase	5.4.3.8	12211	2302	2156/ 25153/ 25154	2156
hemN	coproporphyrinogen III oxidase	1.-.-.-	6651	1507	1632	1632
hemY	porphyrin biosynthesis protein		9741	1797	7599	7599

II.D.13. Proteína Transportadora de Carboxil Biotina (BCCP)						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
accB	biotin carboxyl carrier protein	6.4.1.2	301	48	7155	7155

II.D.14. Cobalamina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
btuR	cob(I)alamin adenolsyltransferase	2.5.1.17			4418	4418
cobB	cobirinic acid a,c-diamide synthase	-				20316
cobC	cobalamin biosynthetic protein	3.1.3.-			25791	
cobS	cobalamin synthase	-			4426	4426
cobT	nicotinate-nucleotide-dimethylbenzimidazole phosphoribisyltransferase	2.4.2.21			4424	4424
cobU		-			4423	4423
pgmA	phosphoglycerate mutase	5.4.2.1	8351	1886	3374	3374

II.D.15. Enterobactina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
entA						
entB						
entC						
entD						
entE						
entF	ATP-dependent serine activating enzyme	-			1163	1163

II.D.16. Biopterina						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
gcpE	gcpE protein		17981	2575	1106	1106
ispD	4-diphosphocytidyl-2c-methyl-d-erythritol synthase	2.7.7.-	4871	1293	3526	3526
ispE	isopentenyl monophosphate kinase	2.7.1.-	18491	2645	3244	3244
ispF	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase		4881	1294	3526	3526
ptr1 ltdH	pteridine reductase 1	1.1.1.253	6161	1457	5070	5070
ygcM	6-pyruvoyl tetrahydrobiopterin synthase	4.6.1.10	1311	193	3163	3163

II.D.17. Outros						
Gene	**Produto**	**EC#**	**XPD**	**XCVC**	**Xcitri**	**Xcamp**
dxr	1-deoxy-D-xylulose 5-phosphate reductoisomerase	1.1.1.-	2971	1048	4892	4892

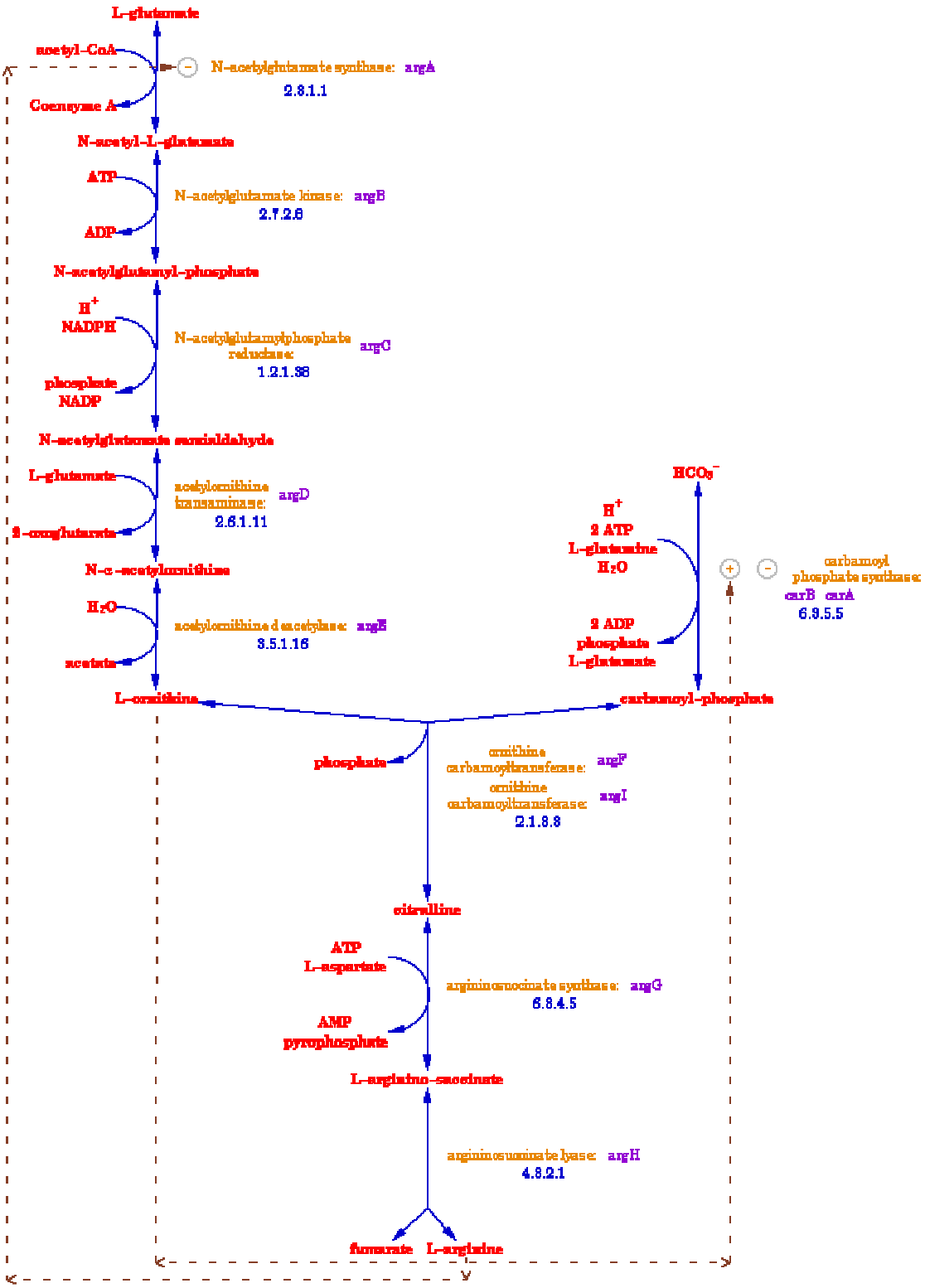
II.E. Biossíntese de Ácidos Graxos e Fosfatídicos

Gene	Produto	EC#	XPD	XCVC	Xcitri	Xcamp
	fatty acyl-CoA synthetase		17281		4159	4159
accA	acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	6.4.1.2	1371	203	1215	1215
accB	acyl carrier protein	6.4.1.2	301	48	7155	7155
accC	biotin carboxylase	6.3.4.14	311	49	7158	7158
accD	Acetyl-CoA carboxylase beta subunit	6.4.1.2	6251	1467	3844	3844
acpD	acyl carrier protein phosphodiesterase	3.1.4.14			25689	
acpE		2.7.8.7				
acpP	acyl carrier protein	1.6.99.3	13701	672	5674	5674
acs	acetyl coenzyme A synthetase	6.2.1.1	11851	2255	3998	3998
cdh		3.6.1.26				
cdsA	phosphatidate cytidylyltransferase	2.7.7.41	2981	1049	4890	4890
cls	cardiolipin synthase	2.7.8.-	3491	1087	7633	7633
cls	cardiolipin synthase	2.7.8.-	4331	1209	4041	4041
dgkA	diacylglycerol kinase	2.7.1.107	12511	2334	2717	2717
drb0080	short-chain alcohol dehydrogenase	1.1.1.-	11931	2269	7310	7310
fabA	beta-hydroxydecanoyl-ACP dehydratase	4.2.1.60	14251	572	217	217
fabB	3-oxoacyl-(ACP) synthase	2.3.1.41	10421	1639	5676	5676
fabB	3-oxoacyl-(ACP) synthase	2.3.1.41	13691	673	4172	4172
fabD	malonyl CoA-ACP transacylase	2.3.1.39	13721	670	5670	5670
fabG	3-oxoacyl-[ACP] reductase	1.1.1.100	13711	671	5672	5672
fabG	3-oxoacyl-[ACP] reductase	1.1.1.100	1151	173	4164	4164
fabG	3-oxoacyl-[ACP] reductase	1.1.1.100	2321	319	2293	2293
fabH	3-oxoacyl-[ACP] synthase III	2.3.1.41	7691	1970	7564	7564
fabH	3-oxoacyl-[ACP] synthase III	2.3.1.41	9531	1817	5664	5664
fabI		1.3.1.9				
fabZ	(3r)-hydroxymyristoyl ACP dehydrase	4.2.1.60	2931	1044	4898	4898
psd	phosphatidylserine decarboxylase	4.1.1.65	5431	1365	3867	3867
tesA	acyl-CoA thioesterase I	3.1.2.-			2726	2726
tesB	acyl-CoA thioesterase II	3.1.2.-	2801	1021	5860	5860

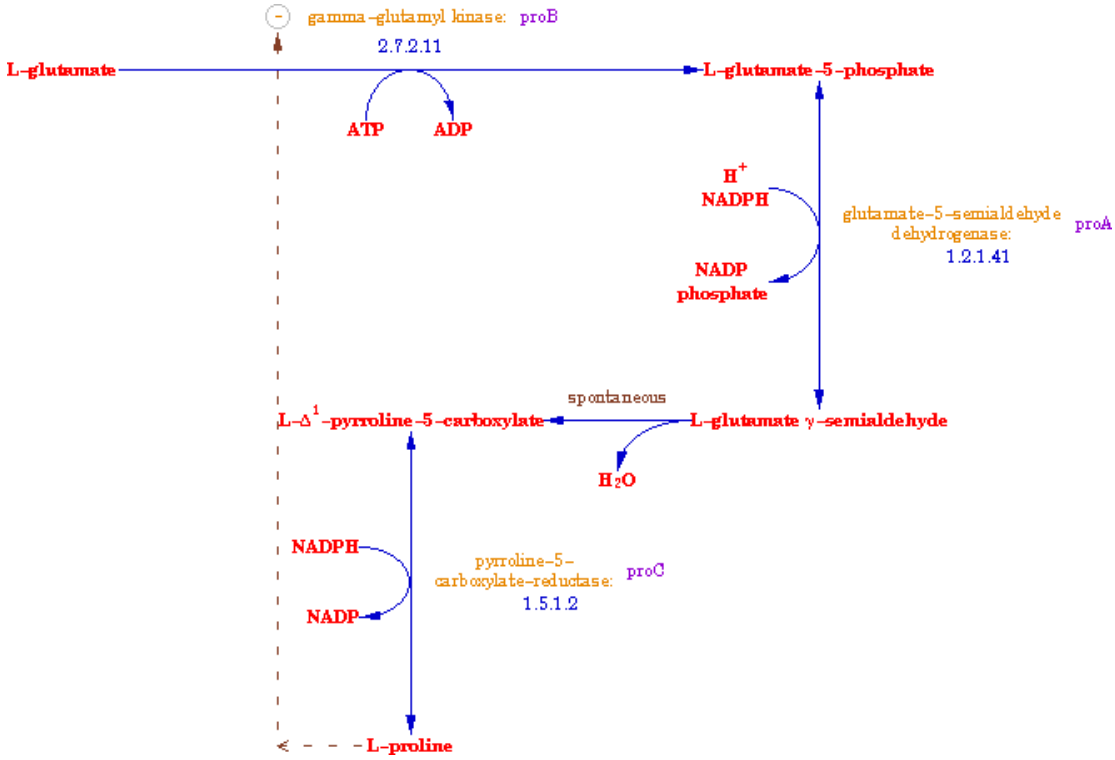
II.F. Poliaminas

Gene	Produto	EC#	XPD	XCVC	Xcitri	Xcamp
speA	biosynthetic arginine decarboxylase	4.1.1.19	941	144	1168	1168
speB		3.5.3.11				
speC		4.1.1.17				
speD	S-adenosyl methionine decarboxylase proenzyme	4.1.1.50	6921	1539	11455	11455
speE	spermidine synthase	2.5.1.16	931	143	1169	1169
speF		4.1.1.17				
speG		2.3.1.57 2.3.1.-				

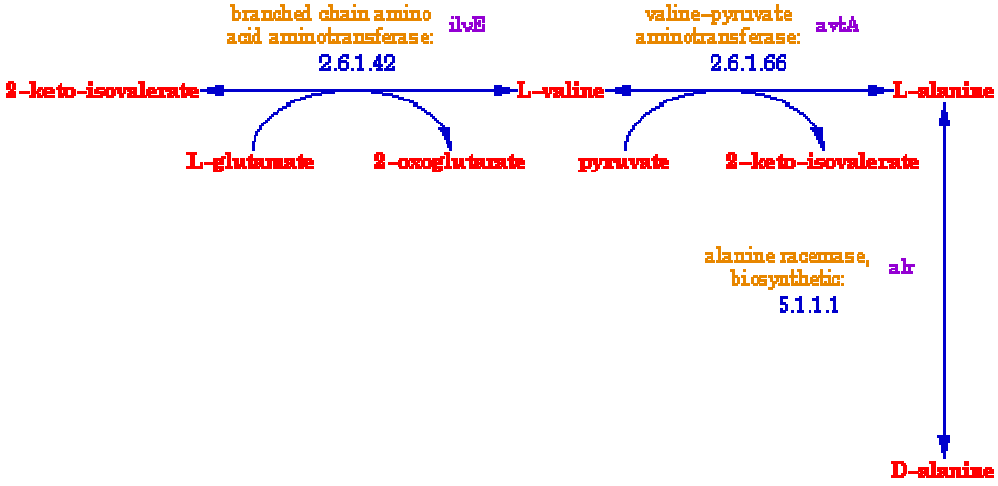
II.A. Aminoácidos / II.A.1 Família do Glutamato / Arginina



