

異種系統樹データベース間の調停作業支援のための球面描画法

岡本 潤一[†] 森 康真^{††} 黒木 進^{††} 北上 始^{††}

[†]広島市立大学情報科学部

^{††}広島市立大学大学院情報科学研究科

〒734-0004 広島県安佐南区大塚東 3-4-1

E-mail: {junichi, mori, kuroki, kitakami}@de.info.hiroshima-cu.ac.jp

あらまし 分子進化学の分野において、1つの系統樹は平面上で根から放射状に広がる平面グラフとして描かれていたため、2種類の系統樹の葉どうしを辺で結んだ調停グラフを描画するとき、葉どうしをつなげる辺の間に交差が生じてしまうという問題がある。すなわち、辺の間に交差を持つ調停グラフは認知的な操作性を低下させてしまう。この交差を避けるため、従来の調停グラフの描画においては、描画が簡略化され、どの葉も同じ深さをもつ木として描かれていた。すなわち、この放射状に広がる系統樹の概形や枝の長さは無視されていたため、十分な生物学的情報を持つ調停グラフを描くことができなかった。

本研究では、調停グラフの描画における簡略化を避けるために、平面上で根から放射状に広がる2つの異なる系統樹から構成される調停グラフを球面上に描画する手法について提案する。

キーワード 調停グラフ, 視覚化, 球, 系統樹

Spherical Drawing Method to Support Reconciliation Work across Heterogeneous Trees Database

Junichi Okamoto[†] Yasuma Mori^{††} Susumu Kuroki^{††} Hajime Kitakami^{††}

[†] Faculty of Information Sciences, Hiroshima City University

^{††} Graduate School of Information Sciences, Hiroshima City University

3-4-1 Ozuka-Higashi, Asa-Minami-Ku, Hiroshima 731-3194, JAPAN

E-mail: {junichi, mori, kuroki, kitakami}@de.info.hiroshima-cu.ac.jp

Abstract In the field of molecular evolution, one phylogenetic tree is drawn as a planar graph that radially spreads from the center of the root on a plane. Therefore, there is a problem that the existing reconciliation graph drawn on the plane has cross overs of edges where the each edge connects two leaf sets in the reconciliation graph. Namely, the reconciliation graph with crossovers of edges make cognitive interface of users turn for the worse. In order to avoid the crossovers of edges, the existing reconciliation graph drawn on the plane disregards both the configuration of the tree radially spread from the center and the length of the tree branch. Therefore, the graph lacks biological information such as the tree configuration and the branch length. This paper proposes a method for drawing reconciliation graphs on a sphere in order to avoid the lack of biological information in the reconciliation graph where the graph is composed of two phylogenetic trees radially spread from the center of the root and drawn on a sphere.

Keywords: reconciliation graph, visualization, sphere, phylogenetic tree.

1. はじめに

分子進化学とは、生物の DNA やアミノ酸などの分子配列データを用いて進化を解明することを目的とした研究分野である。これらの分子配列データには過去に起きた生物の進化に関する情報などが含まれており、それらを含む DNA データベース[1]、及び、タンパク

質データベース[2]を解析することによって、生物の進化や起源、特定の遺伝子に対する同定などを解明することができる。本論文では、これらのデータベースに含まれている生物分類樹データベース、及び、塩基配列やアミノ酸配列といった分子配列データから作成される分子進化系統樹データベースに着目する。これら

の系統樹データベース間の関係を比較する方法のひとつとして調停処理[3],[4],[5]がある。調停処理では、系統樹をそれぞれ順序木とみなし、比較する2種類の順序木を向い合せて配置した調停グラフを作成する。描画された調停グラフに対する認知的な操作性が確保できれば、調停作業の効率化が可能となる。そのためには、調停グラフの2つの葉集合間を結ぶ辺の重なりがない状態か、もしくは、辺の交差が少ない状態にする必要がある。2つの順序木における2つの葉の交換[6],[7]は、この状態を見つけ出すのに重要な役割を演じている。

分子進化学において、系統樹の利用価値を高めるために、系統樹を平面上に根から放射状に広がる平面グラフとして描いている。図1に平面上に描かれた調停グラフの簡単な例を示す。図1は、2つの系統樹を向い合わせたグラフであり、生物種'a', 'b', 'c', ..., 'g'を別々の観点で分類した2つの系統樹を向い合せて配置し、対応する種どうしを1対1で結んだ辺(以後、種対応辺と呼ぶ)によって対応づけている。

