

WAKO

Infomatic

World

September 2009

No.17

実験生物学者・
実験化学者のための
IT活用誌

Contents

Cell Illustratorを用いた遺伝子多重化の効果の解析 2

国立遺伝学研究所 集団遺伝研究部門
特任研究員 佐藤 行人
東京大学海洋研究所 分子海洋科学分野
教授 西田 睦

Cell Illustratorを用いた細胞システム維持に影響する
生命パスウェイ重要部位の推定 4

大島商船高等専門学校 商船学科
講師 北風 裕教

次世代シーケンサーデータ処理ツール
“MetaGenomeGAMBLER”のご紹介 9

インシリコバイオロジー株式会社
代表取締役 大山 彰

パスウェイ描画解析ソフト
“Cell Illustrator【セルイラストレータ】” 8

次世代シーケンサーデータ処理ツール
“MetaGenomeGAMBLER” 12

Cell Illustrator を用いた遺伝子多重化の効果の解析

国立遺伝学研究所 集団遺伝研究部門

特任研究員 佐藤 行人

東京大学海洋研究所 分子海洋科学分野

教授 西田 睦

1 生物の進化と遺伝子重複

ヒトのゲノムには、タンパク質をコードするものだけを見てもおよそ21,400個の遺伝子が含まれる[1]。この2万以上にも及ぶ遺伝子は、当然のことながらヒト進化の過程で忽然と出現したわけではなく、また生命誕生当初からすべて揃っていたわけでもない。生物の進化の過程で、種が分岐して生物が多様化してきたのと同じように、遺伝子も「遺伝子重複」によって分岐し、さらに機能が変化することで多様化してきた[2]。

遺伝子機能の変化は、重複した遺伝子に変異が蓄積することで起きる。タンパク質遺伝子の場合を考えると、重複によって余剰な遺伝子ができれば、そこには様々な変異—置換・挿入・欠失、ドメインの入れ替え、転移因子の挿入などが蓄積する余地が生じる。それらが新しいタンパク質機能をもたらせば、その変異を獲得した重複遺伝子は、進化的に「新しい遺伝子」として存続することになる[2, 3]。

以上から遺伝子重複は、タンパク質などの機能分子の多様性や進化、さらには、生物の多様性と進化を理解する上でたいへん重要な基礎過程であると言える。

2 重複によって「システム」の挙動は変わるか？

上で述べたように、重複は遺伝子やタンパク質の多様性や進化を理解する上で重要であるが、それだけでなく、より高次の生物システム、例えば代謝経路やシグナル伝達経路の進化を考える上でも興味深い。代謝経路を例にとれば、どのような遺伝子の重複と機能の変化によって、既存の代謝経路が組み合わさって大規模な代謝システムが出現したのか、あるいは組織特異的なサブシステムが形成されてきたのかを解明していくことは、これからのエキサイティングな課題の1つだろう。

もう1つ、遺伝子重複がもたらす別の重要な効果として、遺伝子産物の量の変化が挙げられる[3]。タンパク質コーディン

グ遺伝子が重複すると、シンプルに考えれば産物量が増加すると期待される。また、重複後、転写制御領域の機能が変わった場合には、ターゲット遺伝子の発現部位・時期・強度などが変化して、結果的に、ある系におけるターゲット遺伝子の産物量が増えるだろう。

ではそうした重複によって、代謝やシグナル伝達などの「システム」に関する遺伝子の産物量が増減した場合に、そうした変化は、システム全体の挙動や機能に対してどのような効果をもつのだろうか？こうした問いは、(個別遺伝子還元思考でなく)システムのレベルで生物進化を考えていく上でたいへん重要だと思われるものの、今までほとんど検討されてこなかった。複雑なシステムの動態の中での産物量変化の効果を検討するには、知識(反応速度論や制御工学など)や技術(プログラミングやシステム構築など)が必要となる。そうした要件が、主にDNA配列比較やウェット実験から進進理解にアプローチしてきた分子進化研究者にとって、容易には手の出し辛いハードルであったのかもしれない。

3 遺伝子多重化の効果をシミュレートする試み

ところが、近年発展の著しいCell Illustrator (CI) [4]などのパスウェイシミュレータを用い、これを思考支援ツールとして活用することで、上のような課題によりやくアタックできる可能性が出てきた。筆者らもCIを用いることで、重複による遺伝子の多重化がシステムの挙動に与える効果をシミュレートし、その結果に基づいて、魚類のイトヨの生態進化に関する興味深い考察を行うことができた[5]。

筆者らの比較ゲノム解析から、脊椎動物の嗅覚シグナル伝達経路に関する遺伝子群[6]のうち、フォスフォジエステラーゼ1C (PDE1C) 遺伝子がイトヨでのみ高度に重複し多重化していることが判明した(図1) [7]。イトヨは、営業行動やなわばり行動をとりたいへんに興味深い魚であるので、筆者らはイトヨでのみ見られるPDE1Cの多重化に興味を惹かれ、そ

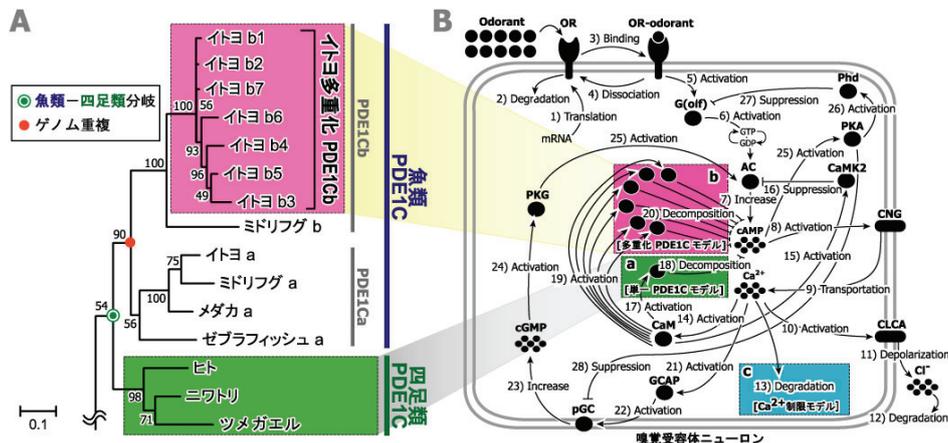


図1. 脊椎動物の嗅覚シグナル伝達パスウェイに関するフォスフォジエステラーゼ1C遺伝子の進化系統樹(パネルA)と、解析に用いた嗅覚シグナル伝達パスウェイ・モデルの模式図(パネルB)

の多重化が嗅覚シグナル伝達経路の挙動に及ぼす影響を、CIを用いて検討した。

シミュレーションの結果、PDE1Cの多重化は嗅覚シグナル伝達経路が出力する脱分極パルスを延長させる効果をもつことが示唆された(図2)。図2Aのシミュレーションモデルは、イトヨの多重化PDE1Cと嗅覚型Gタンパクの共発現解析結果に基づいて構築した。図2Bに示した各条件下でのシミュレーション結果の検討から、増加したPDE1C産物が上流のカルモジュリンを巡って互いに競合するために、上記の延長が起きることが示唆された。筆者らにとって予想もしなかった効果である。

