



第3回 2013年10月25日（金）14:50～16:30

Representation of DNA sequences using Gray code

グレイコードによる DNA 配列の表現

Natsuhiko Ichinose / 市瀬 夏洋

Graduate School of Informatics

Kyoto University

京都大学大学院 情報学研究科

Sorted data structures of DNA sequences are important for the search or the feature detection. Although the sorting procedure is usually performed based on the lexicographical order, it is not always adequate for analyzing DNA sequences. The Gray code is another ordering in which adjacent sequences differ from each other by one nucleotide. Namely, a neighboring region in the code always has a set of similar sequences. Since the similar sequences tend to have a similar function, the Gray code may be more informative than the lexicographical order. In the lecture, I introduce the application of the Gray code to motif discovery. The Gray code is not limited in the 1-dimensional space. I demonstrate the examples of the Gray code in the 2-dimensional space such as surfaces of a tetrahedron or octahedron.

Keywords: Gray code, motif discovery, comparative genomics

DNA配列をソートしたデータ構造は、配列の検索や特徴抽出において重要なデータ構造である。このようなデータ構造に適用される辞書順のソートは、DNA配列解析に対して必ずしも適切な構造ではない。グレイコードとは辞書順とは異なる順序構造であり、隣り合う配列が一塩基のみ異なるような並べ方である。すなわち、コード内で近傍にある配列は、常に相同な配列となる。DNA配列では、相同な配列は同様な機能を担う傾向にあるため、グレイコードは辞書順に比べより有益なコードとなろう。講義では、グレイコードのモチーフ発見問題への応用を紹介する。グレイコードは1次元の並びに限定されるものではない。2次元空間に構成した例として四面体や八面体の表面上のグレイコードを説明する。
キーワード: グレイコード、モチーフ抽出、比較ゲノム