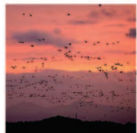


35年間の学際融合研究を振り返って



随 筆

松 田 秀 雄*

Reflections on 35 Years of Interdisciplinary Research

Key Words: Interdisciplinary research, Bioinformatics,

はじめに

筆者は1994年に大阪大学に助教授として着任して以来、32年間大阪大学でお世話になっている。本誌の「随筆」に執筆依頼をされた機会に、これまでの研究生活を振り返ってみたい。

大阪大学に赴任する前の前職は、神戸大学工学部システム工学科(現 システム情報学部)であった。当初は、修士課程と博士後期課程のテーマを継続して、並列計算機の研究開発を行っていた。転機が訪れたのは、1991年4月から10カ月間、米国アルゴンヌ国立研究所に客員研究員として滞在し研究する機会を得たことであった。

アルゴンヌ国立研究所は、米国エネルギー省傘下の研究所であり、1946年に原子力の平和利用を研究する目的で設立された。情報科学の分野では、並列計算を主体とするスパコンの研究開発で著名な実績があり、2024年には世界第2位、2026年7月現在においても第6位のスパコンであるAuroraを開発している。海外での研究先としてアルゴンヌ国立研究所を希望したのは、最先端のスパコンを利用した並列計算の技術を修得したいと思いからであった。

ところが、アルゴンヌ国立研究所で筆者を受け入れてくれたRoss Overbeek博士は、情報科学分野において著名な研究者であったが、筆者には情報科学ではなく生物学の研究をするように強く勧めた。

当時、筆者は認識できていなかったが、1990年

代はヒトゲノムプロジェクトが開始され、ゲノムの解読が急激に進展した時期であった。ヒトゲノムは30億塩基で構成されており、1文字1バイトとすると3Gバイトのデータ量となる。2026年現在ではそこまで大きな量とは感じないが、筆者がゲノム研究を始めた1991年当時は非常に大規模なデータであったため、その解析には並列計算に代表される高度な情報科学技術が必須と考えられ、生物学と情報科学の融合研究が求められていた。今から考えると非常に良いタイミングで学際融合研究をスタートすることができたわけであったが、当時は周りについていくのがやっとで大変苦労した。

アルゴンヌ国立研究所での研究生活

前述のように、並列計算機についての研究開発が専門であった筆者にとって、生物学、特にゲノムの研究は全くの専門外であった。滞在開始当初は、大腸菌のゲノムデータベースのプロジェクトに参加するように言われ、メンバーの研究打合せに出席したが、議論で出てくる用語が全くわからないので閉口した。geneとかであれば遺伝子というイメージがわくが、restriction siteとか何番のcloneがどうしたとかばかりで、何のことかさっぱり理解できなかった。実は、当時は、まだ大腸菌のゲノムは部分的にしか解読されておらず、ゲノムの一部を切り出したクローンを対象としており、クローンのゲノム上の元の位置を制限酵素(restriction enzyme)の認識サイトを手掛かりに決めていたのであった。

とにかく議論に出てくる用語がわからないので、自分で調べようと参考となる解説や書籍を教えてほしいとOverbeek博士にお願いすると、「本で調べるのは時間の無駄だ。わからないことがあれば私に聞け。」と言われてしまった。そこで質問するのだが、そもそも理解の前提となる生物学の基礎知識が不足



* Hideo MATSUDA

1959年11月生まれ
神戸大学大学院 自然科学研究科 システム
科学専攻博士後期課程 (1987年)
現在、大阪大学学際大学院機構 超域イ
ノベーション博士課程プログラム部門 特
任教授 学術博士
TEL: 06-6879-7390
E-mail: matsuda.hideo.itgp@osaka-u.ac.jp

しているのです、説明してもらってもなかなか理解できず苦労した。

それでも何とか研究が続けられたのは、Overbeek博士の作成したプログラムのソースコードを自由に読むことができたからであった。用語の意味はわからなかったが、プログラムのロジックと動作は理解できたので、何とか議論についていけるようになった。このときの成果は、グループ全員の共著で研究所のテクニカルレポートにまとめることができた¹⁾。

