

カラー空間符号モアレマッチング法によるゲノム情報における規則性の抽出

●谷田 純¹⁾ ◆森川 正章²⁾

1) 大阪大学大学院情報科学研究科 2) 大阪大学大学院工学研究科 (現所属：北海道大学大学院地球環境科学研究科)

＜研究の目的と進め方＞

大容量かつ曖昧性を特徴とするゲノム情報に適した新しい情報解析技術の開発をめざして、光演算技術に基づくゲノム情報解析手法に関する研究を行った。光演算技術は、大容量情報に対する並列処理性、人間の視覚系に対する情報可視性など、他の技術には見られない特徴を有している。特に、谷田らが考案した空間符号モアレマッチング法は、ゲノム配列に潜む特徴的配列や異種配列間の類似性を配列情報の符号化とモアレ現象の組合せにより可視化する独創的な光演算技術である。そこで、本研究では、空間符号モアレマッチング法を基盤技術として、その特徴を効果的に活かすゲノム情報解析法に関する集中的な研究を行った。

まず、空間符号モアレマッチング法をもとに、さまざまな技術的な展開を検討し、ゲノム情報解析における適用領域の拡大に努めた。次に、提案手法を効率的に実行する専用ハードウェアの開発を進めた。特に、光演算技術の有効利用の観点から、その可能性を検討した。また、提案手法の有用性を確認するため、超好熱菌をはじめとするゲノム解析の現場での評価を行った。

なお、研究の進展に伴い、本研究にかかる課題名は、空間符号モアレ法に基づく光学的ゲノム情報解析技術の開発(2000年度)、空間符号モアレマッチング法による鳥瞰的ゲノム情報解析(2001年度)、空間符号モアレマッチング法によるマルチスケールゲノム情報解析(2002年度)、カラー空間符号モアレマッチング法によるゲノム情報における規則性の抽出(2003年度)と変更してきた。しかし、研究の主な流れについては、前述の研究方針に沿ったものである。

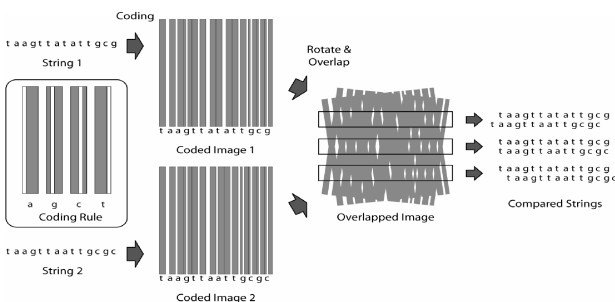


図1 空間符号モアレマッチング法の原理

＜研究開始時の研究計画＞

空間符号モアレマッチング法は、空間符号化技術に基づく光演算技術の一つとして考案された。光演算技術は、光の有する物理的特性の有効利用により、大容量情報に対する並列処理を実現する情報処理技術である。その中で、空間符号化技術は、処理対象となる情報を何らかの符号パターン(時間信号との対比において、空間符号と呼ばれる)に変換した後、光学的なパターン操作技術を

利用して、もとの情報を効率的に処理する演算技術として知られている。

図1に空間符号モアレマッチング法の原理を示す。二つの塩基配列を入力とする。それぞれの塩基配列ごとに、塩基の種類(A, T, G, C)に対応した空間符号パターンを配置し、それらの組みあわせとして符号化画像を構成する。次に、得られた符号化画像をわずかに傾けて重ね合わせる。この操作により、図2に示すように、入力した塩基配列の並びにおいて、配列が一致した部分にモアレ縞が現れる。モアレ縞とは、周期がほぼ等しい格子状の信号を重ね合わせた際に生じる粗い縞信号である。等しい塩基配列が重なりあった領域にのみモアレ縞が発生することにより、二つの塩基配列間の一致情報を得ることができる。この出力結果は、ゲノム情報解析におけるホモロジー検索と等価なものであり、配列の一致を示すモアレ縞を検出することにより、二つの塩基配列間の類似性に関する情報を得ることができる。

空間符号モアレマッチング法は、簡便な手法ながら非常に特徴的な情報を出力するため、興味深い技術である。そこで、本研究課題では、以下の研究項目を設定した。

1) 空間符号モアレマッチング法の拡張

空間符号モアレマッチング法を基本技術として、そこから派生するさまざまな技術を網羅的に検討し、ゲノム情報解析における同手法の役割を明らかにする。

2) 空間符号モアレ情報端末の試作

光演算技術の適用により、空間符号モアレマッチング法を高速かつ効率的に実行する専用処理装置を試作する。試作を通して、ゲノム情報解析における光演算技術の有用性を明らかにする。

