

## 微生物群集のダイナミクスの数理モデリングと数値解析\*

小谷野 仁<sup>†</sup>      澤田 和典<sup>‡</sup>      山本 希<sup>§</sup>      山田 拓司<sup>¶</sup>  
 東京工業大学   株式会社ぐるなび   東京工業大学   東京工業大学

## 1. 配列の集団の時間発展の数理モデル

1つの環境中の生物群集は、群集中の他の個体や周囲の環境と相互作用しながら時間発展していく。我々はこのダイナミクスに興味がある。アルファベット  $A = \{a, c, g, t\}$  上の文字列がなす非可換位相半群を  $A^*$  によって表す。そうすると、DNA や遺伝子の配列は  $A^*$  の元であり、生物群集が持つ DNA やある遺伝子の全体 (母集団) の分布は、 $A^*$  上の確率関数として表される。そこで、生物群集をそれが持つ DNA の集団や 16S リボソーム RNA 遺伝子の集団として捉えることにより、生物群集の時間発展の数理モデリングにアプローチする。進化学では、太古の地球の海底熱水孔域に、現存の生命の共通祖先 (の DNA) が誕生したと考えられている。その後、地球上の悠久の生命史の中で生命は多様化し、現在の地球上には多様な生命、或いは同じことであるが、多様な DNA が存在する。この過程は、地球上の DNA の集団が、 $A^*$  の 1 点である共通祖先の DNA から、配列中に確率的に起こる突然変異と環境から掛かる淘汰圧の下で、世代交代の度に  $A^*$  の中で広がっていったことと見なせる。そこで、[5] では、拡散方程式をモチーフにして、 $A^*$  上で、1つの環境中の生物群集が持つ DNA の集団が、確率的な突然変異と環境からの淘汰圧の下で時間発展していく様子、すなわち DNA の集団の進化を記述する

$$\frac{\partial q(s, t)}{\partial t} = -c(s, t) + b(s, t)(1 - \pi)^{\ell(s)} + \sum_{1 \leq d < \infty} \sum_{s' \in V(s, d)} b(s', t) \frac{\ell(s') C_d \pi^d (1 - \pi)^{\ell(s') - d}}{|V(s', d)|}$$

という形の偏微分方程式を導出した (記号の詳細は略)。また、モデルの数理解析を行って、集団が分化して新しい種が作られるための条件や、集団が平衡

状態を維持し、長期間に渡って変化しないであるための条件を示した。

## 2. 配列の集団への淘汰圧の統計的推定

進化学における淘汰圧は、力学における力のように観測できるものではないが、それを推定したいという欲求は、進化学や生態学においては自然なものとしてある。これまで、数値として表される形質に対しては、その測定値に基づいて環境から集団中の個体に掛かる淘汰圧を推定する方法が色々と考えられ、実際に淘汰圧を推定した事例研究も多くなされてきたが、微生物学においてこれらの方法を用いることはできない。DNA やある遺伝子の配列に対して淘汰圧を推定することが、全ての生物に適用できる統一的ですっきりした接近法であるが、このような方法はこれまで考えられてこなかった。また、 $A^*$  上には、これまで、実数の集合  $\mathbb{R}$  上の正規分布や自然数の集合  $\mathbb{N}$  上の 2 項分布のようなパラメトリックな分布は導入されていなかったが、[1] では、その混合モデルを用いて、ある環境中の配列の集団にその環境から掛かる淘汰圧の分布を表すように設計された

$$q_d(s; \lambda, \rho) = \frac{1}{(\rho + 1) |\partial U_d(\lambda, d(s, \lambda))|} \left( \frac{\rho}{\rho + 1} \right)^{d(s, \lambda)}$$

という形の確率関数を持つ分布を  $A^*$  上に導入し (記号の詳細は略)、その基本的な性質を調べることから始めて、その混合モデルのパラメータを推定する方法を開発し、[2, 3] において証明された結果を応用して、統計的漸近理論の枠組みでその方法に対して数理的基礎付けを与えた後、数値実験を行ってその有効性を確かめた。

## 3. 環境微生物のダイナミクスへの応用

本発表において、我々は、まず、上記の 2 つの研究の結果を簡単に述べた後、それらにおいて開発した理論モデルと統計的方法を組み合わせ、*Brassica rapa* subvar. *hiroshimana* Kitam. という植物の周辺環境を人工的に高塩環境に改変した前後の 2 時

\*Mathematical modeling and numerical analysis of the dynamics of microbial communities

<sup>†</sup>Hitoshi Koyano, Tokyo Institute of Technology

<sup>‡</sup>Kazunori Sawada, Gurunavi, Inc.

