1014-37-747 **Suzanne S Sindi*** (`ssindi@math.umd.edu`), University of Maryland, Mathematics Department, College Park, MD 20742. *A Stochastic Dynamical System for the Evolution of Repeat Strings.* Preliminary report.

The genome (DNA sequence) of an organism is a sequence of letters {A,C,G,T}. Surprisingly, much of the genome of higher organisms consists of sequences that occur repeatedly. For example, it is estimated that nearly 50% of the human genome is a repetitive sequence.

Typically, researchers employ the concept of a *repetitive region* without a precise definition. We introduce the concept of a repeat string. A subsequence S of a genome W and fixed $k \in \mathbb{N}$ is a **repeat string** if every k-length subsequence in S occurs more than once in the genome W. We analyze the structure of repeat strings in two organisms: a nematode and a mustard seed. In both a variety of interesting and thus far unknown properties were found. Most surprisingly, we observed a power law distribution between the length of a repeat string and the number of repeat strings.

Our goal was to determine a theoretical model that could explain how such a distribution could have emerged. We developed a stochastic dynamical system to model the evolution of repeat strings in a genome employing two dominant genetic mechanisms: sequence duplication and point mutation. We show our model is a contraction, in the appropriate space and norm, and converges, to a power law. (Received September 23, 2005)