

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



GUILHERME COUTINHO DE MELO SERRANO

**“SÍNTESE E DEGRADAÇÃO DE LISINA EM ORGANISMOS
SUPERIORES: UMA POSSÍVEL ORIGEM BACTERIANA”**

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
GUILHERME COUTINHO DE MELO
SERRANO
e aprovada pela Comissão Julgadora.

Dissertação apresentada ao Instituto de
Biologia para obtenção do Título de
Mestre em genética e biologia
molecular, na área de genética
molecular de sistemas.

A handwritten signature in blue ink, appearing to read "Arruda".

Orientador: Prof. Dr. Paulo Arruda

Campinas, 2010

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

Se68s

Serrano, Guilherme Coutinho de Melo
Síntese e degradação de lisina em organismos superiores: uma possível origem bacteriana / Guilherme Coutinho de Melo Serrano. – Campinas, SP: [s.n.], 2010.

Orientador: Paulo Arruda.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Biosíntese. 2. Lisina. 3. Lisina - Metabolismo. 4. Inseto. 5. Bactérias. I. Arruda, Paulo. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

(rcdt/ib)

Título em inglês: Synthesis and degradation of lysine in higher organisms: a possible bacterial origin.

Palavras-chave em inglês: Biosynthesis; Lysine; Lysine – Metabolismo; Insects; Bacteria.

Área de concentração: Genética de Microrganismos.

Titulação: Mestre em Genética e Biologia Molecular.

Banca examinadora: Paulo Arruda, Ricardo Azevedo, Gonçalo Amarante Guimarães Pereira.

Data da defesa: 21/02/2010.

Programa de Pós-Graduação: Genética e Biologia Molecular.

BANCA EXAMINADORA

Campinas, 22 de fevereiro de 2010.

BANCA EXAMINADORA

Prof. Dr. Paulo Arruda (Orientador)



Assinatura

Profa. Dr. Ricardo Azevedo



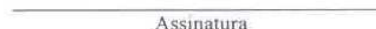
Assinatura

Prof. Dr. Gonçalo Amarante Guimarães Pereira



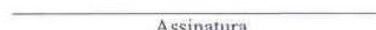
Assinatura

Prof. Dr. Fabio Trindade



Assinatura

Profa. Dra Maria Elena Goldman



Assinatura

A resposta certa, não importa nada: o essencial é que as perguntas estejam certas.

Mário Quintana, *As indagações*

Aos meus pais, que por muitas vezes abdicaram da
própria felicidade para darem condições
de eu seguir o meu caminho.

AGRADECIMENTOS

Agradeço:

Ao Paulo Arruda que em um momento de muito atarefamento em sua vida aceitou me orientar sendo impecável e brilhante nessa jornada.

A FAPESP por acreditar e fomentar esse projeto

A Thais Rezende e Silva por me auxiliar intelectualmente no início desse trabalho além de ser parceira em todos os trabalhos por nosso grupo produzidos.

Ao Eduardo Kyota por todo suporte técnico no laboratório além de ser grande companheiro de sala e amigo para todas as horas.

Ao Mario Paniago, grande empresário do laboratório vizinho que além de amigo foi crítico em discussões valiosas durante o trabalho.

Ao Márcio Jose da Silva e Daniela Stancato pelo apoio e permissão para uso de seu laboratório em vários experimentos.

Ao Professor Nilson e sua técnica Veruska do LNLS por disponibilizar o uso de sua estrutura.

Aos membros da pré-banca Dr. Fabio Trindade e Dr. Gonçalo Pereira pela discussão e crítica de alto nível sobre o tema e contribuições para que a tese atingisse a forma atual.

A secretaria de pós graduação do programa de genética e biologia molecular, Lourdes Aparecida, por me auxiliar prontamente em todo processo burocrático.

Ao curso de graduação de Biologia da Unicamp que me preparou para enfrentar os novos desafios da pós graduação.

A todos os amigos que me acompanharam por esses anos de trabalho e agora ficaram para o resto de uma vida.

ÍNDICE

BANCA EXAMINADORA	iii
DEDICATÓRIA.....	v
AGRADECIMENTOS	vi
ÍNDICE	vii
Índice de figuras e tabelas	viii
Lista de abreviações e termos em inglês.....	xi
Resumo e Abstract.....	1
Introdução Geral.....	4
Materiais e Métodos	
1. Busca por homólogos, alinhamento e filogenia.....	7
2. Busca por domínios conservados.....	8
3. Manutenção da colônia de <i>A. gambiae</i> e <i>A. aegypti</i>	8
4. Extração de DNA e PCR.....	8
5. Extração de proteína.....	9
6. Separação de proteínas em gel de poli(acrilamida) e Western blot...	9
7. Ensaio enzimático da LKR/SDH.....	10
Capítulo I – Síntese de lisina pela via do ácido aspártico em insetos	
1. Introdução.....	11
2. Resultados.....	16
2.1 Identificação de possíveis genes que codificam enzimas da via do ácido aspártico em insetos.....	17
2.2 Análise das sequências de aminoácidos das enzimas do ramo da lisina encontradas em <i>A. gambiae</i> e <i>A. aegypti</i>	22
2.3 Amplificação dos genes a partir do DNA genômico dos insetos.....	31
3. Discussão	35
4. Conclusão.....	39

Capítulo II – Presença da via da sacaropina em bactérias

1. Introdução.....	39
2. Resultados.....	43
2.1 Identificação dos genes que codificam a LKR e a SDH em bactérias	43
2.2 Amplificação do gene LKR/SDH de <i>Silicibacter pomeroyi</i>	53
2.3 Ensaio enzimático das de LKR e da SDH de <i>Silicibacter pomeroyi</i>	54
3. Discussão	55
4. Conclusão.....	59
Bibliografia.....	60
Anexo I.....	71

ÍNDICE DE FIGURAS E TABELAS

Figura 1. Via metabólica do ácido aspártico em plantas.....	15
Figura 2. Estrutura do supercontig 1.3837 de <i>Aedes aegypti</i>	22
Figura3. (A) Alinhamento das sequências de aminoácidos e (B) árvore filogenética das sequências de plantas bactérias e <i>Anopheles</i> para a enzima AK.....	24
Figura 4. (A) Alinhamento das sequências de aminoácidos e (B) árvore filogenética das sequências de plantas bactérias e <i>Anopheles</i> para a enzima ASDH	25
Figura 5. (A) Alinhamento das sequências de aminoácidos e (B) árvore filogenética das sequências de plantas bactérias e <i>A. gambiae</i> para a enzima DHDPS.....	26
Figura 6. (A) Alinhamento das sequências de aminoácidos e (B) árvore filogenética das sequências de plantas bactérias e <i>Anopheles</i> para a enzima DHDPR.....	27
Figura 7. (A) Alinhamento das sequências de aminoácidos e (B) árvore filogenética das sequências de plantas bactérias e <i>Anopheles</i> para a enzima DAPE.....	28
Figura 8. (A) Alinhamento das sequências de aminoácidos e (B) árvore filogenética das sequências de plantas bactérias e <i>Anopheles</i> para a enzima DAPD.....	29
Figura 9. (A) eletroforese em gel de agarose do DNA genômico de bactérias e <i>Anopheles gambiae</i> . (B) amplificações com primers 16S bacteriano das amostras apresentadas em (A).....	32
Figura 10. Eletroforese em gel de agarose com dos fragmentos amplificados dos genes que codificam enzimas da via do ácido aspártico em <i>Anopheles gambiae</i> tratado com tetraciclina	33
Figura 11. Eletroforese em gel de agarose dos fragmentos amplificados do gene que codifica a enzima DHDPS de <i>Aedes aegypti</i>	34
Figura 12. Via da sacaropina de degradação de lisina em plantas e animais.....	43
Figura 13. Alinhamento das sequências de aminoácidos das enzimas LKR e SDH de plantas, animais e leveduras.....	46

Figura 14. (A), Alinhamento das sequências de aminoácidos dos polipeptídios que flanqueiam o ID de plantas e cianobactérias conjuntamente com as sequências de aminoácidos da LKR e SDH de animais, leveduras e α -proteobactérias. (B), árvore filogenética das sequências da LKR de plantas, animais, α -proteobactérias, levedura e da correspondente em cianobactérias. (C) árvore filogenética das sequências da SDH de plantas, animais, α -proteobactérias, levedura e da correspondente em cianobactérias.....	51
Figura 15. Gel de agarose 1% do resultado da amplificação dos genes LKR e SDH em <i>Silicibacter pomeroyi</i> , utilizando diferentes combinações de primers.....	53
Figura 16. Ensaio enzimático de LKR e SDH. (A) Oxidação de NADPH em função do tempo - atividade de LKR. (B) Redução de NAD ⁺ em função do tempo - atividade de SDH. (C) Oxidação de NADPH em função da massa de proteína aplicada. (D) Redução do NAD ⁺ em função da massa de proteína aplicada.....	54
Figura 17. Diagrama explicativo da teoria criada para explicar a distribuição atual dos genes LKR/SDH e do inter-domínio (ID).....	58
Tabela 1. Enzimas da via do ácido aspártico cujas sequências de aminoácidos foram utilizadas como “iscas” na busca de ortólogos em insetos.....	16
Tabela 2. Enzimas da via do ácido aspártico cujas sequências de aminoácidos foram identificadas como possíveis ortólogas em insetos.....	18
Tabela 3. Características das sequências dos genes da via de síntese de lisina em <i>A. gambie</i>	19
Tabela 4. Número de transcritos encontrados no banco de dados do portal Vectorbase para os genes da via de biosíntese de lisina.....	20
Tabela 5. Resultado da busca por BLASTn no banco de dados de todas sequências nucleotídicas do NCBI.....	21
Tabela 6. Similaridade entre os domínios de síntese de lisina com o banco de dados de domínios conservados cdd e pfam.....	31

Tabela 7. Estrutura dos domínios enzimáticos da LKR/SDH em plantas, animais e bactérias.....	48
Tabela 8. Matriz de similaridade entre as sequências da LKR, SDH e ID de plantas, animais, leveduras, cianobactérias e α -proteobactérias.....	52

ABREVIACES E TERMOS EM INGLS

Contig:	Contguo. Sequncia produzida a partir da sobreposio de duas ou mais sequncias para de uma mesma regio genmica.
Read:	sequncia nica produzida a partir de um nico evento de sequnciamento (clone).
Singleton:	Sequncia que no se sobrepe com nenhuma outra.
Primer:	Oligonucleotdeo iniciadores de replicao de DNA.
DNA:	cido desoxiribonucleico
RNA:	cido ribonucleico
cDNA:	DNA obtido a partir da transcrio reversa de um mRNA.
EST:	Expressed Sequence Tags (etiquetas de sequncias expressas)
Kb:	Milhares de pares de base
KDa:	Milhares de daltons (unidade de massa atmica)
PCR:	Polimerase chain reaction (reao da polimerase em cadeia)
EDTA:	cido Etilennodiaminotetractico
DTT:	Ditiotreitol
SDS-PAGE:	Doedecil sulfato de sdio
Western Blot:	Reao de anticorpo com antgeno em membrana.
Tris:	Tris (hodrximetil)aminometano
TLC:	Thin layer chromatography (comomatografia em camada delgada)
Pb:	Pares de base
GI:	Gene Index. Nmero de indexao de um gene no GeneBank.

Resumo

A lisina é considerada um aminoácido essencial, pois é componente fundamental de proteínas e não pode ser sintetizado por animais, sendo necessário ingeri-lo em sua forma final. Sua concentração é baixa em cereais, principal fonte de alimento animal e sua carência pode trazer sérios danos ao organismo, principalmente relacionados ao sistema nervoso. Por outro lado, seu excesso, causado pela deficiência na degradação, também é danoso podendo levar ao retardo no desenvolvimento mental. A síntese da lisina em plantas e bactérias é realizada principalmente pela via do ácido aspártico, que além desse aminoácido é responsável pela produção de treonina, metionina e isoleucina.

A degradação da lisina em animais e plantas ocorre principalmente pela via da sacaropina. Essa via, por sua vez, é utilizada para a síntese de lisina em fungos. Assim, tanto a síntese como a degradação de lisina em diferentes organismos possuem arquiteturas metabólicas particulares que, durante o processo evolutivo, foram selecionadas para adequar-se as necessidades do metabolismo, diferenciação e desenvolvimento.

Até o momento não existia indício da existência e da funcionalidade da via do ácido aspártico em animais e nem da via da sacaropina em bactérias. O presente trabalho apresenta um conjunto de resultados que sugerem a existência da via do ácido aspártico em insetos e a via da sacaropina em bactérias. Foram identificados em *Anopheles gambiae* e *Aedes aegypti*, respectivamente 6 e 5 genes dos 9 que compõem a via do ácido aspártico em bactérias. A análise de similaridade de sequências sugere que os genes da via do ácido aspártico encontrados nos insetos pode ter se originado a partir de bactérias endossimbióticas. A necessidade da via de síntese de lisina nesses insetos pode estar ligada a fonte de alimentação, em alguns casos deficiente em aminoácidos essenciais. Quanto à degradação da lisina, os resultados sugerem a existência, em alguns grupos de bactérias, dos genes que codificam as enzimas lisina-cetoglutarato redutase e sacaropina desidrogenase (*lkr* e *sdh*), responsáveis pelos passos iniciais da via em plantas e animais. Em alfa-proteobactérias, encontramos os

genes *lkr* e *sdh* ligados em tandem na forma de um operon. Em plantas e animais esses genes também são ligados e codificam um polipeptídio bifuncional contendo as atividades da LKR e de SDH. O operon da *lkr* e *sdh* encontrado em alfa-proteobactérias contém também os genes que codificam as enzimas glutathione-S-transferase (GST) e fosfogliceraldeído mutase (FM). Essa estrutura genômica sugere que a LKR e a SDH podem fazer parte de um conjunto de enzimas importantes na defesa contra estresse oxidativo, além da degradação da lisina como pode ser demonstrado pela atividade enzimática de LKR e SDH nesse organismo. A análise da estrutura genômica dos genes da via da sacaropina em bactérias revelou um fato curioso. Em plantas existe um peptídeo de aproximadamente 100 aminoácidos intercalando os domínios polipeptídicos da LKR e da SDH. Esse peptídeo, denominado de interdomínio (ID) é exclusivo de plantas, não tendo sido encontrado em nenhum dos genomas de eucariotos seqüenciados até o momento. Sequências similares ao ID foram encontradas em cianobactérias e algumas archeobactérias, porém em nenhum caso esses IDs puderam ser associados com sequências codificadoras de LKR e/ou SDH. Neste trabalho buscamos remontar uma possível história evolutiva da via da sacaropina e apresentamos uma hipótese evolutiva para a origem dos genes LKR e SDH de plantas e animais.

Abstract

Lysine is an amino acid that is not synthesized by animals and therefore need to be ingested in its final form. Lysine concentration is low in cereals and the lack of this amino acid in diet can cause severe damage to the organism specially associated to the nervous system. On the other hand its excess, caused by the deficiency in the degradation of lysine, can also lead to organism disorder as mental development retardation. Lysine is mainly synthesized through the aspartate pathway in plant and bacteria and the same pathway can also lead to synthesis of threonine, methionine and isoleucine. Lysine degradation occurs through the saccharopine pathway in animal and plant, but other pathways are known for bacteria and fungi. Until now there was not an evidence of the existence of any of the pathways. Here we present new data of the existence of the aspartate pathway in animals and the saccharopine pathway bacteria respectively. We were able to demonstrate the existence of most of the genes that compose the aspartate pathway in *Anopheles gambiae* e *Aedes aegypti* and sequence clustering suggest a possible origin of these genes from endosymbiotic bacteria. Concerning the lysine degradation, we have gathered data that indicates the existence of genes responsible for producing the proteins for two first enzymatic steps, lysine ketoglutarate reductase (LKR) and saccharopine dehydrogenase (SDH), in bacteria. Exclusively in Proteobacteria we found these genes in the exact same structure of animals and plants and we were able to detect enzymatic activity for LKR and SDH. Curiously these genes are inside a single operon that also has glutathione S transferase and phosphoglyceraldehyde mutase, both genes previously associated with defense against oxidative stress. This may suggest a possible new role for LKR/SDH in preventing oxidative stress. The analysis of amino acid sequence and domains of LKR/SDH in plant, animals and fungus, shows that there is a region of approximately 110 aa between LKR and SDH domains that is exclusive to plants, we called this region the inter domain (ID). Interestingly we found only in Cyanobacteria this typical plant domain, the ID. After these novel results we attempted to reconstruct the evolutionary history of pathway, suggesting an explanation for the origin of plants and animals LKR and SDH.

Introdução Geral

Os aminoácidos possuem participação fundamental nos sistemas biológicos, pois são as peças primordiais constituintes das dezenas de milhares de diferentes proteínas que compõem a arquitetura e o funcionamento dos organismos. Além disso, atuam no controle do pH intracelular, servem como reserva de nitrogênio e regulam diferentes aspectos do metabolismo seja agindo em suas formas finais ou provendo moléculas do metabolismo intermediário que atuam como efetores na regulação da expressão gênica (Ferreira et al., 2005). A descrição do primeiro aminoácido foi feita no início do século XIX por dois químicos franceses que isolaram o composto de aspargo e deram o nome de asparagina (Vauquelin e Robiquet, 1806). Os aminoácidos, como o próprio nome diz, possuem em suas estruturas químicas um grupamento amino e uma carboxila. Esses dois grupamentos permitem que os aminoácidos se liguem uns aos outros, através da ligação amina-carboxila, formando as ligações peptídicas. Os aminoácidos mais relevantes para os organismos são os alfa-aminoácidos, que apresentam fórmula geral $H_2NCHRCOOH$, sendo R a cadeia orgânica variável que se liga ao carbono alfa e dá a especificidade de cada aminoácido. Os aminoácidos são comumente classificados em quatro categorias de acordo com a característica da cadeia lateral, eles podem ser ácidos, básicos, polares ou apolares, sendo essas forças as principais moldadoras das estruturas das proteínas (Creighton, 1993). Existem ainda isômeros óticos como L e D-aminoácidos dependendo do desvio ótico gerado no gliceraldeído que lhe deu origem. A grande maioria dos aminoácidos componentes das proteínas é L-aminoácido.

A lisina apresenta fórmula molecular $HO_2CCH(NH_2)(CH_2)_4NH_2$ e tem a cadeia lateral mais longa entre os aminoácidos que compõem as proteínas. A presença de dois radicais amino dá à lisina a característica básica. A existência de um radical amino lateral permite que a lisina incorporada nas proteínas sofra modificações pós-traducionais como metilação, acetilação, ubiquitinação e carboxilação, que podem levar a alterações nas funções das proteínas (Sadoul et al., 2008).

Nove dos vinte aminoácidos naturais constituintes das proteínas não são sintetizados por animais (Ferreira et al., 2005). Esses aminoácidos precisam ser ingeridos pelos animais na sua forma final, pronta para serem incorporados nas proteínas ou desempenhar suas demais funções no metabolismo dos organismos. Por essa razão esses aminoácidos são chamados de aminoácidos essenciais (Fürst and Stehle, 2004). Dentre os aminoácidos essenciais, a lisina é considerada o mais importante sob o ponto de vista nutricional, pois está presente em baixas concentrações nos cereais que são a maior fonte para alimentação animal e humana (Shewry et al., 1981, Bright e Shewry, 1983). As sementes em geral contêm grandes quantidades de proteínas de reserva que, após a germinação, são utilizadas como fonte de nitrogênio para o desenvolvimento dos estágios iniciais do embrião. Nos cereais, a maior parte das proteínas de reserva pertence a uma classe de proteínas solúvel em soluções alcolicas chamada de prolamina (Shewry e Halford, 2002). As prolaminas por sua vez são classificadas em α -, β -, γ -, and δ -prolaminas de acordo com suas características estruturais e de solubilidade (Esen, 1987). Em sementes de milho, sorgo e coix as α -prolaminas são a classe mais abundante (Esen, 1987, DeRose et al., 1989, Ottoboni et al., 1990). Ocorre que lamentavelmente as α -prolaminas não possuem resíduos de lisina em sua composição (Shewry et al., 1981) razão pela qual a alimentação com base nesses cereais é deficiente. Para uma alimentação nutricionalmente adequada, a lisina deve ser ingerida na concentração mínima 103 mg/kg/dia segundo recomendações da FAO (www.fao.org).

A deficiência no metabolismo ou ingestão de lisina ocasiona diversos distúrbios. A falta de lisina na alimentação prejudica a síntese de colágeno e elastina, componentes fundamentais do tecido conjuntivo. Uma alimentação deficiente nesse aminoácido leva a queda do conteúdo total de lisina no cérebro (Mori, et al. 1991) e afeta a atividade do hipotálamo (Smirga et al., 2000). Pode ainda levar em última instância a doenças cardíacas, apoplexia, osteoporose e até câncer (Matthias-Rath, 2003). A fortificação na alimentação com lisina foi associada à diminuição da ansiedade e do stress (Smirga et al., 2004).

A carência de lisina pode torna-se grande problema de saúde pública em locais onde a população depende de alimentação pouco diversificada e a base de cereais como milho e arroz (Jansen, 1962). O mesmo é válido para animais. Apesar de existir diversas fontes de alimentação dentro da biodiversidade de nosso planeta, as espécies são especialistas em algumas poucas fontes de alimentos. Os mosquitos, por exemplo, alimentam-se majoritariamente de néctar durante sua vida pré-copulatória. O néctar de plantas é pobre em proteínas e aminoácidos livres, especialmente em alguns aminoácidos essenciais, dentre eles lisina (Gardener e Gillman, 2001). Nesse contexto é intrigante como um animal especializado pode não sintetizar lisina sendo que sua principal fonte de alimentação possui baixos níveis desse aminoácido. Mosquitos podem ser criados durante todo seu ciclo adulto apenas com a proteína adquirida na fase larval e com alimentação de solução aquosa de sacarose 10%. Se o inseto passa pelas fases larvais, pupa e adulto necessitando de proteínas diferentes, com composição diferente de aminoácido, apenas a base de sacarose, como pode ele providenciar lisina e outros aminoácidos essenciais sem a capacidade de sintetizá-los?

Em plantas, o metabolismo de lisina é finamente regulado tanto na via de síntese como na via de degradação (Galili, 1995; Arruda et al., 2000; Azevedo e Lea, 2001). A síntese de lisina é controlada por inibição alostérica através desse produto final sobre as enzimas AK e DHDPs, respectivamente, o primeiro e quarto passos de nove que compõem a via (Azevedo et al., 2006). Plantas super-expressando o gene bacteriano da Dihidropicolinato Sintase (DHDPs) tiveram grande aumento na concentração de lisina livre acarretando em uma germinação reduzida (Zhu e Galili, 2003). A degradação de lisina é estimulada pela lisina e o gene que transcreve para a enzima Lisina cetoglutarato redutase/Sacaropina desidrogenase (LKR/SDH), os dois primeiros passos da via, tem sua expressão aumentada por um fator de transcrição conhecido como O₂, opaco 2 (Kemper et al., 1999). Plantas super-expressando o gene de degradação SDH tiveram o crescimento de brotos de *Arabidopsis thaliana* inibido por falta de lisina (Zhu e Galili, 2004).

No presente trabalho procuramos contribuir para o entendimento do metabolismo de lisina em diferentes modelos biológicos. Utilizando genômica, bioinformática e biologia molecular investigamos a possível existência da via do ácido aspártico em insetos e da via da sacaropina em bactérias. Os resultados são discutidos no contexto da evolução dessas vias nos diferentes organismos e da possibilidade de aquisição dos genes através de transferência lateral.

Materiais e Métodos

Busca por homólogos, alinhamento e filogenia

A busca por genes homólogos aos da via de biosíntese de lisina foi feita a partir de sequências iscas estabelecidas a partir de genes bem descritos na literatura. Tomou-se como iscas os genes de *E.coli* e *A.thaliana* em todos os casos, exceto em THPA, ADAPAT, ADPD em *A. thaliana* que segundo Hudson et al, 2005 não estão presentes em plantas. Esses genes foram submetidos à busca utilizando o algoritmo tBLASTn no GenBank e nos bancos do Tigr e sequências com e-value acima de $1e-10$ foram adicionados a um banco de dados criado no programa Bioedit versão 7.0. (Hall, 1999) As sequências foram orientadas no sentido correto e alinhadas no programa Clustal X (Thomson et al., 1997), utilizando o algoritmo BLOSUM 62 e valor de bootstrap igual 2000. Seguiu-se o corte manual da parte das sequências sem correspondentes em todos representantes, deixando somente a parte do gene representada em todos os organismos. A análise filogenética foi realizada pelo programa Mega 3.1 (Kumar et al., 2003) utilizando neighbor-joining e o modelo PAM de Dayhoff (Schwarz e Dayhoff, 1979). Foi realizada re-amostragem (bootstrap) de 1000.

Busca por domínios conservados

Uma vez alinhadas, as sequências dos genes possivelmente homólogos foram submetidas a uma busca por domínios conservados no banco CDD (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) e os resultados foram expressos graficamente. De forma semelhante às estruturas 3D das proteínas foram determinadas com o auxílio dos programas Phyre (<http://www.sbg.bio.ic.ac.uk/phyre>) e Esypred3D (www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/), em suas configurações padrão. Dos resultados foram selecionadas as estruturas com identidade superior a 30% e essas passaram por análise visual.

*Manutenção da colônia de *A. gambiae* e *A. aegypti**

Colônias matrizes de *A. aegypti* foram estabelecidas no ICBII-USP, laboratório de mosquitos transgênicos sob responsabilidade da Profa. Margareth L. Capurro, e lá são mantidas em estabelecimentos apropriados e certificados para portar insetos patogênicos. Colônias de *A. gambiae* foram estabelecidas na UC-San Diego. Dois grupos foram criados para cada um dos insetos, um tratado com antibiótico tetraciclina a 25µg/ml e outro controle. Todos os experimentos realizados utilizaram indivíduos de ambos os grupos para determinar uma possível interferência bacteriana nos resultados. Esses pernilongos foram à base de todos os experimentos realizados nesse trabalho.

