

自己組織化マップを用いた生物種の系統解析手法の提案

西牟田 健太郎^{†1} 吉原 郁夫^{†2}
山森 一人^{†2} 安永 守利^{†3}

生物は遺伝情報を DNA 塩基配列に保持している。塩基配列の類似性等を解析することは生物種の進化の歴史を解明するために非常に重要な問題である。解析手法として塩基配列同士のパターンマッチングが主に用いられているが、生物種の形質等の情報が必要であるなどの制限がある。そこで、本論文では塩基配列のみを使用した生物種の進化系統解析手法を提案する。コドンの出現頻度を入力とした自己組織化マップによりマップを作成し、マップ上に現れた領域間の距離と生物種間の距離の対応しているかによって進化系統を推測する。

Analysis of Evolutionary Lineage by Self-organizing map

KENTARO NISHIMUTA,^{†1} IKUO YOSHIHARA,^{†2}
KUNIHIITO YAMAMORI^{†2} and MORITOSHI YASUNAGA^{†3}

DNA base sequences are considered to involve foot prints of evolution of living creatures. Most analysis methods of the base sequences employ pattern matching rules or knowledge, which require a lot of knowledge and data in advance. This paper proposes an evolutionary lineage method of species by using a novel Self-organizing map (SOM) whose inputs are frequency of appearance of codons. The neighboring relation of the SOM is believed to correspond to similarity between species.

^{†1} 宮崎大学大学院
Graduate School of Engineering University of Miyazaki

^{†2} 宮崎大学
University of Miyazaki

^{†3} 筑波大学
University of Tsukuba

1. はじめに

生物の持つ情報は親から子へと受け継がれ、その遺伝情報は細胞内の DNA 塩基配列に保持されている¹⁾。この DNA 塩基配列は、長い間受け継がれていくうちに突然変異などによって変化し、その積み重ねが今日の生物の多様化を生んだ²⁾。多種多様な生物種の変化の流れが進化と考えられる。DNA 塩基配列の中に進化の痕跡が残されていると考えられており、生物の塩基配列に内在する機能的な、あるいは進化的な類似性を解析するのは生物進化の足取りを解明するために非常に重要である³⁾。

塩基配列とは、DNA 中の塩基の並びのことで、塩基にはアミン (A)、シトシン (C)、グアニン (G)、チミン (T) の 4 種類があり、塩基配列は 4 塩基の並びによって構成されている。塩基配列の連続する 3 塩基の並びをコドンと呼び、コドンは $4^3 = 64$ 種類あり、これらは、20 種のアミノ酸に対応している。アミノ酸が順に結合しタンパク質がつけられる。塩基配列のタンパク質の設計図になる領域をエクソン、取り除かれる領域をイントロンという²⁾。

生物種の遺伝子中の塩基配列の中にある特徴的な塩基の並び (パターン) があり、現在様々なパターンの解析が行われている。生物の進化系統解析は、その特徴パターンを用いて塩基配列にパターンマッチングを適用する方法が主に用いられる¹⁾⁴⁾⁵⁾。しかし、特徴パターンは探索途中であり、未知の遺伝子が発見された際にその遺伝子に対しどの生物のどの特徴パターンとマッチングをとればよいかわからないなどの恐れがある。この問題を回避するために我々は、パターンマッチングを用いない解析方法を試みている⁵⁾。

パターンマッチングを使用しない系統解析手法として塩基配列から特徴量を計算しそれを生物種間の距離として用いる方法を試み、特徴量を計算することで、生物の持つ情報 (形質) などをあらかじめ用意しなくても生物種の系統を探ることができるのではないかと考えている。従来研究としては、データ圧縮率を用いた手法⁶⁾、統計力学による長距離秩序を用いた手法⁷⁾⁸⁾、 $1/f$ ゆらぎを用いた手法⁹⁾ データの分散を用いた手法¹⁰⁾、情報エントロピーを用いた方法¹¹⁾¹²⁾¹³⁾¹⁴⁾¹⁵⁾、カオス理論を用いた手法¹⁶⁾ がある。

従来方法は、生物種の塩基配列から特徴量を計算し、それを生物種間の距離とする考えに基づいている。塩基配列から特徴量を前述の方法によって計算するが、1 つの生物種に対し 1 つの特徴量しか算出しない。数千 ~ 数万塩基の配列から 1 つの特徴量では足りないのではないかと考えた。そこで塩基配列から標本を切り出し、標本ごとに特徴量を求め、複数の特徴量から生物種間の距離を測る。本研究では、塩基配列から切り出された標本から特徴量と