II.A. Aminoácidos / II.A.1 Família do Glutamato / Prolina

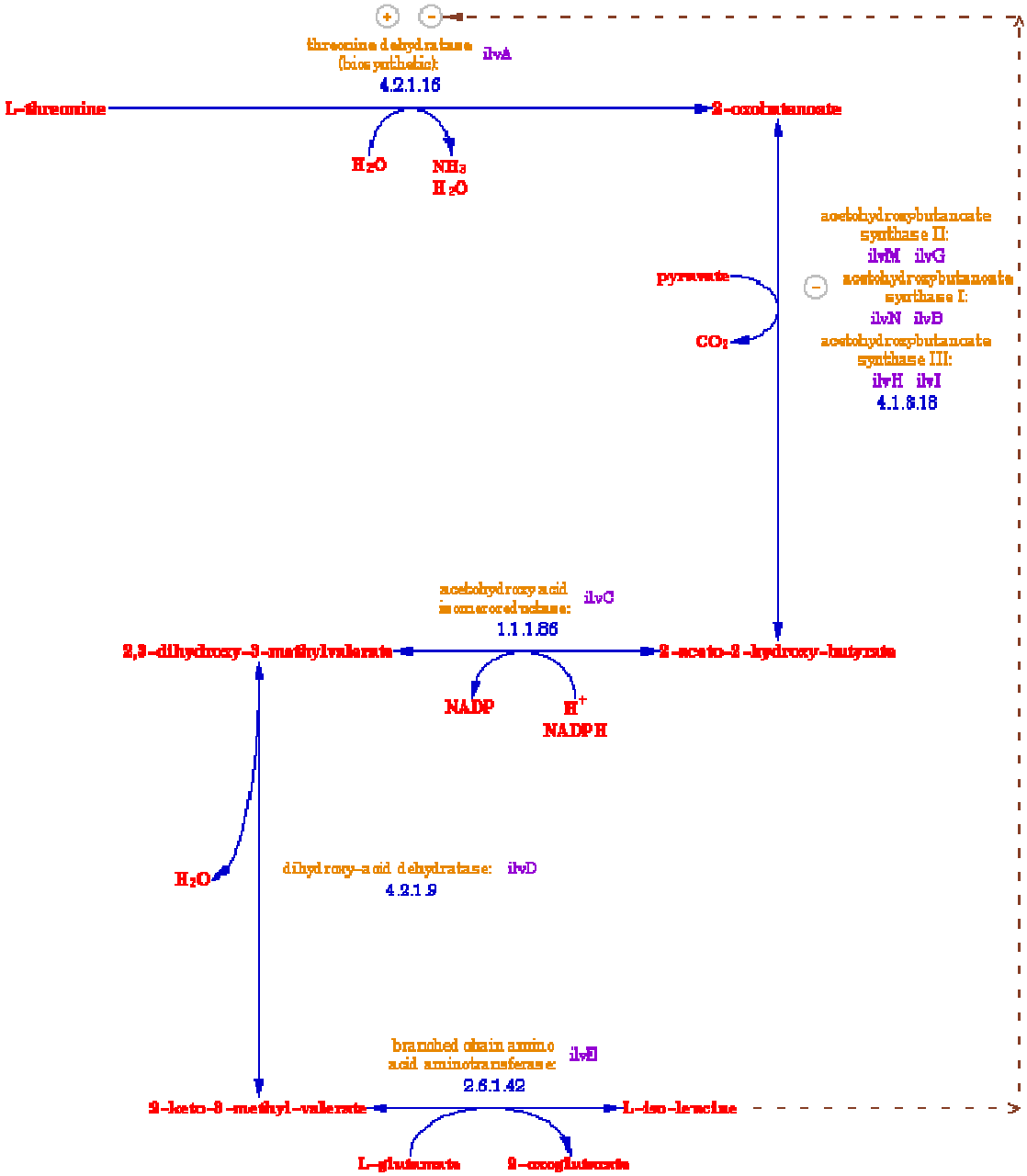


II.A. Aminoácidos / II.A.2 Família do Aspartato Família do Piruvato / Alanina

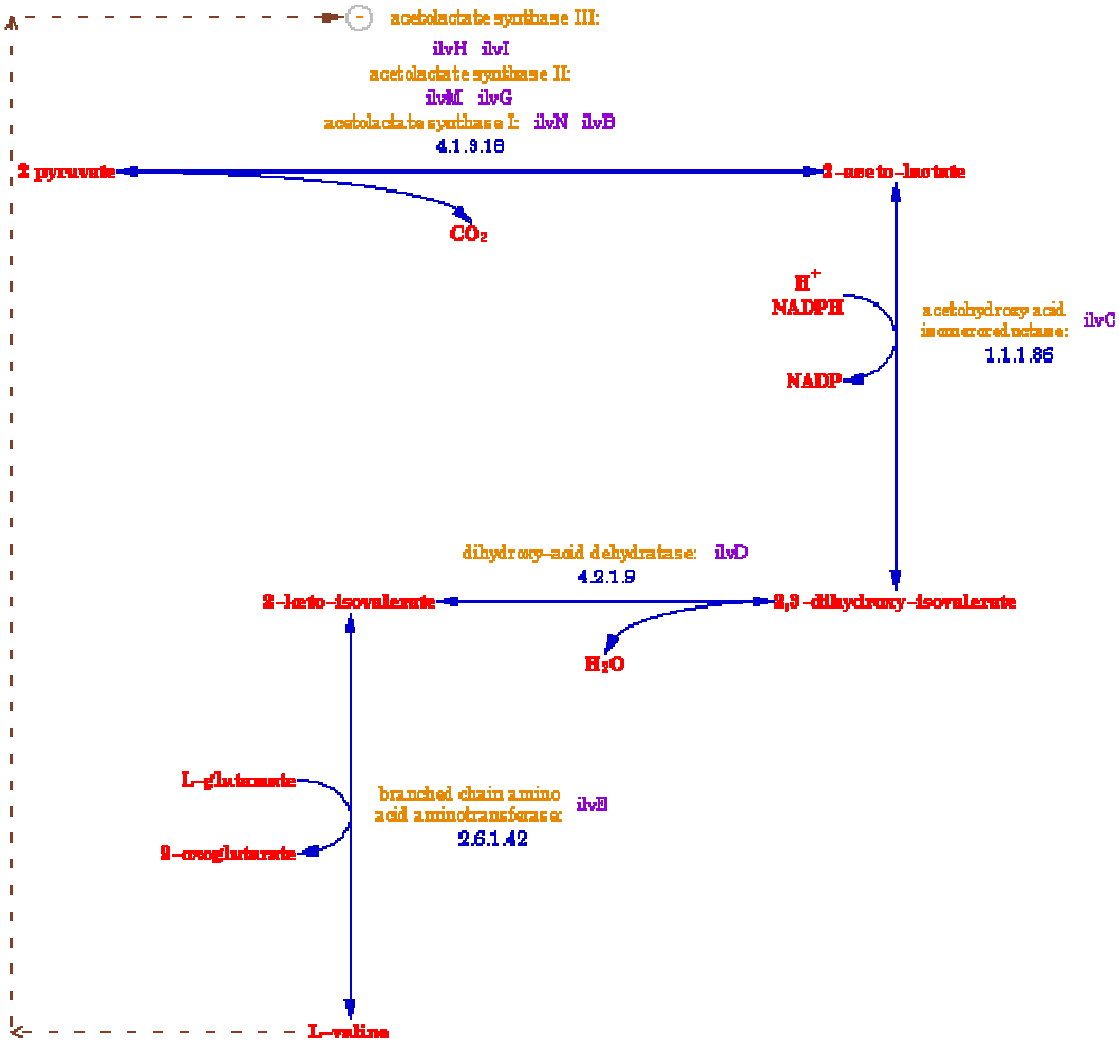


II.A. Aminoácidos / II.A.2 Família do Aspartato Família do Piruvato / Isoleucina/Valina

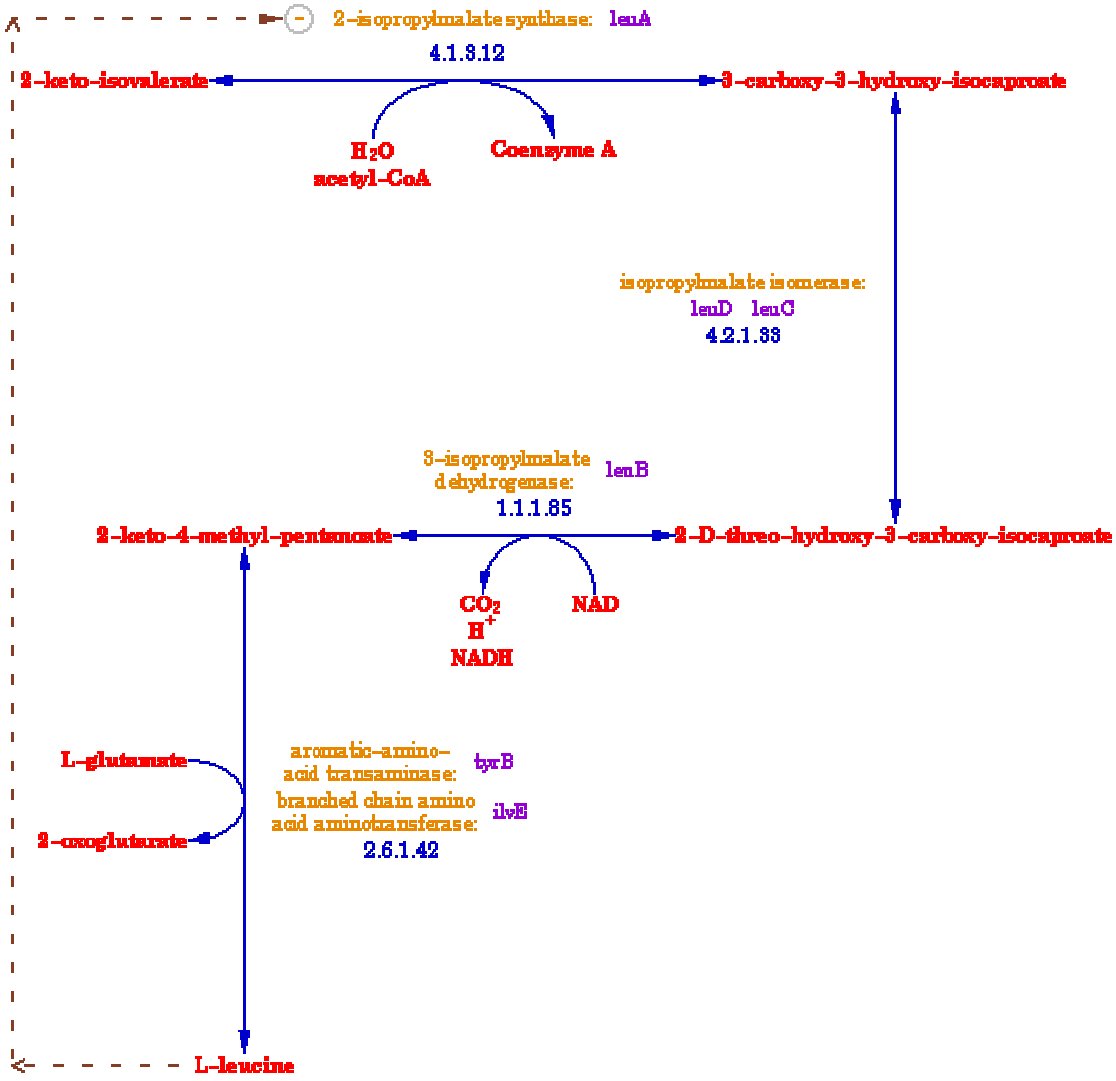
Isoleucina



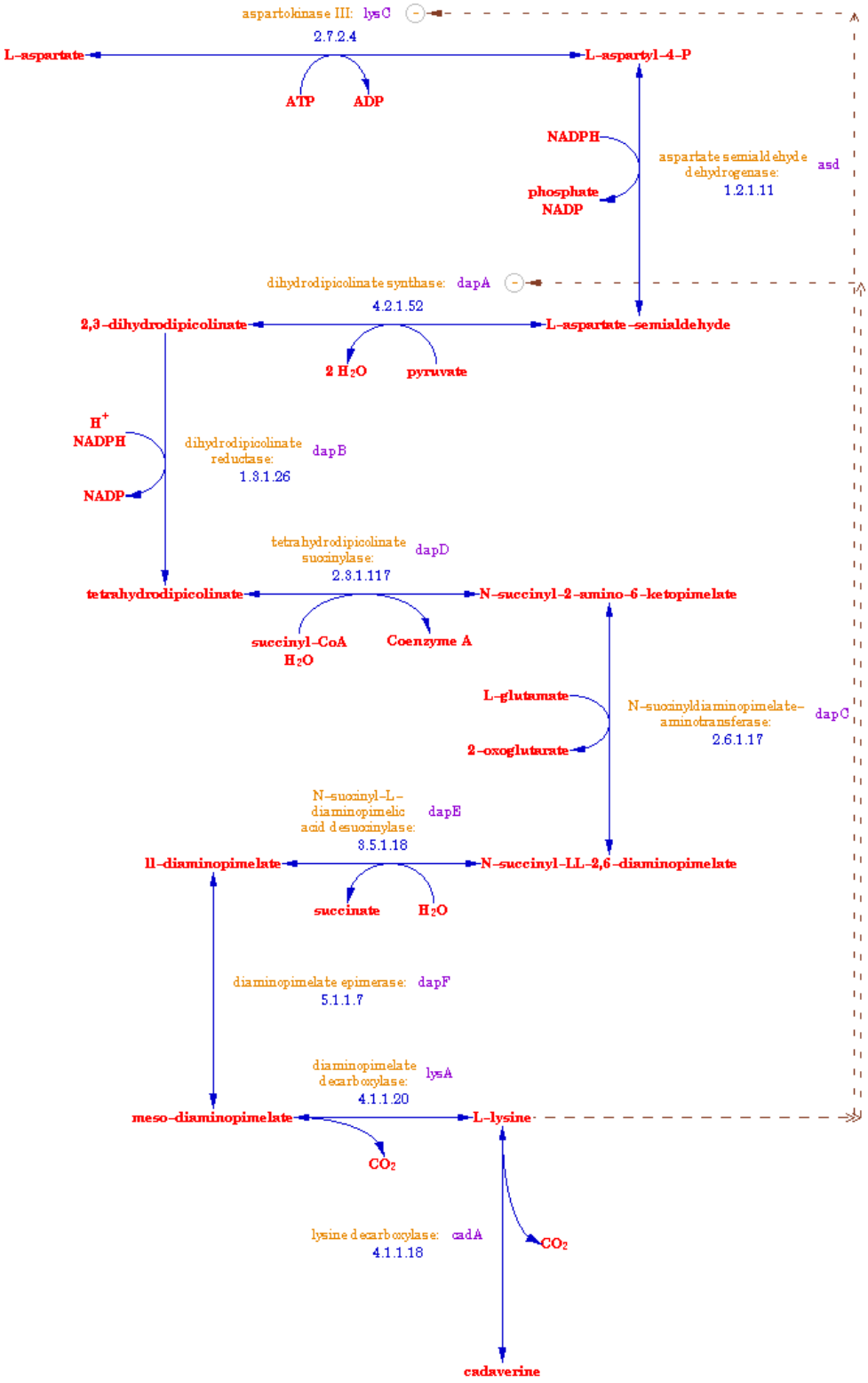
Valina


acetolactate synthase III:
ilvH ilvI
acetolactate synthase II:
ilvM ilvG
acetolactate synthase I: ilvN ilvB
4.1.3.18
2 pyruvate
2-aceto-lactate
CO2
H+
NADPH
NADP
acetohydroxy acid isomeroreductase: ilvC
1.1.1.86
dihydroxy-acid dehydratase: ilvD
4.2.1.9
2-keto-isovalerate
2,3-dihydroxy-isovalerate
H2O
L-glutamate
2-oxoglutarate
branched chain amino acid aminotransferase: ilvE
2.6.1.42
L-valine