Extração de DNA e PCR

O DNA dos organismos utilizados nesse trabalho foram extraído utilizando o reagente DNAZOL® da Invitrogen seguindo o protocolo de fabricante. A PCR foi realizada em 38 ciclos compostos de 40" à 94º C; 30" de 55º-65º C; 2' à 72º C. O resultado foi observado em gel de agarose 1% corado com brometo de etídeo.

Extração de proteína

As proteínas totais de milho, *Silicibacter* e *Nostoc* foram extraídas macerando-se 1g de material em 3 ml de tampão tampão fosfato 25mM pH 7,4 contendo benzamidina 5mM, DTT 1mM, EDTA 1mM na presença de nitrogênio líquido. O conteúdo foi centrifugado a 7.000 rpm por 15 minutos e ao sobrenadante teve o pH corrigido para 5,5, o PI da LKR, então foi adicionado sulfato de amônia 60%. A solução foi deixada descansar por 30 minutos para que ocorresse a precipitação das proteínas. Centrifugou-se a solução a 15.000 rpm por 15 minutos e o precipitado foi ressuspenso em tampão Tris-HCl pH 8,5 contendo DTT 1mM e EDTA 1mM. Seguiu-se então uma diálise em membrana Quiagen com limite de exclusão de 12.000 - 14.000Da.

Separação de proteínas em gel de poliacrilamida e Western blot

O gel de poliacrilamida é composto por duas partes: o gel de separação (resolving gel), que fica na parte inferior e o gel de empacotamento (stacking gel), situado na parte superior. O gel de separação é composto por Tris-HCl 375 mmol/L, pH 8,8; acrilamida/bis-acrilamida (29:1) 15% (v/v); SDS (/sodium dodecyl sulfate/) 0,1% (m/v); APS (persulfato de amônio) 0,1% (m/v) e Temed 0,08% (v/v). O gel de empacotamento é composto por Tris-HCl 126 mmol/L, pH 6,8; acrilamida/bis-acrilamida (29:1) 5% (v/v); SDS 0,1% (m/v); APS 0,1% (m/v) e Temed 0,05% (v/v). Para a corrida no SDS-PAGE, 2 volumes da amostra foram coletados e aos mesmos, foi adicionado 1 volume de tampão de amostra 3X (Tris-HCl 50 mmol/L, pH 6,8; glicerol 45% e SDS 2%). As amostras contendo o tampão de amostra foram então fervidas por 5 minutos e 20 uL das mesmas foram carregados no gel, submetido à diferença de potencial de 120 V por 1 hora. O gel foi então corado em solução de /Coomassie Blue/ R-250, Bio-Rad, (/Coomassie Blue/ R-250 0,25%; ácido acético 10 %; etanol 50%) por 1 hora e em seguida descorado em solução de descorante (etanol 10% e ácido acético 5% em água).

Quando se desejou fazer western blot o gel de poliacrilamida foi embebido em uma solução contendo mM Tris-base, 190 mM glicina, 3.5 mM SDS e 20% metanol por 10 minutos. As proteínas foram então transferidas para uma membrana de nitrocelulose (BioRad Transfer-Blot®) em uma cuba de transferência (BioRad transfer blot®- SD). As membranas foram então bloqueadas “overnight” a 4°C em 20 mM Tris-HCl, pH 7.4, 137 mM NaCl, 2.5 mM KCl, 0.1% (v/v) Tween 20, e 5% (w/v) leite em pó desnatado. Em seguida a membrana foi encubada com Como anticorpo primário foi utilizado anticorpo de coelho crescido contra LKR/SDH de milho. Entre as incubações do anticorpo primário e secundário lavou-se por 5 minutos a membrana com tampão TBS-T. O anticorpo secundário foi de cabra conjugada com HRP crescido contra anticorpos de coelhos. Segui-se incubação com solução de Pierce (SuperSignal® West Pico) e as bandas detectadas em um filme quimioluminescente.

Ensaio enzimático da LKR/SDH

O ensaio enzimático para LKR foi realizado monitorando-se o consumo do cofator NADPH através de seu pico de absorbância em 340nm na reação que transforma lisina em sacaropina. Utilizou-se Tris-HCl pH 7,4 100mM, Lisina 20mM a-cetoglutarato 10mM, NADPH 0,1mM. A atividade de SDH foi medida monitorando-se a formação de NADH, também a 340nm, resultante do consumo de sacaropina. Utilizou-se Tris HCl pH 8,5 100mM, sacaropina 2mM e NAD 2mM.

Capítulo I – Síntese de lisina pela via do ácido aspártico em insetos

Introdução

A lisina é sintetizada em bactérias e plantas através da via do ácido aspártico. O ácido aspártico, precursor inicial da via, pode ser obtido a partir da transaminação do oxalacetato oriundo do ciclo de Krebs na mitocôndria, ou através da ação da fosfoenolpiruvato carboxilase no citoplasma de plantas (Lea et al., 2001). O ácido aspártico participa como precursor de duas vias: uma que leva a síntese de asparagina, composto utilizado para transporte e armazenamento de nitrogênio em plantas, e outra que leva a síntese de lisina, treonina, isoleucina e metionina (Azevedo et al., 2006) (Fig. 1). A via do ácido aspártico em bactérias e plantas é finamente regulada pelos aminoácidos produtos finais. Os aminoácidos lisina, treonina e isoleucina assim como o intermediário S-adenosil metionina (AdoMet) regulam, através de retro inibição, pontos chave da via catalisados pelas enzimas aspartato kinase (AK), dihidrodipicolinato sintase (DHDPS), homoserina desidrogenase (HSDH) e a treonina deaminase (TD) (Fig. 1, linhas tracejadas) (Karchi et al., 1994; Cracium et al., 2000; Azevedo et al., 2005). Em bactérias, o ramo da via do ácido aspártico que leva a síntese de lisina é composto por nove enzimas, AK, aspartato-semialdeído desidrogenase (ASADH), DHDPS, dihidrodipicolinato redutase (DHDPR), tetrahidrodipicolinato acilase (THPA), N-acil-L,L-diaminopimelato aminotransferase (ADAPAT), N-acil-L,L-diaminopimelato deacilase (ADPD), diaminopimelato epimerase (DAPE) e diaminopimelato decarboxilase (DAPD), (Umbarger, 1978; Eggling, 1993; Azevedo et al., 2005). Em plantas, por sua vez, foram encontrados até o momento somente 5 enzimas semelhantes a via de síntese de lisina em bactérias (Hudson et al., 2005) e essas enzimas foram identificadas atuando nos cloroplastos de folhas ou em plastídios de órgão não fotossintéticos (Lea e Azevedo, 2003). Recentemente uma nova enzima, L-diaminopimelato aminotransferase, foi descoberta exclusivamente em vegetais e cianobactérias. Essa enzima é capaz de substituir os três passos enzimáticos que transformam tetrahidrodipicolinato em diaminopimelato (THPA,

ADAPAT, DAPD) (Hudson et al., 2006), elucidando a maneira pela qual a lisina é sintetizada em plantas.

A enzima AK catalisa o primeiro passo da via que envolve a fosforilação do ácido aspártico para formar β -aspartil fosfato (Fig. 1). Esse passo enzimático, assim como o seguinte, é comum para a síntese de todos os aminoácidos da via do ácido aspártico (Fig. 1). Pelo menos duas isoformas de AK são conhecidas em plantas, uma monofuncional sensível a lisina e responsável pela maior parte da atividade de AK (Rognes et al., 1980; Azevedo et al., 1997; Lugli et al., 2002) e uma bifuncional, aspartato quinase/homoserina desidrogenase (AK/HSDH) sensível a treonina (Wilson et al., 1991; Azevedo et al., 1992; Paris et al., 2003). Uma planta transgênica de alfafa contendo uma AK/HSDH insensível a inibição pela treonina mostrou aumento no teor de treonina e diminuição nas quantidades de aspartato e glutamato (Galili et al., 2000). Em plantas transformadas com AK/HSDH insensíveis para produtos finais lisina e treonina é possível identificar mais atividade de AK do que em plantas somente com AK insensível a lisina, comprovando que a isoenzima bifuncional regula a síntese de ambos os aminoácidos (Tewari-Singh et al., 2004).

O β -aspartil fosfato formado pela atividade da AK é convertido em aspartil-semialdeído (ASA) pela atividade da enzima aspartato semialdeído desidrogenase (ASADH) em uma reação dependente de NADPH (Fig. 1) (Paris et al., 2002; Azevedo et al., 2005). A partir desse passo a via do ácido aspártico é dividida em dois ramos, um que leva a síntese de lisina e outro que leva a síntese de treonina, metionina e isoleucina (Fig. 1) (Azevedo et al., 2005; Hudson et al., 2005).

A DHDPS é a primeira enzima do ramo da via do ácido aspártico que leva a síntese de lisina. A enzima catalisa a condensação de piruvato e ASA em dihidrodipicolinato (Fig. 1). Em plantas a DHDPS é inibida por baixas concentrações de lisina (Azevedo et al., 1997 ; Craciun et al., 2000; Lee et al., 2001). Isso faz com o que o nível de lisina seja muito baixo nos diversos órgãos das plantas, pois sua produção é feita em estrita conexão com a demanda (Azevedo et al., 1997 ; Lee et al., 2001). Já em bactérias a DHDPS só é inibida, por retro-inibição, por altas concentrações de

lisina (Cohen e Saint-Girons, 1987; Azevedo et al., 1997 ; Lee et al., 2001). Esse fato tem sido explorado para a obtenção de plantas transgênicas expressando a DHDPs de *E. coli* para acumular lisina (Perl et al., 1992; Falco et al., 1995 ; Brinch-Pedersen et al., 1996 ; Galili et al., 2000 ; Nancy et al., 2007 ; Frizzi et al., 2008 ; Reyes et al., 2009).

Uma série de 6 enzimas em bactérias (DHDPR, THPA, ADAPT, ADPD, DAPE, DAPD) e 4 em plantas (DHDPR, DAPAT, DAPE, DAPD) completam o ramo da via do ácido aspártico que leva a síntese de lisina transformando o tetrahidropicolinato em lisina através de uma série de reações de redução. A enzima DAPAT, recentemente descoberta em plantas, é capaz de sobre passar os três passos existentes em bactérias formando L-diaminopimelato a partir de tetrahidropicolinato (Hudson et al., 2006). Essa enzima é capaz de transferir o grupamento amina do glutamato para o tetrahidropicolinato na presença de amônia e NADPH, formando diaminopimelato. Em bactérias esses passos de redução e transaminação são feitos por enzimas diferentes com inserção e posterior retirada de um grupamento acil (Fig 1).

A via do ácido aspártico também é regulada por retro-inibição pela treonina e pela isoleucina (Fig. 1). O intermediário AdoMet também age sozinho na retro inibição da AK e da treonina sintase (TS) e em conjunto com a lisina na retro inibição da AK (Rognes et al., 1980) (Fig. 1). A inibição da AK pela lisina, pela lisina + AdoMet e pela treonina faz com que o nível celular desses compostos regulem o fluxo da via (Azevedo e Lea, 2001; Galili et al., 2002). Mutantes de várias espécies de plantas contendo AK insensível a retro-inibição pela lisina não mostraram um aumento significativo desse aminoácido sugerindo que a isoforma AK/HSDH possui atividade suficiente para manter o fluxo da via (Frankard et al., 1992). Além disso, a sensibilidade da DHDPs a retro-inibição pela lisina, poderia ser o fator limitante para o desbloqueio da via em direção ao acúmulo de lisina (Galili et al., 1995). Essa hipótese foi confirmada em plantas transgênicas contendo a enzima DHDPs insensível a retro-inibição pela lisina. Essas plantas apresentavam acúmulo significativo de lisina em vários órgãos (Negrutiu et al., 1984; Frankard et al., 1992; Shaul e Galili, 1992a; Ben

Tzvi-Tzchori et al., 1996). Por outro lado, em plantas mutantes ou transgênicas transformadas com AK insensível a lisina, notou-se também o acúmulo de treonina sugerindo que a DHDPs seria o ponto chave de controle do fluxo da via da lisina e que AK estaria mais relacionado a treonina. As plantas com AK insensível desviariam o fluxo para a síntese de treonina (Shaul e Galili, 1992b). Esses dados sugerem que a síntese de lisina é regulada majoritariamente pela inibição alostérica da DHDPs enquanto que a AK seria principalmente regulada pela treonina, apesar de também ser sensível a lisina (Galili, 2002; Azevedo et al., 2005).

Até o momento a síntese de lisina foi descrita em bactérias, fungos, plantas e alguns protozoários. Nem todos esses grupos compartilham as mesmas vias metabólicas para a síntese do aminoácido. Em fungos, por exemplo, a síntese de lisina ocorre pela via da sacaropina, que atua no sentido da degradação desse aminoácido nos demais organismos (Nishida e Nishiyama, 2000; Arruda et al., 2000; Azevedo et al., 2005; Galilli, 2004; Stepansky et al., 2006). A via do ácido aspártico por sua vez está presente numa grande diversidade de plantas e bactérias, mas até o momento nunca foi demonstrada de forma inequívoca em animais. Alguns estudos buscaram detectar a síntese de lisina em animais, entretanto, somente traços dessa síntese puderam ser encontrados em crustáceos e cnidários, mas os trabalhos não descartaram a hipótese de contaminantes nos modelos estudados (Cowey e Forsters, 1971; Lesser e Allen, 1976; Fitzgerald e Szmant, 1997). Até o momento não foi demonstrada a existência de genes codificadores das enzimas da via do ácido aspártico em animais.

O objetivo deste capítulo foi tentar responder as seguintes questões: A via do ácido aspártico realmente não está presente em nenhum animal? Porque a via metabólica do ácido aspártico não existe em animais? A via foi perdida ou nunca existiu? Qual sua origem? Como uma via importante para todos os demais organismos pode ser perdida em animais?

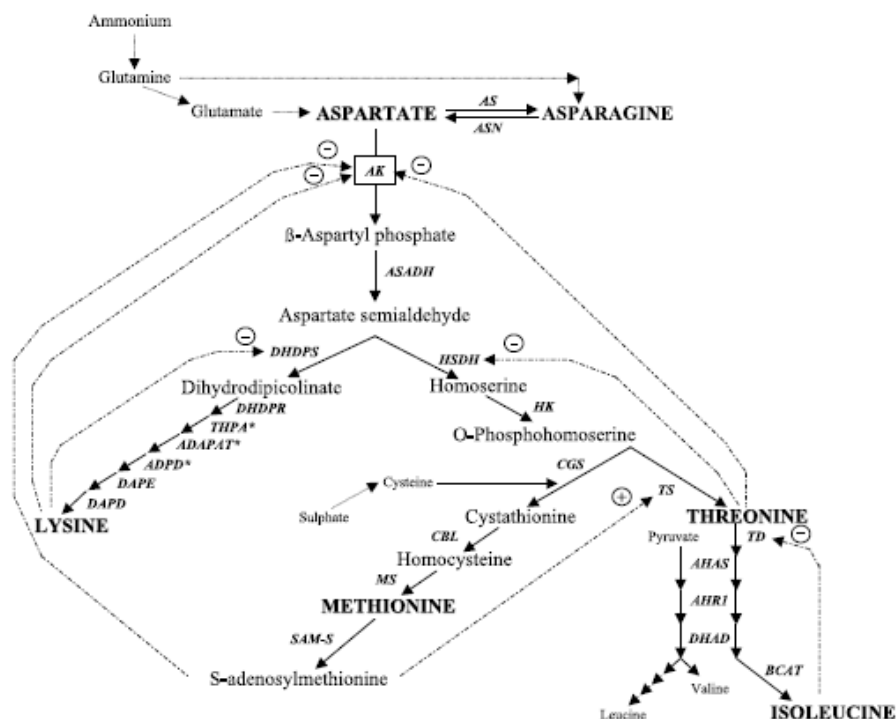


Figura 1 - Via metabólica do ácido aspártico em plantas. AS, asparagina sintetase; ASN, asparaginase; AK, aspartato quinase; ASADH, aspartato semialdeído desidrogenase; HSDH, homoserina desidrogenase; HK, homoserina quinase; CGS, cistationina-sintase; TS, treonina sintase; CBL, cistationina liase; MS, metionina sintase; SAM-S, S-adenosilmetionina sintetase; TD, treonina deaminase; AHAS, ácido acetohidroxi sintase; AHRI, ácido acetohidroxi isomeroreductase; DHAD, ácido dihidroxi desidratase; BCAT, aminotransferase dos aminoácidos da cadeia ramificada; DHDPS, dihidrodipicolinato sintase; DHDPR, dihidrodipicolinato redutase; THPA*, tetrahidrodipicolinato acilase; ADAPAT*, N-acil-L,L-diaminopimelato aminotransferase; ADPD*, N-acil-L,L-diaminopimelato deacilase; DAPE, diaminopimelato epimerase; DAPD, diaminopimelato decarboxilase (extraído de Azevedo et al, 2005)

Resultados

As sequências de aminoácidos das nove enzimas do ramo da via do ácido aspártico que leva a síntese de lisina em bactérias (Velasco, 2002) e das sete enzimas correspondentes em plantas (Hudson et al., 2006), foram utilizadas como “iscas” em buscas para identificar possíveis genes ortólogos, codificadores dessas enzimas, no genoma de *Aedes aegyptis* e *Anopheles gambiae*. A lista das enzimas com os respectivos GIs dos genes utilizados como “iscas” é mostrada na Tabela 1. As sequências “iscas” foram utilizadas em buscas no GenBank na seção de genomas específicos de cada organismo. As buscas foram feitas utilizando-se a ferramenta tBLASTn e, quando necessário, foram feitas buscas escalonadas como descrito por Kasahara e Kato (2003). A utilização de sequências “iscas” oriundas de uma planta e uma bactéria teve como objetivo abordar a hipótese dos genes compartilharem origem similar com um ou outro grupo. Essa abordagem possibilitou também capturar nos bancos de dados possíveis alterações de sequência devido à evolução convergente.

Tabela 1 - Enzimas da via do ácido aspártico cujas sequências de aminoácidos foram utilizadas como “iscas” na busca de ortólogos em insetos.

Enzima	Número EC	GI <i>A. thaliana</i> ¹	GI <i>E. coli</i> ²
Aspartato quinase (AK)	2.7.2.4	2257742	48994873
Aspartato-semialdehyde dehydrogenase (ASDH)	1.2.1.11	42592260	24080789
Dihidrodipicolinato sinthase (DHDPS)	4.2.1.52	30698031	24080789
Dihidrodipicolinato reductase (DHDPR)	1.3.1.26	30698605	48994873
Diaminopimelato aminotransferase (DAPAT)	2.6.1.83	18418270	*
Tetrahidropicolinato succinilase (THPA)	2.3.1.17	#	24080789
Succinildiaminopimelato aminotransferase (ADAPAT)	2.6.1.17	#	48994873
Succinyi-diaminopimelate desuccinilase (ADPD)	3.5.1.18	#	24080789
Diaminopimelato epimerase (DAPE)	5.1.1.7	16504487	91070629
Diaminopimelate decarboxilase (DAPD)	4.1.1.20	30698537	48994873
# Enzimas que não possuem evidência de existirem em plantas (Hudson et al., 2005)			
* Enzima exclusiva de plantas (Hudson et al., 2006)			
1 GI's atribuído a cada enzima de <i>A. thaliana</i> no GenBank			
2 GI's atribuído a cada enzima de <i>E. coli</i> no GenBank			

Identificação de possíveis genes que codificam enzimas da via do ácido aspártico em insetos

Buscas utilizando tBLASTn com as sequências “iscas” nos genomas de *A. gambiae* e *A.aegypti* retornaram um conjunto significativo de possíveis sequências homólogas. Sequências com mais de 50% de cobertura e identidade superior a 30% com as sequências “iscas” dos genes correspondentes foram separadas para análise. A Tabela 2 apresenta as enzimas para as quais foram identificadas sequências de aminoácidos com similaridade significativa e, portanto, foram considerados como possíveis homólogos dos genes da via de síntese de lisina em *A. gambiae* e *A.aegypti*. Como pode ser visto, com exceção da AK, encontrada apenas em *A.aegypti*, as enzimas ASDH, DHDPs, DHDPR, DAPE, DAPD, foram encontradas em ambos os insetos. A maioria dos “hits” que obtiveram e-value mais significativos originaram-se de buscas feitas com as sequências “iscas” de *E. coli*. Devido ao limite de corte em cobertura e similaridade, não é possível afirmar que foram encontradas sequências possivelmente homólogas para as enzimas THPA, ADAPAT, ADPD, em *A. gambiae* e *A. aegypti* apesar de termos obtido resultados com relativa similaridade entre as sequências “iscas” e sequências dos bancos dos dados genômicos dos insetos. A ausência dessas enzimas, típicas de bactérias, em *A. gambiae* e *A.aegypti* sugere a possibilidade do banco de dados analisados não estarem contaminados com sequências de bactérias para as enzimas da via do ácido aspártico. Todavia, o gene exclusivo de plantas DAPAT também não foi retornou “hits” significativos dos bancos genômicos dos insetos, sendo assim difícil especular sobre a origem dos “hits” a partir dessas informações isoladas.

Tabela 2 - Enzimas da via do ácido aspártico cujas sequências de aminoácidos foram identificadas como possíveis ortólogas em insetos

Enzima	Número EC	A. <i>gambie</i>	A. <i>aegypti</i>
Aspartato Quinase (AK)	2.7.2.4	+	-
Aspartato-semialdeído desidrogenase (ASDH)	1.2.1.11	+	+
Dihidrodipicolinato sintase (DHDPS)	4.2.1.52	+	+
Dihidrodipicolinato redutase (DHDPR)	1.3.1.26	+	+
Diaminopimelato aminotransferase (DAPAT)	2.6.1.83	-	-
Tetrahidropicolinato succinilase (THPA)	2.3.1.17	-	-
Succinil-diaminopimelato aminotransferase (ADAPAT)	2.6.1.17	*	*
Succinil-diaminopimelate desuccinilase (ADPD)	3.5.1.18	*	*
Diaminopimelato epimerase (DAPE)	5.1.1.7	+	+
Diaminopimelate decarboxilase (DAPD)	4.1.1.20	+	+
+ Presença de sequências de aminoácidos ortólogas em <i>Anopheles</i> e <i>Aedes</i>			
- Ausência de sequências de aminoácidos ortólogas em <i>Anopheles</i> e <i>Aedes</i>			
* Incerteza quanto à anotação do gene			

As sequências de aminoácidos possivelmente homólogas as enzimas do ramo da lisina de bactérias e plantas foram encontradas nos bancos com sequências genômicas de *A. gambie* e *A.aegypti* (Tabela 2). Uma avaliação preliminar da possibilidade de expressão desses genes foi realizada através de BLASTn das sequências identificadas (Tabela 2) em bancos de ESTs de ambos os insetos (<http://agambiae.vectorbase.org/SequenceData/EST/> <http://aaegypti.vectorbase.org/SequenceData/EST/>). Além de possibilitar a verificação da expressão dos genes correspondentes, a existência de ESTs possibilita, quando comparada com a sequência genômica, inferir a existência de introns nos genes identificados. Para complementar a análise dos genes encontrados nos dois insetos, buscamos pelo posicionamento cromossômico das sequências genômicas encontradas. A Tabela 3 resume os dados, de posicionamento cromossômico de alguns dos genes do ramo da lisina encontrados em *A. gambie*. O genoma de *A.aegypti* ainda não está bem curado e alguns segmentos cromossômicos ainda não estão resolvidos, fato esse que prejudicou a investigação do posicionamento

cromossômico e presença de introns. Em *A. gambie* observa-se a existência de intron no gene DHDPR, fato que enfraquece a hipótese de contaminação, já que bactérias não possuem introns.

Tabela 3 - Características das sequências dos genes da via de síntese de lisina em *A. gambie*

Enzima	Posicionamento cromossômico	Existência de cDNA	Existência de Introns
AK	Desconhecido	Sim	Não
ASDH	Desconhecido	Sim	Não
DHDPS	2L	Sim	Não
DHDPR	Desconhecido	Sim	Sim
DAPE	Desconhecido	Sim	Não
DAPD	3L	Sim	Não

A possibilidade de contaminação das sequências de insetos com sequências de outros organismos é uma preocupação a ser levada em conta na anotação gênica principalmente no caso de bactérias que possuem íntima relação com esses animais seja habitando o trato intestinal ou colonizando o interior de células gonadais na forma de endosimbiontes. Visando excluir a possibilidade de contaminação foram realizadas buscas nos bancos de dados supracitados de transcritos que passaram por purificação através da calda poliA do mRNA (Tabela 4). Acredita-se que RNAs poli-adenilados sejam raros em bactérias já que quando um RNA é poliadenilado ele é sinalizado para a degradação nesse grupo de organismo (Steeg, 2000). Os resultados obtidos mostram “hits” para a maioria dos genes procurados e os transcritos estão presentes em número variado. Esses dados constituem mais um indicativo de que esses genes provavelmente são genes pertencentes ao genoma desses insetos e não contaminantes bacterianos.

Tabela 4 - Número de transcritos encontrados no banco de dados do portal Vectorbase para os genes da via de biosíntese de lisina

Gene	<i>A.gambie</i>	<i>A. aegypti</i>
Ak	0	0
Asdh	0	0
Dhdps	20	7
Dhdpr	0	1
Dape	43	0
Dapd	10	6

Número total de ESTs no banco de *A. gambie* = 216.625

Número total de ESTs no banco de *A. aegypti* = 303.409

Ainda buscando reforçar a exclusão da possibilidade de contaminação, foi realizada uma busca no NCBI utilizando a ferramenta BLASTn tendo como “iscas” as sequências nucleotídicas dos possíveis genes da via de síntese de lisina em *A. gambie* e *A.aegypti* e como universo de busca todas sequências de bactérias. Essa busca visava identificar sequências idênticas em outros organismos que poderiam ser considerados contaminantes dos bancos de dados e que podem ter dado origem às sequências encontradas em insetos. A Tabela 5 mostra os organismos com os melhores resultados na busca. Nota-se que não há identidade de 100% em nenhum caso, o que significa que a sequência considerada homóloga as sequências dos genes da via do ácido aspártico não pertencem a nenhum organismo cujas sequências genômicas, completas ou parciais, estão depositadas no Genbank .