2 自己組織化マップを用いた生物種の系統解析手法の提案

して各コドンの出現頻度を計算する．多量のデータを扱うことのできる自己組織化マップへ入力し，可視化する．

自己組織化マップ (SOM: Self-Organizing Map) は Kohonen によって提案されたニューラルネットワークの 1 種である¹⁷⁾．入力データ間の関係を教師なし学習し，データの類似度を反映したマップを作成することにより，特徴をより可視化する．分類の施されていない未知のデータを扱うことが可能であり，データの量や，構造といった性質に依存しない．自己組織化マップによって可視化されたマップに現れる領域の位置はそれぞれの類似度に関係し，隣接する領域は遠く離れた領域よりも近い性質を持っていると推測できる．つまり，未知の遺伝子を入力した場合，様々な生物種の遺伝子と入力することによって，未知の遺伝子と他の生物種の遺伝子と領域が隣接関係にあれば，隣接関係にある生物種の特徴パターンとマッチングができるなど，解析を行う前の手がかりになるとも考えられる．また，作成されたマップにおける各領域の相対関係を解析するために，各領域を構成するノードから特徴ノードを決定し，特徴ノード間の距離を算出する．さらに，この特徴ノード間の距離を生物種間の距離と考え，距離行列法によって系統樹を作成する．

本研究では，コドンの出現頻度を入力とした自己組織化マップにより可視化し，マップ上の領域の隣接関係によって生物種の系統解析を行い，作成されたマップの領域の相対関係から系統樹を作成する方法を提案する．実際にこの手法を 20 種の生物種のリボソームタンパク質遺伝子を使用し，解析を行った．

自己組織化マップによる可視化では，生物種ごとに領域が作られ，領域間の隣接関係は生物種間の距離と対応していた．系統樹を作成した結果は，大まかにみると動物類，菌類，植物類に分かれ，細かく見るとネズミ科に属するハツカネズミとドブネズミが結ばれ，それらと同じ哺乳綱に属するヒトが結ばれており，系統が反映された系統樹が作成できた．

2. 自己組織化マップによる進化系統解析手法

生物種の塩基配列から標本を切り出し，標本ごとに特徴量としてコドンの出現頻度を計算する．計算した出現頻度を 64 次元のベクトルとして自己組織化マップに入力する．自己組織化マップは入力データの傾向を学習させ，データの類似度を反映した領域を形成する．そのため，マップに形成された領域間の距離と生物種間の距離が対応していると考えた．

2.1 自己組織化マップ

本研究で用いる自己組織化マップ (SOM: Self-organizing Map) は，ニューラルネットワークの 1 種で，Kohonen により記憶やその想起・連想のメカニズムを計算機上で実現するた

めに開発された¹⁷⁾．

自己組織化マップは，高次元空間のデータを 2 次元空間に非線形写像するアルゴリズムである．そのため，高次元データの分布を 2 次元平面上に視覚化する有効なモデルである．

2.2 自己組織化マップの構造

自己組織化マップは図 1 のように，入力層と競合層の 2 層から構成されている．入力層は， n 次元の入力ベクトルと同じ数のノードがある．競合層のノードは，出力を視覚的に見るため通常 2 次元に配列される．競合層では，ノードがそれぞれの近傍の情報をもとに学習を行う．自己組織化マップの学習は競合層の構造に影響を受けることがある．たとえば，四方に境界を持つ長方形型の構造の場合，境界際の学習の近傍領域が減少してしまう．そこでノードの構造は球体や上下左右がつながったトラス構造をとるものがあり，本研究はトラス構造の競合層を用いる．

また，競合層のノードは重みベクトルを持っており，入力層のノードと完全結合している．

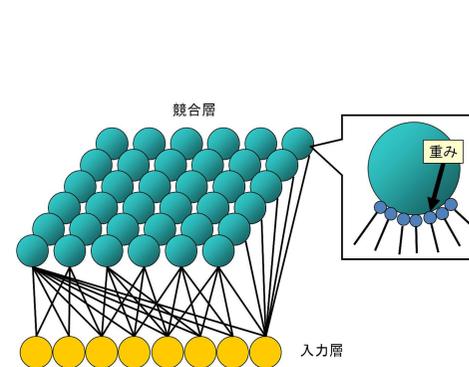


図 1 自己組織化マップの構造
Fig. 1 Structure of SOM

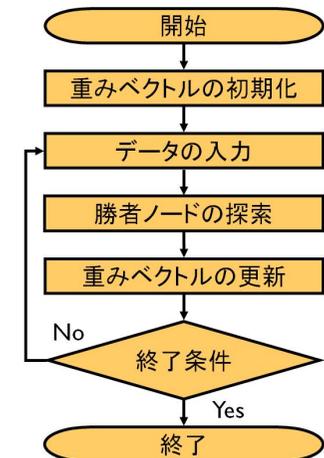


図 2 自己組織化マップの処理の流れ
Fig. 2 Algorithm of som

2.3 学習アルゴリズム

自己組織化マップの学習アルゴリズムは，主に，重みベクトルの初期化，入力ベクトルの入力，ユークリッド距離の計算，勝者ノード（入力ベクトルとのユークリッド距離が最小な

