Abstract Submitted
for the MAR17 Meeting of
The American Physical Society

**Genome-Wide Motif Statistics are Shaped by DNA Binding Proteins over Evolutionary Time Scales** LONG QIAN, EDO KUSSELL, New York University — The composition of genomes with respect to short DNA motifs impacts the ability of DNA binding proteins to locate and bind their target sites. Since nonfunctional DNA binding can be detrimental to cellular functions and ultimately to organismal fitness, organisms could benefit from reducing the number of nonfunctional binding sites genome wide. Using in vitro measurements of binding affinities for a large collection of DNA binding proteins, in multiple species, we detect a significant global avoidance of weak binding sites in genomes. The underlying evolutionary process leaves a distinct genomic hallmark in that similar words have correlated frequencies, which we detect in all species across domains of life. We hypothesize that natural selection against weak binding sites contributes to this process, and using an evolutionary model we show that the strength of selection needed to maintain global word compositions is on the order of point mutation rates. Alternative contributions may come from interference of protein-DNA binding with replication and mutational repair processes, which operates with similar rates. We conclude that genome-wide word compositions have been molded by DNA binding proteins through tiny evolutionary steps over timescales spanning millions of generations.

Long Qian
New York University

Date submitted: 22 Oct 2016 Electronic form version 1.4