Tabela 5 - Resultado da busca por BLASTn no banco de dados de todas sequências nucleotídicas do NCBI

Enzima	<i>A. gambie</i>			<i>A. aegypti</i>		
	Organismo	e-value	identidade	organismo	e-value	identidade
AK	<i>Flavobacterium johnsoniae</i>	2e-60	70%	-		
ASDH	<i>Gluconobacter oxydans</i>	1e -115	80%	<i>Thauera sp</i>	0	96%
DHDPS	<i>Flavobacterium johnsoniae</i>	2e-86	70%	<i>Flavobacterium johnsoniae</i>	2e-87	72%
DHDPR	<i>Flavobacteriaceae bacterium</i>	5e-59	76%	<i>Thauera sp</i>	6e-126	96%
ADPD	<i>Clavibacter michiganensis</i>	7e-120	91%	<i>Aeromonas hydrophila</i>	0	97%
DAPE	<i>Aeromonas hydrophila</i>	7e-140	95%	<i>Thauera sp</i>	1e-120	88%
DAPD	<i>Azoarcus SP</i>	1e-118	70%	<i>Dechloromonas aromatica</i>	9e-130	87%

A inexistência de sequências idênticas às encontradas em insetos em outros organismos sugere ausência de contaminação.

Para testar de outra forma se as sequências do ramo da lisina encontradas nos dois insetos poderiam ser “reads” oriundos de contaminantes bacterianos da amostra de ácidos nucleicos isolados dos insetos, investigamos as adjacências dos genes posicionados em contigs no genoma dos insetos presentes nos bancos de dados. Na hipótese de tratar-se de contaminação as sequências apareceriam em sua maioria como sequências únicas e isoladas contendo apenas um gene por contig além de poucos reads para esse gene. Por outro lado, no caso de uma inserção por transferência lateral, essas sequências estariam flanqueadas por mais sequências bacterianas e possivelmente já com traços de genomas eucariotos. Ainda, seria possível sugerir de qual grupo de bactérias essas sequências poderiam ter sido originadas. A Figura 2 mostra o resultado encontrado no supercontig do gene DHDPS de *A. aegypti*. Adjacentes ao gene da DHDPS foram identificados os genes

bacterianos que codificam as proteínas GCN5-acetiltransferase (GNAT) e flavina mononucleotídeo redutase (FNR) (Fig. 2). As sequências desses dois genes parecem ser oriundas do grupo Flavobacteriaceae. Nota-se também, que a porção do nucleotídeo 1 até o início do gene DHDPs não teve similaridade com sequências procarióticas sugerindo serem tipicamente sequência eucariótica do próprio animal. Não foi possível realizar a busca para outros genes porque eles foram retirados dos contigs que se encontravam pelo curador do NCBI até ser resolvido se se tratam de genes de insetos ou contaminantes.

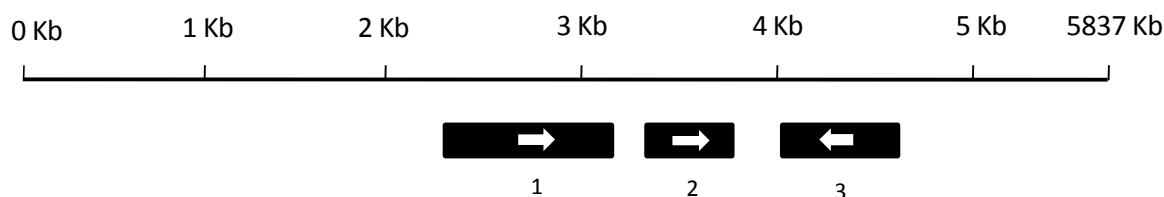


Figura 2: Estrutura do supercontig 1.3837 de *Aedes aegypti*. 1, DHDPs; 2, GNAT-acetiltransferase; 3, flavine mononucleotide redutase. Todos os genes citados tem maior identidade com a de sequência de *Flavobacterium Johnsoniae*, indicando que esse pode ter sido o grupo de bactéria que deu origem a esses genes de *Aedes aegypti*.

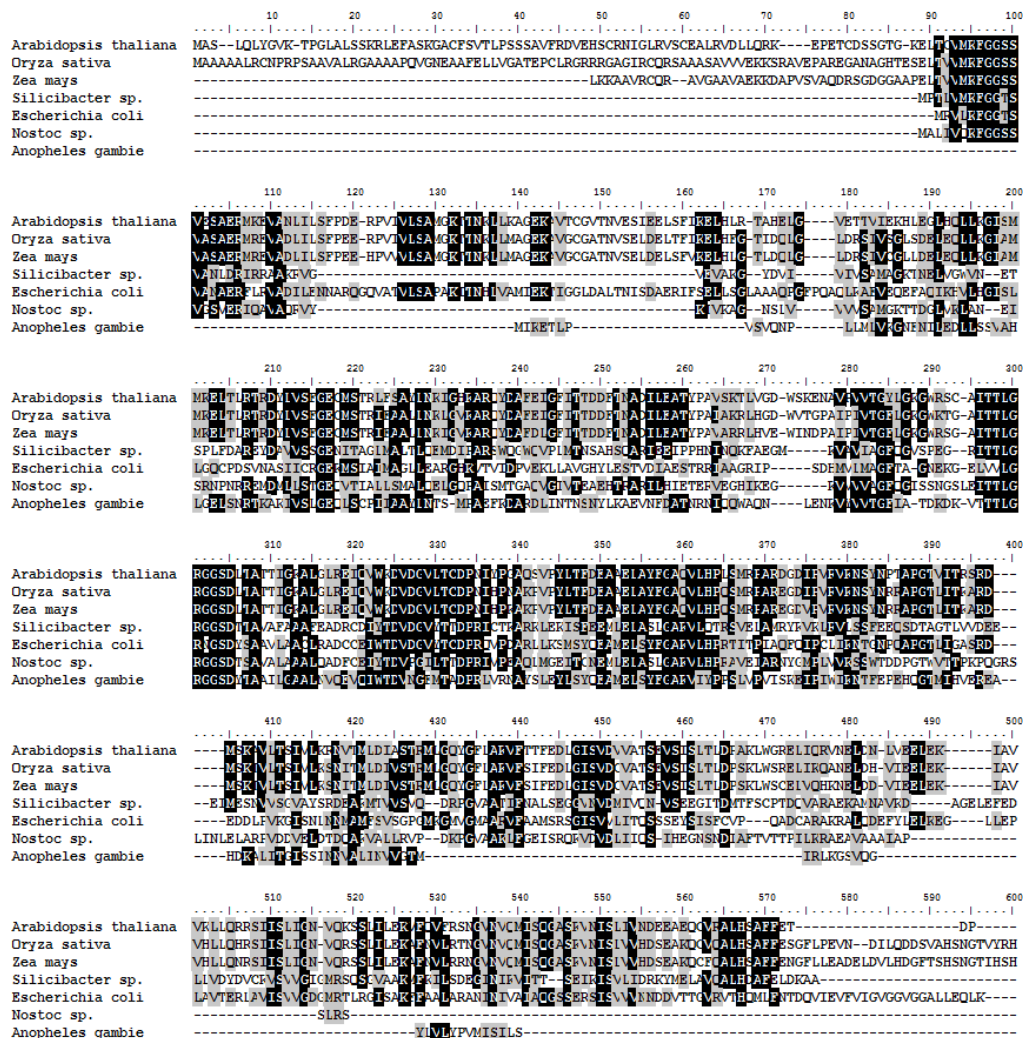
Análise das sequências de aminoácidos das enzimas do ramo da lisina encontradas em A. gambie e A.aegypti

Uma vez identificadas nos dois insetos sequências possivelmente homólogas as sequências de enzimas do ramo da lisina de bactérias e plantas, foram realizados alinhamentos dessas sequências com as de bactérias e plantas para investigarmos a presença de domínios conservados e a relação filogenética entre as sequências dos diferentes organismos. As Figuras 3 a 8 mostram os alinhamentos das sequências de aminoácidos das enzimas AK, ASDH, DHDPs, DHDPR, DAPE e DAPD de plantas, bactérias, *A. gambie* e *A. aegypti*. Os alinhamentos apresentam padrões de motivos

conservados para cada uma das enzimas de planta, bactéria e insetos que sugerem ser plausível a anotação dessas enzimas de insetos como sendo possíveis ortólogos das enzimas que compõem o ramo da síntese de lisina da via do ácido aspártico em plantas e bactérias. As sequências com baixa similaridade foram excluídas dos alinhamentos.

As relações filogenéticas entre as sequências dos diferentes organismos foram obtidas dos alinhamentos mostrados nas Figuras 3A a 8A utilizando o método “neighbor-joining” (Saitou e Nei, 1987) e modelo PAM de Dayhoff (Schwartz e Dayhoff, 1979). As sequências identificadas em insetos foram, invariavelmente, sempre mais próximas aquelas de bactérias do que de plantas. Isso pode ter várias implicações, mas duas parecem mais pertinentes ao nosso estudo. A primeira, é que as sequências identificadas em *A. gambiae* e *A. aegypti* sejam oriundas de contaminantes bacterianos, estejam estes presentes no trato intestinal dos insetos amostrados para a realização do seqüenciamento ou na forma de endosimbiontes, comuns em insetos (Wernegreen, 2002). Na segunda hipótese, os genes encontrados nos dois insetos poderiam ter sido adquiridos através de transferência lateral a partir de bactérias, como ocorrido no caso demonstrado entre *Wolbachia* e *Drosophila* (Kondo et al., 2002), cenário esse que nos parece mais crível.

A



B

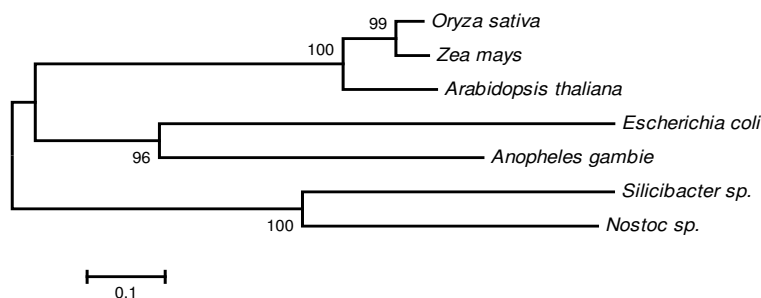
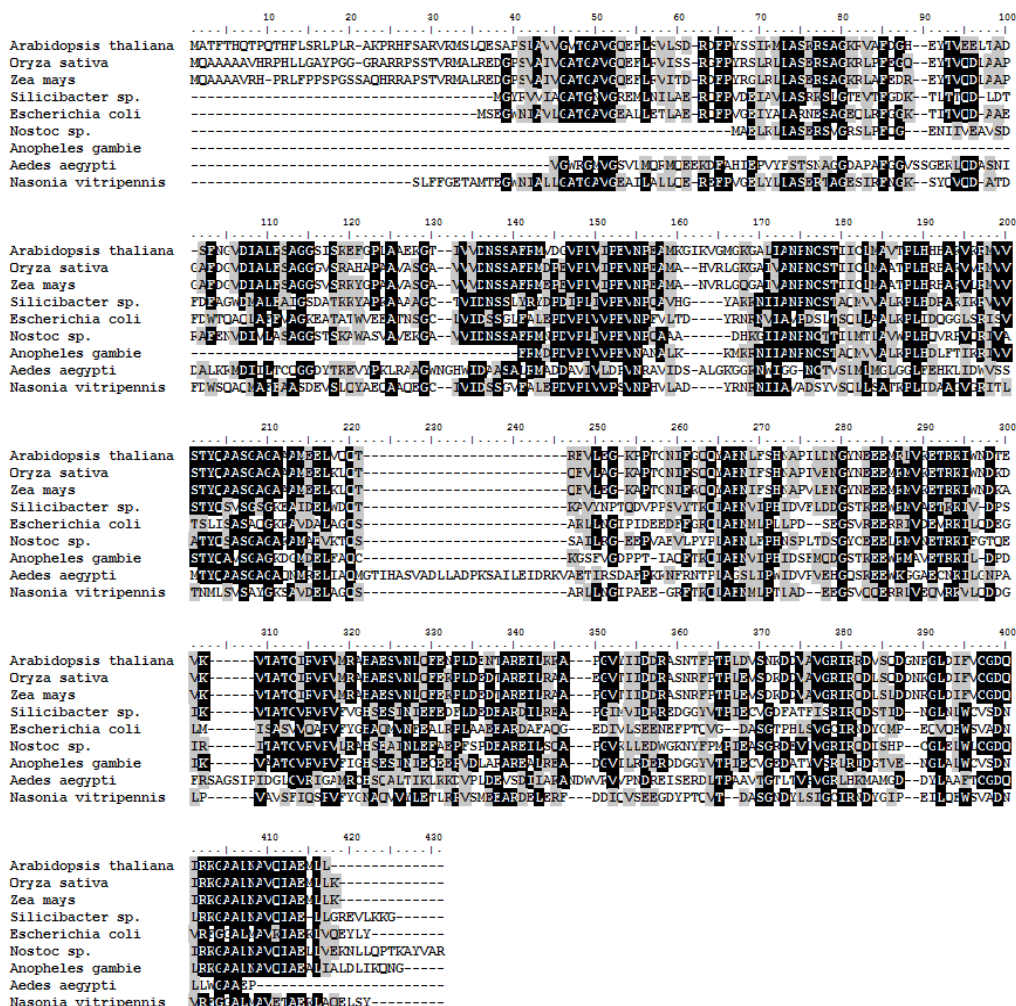


Figura 3 - (A) Alinhamento das sequências de aminoácidos e **(B)** árvore filogenética das sequências de plantas, bactérias e *Anopheles* para a enzima AK.

A



B

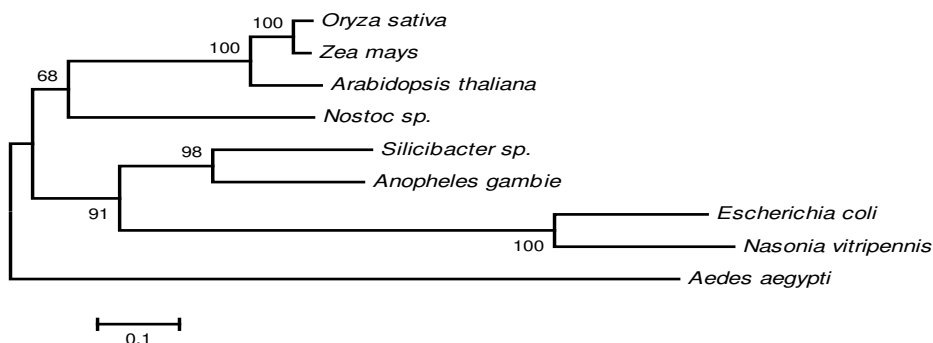
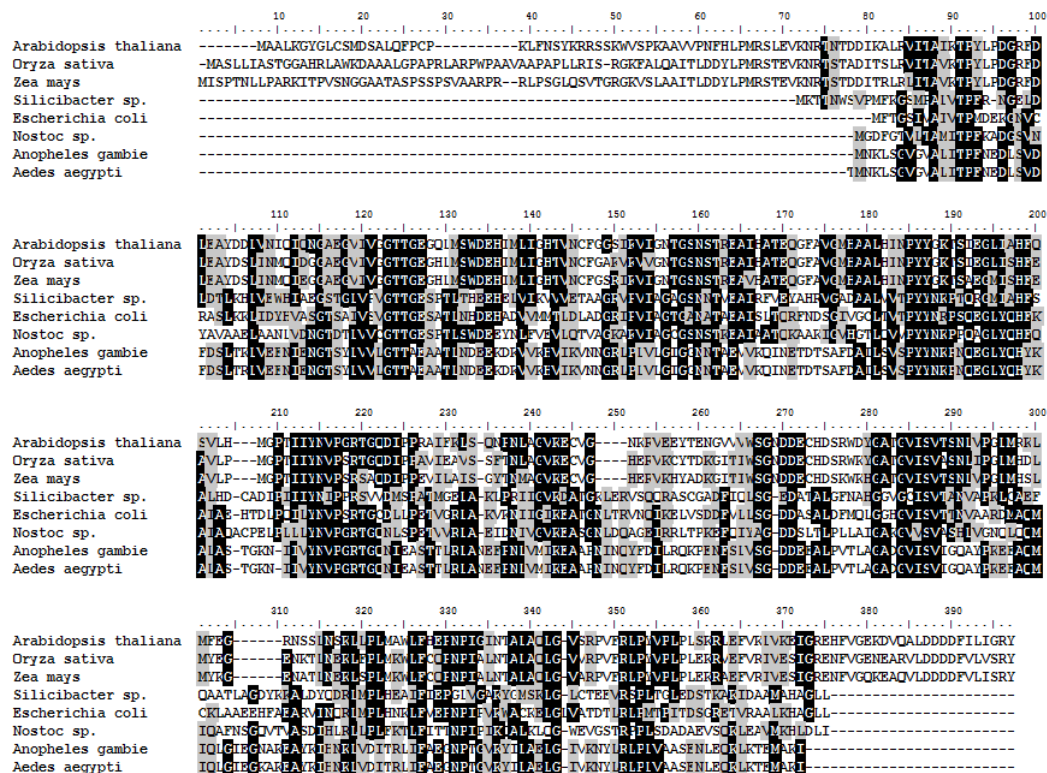


Figura 4 - (A) Alinhamento das sequências de aminoácidos e **(B)** árvore filogenética das sequências de plantas bacterias e *A. gambiae* para a enzima ASDH.

A



B

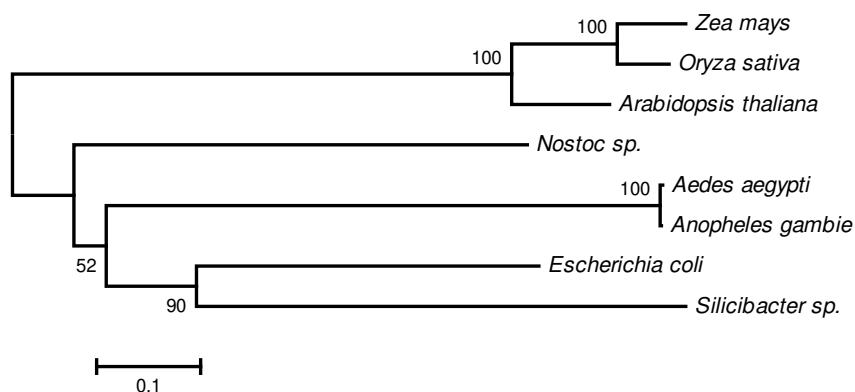
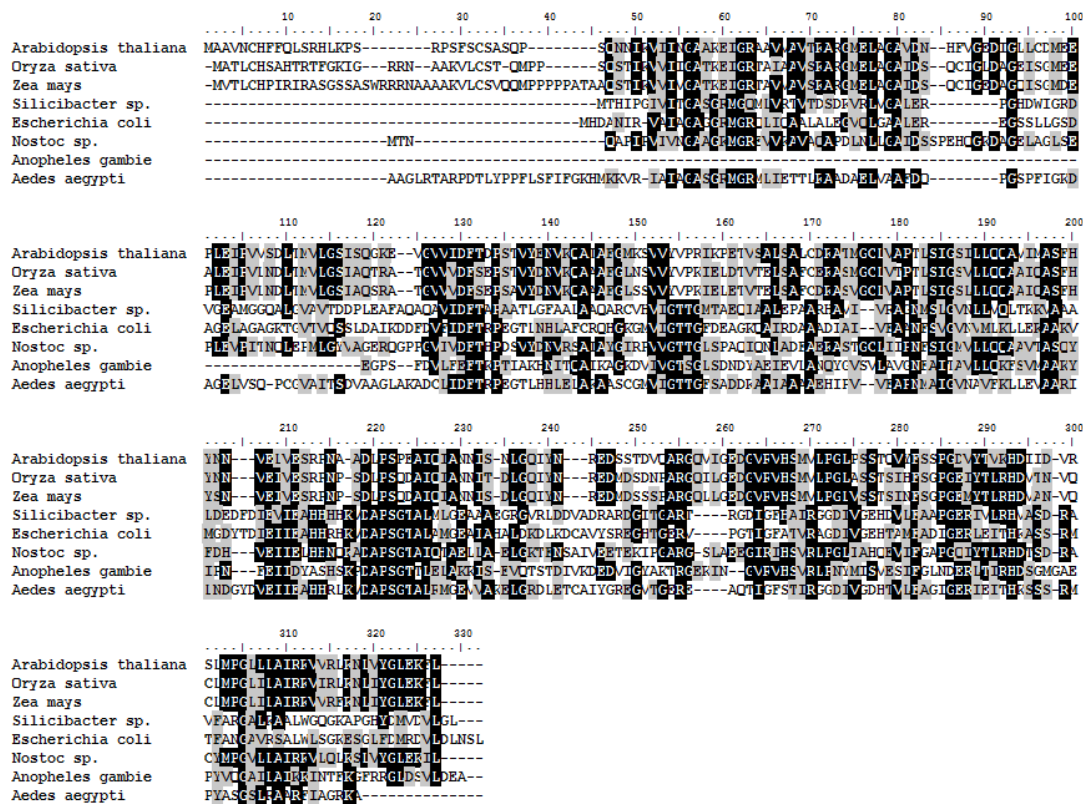


Figura 5 - (A) Alinhamento das sequências de aminoácidos e **(B)** árvore filogenética das sequências de plantas, bactérias e *Anopheles* para a enzima DHDPS.

A



B

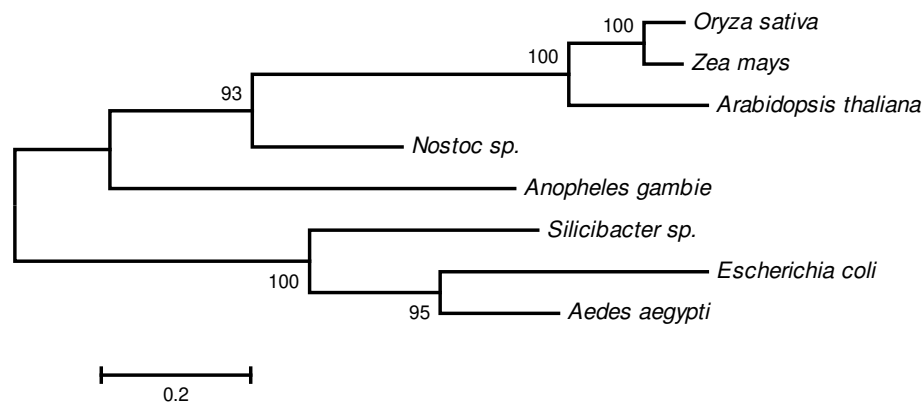
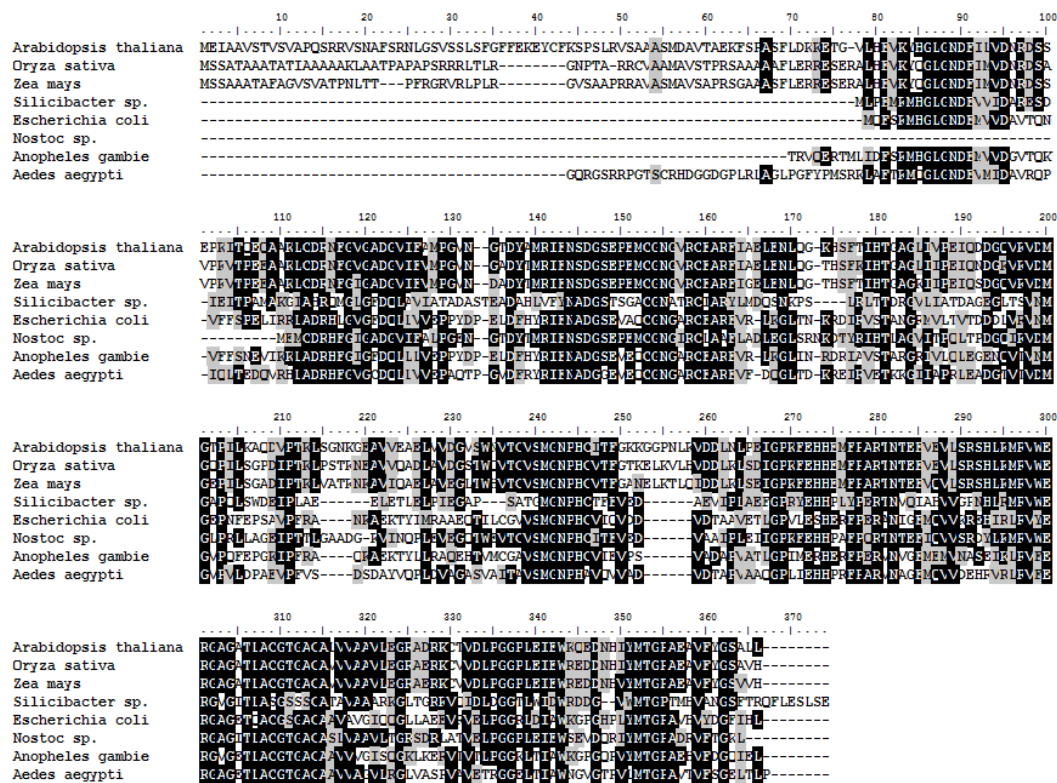


Figura 6 - (A) Alinhamento das sequências de aminoácidos e **(B)** árvore filogenética das sequências de plantas bactérias e *Anopheles* para a enzima DHDPR.