3 自己組織化マップを用いた生物種の系統解析手法の提案

重みベクトルを持つ競合層のノード)の探索, 重みベクトルの更新の5つのステップからなる. 自己組織化マップの学習の流れを図2に示す.

まず, n 個の入力ベクトルを $\mathbf{x}^{(n)} = (x_1^{(n)}, x_2^{(n)}, \dots, x_k^{(n)})$ とし, 競合層のノードが持つ重みベクトルを $\mathbf{w}_{ij} = (w_{ij1}, w_{ij2}, \dots, w_{ijk})$ とする. このとき, 競合層のノードの位置を $i = (0, 1, \dots, I - 1)$ と $j = (0, 1, \dots, J - 1)$ で表す.

Step 1. 各重みベクトルをランダムに初期化する

Step 2. 入力ベクトルを入力する.

Step 3. 入力ベクトルと各ノードの重みベクトルとのユークリッド空間上の距離を計算する.

ユークリッド距離 E は以下の式により定義される.

$$E = \sqrt{\sum_{k=0}^N (x_k^{(n)} - w_k)^2} \quad (1)$$

Step 4. 勝者ノード (ユークリッド距離が最小なノード) を探す.

Step 5. 勝者ノードを中心とした更新範囲に含まれる重みベクトルを入力ベクトルとの差が小さくなるように更新する. 重みベクトルの更新は以下の式により行う.

$$\mathbf{w}_{ij}^{(new)} = \mathbf{w}_{ij}^{(old)} + \alpha(t)(\mathbf{x}^{(n)} - \mathbf{w}_{ij}^{(old)}) \quad (2)$$

$t (= 0, 1, \dots, T)$ は, 学習回数を表しており, $\alpha(t)$ は, 学習係数を表している. また, 更新範囲は, 勝者ノードを中心とした一辺 $\beta(t)$ の正方領域である. $\alpha(t)$ と $\beta(t)$ を次の式で求める.

$$\alpha(t) = \alpha_{init}(1 - t/T) \quad (3)$$

$$\beta(t) = \max\{\epsilon, \beta_{init} - t\} \quad (4)$$

ここで, α_{init} は, 学習係数の初期値で, $(0 < \alpha(t) < 1)$ の範囲の値をとる. β_{init} は, 更新範囲の初期値とし, パラメータとして与える.

2.4 一括学習型自己組織化マップ

先ほど述べた, 従来の自己組織化マップは, データの入力順によっても出来上がるマップが変わってしまう. データの入力順が重要な要素である場合は, 従来の自己組織化マップで問題はないが, ゲノム解析では, データの入力順によってマップが変わってしまうとマップ上での生物種の関係性を解析するのが困難になる. そのためデータ入力に依存しない方法が必要となる.

そこで, データの入力順に依存しない一括学習自己組織化マップ (Batch Learning-Self Organizing Map: BL-SOM) を用いる.

2.4.1 一括学習自己組織化マップのアルゴリズム

一括学習自己組織化マップのアルゴリズムは, 入力ベクトルの分類, 重みベクトルの更新の2つのステップからなる.

2.4.2 入力ベクトルの分類

すべての入力ベクトルを最小のユークリッド距離を有する重みベクトルを持った競合層のノードに分類する.

一つのノードに複数の入力ベクトルが分類されてもよく, また入力ベクトルが分類されなくてもよい.