計算機上で平面に調停グラフを描画する場合、R.D.M.Pageらの提案するGeneTree[8]と呼ばれるソフトウェアツールがある。このシステムでは、両者の木を平面上で向かい合わせに配置し、2つの葉集合間で種対応辺を結んでいる図と展開された調停木とを切り替えて表示できるものである。GeneTreeシステムにおいて、系統樹はピラミッド構造をした有限木を、平面上で向かい合わせたグラフで表されるが、本来、根から放射状に広がる系統樹の概形と枝の長さが無視されているなど、生物学的情報が十分に反映されていない。専門家が調停グラフを手作業で描く場合、系統樹作成に用いられた生物種の数が多ければ多くなるほど調停作業が困難になる上に、計算機と同様に、手作業での調停作業は平面に描かざるを得ないため、系統樹の概形と枝の長さを無視している。ある程度の情報技術を持つ分子生物学の専門家が、より高度な計算機上で調停処理を行うアルゴリズムを考えない限り、生物学的情報の欠落した、すなわち、根から放射状に広がる系統樹の概形と枝の長さが無視された系統樹が描かれる可能性が示唆される。

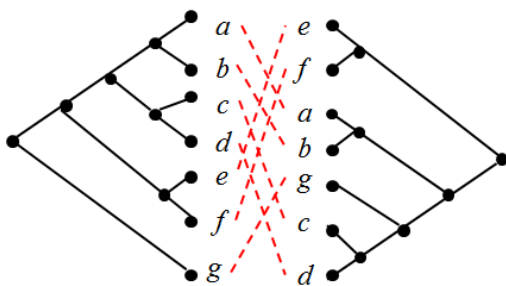


図1 調停グラフの例

そのため、分子生物学の専門家に代わり、従来の平面上に描かれる、生物学的情報の欠落した調停グラフとは異なる、根からの放射状に広がる系統樹の概形と枝の長さを表現した調停グラフを描画する手法について提案することができれば、調停グラフの生物学的情報を調停作業にそのまま反映でき、且つ、調停作業の効率化を実現することができる。

以上より本論文では、調停処理での利用価値をより高めるとともに、調停グラフを見やすくし、調停作業を効率化することを目的とする。そのため本論文では、CGを用いて、根から放射状に広がる異種系統樹から構成される調停グラフを球面上に描画する手法について提案する。

以下、2章では、従来手法の説明として系統樹と調停グラフについて説明し、平面上に描かれた調停グラフの問題点を述べる。3章では、球面上に描かれた調停グラフの利点を述べる。4章では、システムの処理手順と機能について説明し、5章で簡単な描画例を紹介する。6章では、システムのユーザインタフェースの評価を行う。また、7章で本論文のまとめを行う。

2. 従来手法

本章では、調停に用いられる異種系統樹の種類とその特徴について説明し、従来手法である平面上に描かれた調停グラフの問題点を、具体例を挙げて述べる。

2.1 系統樹の種類と特徴

調停に用いられる系統樹には、生物種を形態学的な観点から分類することによって作成された木である生物分類樹や、DNAの塩基配列やタンパク質のアミノ酸配列の違いから推測された木である分子進化系統樹などがある。例として、分子進化系統図の例を図2に示す。図2では、根から放射状に広がる系統樹の概形と枝の長さが表現されている。調停に用いる異種系統樹の組み合わせとしては、生物分類樹と分子進化系統樹、異なる分子進化系統樹などがある。前者の組み合わせは、生物分類樹の視点による生物分類樹の見直し[4]に重要であり、後者は宿主-寄生関係にある生物の間の進化的関係を調べるのに重要である。

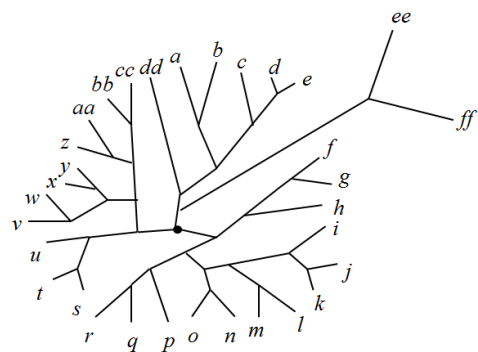


図2 放射上に広がる系統樹の例

また、後者の例として、多種類の鼠とそれらに寄生する蚤のそれぞれの分子進化系統樹を比較検討する研究がある[9].

