

Seqüenciamento de genes expressos de larvas de 1º e 2º estádios de *Anopheles darlingi* (diptera:culicidae), da Amazônia.

Ketlen Christine Nunes de SOUZA¹; Carlos Gustavo Nunes da SILVA³; Míriam Silva RAFAEL².

¹Bolsista PIBIC INPA/CNPq; ² Orientador INPA/CPCS; ³ Colaborador UFAM/CAM

A malária ou paludismo é uma parasitose que afeta mais de 90 países no mundo, onde cerca de 2,4 bilhões de pessoas (40% da população mundial) convivem com os riscos de contágio, segundo o Ministério da Saúde/Fiocruz (2005). *Anopheles darlingi* é o principal vetor dessa enfermidade, que apresenta os maiores registros na região amazônica (Tadei e Dutary-Thatcher, 2000). A análise de DNA genômico de diversos organismos tem permitido a obtenção de bibliotecas de cDNA, por meio de métodos do DNA recombinante (Adams *et al.* 1991). Diante disso, a biblioteca de cDNA de larvas de 1º e 2º estádios de *A. darlingi*, já produzida por (Rafael *et al.*, 2006) foi utilizada para identificar seqüências de genes expressos (*Expressed Sequence Tags - ESTs*), realizar comparação com o genoma de *A. gambiae* e inferir sua função com o uso da bioinformática. Fêz-se a preparação de plasmídeos (*mini-prep*), segundo protocolo de Birnbaim e Doly (1979), seleção e quantidade das colônias recombinantes. Os produtos das *mini-preps* foram analisados em gel de agarose a 0,8% e fotodocumentados em aparelho *Eagle eye*. No seqüenciamento, utilizou-se o *primer* M13R: 5'-TAGGGAAACAGCTATGAC - 3' Reverso. Os produtos da reação de seqüenciamento foram purificados e submetidos à eletroforese por capilaridade, no *MegaBACE 1000 DNA Sequencing System* (Amersham biosciences - *GE heath care*™), pelo método Dideoxi (Sanger *et al.*, 1977). A leitura e análise dos dados gerados foram realizadas pelos programas de *ScoreCard* (MegaBACE™) e *Base Calling* (*Base Caller Cimaron 1.53 Slim Phredfy*). O rendimento das colônias recombinantes foi de $1,9 \times 10^5$ e das placas obtidas a partir do *Scorecard* foi 69% (seqüências de boa qualidade). Os clones seqüenciados apresentaram entre 601 e 700 pares de bases. A avaliação da qualidade das seqüências foi obtida com o programa *Phred* e em seguida o programa *CAP3* (Huang e Madan, 1999) foi utilizado no agrupamento das seqüências para a montagem de *clusters*. As seqüências foram submetidas ao banco de dados do *GenBank* (NCBI), por meio do programa *BLAST* (Altschul *et al.*, 1997) e alinhadas com a modalidade *BLASTx*, onde os resultados foram selecionados pelo índice estatístico de *e-value*, aceitando-se os valores $\leq e^{-10}$. Fêz-se a busca de *ESTs* de interesse, com ferramentas de bioinformática. As *ESTs* foram comparadas com o banco de dados do *GenBank*, onde encontrou-se 91 *contigs* e 71 *singlets*. A quantidade de placas (*mini-preps*), clones seqüenciados, porcentagem da eficiência, seqüenciamento e as seqüências obtidas de boa qualidade estão descritos na Tabela 1.

Tabela 1 – Resultados das mini-preps de larvas de 1º e 2º estádios de *A. darlingi*.

Nome da Placa	Clones seqüenciados	Valor do ScoreCard (%)	Seqüências de boa qualidade
DL 02	96	55	53
DL 03	96	69	66
DL 04	96	45	43
03	288	56	162

Das 288 seqüências analisadas, 162 foram validadas pelo *BLASTx*. As seqüências homólogas encontradas, com *e-value* $\leq e^{-10}$, foram agrupadas por categoria geral de função do *COG-Clusters of Orthologous Groups*, dentre as quais, a categoria geral Armazenagem e Processamento de Informações foi a mais freqüente, com maior número de seqüências (Gráfico 1). Para que uma predição de seqüência protéica seja confiável é necessário inferir relações de ortologia entre genes de diferentes espécies (Tatusov *et al.*, 2000), com a finalidade de classificar proteínas de genomas completamente seqüenciados. Os *COGs* foram classificados em 18 categorias funcionais (Riley, 1993), representando o percentual de organismos relacionados a cada grupo das categorias gerais. Nas categoriais gerais de Armazenagem e Processamento de Informações, a categoria funcional com maior freqüência foi **J**) tradução, incluindo estrutura ribossômica; em Não Caracterizado, **S**) foi a categoria funcional mais freqüente;