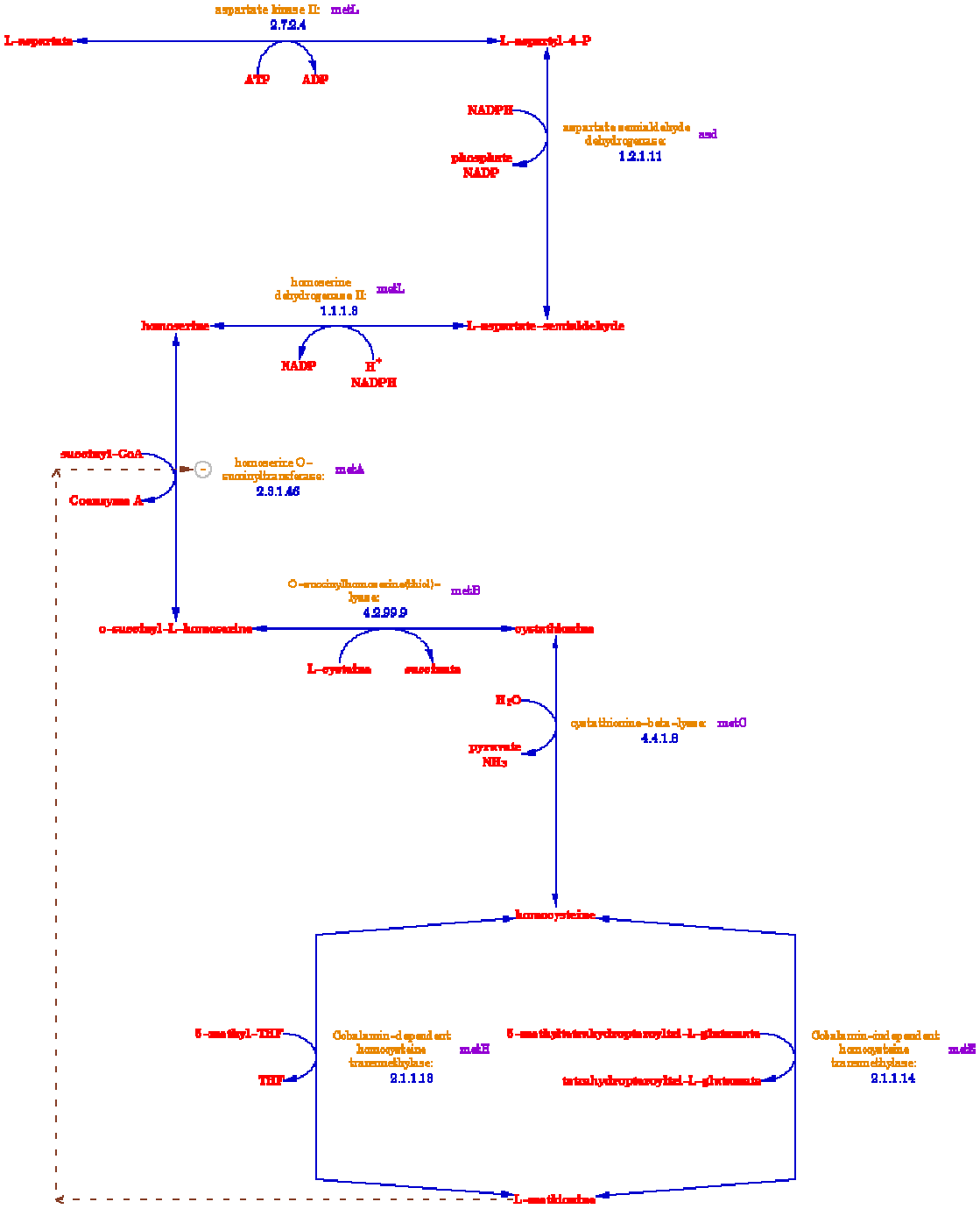
II.A. Aminoácidos / II.A.2 Família do Aspartato Família do Piruvato / Leucina



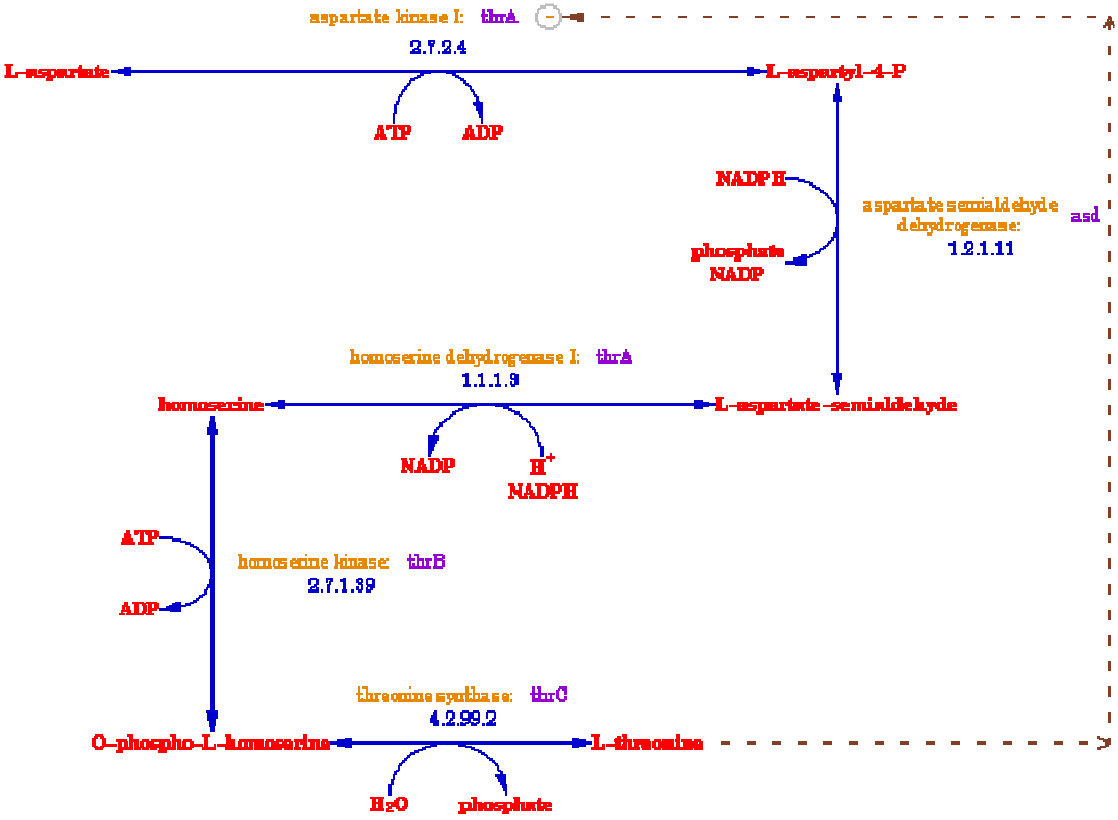
II.A. Aminoácidos / II.A.2 Família do Aspartato Família do Piruvato / Lisina



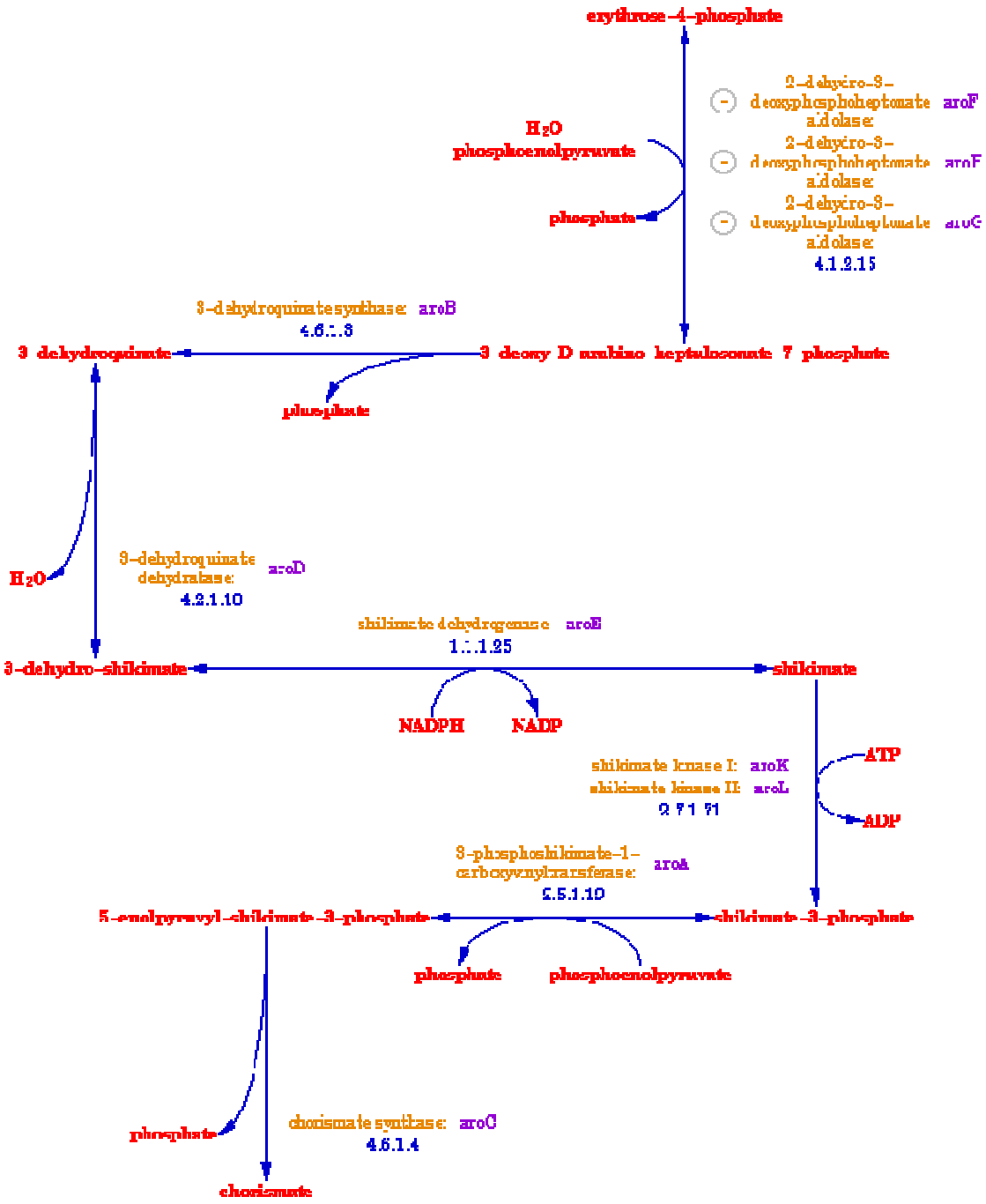
II.A. Aminoácidos / II.A.2 Família do Aspartato Família do Piruvato / Metionina



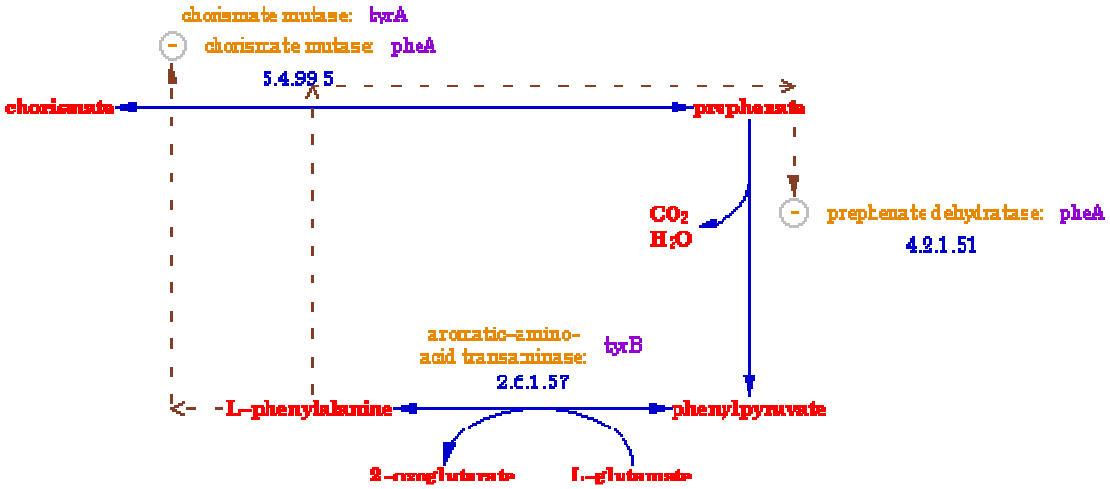
II.A. Aminoácidos / II.A.2 Família do Aspartato Família do Piruvato / Treonina



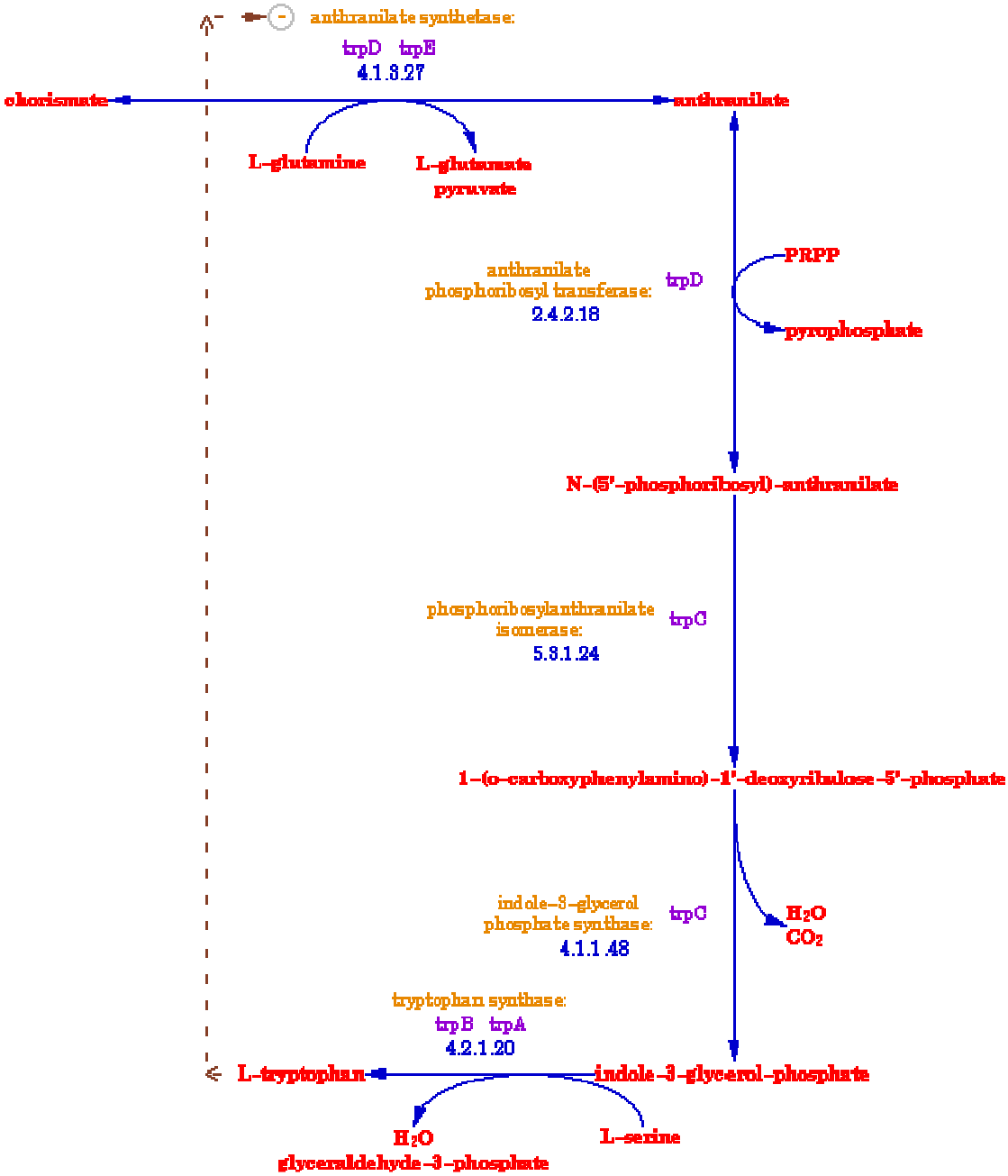
II.A. Aminoácidos / II.A.4 Família dos Aminoácidos Aromáticos / Corismato



