

De novo transcriptome assembly and comparative expression profiling of midgut tissues of four non-model insects

Rajesh Kumar Gazara¹, Christiane Cardoso², Daniel Bellieny-Rabelo¹, Clélia Ferreira², Walter R. Terra², and Thiago Motta Venancio¹

1 Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro; Campos dos Goytacazes, Brazil.

2 Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil.

In the present work we describe the *de novo* transcriptome assembly and expression profiling of midgut tissues from four non-model insect species belonging to different orders: Lepidoptera (sp. *Spodoptera frugiperda*, a maize pest insect), Diptera (sp. *Musca domestica*, house fly, transmits human and animal diseases), Hemiptera (sp. *Dysdercus peruvianus*, a major cotton pest) and Coleoptera (sp. *Tenebrio molitor*, a storage grain pest). Total mRNA samples of posterior midgut, anterior midgut, whole midgut and carcass (i.e. body without midgut tissues) samples were submitted to pyrosequencing using a 454 instrument. Sequencing reads were filtered to remove low-quality and contaminant reads and assembled *de novo* with different algorithms. Different assemblies obtained with MIRA and Newbler were merged with CAP3, as this strategy has been previously shown to provide better results than the usage of a single algorithm. We obtained 6,395, 9,010, 4,005 and 6,833 unigenes for *D. peruvianus*, *T. molitor*, *M. domestica* and *S. frugiperda*. Differentially expressed genes were inferred by comparing different tissues within each species. A stepwise strategy was employed to functionally annotate unigenes. In total, 79.69 to 93.09% of unigenes were assigned with some functional information. A total of 2,970, 5,964, 2,506 and 3,627 unigenes from *D. peruvianus*, *T. molitor*, *M. domestica* and *S. frugiperda* were assigned to eggNOG orthologous groups (OGs), respectively. OG information was used to find functions that are commonly enriched among highly expressed genes the same tissues across different species, as well as to find gene families with divergent expression patterns across species and tissues. Particular gene families were explored in more detail, such as important digestive enzymes. Taken together, our results provide a collection of genes and gene families with critical roles in the development, maintenance and biochemistry of midgut tissues across divergent insect species.

Funding: The authors acknowledge Universidade Estadual do Norte Fluminense Darcy Ribeiro and the following Brazilian funding agencies for their support: A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).