2.4.3 重みベクトルの更新

分類された入力ベクトルから, 重みベクトルを更新する. 更新する競合層のノードを中心とした範囲に含まれる分類された入力ベクトルから平均ベクトルを求める. 以下の式によって重みベクトル \mathbf{w}_{ij} を更新する.

$$\mathbf{w}_{ij}^{(new)} = \mathbf{w}_{ij}^{(old)} + \alpha(t) \left(\frac{1}{N_{ij}} \sum_{\mathbf{x}^{(k)} \in S_{ij}} \mathbf{x}^{(k)} - \mathbf{w}_{ij}^{(old)} \right) \quad (5)$$

S_{ij} は, i, j を中心とした一辺 $\beta(t)$ の正方領域に分類された入力ベクトル $\mathbf{x}^{(k)}$ の集合とし, N_{ij} は S_{ij} の要素数である. $t (= 1, 2, \dots, T)$ は学習回数を示す. また, $\alpha(t)$ は学習係数 ($0 < \alpha(t) < 1$), $\beta(t)$ は近傍範囲を表している. $\alpha(t)$ および $\beta(t)$ を以下の式に定義する.

$$\alpha(t) = \max\{0.01, \alpha_{init}(1 - t/T)\} \quad (6)$$

$$\beta(t) = \max\{0, \beta_{init} - t\} \quad (7)$$

ここで, α_{init} と β_{init} は, それぞれの初期値とし, パラメータとして与える.

2.5 標本の切り出し

自己組織化マップへの入力として, 各コドンの出現頻度を用いる. コドンの出現頻度は塩基配列より切り出した標本ごとに計算する. 標本の切り出しは, 切り出し開始位置をずらしながら塩基配列の最後まで繰り返す. 塩基配列の長さが生物種ごと違うため, 標本を切り出す位置のずらし方を生物種ごとに变化させて同じ数だけ標本を切り出せるようにする. 適切な標本の長さが不明であるため, 予備実験によって標本の長さを決定する.

標本の切り出し方を図3に示す.

4 自己組織化マップを用いた生物種の系統解析手法の提案

2.6 出現頻度

切り出した標本からコドンの出現数を1塩基ずつずらしながら数えていく。標本の最後まで数えたら、出現数から各コドンの出現頻度を計算する。

標本でのコドンの出現数の数え方を図4に示す。

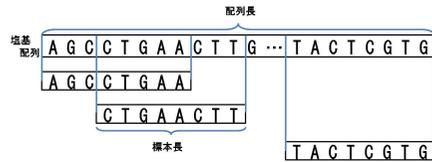


図3 配列の切り出し方

Fig. 3 Shifting of cut-out position

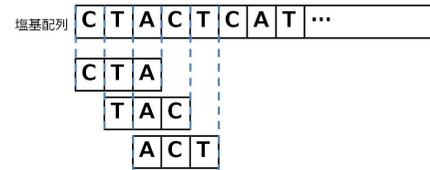


図4 コドンの数え方

Fig. 4 Shifting of codon

3. 系統樹

生物種の進化の解析において、生物種間の進化的関係を表現する方法として良く用いられるのが、進化系統樹と呼ばれるものである。系統樹は木構造で表され、根を持つ有限系統樹と根を持たない無限系統樹の二つに大きく分けられる。

進化系統樹は塩基配列やアミノ酸配列などを基に作成される。DNA データから進化系統樹を作成する方法にはいろいろなものがあるが、近隣結合法 (NJ 法)、最尤法などがある。

本研究では、自己組織化マップによって作成されたマップ上の領域間の距離から系統樹を作成する。作成されたマップから同じ生物種の領域に属するノードの重みベクトルから特徴ノードを決定し、特徴ノード間の距離を距離行列法によって系統樹を作成する。

3.1 距離行列法

距離行列法とは、距離行列を用いて系統樹を作成する方法である。距離行列とは、生物種間の距離すべての距離を並べたものである。

距離行列法の手順を以下に示す。

Step1. 生物種の全ペアから距離行列の計算

Step2. 距離行列の要素中の最小値をとり、それに対応する生物種の組み合わせを1つの群とする。

Step3. Step2の群についての系統樹を作成する。

Step4. Step2の群と他の生物種との距離行列を計算

Step5. 距離行列の要素が1つになるまでStep2. からStep4. を繰り返す。

3.2 距離計算法

学習後のマップに現れる生物種ごとのノードから特徴ノードを決定し、特徴ノード間の距離を関数を用いて計算する。計算した距離を生物種間の距離として距離行列法によって系統樹を作成する

距離行列法に用いる距離の計算手順を以下に示す。

Step1. 学習後のマップ上で同じ生物種クラスに属するノードの重みベクトルの平均を求めると。

Step2. 平均とユークリッド距離の近い重みベクトルを持つノードを特徴ノードとする。

Step3. 特徴ノード間の距離を計算する。

自己組織化マップによって得られる情報は、ノードの座標と重みベクトルの2種類あるため、特徴ノードを決定しノード間の距離を計算する際に使用する関数を3種類準備した。3種類の評価関数は実験によって比較する。使用する評価関数を以下に示す。