3) 空間符号モアレマッチング法の評価

超好熱菌など実際の生物ゲノム解析への適用により同手法の特性・問題点を明らかにする。

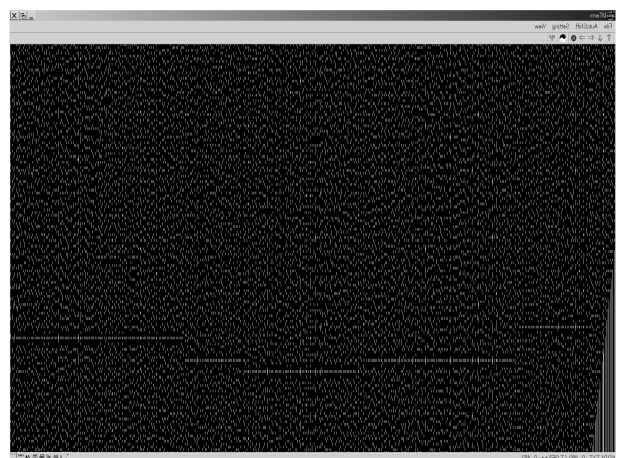


図2 空間符号モアレマッチング法の出力例

なお、本研究開始当初、研究代表者はゲノム情報分野との関わりが薄かったため、同分野の研究状況を把握し、先端技術に追従することも重要な課題であった。その過程において、光演算技術分野と接点を見出し、同分野の研究者に対してゲノム情報処理の重要性を発信することにも重点を置いた。

〈研究期間の成果〉

1) 空間符号モアレマッチング法の拡張

空間符号モアレマッチング法の潜在能力を引き出すため、さまざまな拡張技術を検討し、ゲノム情報可視化において多くの有用な成果を得た。

a) 対称型符号パターンによるステムループ構造の抽出
空間符号モアレマッチング法では、塩基と符号パターン間の対応関係は、互いに区別可能であれば、任意に決めることができる。そこで、相補関係をなす塩基の組 (AとT、GとC) が鏡映対称となる対称型符号パターンを導入することにより、転写RNAにおけるステムループ等の2次構造の予測機能を付加した。

b) 濃度反転符号による一塩基置換検出能の実現[1]-1]
従来、空間符号モアレマッチング法は、塩基の挿入・欠失を明確に検出できたが、一塩基置換などの突然変異に対する感度は低かった。この問題に対して、比較配列から生成される符号化画像の一方の濃度を反転する手法を考案した。これにより、図3に示すように、配列一致部位のモアレ縞の鮮鋭化が実現され、一塩基置換を含む配列比較能力の向上を達成した。

c) カラー符号による多重配列比較手法の開発[1]-2]
空間符号モアレマッチング法は、主に2配列間の情報比較技術として考案された。しかし、ゲノム配列のマルチプルアライメント等の要求に応えるためには、複数配列の同時比較が必要となる。そこで、符号パターンとして透過型のカラーフィルタを用いる技術を開発し、複数配列の同時比較技術を開発した。

d) 特性に基づく符号化によるアミノ酸配列比較 [1]-3]
空間符号モアレマッチング法の適用範囲を拡大するため、アミノ酸配列に対する同手法の適用を検討した。アミノ酸配列では20種類のアミノ酸が存在するが、類似アミノ酸との置換がしばしば見られる。そのため、基本手法の単純な適用では類似配列の同定がうまく行えなかった。そこで、pHや疎水性などアミノ酸の特性値に基づいた符号化の適用により、良好な検出性能が得られることを確認した。