2.2 平面上に描かれた調停グラフ

平面上に描かれた調停グラフは、図1のように、根から放射状に広がる系統樹の概形と枝の長さが無視され、どの葉も同じ深さの木として描かれていた。しかしながら、本来は図2のような系統樹が描かれるべきである。そこで、図2のような2つの異なる系統樹を並べた場合、描かれる調停グラフの問題点について考える。根から放射状に広がる2つの系統樹の調停グラフを図3に示す。図3から、種対応辺どうしの交差や、種対応辺と系統樹の枝の交差が多数できてしまうことがわかる。また、平面上で放射状に広がる系統樹の概形を表現した調停グラフを描くと、どのように種対応辺を描画しても依然として交差は残ってしまう。本来、調停グラフは、種対応辺どうしの交差に着目するために描かれ、その交差はなるべく少ないほうが好ましい。交差を減らす作業を行う際、放射状に広がる系統樹の概形と枝の長さを表現したままの調停グラフでは、見にくい上に作業効率が悪い。よって、このような調停グラフを描くことは望まれないため、平面上に描かれる調停グラフは放射状に広がる系統樹の概形と枝の長さが無視されている。

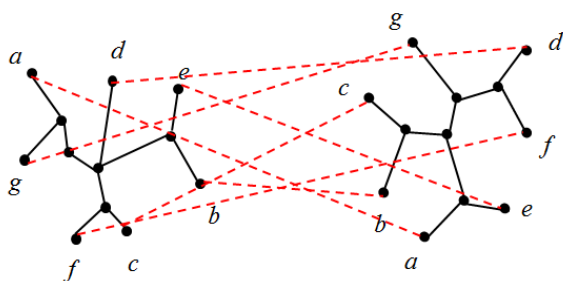


図3 系統樹の概形を表現した調停グラフの例

3. 球面上に描かれた調停グラフ

平面上に描画された調停グラフでは、放射状に広がる系統樹の概形や枝の長さが無視されていた。そこで、放射状に広がる系統樹を向い合せに配置することのできる球面上に調停グラフを描く方法を提案する。球面上で、2つの異なる系統樹の根を北極点と南極点のそれぞれに配置し、根からトップダウンに放射状に広がる系統樹をそれぞれ描画する。そして、種対応辺をそれぞれ結べば、放射状に広がる系統樹の概形と枝の長さを表現した調停グラフを描くことができる。これにより、生物学的情報が正しく反映された系統樹を調停でき、作業効率を上げることができる。

さて、球面に描かれた調停グラフにおいて、まだ交差が残っている場合は、交差数を減少させる方法[6],[7]が提案されている。理論的に除去不可能な交差

がいくつか残った場合は、交差に関連する生物種の中に、進化の過程で絶滅したり、生息域が変化したりする生物種が存在するなどの可能性を示唆することがわかっている[3],[4],[8].

なお、著者らが作成した球面インタフェースの描画システムでは、CGを用いて球面上に調停グラフを描くことによって、球と調停グラフに対して、球の回転や枝の移動・色分けなどといった認知的な操作を行うことができる。これにより、認知的な操作性の向上を実現できる。

4. システム概要

本論文では、OpenGLにより、CGを用いて根から放射状に広がる異種系統樹から構成される調停グラフを、球面上に描画するシステムを作成した。本章では、本システムの処理手順と機能について説明する。

4.1 システムの処理手順

本システムの処理手順について簡単に説明する。

- (1) 入れ子状の入力データを、木構造のデータに変換する。
- (2) 木構造のデータを幅優先探索しながら、系統樹の球面上のノード座標を決定する。
- (3) 決定したノード座標に基づき、系統樹を描画する。
- (4) 最後に種対応辺を描画して、調停グラフの完成とする。
- (5) 描画後、描画した球や調停グラフに対しての操作を各種提供する(機能の詳細については次節参照)。

4.2 システムの機能

本システムの機能について説明する。本システムでは、初期描画後に、球や調停グラフに対して様々な操作を行うことにより、認知的な操作を提供する。以下、各種機能に対する入力をまとめたものを表1に示す。なお、表中の入力はマウス操作またはキーボード操作を表している。

表1 各種機能と入力

機能	入力
球の回転	左ドラッグ
球の拡大縮小	ホイール
球の平行移動	w, a, s, d
球の拡大縮小/平行移動リセット	r
枝の色分け	右クリック
枝の色分けリセット	f
枝の移動	右ドラッグ
枝の移動リセット	v
系統樹の拡大縮小機能	m, n

5. 描画例

描画例に使用するデータには、シトクローム b と 12SrRNA に関する霊長類の分子配列データを、それぞれ 120 種程度収集し、ClustalW[1]を用いて作成された分子進化系統樹を用いる。シトクローム b は酸素呼吸をする生物種にみられるアミノ酸であり、12SrRNA は系統樹解析に適しているとされる DNA である。本来は 16SrRNA を用いる予定だったが、十分なデータ数が集められなかったため、12SrRNA とした。

