

Genetic phonetics could be the trick to sounding out DNA's meaning

Most modern attempts to decipher how portions of genetic code are translated into physical characteristics are akin to a first-grader trying to sound out a word letter by letter — or, in this case, base pair by base pair.

But University of Florida researchers have developed a computational method that's more like reading whole words at a time.

In a world where science's ability to transcribe an organism's genetic code is growing faster every day, the technique could offer much needed efficiency in translating the seemingly endless string of characters into information that can cure disease or create new crops.

The researchers, from UF's Institute of Food and Agricultural Sciences and the UF Genetics Institute, published their verification of the method in Wednesday's PLoS ONE, an online journal produced by the Public Library of Science.

"We worked very hard to find ways to collect genetic information," said Rongling Wu, the project's lead researcher and a UF Research Foundation professor. "We now must work hard to find ways to use it."

In many respects, researchers think of an organism's genome as ticker-tape listings of four letters — representing four amino acid bases — repeated in varying orders. The goal is to find meaning within the sequences, to figure out how variations in the pattern affect the organism's physiology.

Humans, for example, have 3 billion letters in our code. Between any two of us, 99.9 percent of those letters are the same. But it's that last 0.1 percent of difference, peppered throughout our DNA in the form of single-letter changes, that accounts for our unique identities—from eye color to disease susceptibility.

These differences are called single nucleotide polymorphisms, or SNPs (pronounced "snips").

The simplest way to find out how a SNP affects an organism is to collect a group of organisms that have different variations of that letter in their genetic code.

But physical traits are typically affected by multiple SNPs that interact in sometimes unpredictable ways — much like the way an "e" at the end of a word can change its pronunciation.

Fortunately, the rules of genetics say that SNPs that affect the same trait are generally related to each other in some way, such as being near each other.

Wu's model uses these rules in conjunction with statistical analysis of real data from genetically mapped organisms. As a result, the model can find whole groups of SNPs associated with a physical trait.

Just as an understanding of general phonetic principles allows a reader to sound out a whole word, this extra knowledge of genetics allows Wu's model to find whole pictures of genome/physical correlations.

"The real promise of Wu's work is that it could offer the opportunity for a researcher to not spend a really disheartening amount of time parsing out individual nucleotides, and move more directly to doing the type of genetic work that's going to have a greater significance," said Rory Todhunter, a researcher working with canine genetics at Cornell University.

In the paper, the researchers verified their model using genetic and physical information from mice that was

first collected from the Washington University lab of James Cheverud in the mid-1990s. They then compared their results with several years' worth of genetic analysis.

This validation was important, said Wei Hou, the first author of the paper and an assistant professor at UF's department of epidemiology and health policy research. But the analysis of modern data will be the real key to the technique's importance. For example, the mouse genetic information used in this paper featured only a few thousand SNPs. The July 29 issue of the journal *Nature* cited more than 8 million SNPs for the mouse genome.

"This shows how we need to move beyond looking at genomes SNP by SNP," Cheverud said. "Imagine the work that's ahead of us if we don't."

Source: UF

This document is subject to copyright. Apart from any fair dealing for the purpose of private study, research, no part may be reproduced without the written permission. The content is provided for information purposes only.