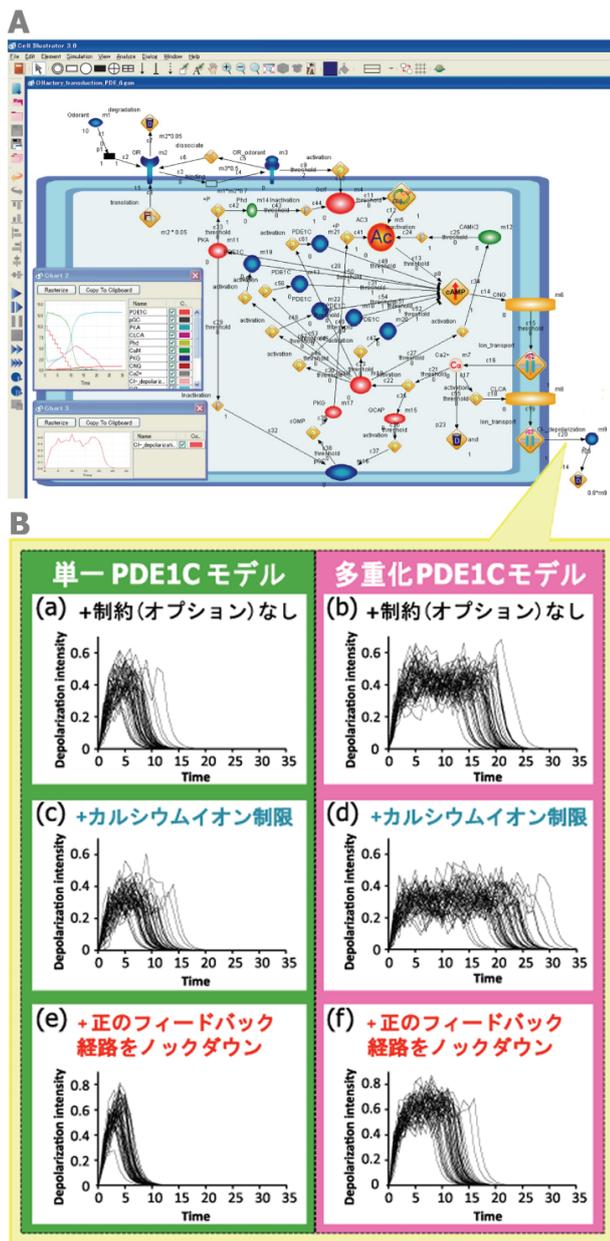


図2. Cell Illustratorで作成した嗅覚シグナル伝達パスウェイの動的モデル(パネルA)、および、PDE1C遺伝子が単一の場合(ヒトやマダカなど)と多重化した場合(イトヨ)とでシミュレートされる脱分極パルスの違い(パネルB)。

さてこの嗅覚シグナル延長の生態的意味であるが、行動学研究的蓄積があるマウスでの知見から類推すると、なわばり行動の進化と関連している可能性がある。ハツカネズミの尿

中に含まれるタンパク質も匂いシグナルを延長する効果をもっており、これによって侵入者がなわばり雄となわばりマークを区別しにくくなり、その結果なわばり周辺に近づくのを避けるようになる。このことが闘争による死亡率を減少させるので、選択的に有利な形質として固定するという説である[8]。イトヨもなわばりをもつ習性があるので、上のマウスの仮説は興味深く、イトヨでもなわばり行動の進化とPDE1Cの多重化が関連していることが想像される。この可能性は行動学実験などによって検証することができるだろう。

4 表現型進化の理解へ：パスウェイ・シミュレーションによる思考支援と仮説創出

以上のように、CIなどのパスウェイシミュレータを用いることで、興味をもった生物の遺伝情報に基づいて、その生物の行動特性や生態特性といった表現型のレベルの進化に関する仮説を抽出できる可能性が示された。今回紹介した研究例では、PDEという一見「ありふれた」酵素遺伝子の重複が、システムのレベルでは予想もしなかった効果をもたらすことが示唆されている。生命システムは、人間が思いもしなかったようなやり方で、様々な適応を実現しているのかもしれない。またこのことは、複雑なシステム動態の中での遺伝子重複の意味などを人間が解釈する上で、パスウェイシミュレータが非常に強力な思考支援ツールとなることを示唆している。シミュレーションによる仮説創出と、仮説の経験的な検証を研究の両輪とすることで、複雑な生命システムや表現型進化の理解が、大いに推進されていくことだろう。

◆ 参考文献 ◆

- [1] Hubbard, T. J. et al.: Ensembl 2009. *Nucleic Acids Res.* 37: D690–D697, 2009.
- [2] Ohno, S.: *Evolution by gene duplication*. Springer-Verlag, New York, 1970.
- [3] Conant, G. C. and Wolfe, K. H.: Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9: 938–950, 2008.
- [4] Nagasaki, M. et al.: *Foundations of systems biology: using Cell Illustrator and pathway databases*. Springer-Verlag, New York, 2009.
- [5] Sato, Y. et al.: Evolution of multiple phosphodiesterase isoforms in stickleback involved in cAMP signal transduction pathway. *BMC Syst. Biol.* 3: 23, 2009.
- [6] Kanehisa, M. et al.: The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32: D277–D280, 2004.
- [7] Sato, Y. et al.: Temporal pattern of loss/persistence of duplicate genes involved in signal transduction and metabolic pathways after teleost-specific genome duplication. *BMC Evol. Biol.* 9: 127, 2009.
- [8] Hurst, J. L. et al.: Proteins in urine scent marks of male house mice extend the longevity of olfactory signals. *Anim. Behav.* 55: 1289–1297, 1998.

Cell Illustrator を用いた細胞システム維持に影響する生命パスウェイ重要部位の推定

大島商船高等専門学校 商船学科

講師 北風 裕教

1 はじめに

生物が生存していくための機能は細胞によって成り立ち、細胞の働きによって生命はその活動を行うことができる。たとえば概日リズムやアポトーシス、細胞分裂などが細胞機能として考えられ、これらの機能を含めて細胞の活動は、すべて細胞内にあるタンパク質やRNAなど生体分子の複雑な相互作用によって生命活動が維持されている。細胞内の相互作用は生物学者によって盛んに調べられており、様々な相互作用の反応経路（生命パスウェイ）の詳細が現在明らかとされつつある。ハイブリッド関数ペトリネット(HFPN)記述で生命パスウェイのモデリングを可能にするツールCell Illustrator (CI)を利用すれば、細胞内の生体分子の振る舞いをシミュレーションにより実現することができ、実際多くの分子機構がこれまでに確認されている[1]。

我々の研究室ではこれらHFPNによって描かれた生命パスウェイから細胞システム維持に影響を及ぼす重要な経路（生命パスウェイ重要部位）を自動で推定するための手法について調べている[2]。ここでは、生命パスウェイ重要部位についてマウスの概日リズムを例にとり解説した後、HFPNを用いた重要部位の推定法について説明する。この手法は重要部位の有無を正しく行うことができる半面、ユーザが重要部位の判定を行うための推定時間が膨大になり、生命パスウェイのサイズが大きい場合には推定が不可能となる問題がある。また、ユーザによる知識の違いによって判定結果が異なるという問題もある。そこで、従来手法で難しかったこれらの問題に対して、現在我々が取り組む手法で、推定率は低下するが生命パスウェイのサイズが大きくなった場合においても自動で連続的に推定を可能にする新しい推定法について解説する。

