



生物学と数学の観点から見た大学院教育 report

大学院教育において「数学」の役割を重要視して欲しいと思っています。それは、「数学は科学の言語」であるからです。専門的に科学を学ぶためには論理思考は欠かせません。数学研究者の立場から大学院教育の在り方をお伝えします。

もう一つの大事な要素は英語教育です。英語は世界の言語であり、科学の言語でもあるといつも過言ではありません。ゼミ、卒業研究を始めとし、特に大学院のゼミや研究に英文の学術誌を読んだり、英文で自分の研究成果を書いたり、英語で発表したりすることは普通のことです。英語が得意でない、日本語の文献しか読めないなどは通用しません。日頃から英語でニュースを読んだり、YouTubeなどで英語を聞いたりすることを薦めます。

私が口癖のようにいつも学生に言っていることは、「数学は大事だ」ということです。大学院教育になると、学部教育以上に数学の役割は重要です。その理由は、まず、近代科学の先駆者の一人であるガリレオ・ガリレイ（一五六四ー一六四二）は、「宇宙という書物は数学の言葉で書かれている」という言葉を残したことです。それ以来数学は科学の言語であると認識されました。要するに、ある実験分野（物理、化学、生物など）はどの時点で数学と呼ぶかと言うと、数学的な理論や論理に基づいて実世界を客観的に説明できる時点となります。そのため、専門的に科学の分野の勉強や研究を行うには論理思考は欠かせません。生物学の場合では、自然のデータを収集、解析して、自然現象を説明しようとするので、データ解析、プログラミングや情報科学の知識は不可欠です。数学の知識なしでデータ分析はできません。例えば、主成分分析や因子分析といった多変量解析の手法は行列の固有値及び固有ベクトルに基づいているので、行列、行列式などの数学の知識がないと理解できません。当然、正しい結論を導くことができませんし、更に解析結果の正しい解釈ができるはずありません。従って、数学的な基礎に情報科学、プログラミング技術、データ分析の理解は大学院教育の基盤となっています。



サバウ
バシレ ソリン

生物学研究科
生物学部 生物学科
教授／博士（理学）

大学院教育の基盤とは、

専門的学問分野と生物学科としてのつながり

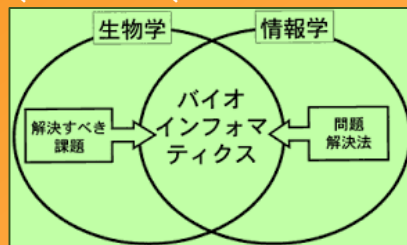
FEATURE

バイオインフォマティクスという学問分野

バイオインフォマティクスとは、生物学的な真実を理解するために数学や情報科学の手法を使う分野です。ヒトゲノム解読を始めとし、多くの生物学の進展に貢献しており、計算生物学などの名称にも使われています。計算機や情報科学は、生物学に大きな変化をもたらした生き物の「情報」の観点からの研究を促進してきました。生物の遺伝的かつ機能的な情報は、DNA、RNA、タンパク質といった高分子の中に化学物質のアルファベットを用いて暗号化されています。DNAは四つのデオキシリボ核酸（アデニン、チミン、シトシン、グアニン）から構成され、RNAは四つのリボ核酸（アデニン、ウラシル、シトシン、グアニン）から作られ、タンパク質は二十種類のアミノ酸で作られています。これらの高分子を構成要素の線形的な鎖と考えたら、それらをパソコンを用いて文字列として表すことができます。従って、情報科学のアルゴリズムを用いて、配列比較より類似性を計算し、それに基づいて相同性、分子構造、生物学的な機能などを推定できます。

配列比較は分子生物学や生命科学の研究にとつて最も有意義なツールと考えられ、インターネットの普及に伴い、ある生物学的な配列を公共データベースで検索し、世界の研究者が共有できる環境の実現が可能になってきました。

現在、ウェブブラウザを利用してBLASTなどのプログラムを使うと、自分の研究したい配列に一致している、あるいは類似している多数の配列を数秒以内に見つけることができます。これらのプログラムは非常に一般的になってきました。しかし、ある程度