しばらくすると、別のプロジェクトとして、分子系統樹による微生物の分類について研究することになった。微生物は形態学的に分類することは困難で、共通する遺伝子配列の類似性(相同性と呼ばれる)で分類することが盛んになりつつあった。この研究は古細菌の提唱者として著名なイリノイ大学のCarl Woese教授のグループとの共同研究であった。分子系統樹を構成する方法の一つとして注目されはじめていた最尤法により系統分類を行おうということになったが、最尤法は他の方法と比べて非常に計算時間のかかる方法で、60種類の微生物を分類する分子系統樹を構成するのに、当時の高性能計算サーバーでも約12日間かかっていた。そこで、Overbeek博士の発案でスパコンによる並列計算で処理時間の短縮を図ることになり、筆者が並列化の実装を担当することになった。並列化の効率を上げるのに苦労したが、研究所には並列計算の専門家が数多くいて快く協力してくれたため、プログラム開発を何とかやり遂げることができた。さらに幸運なことに、当時世界第8位だったスパコンをかなり自由に使えたため、前述の処理時間を約4.5時間にまで短縮することができた。この成果は国際学術論文誌にて論文を出版²⁾することができ、さらにスパコン関連のトップコンファレンスの一つであるICS (International Conference on Supercomputing)に論文が採択され発表することができた³⁾。

グラフ理論による遺伝子の分類

日本に帰国してから数年経った1997年から、JST-CRESTの大腸菌ゲノムプロジェクト(研究代表 森浩禎先生(奈良先端大))に参加することができた。そこでの主な成果は、大腸菌の全遺伝子を、遺伝子を頂点、閾値以上の相同性を持つ遺伝子ペアを辺でつないだグラフで表現すると、相互に相同性

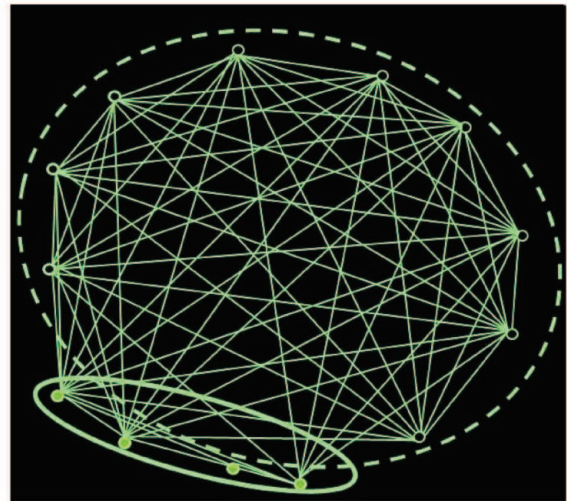


図1 準完全グラフによる遺伝子の分類

を示す遺伝子の集合は特徴的なグラフで表現できることを見出したことであった(図1)。

図1をよく見ると、すべての頂点(遺伝子)のペアが辺で連結されている(一定値以上の相同性がある)わけではなく、一部の頂点どうしの間には辺がないことがわかる。このように、すべての頂点どうしが相互に連結している(完全グラフと呼ぶ)わけではないが、ある一定の割合以上の頂点どうしが相互に連結しているグラフ構造を準完全グラフ(quasi-complete graph)と名付け、これにより遺伝子を分類できることを示した⁴⁾。これはグラフ理論という情報科学でも特に基礎的な理論が、生物学の遺伝子分類に应用できることを示している。

この経験から、生物学と情報科学のように大きく異なる分野間の融合研究では、単純に各分野の最先端の技術をマッチングするのではうまく行かないことがあり、いったん基礎となる技術に立ち返って見直して、それぞれの分野で中核となる概念間をつなぐことから始めた方がうまく行くのではないかという感触を得た。

マウス全遺伝子機能注釈プロジェクトの立ち上げ

これまでは微生物のゲノムを対象に研究してきたが、1998年に理化学研究所の林崎良英先生から、マウスの遺伝子配列(正確には、cDNA(相補DNA)配列)の解析を共同で行わないかというお誘いを受けた。マウスのゲノムは、ヒトゲノムと同等のデータ量を持ち、当時はまだゲノムが解読されていなかったため、まずは発現している遺伝子を研究対象に

する戦略が採られ、遺伝子の転写産物であるRNAの配列を解析することが盛んにおこなわれていた。当時は、RNAの全長の配列ではなく終端の断片配列を取得して解析することが主流で、これだとタンパク質のコード領域が含まれていないため、発現している遺伝子の機能を予測することが困難であった。

これに対して、林崎先生は、転写産物であるRNAの配列の全長について逆転写してcDNAを取得する技術を開発していた。cDNAの配列をシーケンサーで読み取れば、RNAの完全長配列が得られ、そこからコード領域を予測できれば翻訳によりタンパク質のアミノ酸配列が得られる。これを基盤技術として、筆者が代表者になって1999年にJSTの計算科学技術活用型特定研究開発推進事業 (ACT-JST) に応募したところ運よく採択されて、当時の最先端の64ビットCPUであるDEC Alphaプロセッサを4基搭載した計算サーバーを導入するなど、高性能なデータ解析環境を構築することができた。

この間、林崎先生のグループは、順調に完全長cDNAを取得して配列の読み取りを続けており、その数は2万個を越えるまでになっていた。マウスは大腸菌などと比べて遺伝子数もそれらが持つ機能の種類も段違いに多いため、遺伝子の機能について体系的に整理する必要がある。そこで、林崎先生の発案で、世界6カ国から60人以上の生物学者、情報科学者を理研に招き、2000年の夏に2週間かけてマウス遺伝子配列の機能アノテーション (機能注釈付け) であるFANTOM (Functional Annotation of Mouse) の会議を行うことになった。60人以上の研究者が2週間、理研の宿舎に泊まり込んで、21,076個の配列を手分けして1つ1つ目で見て、機能を決めていくという作業を行った (図2)。各配列については事前に様々な手法でコンピュータ解析がかけられており、



図2 FANTOM 会議参加者の写真
(前列右から5人目が筆者)

それらの結果を専門家の目で総合的に判断して最も適切な機能を判定していくという作業であった。林崎先生を始め、参加した多くの研究者に取って初めての試みであり、どのようなコンピュータ解析を行って、結果を視覚的にどのようにわかりやすく提示できるかが試行錯誤の連続であったが、前述のACT-JSTのプロジェクトで整備した計算環境とプロジェクトメンバーの活躍もあり、何とか2週間で全配列の機能アノテーションを完遂することができた。この成果は、Natureに論文として掲載された⁵⁾。

この論文で、筆者は、マウスの遺伝子配列からモチーフと呼ばれる共通部分配列を新たに発見したことを報告した。マウスの遺伝子の中には、ゲノム中で複数のコピーが存在するものがあり、単純に複数配列にまたがる共通部分配列を探索すると、このようなコピーも得られてしまうので、区別するのに苦労していた。林崎先生は、筆者のコンピュータ解析で得られた多数の結果に対して、モチーフと呼べるものかどうかを専門家の目で一つ一つチェックしてくださったため、モチーフの判定基準を作るのに非常に役立った。この協力がなければ、研究は進まなかったと思われる。

FANTOMは、今ではFunctional Annotation of Mammalian Genomeとして、マウス以外の哺乳類ゲノムにまで対象を広げて、現在も継続して活動している⁶⁾。筆者にとってその立ち上げに貢献できたことは非常によい経験となった。

スパコン「京」による大規模遺伝子ネットワーク推定

神戸でスパコン「京」が開発されるのに合わせて、京の高度利用を目的とした文科省の戦略プログラム事業が2011年から始まり、筆者もライフサイエンス分野の事業に参加することになった。当時、東京大学におられた宮野悟先生、玉田嘉紀先生との共同研究で、動的ベイジアンネットワークモデルによる遺伝子ネットワークの推定を研究テーマとした。ここでいう遺伝子ネットワークとは、ある遺伝子の発現が別の遺伝子の発現を活性化、または抑制する、すなわち、遺伝子間の発現制御関係を推定し、遺伝子を頂点、制御関係のある遺伝子のペアを有向辺 (制御する側の遺伝子から制御される遺伝子への向きのある辺) で表したグラフのことである。入力として、各遺伝子の発現量の経時的な変化を与えると、遺伝

子間の制御関係を推定し、グラフの形で出力する。

前述の、アルゴンヌ国立研究所で行った最尤法による分子系統樹の構築がそうであったように、動的ベイジアンネットワークモデルは遺伝子ネットワークの推定手法の中で最も計算時間のかかるものであるため、高性能な計算環境が必要であり、京を利用することで計算時間の短縮を目指した。実際には、1990年代のスパコンと比べて、京の計算能力は圧倒的に高く、約1万個の遺伝子を頂点とする大規模な遺伝子ネットワークを推定することができた。

この解析の入力データとして、筆者は京都大学の河田照雄先生との共同研究で、マウスに寒冷刺激をかけたとき、脂肪細胞中の遺伝子の発現量を経時的に測定したものを使用した。マウスに寒冷刺激を加えると、一部の脂肪細胞は熱産生する脂肪細胞(ベージュ脂肪細胞と呼ぶ)に変化するが、別の脂肪細胞は変化しない現象を遺伝子ネットワークを推定することで解明しようとしていた。京を使って遺伝子ネットワークを推定してみたところ、ベージュ脂肪細胞に変化する脂肪細胞では、遺伝子ネットワーク中で、熱産生に最も重要な遺伝子であるUcp1の近傍に、予想もしなかった遺伝子が見つかった。その遺伝子は、炎症性の生理活性物質であるIL-1 β と結合するが炎症作用は起こさないデコイ受容体(おとりの受容体)をコードしていた。一方で、寒冷刺激をかけても熱産生しない脂肪細胞については、Ucp1の近傍にこの遺伝子は見られなかった。このことから、IL-1 β が脂肪細胞の熱産生に関与していることが示唆された。その後の河田先生による検証実験から、通常の脂肪細胞は、寒冷刺激を受けてもIL-1 β の炎症作用により脂肪細胞の熱産生が抑制されるが、デコイ受容体を発現する脂肪細胞では、炎症作用が阻害されることで、熱産生が抑制されずに生じることが示された⁷⁾。寒冷刺激による脂肪細胞の熱産生とIL-1 β による炎症作用との関連を報告した論文はそれまでにはなく、京を活用して大規模遺伝子ネットワークを構築することで初めて明らかになった事例となった。

おわりに

本稿では、筆者がこれまでに経験してきた、スパコンを利用した高性能並列計算に代表される情報科学分野の技術と、ゲノムデータに代表される大規模

でかつ複雑なデータを生み出している生物学分野という、異分野間の融合研究の例を紹介させていただいた。

このような学際的な融合研究を成功させるには、各分野について基礎概念に立ち返って理解することが重要であり、さらにそれらの基礎概念を取り扱う基礎的な理論も含めて検討することが必要ということを経験した。

筆者は、現在、文理融合の学際大学院教育である超域イノベーション博士課程プログラムにおいて、AIに関する授業などを担当している。文系も含めた大学院教育となると、さらに学際的になり、学生に理解してもらうためには、なるべく基礎的な概念に立ち返って説明するように心がけている。幸いにして、授業で取り上げている生成AIは非常に学際的であり、文系、理系を問わず、幅広く使われている。生成AIの基盤モデルであるLLM(大規模言語モデル)をベースに基礎から応用までの幅広い領域で教育できればと考えている。

参考文献

- 1) Baehr, A. et al., An Integrated Database to Support Research on *Escherichia coli*, Technical Report ANL-92/1, Argonne National Laboratory, 1992.
- 2) Olsen, G.J., Matsuda, H., Hagstrom, R., Overbeek, R. fastDNAmL: a Tool for Construction of Phylogenetic Trees of DNA Sequences using Maximum Likelihood. Computer Applications in the Biosciences. 10(1), pp.41-48, 1994.
- 3) Matsuda, H., Olsen, G.J., Overbeek, R., Kaneda, Y. Fast Phylogenetic Analysis on a Massively Parallel Machine, Proceedings of the International Conference on Supercomputing (ICS 94), pp.297-302, 1994.
- 4) Matsuda, H., Ishihara, T., Hashimoto, A., Classifying Molecular Sequences using a Linkage Graph with Their Pairwise Similarities, Theoretical Computer Science, 210(2), pp.305-325, 1999.
- 5) Kawai, J. et al., Functional annotation of a full-length mouse cDNA collection. Nature. 409(6821), pp.685-690, 2001.

- 6) FANTOM Web site. <https://fantom.gsc.riken.jp/>
7) Goto, T., et al., Proinflammatory Cytokine
Interleukin-1 β Suppresses Cold-Induced

Thermogenesis in Adipocytes, Cytokine. 77,
pp.107-114, 2016.

