Reference assembly of the Indian Cobra *Naja naja* provides a comprehensive view of its venom gland genes

Somasekar Seshagiri

Diretor Associado e Cientista da equipe da Genentech, Califórnia, EUA.

In India snakebites involving four venomous snake species dubbed the "big four" are responsible for over 50,000 deaths annually. Naja naja, the India cobra, is one of the "big four" venomous snakes found through out India. The genome of the Indian cobra has not been sequenced. Obtaining a well assembled genome and venom gland transcriptome will allow the identification of the complex venom components. Current anti-venoms are developed in horses and are often not effective and/or cause sever side effects. A well annotated genome will provide the template for the development of effective synthetic anti-venom antibodies with minimal side effects.

Using a recently developed technique in our lab, we sequenced single chromosomes (SChrom-seq) obtained by laser capture microdissection (LCM). Using genomic DNA we generated ~60 Gb of Oxford Nanopore and ~40 Gb of PacBio single molecule long read data, and ~120 Gb of short read Illumina data. Also, we generated ~300 Gb of BioNano optical mapping data. We also generated Hi-C data. We combined these data to produce a ~1.7 Gbp assembly that contained near chromosomal assemblies. The assembled genome has an N50 of 223Mbp. In this assembly ~96% of the genome is represented by 20 scaffolds that are >1Mbp..

In addition to the genome, we sequenced RNA from nine different tissues. In particular, we sequenced RNA from venom gland using PacBio technology and combined it with short read RNA-seq to obtain full-length transcripts. We used the predicted genes from the genome and used it to assess for differentially expressed across tissues. We have identified ~150 venom gland specific genes that includes, three finger toxins, phospholipase 2, venom metallo proteases, cysteine rich small proteins, natriuretic peptides, cobra venom factor, venom kunitz type proteins, venom factor lectins, and NGF-beta.