2 生命パスウェイ重要部位

2005年に松野らによって報告されたマウスの概日リズムのモデル[3]を例にとり、生体分子の相互作用を考える(図1)。

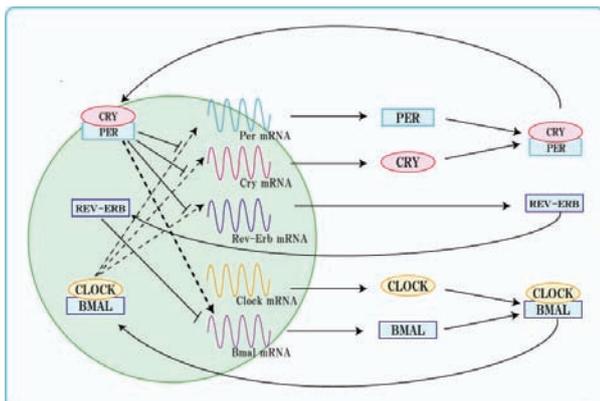


図1. マウスの概日リズムに関わる遺伝子の制御関係

Bmal遺伝子とClock遺伝子から生成されるBMALタンパクとCLOCKタンパクはBMAL/CLOCK複合体を形成し、核内に戻りPer遺伝子・Cry遺伝子・Rev-Erb遺伝子を促進させる。促進されたPer遺伝子とCry遺伝子はPERタンパクとCRYタンパクを生成し、それらがPER/CRY複合体を形成して核内に戻りPer遺伝子・Cry遺伝子・Rev-Erb遺伝子への促進を抑制する。また、Bmal遺伝子を促進させる。一方、Rev-Erb遺伝子はREV-ERBタンパクを生成し核内に戻ってBmal遺伝子を抑制する。このように核内でお互い制御し、関係しあうことで、概日リズム（周期反応）が生み出されている。HFPNを用いてこれらの生命パスウェイを記述したものが図2であり、そのリズムを示したものが図3-①である。

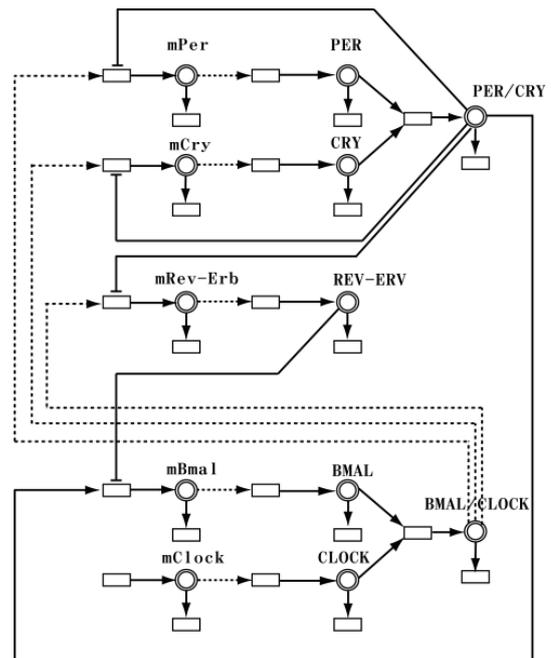
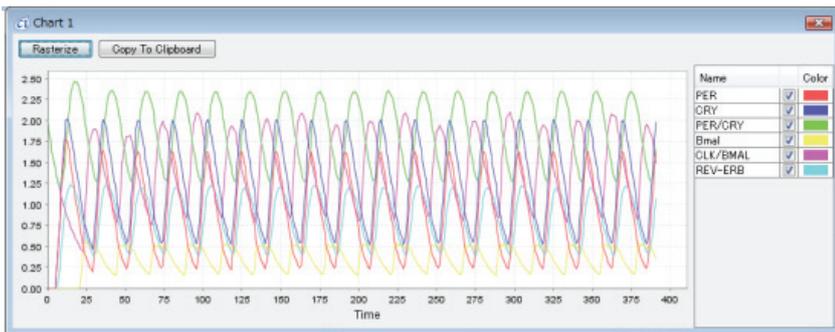


図2. マウスの概日リズムをHFPNを用いて記述

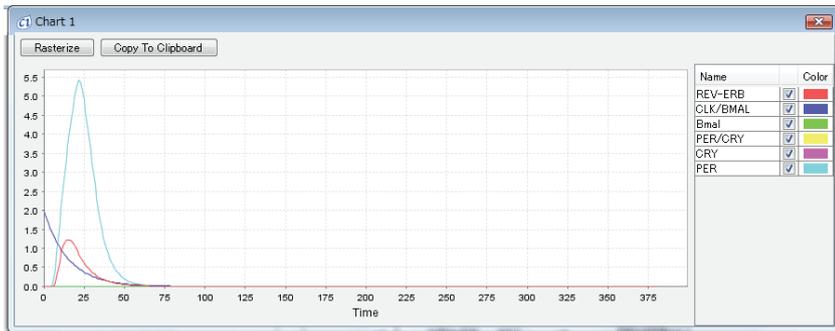
これらによって24時間の周期反応が起こることが知られているが、例えばCry遺伝子をロックアウトすると全体の24時間のリズムが失われることが分かっている。これをHFPNを用いて記述する場合、対応する遺伝子の経路を削除することで実現できる。HFPNのシミュレーション結果を図3-②に示す。この結果は、生物実験に一致する。一方、Rev-Erb遺伝子をロックアウトした場合は、24時間のリズムがそのまま保たれることが知られている。HFPNの生命パスウェイにおけるシミュレーション結果を図3-③に示す。この結果は、生物実験に一致している。ここでいう概日リズムにおける細胞の恒常性とは周期の維持であるので、Rev-Erbをロックアウトしても恒常性は保たれているとここでは判断する。すなわち、ある細

胞システムの恒常性を保つという観点においては、Cry遺伝子からCRYタンパクの生成の経路は重要であるが、Rev-Erb遺伝子からREV-ERBタンパクの生成の経路はそれほど重要ではないと考えられる。つまり、概日リズムという細胞システムの恒常性を、Cry遺伝子からCRYタンパク生成の経路が崩れてしまうことになるので、我々はこの経路のことを生命パスウェイ重要部位と言う。ここで、生命パスウェイというのは概日リズム全体も生命パスウェイと考えることができるほか、アポトーシスや細胞分裂、さらに他のどのような細胞内の機能についてもそれを生み出す分子のネットワークについて、すべて生命パスウェイと呼ぶこととする。

①



②



③

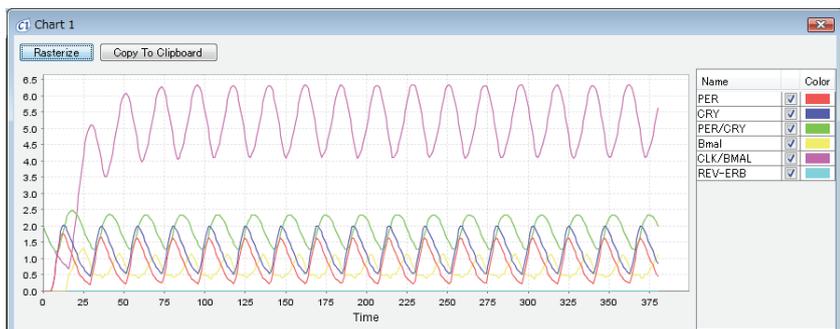


図3. マウスの概日リズムの結果

3 HFPNによる重要部位の従来の推定手法と問題点

HFPNを用いて重要部位を推定する従来の手法について図4に流れ図を示す。

従来手法では、ユーザがシミュレーションツール(CI)を利用してHFPN表記により生命パスウェイを作成した後、生物実験におけるロックアウトの作業をCI画面上でアークの削除により実現する。この作業はCIにおいて手作業で行い、続けてHFPNの実行処理を行う。HFPNの実行は計算機によって自動的に行われ、トークンの時間変化がグラフとして出力さ

れる。この出力結果からユーザは目視により重要部位の判定を行う。ここでの判定は、ユーザのこれまでの経験による判断によって決定される。この作業を全てのアークにおいて繰り返し行うことにより、生命パスウェイ全体における重要部位を確認することができる。

