

フラクタルコーディングによるゲノムデータ解析

●郷原 一寿

北海道大学大学院工学研究科

〈研究の目的と進め方〉

非線形ダイナミクスの手法により、DNA配列の新しい解析法を開発し、未知の文法構造を解明することを目的とする。ATGCの4塩基のひとつひとつに、ある写像を割り当て、これを適当な順番で切り替えると、フラクタル集合を形成する。この原理を、シーケンスが実験的に明らかとされているDNA配列に対して適用すると、一つのDNA配列をフラクタル集合に置き換えることが可能となる。これによって、配列の文法構造をフラクタル集合上のダイナミクスとして解析することができる。この手法は数学的には1980年代半ばに、DNAへの適用は1990年に既に報告され、いくつかの論文も提出されているがまだまだ研究の余地がある。特に、従来の研究では、4塩基を正方形の4分割領域に割り当てる2次元線形写像が用いられてきたが、対応させる写像には任意性がある。この点に着目し、最も単純な1次元線形写像を用いた数理的な解析を行い、実データへ適用する。

〈研究開始時の研究計画〉

- 1) 1次元写像を用いたDNA配列の新しい解析法を提案(図1)。
- 2) 21番染色体を具体的な例として、部分領域と階層構造の段数を変数として網羅的に解析を行う。
- 3) 物理地図、遺伝子地図上に対応する特徴配列をマッピングする。
- 4) 新しい解析法による、遺伝子コード領域の同定法について検討。
- 5) 未知の配列文法の解明を行う。
- 6) 出来上がったアルゴリズムをパッケージ化して公開する。

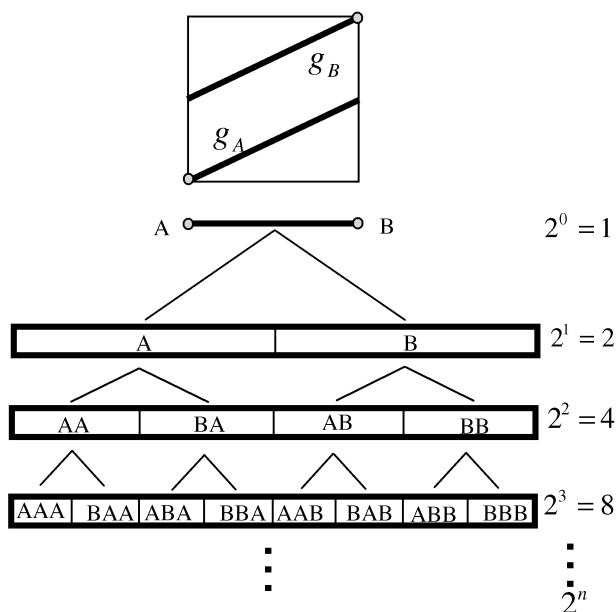


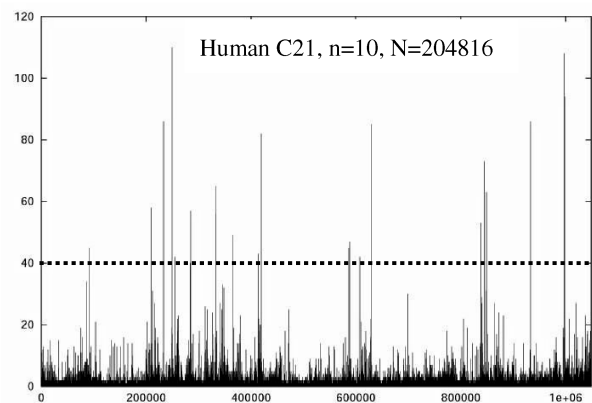
図1. 1次元線形写像への2記号列A, Bの埋め込み

〈研究期間の成果〉

- 1) 1次元線形写像の切り替えによるDNA配列データのフラクタル集合上へのマッピング方を考案。
- 2) 21番染色体の領域21q22.1(B2291C14-T1533) (総数: 798282) に対して、部分領域と階層構造の段数を変数として解析を行った。その過程で、計算量、メモリ容量について効率的な計算法を考案。
- 3) 21番染色体の領域(同上)の物理地図、遺伝子地図上に対応する特徴配列をマッピングするグラフィカルな方法を考案した。この方法の結果の一つとして、特定のアドレスが特異的に数が多いことが明らかとなった。この領域はテロメア内部にあり、多くの遺伝子が集中している特徴的な部分である。
- 4) 新しい解析法による、遺伝子コード領域の同定法について検討。
- 5) 未知の配列文法の解析方法について考察した。
- 6) 出来上がったアルゴリズムをパッケージ化するためのプログラム上のサブルーチン化を検討した。

〈国内外での成果の位置づけ〉

1次元実数直線上にフラクタルコーディングによって塩基配列をマッピングしたことは初めてであり、応用上重要な1次元配列に固有の特徴があることが明らかとなってきた。従来一般的に2次元のフラクタル集合を図形として解析する方法と比較して、より定量的な取り扱いを可能とすることが分かってきた。



〈達成できなかったこと、予想外の困難、その理由〉

計算機パワーの不足のために、大容量データの取り扱いができなかった。このため原理的な考察は深めることができたが、応用に供するための検討が時間的な制約により不足した。結果として、当初目的とした、だれでも使いやすいアルゴリズムとしてパッケージ化することができなかった。

〈今後の課題〉

大量な実データに対して適用できるよう計算機パワーを充実させ、大規模データに対して本手法を適用することが当面の課題である。この課題を解決すれば、塩基配

列の新たな特徴を見出すことができると考える。また、だれでも使いやすいソフトウェアパッケージとして公開することも重要な課題であると考えている。