3.2.1 重み空間上の距離

学習後のノードの重みベクトルから特徴ノード間の距離を計算する。計算式は以下に示す。

$$distance_{ab} = \sqrt{\sum_{n=0}^{63} w_{an}^2 + w_{bn}^2} \quad (8)$$

ここで a と b はすべての生物種の組み合わせが入る。

3.2.2 座標空間上の距離

特徴ノードの座標を用いて計算する。計算式は以下に示す。

$$distance_{ab} = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \quad (9)$$

ここで a と b はすべての生物種の組み合わせが入る。

3.2.3 合成距離

重み関数と座標関数を合わせたものを距離として計算する以下の式で計算する。

4. 検証実験

提案手法を評価するために20種類の生物種を用いて検証を行う。生物種のリボソームタンパク質遺伝子を使用する。自己組織化マップによって作られたマップ上に現れた生物種

5 自己組織化マップを用いた生物種の系統解析手法の提案

ごとの領域間の距離が生物種間の距離と対応しているかによって評価する．領域間の距離を生物種間の距離として系統樹を作成し，隣接関係と生物種間の距離の対応を確認する．

まず，塩基配列から切り出す標本の長さを予備実験によって決定する．次に，系統樹作成を行うための特徴ノード間の距離の算出方法の3種類について比較実験を行う．

予備実験によって求めた特徴ノード間の距離を求める関数を使って20種類の生物種について進化系統解析を行う．

4.1 対象データ

実験に使用する遺伝子はリボソームタンパク質遺伝子を用いる．リボソームは翻訳のための重要な部分で，地球上に現存するほぼ全ての生物に存在する．リボソームは，1～4種類のrRNA(リボソームRNA)と数十個のリボソームタンパク質で構成されている．これまではrRNAが解析によく用いられてきたが，近年の研究でリボソームタンパク質の以上に起因すると見られる疾患や変異の例が報告されており，リボソームタンパク質も何らかの重要な機能があると考えられ，解析が行われている．

リボソームタンパク質遺伝子はRPG(Ribosomal Protein Gene database)に公開されているものを使用した¹⁸⁾．

実験の対象となる生物種を表1に示す．動物から菌類，植物類まで広い範囲の生物種を対象とした．

表1 20種類の生物種
Table 1 Species for experiments

略称	生物名	略称	生物名
Hs	ヒト	Rn	ドブネズミ
Mm	ハツカネズミ	Fr	フグ
Ci	ホヤ	Dm	ショウジョウバエ
Ag	ハマダラカ	Am	セイヨウミツバチ
Ce	センチュウ	Mg	イモチ病菌
Fg	赤カビ	Dd	キイロタマホコリカビ
Yl	アルカン資化酵母	Sp	分裂酵母
Sc	出芽酵母	Cn	クリプトコッカス
Um	トウモロコシ	Ro	クモノスカビ
At	シロイナツナ	Cr	コナミドリムシ

4.2 系統樹作成実験

特徴ノード間の距離の計算法についてそれぞれ系統樹を作成し，比較する．

比較する系統樹作成に用いる距離は以下の3つ

- 重み空間上の距離
- 座標空間上の距離
- 合成距離

4.3 実験条件

使用する生物種は20種類の中から6種類の生物種ヒト・ドブネズミ・ハマダラカ・トウモロコシ黒穂病菌・キイロタマホコリカビ・シロイヌナズナを選択し使用した．

実験に使用するSOMのパラメータを以下に示す．

SOM マップサイズ：24 × 24

SOM 初期学習係数：0.5

SOM 初期学習近傍：12

SOM 学習回数：100回

塩基配列から切り出す標本の長さは事前に予備実験によって決定し，2048塩基とした．使用する生物種の場合，どのような系統樹が作成されるかを生物種の分類情報から予想した．予想した系統樹を図5に示す．

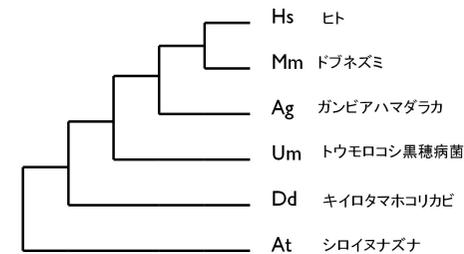


図5 予想系統樹
Fig.5 Ideal tree

4.4 系統樹作成実験結果

学習後のマップを図6に，そのマップから作られた系統樹を図7，図8，図9に示す．

重み関数による系統樹ではヒトとドブネズミが結ばれているが，その2生物種とシロイヌナズナが結ばれており，予想した系統樹と違う結果となった．座標間距離による系統樹では，ヒトとドブネズミの間にトウモロコシ黒穂病菌が入っており予想の系統樹とは離れた結