A



B

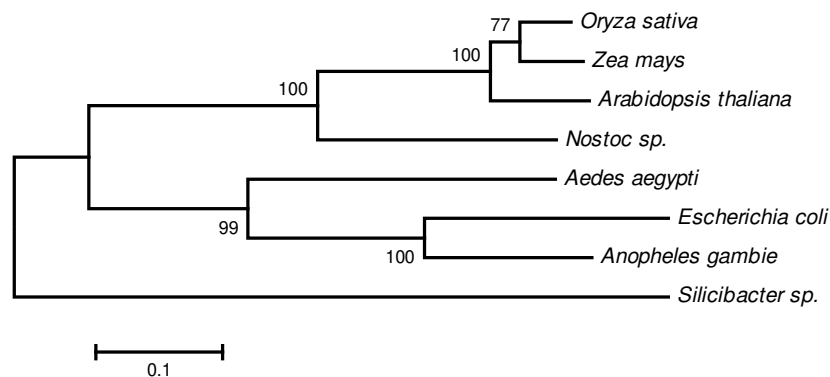
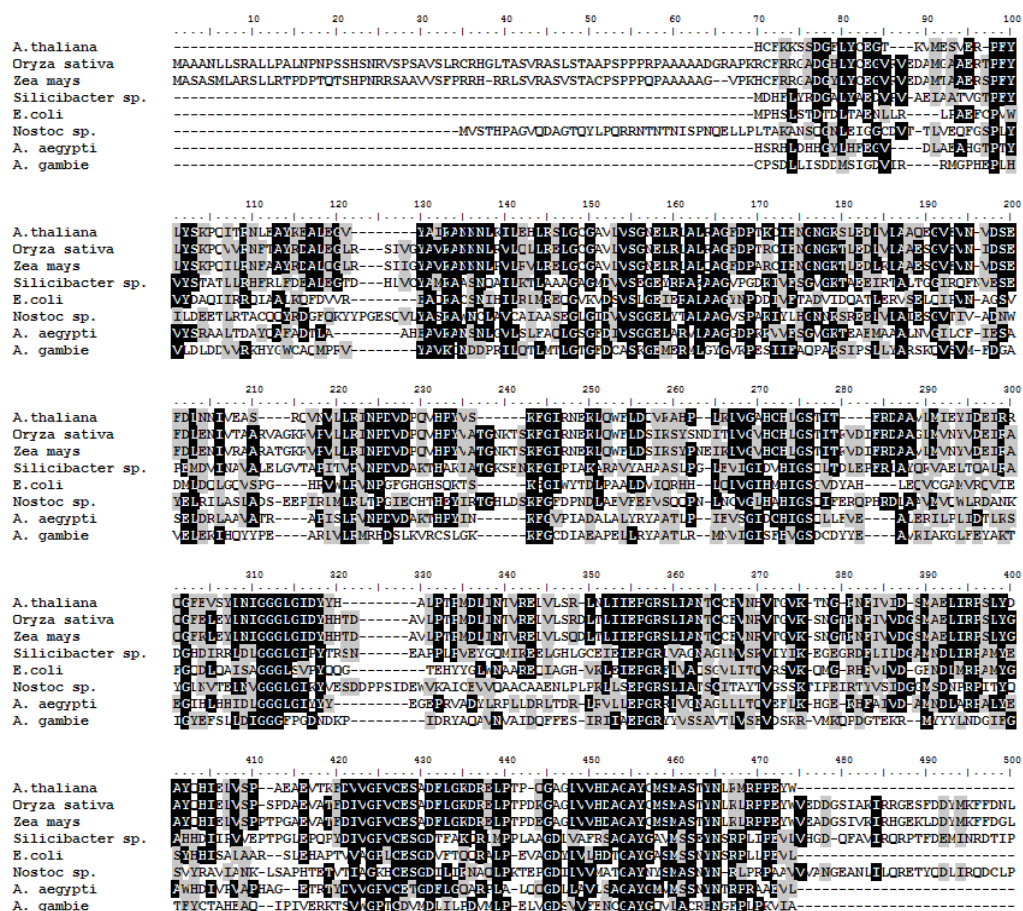


Figura 7 - (A) Alinhamento das sequências de aminoácidos e **(B)** árvore filogenética das sequências de plantas bactérias e *Anopheles* para a enzima DAPE

A



B

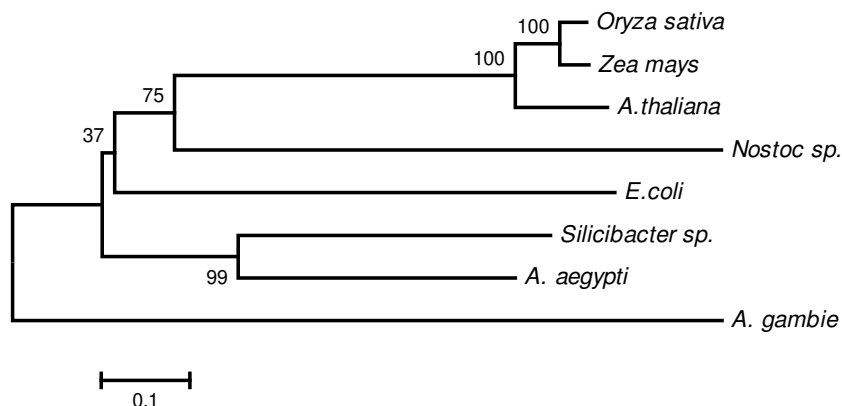


Figura 8 - (A) Alinhamento das sequências de aminoácidos e **(B)** árvore filogenética das sequências de plantas bacterias e *Anopheles* para a enzima DAPD.

A análise das sequências de aminoácidos das possíveis enzimas da via do ácido aspártico encontradas em *A. gambie* e *A. aegypti* revelou, quando essas foram submetidas a buscas no banco de domínios pfam (Finn et al., 2008), que elas compartilham os domínios conservados comuns para essas enzimas de plantas e bactérias. Assim, os domínios conservados encontrados na maioria das enzimas corroboram os dados de alinhamento. Entretanto, verificamos que alguns domínios não foram identificados devido ao fato de não encontrarmos todas as sequências completas de todos os genes nos bancos de dados. Em outros casos mais de um domínio foi encontrado para uma mesma sequência de aminoácidos, dificultando a caracterização por introduzir imprecisão na determinação do subgrupo a que a enzima pertence. Por exemplo, domínios com funções generalistas tais como aminotransferase, redutase, decarboxilase aparecem, mas sem a identificação específica do substrato. Ainda assim, a análise das sequências mostrou que os domínios conservados possuem similaridade superior a 90% com o consenso padrão do banco de dados, o que reforça a anotação dessas sequências como potenciais homólogos das enzimas da via do ácido aspártico. (Tabela 6).

Algoritmos que predizem a estrutura tridimensional de proteínas também foram utilizados como ferramenta para incrementar a anotação das enzimas de insetos. Foram modeladas automaticamente estruturas com base em proteínas com identidade superior a 30%. Em alguns casos, como por exemplo, para a enzima DHDPS e DHDPR foi possível confirmar alta similaridade de estrutura com aquelas de *A. thaliana* e *E. coli* (dado não apresentado)

Tabela 6 - Similaridade entre os domínios de síntese de lisina com o banco de dados de domínios conservados cdd e pfam.

Enzima	<i>A. thaliana</i>	<i>E. coli</i>	<i>A. gambie</i>	<i>A. aegypt</i>
AK	99,1	99,5	97,1	-
AK-HSDH	99,7/94,2	95,7/100	97,0/92,1	96
ASDH	98,8	100	100	98
DHDPS	99,2	96,8	100	99
DHDPR	99,2	99,2	99,2	99
THPA	#	100	-	-
ADAPAT	#	93,1	100	-
ADPD	#	96,8	98,0	-
DAPE	100	100	100	90
DAPD	99,4	100	93,9	90
Os números representam a porcentagem de similaridade com o consenso				
# enzimas que não foram descritas em plantas				
- Enzimas que não forma possível identificar				

Amplificação dos genes a partir do DNA genômico dos insetos

“Primers” específicos para cada um dos genes que codificam enzimas da via do ácido aspártico encontrados em *A. gambie* e *A. aegypti* foram desenhados e utilizados em reações de amplificação por PCR utilizando com molde o DNA genômico desses insetos tratados por várias gerações em meio contendo antibiótico. O tratamento com antibiótico tem como objetivo eliminar possíveis contaminantes bacterianos. Para confirmar a ausência de DNA bacteriano nas amostras de DNA genômico extraídas dos insetos foram utilizados “primers” para o gene ribossomal 16S bacteriano em reação em cadeia da polimerase. Esses primers são comumente utilizados para amplificação do fragmento genômico 16S de uma ampla gama de bactérias (Xiang, 2006). As amostras de DNA genômico de bactérias e de *Anopheles gambie*, quando analisadas através de eletroforese em gel de agarose mostraram-se de boa qualidade sem indícios de degradação que pudessem afetar as reações de PCR com os “primers” específicos de cada organismo (Fig. 9A). A amplificação com primers 16S

resultou em uma banda típica de 180 pb observada para os DNAs extraídos das bactérias *E. coli* (canaleta 2), *X. fastidiosa* (canaleta 3), *Xanthomonas sp* (canaleta 4) e *Silicibacter SP* (canaleta 5) (Fig. 9B). Nas canaletas 6 e 7 estão os produtos amplificados com primers 16S dos DNAs genômicos de *Anopheles gambiae* não tratado e tratado com antibiótico, respectivamente (Fig.9B). Como pode ser visto, ocorreu amplificação do 16S nas amostras não tratadas com antibiótico, indicando a presença de DNA bacteriano possivelmente oriundo de bactérias presentes no trato intestinal ou coabitando o inseto na forma de endossimbionte (Fig.9B, canaleta 6). Porém, não houve produto de amplificação nas amostras tratadas com tetraciclina indicando que o tratamento com antibiótico possivelmente eliminou as bactérias do inseto (Fig. 9B, canaleta 7)

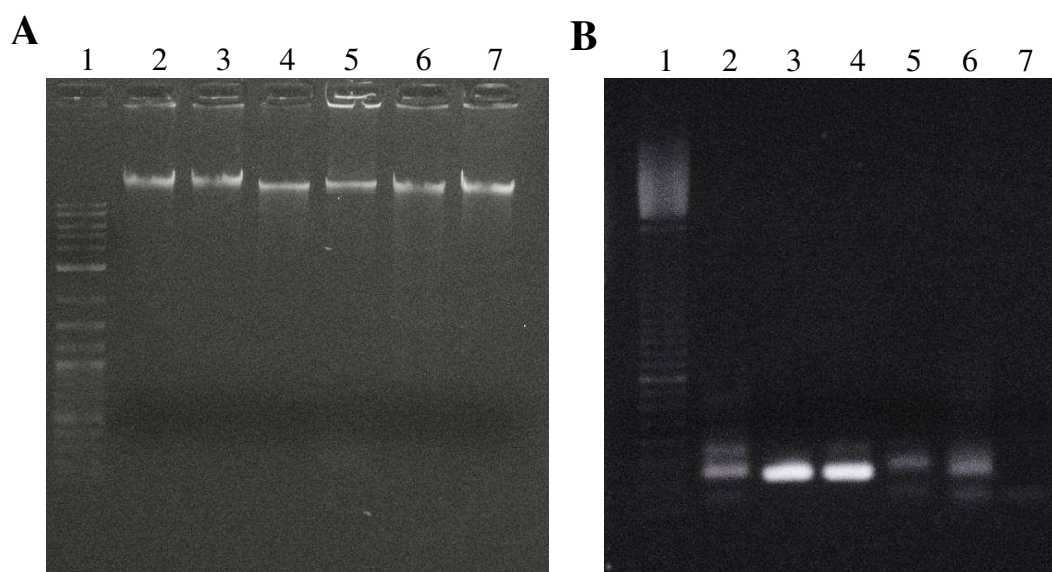


Figura 9 - (A) eletroforese em gel de agarose do DNA genômico de bactérias e *Anopheles gambiae*. 1, padrão de peso molecular 1Kb; 2, *E. coli*; 3, *X. fastidiosa*; 4, *Xanthomonas sp*; 5, *Silicibacter sp*; 6, *Anopheles gambiae* não tratado com antibiótico; 7, *Anopheles gambiae* tratado com tetraciclina por varias gerações. (B) amplificações com primers 16S bacteriano das amostras apresentadas em (A).

As amplificações utilizando DNA genômico de *A. gambiae* como molde foram realizadas na Universidade da Califórnia, San Diego, já que não temos acesso a essa espécie no Brasil. DNA genômico dos insetos não tratados e tratados por 3 gerações com tetraciclina foram utilizados para amplificação com primers específicos dos genes que codificam as enzimas da via do ácido aspártico identificados. Foi possível identificar bandas com tamanhos similares ao esperado para os genes de que codificam as enzimas AK, ASDH, DHDPS, DHDPR e DAPD (Fig 10). O sequenciamento desses amplificados confirmaram a identidade dos sequências obtidas nos bancos de dados com exceção de AK e DAPD que apresentaram sequência diferente, porém ainda com alta identidade na comparação com as identificadas neste trabalho.

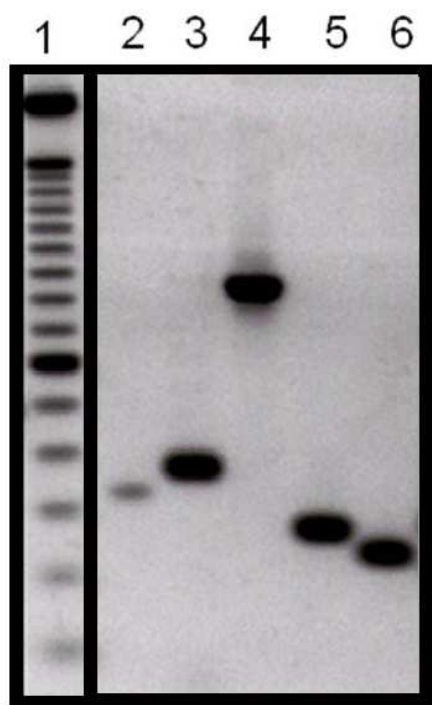


Figura 10 – Eletroforese em gel de agarose com dos fragmentos amplificados dos genes que codificam enzimas da via do aspartato em *Anopheles gambiae* tratado com tetraciclina. 1, padrão de peso molecular 100pb; 2, AK; 3, ASDH; 4, DHDPS; 5, DHDPR; 6, DAPD.

Em *Aedes aegypti* a maioria das amplificações com primers específicos das enzimas AK, ASDH, DHDPS, DHDPR, DAPE e DAPD a partir do DNA genômico não tratado e tratado com antibiótico, não produziram fragmentos amplificados. Somente para a enzima DHDPS foi possível observar um amplicon de ~1Kb nas duas amostras (Fig. 11, canaletas 3 e 4). Isso possivelmente deve-se ao fato de que as a maioria das sequências encontradas nos bancos de dados são cópias únicas (singletons), o que leva a grande imprecisão na determinação da real sequência do gene e dificuldade de desenhar os primers corretamente.

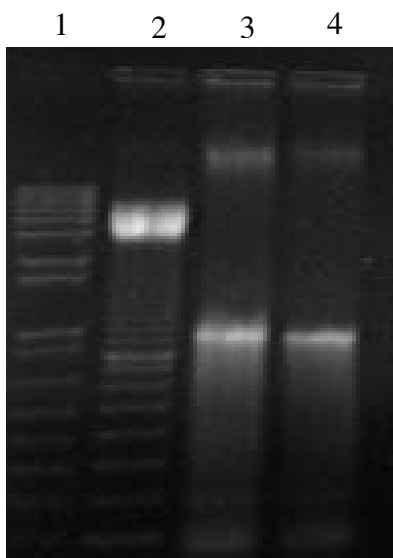


Figura 11 - Eletroforese em gel de agarose dos fragmentos amplificados do gene que codifica a enzima DHDPS de *Aedes aegypti*. 1 e 2, padrões de peso molecular 1kb e 1 kb plus; 3, *Aedes* não tratado com tetraciclina; 4, *Aedes* tratado com tetraciclina por várias gerações.

Discussão

A genômica e a bioinformática foram utilizadas na tentativa de se identificar e caracterizar nos insetos *A. gambiae* e *A. aegypti* sequências de genes que codificam enzimas potencialmente homólogas a enzimas de bactérias envolvidas na síntese de lisina pela via do ácido aspártico.

Foram realizadas buscas utilizando-se como “iscas” as sequências de aminoácidos das enzimas da via do ácido aspártico encontradas em *A. thaliana* e *E. coli*. As buscas utilizaram basicamente as ferramentas de BLAST onde a sequências “iscas” foram confrontadas com sequências genômicas e ESTs dos dois insetos. Em alguns casos foram feitas buscas escalonadas, pois nem sempre buscas com nossas iscas traziam como resultado relevante insetos, mas sim outros organismos como protozoários e a partir desses com novas buscas era possível encontrar as sequências pretendidas. Essa estratégia, bem descrita por Kasahara e Kato em 2003, é muito eficaz, principalmente quando se tem sequências com alta divergência ou de tamanhos muito distintos.

Sequências que apresentam similaridade superior a 30% ou e-value < e-10 têm sido consideradas como potenciais a homólogas apesar de valores inferiores de corte serem utilizados em diversos casos (Reese et al., 2000). Tomando-se esse fato como premissa, todas as sequências provenientes das buscas utilizando tBLASTn com valores iguais ou superiores aos valores de corte acima, foram consideradas como potencialmente homólogas das sequências “iscas” (Tabela 2). Essas sequências foram analisadas uma a uma em busca de domínios conservados como sugerido por (Marchler-Bauer et al., 2009). Analisando os domínios das sequências de insetos selecionadas (Tabela 2) observamos similaridade superior a 90% quando comparados com domínios conservados depositados no pfam e cdd para todos os domínios, indicando mais uma vez tratarem-se de proteínas homólogas.

Os alinhamentos das sequências das enzimas da via do ácido aspártico encontradas nos insetos com aquelas de plantas e bactérias (Fig. 3-8) observa-se que de fato não somente as proteínas apresentam alta similaridade em aminoácidos entre

elas como também esses aminoácidos estão organizados em blocos conservados. Esses dados também foram confirmados pela predição estrutural de proteínas (dado não apresentado).

Uma vez confirmada a anotação *in silico* das sequências de aminoácidos a pergunta que surge é como esses genes estão organizados e qual sua origem? A tentativa de posicionamento das sequências encontradas no genoma dos respectivos insetos genoma foi dificultada pelo baixo número de sequências existentes, que por sua vez não conseguiram formar contigs situados em posições determinadas no genoma.

A grande questão a ser resolvida sob o ponto de vista da origem das sequências dos genes que codificam enzimas da via do ácido aspártico em insetos refere-se sobre uma possível contaminação com sequências bacterianas que sempre colocam em dúvida a montagem das sequências genômicas. Muitas dessas sequências foram excluídas dos bancos de dados dos genomas de insetos, pois os responsáveis pelo seqüenciamento desses genomas tiveram dúvida na montagem e desconfiaram da possibilidade de contaminação. Em geral, sequências quando perfazem contigs onde existe grande cobertura por vários “reads”, são considerados como pouco prováveis contaminantes. Outro fator que permite descartar a hipótese de contaminação bacteriana é a existência de EST oriundos de bibliotecas de cDNA produzidos a partir de mRNA, pois em eucariotos, estes possuem uma cauda poliA, que em bactérias só existe em alguns poucos genes (Dreyfus e Régner, 2002). Outra característica típica de eucariotos que pode ser observada no gene DHDPH de *A. gambie* foi a existência de introns, sugerindo que essa sequência não se trata de um contaminante bacteriano. Por sua vez, se as sequências dos genes que codificam as enzimas da via do ácido aspártico fossem oriundas de contaminação de DNA bacteriano, deveria ser possível encontrar por BLASTn a sequência do gene da bactéria contaminante ou de algum grupo próximo no Genbank. Em nenhum dos casos citados neste trabalho foi possível encontrar sequências idênticas as bactérias presentes no GenBank. Além disso, nenhuma bactéria simbiote de insetos se posicionou entre os 5 melhores resultados

do tBLASTn. Assim, o conjunto de resultados apresentados neste trabalho sugere que as sequências dos genes que codificam enzimas da via do ácido aspártico encontradas em *A. gambiae* e *A. aegypti* (Tabela 2) provavelmente não são oriundas de contaminações a partir de DNA bacterianos presentes nas amostras de DNA extraídos dos insetos e utilizadas no seqüenciamento genômico. Porém não é possível descartar por completo essa hipótese sem a realização de experimentos adicionais.

Assumindo que as sequências dos genes que codificam enzimas da via do ácido aspártico em insetos não são oriundas de contaminações com DNA bacteriano, esses genes deveriam ter características de genes de eucariotos tais como, presença de introns, percentual de GC mais baixo e maior similaridade com genes de grupos filogeneticamente mais próximos, nesse caso com as plantas. Na realidade o que observamos é exatamente o oposto. Apesar da existência de introns no gene DHDPR (Tabela 3), seu conteúdo GC dos genes em geral é alto e as análises filogenéticas para todas as sequências encontradas em insetos mostram invariavelmente que elas são filogeneticamente mais próximas as sequências dos genes da via do ácido aspártico de bactérias do que de plantas (Fig. 3B-8B). As premissas acima nos levam a considerar algumas possibilidades. Levando-se em consideração que as sequências dos genes da via do ácido aspártico encontradas em insetos são mais similares as de bactérias, mas não são contaminantes, esses genes não podem ter sido herdados verticalmente e possuir origem monofilética, já que para isso plantas e insetos deveriam guardar maior similaridade de sequência do que insetos e bactérias, o que não se observa, pois ambos são eucariotos e compartilham um ancestral comum. Além disso, apesar de buscas exaustivas, não observamos a presença de sequências de genes que codificam enzimas da via do ácido aspártico em nenhum grupo relacionado a animais a não ser nos insetos. Esses fatos permitem considerarmos a possibilidade de que as sequências encontradas em insetos podem ter origem num processo transferência gênica lateral entre bactérias, possivelmente simbiontes, e seus insetos hospedeiros. Essa possibilidade é bastante plausível uma vez que transferência gênica lateral de bactérias para insetos e outros eucariotos tem sido observado com certa frequência (Kondo et al., 2002; Hotopp et al., 2007). A

transferência de material genético entre bactérias e insetos é facilitada devido à presença de endosimbiontes bacterianos que habitam as células gonadais (Moret et al., 2001). A transferência gênica lateral para eucariotos a partir de procariotos simbiotes é considerado mecanismo potencial de aquisição de genes e funções (Moret et al., 2001). Em *Drosophila*, por exemplo, foi encontrado o genoma quase que completo da bactéria *Wolbachia* integrado no genoma do inseto (Kondo et al., 2002).

A existência de similaridade entre uma sequência “isca” e as sequências genômicas de um dado organismo depositadas em bancos de dados, não podem ser consideradas suficiente para anotar um gene e tratá-lo como homólogos. Principalmente em casos de organismos que realizam simbiose com diversas bactérias. Assim, é necessário a obtenção de dados experimentais que corroborem os dados de anotação *in silico*. Em nosso estudo, para comprovar a existência das sequências de genes que codificam enzimas da via do ácido aspártico em insetos identificadas a partir de reads do genoma e de ESTs de *A. gambie* e *A. aegypti*, tratamos colônias desses insetos por várias gerações na presença de tetraciclina. Em *A. gambie*, foi possível amplificar por PCR utilizando primers específicos, 5 genes nas amostras de DNA de insetos tratados com antibiótico (Fig. 10), onde não foi possível amplificar sequências de genes que codificam RNA ribossômico 16S nessas amostras sugerindo que elas não estavam contaminadas com DNA bacteriano (Fig. 9). Esses resultados sugerem que pelo menos *A. gambie* possui as sequências dos genes que codificam as enzimas da via do ácido aspártico, integradas no seu genoma. Os resultados obtidos para *A. aegypti* não foram conclusivos. Somente para o gene DHAPS foi possível amplificar um fragmento de tamanho esperado. Essa dificuldade em amplificar os genes de *A. aegypti* reside na imprecisão no momento de desenhar os primers pelo fato de existirem poucas sequências depositadas nos bancos de dados, que na sua maioria tratam-se de reads únicos.

Conclusão

Foi possível demonstrar por análise computacional e experimentação a existência dos genes da via de síntese de lisina em *Anopheles gambiae*. A funcionalidade da via ainda está sendo testada, mas dados de RNA mensageiro indicam que essa via deve estar ativa. O conjunto de dados de diferentes insetos dão robustez a hipótese. A análise das sequências desses genes indicam uma proximidade maior entre bactérias e insetos, sugerindo uma possível herança lateral desses genes, processo que já foi demonstrado anteriormente para outros genes (Kondo et al., 2002). Os dados de *A. aegypti* ainda são preliminares e necessita-se esperar que o projeto genoma esteja mais curado.

Capítulo II – Presença da via da sacaropina em bactérias

Introdução

A via da sacaropina é utilizada em fungos para a síntese de lisina e em eucariotos superiores (Fig. 12) para a degradação de lisina (Zabriskie e Jackson, 2000; Arruda et al, 2000). Na direção da síntese em leveduras, o ácido α -amino adípico (AA) é transformado em amino adípico-semialdeído (AASA) pela atividade da enzima aminoadipato redutase (AAR). Em seguida o AASA é transformado em sacaropina (SAC) pela atividade da enzima sacaropina redutase (SR). Por último a SAC é reduzida formando lisina (LIS) e ácido α -cetoglutarico (α -KG) pela enzima sacaropina desidrogenase (SDH). Na direção da degradação o primeiro passo da via é catalisado pela enzima lisina cetoglutarato redutase (LKR) que condensa lisina e α -cetoglutarato em sacaropina. Em seguida a sacaropina é hidrolisada formando AASA e ácido glutâmico (GLU) pela enzima SDH (Fig. 12).