吉村 奈津佳
生物学部札幌教務課
技術補

の情報科学の知識、統計学、数学の知識がないと、正しい結果を得ることができず、せっかくのデータに含まれている真実を明確にすることが難しくなります。

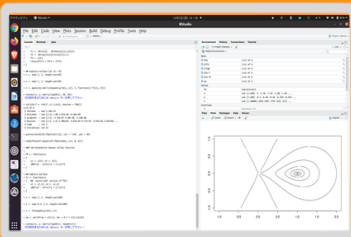
バイオインフォマティクスは、これらの分野の融合の上に立つ学問です。

生物学科としてのサポート・システム

札幌キャンパスでは、それぞれの専門により三名の技術員が配置されています。私は生物学科担当技術員として、主に実験・実習の準備や指導補助、学科施設や機器の管理を担当しています。また、学生から履修や資格取得、就職活動、生活面や体調面における相談を受ける事もあり、必要に応じて各部署への橋渡しを行っています。

一方、卒業研究や教員の研究においては「研究環境を整える」という点を意識しながら、様々な面で関わっています。最近では、大学の改組に伴い、新しく加わった専門分野に対応すべく施設・設備の拡充と研究スペースの確保について、教員間の調整役を担いながら段階的に進めています。活発な研究活動のためには、技術的な面でのサポートはもちろん、広い意味での研究環境の整備は必須なことです。

今後、学科の垣根を越えて、学部や校舎間で機器や施設の共同利用がさらに進められる事が想定されています。実際そのような取り組みが始まっていますが、分野により共同利用が難しいケースもあり、それぞれの事情を、私たち技術員が把握しながら、より研究に取り組みやすい環境を整えることができるよう、工夫していきたいと考えています。



◀データ分析ツール：【R言語】

統計解析に適した命令体系を持っているプログラム言語です。インタープリタ型言語であり、マルチプラットフォームで動作することも特徴の一つです。またオープンソースで無償の言語ですから、誰もが簡単に同じ作業環境を整えることができます。

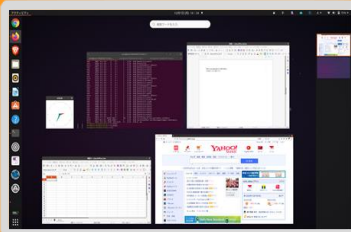
▶オペレーティング・システム：【Linux (リナックス)】

狭義にはUnix系オペレーティングシステムカーネルであるLinuxカーネルを指し、広義にはそれをカーネルとして周辺を整備したシステム全体のことと言えます。



◀オフィス関係：【LibreOffice】

The Document Foundationのプロジェクトが作成する自由・オープンソースソフトのオフィススイートであります。商用ソフトであったStarOfficeをオープンソース化したOpenOffice.orgから派生して作られました。



◆本研究室では、

ドライバイオロジーの研究を行っており、主にパソコンを使用してデータ解析を行います。使用するソフトは、オープンソースソフトウェア (Open source software: OSS) を使用することが特徴です。オープンソースソフトウェアのライセンスとは、誰でも自由にそのソフトウェアの中身を閲覧したり、変更したり、ソフト自体を再配布できるものです。勿論、著作権は廃止されていませんが、基本的に個人的な使用、あるいは教育研究の場合、自由に無料で使用できるソフトウェアです。また、オープンソースソフトウェアには全てのジャンルが揃っています。データ分析や研究を行う前にオープンソースソフトウェアについて勉強することをオススメします。

ここでは使用しているいくつかのソフトを紹介します。

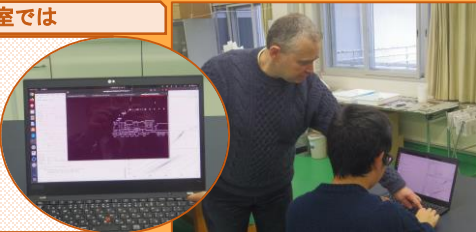
◀幾何学の観点から



インスリンの分子進化の研究について、糖尿病との関連を講義。

▶研究室では

学生向けのオープンソースソフトウェア (OSS)、Linuxなどの研究開発を中心としたゼミ。



PICKUP

生物学研究科 生物学専攻: サバウ バシレ ソリン 教授 研究室

生物の進化を距離空間の性質を用いて研究することによって、今まで分らなかった新たな事実が明かされます。このように数学的な技法を用いて生物を研究しているサバウ バシレ ソリン教授の研究活動を **PICKUP!**

配列比較の幾何学

生物学研究科 生物学専攻
生物学部 生物学科
教授/博士 (理学)
サバウ バシレ ソリン

サバウ バシレ ソリン

一般的に、距離は数学、特に解析学、幾何学の基本的な概念の一つです。三次元空間の二点間A、Bの距離は正の実数であり対称となつています。つまり、点Aから点Bまでの距離は点Bから点Aまでの距離に等しいこととなります。しかし、実世界では、完全に対称な距離関数は存在しません。その理由は、様々な外部の力の影響によって、距離関数が変化するためです。例えば、一定の速度のボートが湖あるいは海を航行しているとします。仮にその速度を一単位にした場合、距離と時間が同じであり、風や波が全く無い場合、二点間の距離は対称です。しかし、風が吹いたり、波が立ったりすると、追い風と向かい風の移動にかかる時間は明らかに異なります。そのため、距離も異なることとなります。生物学やその他の応用的な分野では、距離関数が定義され、それによって類似性が定義されています。例えば、二つの遺伝子の間の進化距離が近いと、その類似性が高い、距離が遠い時に、類似性が低いという解釈ができます。しかし、これはもちろん非対称な距離です。つまり、生化学性質の異なるAというアミノ酸とBというアミノ酸の進化的距離は対称ではありません。

私の研究では、アミノ酸やタンパク質の進化距離の幾何学的な性質を調査し、その結果を生物学的に解釈します。その手法を、現代的な幾何学の理論を生物学配列の類似性に活用し、計量幾何学の観点からインスリンの進化について調べると、疎水性アミノ酸は、進化距離の計算及び遠縁タンパク質の相同性の推定では重要な役割を果たしていることが明らかに。遠縁タンパク質については、疎水性残基を別に扱って研究が必要があるということが分かりました。これによって配列比較から推定される相同性の偽陽性を減らすことが可能になり、この理論を様々な研究課題に応用することが可能となります。

インスリンの進化と糖尿病との関係

現在、数学、特に幾何学の観点からインスリンの分子進化の研究について興味を持ち、糖尿病との関連を調べる研究を行っています。

インスリンは非常に多くの生物が持っているため、進化の研究に相応しいと考えられます。私の研究では、ヒトから線虫の *C. elegans* まで六十種の生物のインスリンの相同配列を幾何学の観点から分析し、相同配列と進化の研究に有意な見解を提供することを目指しています。

まず、幾何学の観点からは、一致性が四十%以上の配列から得られるBLOSUMスコア行列より誘導される進化距離は非対称な距離関数であることに気付きます。この距離関数はアミノ酸空間の幾何学を正確に表していることが分かります。更に、短い進化距離でもこの非対称な距離関数を用いることで、平均対称距離と比較すると、より良い結果が得られることが証明できます。換言すると、このような解析では、常に最適な距離を選ぶことがより良い結果を導くことに繋がります。

次に、幾何学とインスリン相同配列空間の位相の間に高い整合性があることを証明できます。具体的に、このインスリン配列空間の系統解析はClustalX (<http://www.clustal.org/>) でアラインメントした、生物学的に有意な五つのグループに分類されました(図参照)。

結論としては、距離幾何学はまだ十分活用されていない新しい学問であり、これを用いることによって進化的な観点からタンパク質の機能を研究できます。また、距離幾何学の様々な応用によって、地球上の生命の誕生及び進化の理解が深まることを期待しています。

インスリン配列の系統樹