6 自己組織化マップを用いた生物種の系統解析手法の提案

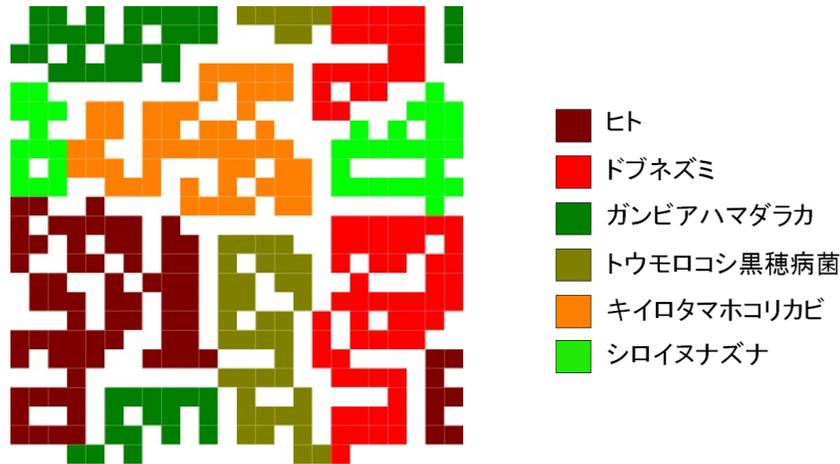


図 6 系統樹作成実験結果マップ
Fig. 6 Result of SOM

果となった．合成距離による系統樹では，予想した系統樹と同じ系統樹が得られた．

5. 20 生物種での検証実験

予備実験によって求めた特徴ノード間の距離の計算法を使用して 20 種類の生物種を対象に系統解析を行い提案手法の検証を行う．

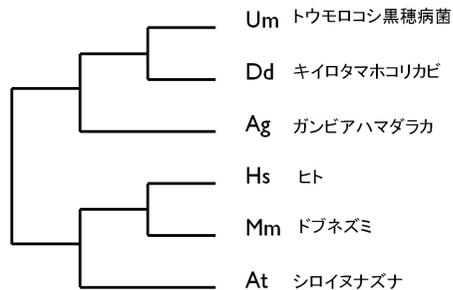


図 7 重み空間上の距離による系統樹
Fig. 7 Evolutionary tree by distance of weight

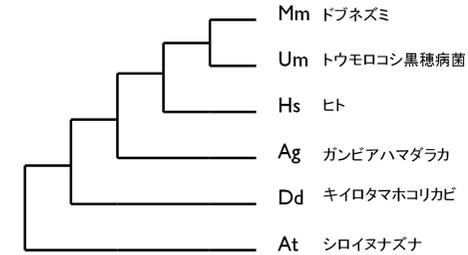


図 8 座標空間上の距離による系統樹
Fig. 8 Evolutionary tree by distance of ordinate and abscissa

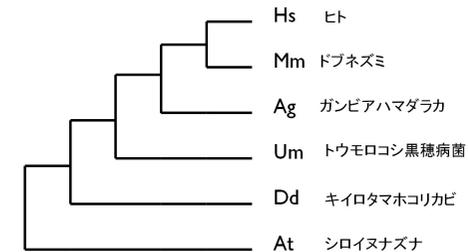


図 9 合成距離による系統樹
Fig. 9 Evolutionary tree by mix distance

5.1 実験条件

実験に使用する SOM のパラメータを以下に示す．

SOM マップサイズ：40 × 40

SOM 初期学習係数：0.5

SOM 初期学習近傍：20

SOM 学習回数：100 回

塩基配列から切り出す標本の長さは 2048 塩基とした．特徴ノード間の距離は合成距離を使用する．

5.2 実験結果

自己組織化マップの結果を図 10 に，そのマップから作られた系統樹を図 11 に示す．

マップを見ると 20 種の生物種ごとの領域が形成されている．ヒト，ドブネズミ，ハツカ

7 自己組織化マップを用いた生物種の系統解析手法の提案

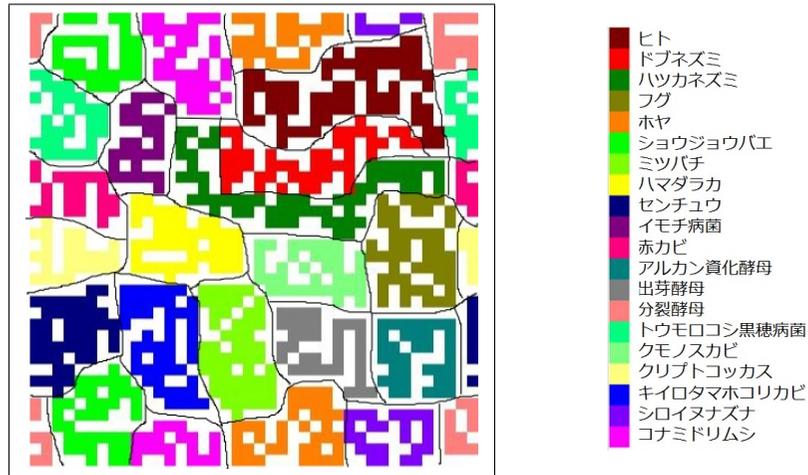


図 10 20 生物種でのマップ
Fig. 10 Result of SOM

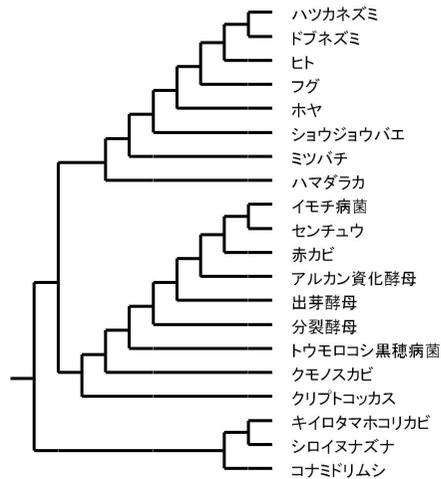


図 11 20 生物種での系統樹
Fig. 11 Result of evolutionary tree

ネズミのそれぞれ領域が近い位置に配置されている。また、アルカン資化酵母，出芽酵母，分裂酵母の領域も近い位置に配置されており，領域の隣接関係と生物種間の距離が対応したマップを得られた。

系統樹の結果をみるとドブネズミ・ハツカネズミが結ばれており，2 生物種と同じ哺乳類であるヒトが結ばれている．系統樹を大きく見ると 3 つの木に分割できる．それぞれ上から動物，菌類，植物に分かれており，進化系統がわかるような系統樹が作成できている．

6. ClustalW による系統樹との比較

パターンマッチングによる系統樹作成法である ClustalW によって作られた系統樹との提案手法によって作った系統樹との比較を行った．ClustalW はパターンマッチングによって系統樹を作成するツールとして知られている．ClustalW による系統樹を図 12 に，提案手法によって作られた系統樹を図 6 に示す．

ヒト・ドブネズミ・ハツカネズミ・フグは ClustalW と同じような結ばれ方をしていたが，植物に分類されるはずのキイロタマホコリカビ・シロイヌナズナ・コナミドリムシは，ClustalW での系統樹ではばらばらになっており，あまりよい系統樹ではない．提案手法による系統樹は ClustalW での系統樹より進化系統が推測しやすいと考えられる．

7. 終わりに

塩基配列のみを用いた生物種の進化系統の解析を行う 1 つの手法として，コドンの出現頻度を入力とした自己組織化マップによる進化系統解析手法を提案した．

20 種類の生物種を対象とし，リボソームタンパク質遺伝子を使用した．自己組織化マップによるマップ作成と作成されたマップから系統樹作成を行った．作成されたマップから 20 種の生物種がマップ上でそれぞれの領域に分かれ，系統的に近い生物の領域間の距離は短いことが確認できた．マップから作成した系統樹は，大まかにみると，動物類，菌類，植物類に分かれており，細かくみると，ヒト，ドブネズミ，ハツカネズミが部分木を形成していたりと，系統的に近い生物種同士が結ばれている．

パターンマッチングを用いたツールである ClustalW によって作成した系統樹と提案手法による系統樹を比較した結果，ClustalW ではあまりよい系統樹が得られない遺伝子データでも提案手法でやると系統関係が推測できることが分かった．

今後の展開として，生物種数を増やしての実験や，SOM の初期重みを固定するなどの工夫が必要であると考えられる．

参 考 文 献

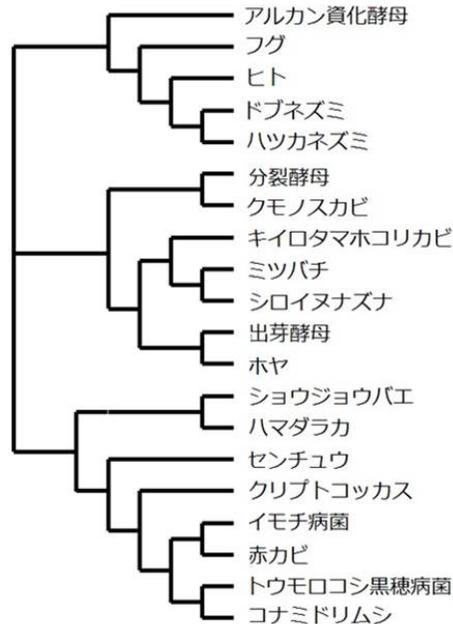


図 12 ClustalW による系統樹
Fig.12 Evolutionary tree by ClustalW

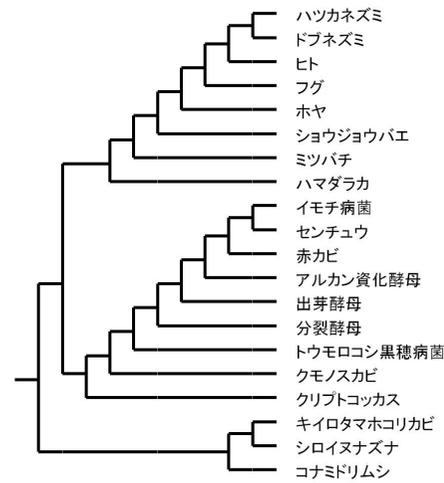


図 13 提案手法による系統樹
Fig.13 Evolutionary tree by proposed method

- 1) 丸山修, 阿久津達也: バイオインフォマティクス-配列データと構造予測-, 朝倉書店. 2007.
- 2) 池内俊彦, 畠中寛: タンパク質と遺伝子, オーム社. 1996.
- 3) 宮田隆, 五條堀孝: ゲノム情報を読む, 共立出版. 1997.
- 4) Gbas, C. and Jambeck, P.: 実践バイオインフォマティクス-ゲノム研究のためのコンピュータスキル-, 株式会社オライリー・ジャパン. 2002.
- 5) 大矢雅則: 情報進化論, 岩波書店. 2005.
- 6) I.Yoshihara, H.Takashima, K.Yamamori and K.Sugawara: Similarity Comparisons of Seeds Using Image Compression Technology, *Memoirs of the Faculty of Engineering University of Miyazaki*, Vol.35, pp.251-256 (2006).
- 7) Y.Sakaguchi, I.Yoshihara, K.Yamamori and N.Kenmochi: Criterion of Evolution Analyzing DNA Sequences based on Statistical Mechanics, *Memoirs of the Faculty of Engineering University of Miyazaki*, Vol.35, pp.269-274 (2006).
- 8) Abe, T., Kanaya, S., Kinouchi, M., Y.Ichiba, T.K. and Ikemura, T.: Informatics for unveiling hidden genome signatures, *Genome Research*, No.13 (2003).
- 9) Y.Koyama, K.Nishimuta, K.Yamamori, I.Yoshihara and M.Yasunaga: Quest fo genetic information hidden behind disorder in DNA sequences, *Proc of 15th Int Symp Artificial Life and Robotics*, pp.824-827 (2010).
- 10) K.Yamamori, Y.Fujita, M.Aikawa and I.Yoshihara: Identification of Exon-intron Boundaries by Integration of Base-oriented Genetic Programming and Statistical Heuristics, *Proc of 12th Int Symp Artificial Life and Robotics*, pp.657-660 (2007).
- 11) I.Yoshihara, Y.Yano, K.Yamamori and M.Aikawa: Analysis of Organism's Evolutionary Relationships using Information Entropy, *Memoirs of the Faculty of Engineering*, Vol.38, pp.289-294 (2009).
- 12) S.Miyazaki, H.Sugawara and M.Ohya: The efficiency of entropy evolution rate for construction of phylogenetic trees, *Genes Genet Syst*, No.71 (1996).
- 13) 多田秀樹, 関田英太郎: HIV-1(V3 loop) とエントロピー, *数理解析研究所講究録*, No.1420 (2005).
- 14) 佐藤圭子, 大矢雅則: HIV の情報論的解析, *信学技報*, No.IT97-52 (1997).
- 15) 平野裕嗣, 山本隆史, 大矢雅則: エントロピー進化率による HIV の解析, *信学技報*, No.IT98-38 (1998).
- 16) 日谷亮, 佐藤圭子, 大矢雅則: インフルエンザ A 型ウイルスにおける HA 蛋白質のエントロピー型カオス尺度による分類, *情報処理学会研究報告* (2007).
- 17) T. コホネン: 自己組織化マップ, シュプリンガー・フェアラーク東京. 1996.
- 18) ribosomal proteingenes database: . <http://ribosome.med.miyazaki-u.ac.jp>.