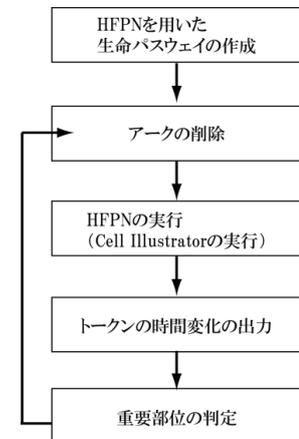


図4. 重要部位推定法の従来手法の流れ図

マウスの概日リズムのモデルにおいて実行した結果を図5に示す。赤色のアークが重要部位に相当し、青色のアークが重要ではない部位に相当する。この結果から、Rev-Erbはそれほど重要ではない経路であることが確認できる。

この手法はネットワーク全体の中から重要部位を確認することができるが、一連の作業に要するシミュレーション時間と、重要部位の確認に要する時間は生命パスウェイが大きくなるに伴い増加するため、逐次実行を繰り返す必要のない自動推定法が求められる。

そこで本研究では次に示すRNNを用いた重要部位の自動推定手法を提案し、これらの問題点の完全に努めた。

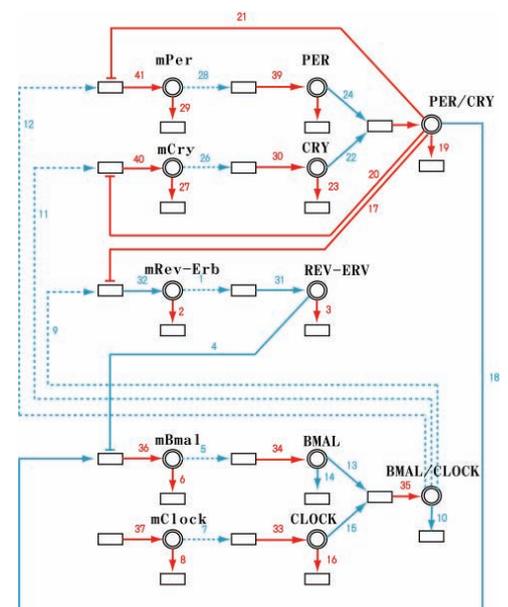


図5. HFPNを用いて確認したマウスの概日リズムの結果

4 RNNを用いた重要部位の自動推定手法への改良

多くの処理時間を要するHFPNから学習によるシステム形成後には想起処理によりそれ程の処理時間を要さないニューラルネットワーク(NN)を用いて重要部位を推定する。HFPNではトークンの時間変化情報が重要部位の判定に必要となる。そこで、NNには時間が記憶可能なリカレントニューラルネットワーク(RNN) [4]を用いる。HFPNからRNNへネットワークの変換を行い、通時的逆伝播法(BPTT法) [4]の学習により自動で連続的に重要部位を推定するシステムとする。この手法はネットワーク移行のためにRNNの学習時間を要するが、連続的にアークの削除が行えるため推定結果を得るまでの時間を改善できる。

図6にRNNを用いた重要部位の推定法の流れ図を示す。この手法は、第1ステージから第4ステージからなり、全て自動的に行うことを可能にする。詳細について以下に示す。

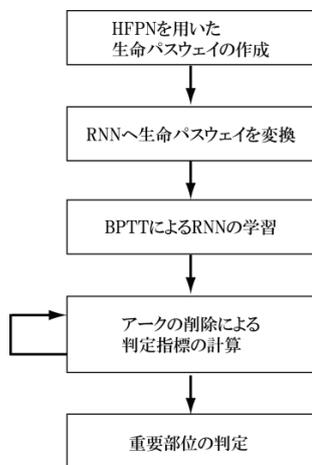


図6. 重要部位推定手法の提案手法の流れ図

4.1 第1ステージ：HFPNからRNNへのネットワーク変換

第1ステージではHFPNからRNNへのネットワークの変換を行う。HFPNによりモデル化した生命パスウェイをRNNに変換する際、ネットワークの連結情報を保持する必要がある。そこで、HFPNにおけるプレース、トランジション、アーク、トークンは次に示すルールで置き換える(図7)。

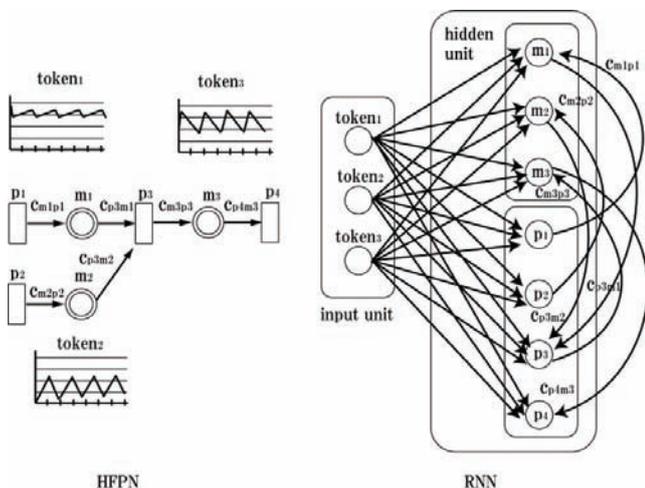


図7. HFPNからRNNへネットワークの変換

- (1) HFPNで抽出したプレースから得られる時系列のトークン値のデータは、RNNの入力ユニットへの入力信号とする。
- (2) HFPNのプレースとトランジションはRNNの隠れユニットとして定義する。(隠れユニットと出力ユニットは区別を行わない。)
- (3) HFPNのプレースからトランジションまたはトランジションからプレースへ結合するアークを、RNNの出力ユニット(隠れユニット)間のネットワーク結合とする。

4.2 第2ステージ：RNNの学習

第1ステージでは、HFPNの連結情報をRNNの連結情報として再定義したが、この段階ではHFPNにおけるトークン値の時系列データをRNNには形成できてはいない。第2ステージでは、トークン値情報をRNNに導入すると同時に重要部位を自動推定できるシステムを構築する。RNNの学習には誤差逆伝播法(BPTT法)を用いる。ここで教師信号として、次の二つのパターンを学習する。

【パターン1】HFPNにおいて生命パスウェイが破綻していない場合(アークを削除していない場合)の入出力関係についてRNNに記憶させる。入力信号はプレース内の濃度量の変化であるトークンの内容で、それに対応する教師信号は全て0とする。

【パターン2】HFPN生命パスウェイが破綻した場合(重要なアークを削除した場合)を想定して全ての反応が停止した状態を記憶させる。入力信号は破綻した状態を想定して全て0で、教師信号は全て1とする。

4.3 第3ステージ：アーク削除による誤差の推定

HFPNでアークを削除して重要部位が否か判定した場合と同様に、RNNのネットワークにおいても、ネットワークの結合を1本ずつ削除することで重要部位を判定する。ここでは、学習後のRNNにおいて結合を削除し、アークの削除処理を施していないRNNの出力信号(標準値)とアークの削除処理を施したRNNの出力信号(判定指標)を比較することで、生命パスウェイにおける経路の重要性を判定する。もし、標準値と判定指標の差が小さい場合は、アークを削除しても影響が小さいことから、対象としたアークは生命パスウェイにおいて重要部位ではなく、他の経路でこの経路の信号を捕うことができるような経路と考えられる。それに対し、標準値と判定指標の差が大きくなることから、対象としたアークは生命パスウェイにおいて重要であるといえる。したがって、他のパスウェイでこの反応を捕うことができない重要部位であると考えられる。この処理を全てのアークにおいて順に行うことで、生命パスウェイ全体の重要部位を判定することが可能になる。