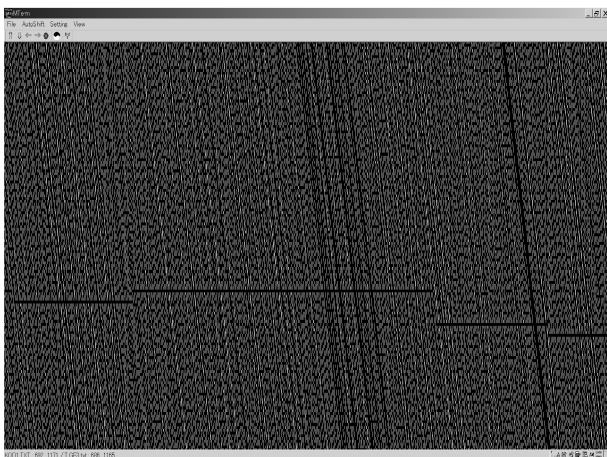


図3 濃度反転符号による縞の鮮鋭化

e) 出現頻度符号化法による大規模ゲノム比較の実現
空間符号モアレマッチング法では、同時に比較可能な配列サイズは、表示デバイスの解像点数によって制限される。しかし、超高精細の出力装置を用いても、高々1000塩基長に満たない配列しか符号化画像として得ることができない。このゲノム情報解析技術としての本質的な問題に対し、図4に示すように、一定長の塩基配列区間における各塩基の出現頻度に基づく出現頻度符号化法を考案した。さらに、モアレ縞の視認性を高めるため色情報を効果的に活用し、比較配列長の劇的な拡大を実現した。

f) 微細構造縞—ゲノム配列間の対応付け
空間符号モアレマッチング法では、比較配列が一致したときに生じるモアレ縞だけに着目していた。しかしながら、符号化画像同士のモアレ縞として得られる出力画像には非常に多くの情報が含まれている。そこで、画像を構成する微細な構造縞と元の配列との対応関係を整理し、モアレ縞画像からの有用な情報抽出の可能性を検討した。

2) 空間符号モアレ情報端末の試作

光演算技術としての空間符号モアレマッチング法の特徴を明らかにするために、提案手法を高速実行する空間符号モアレ情報端末の試作を行った。

a) 光学的空間符号モアレ情報端末の試作[1]-1]

透明フィルムと液晶ディスプレイと重ね合わせた光学的空間符号モアレ情報端末を試作した。一方の配列情報は透明フィルムに印刷して比較を行う。片側の配列情報が固定のため、利用範囲は限定されるが、透明フィルムへの印刷は高密度で行えるため、良好な視認性が得られることを確認した。

b) 空間符号モアレ情報解析支援プログラム: MTermの開発[2]-1]

光学的空間符号モアレ情報端末の制限を排除するため、符号化画像生成から重ね合わせまでをディスプレイ上で実行するプログラムMTermとして実装した。図2、3、5はすべてMTermによる出力結果である。ディスプレイの解像点数不足を懸念していたが、予想に反して比較の妨げとなるエリアジング雑音は発生せず、実用上問題ないことが確認された。出現頻度符号化法など新規に開発された技術を容易に組み込むことも可能であり、空間符号モアレマッチング法の現実的な実装形態として利用価値が高い。

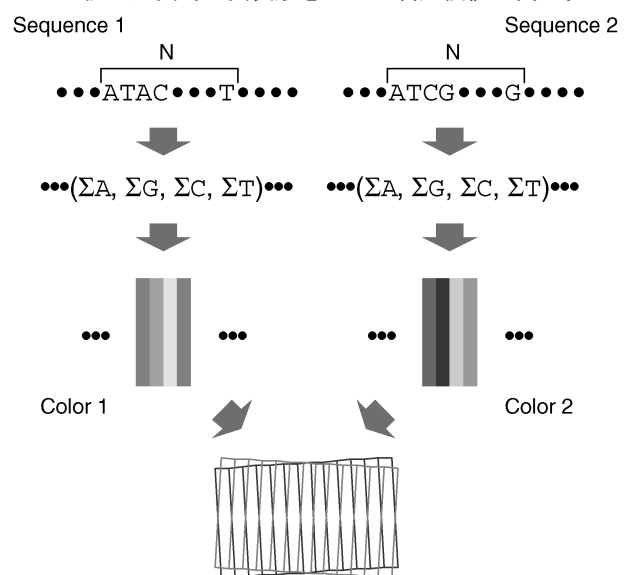


図4 出現頻度符号化法の原理

c) モアレ縞自動解析技術の開発

空間符号モアレ情報端末の利用を考えると、すべての情報比較を人間が行うことは現実的ではない。そこで、モアレ縞画像から特徴的なモアレ縞パターンを自動的に抽出する技術を開発した。フーリエ変換、モフォロジカルフィルタなどの適用により、配列一致部位が抽出できることを確認した。