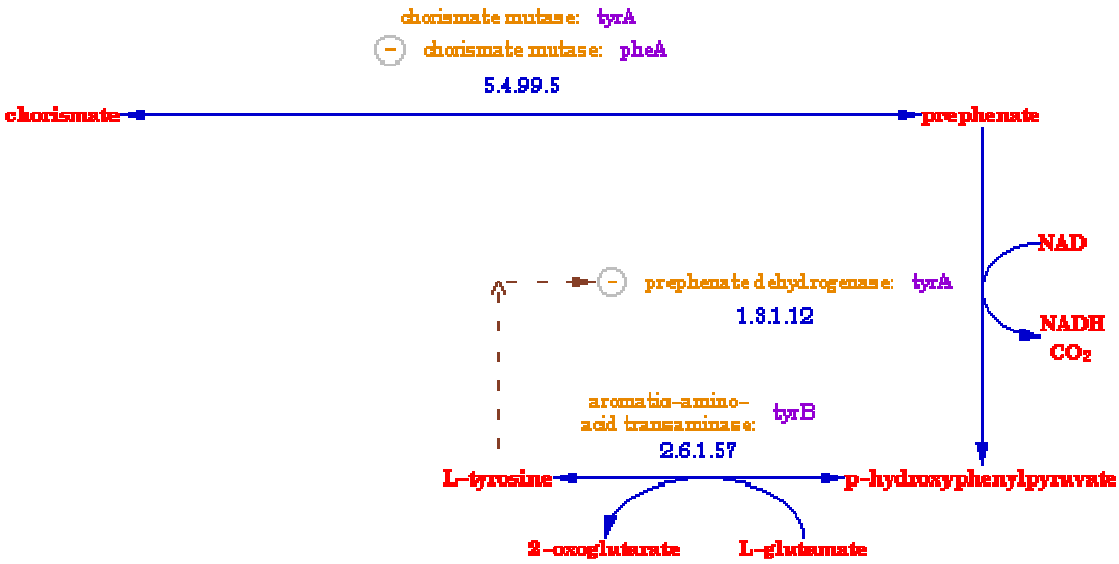
II.A. Aminoácidos / II.A.4 Família dos Aminoácidos Aromáticos / Fenilalanina



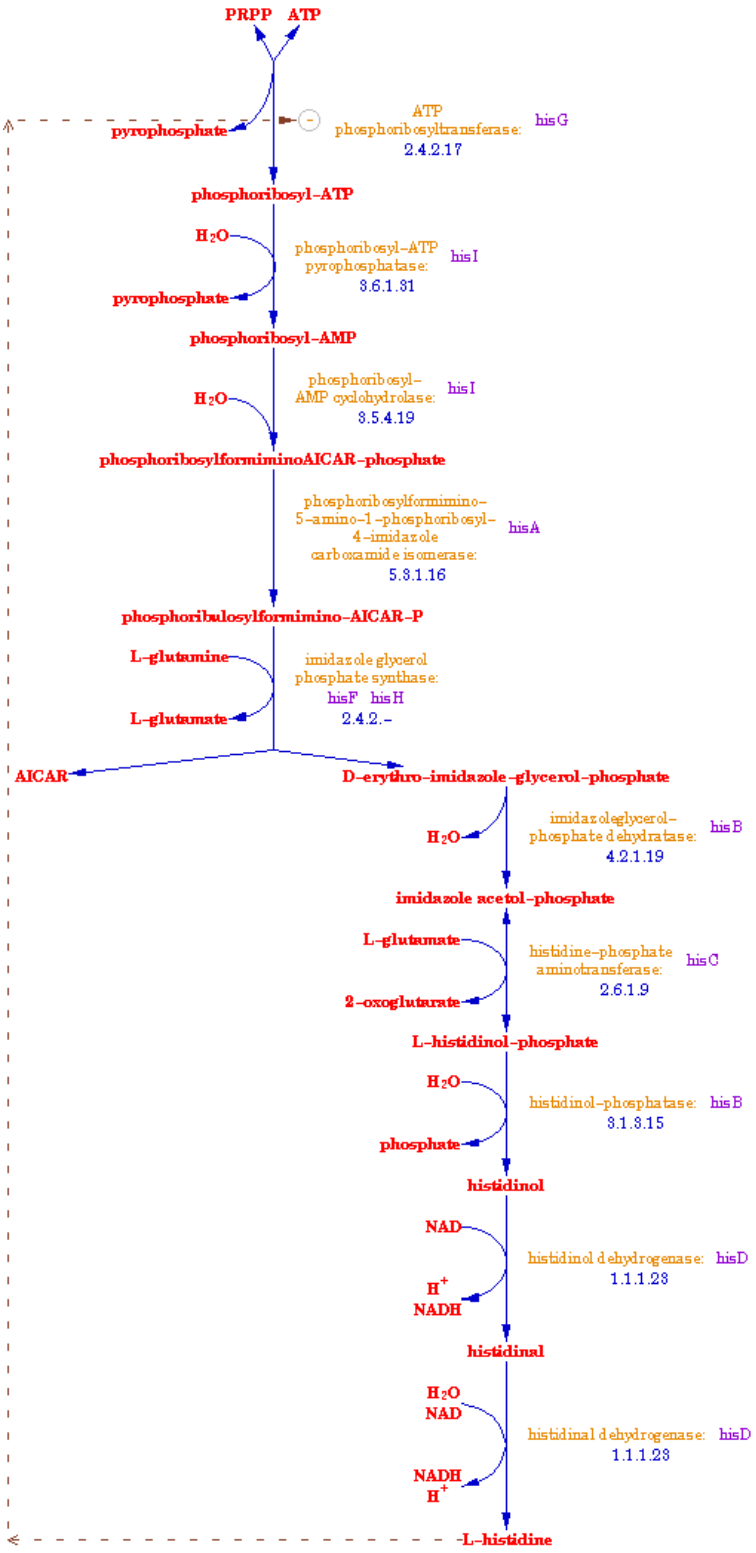
II.A. Aminoácidos / II.A.4 Família dos Aminoácidos Aromáticos / Triptofano



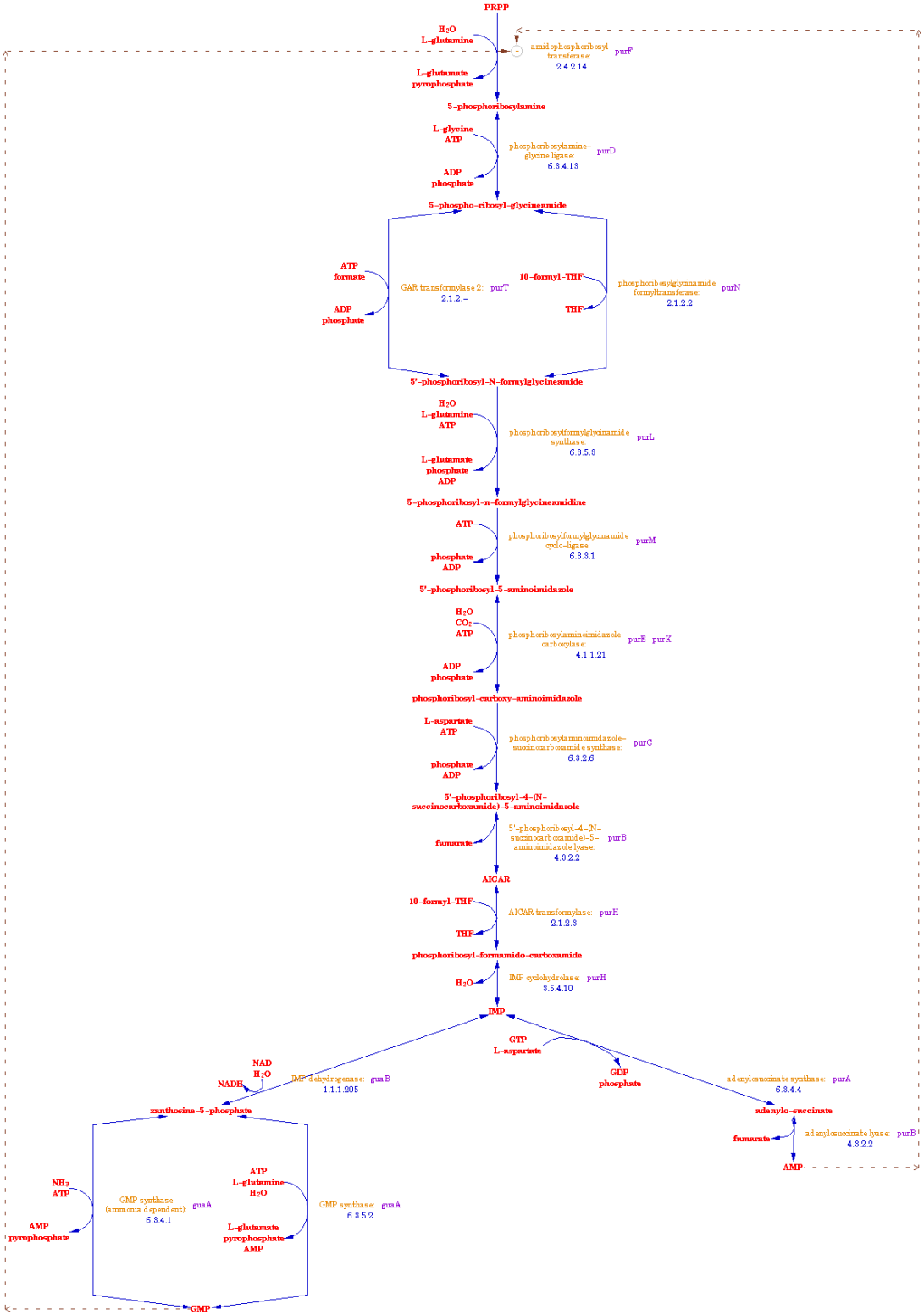
II.A. Aminoácidos / II.A.4 Família dos Aminoácidos Aromáticos / Tirosina