4.4 第4ステージ：重要部位の連続判定

第3ステージで得られた判定指標をソート処理により大きい順に並べ替えた結果を図8に示す。横軸は、ソート後に定義しなおしたアークの番号であり、縦軸は判定指標の値である。ここで、1からP点までは、アークの削除の影響で判定指標の変動が存在するのに対して、P点以降はアークの削除の影響が

あるにも関わらず、判定指標の値が少ない状態で一定値となる。このP点を重要部位と重要ではない部位の判別値として定義する。これにより、各々のアークが重要部位であるか判定することができる。

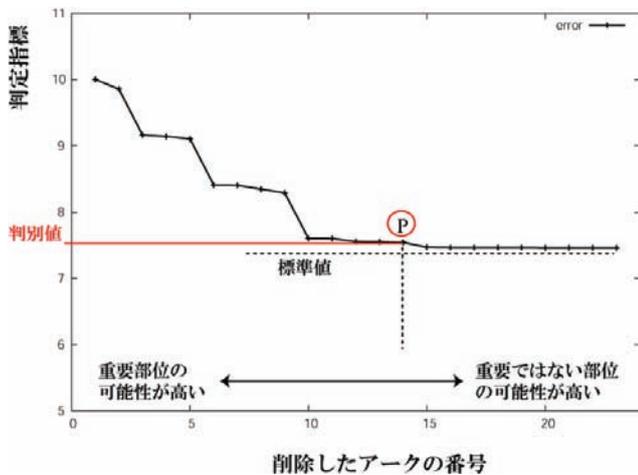


図8. 重要部位と重要ではない部位の判別

5 生命モデルを用いた反応推定実験

HFPNを用いて概日リズムのモデルを描き、推定実験を行った。図5のHFPNの推定手法に対して図9がRNNを用いて自動推定した結果である。赤色のアークが重要部位と推定した結果であり、青色のアークが重要ではない部位と推定した結果である。赤丸で示したアークが自動推定において誤って推定したアークである。結果をまとめたものを表1に示す。この結果から、自動推定では約83%の的中率を示すことができた。

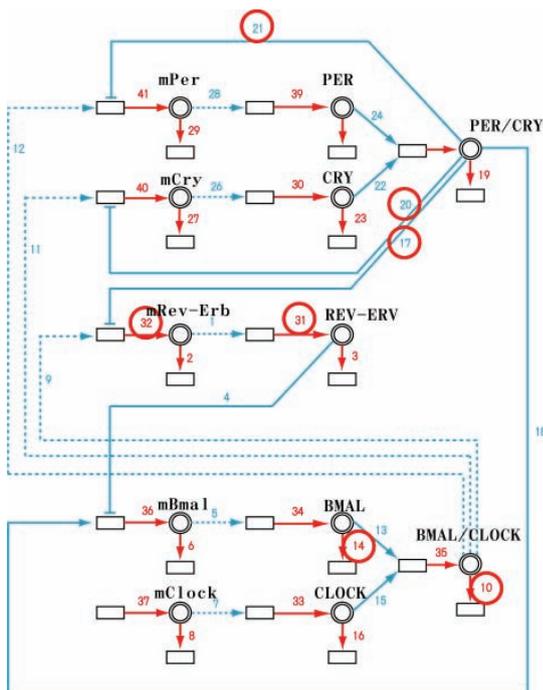


図9. RNNを用いた概日リズムの重要部位の推定結果

表1. RNNによる重要部位の推定結果

	重要部位	重要ではない部位	合計
HFPNの基準指標数	23	18	41
RNN手法の的中率	87.0% (20)	77.8% (14)	82.9% (34)
RNN手法の誤り率	13.0% (3)	22.2% (4)	17.1% (7)

6 おわりに

HFPNを用いて描かれた生命パスウェイから、生命パスウェイ重要部位を推定する手法について示した。現在行っているRNNを用いて推定する手法は推定率が低下する問題があるため、正確な推定率を得ることができない。一方、従来の手法はシミュレーションの実行時間の問題と、人による判定時間の上昇による問題があった。今後、膨大な生命パスウェイから全ての経路の重要部位を推定するためには、人による推定の代わりにコンピュータによる自動推定が必要となる。人による判定部を自動化でき、かつシミュレーション時間を短縮することができれば、RNNの手法を用いるよりも従来の手法の方が正確な推定結果が得られる。本研究ではHFPNのモデルはCI3.0を用いて記述している。CI Online 4.0を用いれば、インターネット経由で常に最新の機能をクライアントが利用でき、さらに高速シミュレーションモジュールが利用できることから10倍から100倍の速度で実行が可能となり、シミュレーションの実行時間についての問題は解決することができる。これらのことから、いかに知識の高い生物学者のレベルでロックアウトによる影響から重要部位の判定を行うことができるかが、今後の大きな課題となるといえる。

謝辞

山口大学大学院理工学研究科の松野浩嗣教授には有益なコメントを頂きました。ここに感謝致します。

◆ 参考文献 ◆

- [1] A. Doi, M. Nagasaki, H. Matsuno, and S. Miyano, "Simulation-based validation of the p53 transcriptional activity with hybrid functional Petri net," In *Silico Biology*, Vol 6, pp.1-13, 2006.
- [2] 北風裕教、中司弘樹、神田全啓、池田信彦、松野浩嗣、宮野悟、リカレントニューラルネットワークによる細胞内反応経路の脆弱・頑健部位の推定手法、電子情報通信学会、Vol.91-D, No.9, pp.2401-2417, 2008.
- [3] H. Matsuno, S. Inouye, Y. Okitsu, Y. Fujii, and S. Miyano, A new regulatory interactions suggested by simulations for circadian genetic control mechanism in mammals, *Proc. The 3rd Asia-Pacific Bioinformatics Conference*, pp.171-180, Imperial College Press, 2005.
- [4] R. Williams, and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, Vol.1, pp.270-280, 1989.

パスウェイ描画解析ソフト シー・イノベーション株式会社

“Cell Illustrator【セルイラストレータ】”

セルイラストレータは、生命をシステムとして理解するというゲノム解読後の生命科学の新たなチャレンジの実現を目指した、パスウェイ描画解析ソフトウェアです。

【機能概要】

- 簡単かつ直感的に生物パスウェイの描画ができる。
生命システムを構成するパスウェイ（代謝経路、遺伝子制御ネットワーク、シグナル伝達経路、細胞間の制御反応など）を描くために必要なアイコンを 350 個以上備えており、これらをドラッグアンドドロップし、コネクタで繋いでいくことで簡単にパスウェイを作成できます。さらに自分でオリジナルのアイコンを作成することもできます。また、これらアイコンには生物学の用語（オントロジー）情報が入っており、情報の整理・共有・再利用に便利です。
- 作成したパスウェイはただちにシミュレーションができる。
作成したパスウェイは、再生ボタンを押せばすぐに簡単なシミュレーションが可能です。仮説の生成・検証をする、遺伝子をロックアウトした影響を予測する、予備実験をするといったことを簡単な操作で実現できます。さらに、数式を入力することで複雑なシミュレーションもできます。
- 遺伝子ネットワークの探索と解析ができる。
マイクロアレイ解析で得られた遺伝子ネットワークを表示、特定の経路を検索、サブネットワークの作成などができます。この機能は、実際にジエヌアイの遺伝子ネットワーク解析に用いられています。