A via de degradação de lisina em plantas começou a ser evidenciada pelos estudos da incorporação de lisina marcada com ^{14}C em sementes em desenvolvimento de milho (Sodek e Wilson, 1970). Os autores verificaram que o ^{14}C proveniente da

lisina era incorporado em ácido glutâmico e prolina e que havia também marcação significativa em ácido α -aminoadípico. Evidências da via da sacaropina para a degradação de lisina já haviam sido obtidas em humanos. Estudando doenças genéticas associadas ao metabolismo de amino ácidos, foi observado que pacientes com lisinúria apresentavam acúmulo de lisina (Oyanagi et al., 1970) no plasma enquanto que pacientes com sacaropinúria apresentavam acúmulo de sacaropina (Carson et al. 1968). As enzimas envolvidas com essas doenças foram isoladas e caracterizadas primeiramente em modelos animais como sendo a lisina-cetoglutarato redutase (LKR) e a sacaropina desidrogenase (SDH) (Hutzler e Dancis, 1968; Hutzler e Dancis, 1970). Verificou-se em seguida que os pacientes com lisinúria e com sacaropinúria apresentavam deficiência nas atividades da LKR e SDH, respectivamente (Dancis et al., 1969; Hutzler e Dancis, 1970). Mais adiante se verificou que em mamíferos a LKR e a SDH são atividades de um polipeptídeo bifuncional (Markovitz e Chuang, 1987).

Em plantas a primeira evidência da atividade da LKR para a degradação de lisina foi observada em milho (Arruda et al., 1982). Posteriormente a enzima foi purificada e caracterizada em milho e arroz (Brochetto-Braga et al., 1992; Gaziola et al., 1997; Gaziola et al., 1999). A LKR mostrou propriedades cinéticas similares à enzima de mamíferos (Brochetto-Braga et al., 1992). A LKR de plantas, assim como a de animais está associada a SDH num polipeptídeo bifuncional (Gonçalves-Butruille et al., 1996) cujas atividades são reguladas diferencialmente por Ca^{+2} , sais e osmólitos (Kemper et al., 1998). As atividades de LKR/SDH estão relacionadas com o metabolismo de aminoácidos durante o desenvolvimento das sementes em forte correlação com a taxa de síntese das proteínas de reserva (Arruda et al., 2000; Lima et al., 2003). Existem evidências que em situações específicas formas monofuncionais da LKR ou SDH foram encontradas (Zhu et al., 2000; Tang et al., 2002).

O papel da via da sacaropina no controle dos níveis de lisina em plantas tem sido evidenciado ao longo dos anos. Apesar das tentativas de se obter plantas com alto teor de lisina através da produção de mutantes contendo enzimas da via de síntese de

lisina pelo ácido aspártico menos sensíveis a retro inibição pela lisina (Azevedo et al., 1990; Shaul e Galili, 1992; Heremans e Jacobs, 1997; Lee et al., 2001), ficou evidente que essa estratégia não era suficiente para manter alto os níveis de lisina. Experimentos com plantas transgênicas expressando as enzimas da via do ácido aspártico, AK e DHDPS de bactérias, que são naturalmente menos sensíveis a retro inibição pela lisina, revelou altas taxas de degradação da lisina pela via da sacaropina (Lee et al., 2001). Plantas transgênicas de milho transformadas com um gene da DHDPS bacteriana insensível a lisina, quando tiveram também a atividade da LKR/SDH reduzida por RNAi apresentaram altos níveis de acúmulo de lisina na semente (Huang et al., 2005; Frizzi et al., 2007; Reyes et al., 2009)

Como citado, a degradação da lisina pela via da sacaropina tem os dois primeiros passos são catalisados pela LKR/SDH, uma proteína bifuncional contendo ambas as atividades (Arruda et al., 2000). Em plantas esses dois domínios enzimáticos são entremeados por um polipeptídeo de cerca de 100 aminoácidos, conhecido como interdomínio (ID) (Kemper et al., 1999). A LKR/SDH é regulada por um complexo que envolve Ca^{2+} , força iônica e osmólitos (Kemper et al., 1998). O gene que codifica a LKR/SDH é regulado transcricionalmente pelo fator de transcrição O2 (Kemper et al., 1999), que regula o fluxo de carbono e nitrogênio na semente do milho (Wang e Larkins, 2001).

A Degradação de lisina é fundamental para o desenvolvimento vegetal e animal. Há evidências que produtos da degradação desse aminoácido possuem importante papel na regulação de outras vias além de participar na formação de glutamato neurotransmissor em cérebro de camundongos (Papes et al., 2001). A degradação da lisina pela via da sacaropina leva a formação de amino adipato que pode então formar acetil CoA ou ácido pipecólico. Ácido pipecólico é um aminoácido não protéico que já foi relacionado a síndrome de Zellweger no sistema nervoso central (Peduto et al., 2004), a desordens peroxissomais (Kerckaert et al., 2000) e a indução de apoptose em células neuronais (Matsumoto et al., 2003). Em plantas o excesso desse aminoácido funciona como um indicador do metabolismo anormal de proteínas e aminoácidos Palf

e Dezsi, 1968). Acredita-se que altos teores desse aminoácido estejam relacionados a altos teores de lisina livre e pode culminar no aumento de proteólise, o que pode ser uma barreira ao aumento de lisina em grãos (Moulin et al., 2006). Em plantas a degradação de lisina pela via da sacaropina também está associada à resposta a diversos tipos de stresses. Plantas submetidas a estresse osmótico apresentaram aumento da quantidade de mRNA da forma bifuncional do gene LKR/SDH e monofuncional SDH produzidos a partir do mesmo gene (Deleu et al., 1999).

Os estudos da degradação de lisina pela via da sacaropina em diferentes organismos evidenciam alguns aspectos interessantes. Porque em plantas e animais a via é utilizada predominantemente para a degradação, enquanto que em fungos a via é utilizada para a síntese de lisina? Porque no polipeptídeo bifuncional de plantas existe um domínio (ID) intercalando os domínios de LKR e SDH que não existe nos animais? Existe algum outro organismo que utiliza a via da sacaropina para a degradação de lisina?

O objetivo deste trabalho foi identificar a existência da via da sacaropina em bactérias. Para isso focamos as buscas nas enzimas LKR e SDH e comparamos as características das enzimas bacterianas com aquelas de plantas, animais e fungos. Os resultados são discutidos no contexto da evolução dessa via nos diferentes organismos e do possível contribuição das enzimas bacterianas na evolução das enzimas de plantas e animais.

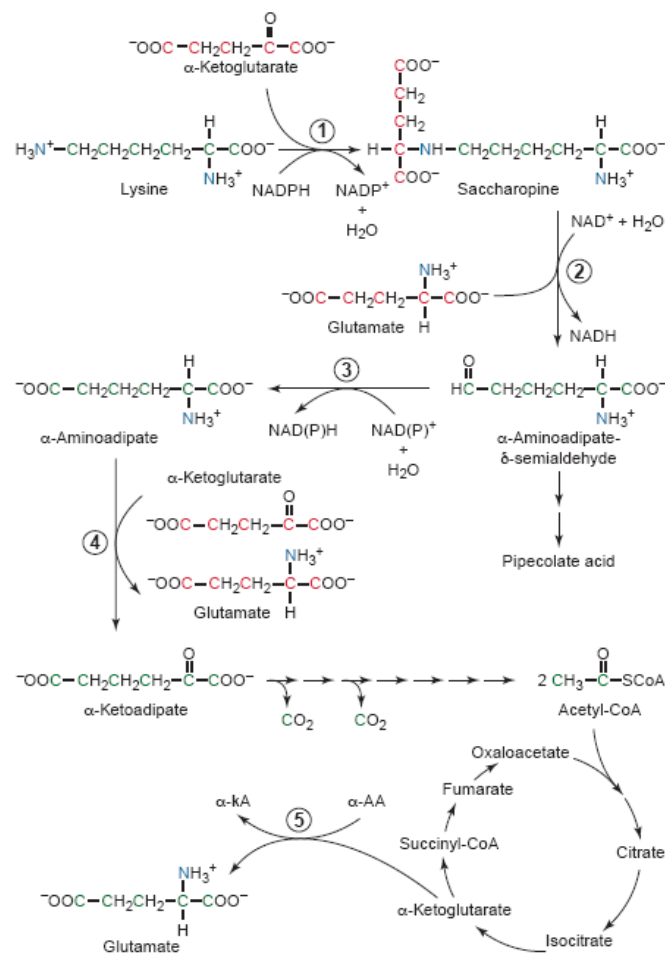


Figura 12 - Via da sacaropina de degradação de lisina em plantas e animais. 1- lisina-cetoglutarato redutase (LKR), (2) sacaropine desidrogenase (SDH), (3) aminoadípico semialdeído desidrogenase (AASDH), (4) aminoadípico aminotransferase (AAA), (5) α-aminoácido transaminase (Adaptado de Arruda et al., 2000).

Resultados

Identificação dos genes que codificam a LKR e a SDH em bactérias

As sequências de aminoácidos das enzimas LKR e SDH de *Arabidopsis thaliana* e *Mus musculus* foram utilizadas como “iscas” para investigar a existência de sequências homólogas em bactérias. Sabe-se que tanto em plantas como em animais as atividades enzimáticas da LKR e da SDH encontram-se em um único polipeptídio

bifuncional (Kemper et al., 1998). Cada um dos dois domínios pôde ser identificado pelo alinhamento com a sequência de aminoácidos da LKR e da SDH de levedura (Fig. 13). Em *Saccharomyces cerevisiae* a LKR e a SDH são codificadas pelos genes *lys1* e *lys9*, respectivamente, que estão localizados em cromossomos diferentes (Bhattacharjee, 1985). Quando comparamos as sequências da LKR/SDH entre plantas, levedura e camundongo (Fig.13A, B), fica claramente evidenciada uma região, de aproximadamente 100 aminoácidos, encontrada exclusivamente em plantas, intercalando os domínios LKR e SDH (Kemper et al., 1999). Existe indicações que essa sequência, denominada de inter-domínio (ID), exerça algum papel regulatório exclusivo de plantas (Kemper et al., 1998). As sequências de aminoácidos da LKR, do ID e da SDH foram utilizadas em buscas para identificar sequências ortólogas em genomas bacterianos disponíveis no GenBank.

A

<i>S.cerevisiae</i> (LYS1)	1	-----MAAVTLHLRAETKPLEARAATPTTVKKLIAK
<i>C.albicans</i> (LYS1)	1	-----MSKSPVTLHLRAETKPLEARAATPTTQQLDA
<i>Z.mays</i>	1	-----MGSAAETGNDTLLGNGVVGHLAETCNWERRAPITPSCARLLLG
<i>A.thaliana</i>	1	-----MNSNGHEEEKLLGNGVVGHLAETVKNWERRITPTPSCARLLHG
<i>O.sativaI</i>	1	MGASEVRSWGFLGHLSPRSGPSGDSMKNQRFVFNHVGVKIVLNSDTLLGNGVVGHLAETANWERRAPITPSCARLLLG
<i>C.elegans</i>	1	-----MRLPHTAGGMMRQISQRFQSSKVFGK-GTPTIGTRRETVNSWERRAPITPSCARLLTKS
<i>M.musculus</i>	1	-----MLRAQRP-R-LARLRACLSRGLHHKPVMAIRREDVNAWERRAPITAPRHIGITTKL
<i>H.sapiens</i>	1	-----MLQVHRT-G-LGRGLVSLSKGLHHKAVLAVRREDVNAWERRAPITAPRHIGITTKNL
<i>S.cerevisiae</i> (LYS9)	1	-----
<i>C.albicans</i> (LYS9)	1	-----
<i>S.cerevisiae</i> (LYS1)	33	GE-----KIVVEDSPQSTFNINISYRQAGATIVPAGSWKTAPDRITIGKEMPEITDTPLVHEHIOAHACYKQCAQWO
<i>C.albicans</i> (LYS1)	35	GE-----EIVVESSQSTFDIKIYEAVGAKIVPEGSWKTAPKERITFGIKELPENETFPPLIHEHIOAHACYKQCAQWO
<i>Z.mays</i>	46	GGKNGPRVNRITIVOPSTRRIHHDAAQMEDACCEISDDLS-----ECGLIIGIKOPKLOMILSDIAYAF-FSHTHKAQKENM
<i>A.thaliana</i>	45	GKDR-TGISRIIVOPSAKRRIHHDAAQMEDVGCETSDLS-----DCGLIIGIKOPKLEMLPBRAYAF-FSHTHKAQKENM
<i>O.sativaI</i>	81	GKRGITGVNRITIVOPSTKRRIHHDAAQMEDVGCETSDLS-----ECGLIIGIKOPKLEMLPBRAYAF-FSHTHKAQKENM
<i>C.elegans</i>	60	GV-----NVLIQPSNRRAYPIQDIYSAGAVRCDLS-----EALIIIGVSRVPEIDOLIENTYAF-FSHTHKAQODNM
<i>M.musculus</i>	54	GV-----KVLIQPSNRRATHDKQYVRAGGILQEDIS-----EACLIIGVSRPPEEKLMRSKTYAF-FSHTHKAQENNM
<i>H.sapiens</i>	54	GV-----KVLIQPSNRRATHDKQYVRAGGILQEDIS-----EACLIIGVSRPPEEKLMRSKTYAF-FSHTHKAQENNM
<i>S.cerevisiae</i> (LYS9)	1	-----
<i>C.albicans</i> (LYS9)	1	-----
<i>S.cerevisiae</i> (LYS1)	106	NVIMRFKKGHTIYDLEFENDQGRVAAFGFIAGFAGAAAGVLDWSFQDNGNTGKGEGERGELPGVTTPMNENELI
<i>C.albicans</i> (LYS1)	108	DVLRFPQNGNGLYDLEFENDQGRVAAFGFIAGFAGAAAGVLDWSFQDNGNTGKGEGERGELPGVTTPMNENELI
<i>Z.mays</i>	120	PLLDKILSERVSLFDYELVGDGGRSLAFGKAGRAGLIDFLHGLGORYLSLGYSTPFI-----SLGQSHMYPSLAAAK
<i>A.thaliana</i>	118	PLLDKILSERVSLFDYELVGDGGRSLAFGKAGRAGLIDFLHGLGORYLSLGYSTPFI-----SLGASYMYPSLAAAK
<i>O.sativaI</i>	155	PLLDKILSERVSLFDYELVGDGGRSLAFGKAGRAGLIDFLHGLGORYLSLGYSTPFI-----SLGQSHMYPSLAAAK
<i>C.elegans</i>	127	EMLDILQNRIRLDYKICDDKGRKRVFEGKNAGNAGSIDLHGLGLRLTALGHNTPEF-----ARSTSQHPQ-SSKRVFOLYG
<i>M.musculus</i>	121	NLLDEILKQERRLIDYKIVDHRGSRIVAFGQVAGVAGINILHGMCLRLTALGHNTPEF-----HGLMAHMYRNSSOAV
<i>H.sapiens</i>	121	GLLDEILKQERRLIDYKIVDHRGSRIVAFGQVAGVAGINILHGMCLRLTALGHNTPEF-----HGLMAHMYRNSSOAV
<i>S.cerevisiae</i> (LYS9)	1	-----
<i>C.albicans</i> (LYS9)	1	-----
<i>S.cerevisiae</i> (LYS1)	177	KDVT KDYK--EALAT-GARKPTVLIGALCRGSGAIDILHKVGIEDANILKWDIKETSR-----
<i>C.albicans</i> (LYS1)	188	KDVIELE--KALTKNGGQYPKCLVIGALCRGSGAIDILFKKGIEDDNPAKWDIAETAK-----
<i>Z.mays</i>	195	AAVIVVAEETATFGIPSGICPVEFVFTGTGNVSGAQEIFKLL--PHTVEVDAEKDPEFQARNLSKQSQ-STKRVFOLYG
<i>A.thaliana</i>	193	AAVISVGEETIASQGPLGICPVEFVFTGTGNVSLGAQEIFKLL--PHTVEVPSKIDPEFVKDKGISQNGISTKRVYQVYG
<i>O.sativaI</i>	230	AAVIAIGEETATFGIPSGICPVEFVFTGTGNVSASTFLEVGMT-----ARSTSQHPQ-SSKRVFOLYG
<i>C.elegans</i>	202	NAIRDICYEIALDRMERSICPVEFVFTGTGNVSGAQEIFEHL--PHEYVDVATPEKVAQKGQLN-----KVYQ
<i>M.musculus</i>	196	QAVRDACYEISLGLMEKSTICPTEFVFTGTGNVSGAQEIFNEL--PCEYVEPHEIREVSKTGDLR-----KVYQ
<i>H.sapiens</i>	196	QAVRDACYEISLGLMEKSTICPTEFVFTGTGNVSGAQEIFNEL--PCEYVEPHEIREVSKTGDLR-----KVYQ
<i>S.cerevisiae</i> (LYS9)	1	-----
<i>C.albicans</i> (LYS9)	1	-----

S. cerevisiae (LYS1) 234 -----G-G-P-FDEI-----EADIFINCYLSKPIAPFTNMEKL-NPN-----
C. albicans (LYS1) 246 -----G-G-P-FDEI-----VDIDIFINCYLSKPIAPFTNMEKL-NEN-----
Z. mays 272 CVVTSRDMVSHKDPTRQFDKGGYYAHPEEYTPVHERIAPYASVIVNCIYWEKRFEPPLNMDQLOQLT-----
A. thaliana 271 CIIITSQDMVEHKDPSKSPDKADYYAHPEEYNPVFHEKISPYTSVLVNCIYWEKRFEPPLNMDQLOQLT-----
O. sativa 292 CVVSSRDMVTPKDPTRCFNKADYYAHPEEYKPVHERIAPYASVIVNCIYWEKRFEPPLNMDQLOQLT-----
C. elegans 269 CVVTRDHFVPRKHGFP-FDKKEEEOFPDRTYSKFAEIAPIYASVILINGIYWDQSPRLITIPDAKNLT-TPVQRYDTP--G
M. musculus 263 TVLSRRHHVVRKTDV-VDPVVEYKYPEYTSRENTDIAPYTTCLINGIYWEONTFRLLRQDAQSLTPVVKSSVVPVEG
H. sapiens 263 TVLSRRHHVVRKTDV-VDPVVEYKYPEYTSRENTDIAPYTTCLINGIYWEONTFRLLRQDAQSLTPVVKSSVVPVEG
S. cerevisiae (LYS9) 1 -----
C. albicans (LYS9) 1 -----

S. cerevisiae (LYS1) 270 -----RR-LRTVVDVSADTNPVHPPIYVATVFN-----KPTVLVPTA-CPKSVISIDHPSLLEPREASEFFSH
C. albicans (LYS1) 282 -----RK-LRTVVDVSADTNPVHPPIYVATVFN-----EPTVEVKLDK-CPKSVCSIDHPSLLEPREASEFFAK
Z. mays 340 -----ETGCPLVGVCDDICDGGGIEFNKSTSIERPFRYDPSKNSYHDDME-CAGVCLAVDILPTEFSKEAQQHFCN
A. thaliana 339 -----KGLPLVGICDDICDGGGIEFNRAALIPSPFRNPSNNSYHDDMD-CGVCLAVDILPTEFAKEAQQHFGD
O. sativa 360 -----KNGCPLVGICDDICDGGGIEFNKSTSIERPFRYDPSKNSYHDDME-CNGVCLAVDILPTEFSKEAQQHFCN
C. elegans 346 CPTLPH-----LLALCDISADPGGSVEFTECTITIERPFCMYDAQOQIHDVSVE-CGILMCSIDNLPALPTEAETCFGD
M. musculus 342 CPALPH-----LVALCDISADTGGGIEFTECTITIERPFCMYDAQOQIHDVSVE-CGILMCSIDNLPALPTEAETCFGD
H. sapiens 342 CPALPH-----LVALCDISADTGGGIEFTECTITIERPFCMYDAQOQIHDVSVE-CGILMCSIDNLPALPTEAETCFGD
S. cerevisiae (LYS9) 1 -----
C. albicans (LYS9) 1 -----

S. cerevisiae (LYS1) 336 DILPSSELTPQPKTAPVVRRAKLLFDRHCAVRKRSRL-----
C. albicans (LYS1) 348 DIMESTLEPNRDTSPVVRRAKQLFDRHCAVRKRSRL-----
Z. mays 414 IISRLVASIASVKQP-----AELPSYRRACIAHACRLTSLLEYTPRMNTMIDLAPAKTNPLPD---KKYSTLVLSLSGH
A. thaliana 413 IISGVGSLASMTET-----SLDPAHRRACISYRCELASLEYTPRMKSNPEEAQDNIIANGVSSQRTFNILVLSGH
O. sativa 434 IISKVARASAKEL-----LELPSHRRACIAHACRLTSLLEYTPRMKIIELPPAPTNLPLPD---KKYNSLVLSLSGH
C. elegans 423 IISYPLWDLNLTNSDQHFDRQLCRTEINAITDQCKLTPNLEYTAQIRKDKAASASANS-----
M. musculus 418 MIYPLVEEILLSDASQPLESONFSPVVRDAVITSNGLLDKKYKQKRESRERI-----
H. sapiens 418 MIYPLVEEILLSDATQPLESONFSPVVRDAVITSNGLLDKKYKQKRESRERI-----
S. cerevisiae (LYS9) 1 -----
C. albicans (LYS9) 1 -----

S. cerevisiae (LYS1) -----
C. albicans (LYS1) -----
Z. mays 486 LFDKFLINEALDIIETAGGSFHLVRCVGVSTDDMSYSELEVGAADDATLDKIIDSLTSLANEHGGDHDAGQEI-ELALK
A. thaliana 488 LFDKFLINEALDIIETAGGSFHLVRCVGVSTDDMSYSELEVGAADDKRVLDQIIDSLTRLANPNEDYISPHREANKISLK
O. sativa 506 LFDKFLINEALDIIETAGGSFHLIRCDVGVSTDDMSYSELEVGAADDATLDKIIDSLTSLANEHGGDHDAGQEI-ELALK
C. elegans 483 -----
M. musculus 473 -----
H. sapiens 473 -----
S. cerevisiae (LYS9) 1 -----
C. albicans (LYS9) 1 -----

S. cerevisiae (LYS1) -----
C. albicans (LYS1) -----
Z. mays 565 IGKVNEYETDVT-----IDKGPKILILGACRVCRPAAEPLASYPDICT-----YGVDDHADQIHRIVIVASLYQKDAEET
A. thaliana 568 IGKVQQENEIKE-KPEMTKKSGVLLGACRVCRPAADPLASVRTISSQWQYKTFYFADSEEEKTDVHRIVIVASLYLKDAKET
O. sativa 584 IGKVNECGTDDS-----MAK-EGSKVLLGACRVCRPAAEPLASYSNIFS-----ESAYDHDIDQIHRIVIVASLYQKDAEET
C. elegans 483 -----RVMGGTTRKRVLLGACRVCRPAAEPLASYSNIFS-----TVATESQRDGGORLCTS
M. musculus 473 -----QFLSMSTKRVLLGSGCYVSGPVLEXTSRDNNIEIT-----TLGSDMTNQMQOLSKK
H. sapiens 473 -----QSLSMGTTRKRVLLGSGCYVSGPVLEXTSRDNNIEIT-----TVGSDMKNQIEQLGKK
S. cerevisiae (LYS9) 1 -----MG-NVLLGSGFVAQPVLDITTAANDDINV-----TVACRTLANAQAALAKP
C. albicans (LYS9) 1 -----MVKKVLLGSGFVAQPVLDITSEQPDIEV-----TVACRTLSKAKELAG-----

S. cerevisiae (LYS1) -----
C. albicans (LYS1) -----
Z. mays 635 VDGIE-N-TTATQLDVA-IGISLSDLVSQVEVVISLLPASFFAAATAGVCTIELKHHVVTASYDES-SNLSQAARKAGVTIT
A. thaliana 647 VEGIS-D-VEAQLDVS-DESELKYVSQVDVVISLLPASCFVAVAKTCTIELKHHVVTASYDDETSMTSMKAKSAGITIT
O. sativa 654 IDGIR-N-ATAQLDVA-BIKNLSNLSVS-----OMKHHVVTASYDES-SKPEQSAEGAGVTIT
C. elegans 534 -----P-N-IQSVVDIAR-SHTTERRLREHDLVVISLLPNTFPLVAKMCTISNORDVTSYSPSEALDKAAKADVTIT
M. musculus 524 -----Y-N-INPVSLTVGKQEAQLQSLVESQDLVISLLPVVLEPVPVAKACIESVNVVTASYTPAKKELEKSVDAGITIT
H. sapiens 524 -----Y-N-INPVSMDCQKEERLGLFVAKQDLVISLLPVVLEPVPVAKACITNKNVTASYTPAKKELEKSVDAGITIT
S. cerevisiae (LYS9) 46 -----S-G-SKASLDVT-DSDALDKVLADNDVVISLIPYTFEPNVVKSATRTKTDVTSYSPALRELEPEIVKAGITIT
C. albicans (LYS9) 45 -----D-K-AQASLDVT-DAALDEQVAKHDLVISLIPYTFEPNVVKSATKKNKKNVTASYNPOKALEKEIEBAGITIT

S. cerevisiae (LYS1) -----
C. albicans (LYS1) -----
Z. mays 712 ICEMGLDPGIDHMSMKMIDEAHRKCKIKATTSYCGGLPSPAAANNPLAYKFSWNPACATRSGRNPAYKFLGETIHDV
A. thaliana 724 ICEMGLDPGIDHMSMKMIDEAHRKCKIKATTSYCGGLPSPAAANNPLAYKFSWNPACATRAGONPAKYKSNQDITHDV
O. sativa 709 ICEMGLDPGIDHMSMKMIDEAHRKCKIKATTSYCGGLPSPASANNPLAYKFSWNPACATRAGRNPAVKFHCETIHDV
C. elegans 608 INEAGLDPGIDHMLAMECDIDIEHGGRTSSESCGGLPAPESDNPTRYKFSWSPEGVLTALMNPAYKYLKNGKIVEVP
M. musculus 598 ICELGLDPGIDHMLAMETIDTAKEICATESVSYCGGLPAPESDNPTRYKFSWSPEGVLMNIMQASVYLLNCKVNVVT
H. sapiens 598 ICELGLDPGIDHMLAMETIDTAKEICATESVSYCGGLPAPESDNPTRYKFSWSPEGVLMNVMQASVYLLDCKVNVVA
S. cerevisiae (LYS9) 119 INEIGLDPGIDHLYAVKTIIDEVRRAGGKIKSLSYCGGLPAPESDNPTRYKFSWSPEGVLLALRNQAKYWDGKQETIS
C. albicans (LYS9) 118 INEIGLDPGIDHLYAVKTIIEVEHKAGGKIVSELISYCGGLPAPESDNPTRYKFSWSPEGVLLALRNQAKYWDGKQETIS