<sup>§</sup>Nozomi Yamamoto, Tokyo Institute of Technology

<sup>¶</sup>Takuji Yamamoto, Tokyo Institute of Technology

点においてその環境中の微生物群集から収集された 16S リボソーム RNA 遺伝子の環境標本を用いて、その集団の時間発展の数値実験を行った結果を示す。

#### 4. 配列の集団の動態解析の統計的方法

粒子の運動や細胞の遊走の研究では、対象は 3 次元 Euclid 空間  $\mathbb{R}^3$  の点として扱われる。つまり、用いられているのは、時間と共に空間  $\mathbb{R}^3$  の点がどのように移動していくのかを調べる、という枠組みである。我々は、環境を改変することにより、微生物群集がどのように時間変化していくのか、変化の特徴に興味がある。今、微生物群集をそれが持つ 16S リボソーム RNA 遺伝子の集団と捉えている。特に、1 つの環境中には同一の 16S リボソーム RNA 遺伝子が多数存在するので、微生物群集を 16S リボソーム RNA 遺伝子の配列の集まりと各配列の相対頻度として表現するという立場を取っている。これは、微生物群集を、アルファベット  $A = \{a, c, g, t\}$  上の文字列の位相半群  $A^*$  上の確率分布 ( $A^*$  は離散的なので、確率関数) として捉えていることに他ならない。つまり、第 1 節において述べたように、“1 つの微生物群集 =  $A^*$  上の 1 つの確率関数”である。(実際には粒子や細胞が移動しない点も含めて) 空間のありとあらゆる点の集まりが  $\mathbb{R}^3$  であった。同様に考えると、(架空のものまで含む) ありとあらゆる微生物群集の集まりは、 $A^*$  上の確率関数の集合 ( $\mathcal{P}$  と書く) ということになる。そこで、微生物群集の動態を、時間が経過するに従って、空間  $\mathcal{P}$  の点がどのように動いていくのか、動き方にどのような特徴があるのかを調べる、という枠組みで研究することを考える。 $\mathbb{R}^3$  は、代数の立場から見ると内積空間であり、解析の立場から見ると距離空間であって、移動距離 (1. 総移動距離, 2. 変位), 速さ (3. 最大値, 4. 最小値, 5. 平均), 速さのばらつき (6. 範囲, 7. 標準偏差, 8. 変動係数), 移動方向 (9.  $(x, y)$  平面上の平均方向, 10.  $z$  軸方向の平均方向), 及び直進度 (11. 方向持続性, 12. 平均合成成長) などの運動解析の基本的な物理量は、これらの空間の構造 (加法, スカラー乗法, 内積, ノルム, 及び距離) を用いて定義されている。集合  $\mathcal{P}$  は、[4] によって  $\mathcal{P}$  上に導入された  $\beta$  多様性という距離によって距離空間をなすが、加法, スカラー乗法, 及び内積は定義されないので、少なくとも自明な仕方で  $\mathcal{P}$  上に上記の 9, 10, 及び 12 に対応する量を定義することはできない。しかし、距離のみを用いて定義される他の量の対応物は、 $\mathcal{P}$  上に導入することができる。但し、1 つの環境中の生物群集が持つ全ての 16S リボ

ソーム RNA 遺伝子を収集することは絶対に不可能であるから、空間における粒子や細胞の位置に基づいて計算される上記の 12 個の物理量と異なり、導入された量は、直接観測できる、或いは観測されたものから直接計算できるものではなく、収集された 16S リボソーム RNA 遺伝子の環境標本から推定するしかない。本発表では、我々は、運動解析に倣って、動態解析の基本的な特徴量を  $\mathcal{P}$  上に導入し、これらの推定量を構成して、それらの統計的性質を調べ、その後、開発した統計的方法を、第 3 節でその動態をコンピューターの中で求めた、*Brassica rapa* subvar. *hiroshimana* Kitam. の周辺環境中の微生物群集に応用した結果を述べる。

#### 引用文献

- [1] H. Koyano, M. Hayashida, and T. Akutsu. Optimal string clustering based on a Laplace-like mixture and EM algorithm on a set of strings. arXiv:1411.6471 [math.ST].
- [2] H. Koyano, M. Hayashida, and T. Akutsu. Maximum margin classifier working in a set of strings. *Proceedings of the Royal Society A*, 2016.
- [3] H. Koyano and H. Kishino. Quantifying biodiversity and asymptotics for a sequence of random strings. *Physical Review E*, 81(6):061912, 2010.
- [4] H. Koyano, T. Tsubouchi, H. Kishino, and T. Akutsu. Archaeal  $\beta$  diversity patterns under the seafloor along geochemical gradients. *Journal of Geophysical Research: Biogeosciences*, 119(9):1770–1788, 2014.
- [5] H. Koyano and K. Yano. Evolutionary model of a population of DNA sequences through the interaction with an environment and its application to speciation analysis. arXiv:1706.01182 [q-bio.PE].