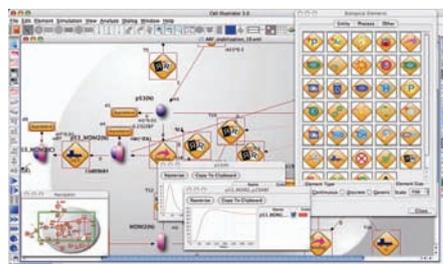


図1. ネットワーク描画例

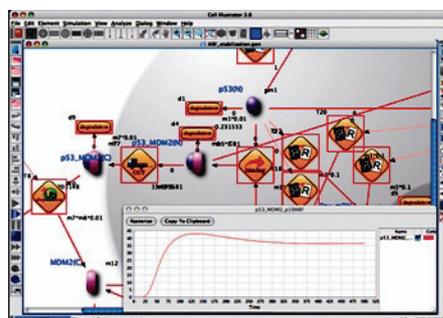


図2. シミュレーション画面例

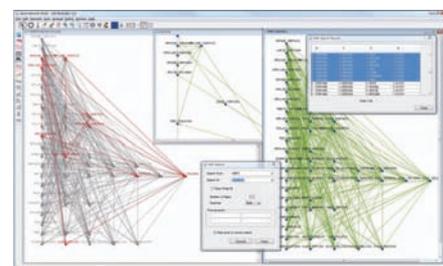


図3. 遺伝子ネットワーク表示例

【特長】

- 最新の CSML フォーマットの入出力形式 CSML3.0 に対応
- あらゆる OS 環境（Windows、Mac OS X、Unix、Linux）に対応
- 350 個以上のオントロジーと関連付けられた、ベクター形式（SVG）の洗練されたアイコン
- 他のパスウェイモデル（SBML、CellML 形式）をインポート可能
- BIOBASE 社の TRANSFAC（遺伝子制御データベース）、TRANSPATH（パスウェイデータベース）に含まれる、ヒト・マウス・ラットの 10 万以上の生体内反応をインポート可能（オプション）
- 作成したパスウェイは直ちにシミュレーションが可能

コード No.	メーカーコード	品名	容量	希望納入価格(円)*	備考
306-33381	GS-CIPC01J	Cell Illustrator Professional Corporate Edition セルイラストレータ プロフェッショナル コーポレート版	1 セット	600,000	プロフェッショナル・ユーザー向け
303-33391	GS-CISC01J	Cell Illustrator Standard Corporate Edition セルイラストレータ スタンダード コーポレート版	1 セット	200,000	一般ユーザー向け
306-33401	GS-CIPA01J	Cell Illustrator Professional Academic Edition セルイラストレータ プロフェッショナル アカデミック版	1 セット	150,000	教育機関のプロフェッショナル・ユーザー向け
303-33411	GS-CISA01J	Cell Illustrator Standard Academic Edition セルイラストレータ スタンダード アカデミック版	1 セット	50,000	教育機関の一般ユーザー向け
300-33421	GS-CISS01J	Cell Illustrator Standard Student Edition セルイラストレータ スタンダード 学生版	1 セット	12,000	学生向け
307-33431	GS-CICA01J	Cell Illustrator Classroom Single Pack セルイラストレータ クラスルーム 1 ライセンス	1 セット	50,000	教育機関向けバック製品
304-33441	GS-CICA10J	Cell Illustrator Classroom 10 License Pack セルイラストレータ クラスルーム 10 ライセンス	1 セット	250,000	教育機関向けバック製品
301-33451	GS-CICA50J	Cell Illustrator Standard Student Edition セルイラストレータ クラスルーム 50 ライセンス	1 セット	1,250,000	教育機関向けバック製品

*：年間のライセンス料となります。

次世代シーケンサーデータ処理ツール “MetaGenomeGAMBLER”のご紹介

インシリコバイオロジー株式会社

代表取締役 大山 彰

次世代シーケンサーからの配列出力

次世代シーケンサーの急速な導入に伴い、出力される膨大な塩基配列情報処理の問題が注目されています。従来のキャピラリーシーケンサーと比較して、数万倍にもおよぶ配列出力スループット能力はこれからのゲノム解析の流れをすっかり変えてしまう可能性があります。しかるに、これら膨大な配列情報のソフトウェア側の処理能力については、シーケンサーの開発に追いついていないのが現状です。

MetaGenomeGAMBLER

インシリコバイオロジー社では、数年前からこの課題に取り組み、SOLiD、Solexa、454などの次世代シーケンサーからの配列データを参照ゲノム塩基配列上にマッピングし、マッピングされた大量の短塩基長（フラグメント）配列と参照ゲノムのアラインメントをグラフィカルに表示するソフトウェアを開発し、販売しています。このソフトウェアMetaGenomeGAMBLERは、当初はキャピラリーシーケンサー出力配列のアセンブルプロジェクト支援用に開発され、メタゲノム解析もできるように改良されていたソフトウェアですが、この上にさらに、次世代シーケンサーの出力配列をマッピングする機能を付加したものです。したがって、従来型のシーケンサー配列のアセンブル処理およびその結果の閲覧と同一の仕組みで次世代シーケンサー配列を扱える利点があります。

Mac Note, Win Noteでも動作

利用者は参照ゲノム塩基配列ファイルと、次世代シーケンサー配列の格納場所を指定するだけで直ちにマッピングを実行することが可能です。数千万件にもおよぶ大量の配列であっても、通常のMacやWindows PCで動作します。CPUやメモリーに余裕があれば、処理の速度や画面スクロールの速度は向上しますが、ノートPC程度の機器でも動作するため、手軽に次世代シーケンサーデータ解析ができることとなります。次世代シーケンサーデータ解析をノートPC上で実行できることは、より多くの研究者や学生が直接データを解析し、それぞれの担当分野や研究目的に応じて、並行に研究を進められる利点につながり、膨大なデータをより効率的に解析することが可能となります。ただし、ノートPC等での実行をスムーズに行うために、参照ゲノム塩基配列を利用者が指定する塩基長で等間隔に分断し、その分断された参照ゲノム部分ごとにマッピングを行っています。

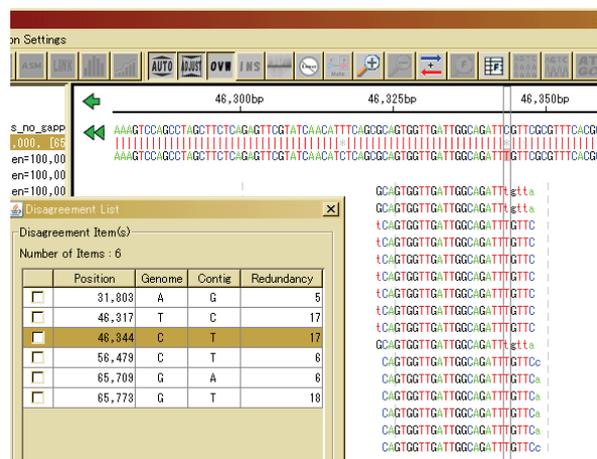


図1. 参照ゲノムとの変異箇所の一覧

Mac OS X用とWindows xp用を同時に改良・更新

MetaGenomeGAMBLERはJava言語で書かれているため、OS間の移植性が高いという特徴があります。現在まで数百回におよぶ改良版をリリースしていますが、いずれのリリースもMac OS X用とWindows xp用を同時に公開しています。

FastQ形式など品質データが付随する塩基配列である場合は、ある一定以上の品質をもつ配列のみを解析対象とすることができます。また、配列のうちの一部をトリミングする処理も可能です。さらに、最初から全件をマッピングするだけでなく、何件目から何件目までと全塩基配列のうち一部だけをマッピングすることも可能となっています。

変異部位の検出とその塩基組成

マッピングされた結果は、参照ゲノムとマップされた各配列とのマルチプルアラインメント形式で表示され、高速でスクロールやズームが可能です。参照ゲノム塩基配列とマッピングされたフラグメントのコンセンサス配列が不一致の部位は、重なっているフラグメント数に占める各ベースコールの割合を表示し、SNPsなどの解析に使用できます。

