PROFILING THE VENOM GLAND TRANSCRIPTOMES OF COSTA RICAN SNAKES BY 454 PYROSEQUENCING

J. Durban [(1)], P. Juárez [(1)], Y. Angulo [(2)], B. Lomonte [(2)], M. Flores-Diaz [(2)], A. Alape-Girón [(2)], M. Sasa [(2)], J.M. Gutiérrez [(2)], A. Conesa [(3)], J. Dopazo [(3)], J.J. Calvete [(1)].

[(1)] Instituto de Biomedicina de Valencia, CSIC, [(2)] Instituto Clodomiro Picado, San Jose, [(3)] CIPF, Valencia.

Central American herpetofauna includes 34 species of venomous snakes. Venom represents a trophic adaptation confering a selective advantage to the snake for the success in the colonization of, and adaptation to, novel hunting territories. However, in regions of sympatry with humans snakebites also represent a relevant, albeit neglected, public health issue. In Central America, 5000 snakebite envenomations occur every year. Application of snake venomics protocols has provided data on the number and distribution of protein families present in the venoms of Costa Rican snakes. To complement and extend this information we have employed the 454 high-throughput technology to investigate the transcriptional activity of the venom glands of *Bothriechis lateralis*, *Bothriechis schlegelii*, *Atropoides mexicanus*, *Atropoides picadoi*, *Crotalus simus*, *Cerrophidium godmani*, and *Bothrops asper* (Caribbean and Pacific variants). Because of the low data compression gained in the assembly step, the lack of a reference genome and the small difference between contigs and reads mean length, bioinformatic processing of the 454 sequence data was performed on whole sets of unassembled reads. To this end, the set of 330,010 nucleotide reads was searched against the non-redundant NCBI database and "best-hits" for entries of snake proteins were identified. The descriptors were analyzed and grouped into 20 documented snake venom protein families. For each species transcriptome, the number of reads for each venom protein family was calculated, generating a profile of relative abundances of the different families. This expression profile allowed us to group the 8 snake species in kinship groups. Besides, a representative full-length amino acid sequences from a phylogenetically closer species were used as template for the relative alignment of the tblastn hits, thus generating consensus sequences. The structural diversity within each gene family in each species as well as an estimation of the minimum number of genes per protein were also addressed.