以上のデータから、CG を用いて球面に調停グラフを描画した例を図 4 に示す。なお、画像作成にあたって、球の回転や系統樹の拡大などの各種操作を任意に行っている。

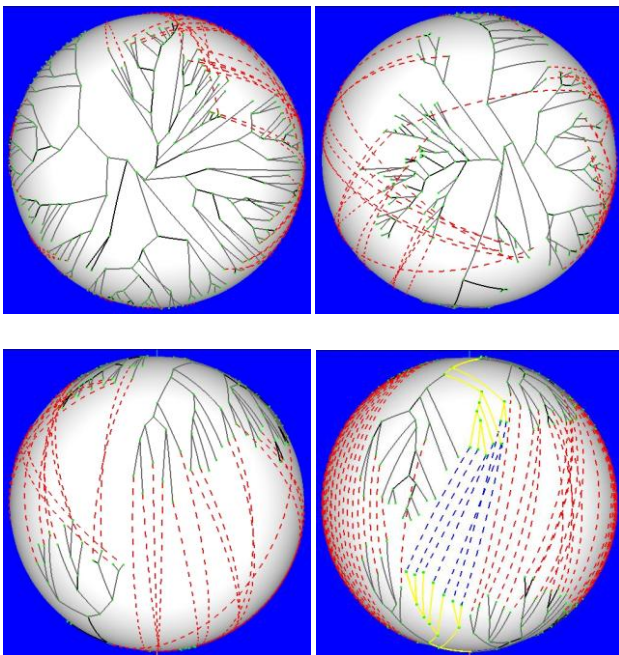


図 4 真上(左上), 真下(右上), 左下(側面)から球を見た調停グラフと、操作を行った調停グラフの例

図 4 から、根から放射状に広がる系統樹の概形と枝の長さが表現されていることがわかる。

6. 評価

本章では、認知的な操作性の確保という観点から著者らが作成した描画システムの評価を行うために、Shneiderman が提唱する「対話型設計における 8 つの黄金律」[10]を用い、本システムのユーザインタフェースが 8 つの黄金律を満たしているかどうかを評価する。8 つの黄金律は以下の通りである。

- (1) 一貫性の保持
- (2) 頻繁に使うユーザのための支援
- (3) 有益なフィードバックの提供
- (4) 段階的な達成感を与える対話を実現

- (5) 簡潔なエラー処理
- (6) 逆操作の許可
- (7) ユーザの主体的な制御権
- (8) 短期記憶領域の負担の軽減

以上の 8 項目について、本システムの環境に従って解釈、検討、拡張を行ったうえで、本システムに適用し評価を行う。

ユーザインタフェースとして、マウス操作、及び、キーボード操作を採用している。似たような操作には同じインタフェースを用いることによって、システム全体で一貫性を持った操作を実現する。これは、黄金律(1)に相当している。ユーザが行える操作は、マウスやキーボードによるポリゴンの操作に限られているため、不正な数値入力などのエラーを防ぐことが可能となる。これは黄金律(5)に相当している。

画面内での操作として、球の回転や拡大縮小、枝の色分けや移動といった機能を、ユーザが自由に調節することができる。これによって黄金律(7)を実現することができる。

ユーザインタフェースによる入力は、全てインタラクティブに画面に反映されるため、これによって、黄金律(3)が満たされている。しかし、キーボードによるコマンドは数が多く、操作によりモードを切り替える必要があるなど、初めて使う人には操作説明がないと使いづらい点が残る。コマンドの一覧を端末内へ表示させているが、初めてこのシステムを使うユーザにはわかりにくいと考えられる。そのため黄金律(8)の考慮は不十分であると考えられる。これを解決する方法として、外部ユーザインタフェースとしてボタンやタブ機能を実装し、マウスのクリックにより全てのコマンドを選択できるようにすることが考えられる。また、操作の結果は即座に知ることができ、次の入力に反映させることができる。これにより、黄金律(4)が実現されている。

黄金律(2)に関して、ユーザインタフェースは頻繁に使うユーザに対しての支援を提供していない。ただし、システムを起動する上で、ユーザがパラメータを調節しなくて済むようなアルゴリズムを実装している。例として、系統樹の枝の長さの自動調整などが挙げられる。

ユーザインタフェースの操作として、誤操作に対する修正を行う機能を実装している。これによって黄金律(6)を実現している。ただし、誤操作の修正は、各種のポリゴン操作全体に対するリセット操作が大部分を占めているため、部分の修正がし難いという点で使い勝手が悪い。そのため黄金律(8)に負担をかけていると考えられる。