ギャップクローズ方式

マッピングできなかった領域は「ギャップ」として定義され、その部位を直接表示およびCSV形式でファイル出力可能です。この様なギャップの発生には、3つの原因が考えられます。ひとつはこの参照ゲノム領域に対応するフラグメント配列が存在しない場合、次は完全に異なる配列に置換されてしまった場合、最後は参照ゲノムと比較して変異が多く、マッピング対象から除外されてしまった場合です。後者2つの場合は、異なる方法で局所的にマッピングを行うことにより、そ

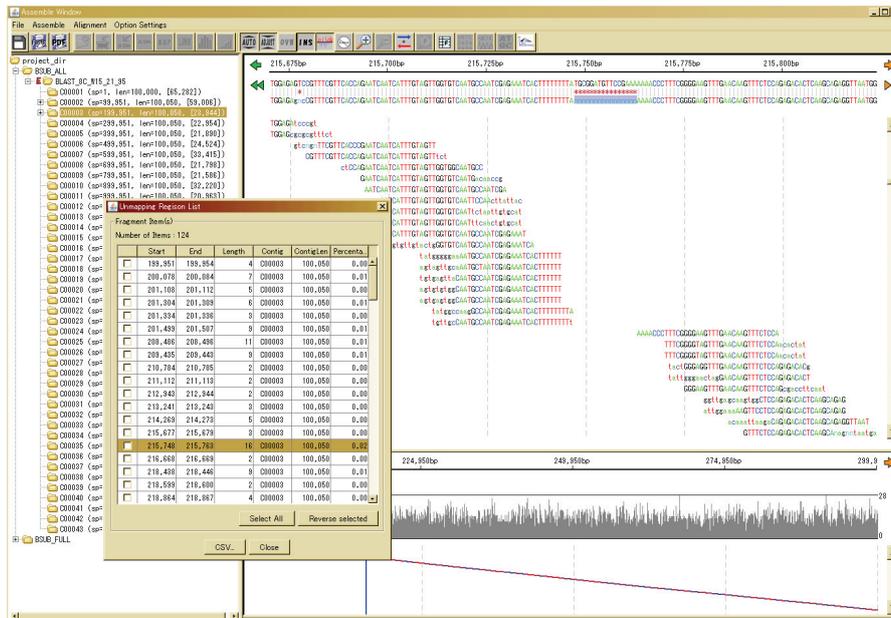


図2. ギャップ位置の検出とアラインメント表示

のギャップを埋めることができる可能性があります。MetaGenomeGAMBLERではこの2つの場合に対応して、2種類のギャップクローズ機能を提供しています。

局所再マッピングによるギャップクローズ

1つ目は、ギャップを含む参照ゲノム領域を切り出し「局所参照ゲノム配列」とします。一回目のマッピングで参照ゲノムに貼りつかなかったフラグメント配列全体をもう一度、今度は「局所参照ゲノム配列」に対して、判定条件を少し緩めて再マッピングします。その結果得られる「局所マッピング結果」は単独で、グラフィカルに閲覧することができます。さらに、「局所マッピング結果」を最初の全体のマッピング結果のギャップ領域部分に挿入することにより、「ギャップクローズ」することが可能となっています。

フラグメントウォーキングによるギャップクローズ

もうひとつは、ギャップ領域の両側に存在するマップできたフラグメント配列から「アンカーフラグメント」をそれぞれひとつずつ指定し、そのフラグメントからギャップ方向に「ウォーキング」していく方法です。まず、左右のアンカーフラグメント配列を問い合わせ配列として、先のマッピングで貼りつかなかったフラグメント配列全体を検索し、相同性部分をつなぎ合わせていく、いわゆる「アセンブル手法」を用います。両側からフラグメントをつないでいき、途中で1つの共通フラグメントにたどり着けば、ギャップクローズが成功することとなります。この方法では、ギャップ部分が参照ゲノム配列から完全に異なる配列に置換されている場合に有効です。

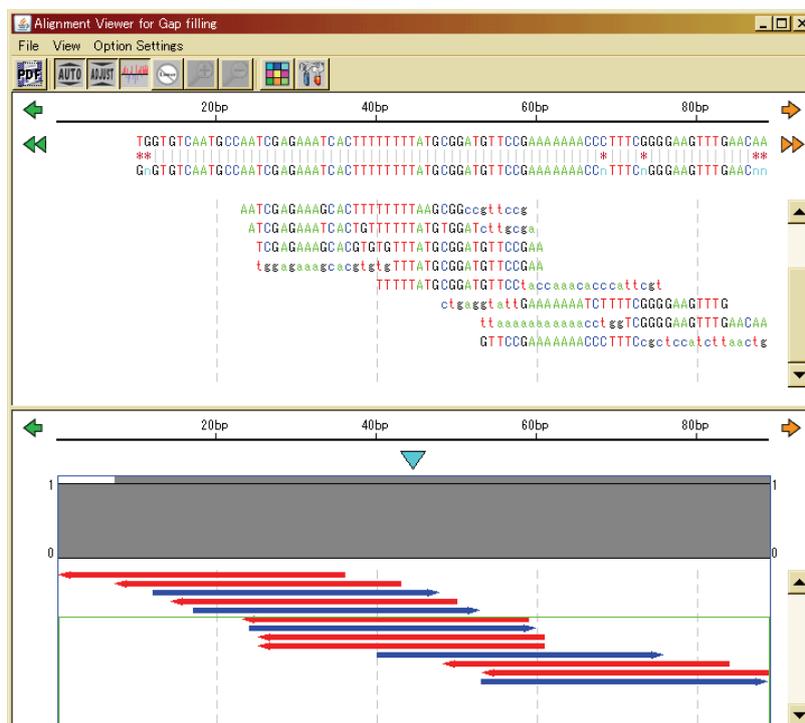


図3. 局所再マッピングと局所アラインメント

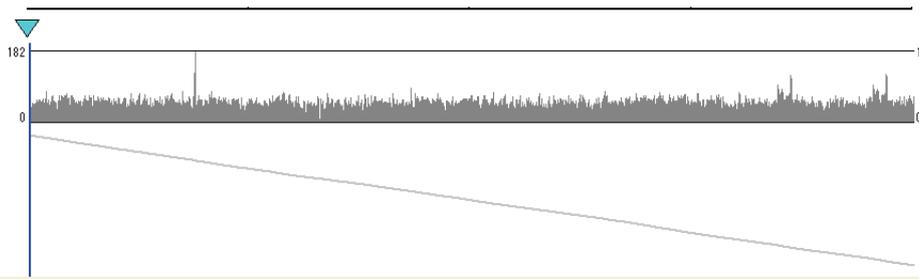


図4. マッピングDepthプロファイル表示 1

ギャップとして顕在しない置換・挿入の検出方法

単一のサンプルに変異がないフラグメントと、参照ゲノムに対して置換や挿入があるフラグメントが混在している場合は、ギャップが顕在的に発見できない恐れがあります。MetaGenomeGAMBLERは、顕在しないギャップ領域から参照ゲノム配列に対する置換・挿入配列を求める方法を実装しています。指定数以上の数のフラグメントがマッピングされている領域で、マッピングされたフラグメント配列の中にある非アラインメントセグメントを抽出し、非アラインメントセグメントのコンセンサス配列とアラインメントコンセンサス配列の一致度数を比較する方法です。この方法により置換・挿入が検出された領域から参照ゲノムにアラインメントされたフラグメント群を除去することにより、置換・挿入を明示的に表示することができます。

