

Computational techniques for protein identification through cyclic codes

Mario Enrique Duarte-González

Leap Metrics

In recent work, error-correcting codes have been applied to model the reliability of genetic information transmission, where DNA, RNA, and proteins are treated as codewords of cyclic codes. Progress in this area could yield insights into evolution and connections between species. The article *Identification of proteins by the use of Chinese remainder theorem codes over finite commutative rings* proposes a procedure for constructing cyclic codes over finite commutative rings with identity, which is based on the decomposition of a commutative ring as a direct sum of local rings and on the juxtaposition of repetition codes with not necessarily equal lengths. In *Construction of Cyclic Codes over $\mathbb{Z}_{20}$ for Identifying Proteins*, three algorithms were designed to identify odd-length polypeptide sequences as codewords of cyclic codes. Here, we review computational algorithms and infrastructure to identify cyclic codes over $\mathbb{F}_4$ and $\mathbb{Z}_4$ for mRNA sequences, as well as codes over $\mathbb{Z}_{20} \simeq \mathbb{Z}_4 \oplus \mathbb{Z}_5$ and $\mathbb{F}_4 \oplus \mathbb{Z}_5$ for proteins. We also discuss the biological and mathematical implications of some specific findings in this topic.

In Special Session 47 on Mathematical Tools with Applications in Quantum and Genetic Coding at the Mathematical Congress of the Americas 2025.