S. cerevisiae (LYS1) -----
C. albicans (LYS1) -----
Z. mays 792 GH-NIYESAKRLRLRELPAFALEHLPNRSLSLYGDIYGISKEASTIYRATLRYEGFSEI-VTLSTKTCFFDAANHPILQDT
A. thaliana 804 GK-NIYDSAAFRRVPLNPAFALECPNDRDSLVEGYGIESEATITFRGTLRYEGFSEI-VTLSTKTCFFDAANHPILQDT
O. sativa 789 G-----SEI-VTLSTKTCFFDAANHPILQDT
C. elegans 688 SG-SVVDNLIDIDF-MPCLNIGPNRDSIRYSVYGLGNDCKTIRGTLRYOGFVDVKALHSVGLSGNIDSFTSN
M. musculus 678 GGVSEFNSVTPMDY-MPCLNIGPNRDSIRYSVYGLGNDCKTIRGTLRYOGFVDVKALHSVGLSGNIDSFTSN
H. sapiens 678 GGISFLDAVTSMDY-MPCLNIGPNRDSIRYSVYGLGNDCKTIRGTLRYOGFVDVKALHSVGLSGNIDSFTSN
S. cerevisiae (LYS9) 199 SE-DLMAAKPYFI-YPCYAFVCPNDRDSTLEKOLYH-PEAETVIRGTLRYOGFPEFVKALVDGMDKDPANETFSKP
C. albicans (LYS9) 198 SE-DLMAAKPYFI-YPCYAFVCPNDRDSTLEKOLYH-PEAETVIRGTLRYOGFPEFVKALVDGMDKDPANETFSKP

S. cerevisiae (LYS1) -----
C. albicans (LYS1) -----
Z. mays 871 SRP---HYKGFIDELNNISTINTDLDIE--ASGGYDDLIARLKLGCCNKKIIVKTYKTKEGLGHEETOIPKGCSSP
A. thaliana 883 KRI---TFGALISNLLNKDADNESEPLA-----GEEISKRILKLGH--SKETAAKAAKTVEGLGFNEEREVPSLCKSV
O. sativa 815 TRP---HYDFLVVDFNACNISTARKEYSEVSGGQGGELISRLISFGHCCKDEIAAKT---KTKEGLGYEETOIPENCSSA
C. elegans 765 IGPDLTWKELIASI-----SN--QKLDIFPDSTRHII---EEKVGGKG---ISATENLGLFS-DKVVDRHGAP
M. musculus 755 ANP---LWTKQLLCOL-----VG--ISRSSPC-KLKEVV---FTKLGGDNTQLEAEWLGILG-DEQVPQAESI
H. sapiens 755 ANP---LWTKQLLCOL-----VG--ISPSS-HDVLKEAV---LKKLGGDNTQLEAEWLGILG-DEQVPQAESI
S. cerevisiae (LYS9) 275 IA---WNEALKQY-----LG--AKSTSKEDLIASIDSKATWKDDDRERISCFAWLGLFS-DKVTTPRGNA
C. albicans (LYS9) 274 GP---WNKATAEI-----VG--AKSSSQDIIDKINQLTKFKSPDQERILACFRWLGLFS-ENQITPRGNP

S. cerevisiae (LYS1) -----
C. albicans (LYS1) -----
Z. mays 947 FDVLCORMEORMAYGHNEODMVLLHHEVEVEYDDCOPAEKHQATLLERCKVENGRSTTAMALTVGTPAATGALLLLKKNV
A. thaliana 952 FDATCYLMEKRLAYSGNEODMVLLHHEVEVEFLESRIEKHQAATLLERCKVENGRSTTAMAKTVGTPAATGALLLLKKNV
O. sativa 893 FDVLCORMEORMAYIHNEODMVLLHHEVEVEYDDGAPTEKHQATLLERCKVENGRPTTAMAKTVGTPAATGALLLLKKNV
C. elegans 824 IDTLAQYDAKILAKDHESDIVVLLNEIDCAQLPGGNSERH-RISLVQYCNPNGF---SAMAKTVGYTTAIVSEMVLNNEI
M. musculus 815 VDAFSKELVSKLSYGPPEKDMIVMDSFCTRHPSGLENK-TIDLVVYCDNNGF---SAMAKTVGYTTAIVSEMVLNNEI
H. sapiens 815 LDAFSKELVSKLSYGPPEKDMIVMDSFCTRHPSGLENK-TIDLVVYCDNNGF---SAMAKTVGYTTAIVSEMVLNNEI
S. cerevisiae (LYS9) 336 LDTLCARLEELQMEDNERDMVVLQHKFCHIEWADGTTERR-TSLVDYCKVGGY---SSMAKTVGTPVAATKFLVLDGFI
C. albicans (LYS9) 335 LDTLCARLEELQYKGERDVLVLOHKFCHIEWANGTKERR-TSLVDYCDPNGY---SSMAKTVGTPCAVAKKQLSGEL

S. cerevisiae (LYS1) -----
C. albicans (LYS1) -----
Z. mays 1027 QTKGVIRPTEPEIYVPALEIDES-SGIKLVEKVT-
A. thaliana 1032 KTRGVIRPTEAEVYDPALEIDQA-YGIKLMEKAE--
O. sativa 973 QKKGIVRPTEPEIYVPALEIDES-SGIKLAERVET-
C. elegans 900 QRAGLQRPILKEIYVPALEIDDFGIPTHTITP-
M. musculus 891 EAKGLMGPTTKETIYCPILERIKAEGITVNTQSTIKL
H. sapiens 891 GAKGLMGPTSKETIYCPILERIKAEGITVNTQSTIKP
S. cerevisiae (LYS9) 412 KGPGLAPMSPPEINDEPKELKDKYGLLKEKIVA-
C. albicans (LYS9) 411 SKGGLAPMSSPINDSINKELKDKYGLLVERKI-

B

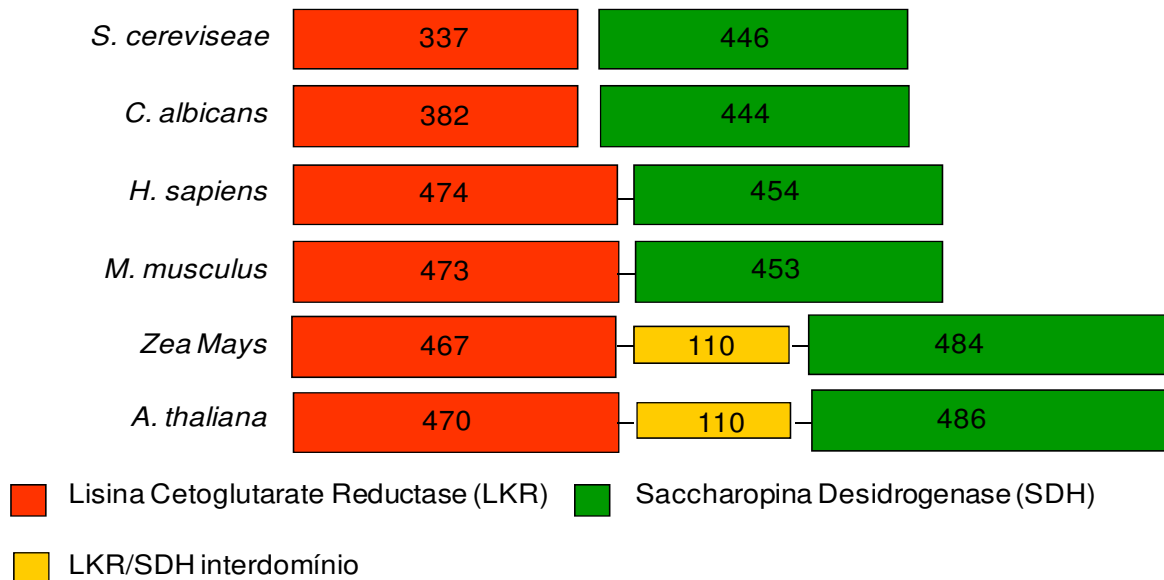


Figura 13 – (A) Alinhamento das sequências de aminoácidos das enzimas LKR e SDH de plantas, animais e leveduras. Realçado em azul estão as sequências de ~100 aminoácidos do ID que intercala os domínios LKR e SDH, existente exclusivamente em plantas. **(B)** Representação esquemática da estrutura dos genes em LKR/ID/SDH em leveduras, animais e plantas.

Através dessa busca, foi possível identificar em diferentes grupos de bactérias proteínas com sequências de aminoácidos similares a LKR e a SDH. Destacaram-se nas buscas as bactérias dos grupos CFB e Proteobactérias pela alta identidade. Curiosamente, exclusivamente em uma das α -proteobactérias, a *Silicibacter sp.*, existem duas cópias para cada um desses genes, em uma forma os genes e possivelmente os domínios enzimáticos da LKR e da SDH encontram-se justapostos no genoma sugerindo a possível formação de um peptídeo bifuncional como ocorre em animais e plantas. Na outra forma os genes estão separados no genoma. Não encontramos nesses grupos bacterianos sequências similares ao ID. Surpreendentemente, sequências similares ao ID foram encontradas em outros grupos bacterianos, principalmente com nas cianobactérias, intercalando domínios protéicos de natureza desconhecida. A Tabela 7 sumariza, através de duas espécies representantes, os grupos de organismos onde foram encontrados os diferentes domínios de LKR, ID e SDH. Os domínios LKR e SDH possuem alta similaridade entre plantas e animais com relação tamanho dos polipeptídios tomados pelo número de aminoácidos. Os domínios LKR e SDH em *Silicibacter* e *Roseobacter* são menores que os de animais e plantas. Em *Gramella* e *Flavobacterium* os dois domínios possuem numero de aminoácidos similares aos de plantas e animais.

Tabela 7 - Estrutura dos domínios enzimáticos da LKR/SDH em plantas, animais e bactérias.

Organismo	Estrutura da enzima	LKR (aa)	ID* (aa)	SDH (aa)	Domínio desconhecido (aa)	Polipeptídio Completo (aa)
<i>H. sapiens</i>	Bifuncional	474	-	454	-	925
<i>M. musculus</i>	Bifuncional	473	-	453	-	924
<i>Z. mays</i>	Bifuncional	467	110	484	-	1059
<i>A. thaliana</i>	Bifuncional	470	111	486	-	1064
<i>Nostoc sp</i>	Desconhecida	-	110	-	587	703
<i>Cianothece sp</i>	Desconhecida	-	110	-	587	703
<i>Silicibater sp</i>	Bifuncional	350	-	380	-	730
<i>Silicibater sp</i>	Monofuncional	355	-	390	-	745
<i>Roseobacter sp</i>	Monofuncional	349	-	380	-	729
<i>Gramella sp</i>	Monofuncional	402	-	457	-	859

*O inter-domínio encontrado até o momento exclusivamente em plantas, intercalando os domínios de LKR e SDH, foi identificado neste trabalho em diversas bactérias. Porém em todos os casos o ID bacteriano foi encontrado intercalando domínios com baixíssima similaridade com os domínios LKR e SDH.

Em *Roseobacter*, *Gramella* e *Flavobacterium* os genes que codificam os domínios LKR e SDH estão posicionados em locais distantes no genoma. Sequências altamente similares ao ID de plantas, com aproximadamente 100 aa, foram encontradas em cianobactérias. A princípio pensávamos que essas bactérias tinham a LKR/SDH numa estrutura polipeptídica semelhante à encontrada em plantas. Mas ao analisarmos mais detalhadamente os polipeptídios que flanqueavam o ID nessas bactérias, verificamos que estes guardavam baixíssima similaridade com as sequências de aminoácidos da LKR e da SDH de plantas, animais e leveduras e mesmo com os domínios encontrados nas α -proteobactérias (Fig. 14A). A Tabela 8 mostra uma matriz de identidade entre as sequências de aminoácidos do alinhamento representado na Fig. 14A. Como pode ser visto a similaridade entre as sequências de aminoácidos dos domínios LKR e SDH de milho e *Arabidopsis* gira em torno de 60 a 70%; entre plantas, animais e leveduras gira em torno de 30 a 40%; entre plantas, animais e α -proteobactérias gira em torno de 20 a 25% e entre qualquer um dos os

organismos e cianobactérias é menor que 10%. Isso indica uma grande divergência nas sequências de aminoácidos dos domínios que flanqueiam o ID de cianobactérias e aqueles de plantas. Já o domínio semelhante ao ID encontrado em cianobactérias mostra uma similaridade em torno de 30% com aquelas encontradas em plantas. A análise filogenética das sequências de aminoácidos da LKR e da SDH de plantas, animais, leveduras, α -proteobactérias e dos domínios que flanqueiam o ID de cianobactérias mostra claramente que esses domínios de cianobactérias posicionam-se com um “out group”; as leveduras formam um grupo próximo as α -proteobactérias e as plantas agrupam-se próximo aos animais (Fig. 14 B e C). Mesmo com a baixíssima similaridade entre as sequências de aminoácidos dos domínios que flanqueiam o ID em cianobactérias com aquelas da LKR e SDH dos demais organismos, não podemos excluir por completo a possibilidade que estes retenham atividades enzimáticas semelhantes a LKR e SDH, apesar de testes de atividade terem apresentado resultado negativo.

A

Nostoc sp	1	-----MTSRRTMCPPDHVDYVNPWMEGNHKSS-----	DRAVQWQGLHQLKQH
Anabaena sp	1	-----MTSRRTMCPPDHVDYVNPWMEGNHKSS-----	DRAVQWQGLYQLKEH
Z.mays	1	-----MGSAAATEGNDTLLGNGVGIETACNMETRAPLPSHCARLGGCKNGPRVNIIVOPSTRRIHHDAAVE	
A.thaliana	1	-----MNSNGHEEEKKLGNGVGIETAVNKEETPTLPSHCARLHGC-KDRTGISIVOPSAKRIHHDALVE	
M.musculus	1	MLRAQRPLRLARLRACLSRGLHHKPAVAVRREVNARERAPLAPKHICG-TKLC-----	YAVLOPSNRRRAIHKDKVY
H.sapiens	1	MLQVHRTGLRGLGVSLSKGLHHKPAVAVRREVNARERAPLAPKHICG-TNLC-----	YAVLOPSNRRRAIHKDKVY
S.cerevisiae	1	-----MAAVTLHARA-TKPLARAAATPTTVKKI-AAC-----	FETVYESPOSTFNNENR
C.albicans	1	-----MSKSPVLLHARA-TKPLARAAATPTTKQI-DAC-----	FETVYESPOSTFDDKEVE
Silicibacter sp	1	-----MTHLWRA-ORLNDVGLTPAGAKAT-EAC-----	IVTVESSSAIPQCHI
Roseobacter sp	1	-----MTHIWRSAARHDEVGLTPQGAAT-AAC-----	YQVVESTORSLPADMA
Nostoc sp	52	AIVDLTSGQ-----KGWPDITFTAN-----AGTGENVLSRFLHKEQCEEPFFKEFE-----	QNGYIV
Anabaena sp	52	AIVDLTSGQ-----KGWPDITFTAN-----AGTGENVLSRFLHKEQCEEPFFKEFE-----	QNGYIV
Z.mays	73	DAGCESE-----LSECGTITCROPKLQMSD-AYAPSTTHKOKENMPTDILSERVSTFDYELVGGDCRS	
A.thaliana	71	DVGCESD-----LSDCGTITCROPKLQMSD-AYAPSTTHKOKENMPTDILSERVSTFDYELVGGDCRS	
M.musculus	74	RAGGIQE-----ITEACTITCRRPPK-SK-TYAPSTTHKOKENMPTDILSERVSTFDYELVGGDCRS	
H.sapiens	74	RAGGIQE-----ITEACTITCRRPPK-SK-TYAPSTTHKOKENMPTDILSERVSTFDYELVGGDCRS	
S.cerevisiae	53	QAGAIIVBAGSWKTAPRDRITC-REMP-TTFFV-EHQAACCYKQCAWODIKREIKHGHTYDLEFENDQGRV	
C.albicans	55	AVCAKIVBAGSWKTAPKERITC-RELP-N-TTFFV-EHQAACCYKQCAWODIKREIPQNGITDLEFENDQGRV	
Silicibacter sp	51	DAGCEAP-----APRDAITC-RELP-GTPTT-RHMGHAYKQPSGKVITNREKAGGGTYDLEYVEDPGRV	
Roseobacter sp	51	AGCDITP-----APDNAITC-RELP-GTPTT-RHMGHAYKQPSGKVITNREKAGGGTYDLEYVEDPGRV	
Nostoc sp	109	NELPKDLP-EGAGDAITLDREGFWT-WAC-----YGRSELDSFVLAKWLD-DEV-SRLITDERFYH-DT-CFCF	
Anabaena sp	109	YELPKDLP-EGAGDAITLDREGFWT-WAC-----YGRSELDSFVLAKWLD-DEV-SRLITDERFYH-DT-CFCF	
Z.mays	147	LATCKACACGLIDFTGL-CQYVSLC-----YSTPLSCASYMPSLAAAKAAIVVVAEATFGPSGICILVEV	
A.thaliana	145	LATCKACACGLIDFTGL-CQYVSLC-----YSTPLSCASYMPSLAAAKAAIVVVAEATFGPSGICILVEV	
M.musculus	148	VAFQACACVACMINITGM-LRLDALC-----HHTPMH-CMA-NRNSQAQAVRDAGYVSLGLPKSIGSLTFV	
H.sapiens	148	VAFQACACVACMINITGM-LRLDALC-----HHTPMH-CMA-NRNSQAQAVRDAGYVSLGLPKSIGSLTFV	
S.cerevisiae	133	AAGCFACAGALG-ADW-F-OT-----HS-----DDEDPAVSP-PNEKAL-KDVKKDYKEALATGARK-----FTVL	
C.albicans	135	AAGCFACAGALG-LDWSF-OT-----NGNTKGTGEGEGE-PGVTP-PNENEL-KDKIELEKALTKNGGQ-YKCI	
Silicibacter sp	125	AAGCYACAGAVTITAW-AQQ-----RGETCPVGVVKNKDALAD-C-RE-DATGADR-----SKAL	
Roseobacter sp	125	AAGCYACAGAVSITMCWMAQQ-----NGKVAQPVVAPSANHLME-Q-QDMALGTDR-----ETAL	
Nostoc sp	177	ANGYLLYPCAFDS-SNRL-EMR-----	
Anabaena sp	177	ANGYLLYPCAFDS-SNRL-EMR-----	
Z.mays	220	FTGVNVSQCAQ-ILKLPHTEFVDAEKLEIFQARN-LSKQSQSTKRVFQLYGCVVTSRDMVSHKDPTRQFDKGD	YAHF
A.thaliana	218	FTGVNVSQCAQ-ILKLPHTEFVDSKLEIFVKDKGISONGISTKRVFQVYVGCITTSQDMVEHKDPSKSFADKAD	YAHF
M.musculus	221	FTGVNVSQCAQ-ILKLPHTEFVDSKLEIFVKDKGISONGISTKRVFQVYVGCITTSQDMVEHKDPSKSFADKAD	YAHF
H.sapiens	221	FTGVNVSQCAQ-ILKLPHTEFVDSKLEIFVKDKGISONGISTKRVFQVYVGCITTSQDMVEHKDPSKSFADKAD	YAHF
S.cerevisiae	199	IG-LRCGSCAID-LHK-G-PDANI-----	LKDIKE
C.albicans	211	IG-LRCGSCAID-LKK-G-PDDNIA-----	KDMAE
Silicibacter sp	185	IG-LRVGTCAAD-CEA-G-AVT-----	KDMNE
Roseobacter sp	185	IG-LRVGTCAAD-CQA-G-GTT-----	KDMAE

Nostoc sp 200 --V PEKRIATTEADA NFAONTNVNDSIV N-KASDAIKTR AD GFO
 Anabaena sp 200 --V PEKRIATTEADA NFAONTNVNENIV N-KASNAIKQS ED GFO
 Z_mays 299 EHYTPVHEHRIAPYAS I NC LWKRFEP NMQLQOMETG CPVVG CD CD IGGSHEF
 A_thaliana 298 EHYNPVHEHRIAPYAS I NC LWKRFEP STKLOQDTKKG LPVVG CD CD IGGSHEF
 M_musculus 289 ERYTSRENTDIAPYTTC L NGIWEQNTNR TRQDAQS ILVPVKSSVVPVEGCPALPHKVAH CD AD TGGSDHF
 H_sapiens 289 ERYTSRENTDIAPYTTC L NGIWEQNTNR TRQDAQS ILAPGKFSFAGVEGCPALPHKVAH CD AD TGGSDHF
 S_cerevisiae 231 TSRGPPEIIPQ AD F NCILSKPIAPFTN MEKNNPN RRRTVDV ADTTNPHNPI
 C_albicans 243 TAKGPPEIIPQ SD F NCILSKPIAPFTN KEI NNENRK RTTVVDV ADTTNPHNPI
 Silicibacter sp 214 TA-GPPEIIPQ HD F NCILARPGTVF P KEALTAA TTAGD ACD PDSPVP
 Roseobacter sp 214 TA-GPPEIIPQ HN F NCILARPGTVF P ASAKVAP TSVIGD ACD PDSPVKV

 Nostoc sp 247 ---V TELTE LKAG CA KC TLRVTEVRDEIHN
 Anabaena sp 247 ---V TELTE LKAG CA KC TLRVTEVRDEVHNVVVSRI
 Z_mays 362 INKSTSRFFR DPSKNSYHDDME CACVCLA DITP EFSKASOHE NIT
 A_thaliana 361 VNRAAL SFRR NPSNNSYDDMD CACVCLA DITP EFAKASOHE DIT
 M_musculus 366 MTECHT HFCM DADQIIHDSVEGSHL CSIDNTPAOLIEA ECE DM
 H_sapiens 366 MTECHT HFCM DADQIIHDSVEGSHL CSIDNTPAOLIEA ECE DM
 S_cerevisiae 292 YTVAVFNKPTVLVPTT ACPCY ISIDHHSILPREASEFESHDI
 C_albicans 304 YEIAVFNKPTVEVKLD KCPK SCIDHHSILPREASEFESHDI
 Silicibacter sp 268 YSRATVWAPVVRVAN DPVD MAIDNHSILPREASEFESHDI
 Roseobacter sp 268 YDRITWGAALRVHD RPVD TAINHSILPREASEFESHDI

 Nostoc sp 281 SEHIREGHILLAGLINRALLIDAGGSFOLNF NLGEONOSTSAAVRNSP SHEVMEITSRIDICA ---VDL
 Anabaena sp 294 SEHIREGHILLAGLINRALLIDAGGSFOLNF NLGEONOSTSAAVRNSP SHEVMEITSRIDICA ---VDL
 A_thaliana 416 NIVNLSGHILFDRFLINEALDMIEAGGSFHLAKC ELGOSADAEYSSEVAGAD DKKVLIQIDSTRLANPNEIYIS
 Z_mays 415 NIVNLSGHILFDRFLINEALDMIEAGGSFHLAKC ELGOSADAEYSSEVAGAD DDTATLKIQIDSTRLANPNEIYIS
 M_musculus 420 STVNSLSGHILFDRFLINEALDMIEAGGSFHLAKC ELGOSADAEYSSEVAGAD DDTATLKIQIDSTRLANPNEIYIS
 H_sapiens 420 STVNSLSGHILFDRFLINEALDMIEAGGSFHLAKC ELGOSADAEYSSEVAGAD DDTATLKIQIDSTRLANPNEIYIS
 S_cerevisiae 338
 C_albicans 376
 Silicibacter sp 314
 Roseobacter sp 314

 Nostoc sp 375 PDER LAIEFVIONG APDFYVSIYPTVEV
 Anabaena sp 371 PDER LAIEFVIONG APDFYVSIYPTVEV
 A_thaliana 496 PDER LAIEFVIONG APDFYVSIYPTVEV
 Z_mays 492 PDER LAIEFVIONG APDFYVSIYPTVEV
 M_musculus 420 KKK ILGSGCYSGSVYINS RDNNITIGSDMTNQM
 H_sapiens 420 RRR ILGSGCYSGSVYINS RDNNITIGSDMTNQM
 S_cerevisiae 338 MGKN ILGSGCYSGSVYINS RDNNITIGSDMTNQM
 C_albicans 376 MVKK ILGSGCYSGSVYINS RDNNITIGSDMTNQM
 Silicibacter sp 314 MT HWCCGLSALGRRILE AGHVAVWNRTPKA
 Roseobacter sp 314 MT HWCCGLSALGRRILE AGHVAVWNRTPKA