マッピングされたフラグメントのDepthプロファイル表示

参照ゲノム上に貼りついたフラグメントの厚さ (Depth) をプロファイルとして表示し、そのズームが可能で、さらに、Depthプロファイルは、インシリコバイオロジー社のタイリングアレイ解析ソフトウェアであるIMCAE (insilicoMolecularCloning Array Edition) にそのデータをエクスポートし、タイリングアレイプロファイルと並行表示が可能です。IMCAEは、その特徴であるフィーチャーのグラフィカルビューアとしての機能を使い、既存の注釈と次世代シーケンサやタイリングアレイのゲノム位置特異的な特徴を表示、印刷することが可能です。

SOLiD方式とSolexaなどその他の方式の取り扱い

MetaGenomeGAMBLERのマッピング機能は、Solexaや454などの配列データの他に、SOLiDによるカラースペースデータを直接取り込みマッピングに使用することができます。また、次世代シーケンサからの配列だけでなく、キャピラリーシーケンサからの波形データやマルチプルFastAデータ、注釈付きGenBank形式ファイルなども混在させてマッピングすることが可能です。ただし、SOLiDのカラースペースデータの場合は、配列データと混在させてマッピングすることは現在のところできません。SOLiDの場合は、SOLiD付属の計算機でマッピングした結果をGFFという形式でファイル出力することができますが、MetaGenomeGAMBLERでは、GFF形式ファイルを直接インポートし、そのマッピング結果をグラフィカルに表示することも可能です。

データの可搬性

次世代シーケンサからの出力ファイルはギガバイト級のサイズをもつことになり、そのコピーにかかる時間も無視できません。通常使用しているMacやPCに内蔵されているHDDに格納する方法は、結果の転送にも時間がかかるためあまり得策ではありません。通常は、ポータブルHDDに一連の配列と結果を格納し、その可搬性を高めています。MetaGenomeGAMBLERでは、実行パラメータなどをプログラムが存在する内蔵ディスク上ではなく、HDD上のプロジェクト情報として格納することができます。この結果、HDDを別のMacやPCに接続し、そこでMetaGenomeGAMBLERを実行し、結果を参照してもまったく同様のパラメータ指定で操作、閲覧することが可能です。

今後の改良と新規開発中製品

インシリコバイオロジー社開発製品は、研究者からの要望を迅速に取り入れ、その改良を反映したVersionを非常に短いインターバルでリリースしています。まだ解析方法が確定していない次世代シーケンサやタイリングアレイなどの領域では、短期間に要望を実現できる仕組みが大変重要であり、今後もこの体制で改良開発が進められる予定です。

なお、インシリコバイオロジー社では、現在次世代シーケンサ配列のマッピング機能とIMC機能をドッキングした新規ソフトウェア製品であるGenomeTravellerを開発中です。また、マッピングだけでなく、次世代シーケンサ配列のDeNovoアセンブル機能をもつソフトウェアも開発しています。近日中に、その製品版がリリースされる予定です。



次世代シーケンサーデータ処理ツール インシリコバイオロジー株式会社

“MetaGenomeGAMBLER”

MGG Edition 別必要ソフトウェア環境

	MGGL	MGGP	in silico Assembler
Mac OS	Mac OS X.4 以上	Mac OS X.4 以上	Mac OS X.4 以上
Windows	Windows xp	Windows xp	Windows xp
Java for Mac	Version 5 以上	Version 5 以上	Version 5 以上
Java for Wind	Version 5 以上	Version 5 以上	Version 5 以上
SHK Driver for Mac			
SHK Driver for Win			

MGG ハードウェア環境

	MGGL	MGGLW	MGGPW
必要メモリ容量 (実装)	512MB 以上	512MB 以上	1024MB 以上
ディスク容量 (空き容量)	5GB 以上	5GB 以上	10GB 以上
	対象とするデータに依存しますが、作業領域などを大きく使いますので、明示的に利用する容量の2倍程度のディスク容量を確保してください。		
ディスプレイ サイズ	MGGL	MGGLW	MGGPW
推奨サイズ	1000W×800H pixel	1000W×800H pixel	1000W×800H pixel
最小サイズ	320W×320H pixel	320W×320H pixel	320W×320H pixel
	現在のバージョンでは MGG の初期画面サイズはディスプレイサイズに比例して、自動的に設定されますが、リサイズ可能です。		
周辺機器：マウス	2ボタンマウスを推奨 (MAC1 ボタンマウスでも操作可能) MGG は 2 ボタンマウスの右ボタンクリックを活用しています。MGG を快適に利用するためには、Mac の場合も 2 ボタンマウスを使うことをお勧めしますが、1 ボタンマウスでもアップルキーを押しながらクリックすることで 2 ボタンマウスと同等の機能を発揮できます。		
周辺機器：プリンター	印刷する場合は必要です。		
ネットワーク環境	ブロードバンド接続環境を推奨 常時使う必要はありません。NET ボタンを使う場合は必要です。また、塩基配列やアミノ酸データベースをダウンロードしたり、MGG の最新バージョンをダウンロード、インストールする場合があると便利です。 一旦、データを取り込んだり、ソフトウェアを更新してしまえば、ネットワークから離れた環境でも充分快適に操作できます。		

品名	略名	容量	希望納入価格 (円)
MetaGenomeGAMBLER LITE Edition (一般)	MGGL	1 セット (ドングル)	220,000
MetaGenomeGAMBLER LITE Edition (アカデミック)		1 セット (ドングル)	170,000
MetaGenomeGAMBLER LITE Edition 年間更新料		1 セット (ドングル)	55,000
MetaGenomeGAMBLER Professional Edition (一般)	MGGP	1 セット (ドングル)	320,000
MetaGenomeGAMBLER Professional Edition (アカデミック)		1 セット (ドングル)	240,000
MetaGenomeGAMBLER Professional Edition 年間更新料		1 セット (ドングル)	80,000
in silico Assembler (一般)	W	1 セット	160,000
in silico Assembler (アカデミック)		1 セット	120,000
in silico Assembler 年間更新料		1 セット	40,000

- 本文に記載しております試薬は、試験・研究の目的にのみ使用されるもので、「医療品」、「食品」、「家庭用品」などとして使用できません。
- 希望納入価格には消費税等が含まれておりません。

和光純薬工業株式会社

本社 ☎540-8605 大阪市中央区道修町三丁目1番2号 ☎(06) 6203-1788 (試薬学術部)
 東京支店 ☎103-0023 東京都中央区日本橋本町四丁目5番13号 ☎(03) 3270-8243 (試薬学術部)
 ●九州営業所 ☎(092) 622-1005(代) ●中国営業所 ☎(082) 285-6381(代)
 ●東海営業所 ☎(052) 772-0788(代) ●横浜営業所 ☎(045) 476-2061(代)
 ●筑波営業所 ☎(029) 858-2278(代) ●東北営業所 ☎(022) 222-3072(代)
 ●北海道営業所 ☎(011) 271-0285(代)

フリーダイヤル：0120-052-099 フリーファックス：0120-052-806

●Wako Chemicals USA, Inc. ●Wako Chemicals GmbH (Neuss)

http://www.wakousa.com http://www.wako-chemicals.de
 Head Office (Richmond, VA) Tel: +49-2131-311-0
 Tel: +1-804-714-1920
 Los Angeles Sales Office
 Tel: +1-949-679-1700
 Boston Sales Office
 Tel: +1-617-354-6772

■ご意見・お問合せ、本誌のDM新規登録・変更等については、
 E-mail : org@wako-chem.co.jp まで

URL : http://www.wako-chem.co.jp