以上から、ユーザインタフェースの操作は、認知的

な操作性に優れている反面、多くのコマンドにより操作方法が分かり難くなっていることがわかる。また、このシステムが使用できる環境は、OpenGL が使用できる環境上に限られているため、計算機の扱いに慣れていないユーザには使いにくいと考えられる。本システムが、どんな環境上でも動くようにできれば、多くのユーザに使ってもらえる可能性が高まると考えられる。また、描画以外にも、操作のログや具体的な数値の変化、枝の交差の情報などをフィードバックできれば、ユーザには有益であると考えられる。

7. まとめ

本論文では、異種系統樹データベース間の調停作業支援のための球面描画法を提案した。従来の手法では、根から放射状に広がる系統樹の概形と枝の長さが無視された調停グラフが描かれていたが、提案手法により、根から放射状に広がる系統樹の概形と枝の長さを表現することができた。それにより、調停グラフを見やすくし、調停作業を効率化するための手段として、調停グラフの球面への描画の有効性を示した。また、調停グラフや球に対しての操作を各種提供することによって、ある程度の認知的な操作性を確保することができた。

本システムでは、枝の長さの反映が完全ではないことが分かっている。現在の描画法では、系統樹の根からトップダウンに系統樹を描画していたが、葉からボトムアップに系統樹を描画することによって、それを解消したい。また、調停作業を効率化する上で、調停グラフの交差を減少させる手法を実装することは重要である。平面上に描かれる調停グラフと球面上に描かれる調停グラフでは、調停グラフの持つ生物学的な情報量が異なっている。そのため、それらの調停グラフにおいて、交差数の減少数がどのように変化するかというのは調停グラフによって様々であると予想される。ただし、交差数減少の手法としては、平面上と球面上とで違いはない。最終的に、分子生物学の専門家が気軽に使用できるように、利用しやすい環境を提供することも課題の1つである。

謝辞 インターネット経由で、DNA 配列、及び、アミノ酸配列データを快く提供して頂いた DDBJ 及び UniProtKB のスタッフに深く感謝致します。なお、本研究の一部は、日本学術振興会、科学研究費補助金(基盤研究(C)、課題番号: 20500137) の支援により行われた。

参 考 文 献

- [1] 日本 DNA データバンク (DDBJ): <http://www.ddbj.nig.ac.jp>
- [2] 国際タンパク質情報データベース (UniProtKB):

<http://www.uniprot.org/>

- [3] Page, R.D.M., and Charleston, M.A.: Reconciled Trees and Incongruent Gene and Species Trees, In: B Mirkin, F R McMorris, F S Roberts and ARzhetsky (eds), *Mathematical Hierarchies in Biology*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, American Mathematical Society, Vol. 37, pp.57-70, 1997.
- [4] Page, R.D.M., and Charleston, M.A.: From Gene to Organismal Phylogenecy: Reconciled Trees and the Gene/Species Tree Pblem, *Molecular Phylogenetics and Evolution* Vol.7, p231-240, 1997.
- [5] Goodman, M., Czelusniak, J., Romero-Herrera, A.E., and Matsuda, G.: Fitting the Gene Lineage into its Species Lineage: A parsimony strategu illustrated by Cladograms Constructed from Globin Sequences, *Systematic Zoology*, Vol. 28, pp.132-168, 1979.
- [6] 北上 始, 森 康真, 大田聡史, 斉藤成也:異種系統樹間の調停のためのゼロ交差制約の充足, 情報処理学会論文誌: データベース, Vol.40 No.SIG05,pp.1-14, 1999.
- [7] 田村慶一, 森 康真, 北上 始:Extremal Optimization による調停グラフの交差数減少, 情報処理学会論文誌:数理モデル化と応用, Vol. 49 No. SIG 4(TOM 20), pp.105-116, 2008.
- [8] Page, R.D.M.: GeneTree: Comparing gene and Species Phylogenies using Reconciled Trees, *Bioinformatics*, Vol. 14, No. 9, pp.819-820, 1998.
- [9] Page,R.D.M.:TemporalCongruenceRevisited: Comparison of Mitochondrial DNA Sequence Divergence in Cospeciating Pocket Gophers and Their Chewing Lice, *Systematic Biology*, Vol. 45, No.2,pp.151-167, 1996.
- [10] Shneiderman, B.: *Designing the user interface: Strategies for effective human-computer interaction*, 3rd ed., Addison-Wesley Publishing, MA, 1998.