com a finalidade de classificar proteínas de genomas completamente seqüenciados. Os COGs foram classificados em 18 categorias funcionais (Riley, 1993), representando o percentual de organismos relacionados a cada grupo das categorias gerais. Nas categorias gerais de Armazenagem e Processamento de Informações, a categoria funcional com maior freqüência foi **J**) tradução, incluindo estrutura ribossômica; em Não Caracterizado, **S**) foi a categoria funcional mais freqüente; e em Metabolismo, a categoria **F**) transporte e metabolismo de nucleotídeo foi a mais freqüente (Gráfico 2).

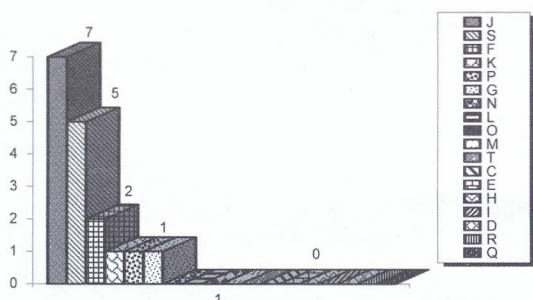


Gráfico 1 - Representação da porcentagem de genes homólogos encontrados por categorias gerais do COG no GenBank.

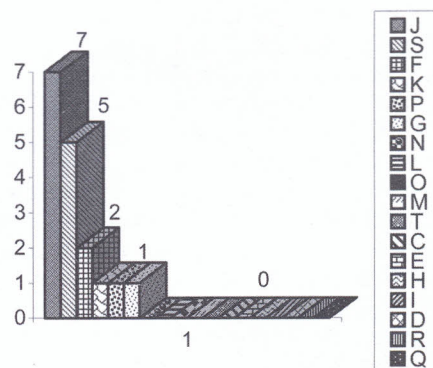


Gráfico 2 - Número de clones ortólogos encontrados em *A. darlingi* e suas prováveis funções.

A biblioteca de larvas de 1º e 2º estádios de *A. darlingi* foi validada, considerando que as seqüências foram comparadas e apresentaram similaridade com genes do complexo *Anopheles gambiae*, *Drosophila melanogaster* e de outros insetos depositados no GenBank. Além disso, a categoria **J** foi a que apresentou maior homologia com as seqüências dos organismos já citados. Comparações desses clones por alinhamento com outras seqüências de proteínas ribossomais poderão ser úteis para estudo de rota metabólica, identificação de genes de resistência a inseticidas, bem como genes reguladores da atividade de proteínas que atuam no desenvolvimento do plasmódio, visando o controle da malária.

Palavras-chave: Mosquitos, ESTs, malária

Fonte Financiadora: PIBIC/CNPq/INPA

Bibliografias citadas

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R., Kerlavage, A.R., McCombie, W.R., Venter, J.C. 1991. Complementary DNA sequencing: "expressed sequence tags" and the human genome project. *Science*, 252: 1651-1656.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. 1997.

Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25: 3389-3402.

Birnbaum, H.C., Doly J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7:1513 -1523.

Huang, X., Madan, A. 1999. CAP3: A DNA sequence assembly program. *Genome Research*, 9 (9): 868-877.

Ministério da Saúde, Fundação Oswaldo Cruz (2005). www.fiocruz.br/ccs/estetica/malaria.htm

Centro de Informação Científica e Tecnológica. Agência Fiocruz de notícias. 2005.

Riley, M. 1993. Functions of the gene products of *Escherichia coli*. *Revista de Microbiologia*, 57(4): 862-952.

Rafael, M. S. ; Bridi, L.C. ; Tadei, W. P. ; Nunes-silva, C. G. ; Astolfi-Filho, S . 2006. Biblioteca de cDNA de Larva de *Anopheles darlingi* (Diptera; Culicidae). *Anais do II Encontro de Genética do Norte: II Engenor*, Belém, PA., 1: 26-26.

Sanger, F.; Nicklen, S.; Coulson, A.R. 1977. DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences. USA.*, 74:5463-467.

Tadei, W.P., Dutary-Thatcher, B. 2000. Malaria vectors in the Brazilian Amazon: *Anopheles* of the subgenus *Nyssorhynchus*. *Revista do Instituto de Medicina Tropical*, 42 (2): 87-94.

Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1): 33-36.