Information theory helps unravel DNA's genetic code

September 12 2012

'Superinformation,' or the randomness of randomness, can be used to predict the coding and noncoding regions of DNA.

DNA consists of regions called exons, which code for the synthesis of proteins, interspersed with noncoding regions called introns. Being able to predict the different regions in a new and unannotated genome is one of the biggest challenges facing biologists today.

Now researchers at the Indian Institute of Technology in Delhi have used techniques from [information theory](#) to identify DNA introns and exons an order of magnitude faster than previously developed methods.

The researchers were able to achieve this breakthrough in speed by looking at how electrical charges are distributed in the [DNA nucleotide](#) bases.

This distribution, known as the [dipole moment](#), affects the stability, solubility, melting point, and other physio-chemical properties of DNA that have been used in the past to distinguish exons and introns.

The research team computed the "superinformation," or a measure of the randomness of the randomness, for the angles of the dipole moments in a sequence of nucleotides. For both double- and single-strand forms of DNA, the superinformation of the introns was significantly higher than for the exons.

Scientists can use information about the coding and noncoding regions of DNA to better understand the [human genome](#), potentially helping to predict how cancer and other diseases linked to DNA develop.

 More information: "Dipole-entropy-based techniques for segmentation of introns and exons in DNA," is published in *Applied Physics Letters*, [apl.aip.org/resource/1/applab/v101/i8/p083701_s1](#)

Provided by American Institute of Physics