 Nostoc sp 410 ITQTPSG VARCK RDP EVGERVIVDVLHR ---T RK ESREQRSTOEF SFMSAGV SSERR ETVEVQVWELRKR
 Anabaena sp 406 ISQTPNG LARCK RDP KAGEQVIVDVLHR ---T RK ESREQRSTOEF SFMSAGV SSERR ETVEVQVWELRKR
 Z_mays 556 IGSQSD SOV V SLPKASFEAA AGVCHL KH VIASY DES SNTSQAAGVAG TCE GLDPGIDHVA S K H
 A_thaliana 562 SESLKY SOVD V SLPKASFEAA AGVCHL KH VIASY DDETSMTHEKAKSAG TCE GLDPGIDHVA S K H
 M_musculus 478 EAKQOS ESOD V SLPKASFEAA AGVCHL KH VIASY DDETSMTHEKAKSAG TCE GLDPGIDHVA S K H
 H_sapiens 478 EAKQOS ESOD V SLPKASFEAA AGVCHL KH VIASY DDETSMTHEKAKSAG TCE GLDPGIDHVA S K H
 S_cerevisiae 390 DSAIDK ADND V SLPKASFEAA AGVCHL KH VIASY DDETSMTHEKAKSAG TCE GLDPGIDHVA S K H
 C_albicans 430 AAKQDEO AKHD V SLPKASFEAA AGVCHL KH VIASY DDETSMTHEKAKSAG TCE GLDPGIDHVA S K H
 Silicibacter sp 366 IARISE SPAD V SLPKASFEAA AGVCHL KH VIASY DDETSMTHEKAKSAG TCE GLDPGIDHVA S K H
 Roseobacter sp 366 INNIAE SPAD V SLPKASFEAA AGVCHL KH VIASY DDETSMTHEKAKSAG TCE GLDPGIDHVA S K H

 Nostoc sp 486 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D
 Anabaena sp 482 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D
 Z_mays 636 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D
 A_thaliana 642 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D
 M_musculus 558 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D
 H_sapiens 558 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D
 S_cerevisiae 470 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D
 C_albicans 510 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D
 Silicibacter sp 446 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D
 Roseobacter sp 446 DAGGK ---VVVTAGPV IHTGGG ---EHLAQLVREGY QATLG ---CNA AVHDIEON MGT LG D

 Nostoc sp 544 KRCVA RGGRHRLKVINR RRYGS IP A EACI RSGVMYEC H
 Anabaena sp 540 KRCVA RGGRHRLKVINR RRYGS IP A EACI RSGVMYEC H
 Z_mays 711 ELEFAEHLNRLSLI G YGHSK EASIR RGRTR EG SMI VTLSTCFDAAHPLLODTSRP TYKGLDEL
 A_thaliana 717 NLEFAEHLNRLSLI G YGHSK EASIR RGRTR EG SMI VTLSTCFDAAHPLLODTSRP TYKGLDEL
 M_musculus 632 YFEGLN EGYPNRDSIK A YGIS SAH RGRTR KCSKA NGFVRLCLINREAPALRPEANPLTWKQLCDL
 H_sapiens 632 YFEGLN EGYPNRDSIK A YGIS SAH RGRTR KCSKA NGFVRLCLINREAPALRPEANPLTWKQLCDL
 S_cerevisiae 543 YEGYV EGYPNRDSITL K YHIP EAEH RGRTR OG PEF KALVDMCLKDDANEIFS KPIAWNEALKQY
 C_albicans 583 YEGYV EGYPNRDSITL K YHIP EAEH RGRTR OG PEF KALVDMCLKDDANEIFS KPIAWNEALKQY
 Silicibacter sp 520 ATPETFEYYPNRDSL P M QHFDKDWV EF RGTIRLNG AAWADVFE
 Roseobacter sp 520 SAPESFEYYPNRDSL P M QHFDKDWV EF RGTIRLNG AAWADVFE

 Nostoc sp 747 ---NNVPFV AGS R DGP
 Anabaena sp 743 ---NNVPFV AGS R DGP
 Z_mays 1015 NN ---ISTINTDLIEASGGYDDDLIARLL NLGCKNKEIAVKTVKTIK LGLHE TO PKGCSPPFV COR EOR A
 A_thaliana 1020 SN ---ILNKDADNESEPLAG EEEISKRII LGLHS KETAAKAAKTIV LGFNE RE PSLCKSVF DATCYL E K A
 M_musculus 924 V G ---ISRSPCEKLEVVFT LGGDN ---TOLEAAE LGLLG EQ PO AESI DAFSKH VSK S
 H_sapiens 924 V G ---ISRSPCEKLEVVFT LGGDN ---TOLEAAE LGLLG EQ PO AESI DAFSKH VSK S
 S_cerevisiae 829 V G ---AKSTSKEDLIASIDSKATW DDEGRE ---RILSGFA LGLFS AK TP RGNA DT CAR E L O
 C_albicans 868 V G ---AKSTSKEDLIASIDSKATW DDEGRE ---RILSGFA LGLFS AK TP RGNA DT CAR E L O
 Silicibacter sp 758 ---V TLEGPEGDAR KE SDQFW ENA
 Roseobacter sp 758 ---V TLEGPEGDAR KE SDQFW ENA

 Nostoc sp 763 LPDPTOMD QAQEQ AKHLEG EM LML SMPSHC GNMTAGVVMVCVD NPA VTK SDRS ESVG TVDG LF
 Anabaena sp 759 LPDPTOMD QAQEQ AKHLEG EM LML SMPSHC GNMTAGVVMVCVD NPA VTK SDRS ESVG TVDG LF
 Z_mays 1092 VGHNEOD V HHEVE EYPD OPAEK OATIE CK ENGRST AMALTVG PAA AL KKK QTKG IRLQD
 A_thaliana 1094 VGHNEOD V HHEVE EYPD OPAEK OATIE CK ENGRST AMALTVG PAA AL KKK QTKG IRLQD
 M_musculus 984 VGPPEKDI V RDSG RHPS H EH TIDIA V GDFNG FAMA TVG PTA PAK DGEHAKG MPFTK
 H_sapiens 984 VGPPEKDI V RDSG RHPS H EH TIDIA V GDFNG FAMA TVG PTA PAK DGEHAKG MPFTK
 S_cerevisiae 893 VEDNPRD V VQHK GIEWAD TKET ISTID CK GG Y SMATTVGYEVA PAK DGEHAKG MPFTK
 C_albicans 932 VEGKPRD V VQHK GIEWAD TKET ISTID CK GG Y SMATTVGYEVA PAK DGEHAKG MPFTK
 Silicibacter sp 783 VDEGPDRV V VQHK GIEWAD TKET ISTID CK GG Y SMATTVGYEVA PAK DGEHAKG MPFTK
 Roseobacter sp 783 VDEGPDRV V VQHK GIEWAD TKET ISTID CK GG Y SMATTVGYEVA PAK DGEHAKG MPFTK

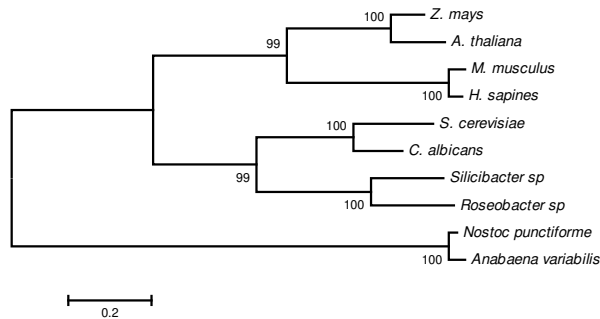
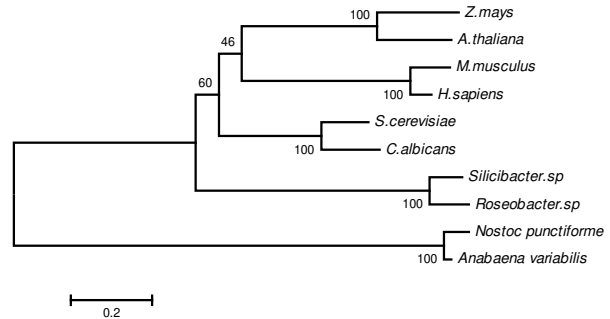
B**C**

Figura 14 - (A), Alinhamento das sequências de aminoácidos dos polipeptídios que flanqueiam o ID de plantas e cianobactérias conjuntamente com as sequências de aminoácidos da LKR e SDH de animais, leveduras e α -proteobactérias. As sequências de aminoácidos do ID de plantas e cianobactérias estão evidenciadas em vermelho. (B), árvore filogenética das sequências da LKR de plantas, animais, α -proteobactérias, levedura e da correspondente em cianobactérias. (C) árvore filogenética das sequências da SDH de plantas, animais, α -proteobactérias, levedura e da correspondente em cianobactérias.

Tabela 8 - Matriz de similaridade entre as sequências da LKR, SDH e ID de plantas, animais, leveduras, cianobactérias e α -proteobactérias

LKR	<i>Nostoc sp</i>	<i>Anabaena sp</i>	<i>Z. mays</i>	<i>A. thaliana</i>	<i>M. musculus</i>	<i>H. sapiens</i>	<i>S. cerevisiae</i>	<i>C. albicans</i>	<i>Silicibacter sp</i>	<i>Roseobacter sp</i>
<i>Nostoc sp</i>		0,91	0,10	0,09	0,07	0,07	0,09	0,10	0,09	0,09
<i>Anabaena sp</i>	0,91		0,09	0,08	0,07	0,07	0,09	0,10	0,09	0,10
<i>Z. mays</i>	0,10	0,09		0,74	0,37	0,38	0,18	0,17	0,18	0,18
<i>A. thaliana</i>	0,09	0,08	0,75		0,36	0,37	0,17	0,17	0,20	0,18
<i>M. musculus</i>	0,07	0,07	0,37	0,36		0,90	0,19	0,17	0,18	0,18
<i>H. sapiens</i>	0,07	0,07	0,38	0,37	0,90		0,19	0,17	0,19	0,19
<i>S. cerevisiae</i>	0,09	0,09	0,18	0,17	0,19	0,19		0,62	0,42	0,40
<i>C. albicans</i>	0,10	0,10	0,17	0,17	0,17	0,17	0,62		0,38	0,37
<i>Silicibacter sp</i>	0,09	0,09	0,18	0,19	0,18	0,19	0,42	0,38		0,69
<i>Roseobacter sp</i>	0,09	0,10	0,18	0,17	0,18	0,19	0,40	0,37	0,69	

SDH	<i>Nostoc sp</i>	<i>Anabaena sp</i>	<i>Z. mays</i>	<i>A. thaliana</i>	<i>M. musculus</i>	<i>H. sapiens</i>	<i>S. cerevisiae</i>	<i>C. albicans</i>	<i>Silicibacter.sp</i>	<i>Roseobacter.sp</i>
<i>Nostoc sp</i>		0,93	0,08	0,08	0,07	0,08	0,10	0,10	0,08	0,10
<i>Anabaena sp</i>	0,93		0,08	0,08	0,07	0,08	0,10	0,10	0,08	0,10
<i>Z. mays</i>	0,08	0,08		0,60	0,30	0,31	0,30	0,29	0,21	0,20
<i>A. thaliana</i>	0,08	0,09	0,60		0,30	0,30	0,30	0,30	0,21	0,21
<i>M. musculus</i>	0,07	0,07	0,30	0,30		0,85	0,37	0,39	0,21	0,20
<i>H. sapiens</i>	0,08	0,08	0,31	0,30	0,85		0,38	0,41	0,22	0,21
<i>S. cerevisiae</i>	0,10	0,10	0,30	0,30	0,37	0,38		0,71	0,25	0,24
<i>C. albicans</i>	0,10	0,10	0,29	0,30	0,39	0,41	0,71		0,25	0,24
<i>Silicibacter.sp</i>	0,08	0,08	0,21	0,21	0,21	0,22	0,25	0,25		0,83
<i>Roseobacter.sp</i>	0,10	0,10	0,20	0,21	0,20	0,21	0,24	0,24	0,83	

ID	<i>Nostoc sp.</i>	<i>Anabaena sp.</i>	<i>Z. mays</i>	<i>A. thaliana</i>
<i>Nostoc sp.</i>		0,85	0,30	0,31
<i>Anabaena sp.</i>	0,85		0,29	0,31
<i>Z. mays</i>	0,30	0,29		0,63
<i>A. thaliana</i>	0,31	0,31	0,63	

Amplificação do gene LKR/SDH de Silicibacter pomeroyi.

Utilizando primers desenhados para os extremos do gene LKR e SDH foi possível amplificá-los a partir de DNA genômico de *Silicibacter pomeroyi* DSS-3. Quando utilizado os dois primers das porções 5'e 3' de LKR em uma reação da PCR, observou-se, após aplicação do resultado da reação em gel de agarose 1%, uma banda única com peso aproximado de 1100pb que se aproxima dos 1050pb que é do tamanho da esperado para o gene segundo dados genômicos (Fig.15). Quando os primers para SDH foram utilizados observou-se a amplificação de um fragmento um pouco maior, porém ainda próximo ao padrão de 1100pb, dado que também está de acordo com os 1140pb esperados segundo dados genômicos. Por fim quando os primers 5' de LKR e 3'de SDH foram combinados em uma reação, uma amplificação de um fragmento com peso pouco superior a 2000pb foi observado, indicando que esses genes estão justapostos no genoma de *Silicibacter pomeroyi* (Fig. 15)

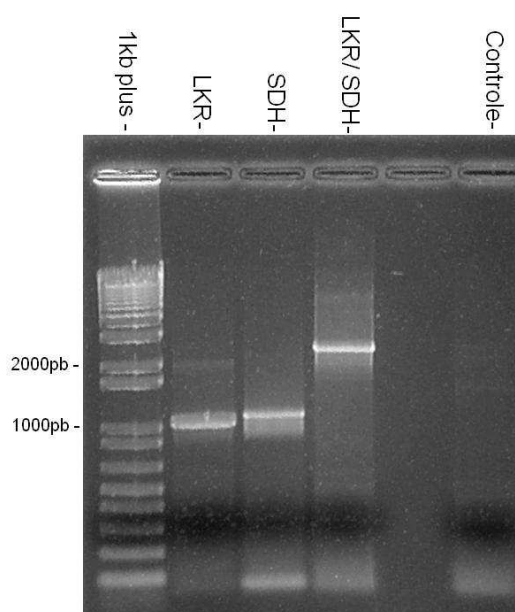


Figura 15 - Gel de agarose 1% do resultado da amplificação dos genes LKR e SDH em *Silicibacter pomeroyi*, utilizando diferentes combinações de primers. Na canaleta LKR/SDH, nota-se a soma do tamanho dos fragmentos presentes nas canaletas LKR e SDH quando se usa os primers de 5' de LKR e 3' de SDH em uma mesma reação.

Ensaio enzimático das de LKR e da SDH de *Silicibacter pomeroyi*

Foi possível observar as atividades de LKR e de SDH em extrato protéico bruto de *Silicibacter sp.* A atividade da LKR é medida através da oxidação do NADPH dependente de lisina e a atividade da SDH pela redução de NAD⁺ dependente de sacaropina assim como descrito em (Kemper et al.,1998) . Tanto para LKR como para SDH, o consumo do substrato aumentou linearmente com o tempo, assim como com a quantidade de proteína adicionada ao ensaio (Fig. 16). Ensaio conduzidos com extrato protéico bruto de *Nostoc sp* não mostraram atividade de LKR nem de SDH. Também não foi possível detectar sinal na reação de western blot utilizando anticorpos anti LKR/SDH de milho em nenhuma das duas espécies.

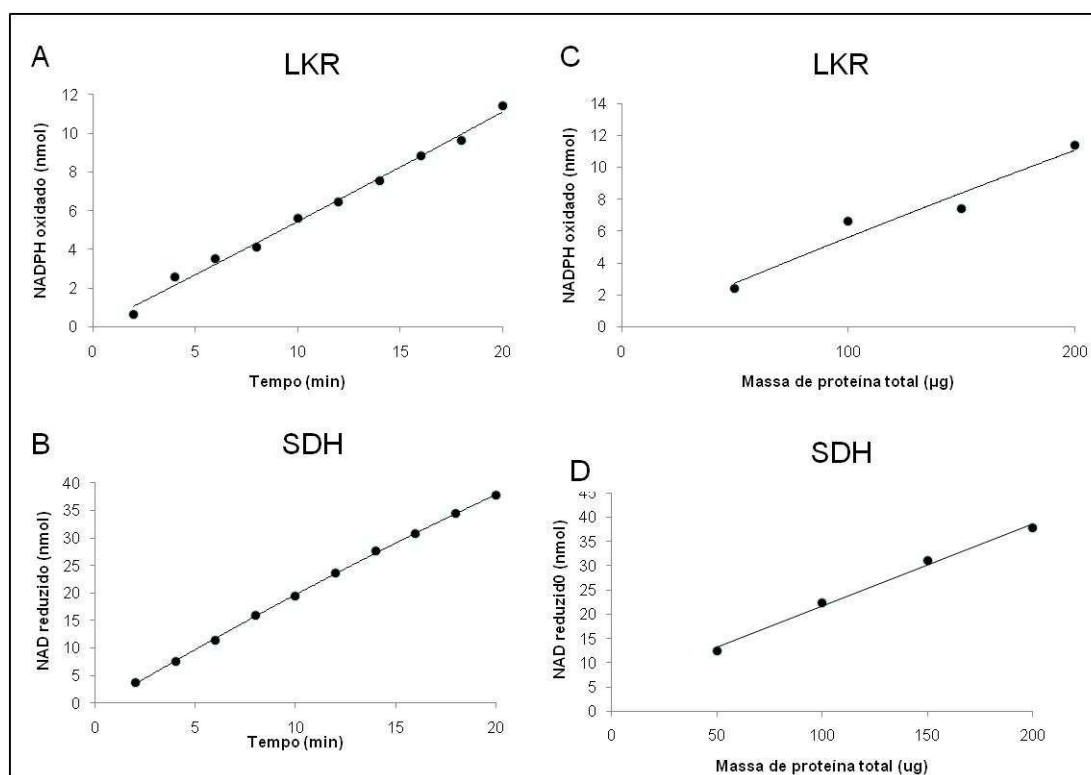


Figura 16 - Ensaio enzimático de LKR e SDH. **(A)** Oxidação de NADPH em função do tempo indicando atividade de LKR. **(B)** Redução de NAD⁺ em função do tempo

indicando atividade de SDH. **(C)** Oxidação de NADPH em função da massa de proteína aplicada. **(D)** Redução do NAD⁺ em função da massa de proteína aplicada.

Discussão

Os resultados obtidos nas buscas utilizando-se tBLASTn com as sequências iscas de LKR, ID e SDH de plantas e animais retornaram hits com valores altos e possibilitaram identificar sequências possivelmente homólogas a LKR e SDH em várias bactérias. Os valores mais significativos apresentados foram observados nos grupos CFB e proteobactérias. Já o ID, foi encontrado principalmente em cianobactérias e não foi encontrado em CFB e proteobactérias. Embora o ID nas cianobactérias estejam intercalando sequências polipeptídicas, não foi possível identificá-las como sendo homólogas a LKR e SDH. Esses resultados podem ser claramente visualizados pelo alinhamento das sequências de aminoácidos mostrando alta similaridade em diversas regiões entre LKR/SDH das α -proteobactérias com plantas, animais e fungos, mas não com as cianobactérias, que apresentam similaridade somente com o ID de plantas. A comprovação da presença dos genes de LKR/SDH no genoma de *Silicibacter sp* foi obtida com a amplificação dos fragmentos correspondentes (Fig.15). A funcionalidade das enzimas LKR e SDH em *Silicibacter sp* foi demonstrada através da atividade das duas enzimas no sentido da degradação encontradas em extratos protéicos brutos da bactéria (Fig. 16). Por sua vez não foi possível observar atividade similar em *Nostoc sp*, sugerindo que os domínios que falqueiam o ID realmente não possuem atividade de LKR/SDH.

Entre os possíveis homólogos a LKR/SDH encontrados em bactérias, o único que apresentou organização que poderia gerar um polipeptídeo bifuncional como o de animais e plantas foi o encontrado em *Silicibacter sp*. (Fig. 15). Nessa bactéria, os genes que codificam a LKR e a SDH encontram-se em um operon que possui também outros dois genes; um que codifica a Glutathione-S-transferase e outro que codifica a Fosfogliceraldeído Mutase. Essas enzimas foram diversas vezes associadas ao controle de espécies reativas de oxigênio quando os organismos são submetidos a condições de estresse oxidativo (Leiers et al., 2003, Motshwene et al., 2003) . Isso nos

leva a supor que em *Silicibacter* a LKR/SDH também pode apresentar funções similares principalmente por que a bactéria habita todos os tipos de ambientes graças a sua grande capacidade de metabolizar compostos tóxicos (Moran et al., 2001). Essa comprovação reforçaria um novo conceito a respeito da via de degradação de lisina atuando também no combate ao estresse oxidativo como já sugerido por alguns autores (Battur et al., 2009; Deleu et al., 1999).

A análise filogenética utilizando os alinhamentos das sequências de aminoácidos da LKR e SDH dos diversos organismos reforça a anotação das sequências encontradas em α -proteobactérias como prováveis ortólogos da LKR/SDH de plantas e animais. As matrizes de identidades (Tabela 5) das sequências de aminoácidos de LKR, ID e SDH dos diversos organismos analisados confirmam esses dados onde fica evidente que α -proteobactérias são mais similares a fungos seguido de animais e plantas, sugerindo uma origem única a partir de um ancestral comum que teria passado esses genes a todos eucariotos. Mas como teria surgido o ID presente somente em plantas e porque ele está presente em cianobactérias intercalando sequências de aminoácidos aparentemente não relacionadas a LKR e SDH?

A análise filogenética da LKR/SDH de plantas, animais e bactérias e a organização desses genes de forma a gerar um peptídeo bifuncional, sugere que os genes que codificam essas enzimas em plantas, animais e fungos podem ter se originado em um ancestral do grupo das α -proteobactérias ao qual pertence a bactéria *Silicibacter sp.* Levando-se em consideração que *Silicibacter sp.* apresenta tanto os genes justapostos como uma segunda cópia desses separados no genoma, é possível que ambas as cópias tenham sido transferidas para os eucariotos, porém, em fungos a “versão” dos genes separados foi mantida, enquanto nos demais eucariotos a “versão” bifuncional foi a que perseverou. Possivelmente isso esteja relacionado com a propriedade de os fungos utilizarem essas enzimas para síntese de lisina, enquanto animais e plantas as usam para degradação.

Analizando as possíveis explicações evolutivas para como esse cenário genômico possa ter surgido chegamos a uma teoria para a herança desse genes que é explicada a seguir e diagramada na figura 17.

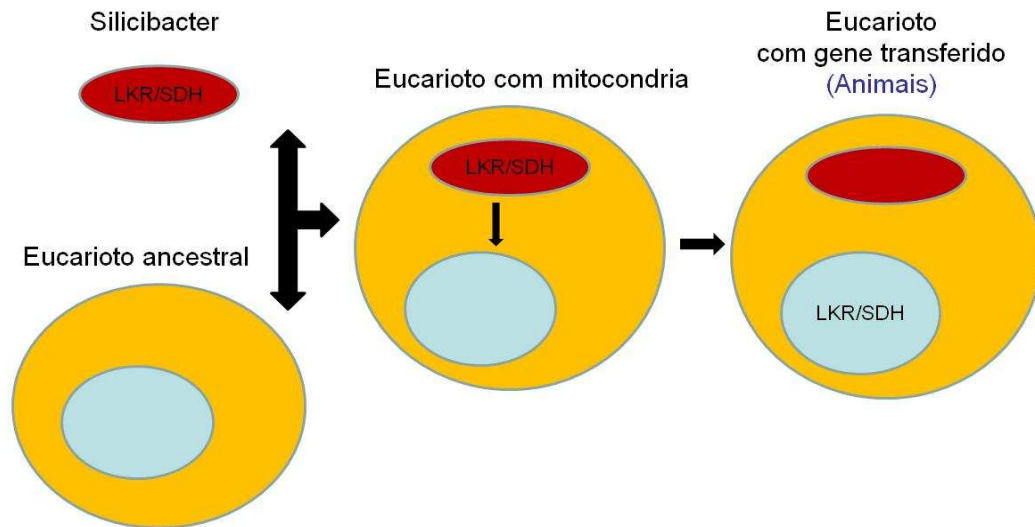
Acredita-se atualmente que o surgimento da mitocôndria, presente somente nos eucariotos, tenha ocorrido a partir de uma endossimbiose gerada após a fagocitose de uma α -proteobactéria por uma célula eucariótica ancestral (Searcy, 2003). A esse processo seguiu-se uma série de adaptações de ambos organismos, sendo uma delas a transferência para o genoma nuclear ou perda de mais de 95% do genoma da bactéria fagocitada (Martin et al., 2002). Assim sendo, é esperado que, da mesma forma que a maioria do genoma, a LKR e SDH do endossimbionte tenha sido transferida para o genoma nuclear do eucarioto. Esse cenário está de acordo com os dados genômicos e filogenéticos obtidos para *Silicibacter sp.*, animais e plantas apresentados nesse trabalho.

Entretanto, como explicar a existência do ID encontrado intercalando os domínios LKR e SDH em plantas? Sabe-se que os cloroplastos surgiram por um fenômeno de endossimbiose semelhante ao do ocorrido com as mitocôndrias, porém posterior e envolvendo procariotos pertencentes ao grupo das cianobactérias (Wang et al., 1999). Sabe-se também que da mesma forma, ao longo da evolução, houve grande fluxo gênico do genoma cloroplasto para o genoma nuclear. Da mesma forma, é esperado que, como a maioria do genoma, o fragmento, hoje denominado ID, tenha sido transferido para o genoma nuclear do vegetal ancestral.

Levando-se em conta o que foi dito, é possível que os genes que codificam a LKR e a SDH dos organismos superiores tenham se originado na aquisição da mitocôndria, a partir da α -proteobactérias fagocitada, nesse trabalho representada por *Silicibacter sp.* Posteriormente, na aquisição do cloroplasto, aqui representado pela *Nostoc sp.*, o inter-domínio foi adquirido, explicando a presença desse fragmento exclusivamente nas plantas. Apesar de existir baixa similaridade entre os domínios que flanqueiam o ID com LKR/SDH, essa similaridade pode ter sido suficiente para favorecer que a inserção do ID tenha ocorrido entre esses domínios por recombinação. Esse cenário evolutivo explica a estrutura atual desses genes, porém, não é a única alternativa possível. O gene completo de cianobactérias pode ter sido o ancestral dos genes de eucariotos e o ID pode ter sido perdido por todos os organismos menos

plantas. Entretanto, essa hipótese é pouco provável já que os dados de similaridade indicam o oposto além de esse caminho evolutivo ser menos parcimonioso.