3) 空間符号モアレマッチング法の評価

a) 超好熱菌類におけるゲノム比較

空間符号モアレマッチング法の有効性を評価するため、超好熱菌類におけるゲノム比較に適用した。16S rRNA遺伝子における類縁関係とモアレ縞との対応関係を明らかにし、両者に相関関係があることを確認した。

b) 大腸菌ゲノムの全ゲノム比較

出現頻度符号による大規模ゲノム比較技術の有効性を明らかにするため、大腸菌K-12株（460万塩基）と病原性大腸菌O157株（550万塩基）の全ゲノム比較を行った。ディスプレイ上で両者の一致関係が視覚化され、提案手法の有効性を確認することができた。

〈国内外での成果の位置づけ〉

モアレ現象を利用したゲノム情報可視化は国内外を見渡しても独自の技術であり、光情報処理、光コンピューティング、情報フォトニクス分野において、有望な先導的研究として評価を受けた。本研究に啓発され、光情報処理技術をゲノム情報解析に応用する試みが国内外の複数の研究グループにより報告された。これらの流れに対応すべく、日本光学会（応用物理学会）情報フォトニクス研究グループにおいて、フォトニックバイオインフォマティクスに関するワーキンググループが設立され、具体的な適用事例の紹介、ゲノム情報処理に関する情報交換など、光情報処理分野における活動が立ち上がった。

光応用技術分野の動向として、医用分野との連携が活発に進められている。現時点では、医用計測、分子生物学の実験支援などの領域に限られているが、光応用技術の適用分野の拡大に伴い、ゲノム情報解析の実施例としての有意義なものと考えられる。

〈達成できなかったこと、予想外の困難、その理由〉

1) 空間符号モアレマッチング法の拡張

方法自体の拡張については、有効な手法の開発を含め、十分な成果が得られたと考えられる。しかし、実際のゲノム情報解析の現場で使用するためには、人間を含めたインタラクティブなシステムを構築する必要がある。本研究では、道具としての空間符号モアレマッチング法の開発に研究の主眼が置かれたが、処理システム内に人間を組み込む場合には、人間の特性を含めたより高度なマン・マシンインターフェースの枠組みが必要となる。時間的制約、研究スコープの相違などにより、これらについては解決されなかった。

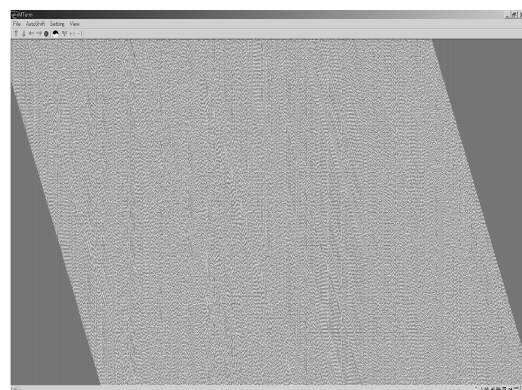
2) 空間符号モアレ情報端末の試作

研究開始当初、提案手法を高速に処理可能な専用ハードウェアの開発をめざしていたが、適切な情報表示デバイスの入手が困難であったため、原理確認のための装置試作に留まった。また、班会議などでの意見交換により、専用ハードウェアの開発が必ずしも利用者の要求を満たすものではないことが明らかになった。実際、空間符号マッチング法のさまざまな拡張により、単一の専用ハードウェアを試作することの意義が薄れている。これらの理由により、研

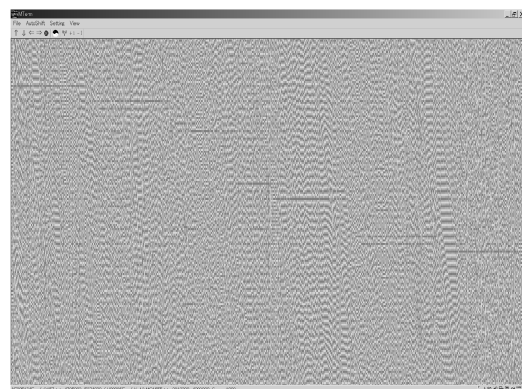
究期間の後半では、ソフトウェア実装に主体を移したため、研究開始当初に計画していた光情報技術の応用によるゲノム情報解析技術は実現できなかった。

3) 空間符号モアレマッチング法の評価

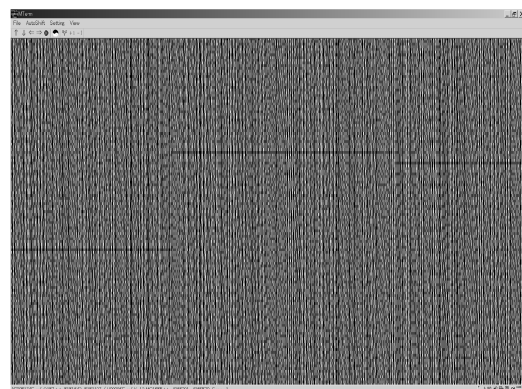
MTerm利用によるゲノム情報解析への適用において、有効性が確認されたものは超好熱菌類、大腸菌などの一部のゲノムに留まる。多様な種に対する適用も試みながら、無意味な情報雑音しか得られない場合が多く、それらから有用な情報を抽出するには至らなかった。画像処理によるモアレ縞自動解析技術は、単純な場合に限られ、処理の高速化も不十分であった。モアレ縞の視認性評価においては、本来、被験者実験を実施すべきであるが、大掛かりな研究計画が必要となり、ゲノム情報解析の主題から離れてしまうため、未実施とした。



符号化単位 11500 塩基
比較配列区間 5.5M 塩基



符号化単位 1000 塩基
比較配列区間 0.5M 塩基



符号化単位 1 塩基
比較配列区間 480 塩基

図5 大腸菌K-12株と病原性大腸菌O157株の全ゲノム比較結果

〈今後の課題〉

1) 空間符号モアレマッチング法の拡張

本研究において開発された各種手法の組み合わせにより、ゲノム解析技術としてのプロトコルを確立する必要がある。現状では、個別技術の集まりに過ぎず、ゲノム解析の現場における利用は容易ではない。また、日進月歩のゲノム情報分野では、新しい知見が得られたときに、即座に対応可能な柔軟性が重要である。技術ライブラリとして整備することにより、得られた手法の利用性を高めることも必要である。

2) 空間符号モアレ情報端末の試作

コンピュータアーキテクチャのトレンドとして、GPU（グラフィック専用プロセッサ）の高速処理能力を活かした汎用処理が注目されている。MTermでは、描画のみならず、各種のデータ処理が必要であるが、それらのGPUによる実装は劇的な性能向上をもたらすものと期待される。また、モアレ縞自動解析処理についても、GPUの利用は有効な手段を提供すると考えられる。

光学的情報処理技術の適用による専用ハードウェアの開発は、現時点では不必要と判断された。しかし、光エレクトロニクスの進展に伴い、光・電子融合デバイスの性能は飛躍的に向上しており、それらの特性を活かした新しい情報システムの開発は近い将来の課題と考えられる。本研究成果は、この問題に対する一つの糸口を与えるものと期待される。

3) 空間符号モアレマッチング法の評価

配列決定技術の進展により、多くの生物種のゲノム情報が明らかにされている。これらの成果に基づいた学問領域として、比較ゲノム学が注目されている。同分野においては、全ゲノムレベルでの配列比較が有力な解析手段となる。空間符号モアレマッチング法はその要求に合致した技術であり、今後の応用に向けた検討が必要である。

分子生物学の進展に伴い、ゲノムに関する新しい知見が次々に得られている。非コード領域を含むゲノム情報解析はますます進展すると考えられるが、全ゲノムレベルでの鳥瞰的な観察が有用な情報を提供する可能性は高い。空間符号モアレマッチング法の適用対象として検討すべきものとしてあげられる。

〈研究期間の全成果公表リスト〉

1) 論文／プロシーディング（査読付きのものに限る）

1. 112261739

Nitta, K., Togo, H., and Tanida, J: Matching information terminal based on spatially coded moire technique, *Optical Engineering*, 40, 2386-2391 (2001).

2. 303241640

Tanida, J., Nitta, K., and Yahata, A: Spatially coded moire matching technique for genome information visualization, *Proceedings of SPIE*, 4929, 26-33 (2002).

3. 303241646

Nitta, K., Yahata, A., and Tanida, J.: Information Extraction from Amino Acid Sequence Using Spatially Coded Moire Matching Technique, *Proceedings of Optics in Computing 2002*, 65-67 (2002).

2) データベース／ソフトウェア

1. 空間符号モアレ情報解析支援プログラム: MTerm.

3) 特許など

該当なし

4) その他顕著なもの

該当なし