Profound Non-Randomness in Dinucleotide Arrangements within Ultra-Conserved Non-Coding Elements and the Human Genome

Emily Crossley,[1]* Larisa Fedorova[2], Oleh A Mulyar[2], Shuhao Qiu[3], Ryan Freeman[2], Alexei Fedorov[1,2,3]

[1]Program of Bioinformatics and Proteomics/Genomics, College of Medicine and Life Sciences, The University of Toledo, Toledo, OH 43614
[2]CRI Genetics LLC, Santa Monica, CA 90404, USA.
[3]Department of Medicine, The University of Toledo, Toledo, OH 43614

*Corresponding author: ecrossl2@rockets.utoledo.edu

Keywords: DNA Structure, UCNE, UCE, Inhomogeneity

Published: 22 November 2024

Introduction: Ultra-conserved non-coding elements (UCNEs) remain fascinating genomic fragments that have maintained almost perfect sequence identity for millions of years. UCNEs are defined as regions of DNA that are longer than two hundred nucleotides in length and at least 95% conserved between humans and chickens. Previous research found that UCNEs have an excess of GpC dinucleotides but a decrease in GpG/CpC dinucleotides compared to the human genome. However, similarities between UCNES or other characteristics that make them unalterable have yet to be unveiled.

Objectives: Based on these findings, we hypothesize that UCNEs have a distinct dinucleotide composition that may contribute to a unique DNA structure. We calculate the distance between all dinucleotide pairs within UCNEs and the human genome to identify patterns in dinucleotide arrangements.

Methods: We purified a publicly available UCNE database, which includes a total of 4,272 sequences, and human genome sequences with masked repetitive elements. Randomly generated UCNE and human genome sequences were created using the dinucleotide frequencies from the real perspective sequences. Statistical analysis was performed to assess the non-randomness in dinucleotide spacing arrangements using relative percentage difference (RPD).

Results: Remarkable non-randomness in dinucleotide spacing arrangements was observed within the entire human genome and UCNEs. Approximately 83% of all dinucleotide pairs within UCNEs showed significant (>10% RPD) non-random genomic arrangements when compared to the rest of the human genome. Most non-random arrangement of dinucleotide pairs occurred at short distances, 2-6 nucleotides. Non-randomness in dinucleotide spacing distances deviated up to 40% from the expected values and were frequently associated with GpC, CpG, ApT, GpG, and CpC dinucleotides.

Conclusion: The described peculiarities in dinucleotide arrangements have persisted for hundreds of millions of years within vertebrates. These distinctive patterns may suggest that UCNEs form a unique DNA structure with distinct properties that contribute to their extraordinary conservation.