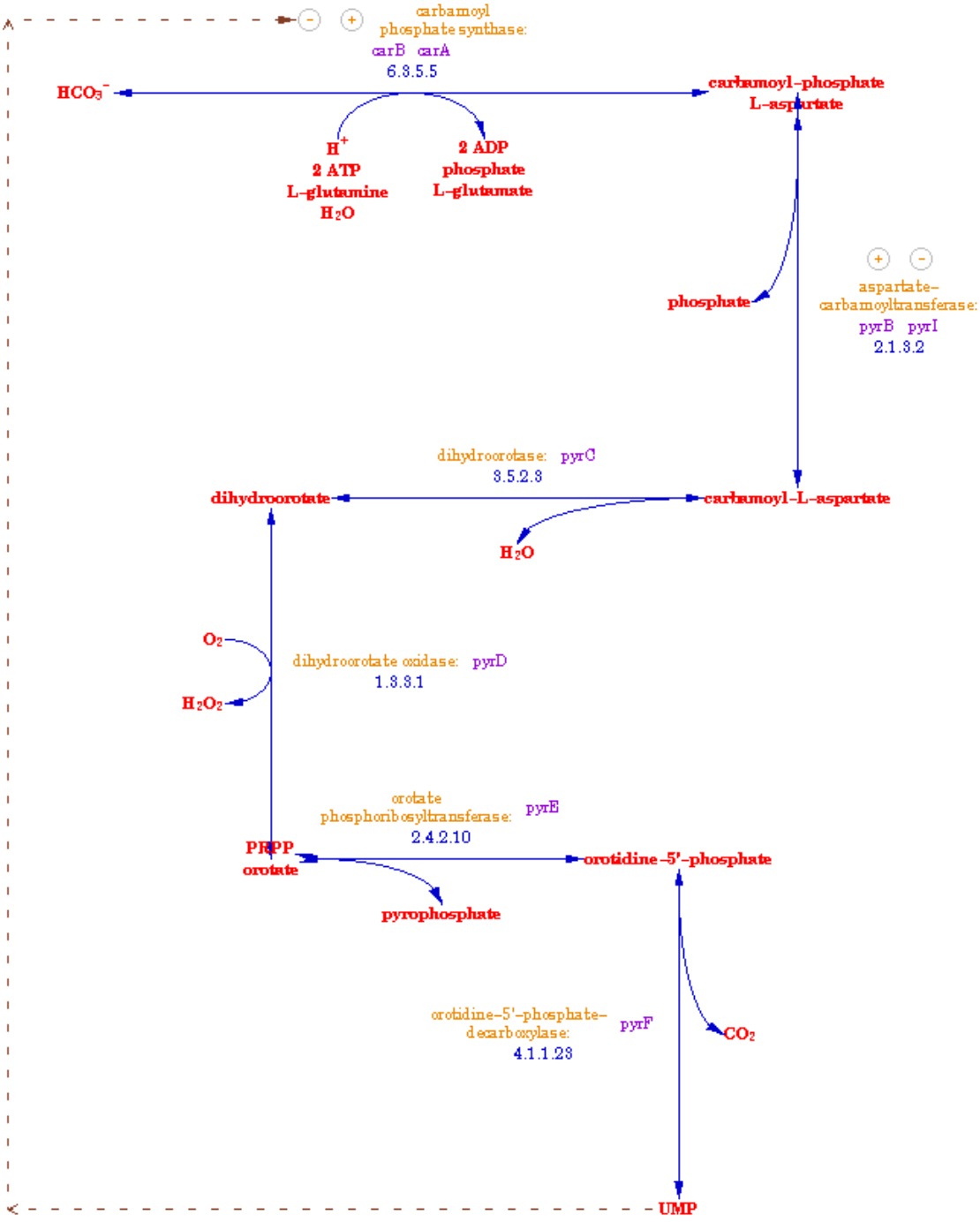
II.A. Aminoácidos / II.A.5 Histidina

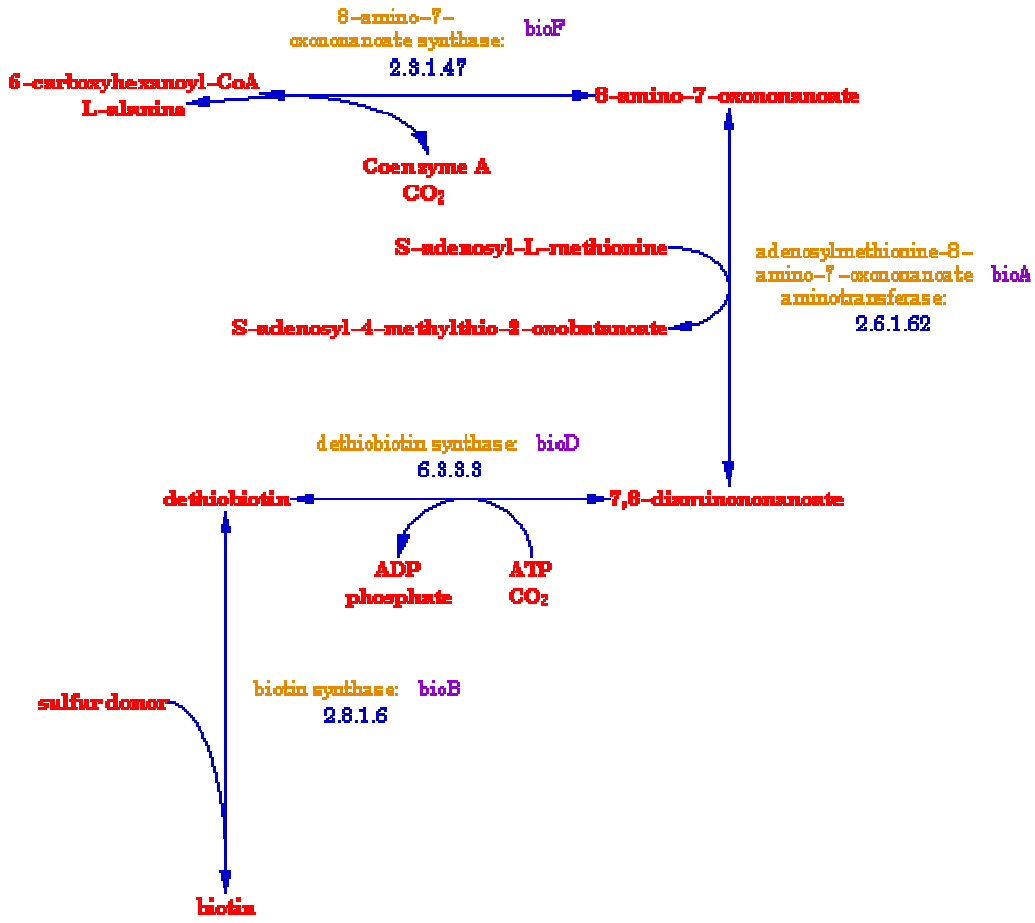


II.B Nucleotídeos / II.B.1 Ribunucleotídeos de Purina

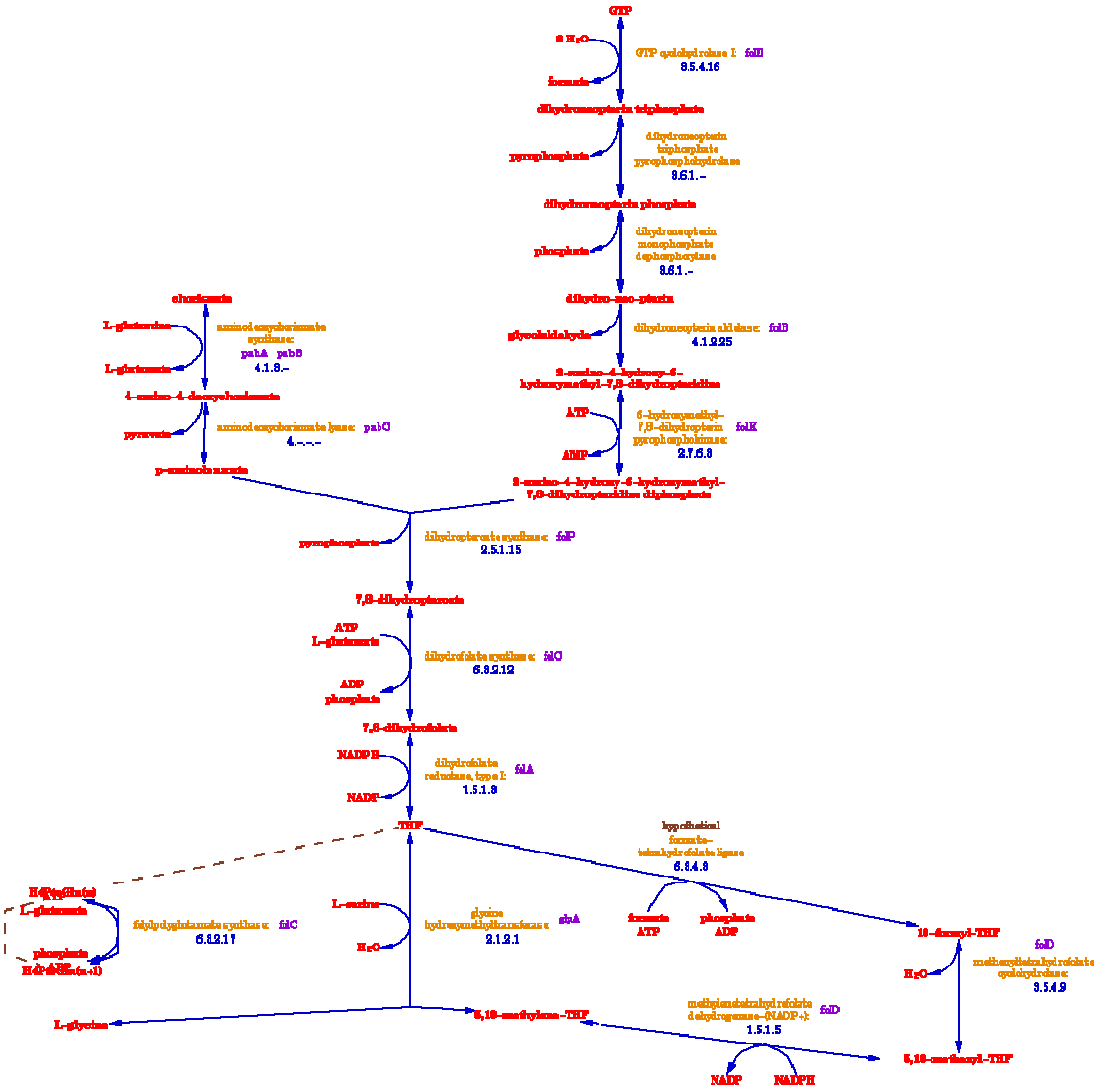


II.B Nucleotídeos / II.B.2 Ribunucleotídeos de Pirimidina

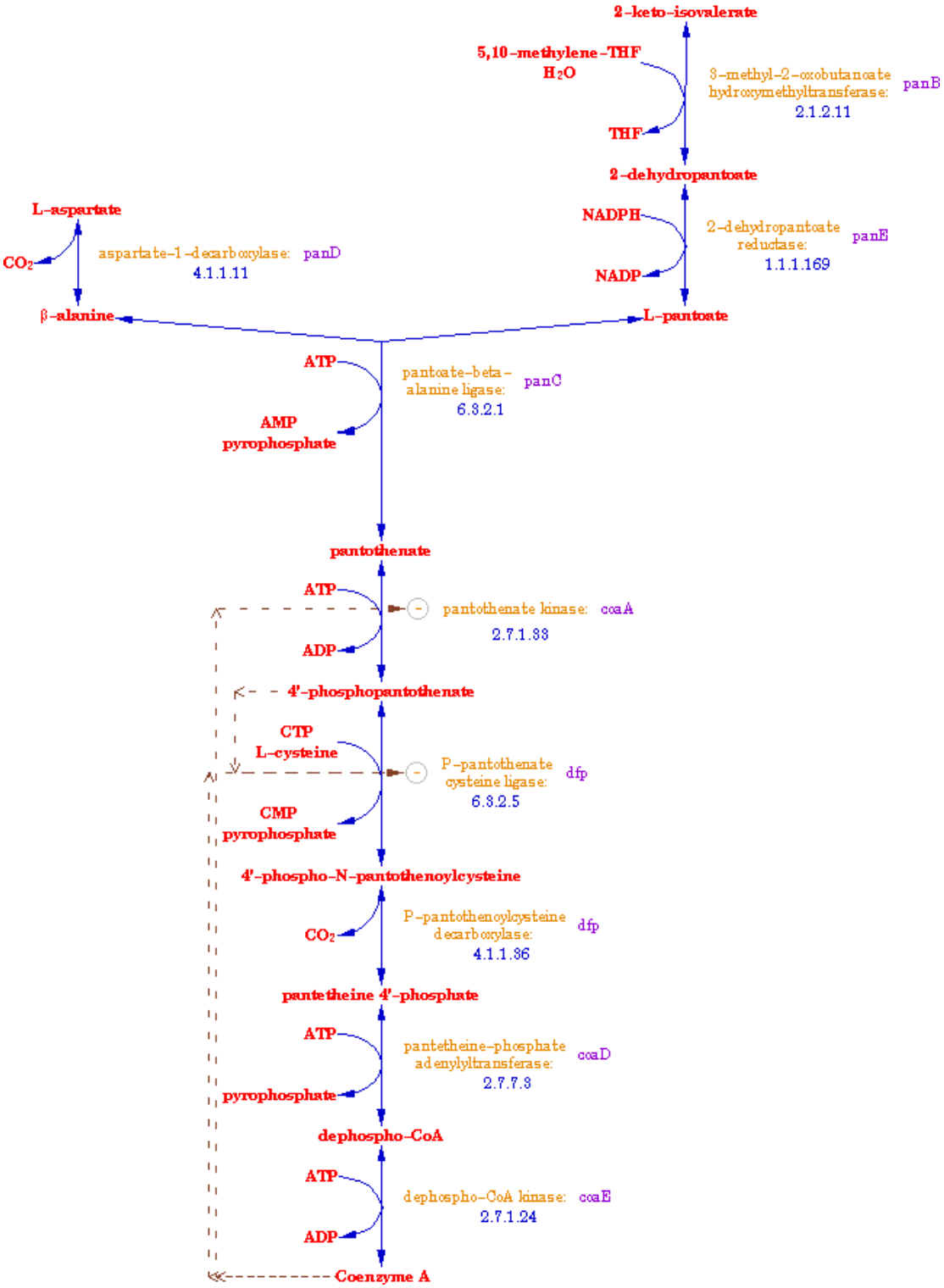


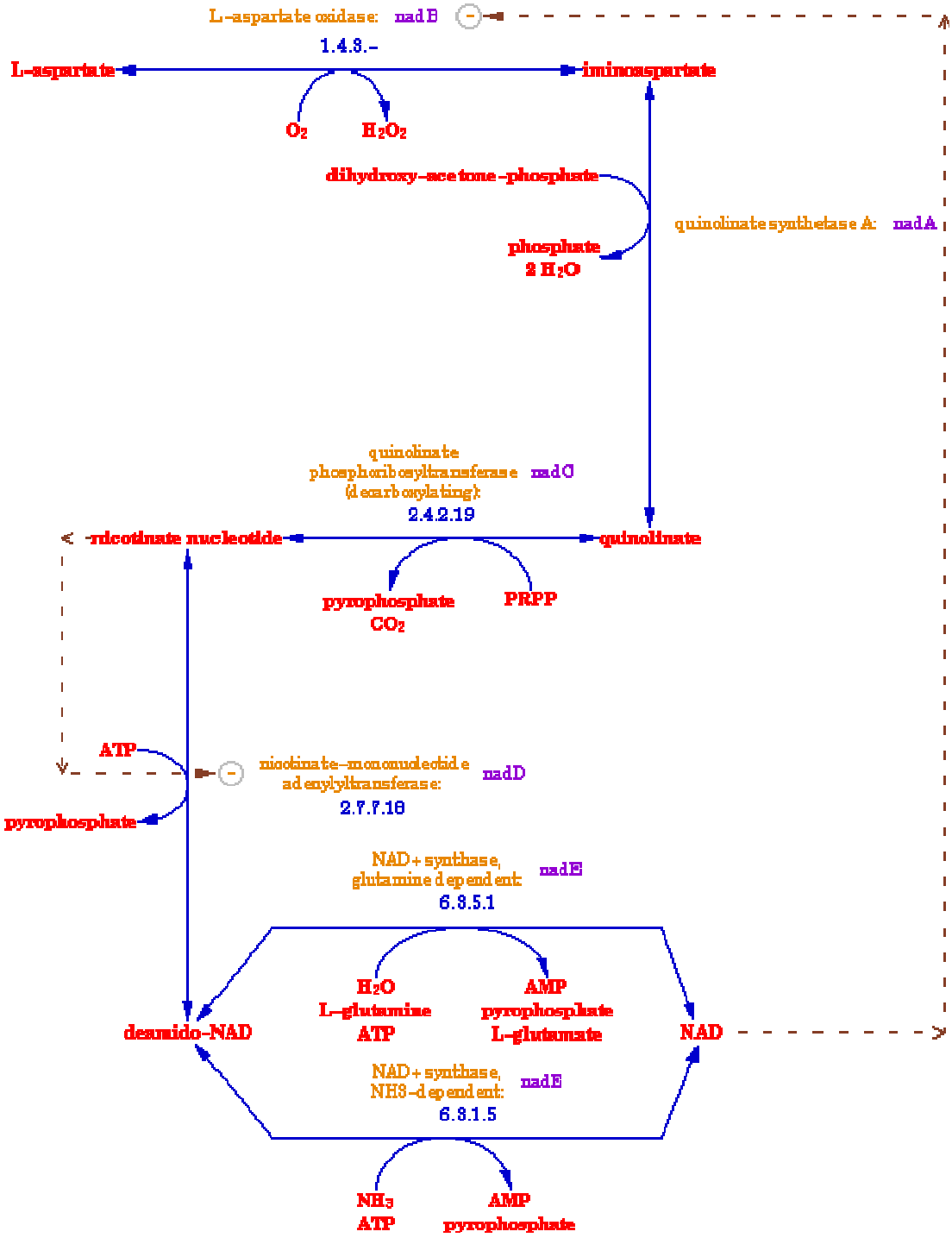

8-amino-7-oxononanoate synthase:
2.3.1.47
bioF
6-carboxyhexanoyl-CoA
L-alanine
8-amino-7-oxononanoate
Coenzyme A
CO2
S-adenosyl-L-methionine
S-adenosyl-4-methylthio-2-oxobutanoate
adenosylmethionine-8-amino-7-oxononanoate aminotransferase:
2.6.1.62
bioA
dethiobiotin synthase:
6.3.3.3
bioD
dethiobiotin
7,8-diaminononanoate
ADP
phosphate
ATP
CO2
biotin synthase:
2.8.1.6
bioB
sulfur donor
biotin

II.D Cofatores, Grupos Prostéticos e Transportadores / II.D.2 Ácido Fólico

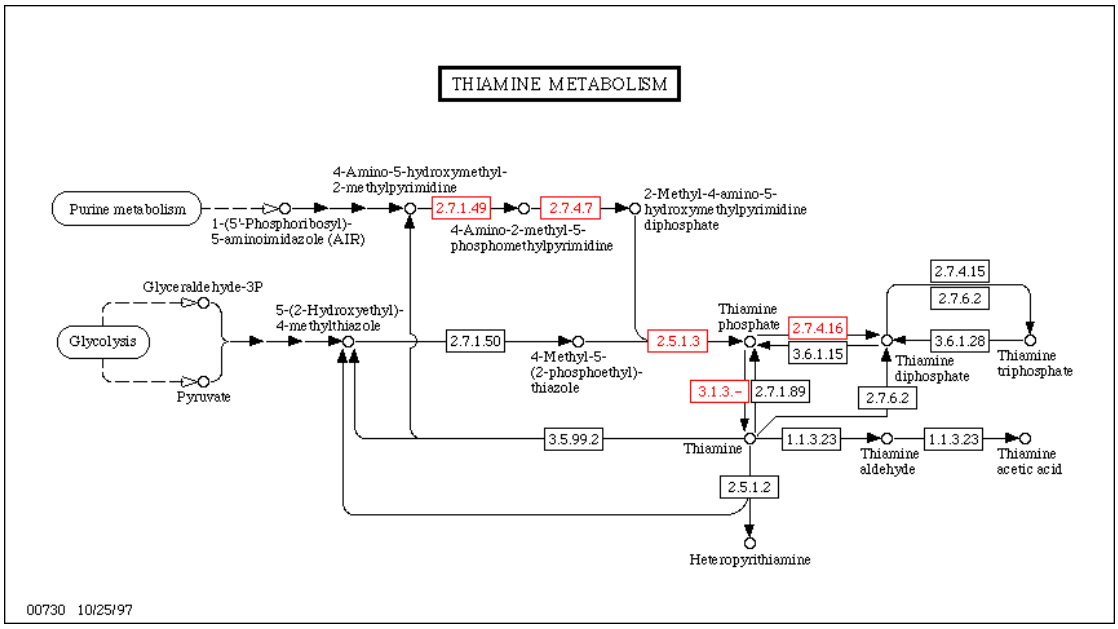


II.D Cofatores, Grupos Prostéticos e Transportadores / II.D.5 Pantotenato

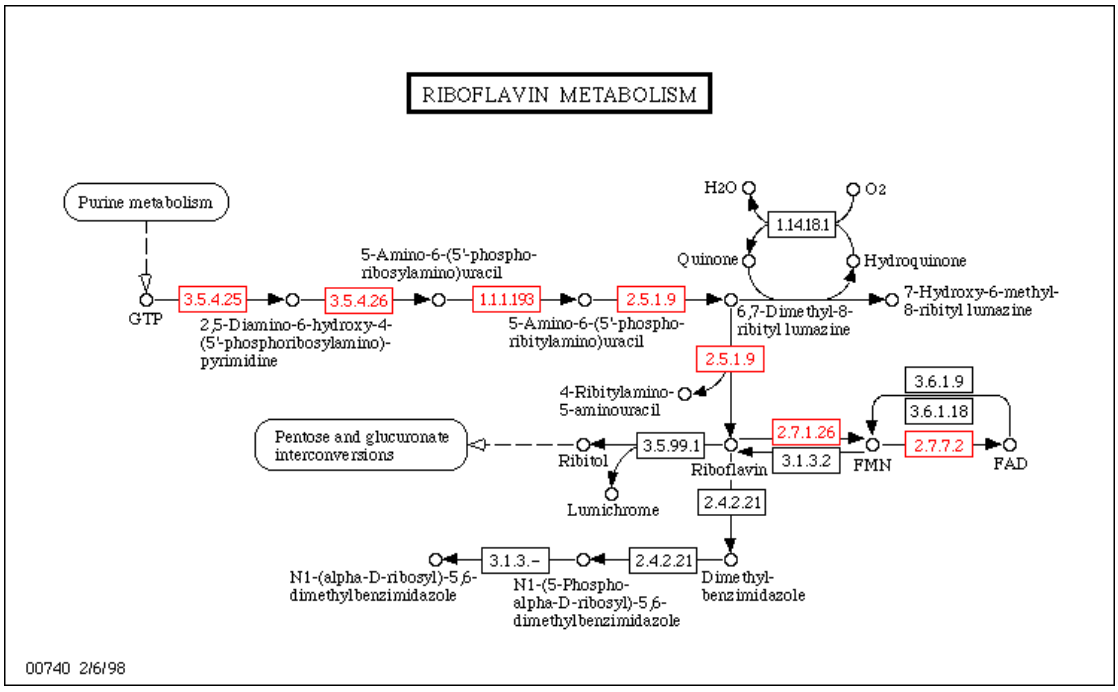



L-aspartate oxidase: nadB
1.4.3.-
L-aspartate
iminoaspartate
O2
H2O2
dihydroxy-acetone-phosphate
phosphate
2 H2O
quinolinate synthetase A: nadA
quinolinate phosphoribosyltransferase (decarboxylating): nadC
2.4.2.19
nicotinate nucleotide
quinolinate
pyrophosphate
CO2
PRPP
ATP
pyrophosphate
nicotinate-mononucleotide adenylyltransferase: nadD
2.7.7.18
NAD+ synthase, glutamine dependent: nadE
6.3.5.1
H2O
L-glutamine
ATP
AMP
pyrophosphate
L-glutamate
deamido-NAD
NAD
NAD+ synthase, NH3-dependent: nadE
6.3.1.5
NH3
ATP
AMP
pyrophosphate

II.D Cofatores, Grupos Prostéticos e Transportadores / II.D.8 Tiamina

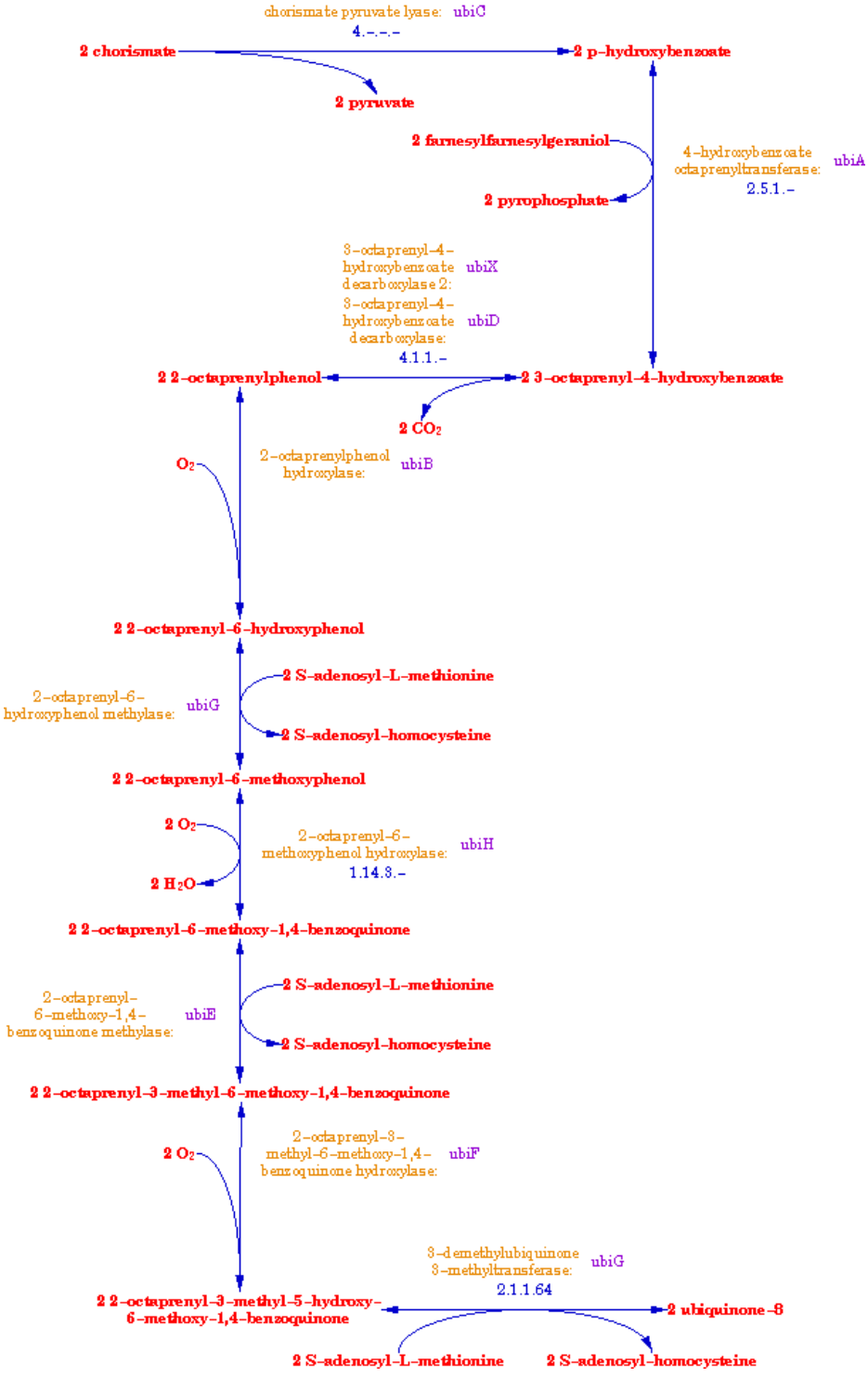


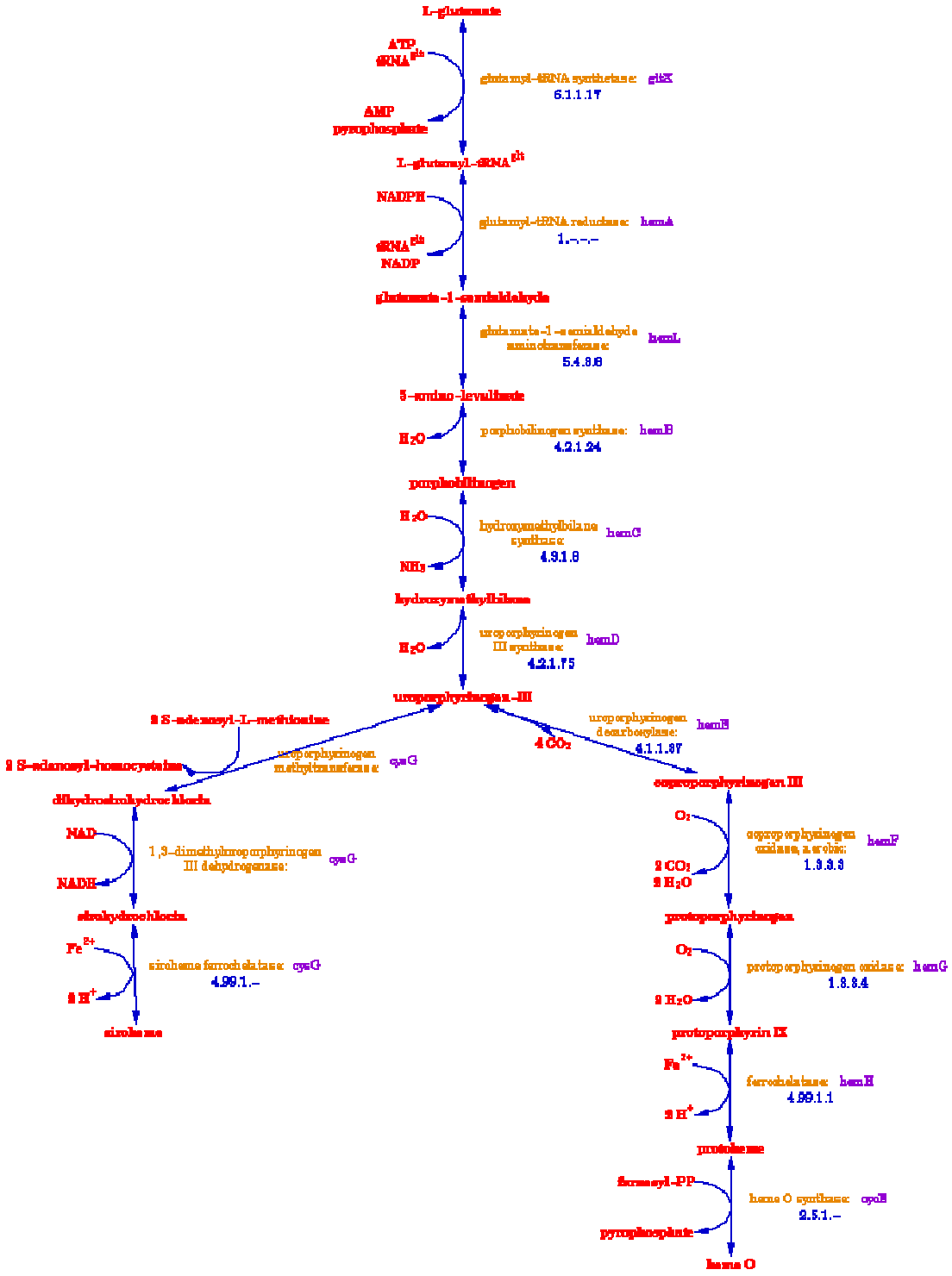
II.D Cofatores, Grupos Prostéticos e Transportadores / II.D.9 Riboflavina



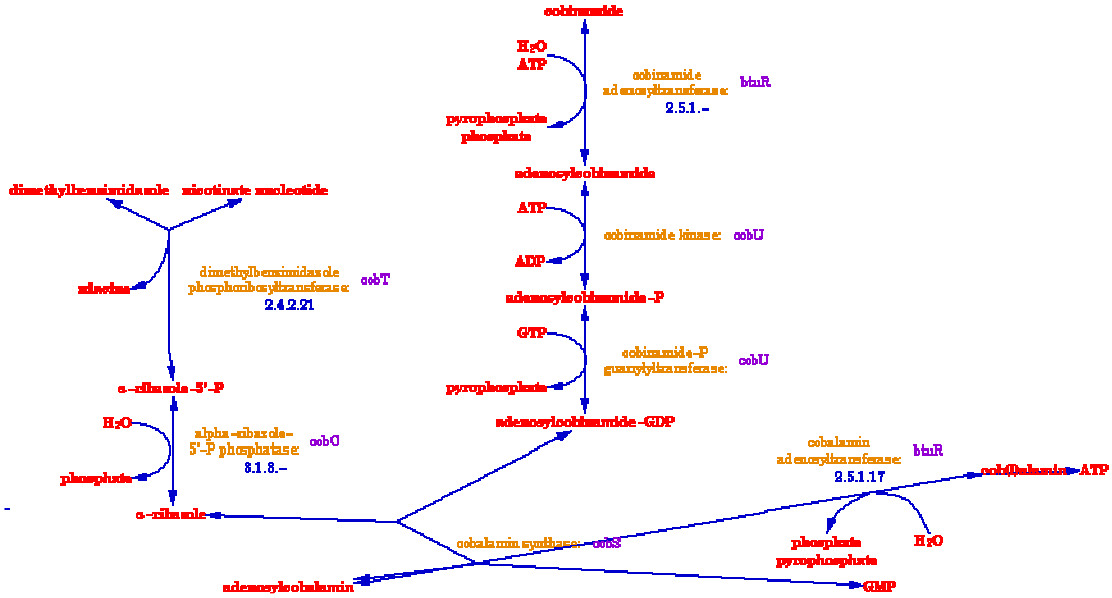
II.D Cofatores, Grupos Prostéticos e Transportadores / II.D.11 Menaquinona e Ubiquinona

Ubiquinona



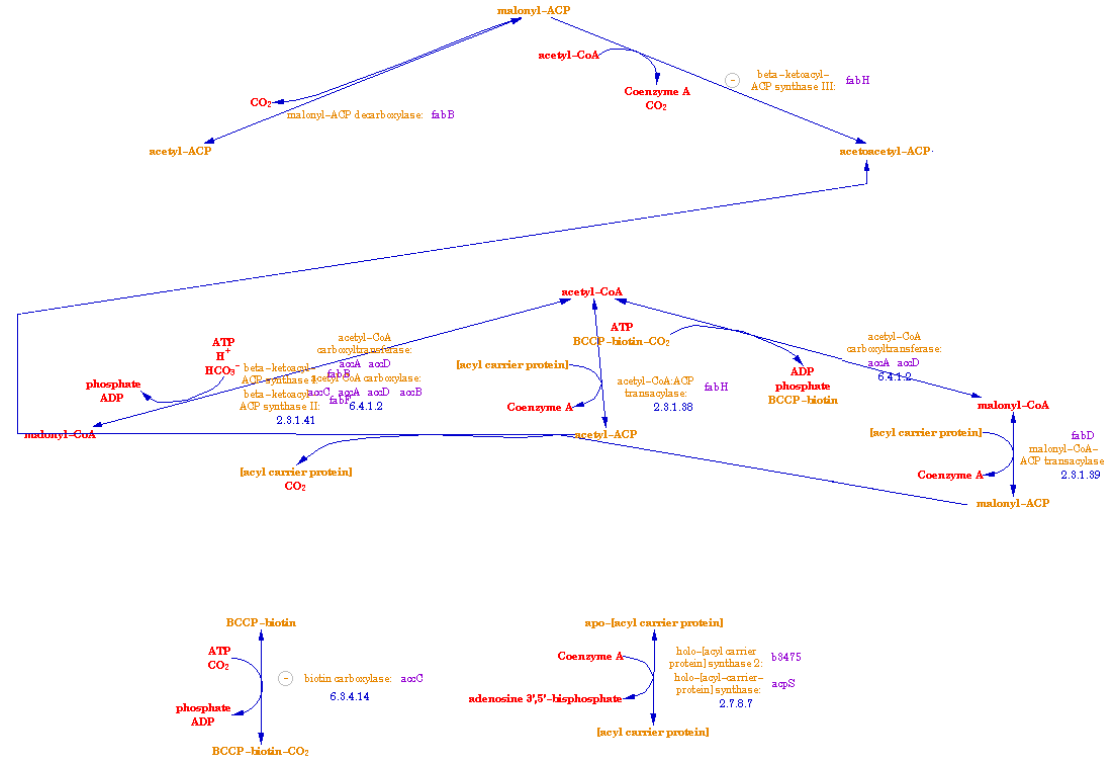

L-glutamate
ATP
tRNA Glu
glutamyl-tRNA synthetase: gltX
6.1.1.17
AMP
pyrophosphate
L-glutamyl-tRNA Glu
NADPH
glutamyl-tRNA reductase: hemA
1.-.-.-
tRNA Glu
NADP
glutamate-1-semialdehyde
glutamate-1-semialdehyde aminotransferase: hemL
5.4.3.8
5-amino-levulinate
H2O
porphobilinogen synthase: hemB
4.2.1.24
porphobilinogen
H2O
hydroxymethylbilane synthase: hemC
4.3.1.8
NH3
hydroxymethylbilane
H2O
uroporphyrinogen III synthase: hemD
4.2.1.75
uroporphyrinogen-III
2 S-adenosyl-L-methionine
uroporphyrinogen methyltransferase: cysG
2 S-adenosyl-homocysteine
4 CO2
uroporphyrinogen decarboxylase: hemE
4.1.1.37
dihydrosirohydrochlorin
coproporphyrinogen III
NAD
1,3-dimethylmuroporphyrinogen III dehydrogenase: cysG
NADH
O2
coproporphyrinogen oxidase, aerobic: hemF
2 CO2
2 H2O
sirohydrochlorin
protoporphyrinogen
Fe2+
siroheme ferrochelatase: cysG
4.99.1.-
2 H+
O2
protoporphyrinogen oxidase: hemG
1.3.3.4
siroheme
2 H2O
protoporphyrin IX
Fe2+
ferrochelatase: hemH
4.99.1.1
2 H+
protoheme
farnesyl-PP
heme O synthase: cyoE
2.5.1.-
pyrophosphate
heme O

II.D Cofatores, Grupos Prostéticos e Transportadores / II.D.14 Cobalamina

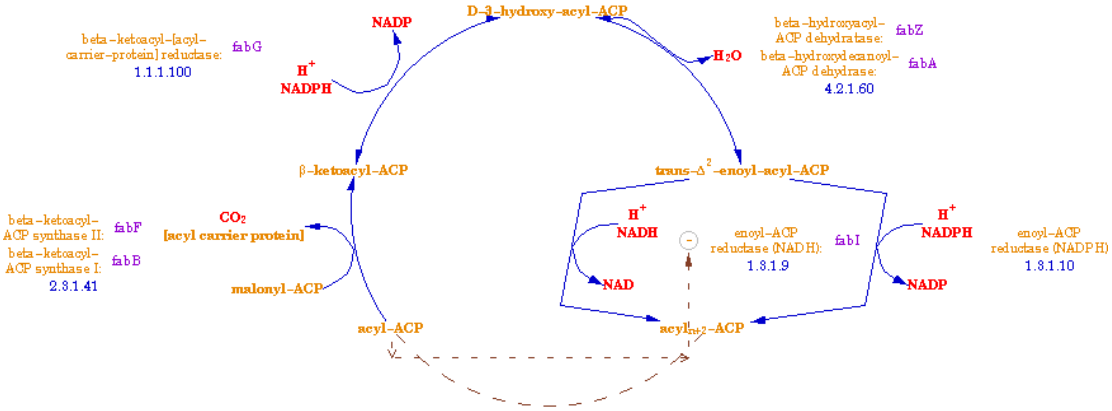


II.E Biossíntese de Ácidos Graxos e Fosfatídicos

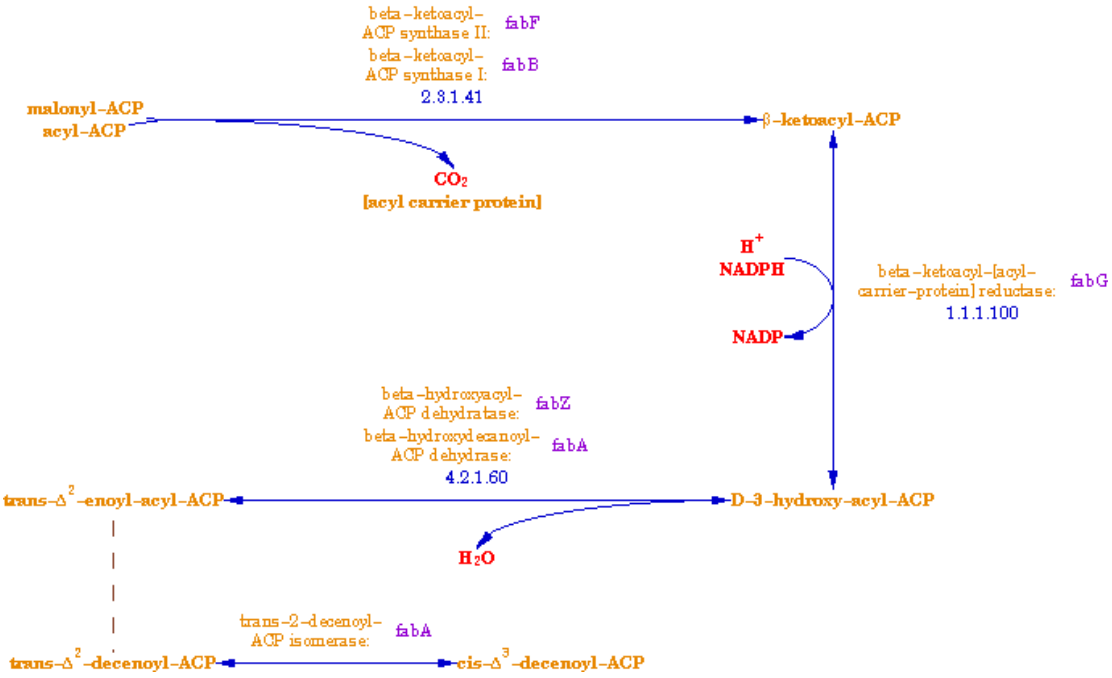
Passos Iniciais



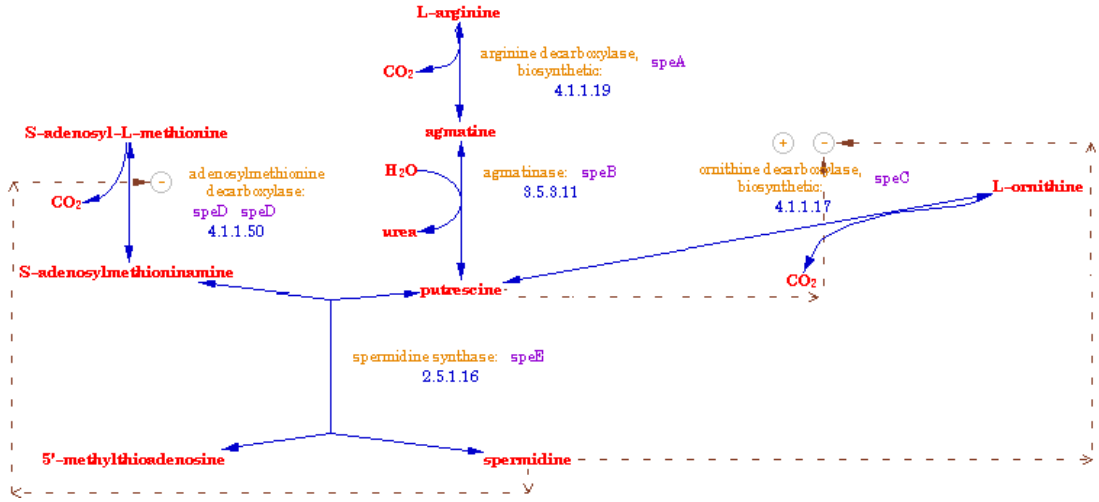
Elongação de Ácidos Graxos Saturados



Elongação de Ácidos Graxos Insaturados



II.F Poliaminas



ANEXO 4

Relatório Final da Categoria
II – Biossíntese de Pequenas Moléculas
(*Xylella fastidiosa* / *Pierce's Disease*)

Small molecule metabolism

X. fastidiosa PD is able to synthesize a wide range of small molecules necessary for survival within the xylem of various host plants. Most of the genes found in *E. coli* necessary for the synthesis of amino acids from chorismate, pyruvate, 3-phosphoglycerate, glutamate and oxaloacetic acid were identified.

However, some genes in *X. fastidiosa* PD are bi-functional, such as phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase (11501, hisI), aspartokinase/homoserine dehydrogenase I (11621, lysC/thrA), imidazole-glycerolphosphate dehydratase/histidinol-phosphate phosphatase (11541, hisB) and a diaminopimelate decarboxylase/aspartate kinase (3671, lysC/lysA) that catalyzes the first and the last steps of lysine biosynthesis. A bi-functional gene (3701) combines the DGTP-pyrophosphohydrolase (mutT) a NTP pyrophosphohydrolase, involved in the GO system responsible for removing an oxidatively damaged form of guanine, and a thiamine phosphate synthase (thiE) responsible for the production of thiamin monophosphate.(Pseudomonas REF)

In addition, the gene for acetylglutamate kinase (2641) has an acetyltransferase domain at its carboxy-terminal end that would compensate for a missing acetyltransferase (Amino-acid N-acetyltransferase EC 2.3.1.1) allowing the incorporation of glutamate into the arginine biosynthesis pathway.

In the serine pathway, phosphoserine phosphatase (serB) this is compensated by the use of the substrates pyruvate and glycine for serine biosynthesis. Other genes are missing, cystathionine b-lyase (L homocysteine can be produced via O-acetyl-L-homoserine using Cystathionine gamma-synthase metB), homoserine O-succinyltransferase (homoserine can be produced via O-acetyl-L-homoserine using Cystathionine gamma-synthase metB and Homoserine O-acetyltransferase metA) and 2,4,5-methyltetrahydrofolate-homocysteine methyltransferase catalyzes the conversion of L-homocysteine to L-methionine which can be carried out by 5-Methyltetrahydropteroyltriglutamate-homocysteine (metE). The first two enzymes are also absent in the *Bacillus subtilis* genome, the third is absent in *Haemophilus influenzae* and the fourth is missing in both genomes. Thus *X. fastidiosa* PD is able to complete these biosynthetic pathways by using alternative vias.

The pathways for the synthesis of purines, pyrimidines, common cofactors and nucleotides are all complete. Whether *X. fastidiosa* PD is also capable of both synthesizing and elongating fatty acids from acetate is open to question, at least one essential enzymes Enoyl-[acyl-carrier protein] reductase. (EC 1.3.1.9) is absent. Two orfs have a significant but limited homology (6161 e-8 and 16211 e-12) by Blast and have the same PFAM hit of 00106, adh-short chain dehydrogenases. These two orfs are related but distinct from the other adh type dehydrogenases in the genome and could therefore serve to complete this pathway.

The *E. coli* enzyme, holo acyl-carrier-protein synthase (also absent in *Synechocystis* sp., *H. influenzae* and *Mycoplasma genitalium*) suggesting that this organism has a limited range of interactions with other biosynthetic pathways that depend on acyl-transfer steps, such as polyketide (REF) and non-ribosomal peptide (REF) biosynthesis. Enoyl-ACP reductase (NADPH) (FabI) is also absent (also absent from *M. genitalium*, *Borrelia burgdorferi* and *Treponema pallidum*). This enzyme is linked to the production of acylated homoserine lactones molecules involved in processes such as quorum sensing (REF) and regulation of virulence gene expression (REF).

X. fastidiosa appears to be capable of synthesizing an extensive variety of enzyme cofactors and prosthetic groups, including biotin, folic acid, pantothenate and coenzyme A, ubiquinone, glutathione, thioredoxin, glutaredoxin, riboflavin, FMN, FAD, pyrimidine nucleotides, porphyrin, thiamin, pyridoxal 59-phosphate and lipoate.

In a number of the synthetic pathways, one or more of the enzymes present in E. coli are absent, but this is also true for at least one other sequenced Gram-negative bacterial genome in each case. We therefore again infer that the missing enzymes are either not essential or replaced by unknown proteins with novel structures.

ANEXO 5

Arquivo utilizado para análise filogenética
com o programa PAUP

input.nex

```
#NEXUS
TIME;
BEGIN TAXA;
DIMENSIONS NTAX=13;
TAXLABELS Borde_fur Ralst_fur Neiss_fur Yers_fur 12591XPD_fur
XF2344XCVC_fur Camp_fur 4702Camp_fur Vesic_fur 4702Citri_fur Oryz_fur
Bruce_fur Rhizo_fur
;
END;
begin characters;
dimensions nchar=176;
format missing=? gap=- datatype=Protein INTERLEAVE;

matrix

Borde_fur           ----------------------------MSDQSELKNMGLKATFPRLKILDIFR-KSDLRH
Ralst_fur           ----------------------------MPSPADLKNIGLKATVPRLKILEIFQ-TSEQRH
Neiss_fur           ------------------------MEKFNNIAQLKDSGLKVTGPRLKILDLFE-THAEEH
Yers_fur            -------------------------MTDNNKALKNAGLKVTLPRLKILEVLQ-NPACHH
12591XPD_fur        ----------------------------MELNDLRKVGLKVTHPRIRILELLEQSSSEHH
XF2344XCVC_fur      MHDDTLLCGCRNRFTRSRMLNGPSIEKSMELNDLRKVGLKVTHPRIRILELLEQSSSEHH
Camp_fur            ----------------------------METHDLRKVGLKVTHPRMRILELLEQKSNQHH
4702Camp_fur        ------------------------MNGERMETHDLRKVGLKVTHPRMRILELLEQKSNQHH
Vesic_fur           ----------------------------METHDLRKVGLKVTHPRMRILELLEQKSNQHH
4702Citri_fur       ----------------------------METHDLRKVGLKVTHPRMRILELLEQKSNQHH
Oryz_fur            ----------------------------METHDLRKVGLKVTHPRMRILELLEQKSNHHH
Bruce_fur           --------------------MNKPYTKPDYEQELRRAGVRITRPRRIILNILN--ETEDH
Rhizo_fur           --------------------MTD--VAKTLEELCTERGMRMTEQRRVIARILE--DSEDH

Borde_fur           LSAEDVYRALIAENVEIGLATVYRVLTQFEQAGILTRSQFDTGKAVFELNDGDHHDHLIC
Ralst_fur           LSAEDVYRILLNEHMDIGLATVYRVLTQFEQAGLLSRNNFESGKAIFELNEGKHHDHLVC
Neiss_fur           LSAEDVYRILLEEGVEIGVATIYRVLTQFEQAGILQRHHFETGKAVYELDKGDHHDHIVC
Yers_fur            VSAEDLYKILIDIGEEIGLATVYRVLNQFDDAGIVTRHNFEGGKSVFELTQQHHHDHLIC
12591XPD_fur        LSAEDIYRQLLEQGNEIGLATVYRVLTQFEAAGLVLKHNFESGQAVYEIDRGGHHDHMVD
XF2344XCVC_fur      LSAEDIYRQLLEQGNEIGLATVYRVLTQFEAAGLVLKHNFESGQAVYEIDRGGHHDHMVD
Camp_fur            LSAEDIYRQLLDHGDEIGLATVYRVLTQFEAAGLVLKHNFEGDQAVYELDRGGHHDHMVD
4702Camp_fur        LSAEDIYRQLLDHGDEIGLATVYRVLTQFEAAGLVLKHNFEGGQAVYELDRGGHHDHMVD
Vesic_fur           LSAEDIYRQLLDHGDEIGLATVYRVLTQFEAAGLVLKHNFEGGQAVYELDRGGHHDHMVD
4702Citri_fur       LSAEDIYRQLLDHGDEIGLATVYRVLTQFEAAGLVLKHNFEGGQAVYELDRGGHHDHMVD
Oryz_fur            LSAEDIYRQLLDHGDEIGLATMYRVLTQFEAAGLVLKHNFEGGQAVYELDRGGHHDHMVD
Bruce_fur           PDALEIFRRAVEEDDSISLSTVYRTMKLLEERGAIHRHAFAGGPSRFEQASGAHHDHIID
Rhizo_fur           PDVEELYRRSVKVDAKISISTVYRTVKLFEDAGIIARHDFRDGGSRYETVPEEHHDHLID

Borde_fur           TNCGTVFEFSDPDIEKRQYKVAKDNGFVLESHAMVLYGICG--NC-----QKGR--

Ralst_fur           LDCGRVEEFFDADIEQRQQSIARERGFALQEHALSLYGNCTKDDCP----HRPRR-
Neiss_fur           VKCGEVTEFHNPEIEALQDKIAEENGYRIVDHALYMYGVCS--DCQ----AKGKR-
Yers_fur            LDCGKVIEFSNESIESLQREIAKQHGIKLTNHSLYLYGHCETGNCREDESAHSKR-
12591XPD_fur        VDTGKIIEFHNEEIELLQRSIAAERGYELEEHSLVLYVRKK----------RGR--
XF2344XCVC_fur      VDTGKIIEFHNEEIELLQRSIAAERGYELEEHSLVLYVRKK----------RGR--
Camp_fur            VDTGHVIEFESEEIEALQRQIAAKHGYELEEHSLVLYVRKK----------RPR--
4702Camp_fur        VDTGHVIEFESEEIEALQRQIAAKHGYELEEHSLVLYVRKK----------RPR--
Vesic_fur           VDTGHVIEFESEEIEALQRQIAAKHGYELEEHSLVLYVRKK----------RPR--
4702Citri_fur       VDTGHVIEFESEEIEALQRQIAAKHGYELEEHSLVLYVRKK----------RAR--
Oryz_fur            VDTGHVIEFESEEIEALQRQIAAKHGYELEEHSLVLYVRKK----------RPR--
Bruce_fur           MDSGDVVEFHSDKIEKLQEEIARSLGFEIVHHRLELYCKKL----------KS---
Rhizo_fur           LKTGTVIEFRSPEIEALQERIAREHGFRLVDHRLELYGVPL----------KKEDL


;
END;
```

```
LOG FILE = output.txt;


SET
AUTOCLOSE = YES
OUTROOT = POLYTOMY
STOREBRLENS = YES
STORETREEWTS = YES;

HSEARCH
nreps=100
ADDSEQ = RANDOM;

savetrees file = trees.TRE brlens=yes;

CSTATUS FULL = YES;

SET CRITERION = PARSIMONY;

EXECUTE TREES.TRE;

CONTREE /
STRICT = yes
SEMISTRICT = NO
MAJRULE = YES
PERCENT = 50
ADAMS = NO
INDICES = YES
GRPFREQ = YES
SHOWTREE = YES
USETREEWTS = YES
TREEFILE = CONSENSO.TRE
APPEND = NO
TCOMPRESS = NO
ROOT = OUTGROUP
OUTROOT = POLYTOMY;

execute consenso.TRE;
savetrees file = treeconsenso.TRE brlens=yes;

[!

Tree number 1 = strict consensu

Tree number 2 = MAJRULE consensu 50%

]

DESCRIBETREES ALL  /
PLOT = BOTH
ROOT = OUTGROUP
OUTROOT = POLYTOMY
BRLENS = YES
```

```
LABELNODE = YES
XOUT = NONE
MPRSETS = NO
APOLIST = YES
CHGLIST = YES
PATRISTIC = YES
HOMOPLASY = YES
FVALUE = YES
DIAG = NO
TCOMPRESS = NO
CMLABELS = YES
CMCOLWID = 1
CMCSTATUS = YES;

[!

Tree number 1 = strict consensu

Tree number 2 = MAJRULE consensu 50%
]

PSCORES ALL /
SINGLE = NO
RANGE = YES
TOTAL = no
TL = YES
CI = YES
RI = YES
RC = YES
HI = YES
GFIT = YES
KHTEST = YES
NONPARAMTEST = NO
TESTDETAILS = NO;


SHOWUSERTYPE;
TSTATUS;
fstatus;

HSEARCH
nreps=100;

BOOTSTRAP
NREPS = 1000
CONLEVEL = 50
NCHAR = CURRENT
FORMAT = NEXUS
BRLENS = YES
REPLACE = YES
KEEPALL = no
;

savetrees file = bootstraping.tre;
```

ANEXO 6

Dados preliminares da anotação de *Xanthomonas*,
Categoria V – Processos Celulares

Tabela de genes de *Xanthomonas axonopodis* pv. *citri* relacionados ao sistema TonB

	Nome do	Nome "alternativo" para		Transporte	
ORF	**Produto**	**Produto**	**Nome**	**TC#**	**valor E**
66	TonB-dependent receptor	FepA ferri-enterobactin receptor	fepA	1.B.14.1.1	1,00E-06
407	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	1,00E-05
630	TonB-dependent receptor	FecA ferric-citrate receptor	fecA	1.B.14.1.2	9,00E-09
911	TonB protein		tonB	hits fracos	3,30
1561	TonB protein		tonB	2.C.1.1.1	2,00E-06
2146	TonB-dependent receptor	FhuA ferrichrome receptor	fhuA	1.B.14.1.4	9,00E-08
2290	TonB-dependent receptor	FyuA Fe-pesticin receptor	fyuA	1.B.14.7.2	1,00E-08
2857	TonB protein		tonB	2.C.1.1.1	1,00E-09
2874	TonB-dependent receptor	FhuA ferrichrome receptor	fhuA	1.B.14.1.4	5,00E-42
3367	TonB-dependent receptor	FecA ferric-citrate receptor	fecA	1.B.14.1.2	2,00E-08
3671	TonB-dependent receptor	IroN ferri-enterobactin receptor	iroN	1.B.14.1.5	2,00E-07
3940	TonB protein		tonB	hits fracos	4,00
4403	TonB-dependent receptor	FyuA Fe-pesticin receptor	fyuA	1.B.14.7.2	2,00E-26
4507	TolQ protein		tolQ	2.C.1.2.1	5,00E-48
4509	TolR protein		tolR	2.C.1.2.1	2,00E-15
4510	TolA protein		tolA	2.C.1.2.1	1,20
4511	TolB protein		tolB	hits fracos	2,00
4578	TonB protein		tonB	2.C.1.1.1	0,19
5074	TonB-dependent receptor	FecA ferric-citrate receptor	fecA	1.B.14.1.2	6,00E-12
5079	TonB-dependent receptor	FecA ferric-citrate receptor	fecA	1.B.14.1.2	0,011
5100	TonB-dependent receptor	IroN ferri-enterobactin receptor	iroN	1.B.14.1.5	2,00E-08
5118	TonB-dependent receptor	FyuA Fe-pesticin receptor	fyuA	1.B.14.7.2	1,00E-05
5303	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	6,00E-08
5413	ExbD2 protein		exbD2	2.C.1.1.1	3,00E-12
5414	ExbD1 protein		exbD1	2.C.1.1.1	3,00E-18

	Nome do	Nome "alternativo" para		Transporte	
ORF	**Produto**	**Produto**	**Nome**	**TC#**	**valor E**
5419	TonB protein		tonB	2.C.1.1.1	0,081
5417	ExbB protein		exbB	**2.C.1.2.1**	4,00E-21
		Para exbB: 2.C.1.1 / 1e-15 O TC# citado bateu com tolQ			
5697	TonB-dependent receptor	FyuA Fe-pesticin receptor	fyuA	1.B.14.7.2	2,00E-16
5701	TonB-dependent receptor	FecA ferric-citrate receptor	fecA	1.B.14.1.2	4,00E-05
5739	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	5,00E-05
5806	TonB-dependent receptor	FyuA Fe-pesticin receptor	fyuA	1.B.14.7.2	1,00E-21
10905	TonB protein		tonB	hits fracos	4,7
10907	TonB protein		tonB	2.C.1.1.1	1,00E-05
3446	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	0,001
3448	TonB-dependent receptor	FhuA ferrichrome receptor	fhuA	1.B.14.1.4	0,012
2437	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	7,00E-10
6861	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	3,00E-04
5603	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	2,00E-17
94	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	6,00E-12
4192	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	3,00E-15
4314	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	1,00E-11
4203	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	9,00E-07
1715	TonB-dependent receptor	FhuA ferrichrome receptor	fhuA	1.B.14.1.4	5,00E-10
1361	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	1,00E-06
6107	TonB-dependent receptor	FhuA ferrichrome receptor	fhuA	1.B.14.1.4	1,00E-31
4551	TonB-dependent receptor	IroN ferri-enterobactin receptor	iroN	1.B.14.1.5	1,00E-07
4480	TonB-dependent receptor	FhuA ferrichrome receptor	fhuA	1.B.14.1.4	6,00E-07

	Nome do	Nome "alternativo" para		Transporte	
ORF	**Produto**	**Produto**	**Nome**	**TC#**	**valor E**
6314	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	8,00E-19
124	TonB-dependent receptor	IroN ferri-enterobactin receptor	iroN	1.B.14.1.5	6,00E-08
4621	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	3,00E-05
5547	TonB-dependent receptor	IroN ferri-enterobactin receptor	iroN	1.B.14.1.5	2,00E-05
1567	TonB-dependent receptor	CirA colicin I/Fe^{3+} catecholate receptor	cirA	1.B.14.1.6	1,00E-07
2119	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	3,00E-24
2113	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	1,00E-16
3077	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	7,00E-18
3162	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	3,00E-14
2851	TonB-dependent receptor	IroN ferri-enterobactin receptor	iroN	1.B.14.1.5	4,00E-07
3893	TonB-dependent receptor	BtuB cobalamim receptor	btuB	1.B.14.3.1	1,00E-07
6219	TonB-dependent receptor	FecA ferric-citrate receptor	fecA	1.B.14.1.2	0,001