A



B

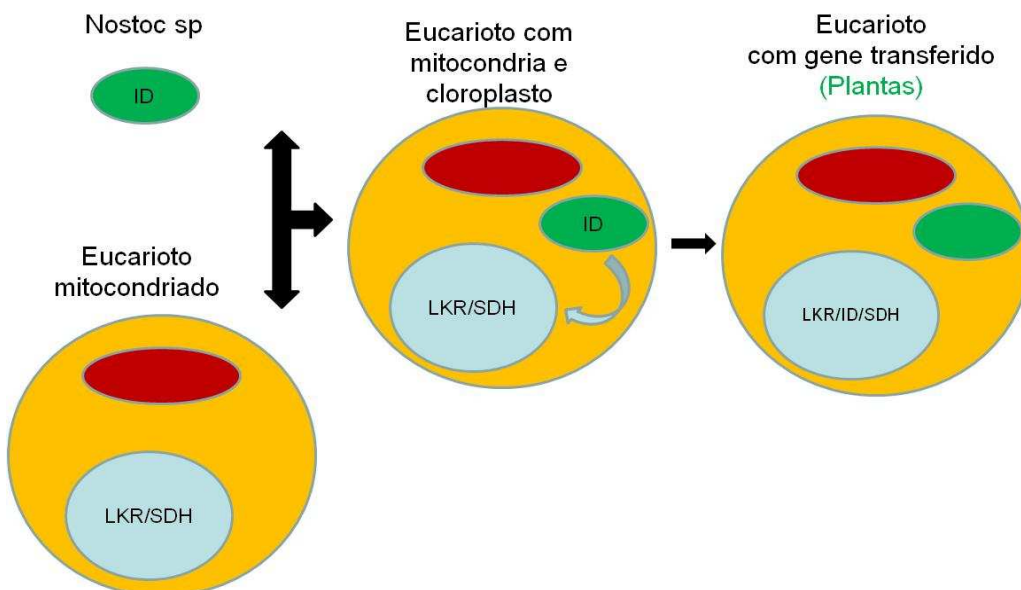


Figura 17 - Diagrama explicativo da teoria criada para explicar a distribuição atual dos genes LKR/SDH e do inter-domínio (ID). **A** – primeiro passo evolutivo, onde um

eucarioto fagocitou uma α -proteobactéria e essa teve seus genes, inclusive LKR/SDH transferido para o genoma do eucarioto; **B** – Segundo passo evolutivo, onde o eucarioto já com a mitocôndria fagocitou uma cianobactéria e essa teve seu genes, inclusive o ID, transferidos para o genoma do vegetal primitivo.

Conclusão

O gene de degradação LKR/SDH está presente em vários grupos de bactérias e ativo em *Silicibacter sp.* A LKR/SDH mais próxima a de animais e plantas presente no gênero *Silicibacter*, pertencente à classe das α -proteobactérias, que se acredita deram origem as mitocôndrias. Logo uma possível origem para esse gene é a transferência do genoma mitocondrial para o genoma dos organismos superiores. O inter-domínio só está presente em cianobactérias. A esse grupo é creditado à origem do cloroplasto e possivelmente essa foi à origem desse peptídeo presente somente em plantas. Há indicativos que LKR/SDH teve origem monofilética.

Bibliografia

Arruda P, Sodek L, Dasilva W J (1982). Lysine-ketoglutarate reductase-activity in developing maize endosperm. *Plant Physiol* 69: 988-989.

Arruda P, Kemper E L, Papes F, Leite A (2000). Regulation of lysine catabolism in higher plants. *Trends Plant Sci* 5: 324-330

Azevedo R A, Arana J L, Arruda P (1990). Biochemical genetics of the interaction of the lysine plus threonine resistant mutant Ltr_19 with opaque-2 maize mutant. *Plant Sci* 70: 81–90

Azevedo R A, Arruda P, Turner W L, Lea P J (1997). The biosynthesis and metabolism of the aspartate derived amino acids in higher plants. *Phytochem* 46: 395–419

Azevedo R A, Lanciwn M and Lea P J (2006). The aspartic acid metabolic pathway, an exciting and essential pathway in plants. *Amino acids* 30: 143-162.

Azevedo R A, Lea P J (2001). Lysine metabolism in higher plants. *Amino Acids* 20: 261–279.

Battur B, Boldbaatar D, Umemiya-Shirafuji R, Liao M, Battsetseg V, Taylor D, Baymbaa B, Fujisaki K (2009). LKR/SDH Plays Important Roles throughout the Tick Life Cycle Including a Long Starvation Period. *Plos One* 4 : e7136.

Ben Tzvi-Tzchori I, Perl A, Galili G (1996). Lysine and threonine metabolism are subject to complex patterns of regulation in Arabidopsis. *Plant Mol Biol* 32: 727–734

Bhattacharjee J K (1985). Alpha-amino adipate pathway for the biosynthesis of lysine in lower eukaryotes, *Crit Rev Microbiol* **12**: 131–151

Bhattacharjee J K. Evolution of α -amino adipate pathway for the synthesis of lysine in fungi. In: R.P. Mortlock, Editor, *The Evolution of Metabolic Functions*, CRC Press, New York (1992). 47–80.

Bright S W J, and Shewry P R (1983). Improvement of protein quality in cereals. *CRC Crit. Rev. Plant Sci.* 1: 49-93.

Brinch-Pedersen H, Galili G, Knudsen S, Holm P B (1996). Engineering of the aspartate family biosynthetic pathway in barley (*Hordeum vulgare* L.) by transformation with heterologous genes encoding feed-back-insensitive aspartate kinase and dihydrodipicolinate synthase. *Plant Mol Biol* 32:611–620.

Carson N A J, Scally B G, Neill D W, Carre I J. (1968) Saccharopinuria: a new inborn error of lysine metabolism. *Nature* 218: 679 only.

Cohen G N, and Salnt-Glrons I (1987). Biosynthesis of threonine, lysine and methionine. In *Escherichia coli* and *Salmonella typhimurium: Cellular and Molecular Biology*, F.C. Neidhardt, ed (Washington, DC: American Society for Microbiology), 429-444.

Cowey C B and Forster J R M (1971). The essential amino-acid requirements of the prawn *Palaemon serratus*. The growth of prawns on diets containing proteins of different amino-acid compositions. *Mar Biol* 10: 77-81.

Craciun A, Jacobs M and Vauterin M (2000). Arabidopsis loss-of-function mutant in the lysine pathway points out complex regulation mechanisms. *FEBS Lett* 487: 234-238.

Creighton T H (1993). *Proteins: structures and molecular properties*. San Francisco: W. H. Freeman. chapter 1

Dancis J, Hutzler J, Cox R P, Woody N C. (1969). Familial hyperlysinemia with lysine-ketoglutarate reductase insufficiency. *J. Clin. Invest.* 48: 1447-1452.

Deleu C, Coustaut M, Niogret M F and Larher F (1999). Three new osmotic stress-regulated cDNAs identified by differential display polymerase chain reaction in rapeseed leaf discs, *Plant Cell Environ.* 22: 979–988

DeRose R T, Ma D P, Kwon I J, Hasmain S E, Kiassy R C, Hall T C (1989). Characterization of the kafirin gene family from sorghum reveals extensive homology with zein from maize. *Plant Mol Biol* 12: 245–256

Dreyfus M and Régnier P (2002). The Poly(A) Tail of mRNAs: Bodyguard in Eukaryotes, Scavenger in Bacteria. *Cell* 27: 611-613.

- Eggeling L (1993). Biology of L-lysine overproduction by *Corynebacterium glutamicum*. *Amino Acids* 6: 261–272.
- Esen A (1987). Proposed nomenclature for the alcohol-soluble proteins (zeins) of maize (*Zea mays* L.). *J Cereal Sci* 5:117–128
- Falco S C, Guida T, Locke M, Mauvais J, Sanders C, Ward R T & Webber P.(1995). Transgenic Canola and Soybean Seeds with Increased Lysine. *Bio/Tech.* 13: 577-582.
- Ferreira R R, Varisi V A, Meinhardt L W, Lea P J, Azevedo R A (2005). Are high-lysine cereal crops still a challenge? *Braz J Med Biol Res* 38: 985–994
- Finn R D, Tate J, Mistry J, Coggill P C, Sammut J S, Hotz H R, Ceric G, Forslund K, Eddy S R, Sonnhammer E L and Bateman A (2008). The pfam protein families database. *Nucl Ac Res Database Issue* 36: D281-D288
- Fitzgerald L M & Szmant A (1997). *Biosynthesis of 'essential amino acids by scleractinian corals*. *Biochem. J.* 322: 213 – 221
- Fjellstedt T A, Robinson J C (1975). Properties of partially purified saccharopine dehydrogenase from human placenta. *Arch Biochem Biophys* 171: 191-196'
- Frizzi A, Huang S, Gilbertson L A, Armstrong T A, Luethy M H, and Malvar T M (2007). Modifying lysine biosynthesis and catabolism in corn with a single bifunctional expression/silencing transgene cassette. *Plant Biotechnology Journal* 6: 13–21
- Frankard V, Ghislain M, and Jacobs M (1992). Two feedbackinsensitive enzymes of the aspartate pathway in *Nicotiana sylvestris*. *Plant Physiol* 99 : 1285-1293
- Fürst P, Stehle P (2004). What are the essential elements needed for the determination of amino acid requirements in humans? *Journal of Nutrition* 134: 1558S–1565S.
- Galili G (1995). Regulation of Lysine and Threonine Syntesis. *The Plant Cell* 7 : 899-906.
- Galili S, Guenoune D, Wininger S, Hana B, Schupper A, Ben-Dor B, Kapulnik Y (2000). Enhanced levels of free and protein-bound threonine in transgenic alfalfa (*Medicago*

sativa L.) expressing a bacterial feedback-insensitive aspartate kinase gene. Transgenic Res 9: 137–144

Galili G (2002). New insights into the regulation and functional significance of lysine metabolism in plants. Ann Rev Plant Biol 53: 27–43

Galili G (2004). New insights into the regulation and functional significance of lysine metabolism in plants. Annu Rev Plant Physiol Plant Mol Biol 53:27–43

Gardener M C and Gikkman M P (2001). Analysing variability in nectar amino acids: composition is less variable than concentration. Journal of Chemical Ecology. 27: 2545-2558.

Gaziola S A, Teixeira C M G, Lugli J, Sodek L, Azevedo R A (1997). The enzymology of lysine catabolism in rice seeds. Isolation, characterization, and regulatory properties of a lysine 2-oxoglutarate reductase - saccharopine dehydrogenase bifunctional polypeptide. Eur J Biochem 247: 364–371

Hall T A (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98.

Heremans B, Jacobs M (1997). A mutant of Arabidopsis thaliana (L.) Heynh. With modified control of aspartate kinase by threonine. Biochem Genet 35: 139–153

Hotopp J C D, Clark M E, Oliveira D C S G, Foster J M, Fischer P, Torres M C M, Giebel J D, Kumar N, Ishmael N, Wang S, Ingram J, Nene R V, Shepard J, Tomkins J, Richards S, Spiro D J, Ghedin E, Slatko B E, Tettelin H, Werren J H (2007). Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes. Science 317: 1753-1756.

Huang S S, Kruger D E, Frizzi A, D'Ordine R L, Florida C A, Adams W R, Brown W E, Luethy M H (2005). High-lysine corn produced by the combination of enhanced lysine biosynthesis and reduced zein accumulation. Plant Biotech J 3: 555-569

Hudson A O, Singh B K, Leustek T and Gilvarg C (2006). An LL-Diaminopimelate Aminotransferase Defines a Novel Variant of the Lysine Biosynthesis Pathway in Plants. *Plant Physiol.* 140: 292-301.

Hutzler J, Dancis J (1968). Conversion of lysine to saccharopine by human tissue. *Biochim Biophys Acta* 158: 62-69

Hutzler J, Dancis J. (1970). Saccharopine cleavage by a dehydrogenase of human liver. *Biochim. Biophys. Acta* 206: 205-214.

Jansen G R (1962). Lysine in human nutrition. Journ of Nutri 76: 1-35.

Jarret H W, Cooksy K D, Ellis B, and Anderson J M (1986). *Anal Biochem* 153: 189–198

Karchi H, Shaul O, and Galili G (1994). Lysine synthesis and catabolism are coordinately regulated during tobacco seed development. *PNAS* 91: 2577-2581.

Kasahara T and Kato T (2003). A new redox-cofactor vitamin for mammals. *Nature* 422: 832

Kemper E, cord-neto G, Capella A, Gonc M, Alves-butruile M, Azevedo R and Arruda P (1998). Structure and regulation of the bifunctional enzyme lysine-oxoglutarate reductase-saccharopine dehydrogenase in maize. *Eur J Biochem* 253: 720-729

Kerckaert I, Poll B T, Espeel M, Duran M, Roeleveld A B C, Wanders R J A and Roels F (2000). Hepatic peroxisomes in isolated hyperpipecolic acidemia: evidence supporting its classification as a single peroxisomal enzyme deficiency, *Archvi Inter J Pat* 43 : 459–465

Kimura M and Ohta T (1972). On the stochastic model for estimation of mutational distance between homologous proteins. *Journal of Molecular Evolution.* 2: 87-90.

Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T (2002). Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *PNAS* 99: 14280–14285.

Kumar S, Tamura K, and Nei M (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5:150-163.

Lasser G W, Allen W V (1976). The essential amino acid requirements of the Dungeness crab *Cancer magister*. *Aquaculture* 7:235-244

Lea P J, Chen Z H, Leegood R C, Walker R P (2001). Does phosphoenolpyruvate carboxykinase have a role in both amino acid and carbohydrate metabolism? *Amino Acids* 20: 225–241

Lea P J, Azevedo R A (2003). Primary products: plant amino acids. In: Thomas B, Murphy SJ, Murray BG (eds) *Encyclopaedia of Applied Plant Sciences*. Elsevier 3: 871–883

Lee S I, Kim H U, Lee Y H, Suh S C, Lim Y P, Lee H Y, Kim H I (2001). Constitutive and seed-specific expression of a maize lysine-feedback insensitive dihydrodipicolinate synthase gene leads to increased free lysine in rice seeds. *Mol Breed* 8: 75–84

Leiers B, Kampkötter A, Grevelding C G, Link C D, Johnson T E, Henkle-Dührsen K (2003). A stress-responsive glutathione S-transferase confers resistance to oxidative stress in *Caenorhabditis elegans*. *Free Radic Biol Med* 34:1405-15.

Lima S T C, Azevedo R A, Santoro L G (2003). Improved procedures for extraction of lysine 2-oxoglutarate reductase/saccharopine dehydrogenase (LOR/SDH) enzyme from *Phaseolus vulgaris* cultivars. *New Zeal J Crop Hort Sci* 31: 261–268

Lugli J, Campbell A, Gaziola S A, Lea P J, Azevedo R A (2002). Enzymes of lysine metabolism from *Coix lacryma-jobi* seeds. *Plant Physiol Biochem* 40: 25–32

Marchler-Bauer A, Anderson J B, Chitsaz F, Derbyshire M K, DeWeese-Scott C, Fong J H, Geer L Y, Geer R C, Gonzales N R, Gwadz M, He S, Hurwitz D I, Jackson J D, Ke Z, Lanczycki C J, Liebert C A, Liu C, Lu F, Lu S, Marchler G H, Mullokandov M, Song J S, Tasneem A, Thanki N, Yamashita R A, Zhang D, Zhang N and Bryant S H (2009). *CDD*: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 37: 205-10.

Markovitz PJ, Chuang DT (1987). The bifunctional aminoadipic semialdehyde synthase in lysine degradation. *J Biol Chem* 262: 9353-9358

Martin W, Rujan T, Richly E, Hansen A, Cornelson S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci* 99: 12246–12251

Matsumoto S, Yamamoto S, Sai K, Maruo K, Adachi M, Saitoh M and Nishizaki T. (2003). Pipecolic acid induces apoptosis in neuronal cells. *Brain Res* 980: 179–184

Moran M A, Buchan A, González J M, Heidelberg H F, Whitman W B, Kiene R P, Henriksen J R, King G M, Belas R, Fuqua C, Brinkac L, Lewis M, Johri S, Weaver B, Pai G, Eisen J A, Rahe E, Sheldon W M, Ye W, Miller T R, Carlton J, Rasko D A, Paulsen I T, Ren Q, Daugherty S C, Deboy R T, Dodson R J, Durkin A S, Madupu R, Nelson W C, Sullivan S A, Rosovitz M J, Haft D H, Selengut J & Ward N. (2004). Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* 432: 910-913

Moret Y, Juchault P, Rigaud T. (2001) *Wolbachia* endosymbiont responsible for cytoplasmic incompatibility in a terrestrial crustacean: effects in natural and foreign hosts. *Heredity* 86: 325-332

Mori M, Kawada T, Ono T & Torii K. (1991) Taste preference, protein nutrition and L-amino acid homeostasis in male Sprague-Dawley rats. *Physiol Behav* 49: 987-996

Motshwene P, Brandt W and Lindsey G. (2003). Significant quantities of the glycolytic enzyme phosphoglycerate mutase are present in the cell wall of yeast *Saccharomyces cerevisiae*. *Biochem. J.* 369 : 357-362

Moulinb M, Deleua C, Larhera F and Bouchereaua A. (2006). The lysine-ketoglutarate reductase–saccharopine dehydrogenase is involved in the osmo-induced synthesis of pipecolic acid in rapeseed leaf tissues. *Plant Physiol and Biochem* 44: 474-482.

- Houmard N M, Mainville J L, Bonin C P, Huang S, Luethy M H, Malvar T M (2007). High-lysine corn generated by endosperm-specific suppression of lysine catabolism using RNAi. *Plan Biotech J* 5:605-614)
- Negrutiu I, Cattoir-Reynearts A, Verbruggen I, Jacobs M (1984). Lysine overproducer mutants with an altered dihydrodipicolinate synthase from protoplast culture of *Nicotiana sylvestris*. *Theor Appl Genet* 68: 11–20
- Nishida H and Nishiyama M (2000). What Is Characteristic of Fungal Lysine Synthesis Through the α -Aminoadipate Pathway? *J Mol Evol.* 51: 299-302
- Ottoboni L M M, Leite A, Targon M L N, Crozier A and Arruda P (1990). Characterization of the storage protein in seed of *Coix-lacryma-jobi* var. *Adlay*. *J Agricult Food Chem* 38:631–635
- Oyanagi K, Miura R, Yamanouchi T (1970). Congenital lysinuria: A new inherited transport disorder of dibasic amino acids. *The Jour of Pediat* 77: 259-266
- Palfi G and Dezsai L (1968). Pipecolic acid as an indicator of abnormal protein metabolism in diseased plants, *Plant Soil* **29** : 285–291.
- Paris S, Wessel P M, Dumas R (2002) Overproduction, purification, and characterization of recombinant aspartate semialdehyde dehydrogenase from *Arabidopsis thaliana*. *Prot Express Purif* 24: 99–104
- Paris S, Viemon C, Curien G, Dumas R (2003). Mechanism of control of *Arabidopsis thaliana* aspartate kinase-homoserine dehydrogenase by threonine. *J Biol Chem* 278: 5361–5366
- Peduto A, Baumgartner M R, Verhoeven N M, Rabier D, Spada M, Nassogne M C, Poll B T T, Bonetti G, Jakobs C and Saudubray J M (2004). Hyperpipecolic acidemia: a diagnostic tool for peroxisomal disorders, *Mol Gen Metabol* 19: 224–230
- Perl A, Shaul O, Galili G (1992). Regulation of lysine synthesis in transgenic potato plants expressing a bacterial dihydrodipicolinate synthase in their chloroplasts. *Plant Mol Biol* 19: 815–823.

- Reese M G, Hartzell¹ G, Harris¹ N L, Ohler U, Abril J F and Lewis S E (2000). Genome annotation assessment in *Drosophila melanogaster*. *Genome Res* 10: 483–501
- Reyes A R, Bonin C P, Houmard N M, Huang S, Malvar T M (2009). Genetic manipulation of lysine catabolism in maize kernels. *Plan Mol Biol* 69 : 81-89
- Rognes S E, Lea P J, Miflin B J (1980). S-adenosylmethionine – a novel regulator of aspartate kinase. *Nature* 287: 357–359
- Sacksteder K A, Biery B J, Morrell J C, Goodman B K, Geisbrecht B V, Cox R P, Gould S J, Geraghty M T (2000). Identification of the alpha-amino adipic semialdehyde synthase gene, which is defective in familial hyperlysinemia. *Am J Hum Genet* 66: 1736–1743
- Sadoul K, Boyault C, Pabion M, Khochbin S (2008). Regulation of protein turnover by acetyltransferases and deacetylases. *Biochimie* 90 : 306–12.
- Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4 : 406-425.
- Schwarz R & Dayhoff M (1979). Matrices for detecting distant relationships. *Atlas of protein sequences* : 353 - 58. National Biomedical Research Foundation.
- Searcy D G (2003). Metabolic integration during the evolutionary origin of mitochondria. *Cell Research* 13: 229-238
- Shaul O, Galili G (1992a). Increased lysine synthesis in tobacco plants that express high levels of bacterial dihydrodipicolinate synthase in their chloroplasts. *Plant J* 2: 203–209
- Shaul O, Galili G (1992b). Threonine overproducing in transgenic tobacco plants expressing a mutant desensitized aspartate kinase from *Escherichia coli*. *Plant Physiol* 100: 1157–1163
- Shewry P R, Halford N G (2002). Cereal seed storage proteins: structures, properties and role in grain utilization. *J Exp Bot* 370:947–958
- Shewry P R, Lew E J L & Kasarda, D D (1981). *Planta* 153: 246-253

- Smriga M, Mori M & Torii K (2000). Circadian release of hypothalamic norepinephrine in rats in vivo is depressed during early L-lysine deficiency. *J. Nutr.* 130:1641-1643.
- Smriga M, Ghosh S, Mouneimne Y, Pellett P L and Scrimshaw N S (2004). Lysine fortification reduces anxiety and lessens stress in family members in economically weak communities in Northwest Syria. *PNAS* 22: 8285 – 8288
- Sodek L, Wilson C M (1970). Incorporation of leucine-C14 into protein in the developing of normal and opaque-2 corn. *Arch Biochem Biophys.* 140: 29–38
- Steege D A (2000). Emerging features of mRNA decay in bacteria. *RNA* 6: 1079–90.
- Stepansky A, Less H, Angelovici R, Aharon R, Zhu X, Galili G (2006). Lysine catabolism, an effective versatile regulator of lysine level in plants. *Amino Acids* 30:121–125.
- Tang G L, Zhu X H, Gakiere B, Levanony H, Kahana A, Galili G (2002). The bifunctional LKR=SDH locus of plants also encodes a highly active monofunctional lysine-ketoglutarate reductase using a polyadenylation signal located within an intron. *Plant Physiol* 130: 147–154
- Tewari-Singh N, Sen J, Kiesecker H, Reddy V S, Jacobsen H J, Guha-Mukherjee S (2004). Use of a herbicide or lysine plus threonine for nonantibiotic selection of transgenic chickpea. *Plant Cell Rep* 22: 576–583
- Thompson J D, Gibson T J, Plewniak F, Jeanmougin F and Higgins D G (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24:4876-4882.
- Umbarger H E (1978). Amino acid biosynthesis and its regulation. *Annu Rev Biochem* 47:532–606.
- Vauquelin L N, Robiquet P J (1806). The discovery of a new plant principle in *Asparagus sativus*. *Ann Chim* 57: 88–93.

- Velasco A M, Leguina J I, Lazcano A (2002). Molecular evolution of the lysine biosynthetic pathways. *J Mol Evol* 55: 445–459
- Wang D Y C, Kumar S, and Hedges S B (1999). Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc Biol Sc* 266:163-171.
- Wang X, Larkins B A (2001). Genetic Analysis of Amino Acid Accumulation in *opaque-2* Maize Endosperm. *Plant Physiol* 125: 1766–1777.
- Wernegreen J J (2002). Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet* 3 : 850–61.
- Woody N C (1964). Hyperlysinemia. *Am J Dis Child* 108: 543
- Wilson B J, Gray A C, Matthews B F (1991). Bifunctional protein in carrot contains both aspartokinase and homoserine dehydrogenase activities. *Plant Physiol* 97: 1323–1328
- Xiang Y H (2006). Bacterial Identification Based on 16S Ribosomal RNA Gene Sequence Analysis. *Advanced Techniques in Diagnostic Microbiology* 2: 323-332
- Young V R (1994). Adult amino acid requirements: the case for a major revision in current recommendations. *J Nutr* 124: 1517S–1523S
- Zabriskie T M and Jackson M D (2000). Lysine biosynthesis and metabolism in fungi. *Nat Prod Rep* 17: 85–97
- Zhu X H, Tang G L, Galili G (2000). Characterization of the two saccharopine dehydrogenase isozymes of lysine catabolism encoded by the single composite at AtLKR=SDH locus of arabidopsis. *Plant Physiol* 124: 1363–1371
- Zhu X, Galili G (2003). Increased lysine synthesis coupled with a knockout of its catabolism synergistically boosts lysine content and also trans regulates the metabolism of other amino acids in arabidopsis seeds. *Plant Cell* 15: 845–853.

Anexo I

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha dissertação/tese de Mestrado/Doutorado intitulada “Síntese e degradação de lisina em sistemas biológicos”:

() não se enquadra no § 3º do Artigo 1º da Informação CCPG 01/08, referente a bioética e biossegurança.

() está inserido no **Projeto CIBio/IB/UNICAMP** (Protocolo nº _____), intitulado _____;

(X) tem autorização da **Comissão de Ética em Experimentação Animal/IB/UNICAMP** (Protocolo nº 1226-1);

() tem autorização do **Comitê de Ética para Pesquisa com Seres Humanos/FCM/UNICAMP** (Protocolo nº _____);

() tem autorização de comissão de bioética ou biossegurança externa à UNICAMP.

Especificar: _____



Aluno:



Orientador: