

バイオインフォマティクス推進センター事業
事業報告書

平成24年1月

独立行政法人
科学技術振興機構

目 次

I	全体要旨	
1.	事業の目的	I-1
2.	事業の成果	I-1
3.	今後について	I-2
II	事業の目的と発足までの経緯	
1 章	事業の目的	II-1
2 章	発足までの経緯	II-3
III	事業の推進と成果	
1 章	事業の推進体制	III-1
2 章	研究開発の推進	III-5
3 章	データベースの提供	III-33
4 章	人材育成	III-46
IV	事業を取り巻く状況の変遷	
1 章	国内の状況	IV-1
2 章	諸外国の状況	IV-11
3 章	論文・特許による研究の変遷の把握	IV-17
V	バイオインフォマティクスの今後の展望	
1 章	基礎研究へのインパクト	V-1
2 章	産業への展開	V-6
3 章	今後の課題とその解決のための提案	V-11
VI	資料編	
1 章	バイオインフォマティクス委員会 委員名簿	VI-1
2 章	研究開発課題の成果	VI-3
3 章	研究開発課題の成果 (データベース、論文、特許等)	VI-21

I 全体要旨

1. 事業の目的

膨大かつ多種多様な生物情報を整理統合し、そこから有用な知識を見出すことにより、新しい産業の創出、新しい医療の開拓、新しい農業の構築へと発展することを可能とする情報生物学（バイオインフォマティクス）の発展の推進及びそれを基盤とした 21 世紀の新しい生物学の創造を目指して、2001 年度（平成 13 年度）からバイオインフォマティクス推進センター事業（BIRD 事業）が開始された。事業内容は、大きく分けると、1. 生命情報データベースの高度化・標準化、2. 創造的な生物・情報知識融合型の研究開発、3. 研究成果情報発信からなる。また、この他に人材育成の活動も行った。

1. の研究課題では、膨大な生物情報からの新しい知識の発見等に不可欠なデータベースの構築や高度化を行い、これらのデータベースを有機的に統合化し、生物をシステムとして理解することを可能とするとともに、それらが広く社会に活用されるような環境を整備することを目的とした。2. の研究課題では、情報科学と生物学とを融合したアプローチにより、多彩な生物情報から生物現象の原理や法則を発見し体系化することを目指す研究開発を推進し、これらの成果がバイオインフォマティクスの発展及び新しい情報生物学創造のインセンティブとして機能し、これを基盤として新たな生物学が樹立されるように支援した。3. では研究開発課題の成果報告会を開催した。また、人材育成活動としてゲノムリテラシー講座の開催などを行った。

2. 事業の成果

(1) 研究開発の推進

「生命情報データベースの高度化・標準化」は第 I 期 4 課題、第 II 期 7 課題が実施された。また「創造的な生物・情報知識融合型の研究開発」は事業期間全体で計 33 課題（継続課題を除くと 22 課題）を実施した。事後評価において当初の研究開発計画をほぼ達成あるいは計画以上の成果が得ることできたと評価された課題は総合評価値が付けられた 32 課題中 24 課題に及ぶ。また論文発表等多くの外部発表実績も残した。以下に特筆すべき成果を挙げる。

「生命情報データベースの高度化・標準化」の中で実施された「ゲノムと環境の統合解析による生命システムの機能解読（金久實）」では KEGG データベースの高度化を通じて、データベースの構築、手法の開発、基礎研究への応用などにバランス良く論文発表されており、また医薬品、医療、環境分野などへの産業利用が想定され得る成果を残している。また「蛋白質構造データバンクの国際的な構築と高度化（PDBj）（中村春木）」では基礎研究への応用に関する論文が非常に多く、基礎生物学への波及が想定されうる成果を残している。一方、「創造的な生物・情報知識融合型の研究開発」においても「ヒト胚の形態発生に関する三次元データベース（塩田浩平）」のように非常に価値の高いデータベースを構築することで、ライクサイエンスや産業応用への波及効果が非常に大きい成果を出しているものがある。

(2) 研究開発成果の公表

研究開発課題の課題終了時に、各課題での新しい試みや得られた知見を多くの方々に直接紹介することを目的として、計8回の研究開発成果報告会を開催した。

(3) 人材育成

人材養成としてゲノムリテラシー講座を計45回開催した。一部の講座についてはゲノム4領域「ゲノム情報科学の新展開」との共催である。受講者は大学院生や若手研究者だけでなく、企業の研究開発担当者にも拡がり、一定の成果が得られた。また、ヒトゲノムに関連する複数のデータベースを一度に検索できる HOWDY や日本人の一塩基多型データのデータベースである JSNP、ライフサイエンス分野のデータベース情報を収録した WING(現 WINGpro)データベースの開発・提供を行い、バイオインフォマティクスの普及に努めた。

3. 今後について

BIRD 事業は KEGG や DDBJ、PDBj のように世界的にも利用されているデータベースの高度化を実施するとともに、メタボロームやインタラクトームのような新規分野においてもデータベースを開発し、基礎研究分野へ大きなインパクトを与えた。また KEGG の DRUG や DISEASE データベースの充実や、ヒト胚の形態発生データベースの構築、メタゲノム研究など、創薬、診断・医療、環境分野など産業への展開も大いに期待できる成果を残した。

今後、より高次の生命現象や、多種多様で大量のデータを扱うためには、データベースや分析手法の研究開発によるバイオインフォマティクスの発展が必須である。また産業への展開のためには、個別化医療や創薬などのニーズに対応した情報基盤の整備が不可欠となってくると考えられ、BIRD 事業の成果を活かして研究開発が進められることが期待される。

BIRD 事業は、2001 年度（平成 13 年度）から 2011 年度（平成 23 年度）までの 11 年間という長期に渡ってバイオインフォマティクスにターゲットを絞って研究開発を推進した点において特筆すべき事業と言える。BIRD 事業は 2012 年（平成 24 年）3 月に終了するが、BIRD 事業と統合データベースプロジェクトが一本化する形で、バイオサイエンスデータベースセンター(NBDC)が 2011 年（平成 23 年）4 月に設置された。NBDC ではデータベースの統合化や基盤技術開発などが進められるが、BIRD 事業で実施した「創造的な生物・情報知識融合型の研究開発」は引き継がれていない。

さらなるバイオインフォマティクスの発展のためには、NBDC での事業活動とともに、BIRD 事業の理念を引き継ぎ、情報科学と生物科学とを融合したアプローチにより、多彩な生物情報から生物現象の原理や法則を発見し体系化することを目指す研究開発を進められる新たな枠組みが必要である。

II 事業の目的と発足までの経緯

1章 事業の目的

1節 事業の概要

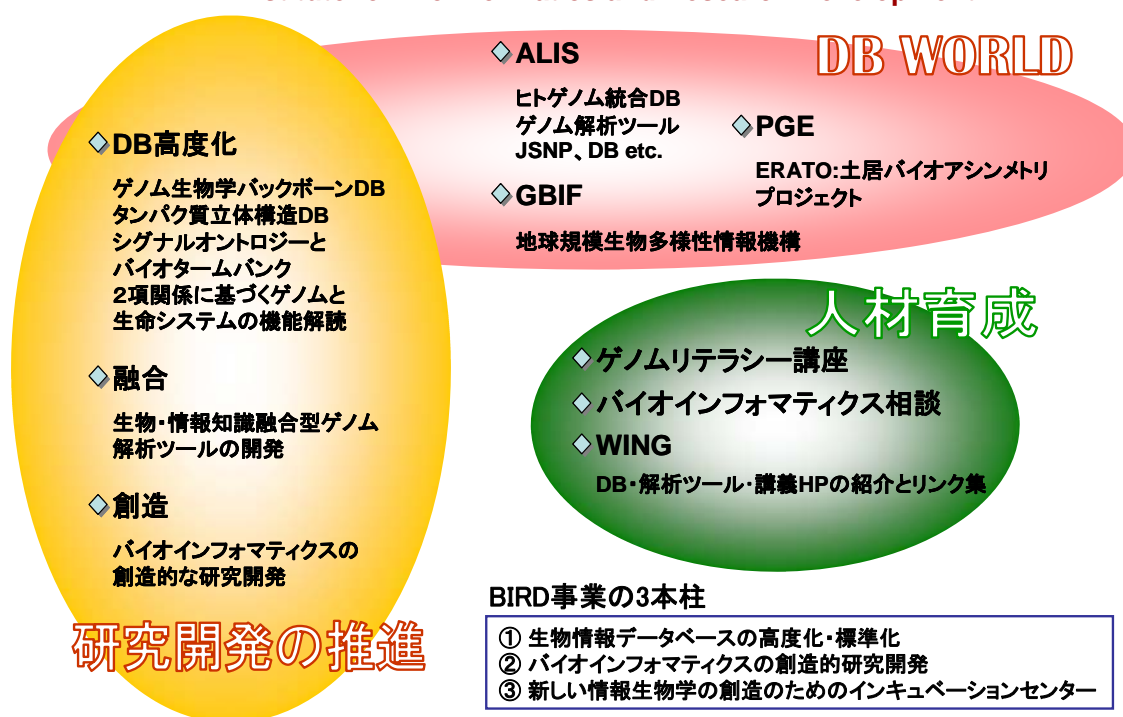
膨大かつ多種多様な生物情報を整理統合し、そこから有用な知識を見出すことにより、新しい産業の創出、新しい医療の開拓、新しい農業の構築へと発展することを可能とする情報生物学（バイオインフォマティクス）の発展の推進及びそれを基盤とした21世紀の新しい生物科学の創造を目指してバイオインフォマティクス推進センター（以下、BIRD）を平成13年4月に設置し、統括、副統括の指導のもと、本事業が運営された。また、事業の推進と課題評価について調査審議するために、バイオインフォマティクス委員会を設置した。事業内容は、大きく分けると、1. 生命情報データベースの高度化・標準化、2. 創造的な生物・情報知識融合型の研究開発、からなる。さらに、本研究開発課題に関連して、3. 研究成果情報発信の役割も果たしている（図表II-1）。

＜統括・副統括＞ （所属機関名役職は平成13年当時）
統括 勝木元也（大学共同利用機関法人 自然科学研究機構 基礎生物学研究所 所長）
副統括 高木利久（東京大学医科学研究所 教授）

図表 II-1 BIRD 事業の全体像

バイオインフォマティクス推進センター(BIRD)

Institute for Bioinformatics and Research Development



2節 事業の目標

1) 事業発足の目的

2000年（平成12年）11月に科学技術会議ライフサイエンス部会ゲノム科学委員会は「バイオインフォマティクス（バイオのための情報科学および情報論的解釈に基づく生命科学）の推進」を目標として、「ゲノム情報科学におけるわが国の戦略について」の方策をまとめた。①人材養成、②データベース構築、③情報解析技術開発の項目について方策を提案している。①では、即戦力の養成、研究交流、試行錯誤の「場」としての拠点形成、大学等における中長期的視点に立った人材の育成などが、②では、国家レベルでの有識者会議による戦略立案、戦略の検討、実施を支える組織の設置などが、③では、情報解析技術開発の在り方等に関する戦略を検討する体制の整備、国家的な支援を要する分野などが謳われた。

BIRD事業は、このうち②データベース構築および③情報解析技術開発を担う目的で2001年（平成13年）4月に発足した。なお、①人材養成に関しては科学技術振興調整費により大学や研究所での人材養成に対する支援が開始された。

2) JST 中期目標の中での事業目標

独立行政法人としてJSTが発足するにあたり、文部科学省は、JSTが達成すべき業務運営に関する目標として、平成15年10月1日、第1期中期目標を定めた。その後、平成19年4月1日からの第2期中期目標が定められた。以下に第1期及び第2期中期目標の中でBIRD事業に関する個別事項を抜粋する。

このようにBIRD事業の目標は、ゲノム情報等の生物情報のデータベースを構築、高度化、統合、維持・運用、普及を図り、ライフサイエンス研究の進展につながる情報基盤を整備することにある。

【第1期中期目標においてBIRD事業に関する個別事項】

ゲノム情報等の膨大な生物情報を整理統合し、有用な知識を見出すことによる新産業の創出等を図るため、新しい生物情報の研究開発によるデータベースの整備等を推進するとともに、当該データベースの普及を促進する。

【第2期中期目標においてBIRD事業に関する個別事項】

ゲノム情報等の生物情報データベースの構築、高度化、活用のための研究開発を行い、研究開発成果を情報発信するとともに、データベースの統合・維持・運用を図ることにより、世界最高水準のライフサイエンス分野の情報基盤の整備の一翼を担うとともにライフサイエンス研究のさらなる進展に貢献する。

2章 発足までの経緯

1節 国内外の科学技術研究の変遷

BIRD 事業の発足の経緯として、90年代から事業発足の2001年までにおける国内外のライフサイエンス研究の変遷を整理する。

1) 大型ゲノム解読プロジェクトの進行

1953年にワトソン・クリックにより発見された二重らせん構造をもつDNAの配列解読がシーケンサーやPCR技術の進歩により現実的になった1986年にヒトゲノム解読計画が提言され、1991年に国際的な解読計画が開始された。その後1996年に日本も参加する国際コンソーシアムが結成され、各国の分担による解読作業が進められ、2000年にドラフト配列の決定が完了し、2001年に国際コンソーシアムとセセラ社からそれぞれ発表された。

なお、ヒトゲノム解読完了以前では、1995年にインフルエンザ菌(*Haemophilus Influenza*)が最初にゲノム全体の解読が行われ、その後様々な真正細菌や古細菌、真核生物のゲノムが解読された。真核生物の中では、1997年の出芽酵母(*Saccharomyces cerevisiae*)が最初で、翌1998年にモデル生物としても注目されていた線虫(*Caenorhabditis elegans*)、2000年にショウジョウバエ(*Drosophila melanogaster*)、シロイヌナズナ(*Arabidopsis thaliana*)が解読され、5番目の真核生物としてヒトゲノムが解読された。

このように大型のゲノム解読プロジェクトが進むことで、DNA配列データなどの大量のデータが生成されることになり、これらのデータを使いやすくするためにデータベース構築を行ったり、配列断片をつなぎ合わせるアセンブリング処理を行ったり、遺伝子のコード領域を同定し遺伝子機能を分析するなど、バイオインフォマティクス技術の高度化の要望がこれまで以上に高まった時期であった。

2) ポストゲノム時代の到来

2000年にヒトゲノムのドラフト配列が決定したわけであるが、ゲノムが解読されたといってもゲノムにどのようなタンパク質の設計図が書かれていて、それがどのようなメカニズムで適材適所に発現し、相互に作用をしながら生命の機能を司っているかは判明していない。

実際、国内では2000年のミレニアム・プロジェクトの一環として、日本人の一塩基多型プロジェクト JSNP を開始しており、ゲノムの個人差を分析し疾患の研究などへ活かす研究も進められていた。また90年初頭から当時大阪大学にいた大久保公策氏を中心としたグループは組織ごとの遺伝子の発現状況を調べる研究を行っており、BodyMap というデータベースを公開していた。

このようなポストゲノム研究の必要性もこの頃既に叫ばれており、ゲノム配列デ

ータだけでなく、タンパク質配列・立体構造データ、遺伝子発現データなど様々なデータを統合化し分析するためのバイオインフォマティクス技術の開発が急務となっていた。

3) バイオインフォマティクスの需要の高まり

このようにゲノム解読による莫大な生成データの蓄積や分析、ポストゲノム研究による多様な種類のデータへの対応のために、ライフサイエンス研究側からのバイオインフォマティクスへの要望が急速に高まることとなった。

一方のバイオインフォマティクス分野においても研究や開発を既に進めており、国内では本事業において研究課題として採択されている DDBJ や KEGG などはそれぞれ 1986 年、1995 年に運用あるいは構築を開始している。DDBJ の場合、米国 NCBI の GenBank と欧州の EMBL と 3 者連携のもとで世界的な核酸配列データベースを構築・運用している。また KEGG は世界にも数少ないパスウェイデータベースとして構築を開始しその後様々な情報を追加し、現在では世界的な地位を確立している。日本が主導的に開催するバイオインフォマティクス関連の国際学会としては、GIW(International Conference on Genome Informatics)があり、第 1 回が 1990 年に開催され、その後毎年開催されてきた。

90 年代から既にバイオインフォマティクス研究は進められてきたものの、大型ゲノム解読プロジェクトやポストゲノム研究などの開始により、よりライフサイエンス研究のニーズにマッチした技術の研究・開発及び人材養成の必要性が高まった時期であると言える。

2節 国内の科学技術政策の変遷

ライフサイエンス分野におけるバイオインフォマティクス関連の科学技術政策・事業について、BIRD 事業が開始される 2001 年までの変遷の概要を省庁別に整理する。なお、90 年代はヒトゲノム解読やイネゲノム解読などの国際プロジェクトへの貢献が中心となることから、事業を中心に整理する。

1) 内閣官房

【ミレニアム・ゲノム・プロジェクト】

1999 年 12 月、政府は、新しいミレニアム（千年紀）の始まりを目前に控え、人類の直面する課題に応え、新しい産業を生み出す大胆な技術革新に取り組む「ミレニアム・プロジェクト」を決定した。今後の我が国経済社会にとって重要性や緊要性の高い情報化、高齢化、環境対応の三分野について、技術革新を中心とした産学官共同プロジェクトを構築し、明るい未来を切り拓く核を作り上げるものである。

ミレニアム・プロジェクトの高齢化分野に該当する「高齢化社会に対応し個人の特徴に応じた革新的医療の実現（ヒトゲノム）」、「豊かで健康な食生活と安心して暮らせる生活環境の実現（イネゲノム）」（これらを「ミレニアム・ゲノム・プロジェクト」と呼ぶ）においては、ゲノムに係る研究開発を国家のイニシアティブの下に、研究者を結集して強力で進めることにより、来るべき新世紀を高齢者にとって活気ある社会への道を切り拓き、安全性の確保と国民の理解の増進を図りつつ、バイオテクノロジーの応用によって幅広い分野における新しい産業の創出を図っていくこととともに、新世紀の人類社会の発展に大きく貢献していくことを目指すとされた。

各個別事業の研究代表者から構成されるプロジェクトチーム（①ヒトゲノム多様性、②疾患遺伝子、③バイオインフォマティクス、④発生・分化・再生、⑤イネゲノム）を部門毎に設け、プロジェクトが推進された。

2) 文部科学省

【科学技術振興調整費】

1991 年から 1995 年にかけて「ヒト遺伝子地図作製技術の開発に関する研究」を JST（当時、日本科学技術情報センター）や大学・研究機関によって実施している。本研究は、1995 年から本格的に開始されるヒトゲノム解読の国際プロジェクトの基礎となる遺伝子地図を作成することを目的とした。

【高機能基盤生体データベース事業】

1995 年から JST（当時、日本科学技術情報センター）がヒトゲノム解読の国際プロジェクトの一翼を担うべく、大学・研究機関（東海大・北里大・慶応大・癌研究会）とともに

にゲノムシーケンシングを実施し、またヒトゲノムの統合データベースや遺伝子予測ツールの開発、ホモロジー検索サービスサイトの構築などを行った。

【科学研究費特定領域研究 ゲノム4領域】

ミレニアム・プロジェクトの一環として、2000年から開始された。統合ゲノム、ゲノム医科学、ゲノム生物学、ゲノム情報科学、の4領域から構成される。特にゲノム情報科学（正式名称は「ゲノム情報科学の新展開」）では、対象データの多様化やゲノムや生命に内在する情動的・数理的構造を捉える際の理論的側面での役割への期待に応えるために、高度データベースの構築と高次生物知識の体系化、ゲノムデータベースからの知識発見、タンパク質高次構造に基づくゲノム情報科学、遺伝子ネットワークのモデル化とシミュレーション、を目的とした。なお、本研究は2004年まで実施され、その後2005年より新たな5年プロジェクトとして実施されている。後半の5年間では生命システム情報、比較ゲノム、応用ゲノム、基盤ゲノムの4領域となり、それぞれの領域でバイオインフォマティクスの研究が取り組まれた。

3) 経済産業省

【完全長 cDNA 構造解析プロジェクト】

東京大学の菅野純夫氏を研究代表としてヒトの完全長 cDNA の取得及び配列解読を実施した。完全長 cDNA の分野においては世界的にリードした事業となり、その後様々な研究で利用される重要なデータとなった。なお、本事業は後述のミレニアム・プロジェクトの一環として行われた。

4) 厚生労働省

【遺伝子解析による疾病対策・創薬推進事業】

後述のミレニアム・プロジェクトの一環として、2004年度を目標に、痴呆、がん、糖尿病、高血圧等の高齢者の主要な疾患の遺伝子の解明に基づくオーダーメイド医療を実現し、画期的な新薬の開発に着手することを目的に実施された。

5) 農林水産省

【イネゲノムプロジェクト】

1991年（平成3年）に、農林水産省は、わが国の基幹穀物であるイネのゲノム研究を開始した。研究基盤の構築を目標にして、イネの遺伝子の大量解析、遺伝地図作成及び物理地図作成に取り組んだ。1998年（平成10年）以降、イネゲノム研究の新たな展開をめざしてイネゲノム全塩基配列解明プロジェクト、ミュータントパネルプロジェクト、イネ完全長 cDNA プロジェクト等の研究プロジェクトを推進し、現在に至っている。

Ⅲ 事業の推進と成果

1章 事業の推進体制

1節 事業の運営体制

BIRD 事業では統括をプログラムオフィサーとし、助言する有識者から構成されるバイオインフォマティクス委員会を組織した体制により事業を運営した（VI. 資料編 1章）。

具体的には、統括は生命情報データベースの高度化・標準化および創造的な生物・情報知識融合型の研究開発課題、ツール開発の評価の取りまとめや、研究者に対する指導・助言、研究開発の進捗状況の総合的な把握・調整等を職務とした。

また、副統括は生命情報データベースの高度化・標準化の総合調整や、創造的な生物・情報知識融合型の研究開発課題、ツール開発の実施者への指導・助言、普及活動の調整、計算機資源活用の調整、統括の補佐を職務とした。

＜統括・副統括＞ （所属機関名役職は平成 13 年当時）

統括 勝木元也（大学共同利用機関法人 自然科学研究機構 基礎生物学研究所 所長）

副統括 高木利久（東京大学医科学研究所 教授）

バイオインフォマティクス委員会は、BIRD 事業に関し、次に掲げる事項を調査審議するため、外部の学識経験者で構成した。

研究開発課題及び代表研究者の選考に関すること

研究開発の実施及び評価に関すること

その他事業の推進に必要な事項に関すること

BIRD は、平成 13 年にゲノム情報等の膨大な生物情報を整理統合し、有用な知識を見出すことによる新産業の創出を図るため、新しい生物情報の研究開発によるデータベースの整備等を推進するとともに、当該データベースの普及を促進するための拠点として設置し、統括、副統括、技術参事、事務参事、事務補助員で構成される組織とした。BIRD では、研究開発経費の管理及び執行、研究員等の雇用、特許等の知財取得の支援、科学技術計算サーバの運用と利用支援など研究開発課題の実施を支援する多様な業務を担った。大学等とは共同研究契約を締結し、消耗品や大学等に雇用される研究者の旅費等は大学等に納入した。

平成 14 年度以降は、それ以前に「高機能生体データベース開発事業」及び「多型情報データベース開発事業」として開発・提供してきた各種情報提供サービスを本事業に組み入れ、サービスを継続することとなった。

平成 17 年度採択課題より、BIRD の事務所機能（研究開発経費の執行、研究員等の雇用、特許等の知財取得の支援など）による推進支援に代え、大学等で経理事務を行う受託研究契約締結に基づき研究開発を推進する体制とした。平成 18 年度末には事務所機能を終了している。ただし、研究員等の雇用は一部継続した。このように受託研究契約を締結して研究機関等で行う事務処理について説明した「事務処理説明書」や「様式」等をホームページに記載した。

2節 研究開発進捗状況の把握と評価

研究開発課題については、統括がバイオインフォマティクス委員会の助言を受け、事前評価、事後評価、5年以上継続する課題については3年目の中間評価を実施した。また、年度単位で、年度報告と次年度計画を毎年3月に提出を受け、バイオインフォマティクス委員会委員からの助言、統括の了承などの進捗状況把握と進捗に応じた研究開発費を含めた統括のマネジメントを要する運営としていた。

また、追跡調査に基づいた追跡評価を課題終了数年後に行うこととなっており、平成13年度採択研究開発課題11課題について、平成19年度は追跡調査の予備調査、平成20年度は追跡調査、平成21年度は追跡評価を実施した。

なお、事業単位では、平成15年度概算要求より総合科学技術会議による優先度判定や、文部科学省に設置された独立行政法人評価部会による毎年度の独立行政法人評価の一環としての事業評価などを受けている。

研究開発課題の進捗については、成果発表報告や知財取得の相談の他、課題毎に行われる打合せへの参加、シンポジウムやワークショップへの参加、その他採択年度単位で行われる成果発表などにより把握に努め、要望や相談を受けて解決策の提示などをおこなった。

3節 研究費の配分

平成13年度の事業開始に際し、生命情報データベースの高度化・標準化については1課題当たり1.5億円/年を目安とし、創造的な生物・情報知識融合型の研究開発のうち、情報科学と生物科学との融合型研究では1課題当たり1億円/年程度、情報科学を主体とした研究開発では1課題当たり3千万円/年という目安を提示し、公募した。審査に当たり3年間あるいは5年間の経費が含まれている提案書において経費の妥当性を示し、提案者が経費についても考える仕組みとした。

採択後は全研究開発期間について提示された予算額を基準とし、研究開発の進捗や一時的な経費の発生など年度毎の報告と次年度計画についてバイオインフォマティクス委員会の助言を得て統括による判断で決定された。

平成17年度以降の採択課題についてはJSTによる研究開発経費の直接執行から大学等研究開発機関での執行に変更すると同時に受託研究契約の締結を行うこととし、間接経費30%の納入を行うこととした。生命情報データベースの高度化・標準化については間接経費30%を含めた1課題当たりの研究費は年間5,200万円から2億円、創造的な生物・情報知識融合型の研究開発については1課題当たり年間3,900万円を目安に公募した。その際も全研究開発期間について提案された研究開発費を基準として、研究開発の進捗や一時的な経費の発生など年度毎の報告と次年度計画についてバイオインフォマティクス委員会の助言を得て統括による判断で決定された。進

捗に伴う人件費や他のプロジェクトとの関係での増額などが実施された。

各年の予算総額の推移は以下の通りである。

平成 13 年度	20.0 億円	
平成 14 年度	21.5 億円	JSNP 等の予算追加
平成 15 年度	20.2 億円	10 月より独立行政法人化
平成 16 年度	17.7 億円	
平成 17 年度	18.6 億円	新規研究開発課題募集の予算追加
平成 18 年度	18.1 億円	
平成 19 年度	16.8 億円	科学技術計算サーバの整理、事務合理化
平成 20 年度	16.8 億円	
平成 21 年度	18.4 億円	文部科学省統合データベースプロジェクト一部移管
平成 22 年度	17.6 億円	文部科学省統合データベースプロジェクト中核機関分の移管
平成 23 年度	0.6 億円	バイオサイエンスデータベースセンターの発足 (ライフサイエンスデータベース統合推進事業の開始)

4節 バイオインフォマティクスの普及推進

生命情報データベースの高度化・標準化のプログラムを遂行するために国内外のデータベースについて調査し、その結果を WING データベースとして提供するとともに、バイオインフォマティクス関係のソフトウェア(ツール)についても調査を行い、ツールリンク集として提供を始め、現在はキーワード検索可能なリンク集として使い方や引用論文情報も付与したサービスとして提供している。

人材養成については、平成 13 年度より科学技術振興調整費にてプログラムとして実施されてきた。こうした大学や研究機関での人材養成とは異なるが、実習を伴う講習会(ゲノムリテラシー講座)を副統括等のアドバイスにより開催した。詳細は、4 章にまとめる。

その他、バイオインフォマティクス相談として、匿名で個別相談を受けるサービスも提供した。

平成 14 年度に一本化した「高機能生体データベース開発事業」及び「多型情報データベース開発事業」において開発・提供してきた各種データベース提供サービスも更新を行い、寄せられる質問に対応にしながら、バイオインフォマティクスの推進に努めた。

5節 研究開発成果の公表

研究開発課題の課題終了時に、各課題での新しい試みや得られた知見を多くの方々に直接紹介することを目的として、下記に示す8回の研究開発成果報告会を開催した。

第1回研究開発成果報告会（2005年1月27日）

創造的な生物・情報知識融合型の研究開発 平成13年度採択新規研究開発課題

第2回研究開発成果報告会（2006年3月13日）

生命情報データベースの高度化・標準化 第Ⅰ期研究開発課題

第3回研究開発成果報告会（2007年11月2日）

創造的な生物・情報知識融合型の研究開発 平成16年度採択 継続研究開発課題

第4回研究開発成果報告会（2008年11月4日）

創造的な生物・情報知識融合型の研究開発 平成17年度採択 新規研究開発課題

第5回研究開発成果報告会（2009年11月17日）

創造的な生物・情報知識融合型の研究開発 平成18年度採択 新規研究開発課題

第6回研究開発成果報告会（2010年11月10日）

創造的な生物・情報知識融合型の研究開発 平成19年度採択 新規研究開発課題

創造的な生物・情報知識融合型の研究開発 平成20年度採択 継続研究開発課題

第7回研究開発成果報告会（2011年3月8日）

生命情報データベースの高度化・標準化 第Ⅱ期研究開発課題

第8回研究開発成果報告会（2011年11月14日）

創造的な生物・情報知識融合型の研究開発 平成21年度採択 継続研究開発課題

2章 研究開発の推進

1節 生命情報データベースの高度化・標準化

膨大な生物情報からの新しい知識の発見等に不可欠なデータベースの構築や高度化を行うことを目的とした。また、これらのデータベースを有機的に統合化し、生物をシステムとして理解することを可能とするとともに、それらが広く社会に活用されるような環境を整備する。このために、平成13～17年度に第I期研究開発を実施し、平成18～22年度に第II期研究開発を実施した。各研究開発課題の名称と代表研究者を図表III-1、図表III-2に示す。

図表 III-1 第I期研究開発課題（研究開発期間：平成13～17年度）一覧

代表研究者（終了時所属）	研究開発課題名
金久 實 （京都大学化学研究所 教授）	2項関係に基づくゲノムと生命システムの機能解読
菅原 秀明 （国立遺伝学研究所 教授）	ゲノム生物学バックボーンデータベースの構築提供
高木 利久 （東京大学大学院新領域創成科学研究科 教授）	シグナルオントロジーとバイオタームバンクの開発
中村 春木 （大阪大学蛋白質研究所 教授）	蛋白質立体構造データベースの高度化

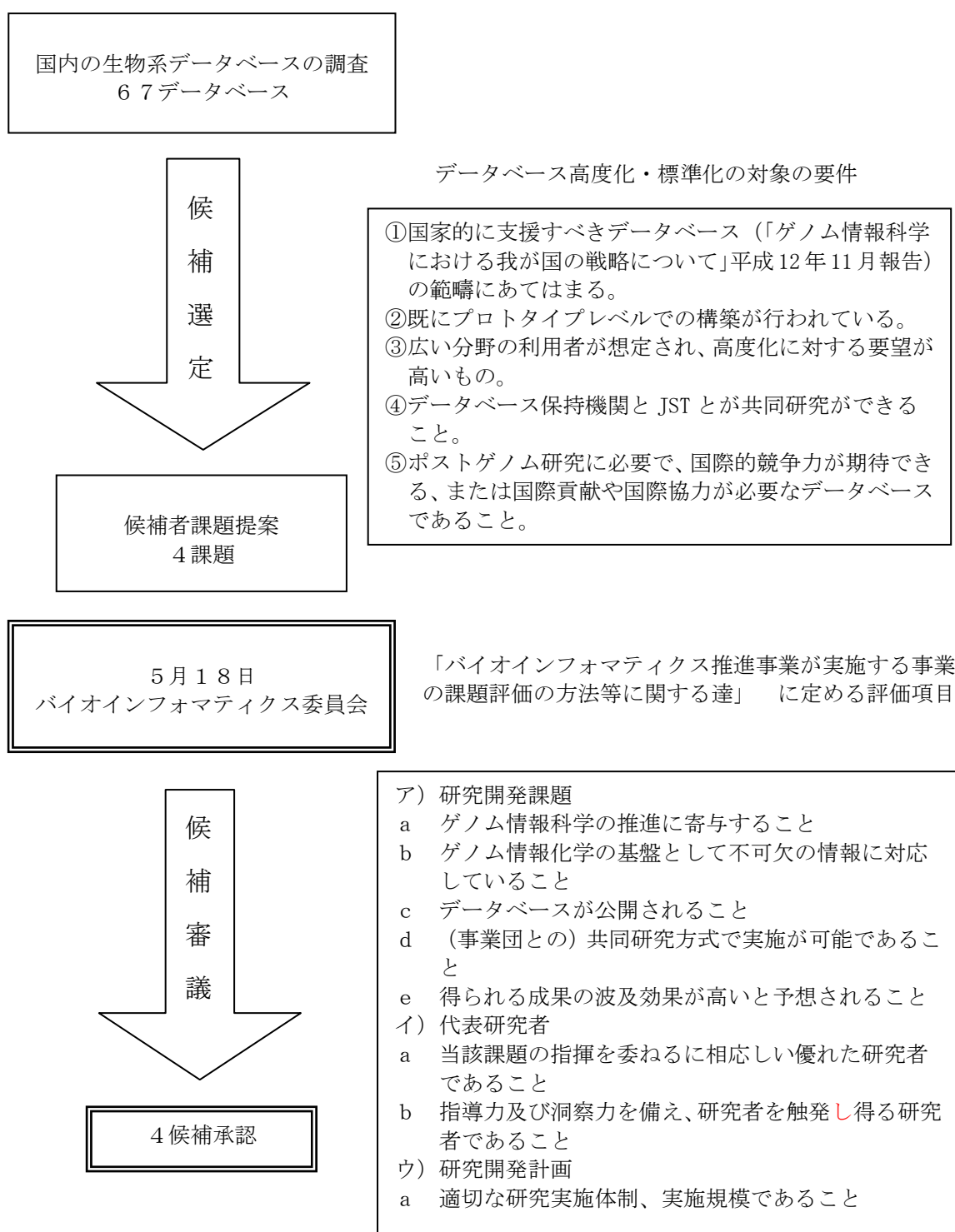
図表 III-2 第II期研究開発課題（研究開発期間：平成18～22年度）一覧

代表研究者（終了時所属）	研究開発課題名
稲葉 一男 （筑波大学下田臨海実験センター 教授）	ホヤプロテイン統合データベースの構築
金久 實 （京都大学化学研究所 教授）	ゲノムと環境の統合解析による生命システムの機能解読（KEGG）
菅原 秀明 （国立遺伝学研究所 教授）	バイオ基幹情報資源の高準化と共用化
高木 利久 （東京大学大学院新領域創成科学研究科 教授）	オントロジーによるパスウェイの高度化および国際標準化（INOHパスウェイデータベース）
中村 春木 （大阪大学蛋白質研究所 教授）	蛋白質構造データバンクの国際的な構築と高度化（PDBj）
西岡 孝明 （慶應義塾大学大学院 教授）	メタボロームMSスペクトル統合データベースの開発（Metabolome-Mass Spectral Database）
森下 真一 （東京大学大学院新領域創成科学研究科 教授）	マルチモーダル統合バイオDB（Multimodal BIODB）

1) 研究開発課題の選考基準と選考経過

第I期研究開発課題の選考基準と選考経過を図表 III-3 に示す。なお、選考候補の審議を行ったバイオインフォマティクス委員会委員についてはVI.資料編 第1章に示す。

図表 III-3 第I期研究開発課題の選考基準と選考過程



第 II 期研究開発課題の選考では、生命情報科学に不可欠で生命情報科学の高度化に資するデータベースの高度化・標準化（データ整備を含む）を対象とした研究開発（特に、国際共同運用データベースを始めとする継続運用が重視されるデータベースについて高度化・標準化を目指す研究開発）であることを必須とし、研究開発計画及び体制が整備されていることを条件とした。第 II 期の選考基準を図表 III-4 に示す。

募集は、膨大な生命情報からの有用知識の発見や、実験計画の立案等に不可欠なデータベースの構築、及び高度化を目指して研究開発を行うものを対象に、平成 17 年 10 月 31 日から 11 月 22 日まで行った。合計で 31 件の応募があり、統括がバイオインフォマティクス委員会の協力を得て 7 課題を選考した。

図表 III-4 第 II 期研究開発課題の選考基準

<p>1. 課題</p> <ul style="list-style-type: none">a 生命情報科学の推進に寄与すること。b 生命情報科学の基盤として不可欠の情報に対応していること。c 既に公開されており今後開発される部分も公開されること。d 得られる成果の波及効果が高いと予想されること。 <p>2. 代表研究者</p> <ul style="list-style-type: none">a 当該課題の指揮を委ねるに相応しい優れた研究者であること。b 指導力及び洞察力を備え、研究者を触発し得る研究者であること。 <p>3. 研究開発計画及び体制</p> <ul style="list-style-type: none">a 適切な研究実施体制、実施規模であること。b 所属機関等と JST が委託研究契約を締結して研究開発を実施することが可能であること。 <p>4. その他</p> <ul style="list-style-type: none">a 国際的競争力が期待できる、または国際貢献や国際協力が必要なデータベースであること。

2) 研究開発課題の成果

第 I 期及び第 II 期の研究開発課題の成果として、事後評価の総合評価と論文等の外部発表数を図表 III-5 に示す。総合評価の詳細を図表 III-6 に示した。なお、VI 資料編 2 章に参考資料がある。

図表 III-5 「生命情報データベースの高度化・標準化」研究課題の総合評価結果及び外部発表数

課題分類	開始年度	研究課題名	代表研究者	総合評価※1	外部発表数※2		
					論文発表数	口頭発表数	ポスター発表数
高度化・標準化	第 I 期 (平成 13～17 年度)	2 項関係に基づくゲノムと生命システムの機能解読	金久實	-	25	46	127
		ゲノム生物学バックボーンデータベースの構築提供	菅原秀明	-	58	77	44
		シグナルオントロジーとバイオタームバンクの開発	高木利久	-	23	20	11
		蛋白質立体構造データベースの高度化	中村春木	-	30	65	48
	第 II 期 (平成 18～22 年度)	ホヤプロテイン統合データベースの構築	稲葉一男	S	31	35	49
		ゲノムと環境の統合解析による生命システムの機能解読 (KEGG)	金久實	S	33	1	24
		バイオ基幹情報資源の高準化と共用化	菅原秀明	A	14	10	27
		オントロジーによるパスウェイの高度化および国際標準化 (INOH)	高木利久	A	9	5	8
		蛋白質構造データバンクの国際的な構築と高度化 (PDBj)	中村春木	S	59	34	70
		メタボロームMS スペクトル統合データベースの開発	西岡孝明	S	22	35	26
	マルチモーダル統合バイオ DB	森下真一	S	19	9	8	

※1 事後評価(あるいは追跡評価)における総合評価

S: 研究開発計画以上の成果が見られ、バイオインフォマティクス研究の進展に大きく貢献した

A: 研究開発計画に沿った成果が見られ、バイオインフォマティクス研究の進展に貢献した

B: 研究開発計画の一部を達成できなかった

C: バイオインフォマティクスの観点から見ると成果は不十分であり、バイオインフォマティクス研究の進展への貢献は今後の課題である

なお、第 I 期については研究終了時の総合評価方法が異なっていたため、S, A, B, C の判定結果の記載がない。該当箇所を「-」で示した。

※2 各代表研究者から提出された成果報告書に記載されている外部発表数(論文、口頭、ポスター)

図表 III-6 「生命情報データベースの高度化・標準化」研究課題の総合評価

課題分類	開始年度	研究課題名	代表研究者	総合評価
高度化・標準化	第Ⅰ期 (平成13～17年度)	2項関係に基づくゲノムと生命システムの機能解読	金久 實	<p>すでに日本を代表する国際的なデータベースとなっている KEGG の高度化・標準化の観点から着実な成果を出している。BRITE データベースを構築し、KEGG のネットワーク構造の階層に基づく高次機能の推論と、KEGG パスウェイでは表現できない知識のコンピュータ化を、階層テキストファイルという語彙の体系で実現したことは非常に意義が大きい。基礎知識の充実のみでなく、産業界への貢献も大いに期待できる。今後さらに研究開発体制を充実させ、日本を代表する国際的なデータベースとしてデータの信頼性の向上などを含めたさらなる発展を期待する。本研究開発の成果は国家戦略的に重要であり、継続して研究開発を進めていくべきものである。</p>
		ゲノム生物学バックボーンデータベースの構築提供	菅原 秀明	<p>三大国際DNAデータバンクのひとつであるDDBJの高度化として、4つのデータベースを構築し、データの品質の問題への取り組み、遺伝子発現データの基盤への取り組み、また、その利用について新しいインターフェイスを作り上げた点が評価できる。高品位データベースの成果から、一旦公開されたデータについても、定期的再評価が必要なことが明らかになったことは、科学研究において重要である。今後、DDBJにおける大規模計算機解析によるアノテーションと研究コミュニティ群による専門的アノテーションを融合する枠組みにより、遺伝子配列データとゲノム配列データのアノテーションが豊かになっていくこと、四次元データベースによって、個々の遺伝子の振舞いや遺伝子のネットワークの振舞いの解析が促進されていくことも期待する。本研究開発の成果であるデータベースとツールを活かしていくためには、研究者コミュニティとの連携を強める必要があるであろう。本研究開発はDDBJの活動とリンクして国際的な対応も行われているが、今後の国際社会への浸透を期待したい。通常業務であるDNAデータバンク構築を含め、更なる研究開発には人員、資金の両面から拡大の必要性があると思われる。国際的拠点として研究開発体制の整備を図り、継続的にDDBJの活動と研究開発に取り組んでいくことを期待する。</p>

高度化・標準化	第Ⅰ期 (平成13～ 17年度)	シグナルオントロジーとバイオタームバンクの開発	高木 利久	<p>挑戦的研究開発課題であり、我が国に欠けていたオントロジーに挑戦し、生物学文献で共有されるパスウェイ知識をデータベース化する情報処理技術、知識基盤の構築を実施したことは高く評価できる。オントロジーベースのパスウェイ検索システムは、ユーザが入力した検索文字列の意図をシステムが理解できる点でキーワードベースの検索と決定的に異なり独自性が高い。なお、マニュアルキュレーションは精度は高いが時間がかかる。この分野は進展が速いので、今後は体制を充実し、スピードアップをはかるべきであろう。本研究開発課題がパスウェイデータの国際標準化を目指す BioPAX パスウェイデータフォーマットの仕様策定メンバーとして日本から参加し、パスウェイデータベースが表現できるべき情報を決めるデータ・モデル部分の設計に積極的に関与していることは意義深い。本研究開発で新しい公共の知識基盤を創出することに成功したが、今後はこれらの高付加価値データを用いて、演繹的な推論、仮説生成に代表される高度な推論を実現することを期待する。</p>
		蛋白質立体構造データベースの高度化	中村 春木	<p>当初の想定を大きく上まわる、研究開発成果を上げた。世界に発信する我が国の蛋白質構造データベースの中核として、PDBjの設立と発展は大きな意義がある。特に当初想定になかったwwPDBの一員として国際的に大きな役割と責任をになうようになったことは、注目に値する。これらの活動を実現した代表研究者のマネジメント努力は高く評価できる。</p> <p>生体高分子の立体構造データは今後さらに増大すると予想される。データベースについて精度の高い品質管理を行うとともに、高度なデータ検索サービスや利用しやすいツールの開発を進め、利用しやすく信頼性の高い蛋白質の構造と機能についての情報源として国内外へ発信し、人材育成、普及・教育活動等も積極的に行って、さらなる社会への貢献を期待するものである。国内唯一の国際拠点である PDBj は国家戦略的に大変重要であり、さらに体制を充実させて他プロジェクトとの連携を深めながら継続して行うべき事業である。</p>
	第Ⅱ期 (平成18～ 22年度)	ホヤプロテイン統合データベースの構築	稲葉 一男	<p>研究開発計画以上の成果が見られ、ライフサイエンス分野の情報基盤整備に大きく貢献した。</p> <p>ホヤに関する有用なデータベース群が開発された。特に中心となるデータベースは国際標準と認められ、広く利用される発生学研究の大きな基盤が構築された。今後は、他の生物種の研究とより密接に連携するなどして、さらに多くの研究領域へ展開することが望まれる。</p>

高度化・標準化	第Ⅱ期 (平成 18～ 22 年度)	ゲノムと環境 の統合解析に よる生命シス テムの機能解 読 (KEGG)	金久 實	<p>研究開発計画以上の成果が見られ、ライフサイエンス分野の情報基盤整備に大きく貢献した。</p> <p>ゲノム情報を基盤とした生命システムと環境との相互作用のデータを統合した、わが国を代表する中核的なデータベースとなっている。新たに疾患や医薬品のデータを加えて統合したことで、ゲノム情報の医学や薬学への応用に向けて、KEGG の活用が期待できる。今後も拡大的継続が必要であり、どのようにして維持、高度化していくか、早急に検討することが必要である。</p>
		バイオ基幹情 報資源の高準 化と共用化	菅原 秀明	<p>研究開発計画通りの成果が見られ、ライフサイエンス分野の情報基盤整備に貢献した。</p> <p>Web サービスに基づくデータベースの標準的な検索インターフェースを整備し、実際の運用レベルで利用できる形で公開していることは意義がある。Web サービスとワークフローに関する研究は、DDBJ のサービスのレベル向上に寄与し、また新世代シーケンサーにも迅速に対応している。しかし、完成度を上げるには更なる努力が必要である。</p>
		オントロジー によるパスウ ェイの高度化 および国際標 準化 (INOH)	高木 利久	<p>研究開発計画通りの成果が見られ、ライフサイエンス分野の情報基盤整備に貢献した。</p> <p>知識処理との組み合わせが一部不十分であったものの、シグナル伝達についての複雑な知識を、推論可能な形で表現する方法を提案し、国際的なパスウェイデータの標準化に貢献した。階層的構造や新たなオントロジー表現を、人の手で読み込み入力していくという丁寧な方法を維持することによって、高品質のデータベースを構築した。このような方法により国際標準化の動きの中で構築したことは、バイオインフォマティクス推進という意味において評価される。</p>
		蛋白質構造デ ータバンクの 国際的な構築 と高度化 (PDBj)	中村 春木	<p>研究開発計画以上の成果が見られ、ライフサイエンス分野の情報基盤整備に大きく貢献した。</p> <p>タンパク質立体構造のデータベース基盤を、厳密なデータ品質の管理のもとに構築するなど、データベースの標準化に大きく貢献し、国際的な協体制を確立した。また、有用なツール群を開発し、さらにそれを用いて生物学的知識の発見を行ったことは計画以上の成果を出したと評価できる。将来後継者を含めどのような仕組みで維持、展開を図るかが今後のポイントとなる。</p>

高度化・標準化	第Ⅱ期 (平成 18～ 22 年度)	メタボローム MS スペクトル 統合データベ ースの開発	西岡 孝明	<p>研究開発計画以上の成果が見られ、ライフサイエンス分野の情報基盤整備に大きく貢献した。</p> <p>分散データベースを用いて学会などのコミュニティによりデータベースを維持する、というアイデアは新規で、かつ、それを広めようとしていることは高く評価できる。受け入れるデータの質のバラツキの問題、幅広いユーザーの要望にどのように答えていくか、日本の主要データベースとして、将来どのようなしくみで維持、展開を図るか、が今後の課題であろう。</p>
		マルチモーダ ル統合バイオ DB	森下 真一	<p>研究開発計画以上の成果が見られ、ライフサイエンス分野の情報基盤整備に大きく貢献した。</p> <p>超高速シーケンサーから得られる大量ゲノムデータを積極的に活用したマルチモーダルな統合検索により、新たな生物学的発見が可能であることを実証した。ライフサイエンスにおける新たな研究スタイルを確立したものであり、非常に高く評価できる。またドライとウェットの融合という BIRD 事業の理念目標が達成された。</p>

2節 創造的な生物・情報知識融合型の研究開発

情報科学と生物科学とを融合したアプローチにより、多彩な生物情報から生物現象の原理や法則を発見し体系化することを目指す研究開発を推進し、これらの成果がバイオインフォマティクスの発展及び新しい情報生物学創造のインセンティブとして機能し、これを基盤として新たな生物科学が樹立されるように支援した。このために、新規採択は、平成 13、17、18、19 年度に行い、それぞれ 3 年間の研究開発を支援するとともに、継続評価することにより 2 年間の継続研究開発を支援した。各研究開発課題の名称と代表研究者を図表 III-7～図表 III-13 に示す。

図表 III-7 平成 13 年度採択 新規研究開発課題一覧

(研究開発期間：平成 13 年 10 月～16 年 9 月)

代表研究者 (終了時所属)	研究開発課題名
伊藤 啓 (東京大学分子細胞生物学研究所 助教授) ※確認中 ()	ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築
伊藤 隆司 (東京大学大学院新領域創成科学研究科 教授) ※吉田美寸夫 (インテック W&G グループサブマネージャー)	インタラクティブ解析からの生物知識獲得
大浪 修一 (慶應義塾大学大学院理工学研究科 助教授) ※北野 宏明 (慶應義塾大学大学院理工学研究科 客員教授)	線虫 <i>C. elegans</i> 発生過程のシステム解析
岸野 洋久 (東京大学大学院農学生命科学研究科 教授)	ゲノム進化とマッピングの階層モデルと解析アルゴリズムの開発
肥後 順一 (東京薬科大学 生命科学部 教授)	高速計算機システムによる蛋白質フォールディングの研究
森下 真一 (東京大学大学院新領域創成科学研究科 教授) ※大矢 禎一 (東京大学大学院新領域創成科学研究科教授)	遺伝子破壊株イメージ・マイニング
矢田 哲士 (京都大学大学院情報学研究科 助教授)	ヒト遺伝子の転写・発現の多様性解明を目指した基盤データベースの開発

※情報科学と生物科学との融合型アプローチによる研究開発の要件、即ち「生物学者と情報学者の協力」において BIRD 代表研究者と協力して研究開発を推進した研究者 (4 課題)

図表 III-8 平成 16 年度採択 継続研究開発課題一覧

(研究開発期間：平成 16 年 10 月～18 年 9 月)

代表研究者（終了時所属）	研究開発課題名
伊藤 啓 (東京大学分子細胞生物学研究所 助教授)	ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築
伊藤 隆司 (東京大学大学院新領域創成科学研究科 教授)	絶対定量オーミックスからの知識発見
森下 真一 (東京大学大学院新領域創成科学研究科 教授)	遺伝子破壊株イメージ・マイニング
矢田 哲士 (京都大学大学院情報学研究科 助教授)	ヒトゲノムにおける広義の遺伝子発見研究

図表 III-9 平成 17 年度採択 新規研究開発課題一覧

(研究開発期間：平成 17 年 10 月～20 年 9 月)

代表研究者（終了時所属）	研究開発課題名
内山 郁夫 (自然科学研究機構基礎生物学研究所 助手)	大規模な比較ゲノム研究を展開するためのワークベンチの構築
太田 元規 (東京工業大学学術国際情報センター 助教授)	タンパク質の構造・機能予測法の開発とヒトゲノム配列への適用
川戸 佳 (東京大学大学院総合文化研究科 教授)	脳スライス中で可視化した神経シナプスの自動解析
塩田 浩平 (京都大学大学院医学研究科 教授)	ヒト胚の形態発生に関する三次元データベース
白井 剛 (長浜バイオ大学バイオサイエンス学部 教授)	実践による超分子複合体モデリングシステムの開発
宮川 剛 (京都大学大学院医学研究科 助教授)	マウスを用いた脳機能表現型データベースの開発

図表 III-10 平成 18 年度採択 新規研究開発課題一覧

(研究開発期間：平成 18 年 11 月～21 年 9 月)

代表研究者（終了時所属）	研究開発課題名
石井 信 (奈良先端科学技術大学院大学情報科学研究科 教授)	予測技術を用いた生命システムの同定手法の開発
黒川 顕 (奈良先端科学技術大学院大学情報科学研究科 助教授)	メタゲノムオーソログ遺伝子統合解析システムの開発
榊原 康文 (慶應義塾大学理工学部 教授)	タンパク質化合物相互作用の網羅的予測手法とデータベースの開発
清水 謙多郎 (東京大学大学院農学生命科学研究科 教授)	高精度タンパク質間相互作用予測システムの開発
中井 謙太 (東京大学医科学研究所 教授)	転写制御領域の構築原理解明

図表 III-11 平成 19 年度採択 新規研究開発課題一覧

(研究開発期間：平成 19 年 10 月～22 年 9 月)

代表研究者（終了時所属）	研究開発課題名
長野 希美 ((独) 産業技術総合研究所 生命情報工学研究センター 主任研究員)	酵素反応分類に基づく酵素反応予測システムの開発
馳澤 盛一郎 (東京大学大学院新領域創成科学研究科 教授)	進化型計算と自己組織化による適応的画像分類法の開発
馬見塚 拓 (京都大学化学研究所 教授)	生命科学上の非構造化データの統合マイニング
水口 賢司 ((独) 医薬基盤研究所 バイオインフォマティクスプロジェクト プロジェクトリーダー)	ダイナミクスを考慮した膜蛋白質の構造モデリング法の開発

図表 III-12 平成 20 年度採択 継続研究開発課題一覧

(研究開発期間：平成 20 年 10 月～22 年 9 月)

代表研究者（終了時所属）	研究開発課題名
内山 郁夫 (自然科学研究機構 基礎生物学研究所 助教)	大規模比較ゲノムワークベンチの実践的応用に向けた研究開発
太田 元規 (名古屋大学 大学院情報科学研究科 教授)	タンパク質の構造・機能・相互作用予測システムの開発と展開
塩田 浩平 (京都大学 大学院医学研究科 教授)	ヒト胚の形態発生に関する三次元データベース
白井 剛 (長浜バイオ大学 バイオサイエンス学部 教授)	実践による超分子ネットワークモデリングシステムの開発

図表 III-13 平成 21 年度採択 継続研究開発課題一覧

(研究開発期間：平成 21 年 10 月～23 年 9 月)

代表研究者（終了時所属）	研究開発課題名
黒川 顕 (東京工業大学大学院生命理工学研究科 教授)	メタゲノム統合解析システムの開発
榊原 康文 (慶應義塾大学理工学部 教授)	タンパク質化合物相互作用の網羅的予測手法とデータベースの開発
清水 謙多郎 (東京大学大学院農学生命科学研究科 教授)	高精度タンパク質間相互作用予測システムの開発

1) 研究開発課題の選考基準と選考経過

平成13年度、17年度、18年度、19年度の新規課題と、16年度、20年度、21年度の継続課題の選考基準についてはそれぞれほぼ同様であることから、ここでは、平成13年度の新規課題の選考基準と、平成16年度の継続課題の選考基準だけを図表 III-14、図表 III-15 に示す。また、各年度の選考経過を図表 III-16 にまとめる。

図表 III-14 平成13年度新規研究開発課題の選考基準

課題選考にあたっては、ゲノム情報科学の推進に寄与すること、創造的で新しいアルゴリズムや仮説提案につながる可能性を有していること、特定分野の研究開発であることを必須とし、研究開発計画及び体制が整備されていることを条件とします。

1. 課題

- a ゲノム情報科学の推進に寄与すること
- b 創造的で新しいアルゴリズムや仮説提案につながる可能性を有していること
- c 研究成果は公開されること
- d 創造的情報生物学の開発を志すこと
- e 研究開発に適時性があること

2. 代表研究者

- a 自らが創造的な研究構想の発案者であること
- b 活力、統率力を有すること

3. 研究開発計画及び体制

- ・研究開発の目標設定が具体的であること。
- ・3年間で所要の成果を達成できること、あるいは、達成する見通しが得られることを基本とし、委員会が必要性を認めた場合、2年延長し得る。
- ・実施体制が整備されていること。

4. その他の条件

- ・同じ時期に、代表研究者は国や特殊法人（科学技術振興事業団、日本学術振興会、新エネルギー・産業技術総合開発機構（NEDO）、情報処理振興事業協会（IPA）、生物系特定産業技術研究推進機構（BRAIN）、医薬品副作用被害救済・研究振興調査機構（医薬品機構）、通信・放送機構（TAO）、運輸施設整備事業団等）が推進する大型の研究開発事業の課題の代表者になっていないこと。
- ・平成13年6月に公募する科学技術振興事業団の戦略的基礎研究推進事業（CREST）、若手研究者研究推進事業（TOREST）及び計算科学技術活用型特定研究開発推進事業（ACT-JST）に同時に応募しないこと。

図表 III-15 平成 16 年度継続研究開発課題の選考基準

I. バイオインフォマティクス推進事業での研究開発実績	
①	当事業で研究開発を始めてからこれまでにきわめて優れた成果が得られているか、または当該研究開発期間内に得られる見通しがあるか。
②	代表研究者が自らの構想実現のために、適切な研究開発実施体制をとってきたか。
II. 研究開発継続後の発展性	
①	研究開発目標・計画が具体的であり、研究開発終了後にきわめて優れた成果が得られる見通しがあるか。
②	継続することにより、バイオインフォマティクス領域はもとよりライフサイエンス分野の発展・展開に対して革新的なインパクトをもたらすことができるか。

図表 III-16 「創造的な生物・情報知識融合型の研究開発」研究開発課題の選考経過

年度	選考経過
13 年度 (新規)	平成 13 年 6 月 20 日から 7 月 31 日まで募集し、101 件の応募の中から、統括がバイオインフォマティクス委員会の協力を得て 7 課題を選考した。
16 年度 (継続)	平成 13 年度公募採択研究開発課題(7 課題)のうち、上記選考基準を満たす 4 課題を選考した。
17 年度 (新規)	募集は、生物科学情報から生物現象の原理や法則を発見し体系化するために、情報科学と生物科学が融合して研究開発を行うものを対象に、平成 17 年 3 月 22 日から 5 月 16 日まで行った。合計で 113 件の応募があり、統括がバイオインフォマティクス委員会の協力を得て 6 課題を選考した。
18 年度 (新規)	募集は、ライフサイエンス分野における大量、複雑化するデータを対象とし、知識を発見する新しい数学的な理論および情報科学の要素技術の開発及び効果的な理解と発想を促すような表現技術など生物情報データベースを活用する情報科学技術を対象に、平成 18 年 8 月 1 日から 9 月 12 日まで行った。合計で 73 件の応募があり、統括がバイオインフォマティクス委員会の協力を得て 5 課題を選考した。
19 年度 (新規)	募集は、ライフサイエンス分野における大量、複雑化するデータを対象とし、知識を発見する新しい数学的な理論および情報科学の要素技術の開発及び効果的な理解と発想を促すような表現技術など生物情報データベースを活用する情報科学技術を対象に、平成 19 年 6 月 14 日から 7 月 31 日まで行った。合計で 45 件の応募があり、統括がバイオインフォマティクス委員会の協力を得て 4 課題を選考した。
20 年度 (継続)	平成 17 年度公募採択研究開発課題(6 課題)で、当初の研究開発期間内にきわめて優れた成果が得られる見込みがあり、研究開発を継続することでさらに発展することが期待される課題として、4 課題を選考した。
21 年度 (継続)	平成 18 年度公募採択研究開発課題(5 課題)で、当初の研究開発期間内にきわめて優れた成果が得られる見込みがあり、研究開発を継続することでさらに発展することが期待される課題として、3 課題を選考した。

2) 研究開発課題の成果

平成 13～21 年度開始の研究開発課題の成果として、事後評価の総合評価と論文等の外部発表数を図表 III-17 に示す。総合評価の詳細を図表 III-18 に示した。なお、VI 資料編 2 章に参考資料がある。

図表 III-17 「創造的な生物・情報知識融合型の研究開発」研究課題の総合評価結果及び外部発表数

課題分類	開始年度	研究課題名※1	代表研究者	総合評価※2	外部発表数※3		
					論文発表数	口頭発表数	ポスター発表数
生物・情報知識融合型	H13	ゲノム進化とマッピングの階層モデルと解析アルゴリズムの開発	岸野洋久	-	37	-	-
	H13	高速計算機システムによる蛋白質フォールディングの研究	肥後順一	-	17	-	-
	H13 H16	ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築	伊藤啓	-	7	-	-
		ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築		-	5	16	24
	H13 H16	インタラクトーム解析からの生物知識獲得	伊藤隆司	-	17	-	-
		絶対定量オーミックスからの知識発見		-	5	16	24
	H13 H16	遺伝子破壊株イメージ・マイニング	森下真一	-	20	-	-
		遺伝子破壊株イメージ・マイニング		-	0	3	19
	H13 H16	ヒト遺伝子の転写・発現の多様性解明を目指した基盤データベースの開発	矢田哲士	-	10	-	-
		ヒトゲノムにおける広義の遺伝子発見研究		-	10	2	15
H16	線虫 <i>C. elegans</i> 発生過程のシステム解析	大浪修一	-	2	-	-	
H17	マウスを用いた脳機能表現型データベースの開発	宮川剛	-	38	21	37	
H17	脳スライス中で可視化した神経シナプスの自動解析	川戸佳	A	69	12	67	
H17 H20	大規模な比較ゲノム研究を展開するためのワークベンチの構築	内山郁夫	A	4	2	7	
	大規模比較ゲノムワークベンチの実践的応用に向けた研究開発		A	8	20	26	

課題 分類	開始 年度	研究課題名※1	代表 研究者	総合評価※2	外部発表数※3		
					論文 発表数	口頭 発表数	ポスター 発表数
	H17 H20	タンパク質の構造・機能予測法の開発とヒトゲノム配列への適用	太田元規	A	53	5	27
		タンパク質の構造・機能・相互作用予測システムの開発と展開		S	28	53	50
	H17 H20	ヒト胚の形態発生に関する三次元データベース	塩田浩平	A	15	12	11
		ヒト胚の形態発生に関する三次元データベース		S	11	4	16
	H17 H20	実践による超分子複合体モデリングシステムの開発	白井剛	A	26	2	62
		実践による超分子ネットワークモデリングシステムの開発		A	32	19	62
	H18	予測技術を用いた生命システムの同定手法の開発	石井信	B	11	6	7
	H18 H21	メタゲノムオーソログ遺伝子統合解析システムの開発	黒川顕	S	10	1	13
		メタゲノム統合解析システムの開発		S	13	7	3
	H18	転写制御領域の構築原理解明	中井謙太	B	19	15	26
	H18 H21	タンパク質化合物相互作用の網羅的予測手法とデータベースの開発	榊原康文	B	10	6	1
		タンパク質化合物相互作用の網羅的予測手法とデータベースの開発		A	10	1	1
	H18 H21	高精度タンパク質間相互作用予測システムの開発	清水謙多郎	B	23	10	33
高精度タンパク質間相互作用予測システムの開発		A		19	16	31	
H19	酵素反応分類に基づく酵素反応予測システムの開発	長野希美	A	9	4	15	
H19	進化型計算と自己組織化による適応的画像分類法の開発	馳澤盛一郎	A	41	44	59	
H19	生命科学上の非構造化データの統合マイニング	馬見塚拓	A	22	22	17	
H19	ダイナミクスを考慮した膜蛋白質の構造モデリング手法の開発	水口賢司	A	18	8	22	

※1 継続課題については1つにまとめて表現している

※2 事後評価(あるいは追跡評価)における総合評価

S: 研究開発計画以上の成果が見られ、バイオインフォマティクス研究の進展に大きく貢献した

A: 研究開発計画に沿った成果が見られ、バイオインフォマティクス研究の進展に貢献した

B: 研究開発計画の一部を達成できなかった

C: バイオインフォマティクスの観点から見ると成果は不十分であり、バイオインフォマティクス研究の進展への貢献は今後の課題である

※3 各代表研究者から提出された成果報告書に記載されている外部発表数(論文、口頭、ポスター)

※2, 3 H13・H16及びH17の一部の課題については事後評価の記載がそれ以降とは異なっていたため、S, A, B, Cの判定結果と口頭・ポスター発表数の記載がない。

該当箇所を「-」で示した。

図表 III-18 「創造的な生物・情報知識融合型の研究開発」研究課題の総合評価

課題分類	開始年度	研究課題名	代表研究者	総合評価
生物・情報 知識融合 型	H13	ゲノム進化とマッピングの階層モデルと解析アルゴリズムの開発	岸野 洋久	バクテリア、ウイルスから動植物のような多細胞生物まで多様な生物について、ゲノムやタンパク質といった分子進化の数々の理論を着実に構築し、30を超える原著論文を公開したことは大きな進展といえる。今後はこうした理論に基づいたソフトウェアやデータベースが実験生物学にフィードバックされ、ライフサイエンスに大きなインパクトを与えるものへ成長することを期待する。
	H13	高速計算機システムによる蛋白質フォールディングの研究	肥後 順一	代表研究者らは蛋白質のフォールディングシミュレーションにマルチカノニカル分子動力学法を取り入れ、力場の正確さと水分子の存在を厳密に考慮する計算に挑戦して一定の成果を修めた。またそこから派生した蛋白質のリガンドの会合シミュレーションにおいて、分子が離れた状態でも会合を促す長距離力が働いていることを示し、当初想定した以上の成果をあげた。本課題は目標自体が非常に難題ではあるが、研究の進展は学術界だけでなく創薬といった産業界にも波及効果が大きいと思われるので、計算機システムとアルゴリズムの両手法の改良による飛躍を期待したい。
	H13	ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築	伊藤 啓	代表研究者らが作製した「GAL4 エンハンサートラップシステム」を用いて、脳の神経細胞の狙った部分だけを可視化する技術の開発や、脳細胞の位置と数の定量データを抽出する技術の開発により、視覚系に加えて嗅覚系、聴覚系など詳細な脳内地図のデータベース化に成功している。この成果はショウジョウバエの神経生物学に対して大きな影響を与え始めているが、脳神経科学全体への寄与はまだこれからであろうと思われる。本課題における情報科学的なアプローチとしては、神経回路の実験的計測を支援する方法の確立に重きを置かれた感があるが、今後は代表研究者の構想にあったように神経回路の持つべき構造および機能を情報科学的に推理する手法の確立を行うことが期待される。
	H16	ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築	伊藤 啓	難しいテーマによく考えた研究戦略で挑戦し、一部達成できなかった点はあるものの神経回路の動作モデルを構築するのに必要な基盤となるデータを集積できたことは大きな意義がある。ただ、インフォマティクスは実験支援として本課題の一部を担うに留まり、成果物の公開も遅れている現状では本事業の趣旨にかなった成果を広く世界に知らしめたとは言い難い。今後は成果のデータベース化と早期公開を強く希望すると同時に生物学的機能データが加えられることを期待する。
	H13	インタラクトーム解析からの生物知識獲得	伊藤 隆司	遺伝子発現を絶対定量化する手法として GATC-PCR 法を開発したことや定量実験を支援するシステム AQUOS を開発したことは、体系的な生物学的解析と情報科学的解析の連携が有効に働いたことを示すものであり優れた成果といえる。今後は、この絶対定量値を生かした蛋白質-核酸相互作用や難易度の高い蛋白質-脂質相互作用のインタラクトーム解析などにより得られる実験データを、その質も吟味しつつデータベース化し、新たな生物知識の発見がなされることを期待する。

生物・情報 知識融合 型	H16	絶対定量オーミックスからの知識発見	伊藤 隆司	実験とインフォマティクスの融合という点で、これからの生命科学のパイロット的な研究であったと思う。その結果として予定の計画を十分にクリアし、新しい問題提起にもつながった。本解析技術やデータは基盤となるものであり、今後の発展とその支援をどうしていくかが評価者と被評価者、両者にとっての課題である。また、直接の目的ではなかったが情報科学と生物科学の融合に向けた人材が育ったことは評価される。
	H13	遺伝子破壊株イメージ・マイニング	森下 真一	出芽酵母遺伝子破壊株のイメージ取得の方法論とイメージ取得アルゴリズムの精度は当初の想定より高度なものとなった。形としての表現型を定量化する試みはバイオインフォマティクスに対する重要な貢献であり、公開された酵母の形態データベースは酵母研究者のサークルでは十分に役立つ段階に達している。今後このデータが生物学にフィードバックされ、遺伝子発現との関連づけで、この手法だからこそ発見に至るという成果にまで達することが期待される。また多細胞生物についてもこの手法が活かされればライフサイエンス分野全体への波及効果が大きいと思われる。
	H16	遺伝子破壊株イメージ・マイニング	森下 真一	従来にはない特徴のある研究アプローチを示し、フェノーム研究の一典型としての評価をうけている。また必須遺伝子の機能予測についても高い可能性を示したことは、極めて意義が大きい。一方、形態と遺伝子のギャップは未だ存在し、そこをうめる必要があると思われる。今後も共同研究など外部との更なる協調により多細胞生物への展開も含めて研究が発展するよう期待する。
	H13	ヒト遺伝子の転写・発現の多様性解明を目指した基盤データベースの開発	矢田 哲士	転写産物を利用したものやab initio 法によるもの、ゲノム配列比較によるものなど世界的に見ても優れた性能を示す遺伝子発見プログラムを開発した意義は大きい。またいくつかの遺伝子発見プログラムやこれらにより予測されたヒト遺伝子候補がデータベース化されて公開されており、様々な生物種のゲノム解読が完了しつつある現在においてライフサイエンス分野の研究者に対しても本研究開発の貢献が見込まれる。本課題ではゲノム配列解析を基盤とした遺伝子予測が進められたが、今後は違った観点からの遺伝子発見の挑戦も期待される。
	H16	ヒトゲノムにおける広義の遺伝子発見研究	矢田 哲士	ヒトの遺伝子発見研究に対して独自の研究アプローチを示せたことは意義がある。しかしながらプロジェクト内の各課題は個別に羅列されたままになっており、目標に向けた体系的な集中が欲しかったのも事実である。また機能性 RNA 遺伝子探索については時間的なことのみを問題点としてあげてあるが、問題の本質は類縁 RNA 探索という考え方にあり、それだけでは非コード遺伝子の解析としては不十分であることを認識するべきであろう。遺伝子予測は古くからのテーマであるが、今も決定版のない新しいテーマである。今度このシステムを使って生物学的な発見がなされることを期待する。
	H16	線虫 <i>C. elegans</i> 発生過程のシステム解析	大浪 修一	細胞分裂パターン自動計測装置の改良や遺伝子機能を阻害した胚の細胞分裂パターンの解析など当初計画されていた細胞期までの計測や、解析すべき遺伝子の数まで到達しなかったものもあるが、研究開発実施場所の整備の遅れを考えればやむを得ない結果といえる。その中でも大規模遺伝子発現量データより遺伝子ネットワークを導出する DBRF-MEGN 法の開発は、論文掲載や特許出願にも結びついており今後のライフサイエンス分野への波及効果に期待がもてる。また線虫の形態形成のデータベース化、その過程での核や細胞内装置の局在についてのデータベース化などは関連分野の今後の研究に大きな影響を及ぼすと思われる。今後は安定した形で研究実施場所を構えて中期展望を持った研究を行ってほしい。

生物・情報 知識融合 型	H17	マウスを用いた脳機能表現型データベースの開発	宮川 剛	疾患に関わる遺伝子について新しい発見などがなされているものの、バイオインフォマティクスの観点から見ると成果は不十分であり、バイオインフォマティクス研究の進展への貢献は今後の課題である。20 系統の遺伝子改変マウスについては、既にデータベースに一般公開されているが、本事業で行った成果としては、より多くの系統データを取得し速やかに公開していく必要がある。今後、標準化されたプロトコルとデータベースを基盤として、国内外のできるだけ多くのマウス表現型解析拠点と連携し、データベースの量的な充実を図る必要があると考えられる。
	H17	脳スライス中で可視化した神経シナプスの自動解析	川戸 佳	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。画像処理の技術により、構造的変化と生物現象との関係を結びつけることに成功した。多数のデータを使って、その関係性を調べる際に有効な汎用性のプログラムが開発されたという点で評価される。
	H17	大規模な比較ゲノム研究を展開するためのワークベンチの構築	内山 郁夫	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。日本発の大規模微生物ゲノム比較データベースとしての存在価値は大きい。微生物ゲノムについては、次世代シーケンサーによるデータ、メタゲノムデータの大幅な増加が見込まれる。比較ゲノムを維持していくことがきわめて重要である。
	H20	大規模比較ゲノムワークベンチの実践的応用に向けた研究開発	内山 郁夫	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。大規模なゲノム比較のためのワークベンチとして開発されたシステムが、近縁ゲノムの比較解析による個別の株の特徴付けや、メタゲノム解析にも適用できるという発展を見せるなど、多方面の応用につながった。ゲノム科学の進展に大きく寄与できるデータベースを構築できた。ただし、最も進展が速い分野と直結しており、今後のケアが重要となる。
	H17	タンパク質の構造・機能予測法の開発とヒトゲノム配列への適用	太田 元規	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。タンパク質の構造予測や相互作用等、焦点を絞った研究開発により、従来よりも高精度の予測システムを構築したことは評価される。今後、実験検証や利用者からのフィードバックが行われ、開発されたツールやデータベースがよりよいものとなっていくことを期待する。
	H20	タンパク質の構造・機能・相互作用予測システムの開発と展開	太田 元規	研究開発計画以上の成果が見られ、バイオインフォマティクス研究の進展に大きく貢献した。タンパク質の複合体構造を考慮した、構造や機能についての種々の特性を検討し、構造・機能の予測法の開発につなげており、しかも多数のタンパク質に適用可能な性能を得ている点が評価できる。独創的解析手法の開発や、地道なデータベース開発の両者ともバイオインフォマティクス研究の進展に貢献したと考えられる。今後の SAHG データベースの充実を期待する。
	H17	ヒト胚の形態発生に関する三次元データベース	塩田 浩平	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。国際的にも他に例のない新規性の高い3次元形態データベースが構築され、高解像度のイメージング技術により得られたデジタルデータやグラフィックスモデルが利用できるようになったことから臨床応用など発展が期待される。

生物・情報 知識融合 型	H20	ヒト胚の形態発生に関する三次元データベース	塩田 浩平	研究開発計画以上の成果が見られ、バイオインフォマティクス研究の進展に大きく貢献した。世界に類例のない規模のヒト胚三次元形態画像データベースを構築しており、バイオインフォマティクス研究への寄与というより、生命科学に広く寄与する重要なデータベースである。公開のための条件が順調に整備され、来年4月に予定通り公開されることを強く期待している。
	H17	実践による超分子複合体モデリングシステムの開発	白井 剛	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。バイオインフォマティクスと実験解析をうまく組み合わせることにより、実際に複数の複合体のモデリングに成功しており、研究成果自体も意義が大きい。開発されたデータベースやツールがより普及することが望ましい。今後の構造データの急激な増加にも対応できるだけの高性能な自動解析方法が確立できれば大きな意義がある。
	H20	実践による超分子ネットワークモデリングシステムの開発	白井 剛	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。構造生物学として、動的な視点から機能生物学への説明を忘れていないことが評価できる。データベースに蓄積された部品の情報から、超分子複合体を構築するという独自性の高いテーマに取り組み、結合部位の予測に発展させている。バイオインフォマティクス技術や開発したツールを用いた実験系研究者との共同研究は順調に進展し多くの成果が得られたこと、モデリングシステムの開発ばかりでなく、その検証のために行われた実験それ自体からも重要な新しい知見が得られていることも評価できる。評価者全員が、計画以上の成果にあと一步であるとの高い評価であった。
	H18	予測技術を用いた生命システムの同定手法の開発	石井 信	研究開発計画に沿った成果が見られ、バイオインフォマティクス研究の進展に貢献した。モデルと実験との融合的な手法による体節および極性におけるパターン形成において新しい仮説の提案に至るなどの展開あり、システム生物学の研究の一つのモデルケースとなると考えられる。より普遍的な方法論として確立できるかが今後の課題である。
	H18	メタゲノムオーソログ遺伝子統合解析システムの開発	黒川 顕	研究開発計画以上の成果が見られ、バイオインフォマティクス研究の進展に大きく貢献した。日本におけるメタゲノム研究の基盤を構築したという面で本研究は高く評価できる。本研究で開発されたシステムは未知の細菌の発見や新しい遺伝子資源探索の手段としての可能性も持つため、今後も引き続き新型シーケンサーに対処するためのシステムの改良を行い、メタゲノム研究における貢献度をさらに増していくことを期待する。
	H21	メタゲノム統合解析システムの開発	黒川 顕	研究開発計画以上の成果が見られ、バイオインフォマティクス研究の進展に大きく貢献した。日本におけるメタゲノム解析の代表的な研究としての位置付けを確立している。メタゲノム情報解析のための有用かつ独自性のあるソフトウェアツール・データベースが構築され、幅広く利用されている。また、それらを利用して実際のデータを解析した結果、有用な生物学的知見も得られた。しかし、これらのデータからどれだけ細菌叢の実態について予見ができるかについては、さらなる実験的検証が必要である。ここで行われたメタゲノム解析は国際的に最先端の研究であり、今後の展開が期待される。
	H18	転写制御領域の構築原理解明	中井 謙太	研究開発計画に沿った成果が見られ、バイオインフォマティクス研究の進展に貢献した。有用なデータベースの開発・更新が行われるなど、研究基盤としての貢献度は十分にあると考えられる。設定した3本柱の成果が互いにフィードバックしあう状態には至らなかったことが惜しまれる。

生物・情報 知識融合 型	H18	タンパク質化合物相互作用の網羅的予測手法とデータベースの開発	榊原 康文	研究開発計画に沿った成果が見られ、バイオインフォマティクス研究の進展に貢献した。高精度、高速性を実現した予測システムを開発しただけでなく、実際にそれを用いて新規リード化合物を予測し、それを実験により確認できたことは高く評価できる。さらなる精度向上のため、見落とししているファクターがないか検討いただきたい。今後は、実用化について企業等を含めて発展させて欲しい。また、継続的開発と啓発活動によって、成果が普及されることを期待したい。
	H21	タンパク質化合物相互作用の網羅的予測手法とデータベースの開発	榊原 康文	研究開発計画通りの成果が見られ、ライフサイエンス分野の情報基盤整備に貢献した。大規模な化合物データが取り扱える網羅的な予測方法で、しかも、実験的な検証に耐えうる相互作用予測方法の開発が行われたことは高く評価される。しかし、予測における拘束条件について今一度検討し、データの取りこぼしおよびノイズの軽減に注力すれば、より高度な知識発見の可能性があると考えられ、今後のさらなる発展が期待される。
	H18	高精度タンパク質間相互作用予測システムの開発	清水 謙多郎	研究開発計画に沿った成果が見られ、バイオインフォマティクス研究の進展に貢献した。従来の手法よりも大幅に高速化された新規のドッキング予測手法の開発や、当初の計画になかったタンパク質間相互作用予測システム、タンパク質-糖鎖結合予測システムの開発が行われたことから、バイオインフォマティクス研究の進展に貢献したと言える。良い成果が得られているだけに、研究成果が十分に周知されていないのは惜しい。今後、酵素の設計に役立てるなど実験へのフィードバックが期待される。
	H21	高精度タンパク質相互作用予測システムの開発	清水 謙多郎	研究開発計画通りの成果が見られ、ライフサイエンス分野の情報基盤整備に貢献した。粗視化モデルを導入した分子動力学シミュレーションや、糖鎖との相互作用など、タンパク質に関わる独自の相互作用予測方法を開発し、高い精度を達成している点が評価できる。
	H19	酵素反応分類に基づく酵素反応予測しすてむの開発	長野 希美	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。派手さはないが地道で着実な研究が行われ、かつ、新たな解析手法も提案された。予測システムの公開が行われ、今後、多くの研究者の利用に供し、有効性の検証が行なわれることが望まれる。今後重要な未知の反応が予測できるかが注目される。
	H19	進化型計算と自己組織化による適応的画像分類法の開発	馳澤 盛一郎	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。本研究は、画像から測定した特徴量を選別し、それにより画像の分類を高精度に行う方法を開発した。汎用性の高い実用的な技術が開発されている点が高く評価できる。しかしながら、情報科学的観点からは、既存ツールと比較してどの程度の優位性があるか不明確であり、今後は開発した手法の更なる改善が期待される。
	H19	生命科学上の非構造化データの統合マイニング	馬見塚 拓	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。多くの問題に適用可能な、独自のマイニング方法を多数提案し、それらが高い性能を発揮できることを示した点が評価できる。実際に生命科学データからの知識発見にどの程度貢献するかは、今後に期待したい。
	H19	ダイナミクスを考慮した膜蛋白質の構造モデリング手法の開発	水口 賢司	研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。創薬をはじめとするライフサイエンス分野の今後の研究開発に貢献しうる知見を得、データベースやウェブアプリケーションといった研究基盤を開発した。膜タンパク質のダイナミクス解析という今後重要性が増す問題に対し様々な角度から研究が行われ、有用な成果が得られた点は高く評価できる。

3節 特筆すべき成果

BIRD 事業の中で特に優れた成果をあげている研究課題を選出し、その成果を示すのに最適な指標を設定した上で、その指標を用いて選定課題の成果を評価する。

1) 評価対象とする研究課題の選定

「BIRD 事業の課題評価の方法等に関する達」の評価項目を参考にし、そのうち、図表 III-5 及び図表 III-17 で示した「総合評価」と「外部発表(論文等)等研究開発成果の状況」の2項目を選定基準とした。その結果、総合評価がSでかつ外部発表のうち論文発表数が特に多い図表 III-19 及び図表 III-20 の5つの課題を選定した。なお、「マルチモーダル統合バイオDB」については総合評価がSの課題の中で、論文発表数は他と比べて必ずしも多くないが、他の多くの研究課題がDBの新規構築や高度化をテーマとしているのに対して本課題は今後のDB統合化の基盤となるテーマであることから特別に選出した。

図表 III-19 「生命情報データベースの高度化・標準化」で選定した研究開発課題

課題分類	開始年度	研究課題名	代表研究者
高度化・標準化	第II期 (H18~22)	ゲノムと環境の統合解析による生命システムの機能解読 (KEGG)	金久實
		蛋白質構造データバンクの国際的な構築と高度化 (PDBj)	中村春木
		マルチモーダル統合バイオDB	森下真一

図表 III-20 「創造的な生物・情報知識融合型の研究開発」で選定した研究開発課題

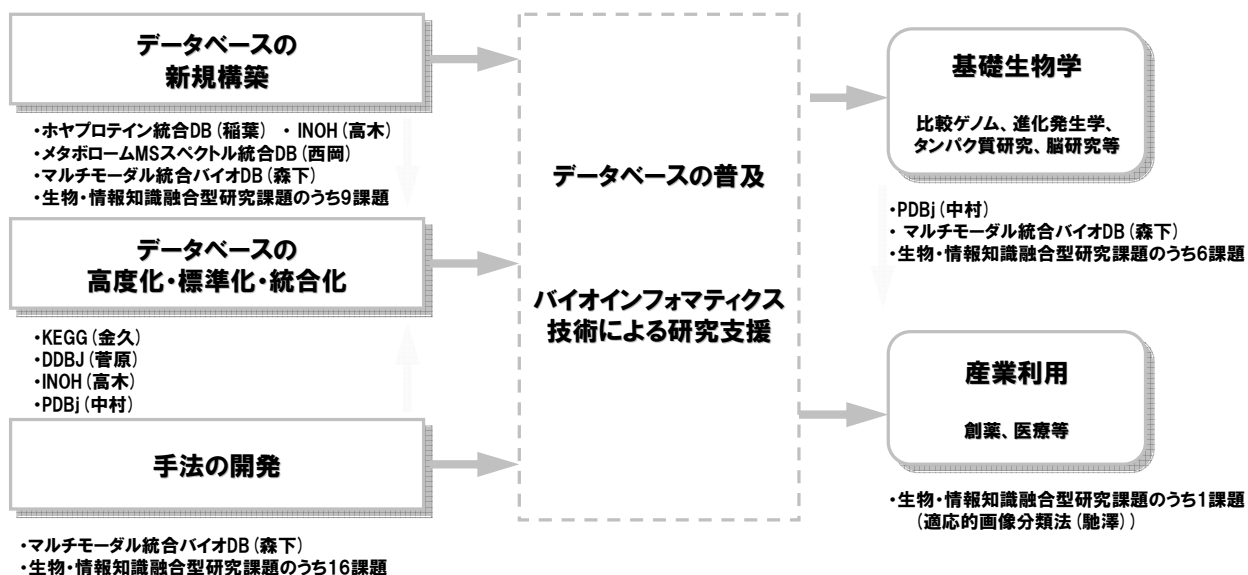
課題分類	開始年度	研究課題名	代表研究者
生物・情報知識融合型	H17	タンパク質の構造・機能予測法の開発とヒトゲノム配列への適用	太田元規
	H20	タンパク質の構造・機能・相互作用予測システムの開発と展開	
	H17	ヒト胚の形態発生に関する三次元データベース	塩田浩平
	H20	ヒト胚の形態発生に関する三次元データベース	

2) 特筆すべき成果の評価

選定課題の評価の前に、まず全研究課題に対して成果と波及効果を分類し、BIRD 事業の成果の全体像を図表 III-21 のように整理した。図の左から右の方向に研究から応用が位置づけられ、ボックスは研究や応用の大まかなフェーズを表し、矢印はそれらの相互の波及効果を概念的に表している。例えば、データベースの新規構築や高度化・標準化・統合化によりデータベースの普及が進み、その結果基礎生物学の進展、作業利用が加速するようなシナリオが想定される。また手法開発はデータベースの高度化等にも貢献するとともに、バイオインフォマティクス技術による研究

支援による基礎研究や産業利用につながるものが想定される。この流れの中に BIRD 事業の研究課題を位置づけると、「生命情報データベースの高度化・標準化」のうち、ホヤプロテイン統合 DB やメタボローム MS スペクトル統合 DB、マルチモーダル統合 DB、及び「生物・情報知識融合型研究課題」の多数は、データベースの新規構築に位置づけられる。一方、KEGG や DDBJ、PDBj は既存データベースの高度化・標準化・統合化に当たる。また「生物・情報知識融合型研究課題」の多くは手法の開発の成果をあげている。BIRD 事業の一部の研究課題は研究期間の間に既に基礎生物学や産業利用において応用研究の成果をあげているものもあるが、多くは今後の波及効果として期待されている段階である。基礎研究については比較ゲノムやタンパク質研究など、また産業利用では主に創薬や医療へ貢献が期待されている。

図表 III-21 BIRD 事業の成果の全体像の整理



選出した 5 つの研究課題に対して研究成果の発表論文の抄録を精査し、図表 III-21 に示した成果の全体像の枠組みに従って論文を分類することで、研究課題がどのような分野での成果をあげているかを定量的に表し評価する。またその成果がどのような分野に波及効果があり得るかについて論文中の記載より抜粋し、波及効果事例として整理する。

この分類結果を集計した結果を図表 III-22 に示す。成果・波及効果ともに各分類の該当論文数を示している。また波及効果については基礎生物学、産業利用の具体的な分野を事例数とともに記載している。以下に各選定課題の特徴をまとめる。

- **ゲノムと環境の統合解析による生命システムの機能解読 (KEGG) 【金久實】**

成果については、データベースの構築、手法の開発、応用研究(基礎生物学)に関する論文が満遍なく出されており、バランスの取れた研究成果といえる。

波及効果については、評価委員による課題評価においてもバイオインフォマティクスやライフサイエンス分野に広く寄与することが期待されている通り、医薬品分野、環境分野、医学分野の多くの産業利用が想定されている。

- **蛋白質構造データバンクの国際的な構築と高度化 (PDBj) 【中村春木】**

成果については、データベースの構築、手法の開発、応用研究(基礎生物学)いずれとも論文数は多いが、その中でも応用研究(基礎生物学)が多い。タンパク質構造 DB 分野は非常に成熟しているため、応用研究に志向が移っているといえる。

波及効果については、論文中での言及はなかったが、基礎生物学はもちろんであるが、創薬等の産業分野にも波及は想定されうると考えられる。

- **マルチモーダル統合バイオ DB 【森下真一】**

成果、波及効果ともに基礎生物学の応用研究に関する論文が多い。研究課題の中で開発された解析ツールなどが既に応用研究へ適用されているものと考えられる。

- **タンパク質の構造・機能・相互作用予測システムの開発と展開 【太田元規】**

マルチモーダル統合バイオ DB と同様に、成果、波及効果ともに基礎生物学の応用研究に関する論文が多い。研究課題の中で開発された解析ツールなどが既に応用研究へ適用されているものと考えられる。

- **ヒト胚の形態発生に関する三次元データベース 【塩田浩平】**

成果・波及効果ともに基礎生物学への応用研究が多い。

本研究課題はもともと発生学や医学系の研究であり、本事業においてその成果を DB 化が成されたといえる。

以上のように、BIRD 事業の研究課題、特に選定された 5 課題については、データベース構築、手法開発とともに、基礎生物学や産業利用への応用研究の成果が大きい。また創薬や医療、環境などの分野への波及効果も大いに期待される。

図表 III-22 選定した研究開発課題の成果論文の分類集計結果

課題分類	開始年度	研究課題名	代表研究者	全論文数	分類対象論文数	成果					波及効果			
						データベースの新規構築	データベースの高度化等	手法の開発	応用研究(基礎生物学)	応用研究(産業利用)	基礎生物学		産業利用	
高度化・標準化	第II期 (平成18~22年度)	ゲノムと環境の統合解析による生命システムの機能解読(KEGG)	金久實	33	29	0	11	11	11	0	4	ゲノム解析: 3 タンパク質構造解析: 1	11	医薬品: 7 環境: 3 医学: 1
		蛋白質構造データベースの国際的な構築と高度化(PDBj)	中村春木	59	43	0	8	14	23	0	0	—	—	
		マルチモーダル統合バイオDB	森下真一	19	17	2	1	4	11	0	3	進化解析: 3	—	—
生物・情報知識融合型	2008	タンパク質の構造・機能・相互作用予測システムの開発と展開	太田元規	28	20	1	1	6	14	0	2	タンパク質機能予測: 2	—	—
	2008	ヒト胚の形態発生に関する三次元データベース	塩田浩平	11	8	2	0	0	3	3	4	発生学: 4	2	医学: 2

※本表の数値は、各研究課題の事後評価報告に記載されていた成果論文で言及されていた成果と波及効果を集計している。波及効果については著者によっては必ずしも言及がない場合、“—”で示している。

4節 人材の輩出

1) 調査の概要

BIRD 事業による研究人材育成の成果を確認するため、各課題の研究開発終了報告書における「研究開発実施体制」の項に記載された研究者に関する情報をもとに、研究終了時の所属、職階を整理した。また、その後の所属機関、職階の移動を確認するため、研究者データベース JST ReaD を用いて最新の研究者情報を調査し、終了時からの機関の移動、職階の変化を調査した。

2) 調査結果のまとめ

(a) 全般

BIRD 事業の各研究課題の終了時報告書に記載された研究体制に関する情報より、個人名を抽出した。名前ベースで集計した結果、少なくとも 723 名の研究者等（大学院生や補助者を含む）が BIRD 事業に参加したことがわかった。参加研究者等が 10 名以上 30 名未満の課題が全課題の半数以上を占めていた。

ここでは研究人材が備える望ましい要素として、若手研究者の代表的な職階である研究員および助手のキャリアアップを「独立性」、教授へのキャリアアップを「主導性」、機関間の移動を「流動性」と定義し、本調査結果を元に以下でそれぞれの指標について考察する。

(b) 独立性

参加研究者等の職階について、終了時報告書記載時点と、JST ReaD 等で調査した時点（2011 年 10 月）とを比較した。若手研究者の中核的な職名と考えられる「研究員」と「助教」に着目すると、課題終了時では「研究員」が 47 名、「助教」が 53 名であったが、最新の調査時点では、それぞれ 43 名と 51 名にわずかに減少していた（図表 III-23）。

図表の対角線上の数字は職階に変化がない者の数を表しており、対角線の上は職階に変化を生じた者の数である。表における職階の並びが実質的にキャリアアップの順となっていると仮定すると、対角線よりの左下はキャリアアップを果たした者の数、右上はその逆を表すことになる。表では、大学において一般的な職階である「教授」「准教授」「講師」「助教」「研究員」について、この順に職位が上位であるとみなし、網掛けによってキャリアアップ領域を示している。その結果、これらの職の間ではかなりの割合でキャリアアップを果たしており、その逆の変化はわずかであることがわかった。例えば、研究員 47 名のうち、19 名が何らかの職階の変化を生じており、そのうち 9 名が助教になっている。同様に助教 53 名のうち 18 名が移動、14 名がキャリアアップするなど、着実に上位の職に移動している。

本集計は、現時点で所属が判明している参加者のみを対象としたものであり、限定的ではあるが、若手研究者が BIRD 事業をきっかけにキャリアアップを図っていると推察できる。

(c) 主導性

研究開発において主導的役割が期待される職である准教授、教授に着目すると、准教授は課題終了時から最新調査時点までの間、49名から38名に減少した。一方教授は74名から79名に増加した(図表 III-23)。職名の変化の内容を見ると、准教授のうち13名は教授に昇格していた。教授については1名が准教授に、その他の職に4名などの変化が見られた。上述のように本集計は限定的ではあるが、少なくとも准教授職については、13名が教授への昇格を果たしており、その点からは、BIRD 事業は、より主導性を発揮する立場の人材を輩出した、ということができらるだろう。

(d) 流動性

終了時と調査時点との所属機関を比較することにより、研究者の移動状況を把握した。ただし、本集計は2時点の比較をしたものであり、所属が同一である場合でも、その間に移動が無かったと判断することはできないことに留意が必要である。また、終了年は課題によって様々であり、課題間で時間経過に差があるが、それを等しく評価している。

機関別の移動状況の集計結果を見ると(図表 III-24)、多数の参加者を輩出した東京大学、大阪大学などでは20%程度の者が機関を移動していた。その他についても概ね10%以上の移動が見られた。

研究者の流動性については研究者全体を対象とした統計調査が存在せず、一般と比較してこの結果を評価することは困難である。その中でも1000人を超える規模の研究者を対象とした調査例として科学技術政策研究所が2008年度に行った調査(サンプル数1,259) [1]、および米国カーネギー財団が1993年に行った調査(サンプル数1,699) [2]を挙げることができる。その報告によると、日本における大学教授職の生涯(在職20年と仮定)における移動回数期待値は0.78回(1993年)～1.51回(2008年)程度であり、1年あたりに換算すると0.039～0.076回である。本調査では経過年が様々であるが、これを仮に平均5年と見なした場合、教授職は0.032となり、前提条件が異なるが、教授職の一般的な水準よりはやや低い水準であると推察される。

[1] 科学技術政策研究所 NISTEP REPORT123、「科学技術人材に関する調査」、2009年

[2] 有本、江原「大学教授職の国際比較」、玉川大学出版部、1996年

図表 III-23 BIRD 事業に参加した研究者の職階の移動（終了時から 2011 年 10 月）

	最新(2011年10月)の職階										うち職階が変化した数	うちキャリアアップを果たした数	変化した割合		
	教授	准教授	講師	助教	研究員	ポストドクター	技術員	補助員・補佐員	大学院生	その他				合計	
課題終了時の職階	教授	67	1			1	1				4	74	7	—	9.5%
	准教授	13	32	1	1	2						49	17	13	34.7%
	講師		1	4	1							6	2	1	33.3%
	助教	2	7	5	35	3					1	53	18	14	34.0%
	研究員	1	3	2	9	28	1			1	2	47	19	15	40.4%
	ポストドクター	1			1	1	1					4	3	—	75.0%
	技術員				1			3				4	1	—	25.0%
	補助員・補佐員			1	1	2		1				5	5	—	100.0%
	大学院生				2	3	3			1	2	11	10	—	90.9%
	その他	3				3		1			16	23	7	—	30.4%
	合計	87	44	13	51	43	6	5		2	25	276	89	—	30.4%

※集計対象は「課題終了時」および「最新」の職階が判明している者のみである。

※「教授」「准教授」「講師」「助教」「研究員」の間でのキャリアアップを示す領域を網掛けで表示した。

※ 集約後の「教授」には「特任教授」、「特命教授」が含まれ、「准教授」には「助教授」「特任准教授」が含まれる。また「研究員」には「主任研究員」「BIRD 研究員」が含まれる。「ポストドクター」には「特別研究員」「ポストドクトラルフェロー」「博士研究員」「ポストドク」「ポスドク」の文字列に部分一致する職名が含まれている。

※課題終了時の職階は重複カウントが含まれる。そのため、縦の数の合計は最新の職階別数を超えることがある。

※空白のセルは「ゼロ」を表す。

※「-」のセルは職階の上下関係が必ずしも定かでないため、「キャリアアップ」を評価していないことを表す。

図表 III-24 BIRD 事業に参加した研究者の所属機関の移動

	最新(2011年10月)の所属機関																	合計	うち所属機関が変化した数	変化した割合				
	東京大学	京都大学	大阪大学	科学技術振興機構	慶應義塾大学	産業技術総合研究所	国立遺伝学研究所	理化学研究所	筑波大学	奈良先端科学技術大学	九州大学	名古屋大学	北海道大学	神戸大学	横浜市立大学	広島大学	東北大学				長浜バイオ大学	海洋研究開発機構	高知大学	その他
東京大学	30	3															2				3	38	8	21.1%
京都大学	3	26	1			1																31	5	16.1%
大阪大学			16			1				1			1				1					20	4	20.0%
科学技術振興機構	3	1				1									1	1					3	10	10	100.0%
慶應義塾大学					11		1	1		1												14	3	21.4%
産業技術総合研究所		1				12																13	1	7.7%
国立遺伝学研究所			1				5							1				1				8	3	37.5%
理化学研究所							1	6														7	1	14.3%
筑波大学									7													7		
奈良先端科学技術大学										6												6		
九州大学						1					5											6	1	16.7%
名古屋大学												5										5		
北海道大学													4									4		
神戸大学	1													3								4	1	25.0%
横浜市立大学						1									1		1				1	4	3	75.0%
広島大学																3						3		
東北大学																	3					3		
長浜バイオ大学																		3				3		
海洋研究開発機構																			3			3		
高知大学																				3		3		
その他	4	2	2	1		2	3					2			1	2					3	75	19	20.2%
合計	41	33	20	1	11	18	11	7	7	7	6	7	4	4	2	5	10	4	3	3	82	286		

※「課題終了時」および「最新」の所属機関が判明している研究者のみを集計の対象としている。

※空白は「ゼロ」を表す。

3章 データベースの提供

1節 BIRD 研究課題の成果としてのデータベース提供

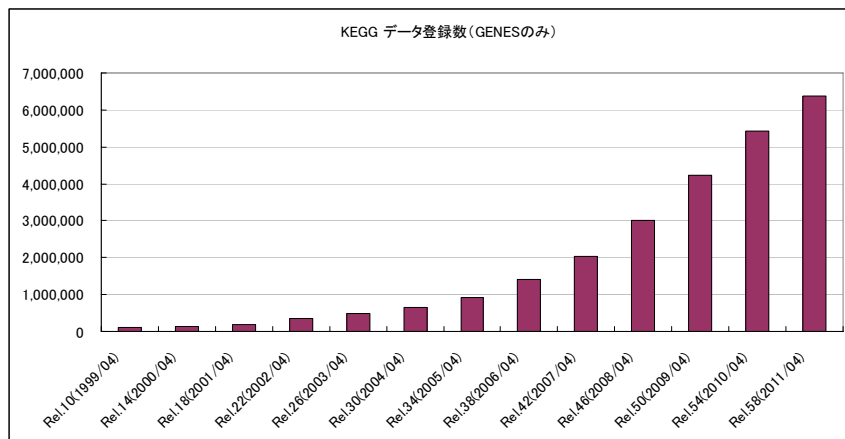
1) KEGG

KEGG (Kyoto Encyclopedia of Genes and Genomes) は、生命システム情報統合データベースであり、京都大学の金久研究室により、運営されている。1995年12月に Release 0.1 が発表され、現在もデータは更新されており、国内外で広く利用されている。BIRD 事業では「生命情報データベースの高度化・標準化」において以下の研究開発課題が採択されている。

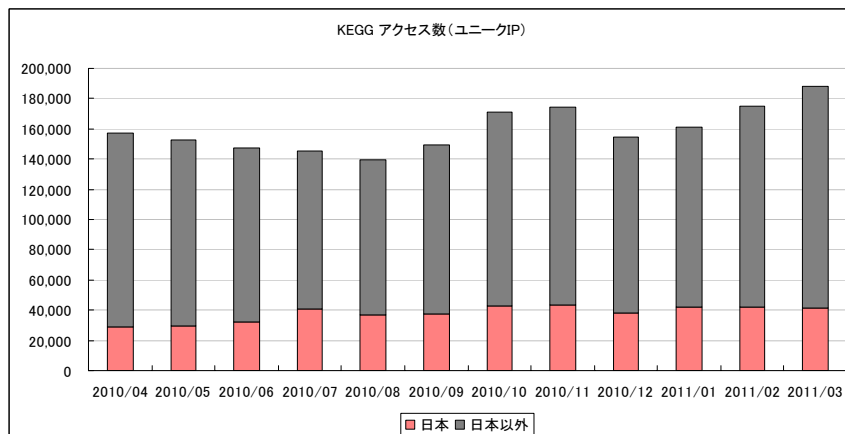
- 第 I 期「2 項関係に基づくゲノムと生命システムの機能解読」
- 第 II 期「ゲノムと環境の統合解析による生命システムの機能解読 (KEGG)」

図表 III-25、図表 III-26、図表 III-27 に KEGG のデータ登録数とアクセス数、主要関連論文の被引用回数の年推移を示す。

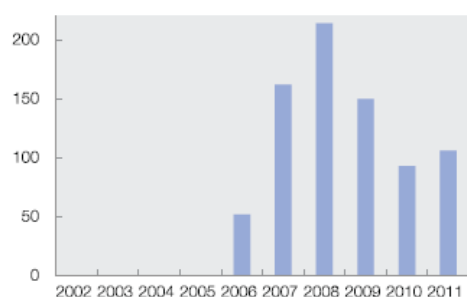
図表 III-25 KEGG のデータ登録数 (GENES のみ)



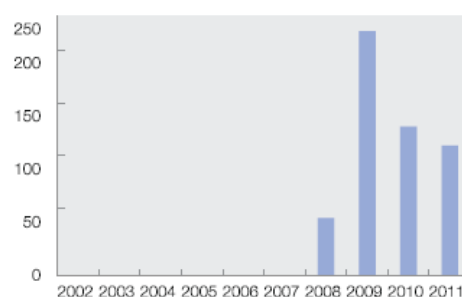
図表 III-26 KEGG のアクセス数 (ユニーク IP 数)



図表 III-27 KEGG の主要関連論文の被引用回数



From genomics to chemical genomics: new developments in KEGG
 Kanehisa, M; Goto, S; Hattori, M; Aoki-Kinoshita, KF; Itoh, M; Kawashima, S; Katayama, T; Araki, M; Hirakawa, M
 Nucl. Acids Res. (2006) 34 (suppl 1): D354-D357



KEGG for linking genomes to life and the environment
 Kanehisa, M; Araki, M; Goto, S; Hattori, M; Hirakawa, M; Itoh, M; Katayama, T; Kawashima, S; Okuda, S; Tokimatsu, T; Yamanishi, Y
 Nucl. Acids Res. (2008) 36 (suppl 1): D480-D484



KEGG for representation and analysis of molecular networks involving diseases and drugs
 Kanehisa, M; Goto, S; Furumichi, M; Tanabe, M; Hirakawa, M
 Nucl. Acids Res. (2010) 38 (suppl 1): D355-D360

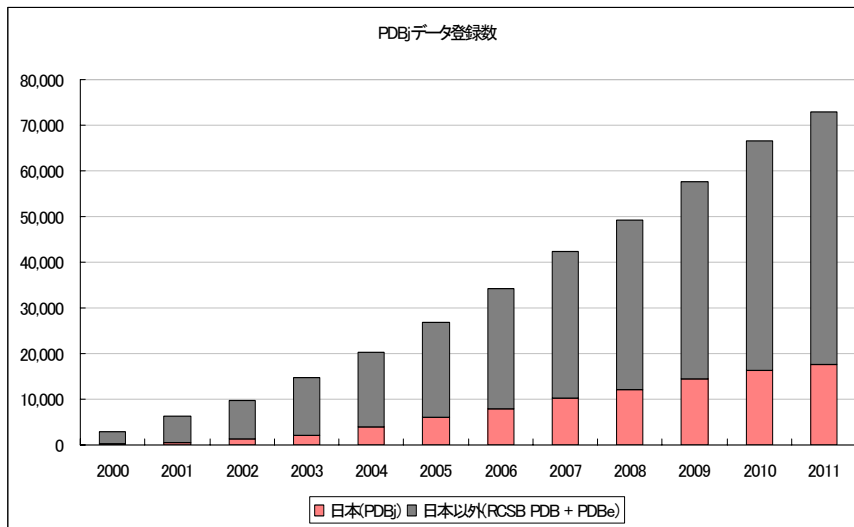
2) PDBj

日本蛋白質構造データバンク (PDBj: Protein Data Bank Japan) は、JST-BIRD (現在は JST-NBDC) と大阪大学の支援を受け、米国 RCSB、BMRB、および欧州 PDBe と協力して、生体高分子の立体構造データベースを国際的に統一化された PDB アーカイブとして運営するとともに、様々な解析ツールを提供している。BIRD 事業では「生命情報データベースの高度化・標準化」において以下の研究開発課題が採択されている。

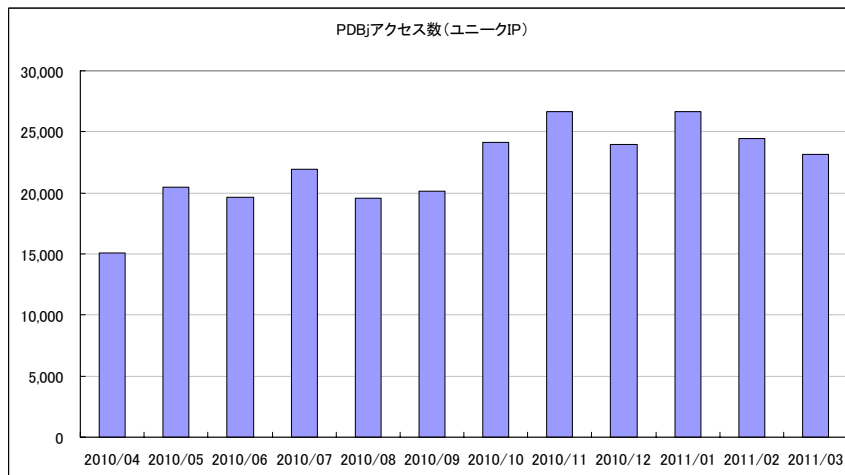
- 第I期「蛋白質立体構造データベースの高度化」
- 第II期「蛋白質構造データバンクの国際的な構築と高度化 (PDBj)」

図表 III-28、図表 III-29、図表 III-30 に PDBj のデータ登録数とアクセス数、主要関連論文の被引用回数の年推移を示す。

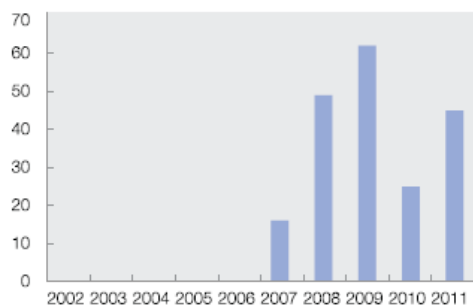
図表 III-28 PDBj のデータ登録数



図表 III-29 PDBj のアクセス数(ユニーク IP 数)

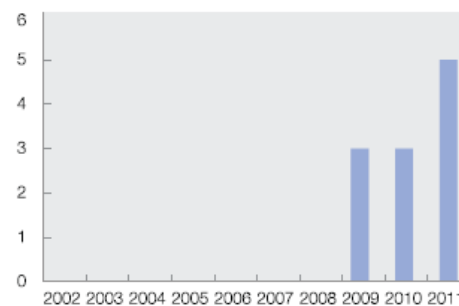


図表 III-30 PDBj の主要関連論文の被引用回数



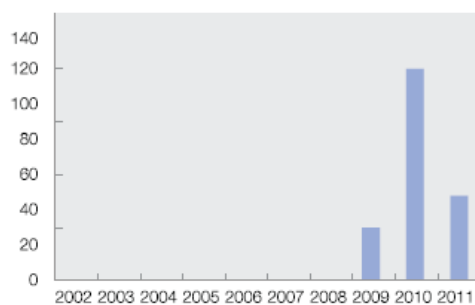
The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data

Berman, H; Henrick, K; Nakamura, H; Markley, JL
Nucl. Acids Res. (2007) 35 (suppl 1): D301-D303



Protein structure databases with new web services for structural biology and biomedical research

Standley, DM; Kinjo, AR; Kinoshita, K; Nakamura, H
Brief Bioinform (2008) 9 (4): 276-285



Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay

Matsushita, K; Takeuchi, O; Standley, DM; Kumagai, Y; Kawagoe, T; Miyake, T; Satoh, T; Kato, H; Tsujimura, T; Nakamura, H; Akira, S
Nature 458, 1185-1190 (30 April 2009)

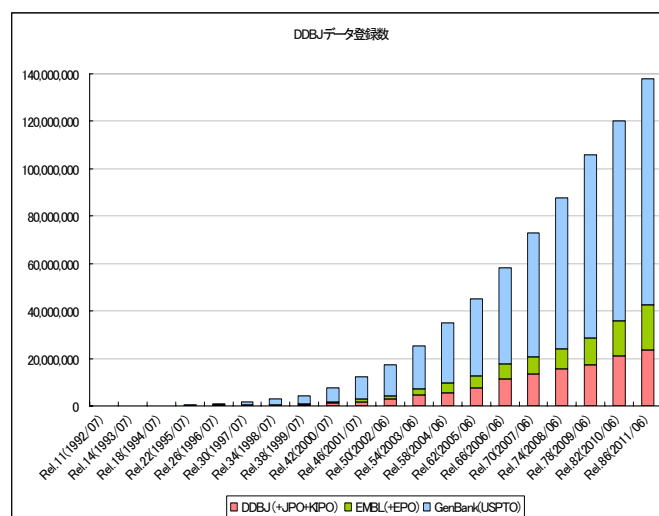
3) DDBJ

DDBJ (DNA Data Bank of Japan) は、欧州の EMBL-Bank/EBI および米国の GenBank/NCBI と共に、INSD (International Nucleotide Sequence Database) を構築・維持している 3つの機関の 1つである。INSD データは目的や国籍に拘わらず閲覧転用することが可能な世界科学の共有財であり、世界中の研究者は 3つの機関のいずれかを通じて INSD にデータを登録することが可能である。DDBJ は、文部科学省からの運営予算で国立遺伝学研究所の生命情報・DDBJ 研究センターにより運営されており、日本からの登録の 99%以上が、DDBJ を通じて行われている。BIRD 事業では「生命情報データベースの高度化・標準化」において以下の研究開発課題が採択されている。

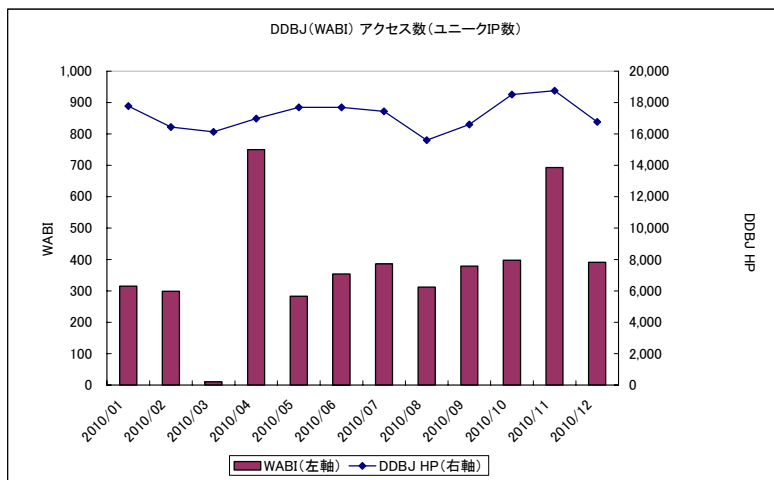
- 第I期「ゲノム生物学バックボーンデータベースの構築提供」
- 第II期「バイオ基幹情報資源の高準化と共用化」

図表 III-31、図表 III-32、図表 III-33 に DDBJ のデータ登録数とアクセス数、主要関連論文の被引用回数の年推移を示す。

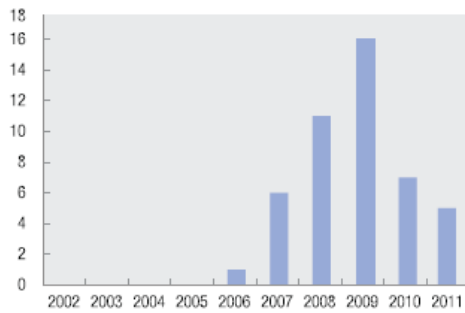
図表 III-31 DDBJ のデータ登録数



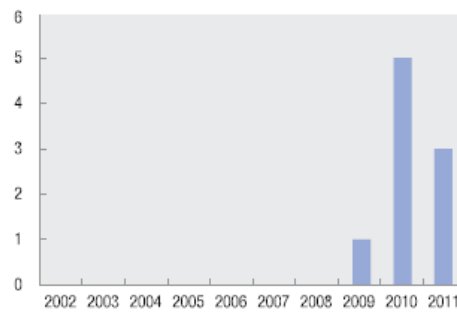
図表 III-32 DDBJのアクセス数(ユニーク IP 数)



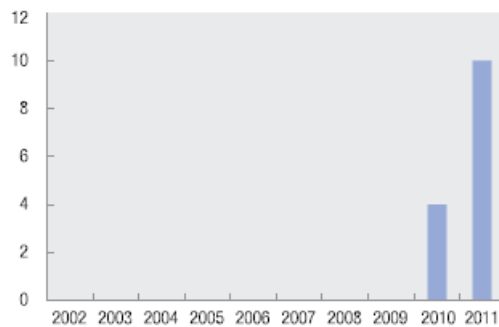
図表 III-33 DDBJの主要関連論文の被引用回数



Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples
 Abe, T; Sugawara, H; Kinouchi, M; Kanaya, S; Ikemura, T
 DNA Res (2006) 12 (5): 281-290



The GTOP database in 2009: updated content and novel features to expand and deepen insights into protein structures and functions
 Fukuchi, S; Homma, K; Sakamoto, S; Sugawara, H; Tateno, Y; Gojobori, T; Nishikawa, K
 Nucl. Acids Res. (2009) 37 (suppl 1): D333-D337



Archiving next generation sequencing data
 Shumway, M; Cochrane, G; Sugawara, H Shumway, Martin;
 Cochrane, Guy; Sugawara, Hideaki
 Nucl. Acids Res. (2010) 38 (suppl 1); D870-D471

2節 JST が独自に作成したデータベースの提供

1) JSNP

JSNP は、2000 年 4 月から JST と東京大学医科学研究所ヒトゲノム解析センター が共同で開設・運営している。JSNP は、2000 年から実施されているミレニアム・プロジェクトの高齢化分野におけるテーマ「高齢化社会に対応し、個人の特徴に応じた革新的医療の実現（ヒトゲノム）」の一環として開発を行ってきており、次の 3 種の DNA 多型情報データベースをネットワーク化していた。

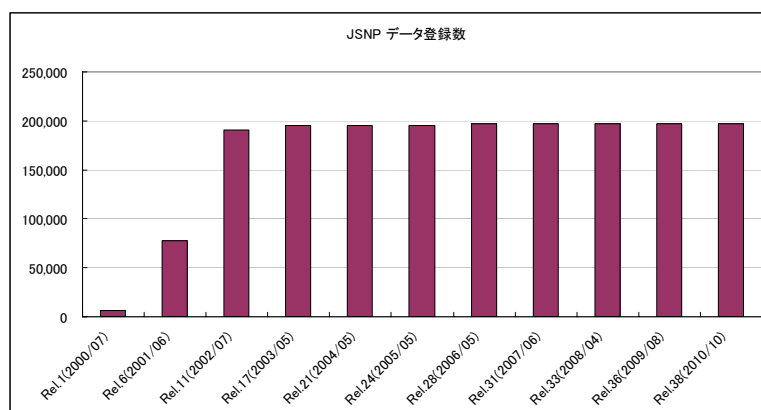
老年病 SNP データベース（東京都老人医療センター）

ヒトミトコンドリアゲノム多型データベース（財団法人岐阜県国際バイオ研究所）

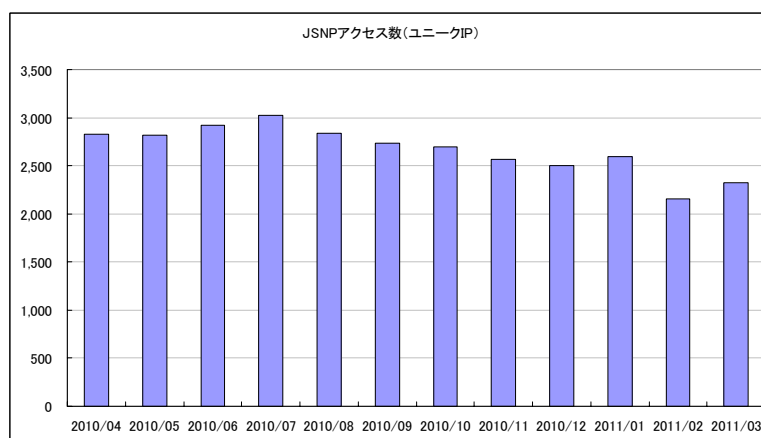
蛋白質多型データベース（独立行政法人 放射線医学総合研究所）

図表 III-34、図表 III-35、図表 III-36 に JSNP のデータ登録数とアクセス数、主要関連論文の被引用回数の年推移を示す。

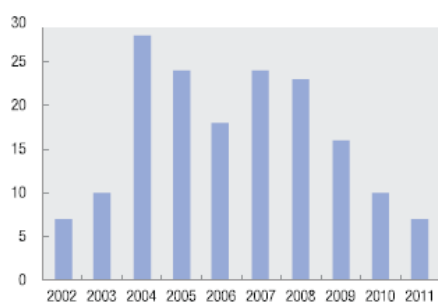
図表 III-34 JSNP のデータ登録数



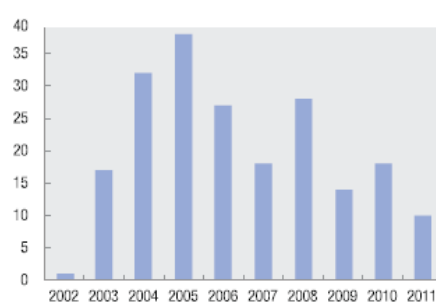
図表 III-35 JSNP のアクセス数（ユニーク IP 数）



図表 III-36 JSNP の主要関連論文の被引用回数



JSNP: a database of common gene variations in the Japanese population
 Hirakawa, M; Tanaka, T; Hashimoto, Y; Kuroda, M; Takagi, T; Nakamura, Y
 Nucl. Acids Res. (2002) 30 (1); 158-162



Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190562 genetic variations in the human genome
 Haga, H; Yamada, R; Ohnishi, Y; Nakamura, Y; Tanaka, T
 Journal of Human Genetics (2002) 47, 605-610

2) HOWDY

HOWDY (Human Organized Whole genome Database:ヒトゲノム情報統合データベース) は、ヒトゲノムに関連する情報を持つ複数のデータベースを、1つのデータベースのようにまとめて検索できるデータベースである。HOWDYは、現在14の公共データベースを情報源としており、最新のヒトゲノム情報を網羅的に検索・取得することができる。

図表 III-37、図表 III-38 に HOWDY の現在のデータ登録数とアクセス数の年推移を示す。

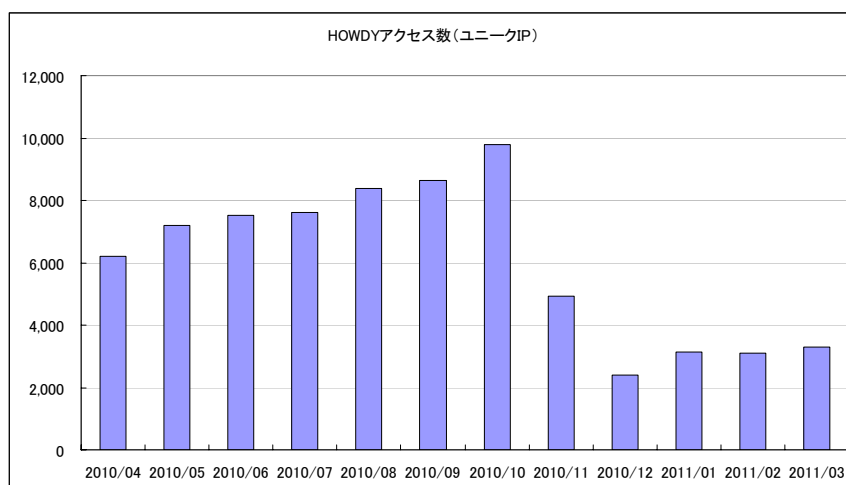
図表 III-37 HOWDY のデータ登録数

Data Class		Entries
Gene	GDB(Data)	-
	HGNC	35,004
	EntrezGene	166,945
Protein	UniProtKB/Swiss-Prot	20,235
	RefSeq(NP)	31,304
	RefSeq(XP)	3,816
SNPs	dbSNPs	30,443,446
	JSNPs	167,195
Marker	UniSTS	303,294
	GDB(Amplimer)	-

Data Class	Entries
Nucleic Acid Sequence	675,128
Refseq(NM)	31,304
Refseq(XM)	3,816
Unigene	122,727
OMIM	21,394
Contig	11,495
ContigMap	26
ALIS-HGS	46
e-PCR	790
GDB(Map)	-

※2011年9月27日時点

図表 III-38 HOWDY のアクセス数（ユニーク IP 数）

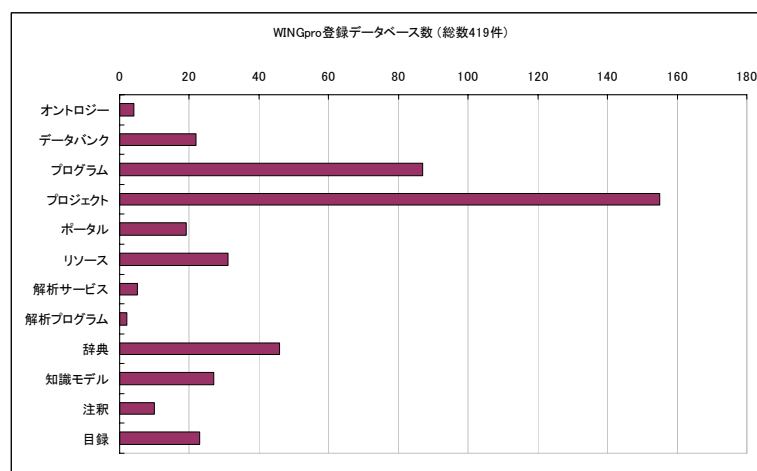


3) その他

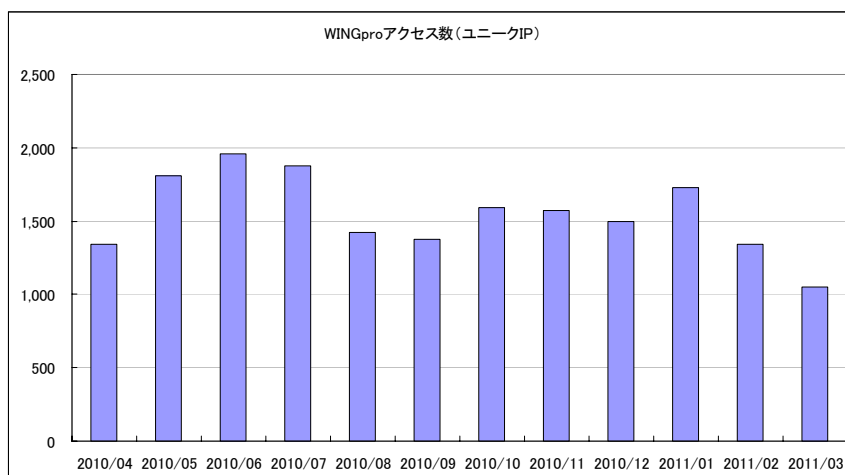
(a) WINGpro

WINGpro（データベースポータルサイト）は、ライフサイエンス分野のデータベースについての情報が収集、整理、分類されている。ディレクトリからは、データベースの構築法による分類と生物種および対象による分類でデータベースを一覧することが可能である。2007年7月2日よりユーザが記事を投稿および編集できる機能を公開しており、新規データベースに関する記事の投稿、既登録データベース関しても利用方法やTipsに関して追記することができる。WINGproは文部科学省「ライフサイエンス分野の統合データベース整備事業 平成18年度」の受託事業の一環として作成されたものである。図表 III-39、図表 III-40 に WINGpro の現在のデータ登録数とアクセス数の年推移を示す。

図表 III-39 WINGpro のデータ登録数



図表 III-40 WINGpro のアクセス数（ユニーク IP 数）

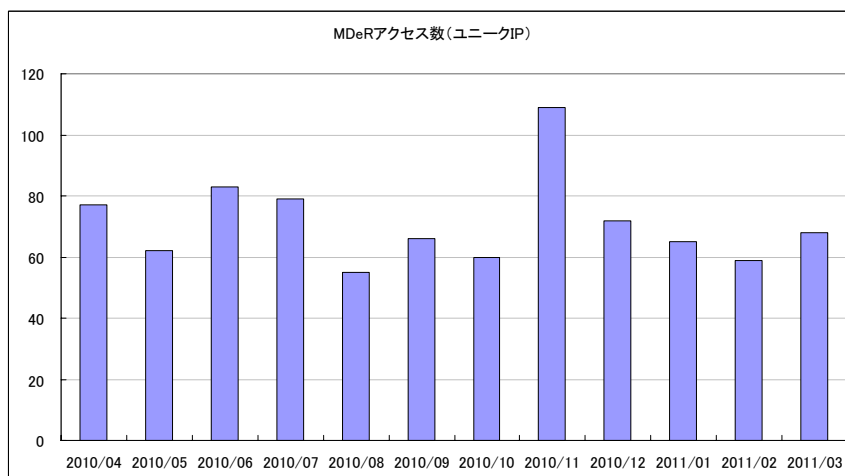


(b) MDeR

ライフサイエンス分野のメタデータ要素レポジトリ (MDeR) は、2008年3月26日に公開された。MDeRでは、メタデータ要素の検索、検索結果の比較、収録メタデータのメタデータ要素の一覧表示を行うことができる。MDeRには、ISO/IEC 11179 Part3(Registry metamodel and basic attributes)に準拠した形で、メタデータ要素が収録されている。

図表 III-41 に MDeR のアクセス数の年推移を示す。

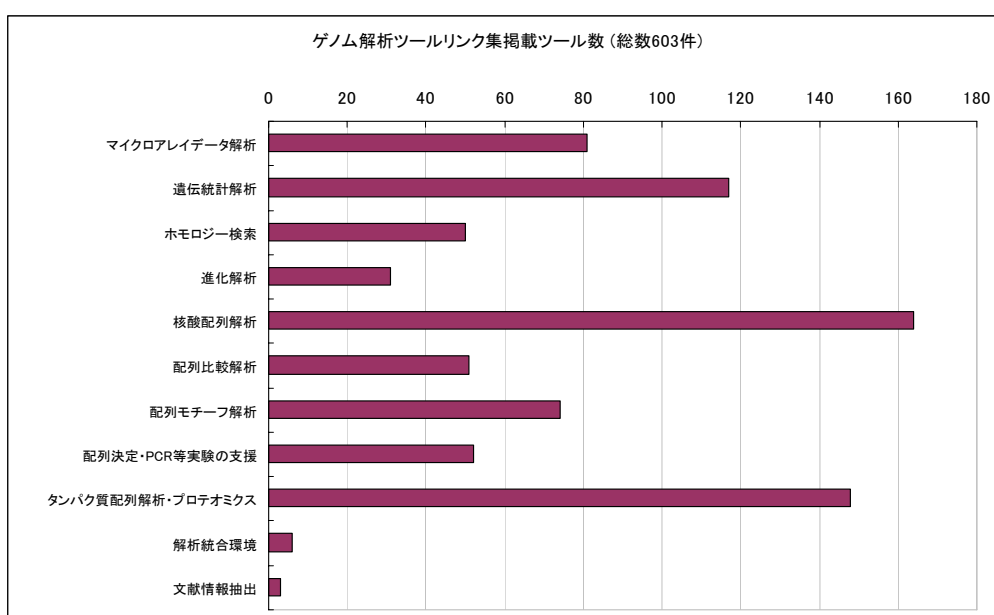
図表 III-41 MDeR のアクセス数（ユニーク IP 数）



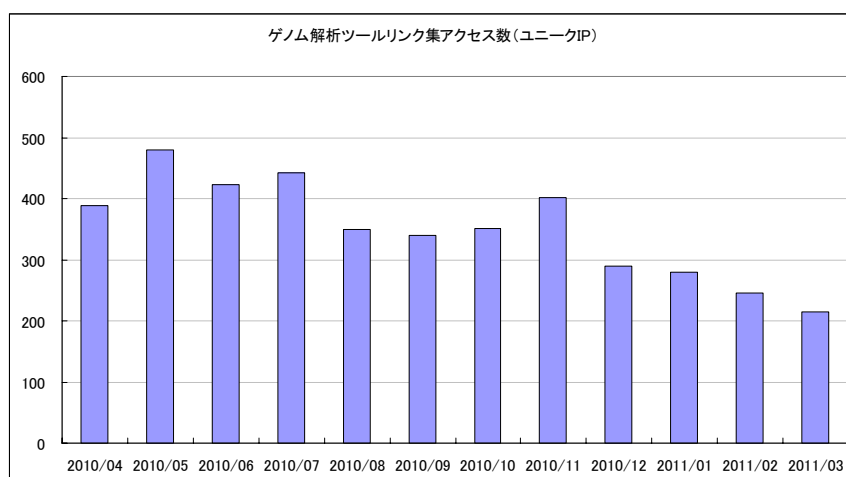
(c) ゲノム解析ツールリンク集

今日、多くの研究機関が分子生物学に関わるデータ解析ツール(以下、ゲノム解析ツール)を提供している。これらは分子生物学研究を押し進めるために必要不可欠となっており、様々な場面で目的・用途に適切なゲノム解析ツールを選択し、場合によっては組み合わせて使用する必要がある。ゲノム解析ツールリンク集では、そのサポートのために、2006年からツール提供サイトへのリンク・簡単な解説を提供している。図表 III-42、図表 III-43 にゲノム解析ツールリンク集の現在のデータ登録数とアクセス数の年推移を示す。

図表 III-42 ゲノム解析ツールリンク集のデータ登録数



図表 III-43 ゲノム解析ツールリンク数のアクセス数（ユニーク IP 数）



(d) GBIF

GBIF(Global Biodiversity Information Facility、地球規模生物多様性情報機構)は、生物多様性に関するデータを各国・各機関で分散的に収集し、ネットワークを通じて全世界的に利用することを目的とする国家・経済体・国際機関の多国間協約に基づく国際的科学協力プロジェクトである。図表 III-44 に GBIF のデータ登録数を示す。

図表 III-44 GBIF のデータ登録数

標本・観察データ概要

項目	件数
データ件数	294,885,721
データセット件数	11,340
データ提供機関数	350
データ提供国数	47

国別標本・観察データ件数

順位	国名	データ提供機関数	データ件数
1	United States	94	102,761,366
2	Great Britain	11	38,500,518
3	Sweden	2	24,277,686
4	South Africa	2	14,371,142
5	France	18	12,737,641
6	Netherlands	4	11,917,383
7	Australia	12	10,608,506
8	Germany	19	9,532,643
9	Norway	4	8,242,710
10	Spain	6	5,949,608
11	Costa Rica	2	5,755,456
12	Denmark	3	5,587,570
13	Belgium	5	3,375,704
14	Finland	8	2,978,533
15	Mexico	4	2,932,263
16	Austria	11	2,872,901
17	Japan	3	2,707,575
18	Canada	10	1,905,558
19	Poland	28	1,756,781
20	South Korea	21	1,654,597
21	Ireland	2	1,590,571
22	New Zealand	1	1,508,826
23	Switzerland	3	1,468,197
24	Taiwan	5	1,162,704
25	Luxembourg	1	660,040
26	Colombia	2	466,534
27	Iceland	1	458,396
28	Israel	1	430,857
29	Argentina	9	344,662
30	Slovenia	3	266,295
31	Andorra, Principality of	1	75,142
32	Portugal	5	66,915
33	Estonia	2	51,547
34	Peru	4	40,078
35	India	2	36,823
36	Madagascar	3	18,707
37	Togo	1	7,983
38	Benin	1	7,397
39	Cameroon	2	6,214
40	Tanzania	1	6,013
41	Hong Kong	1	1,901
42	Pakistan	1	853
43	Guinea	1	493
44	Mauritania	1	177
45	China	1	0
46	Nicaragua	1	0
47	Slovak Republic	1	0

データ形式標準

採用しているデータ形式	採用機関数	データセット件数
Darwin core	237	9,250
ABCD Schema	59	1,492

データ公開標準

採用しているプロバイダ	採用機関数	データセット件数
DiGIR protocol	179	8,979
TAPIR protocol	62	272
BioCASE protocol	58	1,491
IPT	2	3

(2011 年 11 月 1 日現在)

4章 人材育成

1節 概要

BIRD 事業では、「ゲノムリテラシー講座」を開催し、バイオインフォマティクス分野での人材育成を目的とした研修や普及活動を行った。

2001 年度（平成 13 年度）から 2004 年度（平成 16 年度）には、文部科学省科学研究費特定領域研究 ゲノム 4 領域の「ゲノム情報科学の新展開」との共催で、バイオインフォマティクスの基礎講義や、多型解析・ゲノムアノテーション・テキストからの情報抽出等、新しい研究成果を活用した実習を含む講義が実施された。

また、2009 年度（平成 21 年度）と 2010 年度（平成 22 年度）には、「バイオインフォマティクス技術者認定試験」の出題範囲から、情報学、配列解析、パスウェイ解析・システム生物学等に関する講義を実施し、その講義内容を e ラーニング教材としてストリーミング配信した。

この他に、バイオインフォマティクスに係る相談や、NCBI ミニコース日本語版・ライフサイエンス分野等の Web ラーニングプラザの提供を行った。

2節 実施結果

ここではゲノムリテラシー講座の実施結果について、受講者のアンケート結果などをもとにまとめる。

ゲノムリテラシー講座の主な目的としては、バイオインフォマティクスの基礎講義と関連する新しいツールの実習により、バイオインフォマティクスの普及させることであった。まずは、バイオインフォマティクスの普及のためには、学習意欲の高い若手研究者を対象とすることが望ましい。実際の参加者の属性は、バイオインフォマティクス初級者、大学院生、企業等の研究者などの 20 代から 30 代の若手研究者が中心であり、参加者の募集方法が適切であったことがうかがえる（図表 III-45、図表 III-46）。また、2007 年度（平成 19 年度）から 2008 年度（平成 20 年度）にかけて、バイオインフォマティクス授業を担当する教官、都道府県の衛生試験場研究者等の参加が見られるようになり受講者層が拡大しており、バイオインフォマティクスが普及していったことを示唆している。

さらに、普及という観点から、2009 年度（平成 21 年度）から講義内容のストリーミング配信が行われており、時間的地理的問題で講義に参加できない方々にも、学習の機会が与えられたことは特筆すべきである。

図表 III-45 年度別の受講者属性（職業）

年度	開催数	延べ回答数	学生	教育職	研究開発	その他
H16	5	-	-	-	-	-
H17	7	134	19	7	75	33
H18	6	105	17	10	54	24
H19	3	55	12	10	19	14
H20	3	67	29	11	24	3
H21	4	73	3	2	48	20
H22	2	38	9	0	27	2

図表 III-46 年度別の受講者属性（所属機関）

年度	開催数	延べ回答数	大学	民間企業	その他 研究機関
H16	5	-	-	-	-
H17	7	134	32	69	33
H18	6	103	32	44	27
H19	3	54	26	12	16
H20	3	69	41	21	7
H21	4	71	12	44	15
H22	2	40	15	21	4

IV 事業を取り巻く状況の変遷

1章 国内の状況

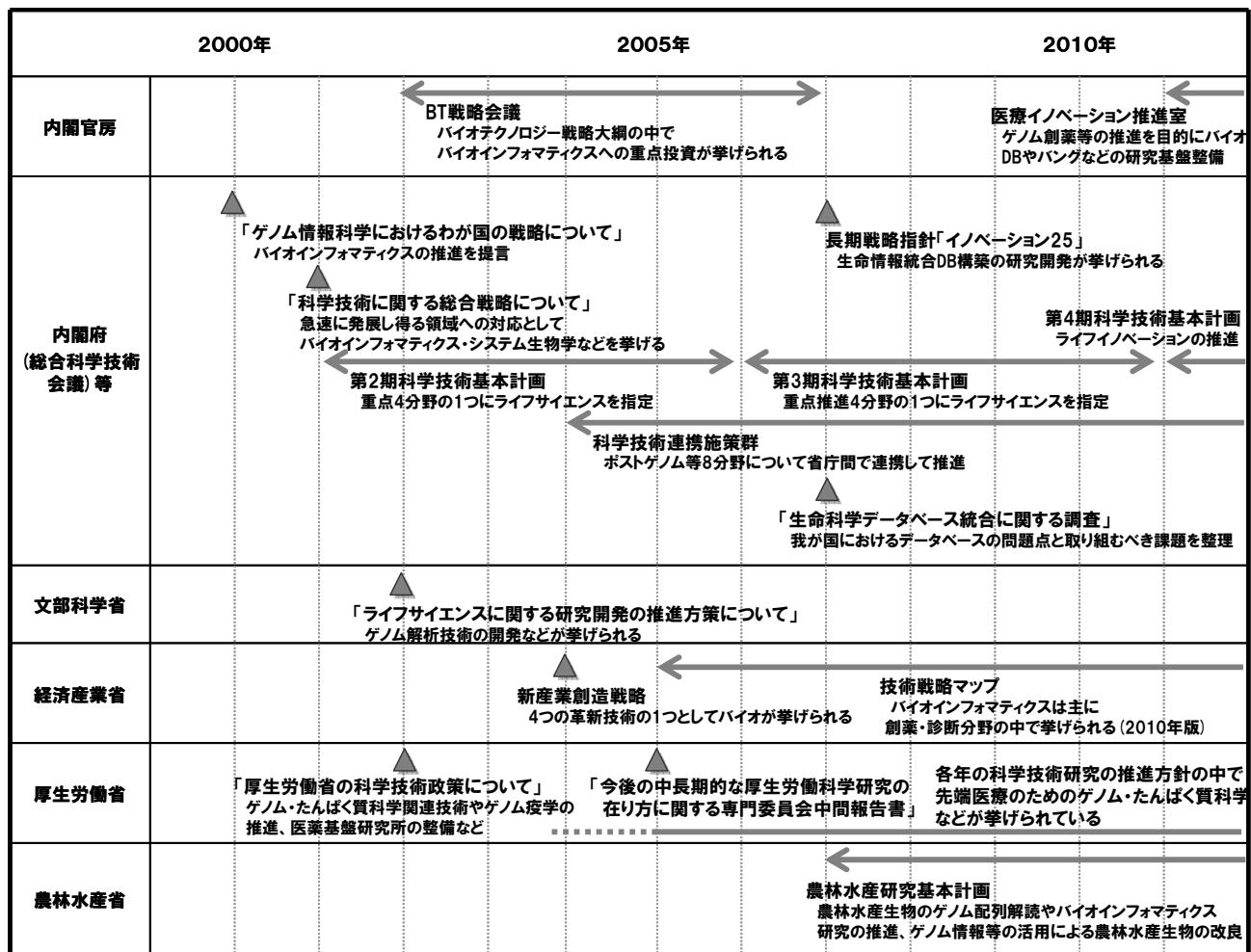
1節 国内の科学技術政策の変遷

BIRD事業が開始されてから現在に至るまでの期間における国内のバイオインフォマティクス関連科学技術政策の変遷を 図表 IV-1 に整理する。

まずバイオインフォマティクス関連政策として、内閣府の前身である総理府に設置された科学技術会議ライフサイエンス部会ゲノム科学委員会によって「ゲノム情報科学におけるわが国の戦略について」（2000年）の中でバイオインフォマティクスの推進が提言された。その後内閣府（総合科学技術会議）等によって諮問第1号「科学技術に関する総合戦略について」に対する答申（2001年）や第2期、第3期科学技術基本計画（それぞれ2001～2005年、2006～2010年）などにおいてライフサイエンスの重点化が進められた。また文部科学省、経済産業省、厚生労働省、農林水産省においてもそれぞれの立場からゲノム科学やタンパク質科学等のライフサイエンス関連の戦略などが練られた。例えば、文部科学省では「ライフサイエンスに関する研究開発の推進方策について」（2002年）のように研究開発の立場から方策が検討された。また経済産業省の場合は、「新産業創造戦略」（2004年）や「技術戦略マップ」（2005年～）のように技術戦略や産業化の立場からの戦略が練られた。

これらの各省庁の政策の一方で、内閣府の科学技術連携施策群においてポストゲノム等の重点課題については省庁間で連携して推進することが決定され、2007年には連携群の中で「生命科学データベース統合に関する調査」の実施結果が公表され、ライフサイエンス統合データベース構築への舵取りがなされた。近年では内閣官房において医療イノベーション推進室が設置される等、創薬等のより応用寄りへの取り組みもなされている。

図表 IV-1 国内の科学技術政策の変遷 (BIRD 事業開始時から現在まで)



※年の軸はその年の開始月を表す。但し、政策の期間あるいは時点の表示については月単位の区別はせず配置している。

2節 大学・研究機関、国家プロジェクトの変遷

1) 大学・研究機関の変遷

BIRD事業が開始された2001年時点に既に設立されていた、あるいはその後現在までに設立されたバイオインフォマティクス関連の大学研究室や研究機関の変遷について整理する（図表 IV-2）。

(a) 大学の動向

1991年には現在も日本を代表する3つの研究所・研究科が設立された。1つは東京大学医科学研究所に設立されたヒトゲノム解析センターで、文部科学省が発足させた創成的基礎研究「ヒトゲノム解析研究」および重点領域研究「ゲノム解析に伴う大量 知識情報処理の研究」と連携し、大学等におけるゲノム解析研究を促進させるために設立された。ゲノムデータベース分野、ゲノム構造解析分野、DNA 情報解析分野、ゲノムシーケンス解析分野、シーケンス技術開発分野、シーケンスデータ情報処理分野、機能解析イン・シリコ分野などの研究室がある。2つ目は京都大学化学研究所にスーパーコンピューターシステムが導入され、また同年度に発足した文部科学省「ヒト・ゲノムプログラム」の成果をゲノムネットと名付けた情報サービスとして提供し始めた。KEGGなどの世界的に利用されているデータベースを提供している。3つ目は奈良先端科学技術大学院大学で、情報科学研究科やバイオサイエンス研究科が設置された。

その後、1996年に慶応大学湘南藤沢キャンパスにバイオインフォマティクス・ラボが設置された。遺伝情報をコンピュータで比較・解析して、遺伝子の暗号文を解き明かすことを目的とした、ゲノム配列情報のコンピュータ解析や、細胞生物学の知識をもとに、コンピュータ上に細胞を再構築し、遺伝子操作や薬物添加の反応を予測するためのバーチャルな実験を行うことを目的とした細胞内代謝のコンピュータシミュレーション「E-CELL」を研究し、その後の先端生命科学研究所(山形県鶴岡市)の設立につながる。また北陸先端科学技術大学院大学知識科学研究科にバイオインフォマティクス関連の研究室が複数設置された。

2001年からはさらに多くの研究室・研究科が設置された。2001年にはまず大阪大学蛋白質研究所に日本蛋白質構造データバンク(PDBj)がBIRD事業の支援も受けて設立された。その他、九州大学生体防御研究所に遺伝情報実験センター、慶応大学が山形県鶴岡市にメタボローム解析やシステムバイオロジーを研究する先端生命科学研究所を、理工学部生命情報学科を設立した。

その後は多くの大学にバイオインフォマティクス関連の研究室が設置され、2003年には長浜バイオ大学がバイオロジーを総合的に学べる日本で唯一の単科大学として設立された。2007年には2001～2005年の科学技術振興調整費「生物

情報科学学部教育特別プログラム」の発展形として東京大学に理学部では 30 年ぶりに学科(生物情報科学科)を新設した。

また科学技術振興調整費によって、東大、京大、奈良先端大など大学を中心にバイオインフォマティクス人材養成ユニットが設置され人材養成が行われた。

(b) 公的研究機関の動向

1995 年に国立遺伝学研究所に生命情報研究センターが設立され、2001 年に生命情報・DDBJ 研究センターとなる。遺伝情報分析研究室、遺伝子機能研究室、大量遺伝情報研究室、データベース運用開発研究室、遺伝子発現解析研究室という 5 つの研究室で構成されており、国際塩基配列データベース (DDBJ/EMBL/GenBank) の一つである DDBJ の運営も行う。

1998 年には理化学研究所にゲノム科学総合研究センターが設立され、ヒトゲノム解読、マウス遺伝子百科辞典の完成、RNA 大陸の発見、タンパク 3000 プロジェクトの完了など、生命科学の節目となる、国際的な成果を達成した。その後 2008 年に解散し、各々を引き継ぐ新たな組織として発展した。

一方、経済産業省では 2001 年に産業技術総合研究所に生物情報解析研究センター (JBIRC) と生命情報科学研究センター (CBRC) の 2 つのセンターを設立した。JBIRC では大量のゲノム情報に含まれる生物情報の取得、取得に関する新技術の開発、取得した情報の整理および統合を生物科学の立場から推進することを目的とし、膜タンパク質の立体構造解析やヒト完全長 cDNA の機能解析、ヒトゲノム統合データベースの構築を中心に研究された。2008 年からはバイオメディシナル情報研究センターに改組された。CBRC はゲノム情報、生体高分子の構造と機能、細胞内ネットワークなどを総合的に解析し、産業応用に貢献できる技術を開発し、現在も研究が進められている。

また農林水産省ではこれまでの研究所を再編するなどし、2001 年に農業生物資源研究所を設立し、農業生物の遺伝資源の充実と活用の強化、農業生物のゲノムリソース・情報基盤の整備・高度化、農作物や家畜の生産性向上を目指した生物機能の解明、農作物や家畜の生物機能を高度に利用するための、病原菌等との生物間相互作用の解明と利用技術の開発、新たな生物産業を創出するための生物機能の利用技術の開発を実施している。国際プロジェクトの中心としてイネゲノム、カイコゲノムの解読なども実施している。

そのような中で JST の BIRD は 2001 年に設置され、2010 年までの 10 年の間バイオインフォマティクスを推進してきた。また 2007 年にはライフサイエンス統合データベースプロジェクトが実施された。2011 年からは統合データベースプロジェクトと BIRD 事業を一本化し、バイオサイエンスデータベースセンター (NBDC) が設置され、ライフサイエンスデータベース統合推進事業が進められる

こととなった。以下にNBDCの概要をまとめる。

文部科学省の統合データベースプロジェクトと、JSTのBIRDとが一本化され、NBDCが設置された。NBDCは、データベース研究開発戦略の立案、データベースの充実・公開、国際協力を事業内容とする組織である。この実施体制は図表 IV-3 の通りとなっている。この体制によって、基盤技術開発、府省間データベース統合、ファンディングプログラムの実施機能を果たすこととなる。このファンディングプログラムに関しては、長洲毅志研究総括の下に基盤技術開発プログラム、高木利久研究総括の下に統合化推進プログラムが生まれ、それぞれに研究アドバイザーが付くこととなった。研究開発プログラムには、目的別に大きく分けて、以下に示す2通りが存在する。

- ・ 基盤技術開発プログラム
- ・ 統合化推進プログラム

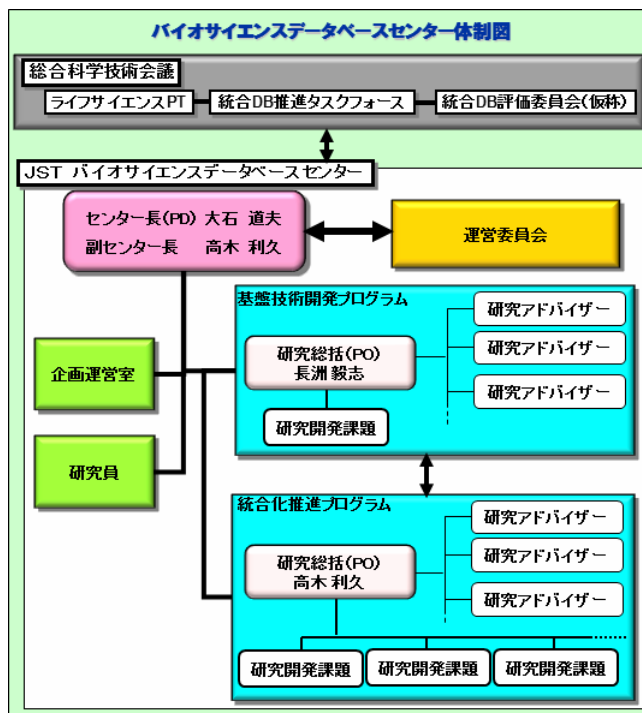
基盤技術開発プログラムは、データベース統合化の実現に向けて基盤となる技術開発を行い、ポータルサイトへの実装までを行うものであり、具体的にはデータベース統合検索技術、大規模データの活用技術、データベース解析統合利用環境の整備など、およびこれらに付随する各種技術開発とその実装を行う。

統合化推進プログラムは、データベースの分野別統合化または目的別統合化、ならびに散在しているデータベースの統合化を行うものであり、生物種別、分野別、目的別またはデータ種類別などで統合化を目的とする。ここで採択される課題は、その分野の研究コミュニティや関連研究機関をとりまとめて、それらを代表する（日本の中核・拠点となる）データベースの構築、整備、統合化を目指すことが求められる。

図表 IV-2 国内の大学・研究機関の設立の推移

	1990年	1995年	2000年	2005年	2010年
大学	<p>← 東京大学医科学研究所ヒトゲノム解析センター</p> <p>← 京都大学化学研究所バイオインフォマティクスセンター（前身のゲノムネット提供時代から）</p> <p>← 奈良先端科学技術大学院大学（バイオサイエンス研究科、情報科学研究科など）</p>	<p>← 慶応大学湘南藤沢キャンパスバイオインフォマティクス・ラボ</p> <p>← 北陸先端科学技術大学院大学知識科学研究科バイオインフォマティクス関連研究室</p>	<p>← 大阪大学蛋白質研究所 日本蛋白質構造データバンク (PDBi)</p> <p>← 九州大学生体防御研究所遺伝情報実験センター</p> <p>← 慶応大学先端生命科学研究所</p> <p>← 慶応大学理工学部生命情報学科</p> <p>← 大阪大学情報科学研究科バイオ情報工学専攻</p> <p>← 長浜バイオ大学</p> <p>← 九州工業大学生命情報工学科</p> <p>← 北海道大学情報科学研究科バイオインフォマティクス講座</p> <p>開始年不明</p> <p>早稲田大学ITバイオ研究所</p> <p>早稲田大学バイオインフォマティクス研究所</p> <p>明治薬科大学バイオインフォマティクス研究室</p> <p>← 東京大学理学部生物情報科学科</p> <p>← お茶の水女子大バイオインフォマティクス研究室</p> <p>← 立命館大学生命情報学科</p>		
文部科学省		<p>← 国立遺伝学研究所生命情報・DDBJ研究センター</p>	<p>← 理化学研究所ゲノム科学総合研究センター</p>	<p>← JSTバイオインフォマティクス推進センター (BIRD)</p> <p>← 科学技術振興調整費バイオインフォマティクス人材養成ユニット (東大、慶応大、奈良先端大、産総研など。各5年間)</p> <p>← ライフサイエンス統合データベースセンター (DBCLS)</p> <p>← JSTバイオサイエンスデータベースセンター (NBDC)</p>	
経済産業省			<p>← 産総研生物情報解析研究センター (JBIRC)</p> <p>← 産総研生命情報科学研究センター (CBRC)</p> <p>← 産総研バイオメディシナル情報研究センター</p>		
厚生労働省				<p>← 医薬基盤研究所</p>	
農林水産省			<p>← 農業生物資源研究所</p>		

図表 IV-3 バイオサイエンスデータベースセンター (NBDC) の事業運営体制



2) 国家プロジェクトの変遷

BIRD 事業が開始された 2001 年時点で既に設立されていた、あるいはその後現在までに実施された国家プロジェクトの変遷を整理する。

(a) 文部科学省

1998 年に理化学研究所のゲノム科学総合研究センターの運営事業として「理研ゲノム」のプロジェクトが開始された。ゲノム科学総合研究センターはヒトゲノム解読やマウス遺伝子百科辞典の完成、RNA 大陸の発見、タンパク 3000 プロジェクトの完了など、生命科学の節目となる、国際的な成果を達成した。また理化学研究所には 2000 年に植物科学ゲノムセンター、2003 年には遺伝子多型センター(2008 年からはゲノム医科学研究センター)などが設置され、国内のゲノム研究の中心的機関となった。

遺伝子多型研究については、2000 年にミレニアム・プロジェクトの一環として東京大学医科学研究所・JST の共同研究の形で日本人多型研究プロジェクト(JSNP)が開始された。その後、2003 年からは理化学研究所に遺伝子多型センター(2008 年以降はゲノム医科学研究センター)が設置されるとともに、同年に「個人の遺伝情報に応じた医療の実現プロジェクト(オーダーメイド医療実現化プロジェクト)」が開始された。

タンパク質構造研究については、2002 年に解析目標数が掲げられた「タンパク 3000 プロジェクト」が開始され、2007 年からはよりターゲットを絞った「ターゲットタンパク研究プログラム」に継承された。

ゲノムやタンパク質の単体の研究だけでなく、それらの生体分子間の相互作用を体系的に研究することを目的とした研究が 2004 年から「ゲノムネットワークプロジェクト」という名称で開始された。2009 年からは「革新的細胞解析研究プログラム(セルイノベーション)」として継承されている。

この他に、2002 年から 2011 年に渡って実験動物や遺伝子材料などのバイオリソースを整備する「ナショナルバイオリソースプロジェクト」や、2003～2007 年の「細胞・生体機能シミュレーションプロジェクト」、2003～2006 年の「糖鎖機能解明とデータベース構築プロジェクト」など、様々な大型研究プロジェクトが実施された。

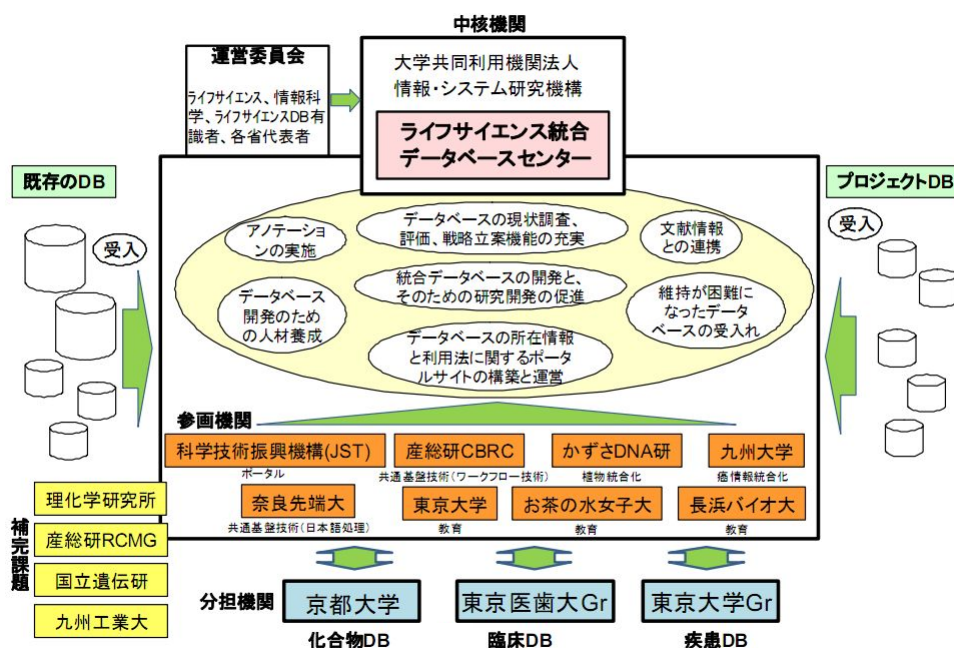
バイオインフォマティクスに特化したものとしては、ミレニアム・プロジェクトの一環として 2000 年から開始された科学研究費特定領域研究「ゲノム 4 領域のうちの一つ「ゲノム情報科学の新展開」などがある。

このような中で 2006～2010 年に「統合データベースプロジェクト」が実施されることとなった。以下の統合データベースプロジェクトの概要をまとめる。

文科省による「ライフサイエンス分野の統合データベース整備事業」は、「戦

略立案・実行評価」、「統合データベース開発」、「統合データベース支援」を3本柱とし、基本的な組織体系は、プロジェクトの中心的な機関として位置づけられる中核機関及び参加機関、各分野および各組織の統合を進めることを目的とした分担機関、事業の適正かつ円滑な実施に必要な重要事項を議論するための研究運営委員会とから構成された。平成19年度以降の本格実施期間からは補完課題を実施する機関も含めた体制となった。中核機関は上述した3本柱全てを担当し、分担機関は中核機関と連携して化合物・医薬品や臨床疾患等の医薬に関わるデータベースの統合化を進めた。研究運営委員会は、ライフサイエンス、情報処理技術、ライフサイエンスデータベースの3分野から選ばれた専門家で構成された(図表 IV-4)。

図表 IV-4 統合データベースプロジェクトの実施内容と運営体制



(b) 経済産業省

1998年に「ゲノムインフォマティクス・プロジェクト」が開始され、翌1999年には「完全長cDNA構造解析」、2000年には「タンパク質の機能解析」が開始された。

統合データベースを目標としたプロジェクトは文部科学省よりも早くから開始されており、2000年からの「バイオインフォマティクス関連データベース整備」、2005年からの「ゲノム情報統合プロジェクト」、2008年からの「統合データベースプロジェクト」、そして2011年から進められている「ライフサイエンスデータベースプロジェクト」へと継承されている。

また経済産業省の場合、文部科学省の基礎研究よりも応用寄りのプロジェクトが多く、例えば1999年からの「糖鎖エンジニアリング」に関するプロジェクトや、2000年からの「生物機能を活用した生産プロセス基盤技術開発」、2004年からの「バイオプロセス実用化」、2007年からの「創薬加速に向けたタンパク質構造解析基盤技術開発」などがある。

(c) 厚生労働省

医療や製薬関連のプロジェクトが大半で、2000年に開始された「遺伝子解析による疾患対策・創薬推進事業」、同じく2000年に開始された「ヒトゲノム・再生医療等研究（ヒトゲノム・遺伝子治療分野）」がある。その後、2002年からの「トキシコゲノミクス研究」、同年開始の「ナノメディシン」、2003年開始の「疾患関連タンパク質解析研究」、2005年の「ファーマコゲノミクス」、2007年の「トキシコゲノミクス・インフォマティクスプロジェクト」、2007年の「創薬基盤推進研究（ヒトゲノムテーラーメイド研究）」などがある。

(d) 農林水産省

農林水産省の場合、農業生物資源研究所を中心に2000年からのイネゲノム解読があり、2003年からの「カイコゲノム解読プログラム」など、ゲノム解読プロジェクトが行われてきた。また実用化を意識した2004年からの「アグリバイオの実用化・産業化研究」や2005年からの「ゲノム育種による効率的品種育成技術の開発」、2008年からの「新農業展開ゲノムプロジェクト」などがある。

農業分野においても統合データベースプロジェクトが2006年から「新産業創出支援のための農林水産生物ゲノム情報統合データベースの構築」を開始した。

(e) 省庁連携による取り組み

生命科学におけるデータベースは、研究のインフラとして世界中で活発に開発、利用されており、現在では多数のデータベースが乱立しているため、それを使いやすくするための統合化が重要な課題となっている。

データベース統合化の取り組みは、当初、各省で独自に進められていたが、平成18年頃から内閣府の主導のもと、省庁連携によるオールジャパンの協力体制を構築して進められている。平成23年12月には、内閣府総合科学技術会議における議論を受けて、文部科学省、厚生労働省、農林水産省、経済産業省が取り組むデータベースの統合化の方針や成果を紹介する合同ポータルサイト「integbio.jp（インテグバイオ）」(<http://integbio.jp/>) が開設され、4省連携によるデータベース統合化の方針と成果について情報発信し、この取り組みに対する研究コミュニティの関心や理解が深まるよう努めている。

2章 諸外国の状況

1節 米国

米国のライフサイエンス／バイオテクノロジーに関する政策は、米国国立衛生研究所（National Institute of Health、NIH）による生命科学・医学研究に対するイニシアチブ（NIH ロードマップ）、米国エネルギー省（DOE）による資源・エネルギー関連政策（GTL（Genomes to Life）プログラム）が著名なものであるが、米国国立科学財団（NSF）や米国国防高等研究計画局（DARPA）などによるものもある。

1) NIH

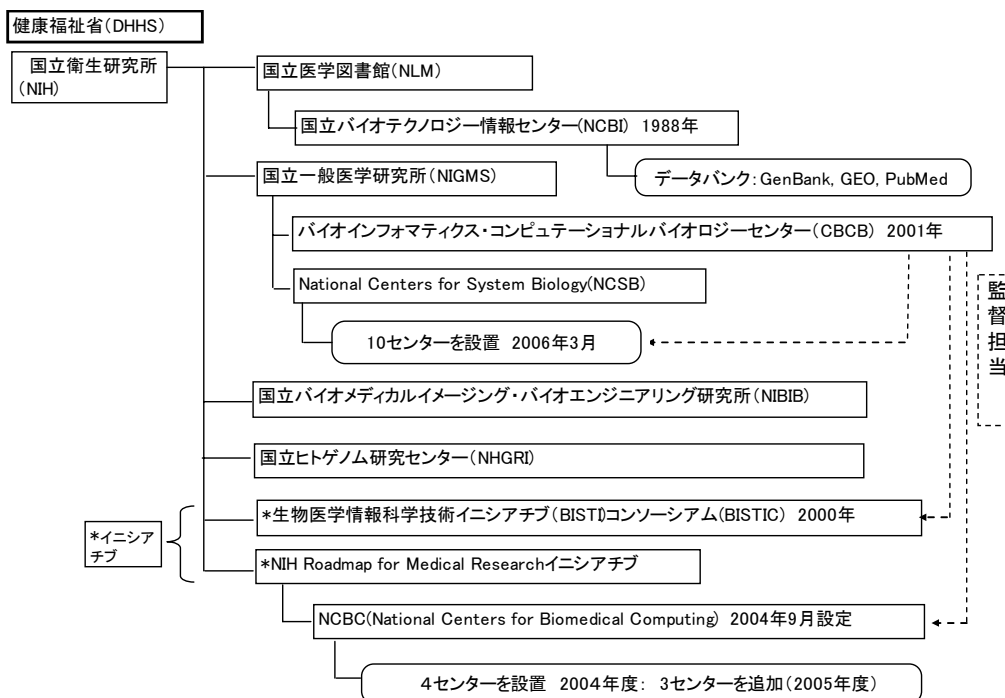
NIHではヒトゲノム計画等のこれまでの成果に加えて更に新たな発見をすることを加速し、その成果が病気に対する、より新しい優れた予防、診断、そして治療に役立つよう、新規に“NIH Roadmap for Medical Research”（NIH ロードマップ、<http://nihroadmap.nih.gov>）という大型プロジェクトを2003年10月（2004年会計年度開始時）からスタートさせた。

NIH ロードマップはNIH傘下の研究所、センターを横断的に結び付けた広範なプロジェクトであり、種々の分野の専門家がチームの一員として参加することで、米国民の求める緊急な医療のニーズ及び健康に関する懸念に対応できる成果を挙げることが期待されている。ロードマップではこれを下記のように3つの分野に分類して各種の課題を設定している。

- | |
|--|
| <ol style="list-style-type: none">1. 発見のための新たな手段の構築（New Pathways to Discovery）2. 未来に向けた研究チームの構築（Research Teams of the Future）3. 臨床研究体制の再編成（Reengineering the Clinical Research Enterprise） |
|--|

NIHは1988年に設立されたNCBI（国立バイオテクノロジー情報センター）を始め、2001年にNIGMS（National Institute of General Medical Sciences、国立一般医学研究所）に親切されることが発表されたCBCB（Center for Bioinformatics and Computational Biology）、2004年にNIHロードマップの一環としてNCBC（National Centers for Biomedical Computing）が創設されるなど、米国のバイオインフォマティクスを推進してきた（図表 IV-5）。

図表 IV-5 NIHのバイオインフォマティクス関連研究所とイニシアチブ



2) DOE

米国では米国エネルギー省 (DOE) が中心になり、エネルギー、資源及び環境問題への対応という視点で、大規模なプログラムが精力的に進められてきた。その一つは多様な微生物のゲノムを基点とする Microbial Genome Program であり、既に100種近い微生物のゲノム解析を完了している。さらにこのプログラムを包含する形で開始されたのが2002年7月に開始された Genomes to Life Program である。これらのプログラムは、世界の環境問題にアプローチする環境微生物学であり、今後10～20年に亘って、自然環境における数千種という微生物や微生物系に対する基本的な理解を得ることを目標にしている。それを通じてDNA配列解析や生物のシミュレーションモデリングなどのシステムバイオロジーを安全・クリーンなエネルギー源の確保や、二酸化炭素の隔離による温暖化対策等に役立てる構想になっている。2005年8月、この計画をさらに発展・拡張した「Genomics GTL(Genomes to Life)」のロードマップを発表し、次世代の生物科学であるシステムバイオロジーをエネルギー・環境分野に適用して行く、今後20～30年間の中期目標とスケジュールを示した。

3) その他

NSF(National Science Foundation)は2007年に1998年から開始した植物ゲノム研究プログラム(Plant Genome Research Program (PGRP))に総額85.8百万ドルの予算を付け、植物のシステムバイオロジーに特に力を入れて推進している。

DARPA(米国国防高等研究計画局) はBio-Computationの研究開発を2000年からの5年間実施し、DNA コンピュータの開発及び、システムバイオロジーのソフトウェアツールとそのオープンソースの枠組みにおいて、生物細胞の時空間での挙動のシミュレーションとモデル化を行った。

この他に、NIST(米国標準技術局)やFDA(米国食品医薬品局)傘下のNCTR(国立毒性研究センター)、Army Research Office などでも、ゲノムやシステムバイオロジーなどの研究が進められた。

2節 欧州

1) EMBL-EBI

欧州では、欧州分子生物学研究所(European Molecular Biology Laboratory: EMBL)及び、その傘下の研究機関としての一部門である欧州バイオインフォマティクス研究所(European Bioinformatics Institute: EBI)、更には英国のサンガーセンターが中心になってバイオインフォマティクスを推進してきた。EMBLは、欧州19か国の出資により1974年に創設され、本部はドイツのハイデルベルクに所在し、イギリスのケンブリッジ近郊のヒンクストンにはEBIがある。これらに加え、フランスのグルノーブル、ドイツのハンブルク等にも研究施設がある。EBIは核酸配列データベースに関する世界の3極体制の1つであるEMBL核酸配列データベースの維持もを行っている。EMBL-EBIは、2006年には組織内外からの研究開発資金として3200万ユーロを受け取っている。

2) 欧州研究開発フレームワーク計画(FP)

欧州研究開発フレームワーク計画(FP)は1984年に開始されたEUの最大の研究開発支援制度である。現在は第7次計画(2007年~2013年)を実施中である。この中でライフサイエンス関連のプログラムが実施されており、図表IV-6に示すような予算が投入されてきた。2002~2006年のFP6ではライフサイエンス分野22.55億ユーロの予算のうち、約11億ユーロは先進ゲノミクス及び健康への応用分野に使われており、その中で様々なプロジェクトが実施された。またシステムバイオロジーが振興科学技術領域(NEST)に取り上げられ、研究の振興が図られた。その結果、この数年で欧州のシステムバイオロジーの研究は、その内容と規模において急速な拡大傾向を示している。

図表 IV-6 欧州研究開発フレームワーク計画（FP）のライフサイエンス関連予算

プログラム	期間	FP全体の予算 (百万ユーロ)	ライフサイエンス関連 分野・プログラム	予算 (百万ユーロ)
FP3	1990-1994	6,600 (5年間)	BIOTECH1 プログラム	184
			BIOMED1 プログラム	149
FP4	1994-1998	11,900 (5年間)	BIOTECH2 プログラム	596
			BIOMED2 プログラム	374
FP5	1999-2002	13,700 (4年間)	生活の質・生活資源の管理分野	2,470
FP6	2002-2006	17,500 (5年間)	健康のためのライフサイエンス・ゲノ ミクス・バイオテクノロジー分野	2,255
FP7	2007-2013	50,500 (7年間)	健康分野	6,050
			食糧・農業・バイオテクノロジー分野	1,935

3) その他

欧州の各国でもライフサイエンス／バイオインフォマティクス関連の研究が進められており、例えば英国ではBiotechnology and Biological Sciences Resesarch Council (BBSRC:生物科学研究評議会)とMedical Research Council (MRC:医学研究評議会)が中心となって推進・実施している。総予算は、2007/2008年はそれぞれ3億8,200万ポンド、5億4,600万ポンドで、2002/2003年は合計6億500万ポンドであった。BBSRC (Biotechnology and Biological Sciences Research Council)は2005年～2008年の計画(Delivery Plan)として統合・システムバイオロジー、植物に対する基礎研究及びゲノム解析等の分野を優先分野として位置づけ、6ヶ所のシステムバイオロジーセンターを設立した。

ドイツでは2000年にドイツ連邦教育研究省 BMBF (Federal Ministry of Education and Research) が同国のバイオテクノロジー研究助成の枠組みを策定し、2001年～2005年までの5年間に8億300万ユーロの資金を投入する「バイオテクノロジー」プロジェクトを実施した。このプログラムは後に、「Biotechnology - using and shaping its opportunities」と呼称されるフレームワークプログラムに編入され、プログラム内容は、分野的に、基礎整備、基盤革新、応用研究、予防研究に分割された。

その他、フランスやスイスなどでも研究機関の設立やシステムバイオロジープロジェクトの実施が進められた。

3節 その他

1) 中国

中国における科学技術開発の政策（プロジェクト）は、以下に述べる主要な計画（プロジェクト）・基金委員会などとの関連で決定される。主要な計画などとして以下のようなものがある。

- ・ 863 計画（ハイテク研究発展計画）
- ・ 973 計画（国家重点基礎研究発展計画）
- ・ 国家中長期科学技術発展計画
- ・ 国家科技攻関（難関攻略）計画
- ・ 国家自然科学基金委員会
- ・ 各 5 カ年計画

2006 年から始まった中国第 11 次 5 カ年計画（2006 年～2010 年）での 863 計画では、バイオインフォマティクスは特定テーマ「生物情報と生物計算技術」に位置付けられ、2006 年～2008 年に渡っておよそ 1.65 億元の予算が割り当てられた。

2007 年には、中国人のヒトゲノムマップ「炎黄 1 号」が中国研究者により始めて完成されたと発表された。このプロジェクトは深セン華大ゲノム研究院及びバイオインフォマティクス国家工程研究センター、中国科学院北京ゲノム研究所の研究者が共同して進められた。この他、図表 IV-7 に示すように様々なバイオインフォマティクス関連のプロジェクトが進められた。

図表 IV-7 中国の主要なバイオインフォマティクス関連プロジェクト

担当計画名	プロジェクトの名称
国家重点計画	・ ヒト肝臓プロテオームプロジェクトにおけるバイオインフォマティクス研究と関連したデータベースの構築
973 計画	<ul style="list-style-type: none"> ・ ヒトの重篤な病気のプロテオミクスの研究 ・ 幾つかの生命科学的疑問に対する先端のおよび学際的研究 ・ 遺伝子機能予測のための系統的解析モデルと検定方法 ・ 多遺伝子病のシステムバイオロジー的研究
863 計画	<ul style="list-style-type: none"> ・ 国家バイオインフォマティクス研究基盤の確立 ・ プロテオミクスプラットフォームの構築とそのガン研究における応用 ・ バイオインフォマティクスの理論の研究とその応用 ・ 中国全住民（民族）のための遺伝子多様性データベースの構築 ・ プロテオミクス技術による抗原抗体反応を含んだヘテロタンパク質の同定
中国科学院計画	<ul style="list-style-type: none"> ・ 中国人の DNA データベース ・ ポストゲノム時代における大規模機能解析
国家自然科学基金計画	・ Signal Transduction の構築、それらモデル化および再配列

2) 韓国

韓国では、2008年国家科学技術委員会が提出した「新技術融合型成長動力推進企画」(以下「企画」と略)において、始めて正式に国家レベルでバイオインフォマティクスを取り上げた。この「企画」によると今後2009年～2013年の5年の間、バイオインフォマティクス技術基盤、バイオテクノロジー基盤、ナノテクノロジー基盤、環境技術基盤を中心に研究を進めている。

また近年になってゲノミクスに対する積極姿勢が伺われるようになった。2009年には、ソウル大学が、英との共同研究で、30代の健常な韓国人男性のゲノム配列を決定している。次世代シーケンサーの導入台数は、日本と同等程度である。JST 研究開発センターが2009年に発表した「ライフサイエンス分野 科学技術・研究開発の国際比較」では、幹細胞、クローン動物、コホート研究、植物科学などの分野で日本に比肩する研究水準を持つとしており、今後の融合研究への動向に注目している。また、サムスングループがシーケンサーの開発に乗り出したことを発表している。

3) インド

2004年インドは、ITとバイオテクノロジーを一体化してバイオインフォマティクスの世界の中心とするためにBiotech-IT parkを作った。また2007年、C-DAC (Center for Development of Advanced Computing) はバイオインフォマティクスのアプリケーション専用のスーパーコンピュータを立ち上げると発表し、2011年末の情報では、National Param Supercomputing Facility (NPSF)に100ギガFLOPSの施設と、C-DAC's Terascale Supercomputing Facility (CTSF)に1テラFLOPSの施設を有している。

インド科学技術省バイオテクノロジー局は、バイオインフォマティクス分野における研究拠点として5つの研究機関を選定し、バイオインフォマティクス研究を振興している。

4) シンガポール

シンガポールは周辺の他のアジア諸国と異なりバイオ産業基盤としての生物資源が少ないという不利を補うために情報技術を活用するバイオインフォマティクスに古くから力を入れ始めたと言われている。2001年にバイオ情報研究所をBiopolisに設立し、バイオ関連データの収集、バイオ医療分野でのコンピュータの高度利用、バイオ分野の分子構造の画像処理、医薬品設計などを研究する役目を担っている。その他、NTU(Nanyang Technological University)は最近BIRC (Bioinformatics Research Centre) を設立し、HP (ヒュレットパッカードコンピュータ・アジア) と連携し、バイオインフォマティクス研究を進めている。

3章 論文・特許による研究の変遷の把握

1節 論文調査

1) 調査の概要

バイオインフォマティクス関連分野の最新の学術研究の状況を把握するため、学術論文データベースを用いて、論文発表の状況を調査した。対象領域は BIRD 事業による発表論文に基づき定義し、世界全体の傾向および主要国間の比較を行った。また、バイオインフォマティクス内に 14 の領域を設定し、領域別の傾向を把握した。

2) 調査結果

学術論文データベース “Thomson Reuters Web of Knowledge” を用い、世界、日本および主要国（米国、ドイツ、英国、中国、韓国）別にバイオインフォマティクスおよびその応用分野の論文を抽出した。抽出には、BIRD事業の各課題による論文の投稿先学術誌を調査し、論文数上位 5 誌を主要 5 誌と位置づけ、主要 5 誌に掲載された論文の推移を比較した(図表 IV-8)。

キーワードを設定し(図表 IV-9)、最近数年の分野の推移を見ると、DNA・ゲノム、タンパク構造、データベース、大腸菌といった分野における論文数は 2005 年頃までの一時期の勢いが衰え、その一方でアルゴリズムは急速な増加、パスウェイは漸増傾向が見られることから、バイオインフォマティクス分野の研究の傾向が、データの蓄積から徐々にデータの応用、利活用に移行していることが推察される(図表 IV-10)。

世界のバイオインフォマティクス関連論文数は、1991 年から 2002 年まではほぼ横ばい、2003 年から 2005 年までは急増し、その後は漸減傾向にある(図表 IV-11)。

バイオインフォマティクス関連論文の 2010 年時点のシェアは米国が約 40%以上と最大である。日本は米国に次いで大きく約 20%弱である(図表 IV-12、図表 IV-13)。

著者が日本に所在する論文（日本論文）は 1991 年から 2010 年の間、漸増傾向にある。これに対して米国は 2003 年から 2005 年まで急増したものの、その後減少するなど大きな論文数の変化が見られる。この間日米ともにライフサイエンス分野の政府予算が急増し、その後減少に転じていることから、この影響を受けた可能性が考えられる(図表 IV-14、図表 IV-15)。

日本と米国の共著は 1991 年以降、増減はあるが 2006 年ころまでは増加傾向にあったといえることができる。しかし、その後は減少傾向にある。日本と中国の共著数は 2002 年から 2007 年にかけて増加したが、その後は減少している。米国との共著数は中国との共著数よりもかなり高い水準である。米国と中国との共著は 1999 年から急速に増加しており、現在では日本と米国の共著の約 2 倍の水準である(図表 IV-16)。

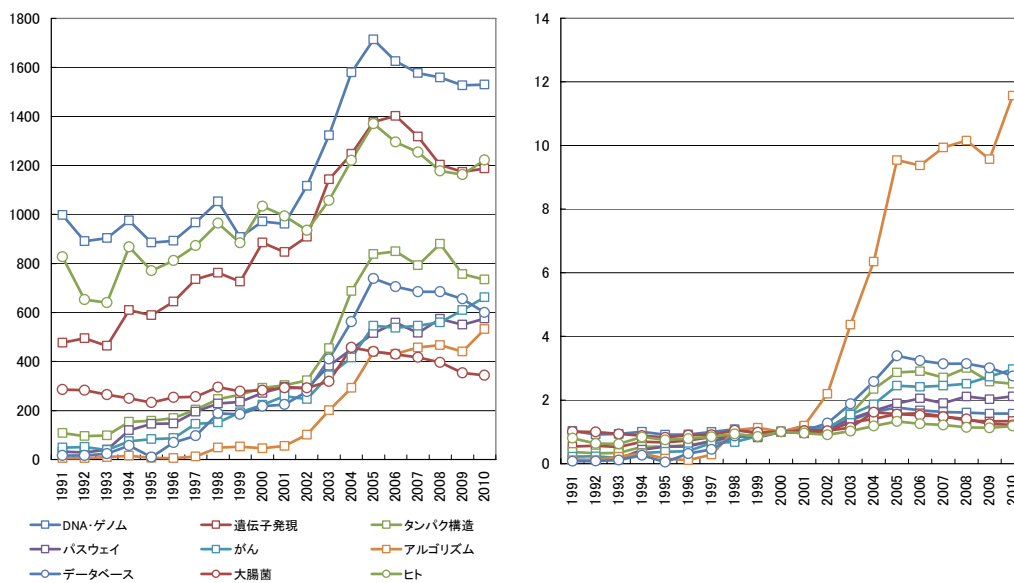
図表 IV-8 BIRD 事業の代表論文による主要 5 誌の選定

順位	掲載誌名	論文数	割合	課題数	5誌 合計 課題数	全課題 に対する 比
1	NUCLEIC ACIDS RESEARCH	23	11.9%	12	29	0.725
2	BIOINFORMATICS	20	10.3%	14		
3	PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS	8	4.1%	6		
4	BMC BIOINFORMATICS	7	3.6%	7		
5	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS	6	3.1%	5		
6	DNA RESEARCH	5	2.6%			
7	GENE	4	2.1%			
7	NATURE	4	2.1%			
7	PROTEIN SCIENCE	4	2.1%			
10	CHEMICAL PHYSICS LETTERS	3	1.5%			
10	DEVELOPMENTAL BIOLOGY	3	1.5%			
10	JOURNAL OF BIOLOGICAL CHEMISTRY	3	1.5%			
10	MOLECULAR BIOLOGY AND EVOLUTION	3	1.5%			
10	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	3	1.5%			
11位以降		98	50.5%			
計		194	100.0%			

図表 IV-9 領域分類と設定のためのキーワード

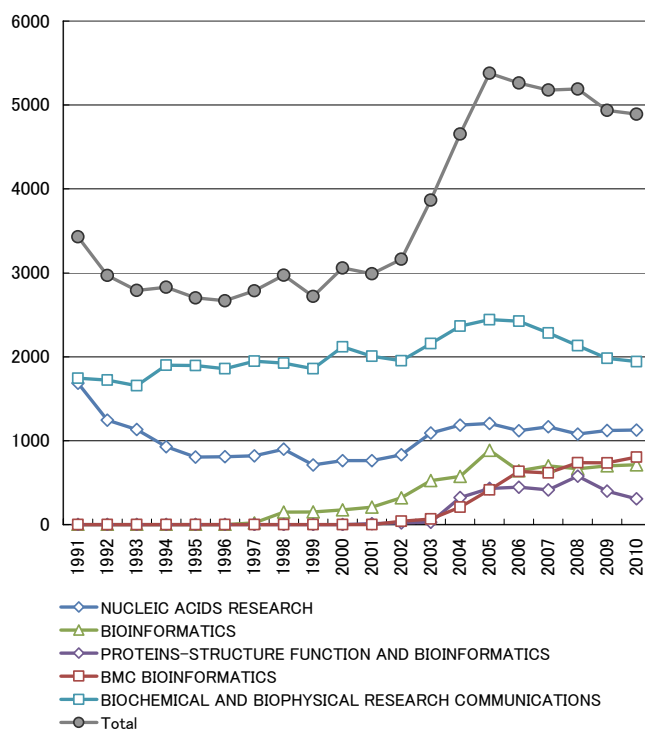
領域・分野	キーワード
DNA・ゲノム	DNA or GENOME
遺伝子発現	GENE and EXPRESSION
タンパク構造	PROTEIN and STRUCTURE
パスウェイ	PATHWAY
がん	CANCER
アルゴリズム	ALGORITHM
データベース	DATABASE
大腸菌	ESCHERICHIA-COLI
ヒト	HUMAN

図表 IV-10 分野別の論分数の推移

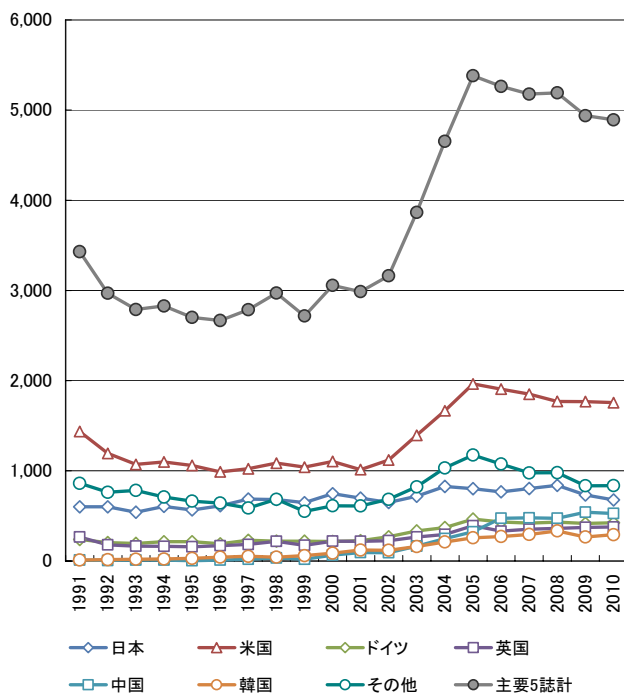


※左図:論文数、右図:2000年を1.0とした時の指数

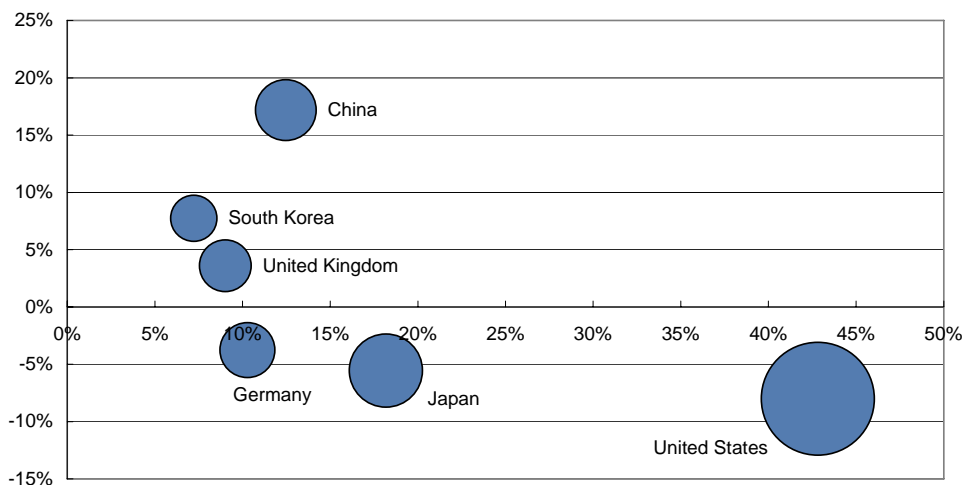
図表 IV-11 主要5誌記載論分数の推移



図表 IV-12 各国の論分数の推移 (主要5誌合計)



図表 IV-13 各国の論文シェアと論文数の増加率

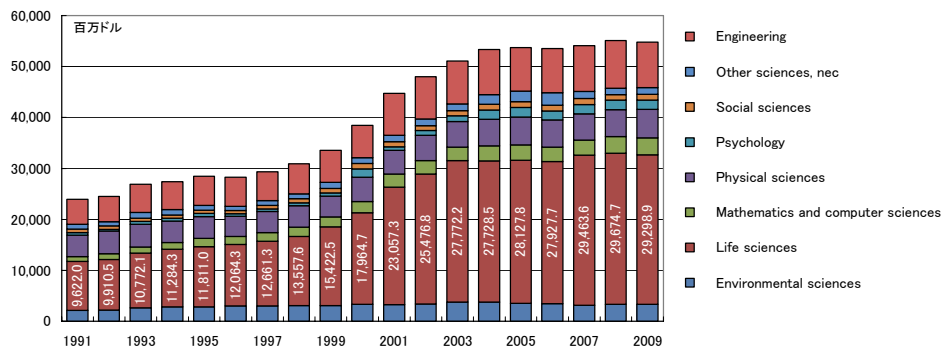


横軸：論文シェア（2008～2010年）

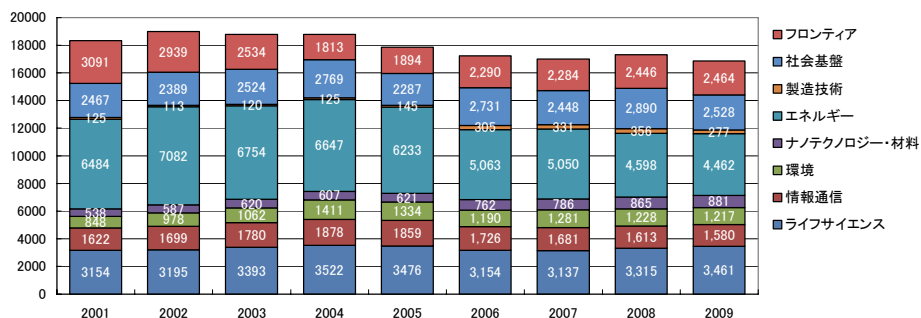
縦軸：増加率（2008～2010年）

円の半径：論文数（2008～2010年）の規模

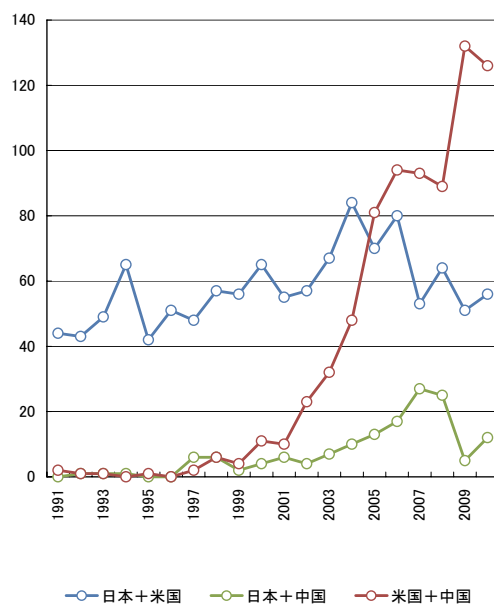
図表 IV-14 米国の科学技術予算の推移（連邦予算、分野別）



図表 IV-15 日本の科学技術予算の推移（国の予算、分野別）



図表 IV-16 日本、米国、中国の共著者数の比較（主要5誌合計）



2節 特許調査

1) 調査の概要

バイオインフォマティクスに関連する研究開発の産業への応用の状況を把握するため、同分野の特許出願状況を調査した。従来、バイオインフォマティクスは生物学、医学の基礎研究における、情報技術を活用した研究支援として認識されることが多かったが、近年、診断、治療、予防などのヒト臨床応用、また、農林水産、食品産業への応用例が増加しつつある。このため、調査の対象を従来のバイオインフォマティクスから、周辺分野を含む領域に広げ、そこへのバイオインフォマティクス技術の応用の状況を把握することに努めた。

世界の出願特許を網羅したデータベース“PATSTAT”を用い、世界、日本および主要国（米国、ドイツ、英国、フランス、中国、韓国）別にバイオインフォマティクスおよびその応用分野の特許を抽出した。特許の抽出には、世界知的財産機構（WIPO）が定義する 35 技術区分を利用し、加えてキーワードを指定することにより、バイオインフォマティクス関連特許を特定した(図表 IV-17)。

2) 調査結果

全体では 1990 年代に急激な出願数の増加があったが、2000 年をピークとして減少傾向に転じている（図表 IV-18）。出願人所在国別の状況では、1991 年から 2007 年の累積件数では依然として米国がもっとも大きな出願国であるが、近年米国の出願は減少し、2007 年単年でみるとフランスが最大の出願国となっている（

図表 IV-19）。日本の 1990 年代から 2007 年までの出願数は 2004 年にやや減少があったが、全体として微増傾向にあり、2007 年には米国、ドイツを上回る出願があった。

2001 年から 2007 年出願分について、主要出願人を見るとフランスの公的研究機関が上位（1, 3 位）を占めている（図表 IV-20）。日本については、理化学研究所、JST、産業技術総合研究所といった公的研究機関が上位（3, 5, 12 位）にあり、これに加えてシスメックス、日立製作所といった民間企業が上位出願人となっている。

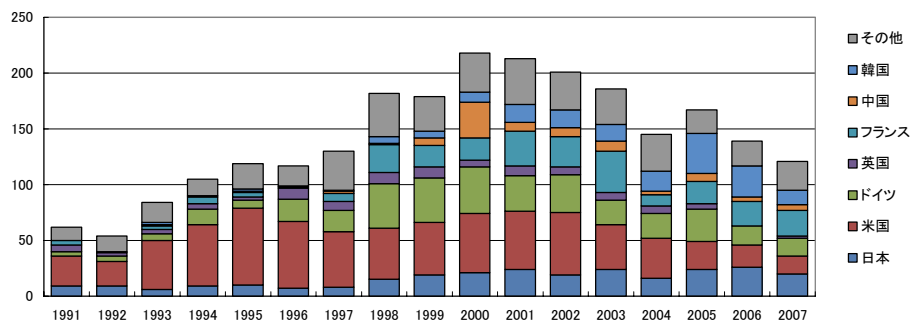
分野別の出願数の傾向では、DNAゲノム、遺伝子解析の出願数は 2000 年ころを境に減少傾向が顕著であるが、その一方、がんといった応用に近い分野についてははっきりとした減少傾向は見られない（図表 IV-21）。

上述のように、JST は出願人として理化学研究所と並んで上位にあり、BIRD 事業を含む研究助成事業が出願に結びついた可能性が大きい。実際「ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築(伊藤啓)」や「線虫 *C.elegans* 発生過程のシステム解析(大浪修一)」の成果は特許成立あるいは申請中のものがある。その成果の社会還元を確認する観点から、これらの特許が具体的にどう活用されたかを追跡し、振り返って BIRD 事業の有効性、実効性を評価することも必要となろう。

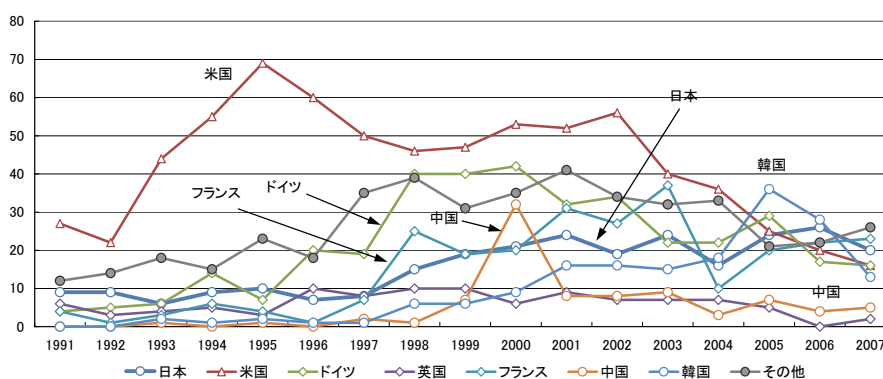
図表 IV-17 バイオインフォマティクス関連特許の定義

条件 A	WIPO 35 技術分類	生物材料の分析
		バイオテクノロジー
		薬品・製薬
条件 B	発明の名称に右の いずれかを含む	DNA or genome
		gene and expression
		protein and structure
		pathway
		algorithm
		database

図表 IV-18 出願数の推移（主要国別）



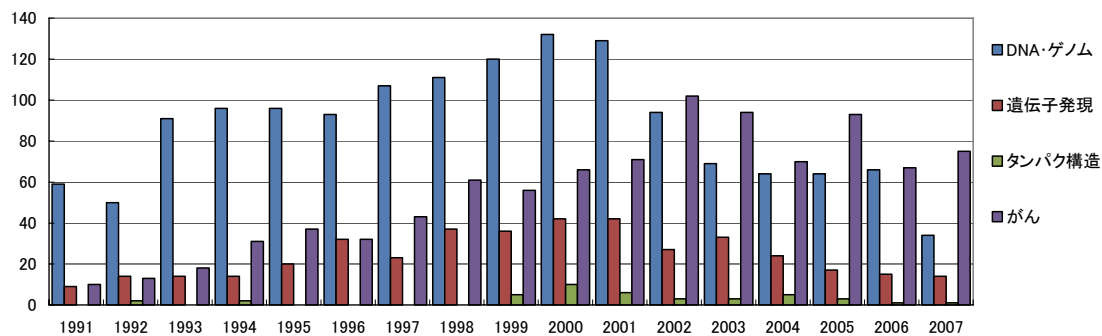
図表 IV-19 出願数の推移（主要国別、国別）



図表 IV-20 上位出願人（世界、2001～2007年）

順位	出願人名称	所在国	出願数
1	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)	フランス	24
1	EPIGENOMICS AG	ドイツ	24
3	INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM)	フランス	17
3	理化学研究所	日本	17
5	科学技術振興機構	日本	16
6	BIOMERIEUX SOCIETE ANONYME	フランス	14
6	SAMSUNG ELECTRONICS CO., LTD.	韓国	14
8	KOREA RESEARCH INSTITUTE OF BIOSCIENCE AND BIOTECHNOLOGY	韓国	13
9	SYSMEX CORP	日本	10
10	COMMISSARIAT A L'ENERGIE ATOMIQUE	フランス	9
10	SIEMENS AG	ドイツ	9
12	HITACHI LTD	日本	8
12	INSTITUT PASTEUR	ドイツ	8
12	産業技術総合研究所	日本	8
15	AGENSYS, INC.	米国	7
15	THE REGENTS OF THE UNIVERSITY OF CALIFORNIA	米国	7

図表 IV-21 分野別出願数の推移



V バイオインフォマティクスの今後の展望

1章 基礎研究へのインパクト

BIRD 事業が生物科学基礎研究の各分野に対して与えたインパクトについてまとめる。

① ゲノム

BIRD 研究開発課題としては、「2 項関係に基づくゲノムと生命システムの機能解読 (金久實)」、「ゲノム生物学のバックボーンデータベースの構築提供 (菅原秀明)」、「マルチモーダル統合バイオ DB (森下真一)」をはじめ多くの研究開発課題がゲノム研究に関連していた。上述の 3 課題は、ゲノム研究にのみ寄与するに留まらないが、ゲノム研究にインパクトを与える代表格としてここに挙げた。これらのテーマには、ゲノム配列解析支援が含まれ、データベースに格納されるデータの充実と、それに対する解析技術や検索技術の高度化が行われた。「2 項関係に基づくゲノムと生命システムの機能解読」の成果としては、ゲノムデータに加えて代謝化合物データや、代謝・シグナル伝達系のパスウェイデータを関連付けたことデータの整理と検索ツールの拡充が進めたことによって、幅広い生命現象についてのデータを容易に抽出可能としたことが挙げられる。このパスウェイデータは、生命現象のモデル化・シミュレーションに利用されており、システム生物学の基盤整備に貢献した。「ゲノム生物学のバックボーンデータベースの構築提供」では、DDBJ、PDB、KEGG、GTOP 等のデータベースを統合する技術が成果として挙げられている。次世代シーケンサーから生産されるデータに対してのメタデータを作成するツールが国際的にも評価され、NCBI/EBI にもそのソースコードが提供されており、次世代シーケンサーに関わる情報基盤整備に貢献した。GTOP の技術により微生物ゲノムの ORF 予測精度を向上させることに成功させており、微生物研究対してもインパクトを与えた。微生物研究に対する寄与という観点では、「大規模な比較ゲノム研究を展開するためのワークベンチの構築 (内山郁夫)」により開発されたオーソログ解析システムがピロリ菌の解析に活用された。「マルチモーダル統合バイオ DB」では、超高速シーケンサー (いわゆる次世代シーケンサー) によって生成される大量のゲノム配列データに対処可能な大規模並列計算手法が開発された。これを活用することによって遺伝的多様性における周期性とクロマチン構造とが相関するという知見も成果として挙げており、次世代シーケンサーに対する情報基盤整備とそれを活用したゲノム構造研究にインパクトを与えた。

ゲノム研究の発展系としてメタゲノム研究が、近年になって盛んに研究されるようになった。世界的には、人体上の細菌叢に対するメタゲノム解析プロジェクトとして MetaHIT や HMP が実施されており、それに対応するものとして「メタゲノムオーソログ遺伝子統合解析システムの開発 (黒川顕)」が実施された。この中で、メタゲノム配列データに対する配列解析技術と、メタゲノムデータを対象とするデータベースが成果として得られており、日本としての精力的な取り組みがあることを国際的な研究コミュニティに示すことができた。

② メタボローム

「メタボローム MS スペクトル統合データベースの開発（西岡孝明）」によって、二次代謝物質の質量、化学構造式、生物種等の収集データベースと、生体成分を各種質量分析計によって測定した MS スペクトルデータベースと、ピーク同定技術が成果として得られた。ここで開発された MassBank が日本質量分析学会の公式データベースとして認められ、日本のメタボローム研究にとって重要な研究基盤となった。

③ トランスクリプトーム

この研究カテゴリに対する直接的な研究開発課題として、「ヒト遺伝子の転写・発現の耐用性解明を目指した基盤データベースの開発（矢田哲士）」、「転写制御領域の構築原理解明（中井謙太）」が挙げられる。転写制御領域・遺伝子コード領域のモデル化・予測と、そのデータベース化が行われ、その結果としてトランスクリプトーム研究基盤がより強固なものとなった。「転写制御領域の構築原理解明」では、転写制御領域に対する統合的な解析結果が成果として得られた。また、副産物としてエピジェネティクスに関する研究成果も得られており、この萌芽的研究領域に刺激を与えた。また、「ゲノム生物学のバックボーンデータベースの構築提供（菅原秀明）」において、国際標準フォーマットに準拠した発現プロファイルデータベース（CIBEX）が得られた。EBI の ArrayExpressn ならびに NCBI の GEO との国際的に協力関係を築くにまで至っており、国際的な研究基盤整備に貢献した。

④ インタラクトーム

「インタラクトーム解析からの生物知識獲得（伊藤隆司）」や「高精度タンパク質間相互作用予測システムの開発（清水謙多郎）」をはじめとして、「創造的な生物・情報知識融合型の研究開発」には、生体高分子の相互作用についての様々な角度からの取り組みが見られた。「生命情報データベースの高度化・標準課題」においても、「シグナルオントロジーとバイオタームバンクの開発（高木利久）」と「2 項関係に基づくゲノムと生命システムの機能解読（金久實）」で、その基盤技術開発の一環としてのオントロジーの整備と、データベース高度化とが行われた。

「インタラクトーム解析からの生物知識獲得」では、GATC-PCR 法の確率が成果として挙がっており、遺伝子発現解析だけでなく転写因子-DNA 相互作用の解析をはじめ様々な応用を可能にしたことによりインタラクトームをはじめとする分子生物学の幅広い領域にインパクトを与えた。

「シグナルオントロジーとバイオタームバンクの開発」では、パスウェイ情報の新しい表現方法が成果として得られた。国際標準プロジェクトの BioPAX で採用され、国際的な研究基盤構築に貢献した。

⑤ 蛋白質立体構造

「蛋白質構造データバンクの国際的な構築と高度化(PDBj) (中村春木)」をはじめ、「高速計算機システムによる蛋白質フォールディングの研究 (肥後順一)」や「蛋白質の構造・機能・相互作用予測システムの開発と展開 (太田元規)」等がこの研究カテゴリに含まれる。「蛋白質構造データバンクの国際的な構築と高度化(PDBj)」は、国際的研究推進に対する日本としての取り組みを体現したものである。これらの研究開発課題によって幾つもの解析ツールが開発され、研究開発基盤が強化された。「蛋白質の構造・機能・相互作用予測システムの開発と展開」の成果である予測ツールが、複合体予測コンテストである CAPRI において正解に近い構造を返答できるにまで至っており、国際的な研究コミュニティに対して日本の研究レベルの高さを示すことができた。

「実践による超分子ネットワークモデリングシステムの開発 (白井剛)」や「タンパク質化合物相互作用の網羅的予測手法とデータベースの開発 (榊原康文)」、「酵素反応分類に基づく酵素反応予測システムの開発 (長野希美)」でも有用且つユニークな取り組みがあった。「タンパク質化合物相互作用の網羅的予測手法とデータベースの開発」では、新規リード化合物発見が成果として挙がっており、創薬研究に対してインパクトを与えた。「酵素反応分類に基づく酵素反応予測システムの開発」では、酵素反応を予測する機械学習法を成果として得ており、配列データや立体構造データを考慮して酵素反応を再定義する試みが、独自性の高い取り組みとして評価された。

「生命科学上の非構造化データの統合マイニング (馬見塚拓)」では、頻出パターンマイニングとクラスタリングといった機械学習手法が成果として得られており、タンパク質と低分子化合物の結合に重要となる分部構造を抽出できることを示した。

⑥ 進化

進化は生物学の全体に跨る興味対象であり、多くの BIRD 研究開発課題が関連している。その中でも「ゲノム進化とマッピングの階層モデルと解析アルゴリズムの開発 (岸野洋久)」、「大規模な比較ゲノム研究を展開するためのワークベンチの構築 (内山郁夫)」が代表的なものである。前者では進化研究に当たって中心的な手法となる系統樹推定アルゴリズムが得られ、後者では解析プラットフォームが成果として得られた。ゲノムデータが増加する中、これらの成果に含まれる比較ゲノム解析的アプローチが進化研究で取り入れられた。

⑦ 発生

「ホヤプロテイン統合データベースの構築（稲葉一男）」、「線虫 *C. elegans* 発生過程のシステム解析（大浪修一）」、「ヒト胚の携帯発生に関連する三次元データベース（塩田浩平）」といった研究開発課題が発生に関連するものとして挙げることができる。この研究課題によって、モデル生物種の基盤データベースが開発された他、発生過程を時系列的・形態的に解析するための新技術として時系列解析技術や画像解析技術の開発が行われた。「ホヤプロテイン統合データベースの構築」ではCIPROデータベースが成果として得られた。ホヤ研究者コミュニティにおいて国際標準として取り扱われるに至った。「ヒト胚の携帯発生に関連する三次元データベース」では、ヒト胚の3次元形態データベースが成果として得られた。これは、京都大学のヒト胚標本コレクションを対象として、高解像度の3次元形態データをデータベース化したものであり、新規性が高い研究として評価された。関連する取り組みとして「遺伝子破壊株イメージ・マイニング（森下真一）」がある。対象生物種が酵母であり、単細胞生物であるため発生とは言えないが、遺伝子ノックアウトによる形態的変異を画像解析によって定量化する手法・データベースが成果として得られた。線虫 *C. elegans* 発生過程のシステム解析」では、発生に関わる時空間的なシステムのモデル化・シミュレーション手法が成果として得られた。これらは発生や個別生物種に関する基礎研究のみならずシステム生物学にもインパクトを与えた。

⑧ 脳

これまでのバイオインフォマティクス研究に対して、BIRD事業の研究開発課題には脳に関するものが含まれていることが特徴的である。「ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築（伊藤啓）」、「脳スライス中で可視化した神経シナプスの自動解析（川戸佳）」、「マウスを用いた脳機能表現型データベースの開発（宮川剛）」と数も多い。これらには画像解析による表現型の定量化が取り入れられるなど新しい取り組みが含まれている。「マウスを用いた脳機能表現型データベースの開発」では、遺伝子ノックアウトによる90系統以上のマウスの行動解析が行われており、成果としてそれらデータを格納したデータベースが得られた。「進化型計算と事故組織化による適応的画像分類法の開発（馳澤盛一郎）」では、反応予測、画像解析技術が成果として得られた。機械学習、画像解析技術の高度化が主たる目的であったが、ここで開発された技術が脳神経組織に関する時系列画像の解析に用いられるという展開があった。これらの研究成果は、高次生命現象を理解するための新しい方法論を模索したものであり、脳を始めとする関連研究に対してインパクトを与えた。

BIRD 事業の研究開発課題は生物学の幅広い領域にまたがっており、ツール開発やデータベースの構築といった基盤整備から新知見の獲得研究に直接的に貢献するものが存在した。本章では、ゲノムから脳までの8領域について、基礎研究へ与えたインパクトをまとめた。各領域でデータベースの構築・拡張が行われており、KEGG、PDBj、CIBEX、CIPROらが国際的にも認められるようになったことを始め、MassBankが日本質量分析学会の公式データベースとなった。トランスクリプトームではGATC-PCR法が、タンパク質立体構造研究ではフォールディングやドッキングの解析・シミュレーション技術が、脳研究では画像解析を含めた表現型解析技術が開発されるなど、これまでの技術レベルを押し上げる研究開発はもちろん、新しい方法論を模索するものまでが存在した。ゲノムでは遺伝的多様性の周期性とクロマチン構造とに相関関係があることや、トランスクリプトームではエピジェネティクスに関する萌芽研究が行われるなど、新規生物学的知見の獲得や、将来を先導する研究が行われた。BIRD研究課題全体としては、国際的に認められる研究基盤構築と、システム生物学や高次生命現象を対象とする研究等の近年における研究トレンドを牽引する活動が行われたことが基礎研究における成果と要約することができる。

2章 産業への展開

III2章3節「特筆すべき成果」で整理したように、BIRD事業の研究開発課題の中には創薬や医学、環境分野への応用が期待される課題がいくつかあることが分かった。本章では①創薬、②診断・医療、③環境の3分野についてBIRD事業成果の産業展開の可能性についてまとめる。また、バイオインフォマティクスの解析技術やデータベースが商業製品として産業展開する可能性について述べる。最後に近年開発・導入が進みつつある次世代DNAシーケンサーが与えるバイオインフォマティクスへの影響と、その結果としての産業展開についても最後にまとめる。

① 創薬分野への展開

BIRD事業の研究課題のいくつかが本事業の事後評価結果または発表論文の中で創薬分野への展開の可能性が言及されている。その中でも特に創薬との関連性が深い研究がなされているのが「ゲノムと環境の統合解析による生命システムの機能解読(KEGG)(金久實)」である。本研究課題では、疾患パスウェイマップを多数追加するとともに、疾患や医薬品の情報を蓄積したKEGG DISEASEやKEGG DRUGの充実が図られている。実際、代表研究者らによる発表論文[1]では創薬ターゲットの相互作用を予測することを目的に、KEGGデータベースのうちKEGG GENES、KEGG DRUG、KEGG LIGAND、KEGG BRITEから遺伝子、化合物、相互作用の情報をもとにターゲットの予測を行っている。また薬理学的な情報として日本医薬情報センター(JAPIC)が更新する医薬品添付文書の情報を用いている。また、発表論文[2]では医薬品の有害な相互作用の分類・分析を行っており、創薬において非常にさらに有用な成果といえる。

このようにKEGGは従来のパスウェイデータベースから本事業の中で開発したKEGG DISEASEやKEGG DRUGなどの整備によって創薬分野への展開がより近くなったものと考えられる。なお、KEGG以外の研究課題についても、例えば、「ダイナミクスを考慮した膜蛋白質の構造モデリング法の開発(水口賢司)」や「タンパク質化合物相互作用の網羅的予測手法とデータベースの開発(榊原康文)」は創薬のターゲット探索に関わる研究であり、KEGGと同様に創薬への展開の可能性が期待される。さらに、「タンパク質の構造・機能・相互作用予測システムの開発と展開(太田元規)」は立体構造予測の観点からリガンド結合情報は創薬に非常に有用である。

[1] Yamanishi, Y., Kotera, M., Kanehisa, M., and Goto, S.; Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26, i246-i254 (2010).

[2] Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., and Kanehisa, M. Characterization and classification of adverse drug interactions. *Genome Informatics* 22, 167-175 (2009).

② 診断・医療分野への展開

診断・医療分野への産業展開の可能性が期待される BIRD 事業の研究課題としては、「ヒト胚の形態発生に関する三次元データベース（塩田浩平）」などが挙げられる。本研究課題では、ヒトの発生と先天異常のゲノムワイドな解析を行うためのインフラとなる形態データベースを構築するため、京都大学医学研究科附属先天異常標本解析センターに所蔵される世界最大のヒト胚標本コレクションの標本を用いて、器官形成期を中心としたヒト胚の画像ならびに各症例の臨床データのデータベースを構築することを目的とした。本課題で構築したデータベースは世界に類例のないヒト胚三次元形態画像データベースであり、また今後得ることの困難で重要なデータのデータベースでもあり、ほぼ確実に医療等の分野に大きな波及効果をもたらすと期待される。

③ 環境分野への展開

環境分野への産業展開の可能性が期待される BIRD 事業の研究課題としては、まず「ゲノムと環境の統合解析による生命システムの機能解読 (KEGG) (金久實)」が挙げられる。本研究課題では、ゲノムと環境との相互作用の1つの側面として、植物・微生物による二次代謝物質合成経路と、微生物による環境物質分解経路の知識集約を行い、またこれら物質の化学構造を KEGG BRITE 機能分類とともに蓄積されている。また発表論文[1]の研究においては、KEGG データベースを使ったエストロゲン様内分泌かく乱物質候補の予測を行っている。このように KEGG データベースは環境分野において生分解の研究や環境ホルモンの予測・対策の研究に活用できることが期待される。

この他に BIRD 事業の研究課題の中で環境分野への展開が期待されるものとしては、「メタゲノムオーソログ遺伝子統合解析システムの開発（黒川顕）」がある。本研究課題では、環境中の細菌叢やヒトの腸内細菌叢のメタゲノムを分析しデータベース化することを目的としており、環境分野だけでなく、食品や医療・健康分野への貢献が大いに期待される。

[1] Kadowaki, T., Wheelock, C.E., Adachi, T., Kudo, T., Okamoto, S., Tanaka, N., Tonomura, K., Tsujimoto, G., Mamitsuka, H., Goto, S., and Kanehisa, M.; Identification of endocrine disruptor biodegradation by integration of structure-activity relationship with pathway analysis. *Environ. Sci. Technol.* 41, 7997-8003 (2007).

④ 商業製品としての展開

BIRD 事業で開発されたバイオインフォマティクス技術やデータベースの中で商業製品として利用されているものとして KEGG がある。その事例としてタカラバイオのプライマー設計・注文システム「Perfect Real Time サポートシステム」を紹介する。本製品はプライマーを設計し、オンライン注文ができるシステムであり、遺伝子名からプラ

イマー検索を行い、検索結果の遺伝子が属するパスウェイをKEGGデータベースで参照することができる。また逆にパスウェイマップを使って、同じパスウェイ上に属する遺伝子を関連遺伝子として追加検索ができる機能が設けられている。このようにプライマー設計のように欲しい情報を調べたい際にベースとなる知識として非常に活用されている。

この他にKEGGは様々なサービスプロバイダーに利用されている。多くはバイオインフォマティクス分野の分析ツール会社であり、KEGGを利用することで提供サービスの付加価値をより高めていると考えられている。その他にはSAS Instituteなどのように汎用統計パッケージ会社などにも利用されている。このようにKEGGは商業的にも大いに産業展開できているといえる。

一方、他の研究課題については現状なかなか商業製品にまでは展開できてないが、今後展開していくためには、1)利用価値を高める、2)製品としての完成度を高める、3)認知度を高める、などの取組みを行っていく必要があると考えられる。

⑤ DNAシーケンサーに係わる動向とバイオインフォマティクスとの係り

ヒトゲノム計画完了の2003年までは従来型のキャピラリーシーケンサーが主流であったが、基礎研究・臨床研究現場における更なるハイスループット化の要望に応えられないことから、NHGRI(米国立ヒトゲノム研究所)が”Advanced Sequencing Technology Awards”として大規模なグラントを設け、よりハイスループット且つ低コストなDNAシーケンサーの開発を後押しした。このグラントで、2004年から2010年までで、合計60の技術開発プログラムに総額104.3百万ドルが授与されている。この研究成果は、現在販売されているスイスロシュ社のGSシリーズや、米国アプライドバイオシステムズ社のSOLiDシリーズに結実している。現在、これらに加えて米国イルミナ社のGenome Analyzerシリーズ・HiSeqシリーズが大規模ゲノム解析用途での主流となっている。これらはキャピラリーシーケンサーの100倍から10万倍のスループットを誇り、「次世代シーケンサー」という総称が与えられている。上述のAdvanced Sequencing Technology Awardsでは、ヒトゲノムを1000ドルで解読できる技術の開発を目標としており、2009年には10万ドルで解読できるところまでが達成された。これらの世界的な普及状況に関しては、2007年以降に導入がはじまり、2009年には世界の4大センターが各数十台規模を導入している。新たな展開として、中国BGI(Beijin Genome Institute)次が2010年に120台規模に拡大することを発表し、世界トップセンターに躍り出るようになった。

DNAシーケンサーの高性能化は現在進行中であり、米国パシフィックバイオサイエンス社から新製品が登場し、研究現場が導入を開始している。スループットの向上の他、これまでの次世代シーケンサーと比較して連続して長い配列(ロングリード)を決定できることが特徴となっている。

更に次世代を見据えた研究開発も精力的に行われており、Nanopore、Nanofluidics、

原子間力顕微鏡、透過型電子顕微鏡、質量分析などを原理としたものが報告されている。これらは、更なるハイスループット化、高速化、低コスト化、ロングリード化、サンプル調整の容易化を実現するものとして期待されている。スループットに関しては、この中でも有力視される Nanopore において、現行のロシュ社 454 シリーズの約 1 万倍のスループットが見込めるとの試算がある。

次世代シーケンサーの普及により、世界的な傾向として DNA データが爆発的に増えることは確実である。現状においてさえ、記憶容量・通信能力が不足していること、データ解析技術が未成熟であることが問題となっている。将来的には、計算資源の拡張コストが限界に達すると考えられている。データ解析に関しても、DNA シーケンサーの性能向上によって、エラー率が高い、リード長が短いといった問題は軽減されつつあるが、大量のデータに対して現実的な計算時間でエラー除去、アセンブリ、多型同定等を完了させるための技術は確立されていない状況にある。

次世代 DNA シーケンサーは、バイオ関連産業に以下のインパクトを与えると考えられる。

- 医療に与えるインパクト

Advanced Sequencing Technology Awards では、一人のゲノム配列決定にかかる費用を 1000 ドルから 100 ドルに、その決定にかかる時間を 1 週間から 1 時間に引き下げることが目標としている。この目標は、医療現場での検査に DNA シーケンサーを用いるために設定されたものであり、近い将来達成される見込みであることから、医療現場での利用がそう遠くないことであると考えられる。ゲノムデータを検査に利用することを考えると、データ解析に絶対的な信頼性が求められることになる。多型と表現型との関係性を調べるに当たっては、大規模サンプルからデータ収集を行うことや、解析アルゴリズムの精査はもちろん、人的資源の確保・教育、情報インフラ整備が必須となると考えられる。また、検査の場においては、医療従事者が被検査者に対してデータ解析結果を説明する必要があるため、そのためのソフトウェアが必要となるのはもちろん、医療従事者にはその専門性も求められることとなる。医療現場には電子カルテが浸透しつつあり、ゲノム関係のデータ解析結果をそれと統合することによって、さらなる付加価値を埋める可能性がある。データベースが更に大規模化することによって、処理の高速化はもちろんのこと、データ圧縮やセキュリティに対してもこれまで以上の技術レベルが求められることとなる。

- 農業、水産業、環境産業に与えるインパクト

次世代シーケンサーによって、メタゲノムに対する取り組みが本格化し、土壌や海水中の微生物によってその状態を特徴づけようとする研究が行われている。メタゲノムによって環境モニタリングを行い、その場所の利用可能性の評価や、

改善のための指針提供、収穫量の予測などに発展する可能性がある。これらのサービスが、メタゲノムという統一的な手法で、低価格に提供できるようになれば、農業等に与えるインパクトは大きい。

- DNA マイクロアレイの代替としてのシーケンサー

これまで、発現プロファイル解析や網羅的エクソン探索には DNA マイクロアレイが使用されてきたが、次世代シーケンサーは未知の転写産物を検出できることから、この用途に使われるようになった。次世代シーケンサーの利用コストが下がっており、網羅的定量・探索には次世代シーケンサーが選択されることが多くなっている。DNA マイクロアレイの場合には、実験して得られたデータとその解析結果とをサービス提供者のサーバに保管し、ユーザには WEB ページから閲覧させるサービスが存在している。次世代シーケンサーの場合、計測して得られるデータがマイクロアレイに対してあまりにも膨大であり、データ保管コストが高く付きすぎるためにそのようなサービスは成立し辛くなっている。

- バイオインフォマティクス産業の拡大可能性

次世代シーケンサーによって生成されるデータは従来と比較して遥かに膨大であり、現状においては専用の計算資源および人的資源が必要とされている。計算資源に関しては低コスト化のために PC クラスタが選択されることが多くなったが、用途によって必要とされるメモリ量・ハードディスク量が大きく異なること等、この用途における PC クラスタの設計にはバイオインフォマティクスと次世代シーケンサーとに関する専門知識が求められる。人的資源に関しては、基本的なデータ処理に専従者が必要であるとされており、その専従者にはバイオインフォマティクスに関する専門知識に加え、一般的なシステム管理の専門知識はもちろん、並列計算環境を有効活用するための専門知識が求められる。このような背景から、バイオインフォマティクス産業においては、コンサルティングを含めたシステムインテグレーション事業が拡大すると考えられる。また、新型シーケンサーに特化したツール開発・販売が行われるようになっている。次世代シーケンサーが生成したリードデータをマッピングするツールとしては、ノボアライン社の Novoalign が高い性能を誇っている。新型シーケンサー用の統合解析環境や解析結果可視化ツールも販売されている。現状、これらは生命科学に従事する専門家のためのものであるが、上述した通り医療従事者等がこれらのデータに接するようになることが有り得るため、そのインターフェースとなるソフトウェア・ハードウェアの開発・販売に新しい事業機会が生まれる可能性がある。

3章 今後の課題とその解決のための提案

1節 基礎研究に係る課題

バイオインフォマティクス関連基礎領域における明らかな研究トレンドとして、研究対象がより高次の生命現象にシフトしていることと、計測データがより大規模に生成され、様々なタイプのデータを組み合わせて解析する方向へと向かっている。バイオ研究の具体的な研究対象としては、iPS細胞やES細胞など、医学的にも重要な対象に関心が高まっており、遺伝子レベルの改変・導入と、細胞レベルの形態・動態との関係性が正に興味の対象となっている。これらを踏まえ、本章では、以下の項目について提案を行う。

- ・高次生命現象理解のための研究開発
- ・多種多様なデータを有効活用するための研究開発
- ・大量データ処理のための研究開発

1) 高次生命現象理解のための研究開発

BIRD事業の研究開発課題には高次生命現象の理解を研究スコープに置いたものが多く、世界的な研究トレンドをタイムリーに捉えていた。しかしながら、近年になって、欧米のシステム生物学に対する取り組みは、幾つものコンソーシアムの存在からも特に力が入っていることが伺える。欧米に追従することが必ずしも正しいことではないが、世界的な動向を踏まえて日本としての戦略を策定することが望まれる。高次生命現象の数理モデル化やシミュレーションが注目されがちであるが、ノイズ除去や正規化、キャリブレーションといった生データに対する処理や、高次元入力データに適用可能な機械学習手法の開発など、基本的な部分においても不足している要素は多い。前者には、実験生物学に即したデータ解析技術が必要である。以前から、実験生物学のエキスパートと対等に議論できる情報系研究者の不足が叫ばれており、その教育・人材確保が望まれる。後者に関しては、純粋な情報科学的アプローチの改良だけでなく、生命現象を再現することを目的としたヒューリスティクスが解決策のオプションである。バイオインフォマティクスの専門家を育成することはもちろん、統計や大規模計算の専門家との交流といった情報科学における領域横断的協力体制の構築が必要である。

2) 多種多様なデータを有効活用するための研究開発

ここ数年で生命科学に関する多種多様なデータベースが構築されており、そこに格納されているデータ量も非常に大規模になっている。それと並行して、研究トレンドが、より高次の生命現象を対象とするようになっており、様々な物質・現象のデータを統合的に解析することが必須となっている。つまり、如何にして多種多様な大規模なデータベースを活用できるかが研究のキーポイントとなっている。しかし

ながら、このためのデータハンドリングは容易なものではなく、これまでに多くの人的コストを消費してきた。この対策として、セマンティック・ウェブのための技術として確立されたオントロジー・RDF が生命科学のデータベース構築に取り込まれるようになった。これらによって、データの体系的分類や関係性の記述、自動的な相互リンクが可能となり、上述したコストを削減することが可能となっている。BIRD 事業の研究開発課題である「2 項関係に基づくゲノムと生命システムの機能解読」による KEGG データベースの構築や、「インタラクティブ解析からの生物知識獲得」がオントロジー・RDF の整備を行ってきており、研究コミュニティの全体的な研究コストを削減することに貢献している。これらに格納されているデータは独自性も高く、世界と比較しても競争力のある研究基盤にまで至っている。日本の研究レベルを維持するためには、バイオサイエンスデータベースセンターがこれらの運用と更なる高度化のための研究開発とを安定的にサポートできることが大切であると考えられる。

3) 大量データ処理のための研究開発

次世代シーケンサーの登場により、取り扱わなければならないデータ量が指数関数的に増加しており、今後ともその増加傾向は変わらないと言われている。データの増加速度が計算機の性能向上の速度を上回っており、単純に計算を並列するだけでは対処不可能になっている。また、次世代シーケンサーに特有のノイズや、現状では連続して読める塩基配列長が短いことから、従来のアセンブリングアルゴリズムが適用できないのが問題となった。欧米のバイオインフォマティクス研究者は、次世代シーケンサーの登場に前後してこの問題に取り組み始め、このデータ解析に必要となるソフトウェアを研究コミュニティに提供した。その結果、次世代シーケンサーの解析に利用されるソフトウェアのほとんどが、欧米の研究機関が提供するものとなっている。最新の研究トレンドに、バイオインフォマティクスの立場から貢献するには、次世代において計測装置に必要とされるソフトウェア開発に如何に早く取り組むかが重要である。そのためには、バイオインフォマティクス研究者と、計測装置開発者との交流はもちろん、海外の研究動向に対する網羅的な調査と研究者へのフィードバックが不可欠である。

2節 産業化に係る課題

バイオインフォマティクスの産業化は徐々に進展しているが、創薬、診断、環境、食品等の分野における具体的なニーズへの対応、それを担う人材の育成、複数の関連情報のリンクとその情報解析等は、今後の課題と考えられる。

これらを踏まえ、本節ではバイオインフォマティクスの有望産業分野である創薬、診断に係わる課題対応として、以下の項目について提案を行う。

- ・ゲノム情報と臨床情報の連動
- ・大量の解析データに対応するソフトや人的能力拡大の必要性
- ・創薬に係るニーズへの対応
- ・環境、食品分野に係るデータベース・インフォマティクス技術の充実

1) 遺伝情報と個人情報の結びつけによる個別化医療等の実現

特に個別化医療を実現するにあたっては、個人の遺伝情報と臨床中心の医療情報を連動させ、解析することが求められている。臨床情報中心の医療情報については、電子カルテが整備されつつあり、個人別データともに、統計的な解析情報の活用も可能になりつつある。個人のゲノム情報については、次世代 DNA シーケンサーによる解析能力拡大と解析コストの低減が可能になり、5年以内に10万円で個人全ゲノムデータが解析でき、10年以内には1万円・1時間を切る可能性が高まっている。また、特定の疾病に関する検体とそれに付帯する情報を蓄積するバイオバンクの整備が進められつつある。

上記により、特定の検体に係る情報解析を行い、遺伝情報と個人情報を結びつけることが可能になりつつあるが、以下のような課題が存在する。

(「ヒトゲノム・遺伝子解析研究に関する倫理指針の見直しに関する専門委員会」

<http://www.mhlw.go.jp/stf/shingi/2r98520000008f2q.html#shingi29> 参照)

- ・検体の由来を明らかにして個人の診療情報に生かす連結可能匿名化のニーズがあるが、制度や倫理面からの検討が必要
- ・遺伝子検査／診断を医療・産業として推進するためには、遺伝子差別や遺伝情報保護に対する法制度整備等が必要だが、まだ不十分
- ・必要な情報解析を企業に外注することで産業化が進むが、それが妥当であるかどうか
- ・電子カルテと遺伝情報を連携するフォーマットや用語の統一化が必要

以上の課題は最後の点を除けばバイオインフォマティクス自体に由来する課題とはいえないが、産業化や市場拡大を実現するには避けられない課題といえ、先進的な国の事例検討や専門家による議論の深化、結果としての法制度やガイドライン整備が必要と考えられる。

2) 大量の解析データに対応するソフトや人材育成の必要性

次世代 DNA シーケンサーにより大量の DNA 塩基配列データが得られても、その機能解析を行い、さらに臨床応用や創薬に生かすことが求められる。そのためには、新たな解析ソフトの開発とともに、関連人材の育成やスキルの向上が必要になる。ソフトについては、研究用や汎用ソフトは充実しているが、個別の産業や事業に直接生かせるソフトはまだ少ない。そのためには、有償であっても具体的なニーズに対応したソフトの作成、事業化が必要と考えられる。また、人材についても基礎研究的なバイオインフォマティクス人材は増加しつつあるが、例えば医療分野において臨床面の情報解析も含めた人材や創薬を行えるバイオインフォマティクス人材はまだ少ないといえ、その育成が必要である。

3) 創薬に係るニーズへの対応

製薬企業の研究者には、医薬品開発のデータベースや情報源、バイオインフォマティクスに対して、以下のようなニーズがある。

- ・臨床検体の生データ、さらに診療情報や動物データとの関係付け
- ・個人別タイピングデータ
- ・疾患のサブタイプ同定、医薬品の副作用と関係付けられた遺伝子タイプの集団データ
- ・特定の薬剤を投与された集団（コホート含む）の SNP や発現プロファイリング情報
- ・開発する化合物の代謝経路の遺伝子の多型に関する情報
- ・創薬に直結するデータベースは難しいが、ゲノムと疾患を結び付ける情報
- ・（疾患の本態解明の）DB としては、①個人別タイピングデータが見れること、②データの clean up がされていること、③データの品質が確認できること、が重要。
- ・国家プロジェクトで実施したデータにアクセスしたいという要望はあり、その際は、raw データ、個別のデータにアクセスしたい。臨床情報のついていないデータを公開しても何も意味がない。

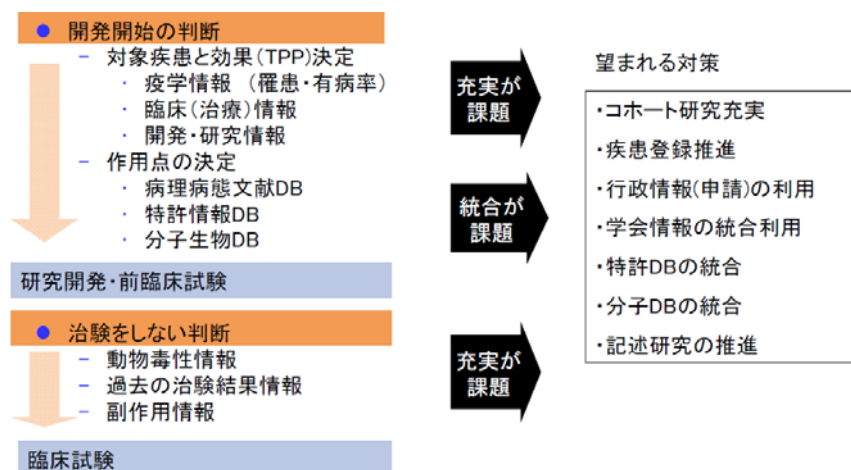
その他、図表 V-1 に示す創薬過程と利用情報などから、以下の点も求められる。

- ・コホート研究の充実と連携、既存コホート情報の活用
- ・特許や論文情報も含めたデータベースの統合
- ・化合物からの探索と、標的となる蛋白質からの検索が一体化したデータベース
- ・Chemical Genomics のデータベース（NIH は実施）
- ・増加する抗体医薬品に係るデータベースやインフォマティクス技術
- ・前臨床試験段階で開発が中止された薬に関する毒性試験等のデータの蓄積
- ・PMDA（独立行政法人医薬品医療機器総合機構）に蓄積されているデータの活用（画像情報でなくテキスト・データとしての検索性、副作用情報のデータベース）

以上について、バイオインフォマティクスにおいて以下のような検討が必要と考えられる。

- ・関連データベースの統合
- ・遺伝情報、疾患、個人別情報（多型等）の関係付け
- ・個別情報、実験の原データへのアクセス

図表 V-1 創薬過程と利用情報



出典：「生命科学データベース統合に関する調査研究」第5回ライフサイエンスPT（4/19/07）資料

4) 環境分野、食品分野に係わるデータベースとインフォマティクス技術の充実

バイオインフォマティクスの活用は、創薬や医療分野中心に進んでおり、環境分野や食料分野では研究は行われていても、産業化を想定して利用できるデータベースやインフォマティクス技術は少なく、その充実が望まれる。

具体的な産業化イメージとしては、環境分野では土壌等の有害物質を分解除去する微生物の機能解明と、それを活用したより高い効果を挙げられる微生物の育種が挙げられる。実際に有害物質分解菌の有効性を検証するマイクロアレイの製品化例がある。

食品分野では、Nutrition（栄養）+Genomics（遺伝学）を融合させたニュートリゲノミクスが注目される。ニュートリゲノミクスは、食品に対する生体応答を遺伝子発現変動として捉え、網羅的に解析することで、様々な機能性食品成分の作用メカニズムを探ったり、生体内タンパク質・代謝物などを解析して、食品が体に与える影響を研究する。その研究ではDNAマイクロアレイとバイオインフォマティクス技術を利用して、網羅的解析を行うといった手法がとられる。今後は多様な機能を有する機能性食品の開発やその効果の検証等に広く利用されると考えられる。

以上に示したように、産業化においては、個々の産業分野に対応したバイオインフォマティクスの技術開発や製品化、そのための人材育成、用途に応じた統合的なデータベースの作成等が、今後の重要な課題といえる。

VI 資料編

1章 バイオインフォマティクス委員会 委員名簿

図表 VI-1 バイオインフォマティクス委員会 委員名簿

役職	氏名	所属・役職名※	就任期間
委員長	勝木 元也	自然科学研究機構基礎生物学研究所 所長	平成 13～23 年度
委員	榊 佳之	東京大学医科学研究所 教授	平成 13～19 年度
	堀田 凱樹	国立遺伝学研究所 所長	平成 13～18 年度
	松原 謙一	奈良先端科学技術大学院大学バイオサイエンス研究科 教授	平成 13～19 年度
	吉田 光昭	万有製薬株式会社つくば研究所 所長	平成 13～23 年度
	秋山 泰	東京工業大学大学院情報理工学研究科 教授	平成 19 年度 (17～18 年度は分科会委員)
	田畑 哲之	(財)かざさ DNA 研究所 副所長	平成 19～23 年度
	藤山 秋佐夫	情報・システム研究機構国立情報学研究所 教授	平成 19～23 年度
	深海 薫	理化学研究所バイオリソースセンター情報解析技術室 室長	平成 20～23 年度 (19 年度は分科会委員)
	松田 秀雄	大阪大学大学院基礎工学研究科 助教授 → 情報科学研究科 教授	平成 20～23 年度 (13, 16～19 年度は分科会委員)
	分科会委員	高木 利久	東京大学医科学研究所 教授
松田 秀雄		大阪大学大学院基礎工学研究科 助教授	平成 13, 16～19 年度 (平成 20～23 年度は委員)
久原 哲		九州大学大学院農学研究院 教授	平成 16～18 年度
藤 博幸		京都大学化学研究所 教授	平成 16～23 年度
秋山 泰		産業技術総合研究所生命情報科学研究センター センター長	平成 17～18 年度 (平成 19 年度は委員)
阿久津 達也		京都大学化学研究所 教授	平成 18, 21～23 年度
宮野 悟		東京大学医科学研究所 教授	平成 18～19 年度
深海 薫	理化学研究所バイオリソースセンター情報解析技術室 室長	平成 19 年度 (平成 20～23 年度は委員)	

※ 所属機関・役職名は、就任当時のもの。

図表 VI-2 バイオインフォマティクス委員会 委員の各年度の役割

No		H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23
1	勝木元也	委員長(※1)★(※4)	委員長	委員長	委員長	委員長	委員長★	委員長★	委員長	委員長	委員長	委員長
2	榊佳之	委員(※2)★	委員	委員	委員	委員	委員★	委員				
3	堀田凱樹	委員★	委員	委員	委員	委員	委員★					
4	松原謙一	委員★	委員	委員	委員	委員	委員★	委員				
5	吉田光昭	委員★	委員	委員	委員	委員	委員★	委員	委員	委員	委員	委員
6	秋山泰					分科会(※3)	分科会★	委員★				
7	田畑哲之							委員★	委員	委員	委員	委員
8	深海薫							分科会★	委員	委員	委員	委員
9	藤山秋佐夫							委員★	委員	委員	委員	委員
10	松田秀雄	分科会			分科会	分科会	分科会★	分科会	委員	委員	委員	委員
11	阿久津達也						分科会			分科会	分科会	分科会
12	久原哲				分科会	分科会	分科会					
13	高木利久	分科会			分科会	分科会	分科会	分科会★				
14	藤博幸				分科会	分科会	分科会	分科会★	分科会	分科会	分科会	分科会
15	宮野悟						分科会	分科会				
	課題採択	高度化 I 期(※5) H13 新規(※6)			H16 継続(※7)	H17 新規(※6)	高度化 II 期(※5) H18 新規(※6)	H19 新規(※6) H20 継続(※7)	H21 継続(※7)			

※1 委員長：バイオインフォマティクス委員会委員長

※2 委員：バイオインフォマティクス委員会委員

※3 分科会：バイオインフォマティクス委員会分科会委員、事前評価、事後評価等の課題採択や評価を担当

※4 各年度の委員、分科会委員が評価に携わるが、限定メンバーとなっているところは★を付した

※5 高度化 I 期、II 期は、「生命情報データベースの高度化・標準化」の第 I 期、II 期の課題

※6 H13 新規、H17 新規、H18 新規、H19 新規は、H13、H17、H18、H19 年度に新規採択された「創造的な生物・情報知識融合型の研究開発」の課題

※7 H16 継続、H20 継続、H21 継続は、H13、H17、H18 年度の新規採択課題の中で、それぞれ H16、H20、H21 年度よりさらに 2 年間継続される課題

2章 研究開発課題の成果

1節 生命情報データベースの高度化・標準化

図表 VI-3 及び 図表 VI-4 に第I期及び第II期の研究開発課題の成果を示す。

図表 VI-3 第I期研究開発課題（研究開発期間：平成13～17年度）の成果

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た達成度
金久 實 (京都大学 化学研究所 教授)	2項関係に基づく ゲノムと生命シ ステムの機能解 読	2項関係データベース BRITE として、さまざまなタイプの2項関係をデータベース化する。具体的には、生物学的意味解釈に利用するため、分子レベル、細胞レベル、個体レベルでの機能を表現する語彙の体系化（オントロジー）転写制御関係データベース、遺伝子疾患関係データベースの構築を行う。またデータベース用アノテーションシステム、テキスト処理システム、利用システムの開発を行う。	研究開発計画の個々の項目の実施という観点から見ればいくつか変更があったことから、当初計画に記載したことがすべて実現されているわけではない。しかし、BRITE の位置づけとしてオントロジーのような機能階層に焦点を絞ったこと、つまり BRITE に KEGG を取り入れるのではなく、BRITE を KEGG の一部としたことで、本研究開発の目的が明確にされた。KEGG に GO のような語彙の階層表現を取り入れ、KEGG BRITE として完成させたことから、大きな枠組みの中で当初計画は充分達成されたと評価できる。
菅原 秀明 (国立遺伝 学研究所 教授)	ゲノム生物学バ ックボーンデー タベースの構築 提供	DDB ならびに研究コミュニティによる DNA 配列データに対する評価とアノテーション付加を実現するシステムと全ゲノム蛋白質立体構造予測データベース構築による塩基配列データの高品位化を行うとともに、時間軸を含む三次元可視化を実現するシステム開発を行い、重度の高い情報探索システムとイメージデータを含む大容量の結果を表示する。データとしては、DNA 配列やアミノ酸配列はもちろんのこと、マイクロアレイに代表される各種遺伝子発現データを収集し、各種生物のイメージデータの上で遺伝子の発現パターンを表示するシステムを開発する。	高品位データベースと四次元データベースという概念を、OASYS、GTOP、CIBEX および BSD として具体化した。OASYS は、DDBJ から入手可能なゲノム配列データのアノテーションを再評価することを目的としていたが、新規のゲノム配列に対して比較ゲノム解析を行なってアノテーションを付与する G-InforBIO の開発へと広がった。また、公開されている微生物ゲノム由来の ORF の再評価を行い新規 ORF 候補を発見した。GTOP 解析によって、ORF の信頼性評価に加えて ORF の構造やタンパク質の立体構造について新たな知見を得ることができた。MADB は CIBEX として DDBJ 固有の事業から国際公共データベース事業に中に位置づけられた。さらに、4 データベースに KEGG と PDB を加えた情報資源の統合検索システムが構築された。本研究開発課題で利用した XML 技術と Web サービス技術は、本研究開発課題と DDBJ のデータサービスの間の相乗効果をもたらした。

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た達成度
<p>高木 利久 (東京大学 大学院新領 域創成科学 研究科 教 授)</p>	<p>シグナルオン トロジーと バイオタ ームバン クの開発</p>	<p>ゲノムにコードされた生命のメカニズムを解読するには、これまで蓄積されてきた膨大であるがバラバラな生物知識を整理統合し、計算機が理解し解析できるような形に整備すること、すなわち、生物知識の枠組みの体系化(オントロジー、辞書作成)とデータベース化が必要である。本研究では、このような動機のもとに、生命科学の中でもその要であるシグナル伝達系を対象としてオントロジーとデータベースを構築する。また、生命科学の知識の全容をカバーするために必要となる網羅的な用語辞書の構築もあわせて行う。</p>	<p>数年先の研究動向や需要をにらんだ、先進的かつ、高度で、標準化を目指したデータベースを開発する、という挑戦的な計画に基づいた研究開発である。その結果、既存のデータベースにはない高付加価値のデータを新たに電子化したデータベースが構築された。現在の公開までに変遷を要したのは、枠組み構築は研究的側面が非常に強く、また人材育成をしながらの研究開発であったことから考えれば、相応であったと考えられる。</p>
<p>中村 春木 (大阪大学 蛋白質研 究所 教授)</p>	<p>蛋白質立体 構造デー タベース の高度化</p>	<p>PDB へのデータ登録作業に積極的に協力して蛋白質の立体構造とゲノム情報との結びつきを強める一方、XMLなどの最新情報技術を利用し、国際的な連携のもとに世界標準としての新しいデータ記述(PDBML)と解析ツールや二次データベースを開発して付加価値を付け加え、構造生物学者だけを対象とする専門的なデータベースから、広く生命科学の研究者、産業界、さらには一般の人にも役立つデータベースに高度化することを目的とする。</p>	<p>国内では、「日本蛋白質構造データバンク (PDBj:Protein Data Bank japan)」を設立し、複数の国内の大学の研究室をサテライト・サイトとする研究開発体制を構築する一方、国際的には、米国 RCSB や欧州 EBI と共同で「国際蛋白質構造データバンク (worldwide Protein Data Bank:wwPDB)」を設立して国際的な連携を強化させ、研究開発を推進した。</p> <p>当初計画の、1)新規蛋白質立体構造データベース構築、2)蛋白質構造を基にした解析システムの開発と二次データベースの構築、3) PDB データベース業務の日本の分担作業の実施、4) BioMagResBank (BMRB) の日本の分担作業の実施、5) 日本国内での教育用データベースの作成と公開、はどの項目も実施され、その成果が Web を通して公開されている。それぞれの内容が充実しており、想定以上の成果が得られ、望ましい展開となった。</p>

図表 VI-4 第 II 期研究開発課題（研究開発期間：平成 18～22 年度）の成果

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
稲葉 一男 (筑波大学下田臨海実験センター 教授)	ホヤプロテイン統合データベースの構築	ホヤにおけるポストゲノム研究の成果を集結したホヤプロテインデータベース及び新たな統合データベースの構築を行い、機能未知遺伝子/タンパク質の機能解明や、疾患遺伝子の機構解明への貢献を目指す。	配列データ、cDNA データ、電気泳動画像データ、タンパク質発現データなどの各種データが収集、整理され、それらが CIPRO を中心とするデータベース群にまとめられ、公開、もしくは、公開予定となっており、計画通りに進んだ。	ホヤゲノムデータベースの統合化を行わなかったが、現在進行中の研究の多様性を考慮した点ではむしろ望ましい判断だったといえる。また、ホヤ研究者コミュニティの国際標準データベースとなったことは、想定外の望ましい展開である。
金久 實 (京都大学化学研究所 教授)	ゲノムと環境の統合解析による生命システムの機能解読 (KEGG)	生命システム情報統合データベース KEGG に、新たに生体外物質との関連情報を蓄積し、生命システムの機能を生体内だけでなく、環境との相互作用も含めて理解することを目指す。	KEGG の高度化が行われており、基盤的な部分は計画通り達成されている。応用部分については、高い達成度を示すものや、フレームワークの構築はできている一方、エントリ数が十分ではないものもあり、今後の充実が望まれる。	メタゲノム解析の進歩、がん化をはじめ代謝系の重要性に対する再評価、いずれも代謝パスウェイマップを基盤とする KEGG の重要性につながる展開となった。研究コミュニティからの要望に応える形で、当初計画を修正して開発を進めている。
菅原 秀明 (国立遺伝学研究所 教授)	バイオ基幹情報資源の高準化と共用化	日本 DNA データバンク DDBJ を中心に、バイオ基幹情報資源の高準化・共用化や、バイオロジー研究開発における方法・手段の表現方法及び情報技術系方法論の規格化を目指す。	メソッド・オントロジーなどやや遅れ気味の計画もあったが、Web API の WABI を拡充し、関連するツール群も開発しており、計画通りに進んだと評価できる。	新世代高速シーケンサーから産出される膨大なデータへ対応した。特にそのメタデータ作成ツールである MetaDefine は NCBI および EBI からも高く評価され、ソースコード提供を行うなど、望ましい展開となった。
高木 利久 (東京大学大学院新領域創成科学研究科 教授)	オントロジーによるパスウェイの高度化および国際標準化 (INOH パスウェイデータベース)	シグナル伝達パスウェイデータベース INOH をより高次の生命現象まで扱えるように発展させる。さらに高い精度のパスウェイデータの抽出、オントロジーの拡充に取り組み、これら成果の国際標準への採用を目指す。	シグナル伝達についての複雑な知識を、推論可能な形で表現する方法を提案し、国際的なパスウェイデータの標準化に貢献した。	国際標準 BioPAX で表現できないデータについて、新たな表現方法を開発し、それが BioPAX にフィードバックされて国際標準に取り込まれた。データベース自体の充実は図れたが、それをういた推論に関しては十分に開発が進展していない。

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
中村 春木 (大阪大学蛋白質研究所教授)	蛋白質構造データバンクの国際的な構築と高度化 (PDBj)	蛋白質立体構造データベースPDBのデータ登録・編集技術の改良を行い、品質精査を伴う登録業務作業の更なる合理化・自動化を目指す。また、NMRデータ・電子顕微鏡画像等の新たなデータベースを構築する。	第一の使命であるPDBの国際分担機能を十分に果たすとともに、タンパク質配列および構造解析のための各種ツールを開発し、さらに、実験系研究者らと共同で生物学的に重要な知見の発見も行っており、高く評価できる。	PDBj Mine や ProMode-Elastic など当初の計画になかったシステムも開発された。また、開発したツールを用いた共同研究によりレベルの高い成果を得たことは望ましいと考えられる。
西岡 孝明 (慶應義塾大学大学院教授)	メタボロームMSスペクトル統合データベースの開発 (Metabolome-Mass Spectral Database)	二次代謝物質の質量、化学構造式、生物種等の収集データベースと、生体成分を各種質量分析計によって測定したMSスペクトルデータベースを統合し、ピーク同定に必要な化学構造(MS) _n 関連知識データベースの開発に取り組む。	MS スペクトルの分散データベースである MassBank を開発し、同時に関連データベースの開発、充実を図っており、計画通りに進んだと評価できる。一方、リポジトリ形式であることから提供されるデータの質の問題が顕在化してきており、今後対応することが検討されている。	MassBank が日本質量分析学会の公式データベースとして認められたことは望ましい展開として評価できる。
森下 真一 (東京大学大学院新領域創成科学研究科教授)	マルチモーダル統合バイオDB (Multimodal BIODB)	出芽酵母、メダカ、ヒトを主としたマルチモーダル統合バイオデータベースの開発に取り組み、蓄積データを有効に統合化出来るシステムの実現を目指す。	超高速シーケンサーの出現・普及に合わせ計画を柔軟に変更し、迅速に研究開発を行い、レベルの高い成果を出したことは特筆に値する。	超高速シーケンサーの急速な普及に対応し、大規模並列計算手法を確立するとともに、クロマチン構造に関して重要な生物学的成果を得るなど非常に望ましい展開となった。

2節 創造的な生物・情報知識融合型の研究開発

図表 VI-5～図表 VI-11 に平成 13～21 年度開始の研究開発課題の成果を示す。

図表 VI-5 平成 13 年度採択 新規研究開発課題（研究開発期間：平成 13 年 10 月～16 年 9 月）の成果

代表研究者	研究開発課題名	研究開発計画	研究開発成果
伊藤 啓 (東京大学分子細胞生物学研究所 助教授)	ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築	脳全体の回路図を作成し、計算機上での動作再現することを目的として、ショウジョウバエ脳神経回路の解析結果をデータベース化することにより、脳回路図を作成し、感覚情報のコード化、異種の信号の統合過程等について計算機シミュレーション及び実験的検証を行う。	ショウジョウバエ脳全体の回路図を作成してコンピューター上での動作再現を行うことを究極の目標とし、視覚、嗅覚、聴覚の 3 つの感覚経路について、感覚細胞から順にステップを追って神経回路の網羅的同一化を行った。この過程で現実の神経回路を定量的にコンピューター内にデータベース化するために、細胞体の位置情報を抽出するアルゴリズムおよび細胞体の個数を計測するアルゴリズムを開発し特許出願を行った。
伊藤 隆司 (東京大学大学院新領域創成科学研究科教授)	インタラクトーム解析からの生物知識獲得	蛋白質-蛋白質、蛋白質-核酸及び蛋白質-脂質相互作用のカタログ化から機能解析にわたる系統的解析を行う。これと連動して「既存知識に基づく解釈支援システム」及び「生物知識発見支援ツール」を開発し、相互作用データからの生物知識獲得に必要な技術の確立を目指す。	インタラクトーム解析を広義に捉えてその更なる発展を促す為に、様々な実験手法や計測技術の開発を情報科学技術と連携のもとに進め、それぞれユニークな基盤技術の開発が達成された。とりわけ遺伝子断片の絶対定量法に解析支援的なバイオインフォマティクス技術を組み合わせて構築したアダプタ付き付加競合 PCR 法 (GATC-PCR) システムは、遺伝子発現、転写因子-DNA 相互作用をはじめ様々な応用が可能なものであり、これまでに例のない種類のデータの創出を可能にした。
大浪 修一 (慶應義塾大学大学院理工学研究科 助教授)	線虫 <i>C. elegans</i> 発生過程のシステム解析	線虫の細胞系譜データを収集し、多細胞生物の線虫初期発生システムの計算機科学的解析等を目指す。このために、大規模突然変異体細胞系譜測定、細胞系譜データを用いた遺伝子機能ネットワーク推定アルゴリズムの開発、シミュレーション技術開発、パラメータ測定技術開発等を行う。	時間的空間的に動的な多細胞生物の発生機構を解明するために、遺伝子機能を阻害した <i>C. elegans</i> 初期胚を用いて細胞分裂パターンの実験的解析および計算科学的解析を並行して行った結果、機能破壊により細胞分裂のタイミングあるいは胚内の細胞の空間配置のいずれかのみに顕著な異常を生じる遺伝子を 10 種類以上検出することに成功した。この過程で大規模な遺伝子発現量データより遺伝子制御ネットワークを導出するアルゴリズムの開発に成功し特許出願を行った。

代表研究者	研究開発課題名	研究開発計画	研究開発成果
<p>岸野 洋久 (東京大学大学院農学生命科学研究科教授)</p>	<p>ゲノム進化とマッピングの階層モデルと解析アルゴリズムの開発</p>	<p>分子進化速度の変動、ウィルスの免疫適応過程、進化の過程での遺伝子の得失、蛋白質の進化等のモデルから、ゲノムに記録されている生物の多様化と適応進化の痕跡を検出する確率論的手法を構築するとともに、ゲノム上の重要遺伝子をマッピングするアルゴリズムを構築する。</p>	<p>一般に利用されているゲノムデータベースを用いて独自のモデルを構築し、ゲノムに映し出された生物多様化と適応のプロセスを検出するとともに、ゲノム上の重要遺伝子をマッピングする手法を構築することを目的とした。主な成果として広く多細胞生物における分子進化の速度の変化と分岐年代を推定するプログラムを開発し公開した。</p>
<p>肥後 順一 (東京薬科大学 生命科学部 教授)</p>	<p>高速計算機システムによる蛋白質フォールディングの研究</p>	<p>ペプチドの立体構造解析のための並列型専用計算機システムと並列化アルゴリズムを開発し、水分子存在下での蛋白質フォールディング計算を行う。また、分子会合、蛋白質機能等の計算アルゴリズムを用いた手法を確立する。</p>	<p>水分子存在下での蛋白質フォールディング計算に着目し、ペプチドの立体構造解析のための並列型専用計算機システムと並列化アルゴリズムの開発を行った。その主な成果として 30 残基長ペプチドのフォールディングシミュレーションが可能となり、また蛋白質とリガンドのドッキングシミュレーションにより分子同士が離れた状態でも会合を促す長距離力が働くことを示した。</p>
<p>森下 真一 (東京大学大学院新領域創成科学研究科教授)</p>	<p>遺伝子破壊株イメージ・マイニング</p>	<p>出芽酵母全遺伝子約 6000 種の遺伝子破壊株の細胞状態変化を顕微鏡により追跡した画像情報を収集し、画像から特徴的な細胞状態の変化を追跡・抽出するイメージ・マイニング技術の開発、および遺伝子情報との因果関係を推測するデータ・マイニング技術の開発を行う。</p>	<p>酵母の遺伝子破壊株画像解析ソフトウェアを開発し、4780 の非必須遺伝子破壊株中で約 2,000 株が野生株に比べ形態が異常に変異している事実をつきとめた。また新規の細胞周期チェックポイントを発見した。</p>
<p>矢田 哲士 (京都大学大学院情報学研究科 助教授)</p>	<p>ヒト遺伝子の転写・発現の多様性解明を目指した基盤データベースの開発</p>	<p>信頼性の高い独自のプログラム群を活用し、ヒト遺伝子の高品質カタログを作成、さらに、それらの転写や発現に関する in vitro/vivo/silico データをデータベース化する。これにより、選択的スプライシングやプロモータ強度の予測等の新しい研究展開を促す。</p>	<p>ヒトゲノム配列中に潜むタンパク質をコードする遺伝子を網羅的に探索することを目標とし、3つの異なった手法に基づいた遺伝子発見プログラムを開発、改良を進めた。また、これらプログラムを効果的にゲノムに適用するためのプロトコールも合わせて開発し、ゲノム全体に適用することで信頼度の高い遺伝子候補の抽出に成功した。</p>

図表 VI-6 平成 16 年度採択 継続研究開発課題（研究開発期間：平成 16 年 10 月～18 年 9 月）の成果

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
伊藤 啓 (東京大学分子細胞生物学研究所 助教授)	ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築	これまでの研究開発では、ショウジョウバエの脳を用いて視覚、聴覚、味覚情報処理機構の比較的下位のレベルの神経回路構造を、体系的に明らかにした。また神経細胞の位置を自動認識して定量座標データ化するソフトウェア手法を開発した。これを発展させて本課題では、感覚情報処理系と高次統合処理回路の間を結ぶ高次情報回路の構造の解析と、細胞の位置だけでなく回路の投射構造を自動解析するソフトウェア手法の開発をめざす。	脳神経回路の視覚、嗅覚、聴覚、味覚の 4 情報処理系経路を地道に解析し、脳内神経回路の総合的理解を高めることができた意義は大きい。ただし、計画に書かれているシミュレーションを行うにはいたらなかった。	脳神経細胞数が想定されたよりもはるかに少なかったことやグリア細胞の役割の大きさ、キノコ体の役割など、当初には予想されていなかったことが明らかにされた。これらは、現段階までの研究展開に大きな影響は与えなかったにしても、今後に向けて望ましい展開となることが期待される。
伊藤 隆司 (東京大学大学院新領域創成科学研究科 教授)	絶対定量オーミックスからの知識発見	これまでの研究開発により、出芽酵母の遺伝子発現絶対定量システムが構築された。そこで本課題では、このシステムを駆使してトランスクリプトームと転写因子-プロモータ相互作用の絶対定量計測を行う。得られた絶対値データから算出される比率と差分の双方を統合的に利用する複眼的解析法を考案するとともに、定量性を重みとして活用することで遺伝子クラスターの解釈や相互作用モチーフの発見などを高精度に行う方法論の確立を目指す。	蛋白質-核酸相互作用絶対定量解析では、GATC-PCR による定量的トランスクリプトーム解析、そのためのプライマー設計法の開発、酵母の転写開始点の網羅的同定など、途中で生じた問題点の克服も含め十分な達成度が得られた。しかし、蛋白質-脂質相互作用の定量解析における技術開発の点では、解析対象とする物質の性質上困難な点も多く十分に達成したとは言えなかった。知識発見手法や支援システムの開発ではインフォーマティクスとしての新規性は少ないものの、実験支援技術としては、実用的かつ有効な方法が開発されたといえる。	当初想定されていなかった展開として、ORF 単位とは無関係に多様な部位から、しかも DNA のセンス・アンチセンスの両鎖からの転写が予想外に多く見出され、転写の複雑性を見出したことがあげられる。このことから、トランスクリプトーム自身の考え方や解析法の再検討を迫られる結果となったが、科学的な価値は高い。本研究の大いなる成果と見ることも出来る。非コード RNA の問題への展開は興味深い。

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
森下 真一 (東京大学大学院新領域創成科学研究科教授)	遺伝子破壊株イメージ・マイニング	これまでの研究開発では、出芽酵母の約 3000 の非必須遺伝子破壊株について細胞の顕微鏡画像（イメージ）を収集し、形態パラメータを正答率 98%以上で定量化するイメージ処理技術の研究開発に成功した。また計算科学的な手法で出芽酵母の先端成長に関する遺伝子を推定して、生物学的実験により検証することにも成功した。本課題ではこれを発展させて、必須遺伝子破壊株の形態パラメータの定量化に取り組み、解析を行うことで出芽酵母の全遺伝子破壊株について、形態パラメータ測定による遺伝子機能の発見（マイニング）をめざすとともに、他のモデル生物の細胞解析への応用を検討する。	出芽酵母形態異常の網羅的解析を行った結果、表現型異常が遺伝子機能を推定する一つのモデルになり得ることを示すことが出来たという点で計画通りに進行したと言える。このように単純と思える測定を基本から行った点は重要である。	形態異常のパラメータや異常を示す遺伝子破壊株の数が想定以上に多かったが、それらが成果の質に反映され望ましい展開となった。パラメータが予想以上に収集されているが、正規分布という仮定のため、その解析という意味では十分とは言い難い部分がある。今回の成果が外部との共同研究に発展しており、その意味でも望ましい方向に進んだと言える。
矢田 哲士 (京都大学大学院情報学研究科 助教授)	ヒトゲノムにおける広義の遺伝子発見研究	これまでの研究開発では、ヒトタンパク質遺伝子の網羅的で信頼性の高い発見を行ってきた。また一早くプロセス型偽遺伝子の体系的な発見にも着手してきた。これらの研究開発では、独自に開発した信頼性の高いプログラムを積極的に活用することで、他に類をみない高精度の情報を提供してきた。本課題では、これまでの成果を更に深化させるとともに、RNA 遺伝子や DNA 重複型偽遺伝子の発見といった広義の遺伝子発見研究の基盤を構築する。	タンパク質遺伝子のアノテーションは大いに改良されたが、プロセス型偽遺伝子の探索法についてはさらなる改良が必要と思われる。また、RNA 遺伝子探索法については、現在問題となっている非コード遺伝子をカバーできるものとはなっていない。	バイオインフォマティクスの側面からは新たな展開が見えないが、生物学的側面からは従来の見方では不十分である点が示唆されている。

図表 VI-7 平成 17 年度採択 新規研究開発課題（研究開発期間：平成 17 年 10 月～20 年 9 月）の成果

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
内山 郁夫 (自然科学研究機構基礎生物学研究所助手)	大規模な比較ゲノム研究を展開するためのワークベンチの構築	急速な蓄積を続けるゲノム配列を最大限に活用した比較解析研究を展開するための比較ゲノムワークベンチの開発を行う。多数のゲノムの同時比較を可能とするとともに、近縁ゲノム間比較に重点を置いた新しい比較解析手法の開発等を目指す。	ゲノム比較を行うためのオーソログ分類の改良や補助ツールの開発が行われ、包括的な微生物ゲノム比較を可能とする微生物ゲノムデータベースが開発された。	計画後にさらに多くの生物種のゲノムが明らかにされたので、近縁ゲノム比較ツール CGAT が開発され、近縁種の比較も同時に行えるシステムの開発につながった。
太田 元規 (東京工業大学学術国際情報センター助教授)	タンパク質の構造・機能予測法の開発とヒトゲノム配列への適用	タンパク質のアミノ酸配列から立体構造および機能を推定する総合的なシステムを開発する。開発したシステムをヒトゲノム由来のタンパク質配列について適用し、その結果を集積した「ヒトゲノム構造・機能アノテーションデータベース」を構築、公開する。	タンパク質の構造予測システムを開発し、従来よりも高い確率で予測できることを確認した。また、これまで個別に実施された構造予測、分子機能に関する研究結果を集約し、構造予測パイプラインを核とした構造・機能アノテーションシステムとその結果を格納するデータベースが構築された。	研究開発を通して、新たに発見された事をツールの機能に追加し、改良した。
川戸 佳 (東京大学大学院総合文化研究科教授)	脳スライス中で可視化した神経シナプスの自動解析	共焦点レーザー顕微鏡の画像から、神経樹状突起の分岐、スパインの位置及び形態を特定し、自動的に高速解析できる、新規なアルゴリズムを持つソフトウェアの開発を行う。	神経シナプスの 3 次元自動解析プログラムが開発され、3 次元スパイン自動検出のツールとして実用可能な状態となった。プログラムを用いて、大量の実験データの解析が可能となり、神経シナプスの解析を大きく加速させた。	マウスの神経細胞での解析が行われ、ホルモンシグナル等神経栄養因子やストレスによるスパイン変動にスパイン内部や樹状突起に存在する多くのキナーゼ群が関係していることを明らかにした。また、実時間スパイン追跡により、スパインが短時間に形成・消滅を繰り返していることが発見された。
塩田 浩平 (京都大学大学院医学研究科教授)	ヒト胚の形態発生に関する三次元データベース	ヒトの発生と先天異常のゲノムワイドな解析を行うためのインフラとして、ヒトの形態形成過程の詳細な三次元画像データベースを構築する。	約 1200 例のヒト胚標本の撮像を行い、ヒト胚 3 次元形態データベースを構築した。構築されたデータベースは、京都大学のヒト胚標本コレクションに基づくもので、他で例を見ない新規性の高い内容となっている。	MR 顕微鏡の開発や EFIC イメージングの導入により、当初の想定よりも解像度の高い画像データが得られ、きわめて実際に近いヒト胚の 3 次元モデルが作製された。

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
白井 剛 (長浜バイオ大学バイオサイエンス学部教授)	実践による超分子複合体モデリングシステムの開発	古細菌 3R 系をターゲットとして、現在進行形の超分子複合体構造・機能解析実験と照らし合わせながら、実践的で一般的な超分子複合体モデリングシステムの実用化を目指す。	超分子複合体モデリングシステムを整備し、データベースサーベイトールからフィッティングツールまでを開発した。複数のタンパク質複合体の構造解析の実験結果との比較により、モデリングに成功したことから、モデリング手法の有効性が示された。	研究開発を通して、システムを構造変化や相互作用変化に十分に利用できるよう、改良する糸口を見つけた。
宮川 剛 (京都大学大学院医学研究科 助教授)	マウスを用いた脳機能表現型データベースの開発	各種遺伝子改変マウスについて、網羅的行動テストバッテリーを用いた表現型データ取得とその情報化のためのプロトコル標準化等を行い、バイオインフォマティクスを適用した脳機能表現型の体系的な比較・解析を目指す。	当初の計画の通り、データ収集とマウスの脳機能表現型データベースの構築を行い、90 系統以上の遺伝子改変マウスを解析した。	90 系統以上の遺伝子改変マウスの網羅的な行動解析により、疾患に関わる遺伝子について新しい発見がなされている。また、解析に使用されたマウスの脳切片も供給する体制を整備している。

図表 VI-8 平成 18 年度採択 新規研究開発課題（研究開発期間：平成 18 年 11 月～21 年 9 月）

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
石井 信 (奈良先端科学技術大学院大学情報科学研究科 教授)	予測技術を用いた生命システムの同定手法の開発	近年、生物学情報は膨大に蓄積されているが、個別の生物機能を発現するシステムの要素の実験による確認は多くの労力が必要である。そこで、この構成に関して不足する情報を提示することのできる情報科学的予測技術を開発する。生体分子データから生物機能の予測、あるいは情報抽出（コード）・解読（デコード）技術の開発を行い、実験への仮説提供、分子から機能までのプロセス同定、機能発現のためのミニマムモデルの提示を目指す。開発手法は、神経の極性形成、および多細胞生物の体節形成の系に適用することで妥当性を評価する。	脊椎動物発生過程における体節形成予測において、時計遺伝子振動モデルの構築、シミュレーションからの予測結果をさらに実際の発生系で確認するところまで進んだ。また、もう一つのテーマである神経細胞の極性予測については、実験データに基づくモデル構築とシミュレーション実験まで研究が進んだ。このようにテーマにした生命現象については高い成果が上がっており、ほぼ当初の計画通り研究が進行したと言える。	モデルと実験との融合的な手法による体節および極性におけるパターン形成機構の解明が進展した。一般的にどこまでこの戦略が適用できるかについては今後の研究が待たれる。
黒川 顕 (奈良先端科学技術大学院大学情報科学研究科 助教授)	メタゲノムオーソログ遺伝子統合解析システムの開発	環境中の細菌は、数百から数万種の集団（細菌叢）を形成し、細菌間相互作用だけでなく、環境と密接に関連しながら複雑なシステムを構成している。これら細菌叢全体をゲノム解析する「メタゲノム解析」により、環境の根幹を形成する細菌叢の生命システムを明らかにすることが可能となりつつある。本課題ではメタゲノム解析により得られた大量のメタゲノムデータから有用な知見を効率よく発見するためのインフラ構築を目的に、メタゲノムデータからの遺伝子配列の予測およびクラスタリングを行い、それらを機能別にデータベース化、可視化する統合解析システムの開発を目指す。	メタゲノム解析のための遺伝子発見ソフトウェアや可視化ツールなど、当初の研究開発計画で目標としていたデータベースや解析システムは全て開発されている。また、ヒトゲノム BodyMap や代謝マップなど、新規の可視化への試みがなされている。	当初計画では想定していなかった新型シーケンサーの登場により、大量の配列データが産出されるようになり、データ量の問題からヒトメタゲノムに絞った解析となった。このような大量のデータに対処すべく、計算機システムならびにアルゴリズムの改良が必要となったが、本研究はこのような環境変化に対応する方向に進んでおり、我が国を代表するメタゲノム情報解析システムを作りつつあるだけでなく、実験研究へのフィードバックも意識されている点でも高く評価できる。

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
<p>榎原 康文 (慶應義塾大学理工学部教授)</p>	<p>タンパク質化合物相互作用の網羅的予測手法とデータベースの開発</p>	<p>創薬の初期ステップであるリード化合物の探索において、計算機によるタンパク質-化合物間相互作用予測は有用な手法であり、立体構造結合エネルギーを計算するドッキング解析などの手法が中心的に研究されてきた。本課題ではより汎用性が高く、入手が容易であるアミノ酸配列データ及びマススペクトルデータを用いた、タンパク質-化合物相互作用予測手法の開発を行う。膨大なデータの中から注目する特性を識別して説明する特徴や規則を発見し、未知のデータに対して意義のある予測を行う統計学的学習手法の一つである SVM を適用して、網羅的な相互作用予測を行う。</p>	<p>当初計画から若干の計画変更はあったが、タンパク質化合物相互作用の高精度の網羅的予測手法を開発し、開発した予測法の実験による検証がなされたことから、ほぼ計画通りに進捗したと言える。また、新規リード化合物候補の発見という当初計画以上の成果を生み出したことも評価できる。</p>	<p>化合物のデータベースである PubChem の膨大なデータから迅速、高精度の絞り込みが可能になり、結果、新規リード化合物候補が得られ、その活性や効果を実験により検証できた事は望ましい展開である。</p>
<p>清水 謙多郎 (東京大学大学院農学生命科学研究科教授)</p>	<p>高精度タンパク質間相互作用予測システムの開発</p>	<p>生命システムの解明には原子レベルでの詳細なタンパク質間相互作用解析が重要である。しかし、生化学実験や構造決定などの実験的手法は多大な時間とコストを必要とし、また、既存のバイオインフォマティクスの手法は詳細な機能解析を行うのに十分な精度を得ることが困難である等、新たな手法の開発が必要である。本課題では相互作用部位予測、ドッキングシミュレーション、高精度複合体モデリング及びダイナミクス解析の3つのアプローチから、高速かつ高精度のタンパク質-タンパク質、タンパク質-低分子の相互作用予測・解析システムの開発を目指す。</p>	<p>機械学習法を応用したいくつもの予測システムが開発され、また、ドッキング予測についても高速で高精度のものが開発されるなど、研究は計画通り順調に進展し、有用な成果が得られたと言える。</p>	<p>配列データからのタンパク質間相互作用予測やタンパク質-糖鎖相互作用予測の研究が新たに行われ、また、タンパク質-リガンド複合体データベースの構築などの当初の予定にない新たな成果が得られた。しかしながら、本研究プロジェクトにおける情報系と実験系との共同研究については、まとまった成果があまり出ていないように思われる。</p>

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
<p>中井 謙太 (東京大学医科学研究所教授)</p>	<p>転写制御領域の構築原理解明</p>	<p>ゲノム塩基配列に書き込まれた遺伝情報の解釈、中でも転写制御情報の解釈は、バイオインフォマティクスのみならず、現代分子生物学の最大の課題の一つである。本課題ではゲノム比較に基づくプロモーターの進化モデル構築、組織特異的発現を示すプロモーター群の持つ共通構造のモデル化、培養細胞から得られる遺伝子活性データを用いた人工プロモーター解析を行う。これによって、遺伝子発現制御機序の構築原理の解明及びその分子進化的変遷の解明を目指し、得られた結果はデータベース化する。</p>	<p>転写制御領域の総合的な解析研究としての計画は十分に達成できていると考えられ、それに加えて副次的な研究成果も得られている。</p>	<p>近年のエピジェネティクス研究の進展により、その大規模データの入手が可能になるなどの外的要因があり、その解析により新たな知見が得られるなど、副産物的ではあるがいくつかの新たな萌芽的研究が生まれている。</p>

図表 VI-9 平成 19 年度採択 新規研究開発課題（研究開発期間：平成 19 年 10 月～22 年 9 月）の成果

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
<p>長野 希美 ((独)産業技術総合研究所生命情報工学研究センター主任研究員)</p>	<p>酵素反応分類に基づく酵素反応予測システムの開発</p>	<p>従来の酵素分類 EC 番号は、主に基質・産物の化学構造や補酵素などに基づいた分類であり、触媒機構において重要な蛋白質の配列情報、立体構造が考慮されていない。本課題では、EC 番号に代わる、より精密な階層的な酵素反応分類を網羅的に行う。これは詳細な酵素触媒機構、立体構造、リガンドの反応部位に基づいた酵素反応クラスの新しい階層分類である。また、分類を促進するために、機械学習の最新技術を駆使して、自動酵素反応予測システムを開発する。予測システムを用い機能未知酵素の酵素反応クラスを予測し、その結果を分類に反映させる。最終的には「酵素反応予測システム」の一般公開を目指す。</p>	<p>当初の計画でのメインテーマであった酵素反応予測法の開発を、着実に達成され、類似反応を担う酵素の予測における原子の重みづけなど、独自の工夫によって、優れた成果を挙げている。開発したシステムの早期公開が望まれる。</p>	<p>当初の計画にはなかった酵素の生物種分布を解析されたが、現時点では解析結果を新たな研究開発にどのように結び付けていくかがはっきりとは示されていない。また、重みづけによってクエリが精密化できる可能性を示した。</p>
<p>馳澤 盛一郎 (東京大学大学院新領域創成科学研究科教授)</p>	<p>進化型計算と自己組織化による適応的画像分類法の開発</p>	<p>生命科学研究用、医用画像解析システムの多くは、解析の対象や目的を絞り、分野固有のドメイン知識に基づいて開発されているため、汎用性が低く、多種多様な画像やニーズへの対応が困難である。本課題では、高い汎用性を持った画像自動分類技術を新たに開発する。撮影対象、分類目的、撮影法など条件が異なる多様な画像群に対して、進化型計算法と自己組織化写像法を用いて、(1) 最適な特徴量を創出し選択するアルゴリズム、(2) 画像の傾向や分布を見通しよく可視化するアルゴリズム、(3) 専門家によるアノテーションに基づいて画像自動分類を行う部分教師付学習アルゴリズムの確立を目指す。</p>	<p>開発した画像解析技術が実際のデータの解析に適用され、有用な生物学的知見が得られた。予期していなかった異分野への展開もあった。</p>	<p>本来の生物学ばかりでなく、工学的応用への展開があった。脳神経組織に関する時系列画像の解析という新たな展開があり、神経組織の応答特性などに関する新たな知見が得られた。また、要素技術となり得るような汎用性の高い分類方法の開発にも成功した。</p>

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
馬見塚 拓 (京都大学化学研究所 教授)	生命科学上の非構造化データの統合マイニング	生命科学においては多種多様なデータが大量に蓄積されつつあり、これら様々な形式の大量データから効率よく知識発見する手法が強く望まれている。本課題では、頻出パターンマイニングおよびクラスタリングという2つの知識発見手法に絞り込み、構造化データ(cDNA マイクロアレイデータ等)と非構造化データ(遺伝子制御や代謝パスウェイのネットワーク等)を効率よく組み合わせるシステムティックな手法を開発する。より具体的には、特に遺伝子と低分子化合物に着目し、個別の生体分子を事例とするだけでなく、遺伝子-低分子化合物の組み合わせをも事例とした新しい知識発見手法の構築を目指す。	当初あげていた2つの課題どちらにおいても、非構造化データからのデータマイニングとして、いくつかの個別のテーマに対して新規アルゴリズムを開発した。	当初計画以外の手法においても、有意義な成果が得られた。
水口 賢司 (独)医薬基盤研究所 バイオインフォマティクスプロジェクトリーダー)	ダイナミクスを考慮した膜蛋白質の構造モデリング法の開発	膜蛋白質は薬物ターゲットの大多数を占めるため、その立体構造情報は非常に重要である。近年X線構造解析により、構造が決定されるようになってきたが、解像度が低い、膜蛋白質複合体の立体構造が解かれていない、脂質分子との相互作用が明らかでない等の構造解析における困難が依然として存在する。本課題では、これらを克服するために、蛋白質データベースを用いた統計解析(バイオインフォマティクスの手法)と、物理化学的ポテンシャル関数を用いた分子シミュレーションを組み合わせ、ダイナミクスを考慮した膜蛋白質の構造モデリング技術の確立を目指す。	膜タンパク質のダイナミクスに関して実験的研究も含めて様々な角度から研究がなされ有用な手法が開発され、その適用が十分に達成されている。	実験データの蓄積に応じてデータベース開発に主軸を移したことは望ましい展開である。また、既存データの収集分類の重要性が認識され、独自性の高いデータベース開発がなされた点は望ましい。

図表 VI-10 平成 20 年度採択 継続研究開発課題（研究開発期間：平成 20 年 10 月～22 年 9 月）の成果

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
<p>内山 郁夫 (自然科学研究機構 基礎生物学研究所 助教)</p>	<p>大規模比較ゲノムワークベンチの実践的応用に向けた研究開発</p>	<p>これまでの研究開発で、大規模な比較ゲノム解析に向けたワークベンチを開発し、オーソログ分類結果全体の効果的な表示、利用者自身の持つゲノムを取り込んだ解析、内群・外群の概念導入による興味ある生物種を中心とした比較解析、近縁ゲノム間の遺伝子の並び順の保存性に基づくコア構造アライメント等の機能を実現してきた。本課題では、開発した比較ゲノムワークベンチの応用可能性をさらに広げるため、新規に決定されたゲノムとその近縁種を取り込んだ比較解析を容易にするための改良、メタゲノムデータへの適用を可能にするための開発、ならびにデータベースのさらなる大規模化に向けた効率的なデータ管理方式の開発を行う。</p>	<p>当初予定していた開発や解析がほぼ終了し、論文準備中の段階に来ている。ゲノム解読の進展が想定以上に早く、データの量、質が当初の予想と大きく異なる状況で、適切に対応してきたことは評価できる。</p>	<p>近縁ゲノム比較に有用なアプローチをうまく取り入れることが出来、ピロリ菌の解析への応用研究など、本研究の有用性を示す研究が行なわれた。また、オーソログ解析については情報を付加することにより同種ゲノム比較にも有用であることがわかった点は望ましい展開となった。</p>
<p>太田 元規 (名古屋大学大学院情報科学研究科 教授)</p>	<p>タンパク質の構造・機能・相互作用予測システムの開発と展開</p>	<p>これまでの研究開発で、タンパク質の構造を予測し、機能アノテーションを付与するシステムを製作した。このシステムをヒトゲノム由来の配列に適用し、ヒトゲノム構造・機能アノテーションデータベースを構築した。本課題では研究対象をタンパク質複合体にまで拡張し、時間概念を取り入れた複合体構造のモデル、および複合体比較と分類を提供する。また、確率的プロファイル比較法による構造予測を高速化し、実用化をはかる。</p>	<p>確率プロフィールアライメントで十分な高速化を達成できなかったなど一部、計画どおりにいかなかった点もあった。しかし、2年間という限られた期間で行うべき研究・開発を厳選した結果、本課題に重要と思われる問題について、ほぼ計画通りの成果が得られた。網羅的な解析を行ったデータベース構築など数多くの成果が出た。</p>	<p>精度の高い予測システムを開発した結果、国際的な複合体予測コンテストCAPRIにおいて正解に近い構造を返答できるようになったことは高く評価できる。また、Motion Treeの開発は想定されていなかった望ましい展開であると考えられる。</p>

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
塩田 浩平 (京都大学 大学院医学研究科 教授)	ヒト胚の形態発生に関する三次元データベース	<p>これまでの研究開発で、京都大学のヒト胚標本コレクションのうち約 1200 例の正常胚について MR 撮像を行い、高解像度の形態計測データを得た。本課題では、これを発展させて、超高磁場 7T(テスラ)の MR 装置と新規のイメージング手法である EFIC(Episodic Fluorescence Image Capture)法を用い、発生早期のヒト胚の超高解像度の画像データベースを構築する。また、ヒトの発生と先天異常のゲノムワイドな解析を行うためのインフラとなる、発生段階ごとのヒトの詳細な三次元画像データベースを構築する。</p>	<p>MR および EFIC の画像に加え、ヒト胚の連続組織切片標本や外表写真も画像データベースに加えることができ、目標以上に充実したデータベースを構築することができた。他の追随を許さないユニークなデータベースとなっている。</p>	<p>多数の連続組織切片標本や書誌データが加わり、想定していた以上に充実したデータベースとなった。</p>
白井 剛 (長浜バイオ大学 バイオサイエンス学部 教授)	実践による超分子ネットワークモデリングシステムの開発	<p>これまでの研究開発で、複合体モデリングに必要なモデリングツールを開発した。これを古細菌 3R(複製・修復・組み換え)複合体の構造解析に適用し、モデリングと実験データとの一致を確認した。本課題では、これを分子ネットワークモデリングへと発展させる。実験的に古細菌 3R 複合体のネットワークを繋ぐ複合体の発見と構造解析を行うとともに、モデリングツールにおいてもネットワークモデリングに適したツールを開発し、複合体構造の解明をめざす。</p>	<p>情報解析からのアプローチによる超分子複合体の構成を解析するシステムの構築と、電顕、生化学研究等の実験からのアプローチによるその実証から構成されており、その両輪がうまく生かされた計画である。モデリングシステムの開発と実践(実験)による検証を行う、という当初目標が達成されている。</p>	<p>超分子複合体が静的なものではなく非常に動的であることは当初予想していなかったが、構造変化を起こしたモデルの検索機能を追加することにより、対応している。動的であることは、超分子複合体の機能機序を解明する上で重要な知見であり、望ましい展開といえる。</p>

図表 VI-11 平成 21 年度採択 継続研究開発課題（研究開発期間：平成 21 年 10 月～23 年 9 月）の成果

代表研究者	研究開発課題名	研究開発計画	当初の研究開発計画から見た進捗状況や達成度等	当初計画では想定されていなかった新たな展開
黒川 顕 (東京工業大学大学院生命理工学研究科教授)	メタゲノム統合解析システムの開発	これまでの研究開発で、メタゲノムデータから遺伝子配列を予測、クラスタリングし、それらを機能別にデータベース化および可視化する、統合解析システムの開発を実施してきた。本課題では、これまでのシステムを発展させ、メタゲノム解析ワークフローを確立するとともに、メタゲノムを研究する際に重要な情報となる、細菌叢を取り巻く環境因子の各要素を記述したメタデータと、メタゲノムデータを統合し、メタ比較解析を可能とするシステム開発を目指す。さらにヒトメタゲノム研究における細菌叢とヒト免疫系との関係の解明に資するシステムの開発を目指す。	当初の研究開発計画のうち、②免疫研究に資するヒトメタゲノム情報解析技術の開発をコスト面から困難であると断念したが、その代わりに、高速シーケンサーからの大量のヒトメタゲノム配列データに対応できる可視化手法や解析パイプラインなど当初想定外の研究開発の実現を達成できた。	新型シーケンサーによるデータ産出量が想定を大幅に上回っていたが、スパコンと GPGPU を活用することにより対応可能となり、その結果、「リアルタイムメタゲノム解析」を世界に先駆けて実現したことは非常に望ましい展開になったと評価できる。
榎原 康文 (慶應義塾大学理工学部教授)	タンパク質化合物相互作用の網羅的予測手法とデータベースの開発	これまでの研究開発で、タンパク質化合物相互作用予測システムの開発を行い、wet 検証実験を利用したフィードバック戦略を確立し、それを応用することでアンドロゲンアンタゴニストの新規リード化合物を発見した。本課題では、これを発展させて、さらに、予測の機能だけを取り出して化合物探索に特化した Chemical BLAST を構築する。また、新たに 2 つのがん関連タンパク質に対するリード化合物の探索とその機能解析を目指す。	主要な課題については計画どおり進展した。テキストマイニングについては多少の遅れがあったが最終的にはシステムに組み込むことができ、全体として計画通りに進んだと評価できる。	網羅的インシリコ予測を組み合わせることで、規模や効率の点で従来の方法を大きく上回る新たなリード化合物探索法を提案・開発できたことは、望ましい展開となったと思われる。
清水 謙多郎 (東京大学大学院農学生命科学研究科教授)	高精度タンパク質間相互作用予測システムの開発	これまでの研究開発で、独自の手法によるタンパク質間相互作用予測・解析システムを構築してきた。本課題では、これを発展させて、より精度の高い統合的な相互作用予測・解析システム（相互作用予測、相互作用部位予測、ドッキングシミュレーション）を開発し、その予測・解析結果を利用して相互作用データベースを構築する。また、多様な電子伝達系を有する芳香環二原子酸素添加酵素について、実験的手法と開発するシステムの両方を用いて解析し、電子伝達時の相互作用の解明・体系化を図る。	タンパク質相互作用予測については高い予測精度を得ることができ、当初の計画通りに進んだといえる。	分子動力学シミュレーションに粗視化モデルを導入したこと、脂質結合タンパク質予測の手法の開発も行ったことなど、計画外の進展があったがこれらは好ましいものと評価できる。

3章 研究開発課題の成果 (データベース、論文※、特許等)

※論文件数は、発表したすべての件数を記しているが、主要な論文のみリストにまとめている。

1) 代表研究者：金久實

生命情報データベースの高度化・標準化 第I期

研究課題名 2項関係に基づくゲノムと生命システムの機能解読

(a) データベース・ツール

データベース・ツール名	公開サイト等
KEGG	http://www.genome.jp/kegg/
BRITE	http://www.genome.jp/kegg/brite.html
KAAS (KEGG Automatic Annotation Server)	http://www.genome.jp/tools/kaas/

(b) 論文、学会発表等

論文発表 国内 0 件、海外 25 件

口頭発表 国内 13 件、海外 33 件

ポスター発表 国内 120 件、海外 7 件

- (1) * Kanehisa, M. and Bork, P.; Bioinformatics in the post-sequence era. *Nature Genetics* 33, 305-310 (2003).
- (2) * Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M.; Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* 125, 11853-11865 (2003).
- (3) * Kotera, M., Okuno, Y., Hattori, M., Goto, S., and Kanehisa, M.; Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *J. Am. Chem. Soc.* 126, 16487-16498 (2004).
- (4) * Hashimoto, K., Goto, S., Kawano, S., Aoki-Kinoshita, K.F., Ueda, N., Hamajima, M., Kawasaki, T., and Kanehisa, M.; KEGG as a glycome informatics resource. *Glycobiology*, in press (2005).
- (5) * Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M.; From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34, D354-357 (2006).

生命情報データベースの高度化・標準化 第II期

研究課題名 ゲノムと環境の統合解析による生命システムの機能解読(KEGG)

(a) データベース・ツール

第I期と同様。

(b) 論文、学会発表等

論文発表 国内3件、海外31件

招待、基調講演 国内9件、海外24件

口頭発表 国内1件、海外0件

ポスター発表 国内21件、海外3件

特許 国内0件、国際0件

- (1) *Erguner, B., Hattori, M., Goto, S., and Kanehisa, M.; Characterizing common substructures of ligands for GPCR protein subfamilies. *Genome Informatics* 24, 31-41 (2010).
- (2) *Kotera, M., Kobayashi, T., Hattori, M., Tokimatsu, T., Goto, S., Mihara, H., and Kanehisa, M.; Comprehensive genomic analysis of sulfur-relay pathway genes. *Genome Informatics* 24, 104-115 (2010).
- (3) *Yamanishi, Y., Kotera, M., Kanehisa, M., and Goto, S.; Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26, i246-i254 (2010).
- (4) *Hattori, M., Tanaka, N., Kanehisa, M., and Goto, S.; SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic Acids Res.* 38, W652-W656 (2010).
- (5) *Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S., and Kanehisa, M.; PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res.* 38, W138-W143 (2010).
- (6) * Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M.; KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38, D355-D360 (2010).
- (7) * Shigemizu, D., Araki, M., Okuda, S., Goto, S., and Kanehisa, M.; Extraction and analysis of chemical modification patterns in drug development. *J. Chem. Inf. Model.* 49, 1122-1129 (2009).
- (8) * Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y.; KEGG for linking genomes to life and the environment.

Nucleic Acids Res. 36, D480-D484 (2008).

- (9) * Hashimoto, K., Yoshizawa, A. C., Okuda, S., Kuma, K., Goto, S., and Kanehisa, M.; The repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes. J. Lipid Res. 49, 183-191 (2008).
- (10) * Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., and Kanehisa, M.; Network-based analysis and characterization of adverse drug-drug interactions. J. Chem. Inf. Model. 51, 2977-2985 (2011).

2) 代表研究者：菅原秀明

生命情報データベースの高度化・標準化 第I期

研究課題名 ゲノム生物学バックボーンデータベースの構築提供

(a) データベース・ツール

データベース・ツール名	公開サイト等
GIB (Genome Information Broker)	http://gib.genes.nig.ac.jp/
G-InforBIO	http://www.wdcm.org/inforbio/G-InforBIO/download.html
OASYS	http://althea.ddbj.nig.ac.jp/index.jsp
Gene Trek in Prokaryote Space (GTPS)	http://gtps.ddbj.nig.ac.jp/
DDBJ-XML	http://gtps.ddbj.nig.ac.jp/
GTOP	http://spock.genes.nig.ac.jp/~genome/gtop.html
CIBEX	http://cibex.nig.ac.jp/index.jsp
BioSimulated DataBase (BSD)	http://bsd.genes.nig.ac.jp/bsd_web/Top.jsp
BMRB	http://bmrp.protein.osaka-u.ac.jp/
WABI	http://www.xml.nig.ac.jp/
CookBook	http://wabi.ddbj.nig.ac.jp/
GTPS	http://gtps.ddbj.nig.ac.jp/
GIB-V	http://gib-v.genes.nig.ac.jp/
WNS	http://wns.nig.ac.jp/
GIB-IS	http://gib-is.genes.nig.ac.jp/
DDBJ Sequence Read Archive	http://trace.ddbj.nig.ac.jp/dra/index.shtml
DDBJ Trace Archive	http://trace.ddbj.nig.ac.jp/dta/dta_index.shtml
MiGAP	http://migap.lifesciencedb.jp/
パイオ・メタデータベース	http://www.ps.noda.tus.ac.jp/biometadb/index.html
メソッド統合検索サイト	http://lifesciencedb.jp/dbsearch/bird/

(b) 論文、学会発表等

論文発表 国内 9 件、海外 49 件 (原著論文のみ)

口頭発表 国内 45 件、海外 32 件

ポスター発表 国内 34 件、海外 10 件

特許出願 なし

- (1) * Sugawara, H., Miyazaki, S., Abe, T., and Shigemoto, Y. (2005).
Biological Data Analysis using DDBJ Web services. Proceedings of
BIOINFO2005. 379-382.
- (2) * Riley, M., Abe, T., Arnaud, B.M., Berlyn, M., Blattner, R.F.,
Chaudhuri, R.R., Glasner, D.J., Horiuchi, T., Keseler, M.I., Kosuge,
T., Mori, H., Perna, T.N., Plunkett, G., Rudd, E.K., Serres, H.M.,
Thomas, H.G., Thomson, R.H., Wishart, D., and Wanner, L.B. (2006).
Escherichia coli K-12: a cooperatively developed annotation
snapshot-2005, Nucleic Acids Research, 34, 1-9.
- (3) * Homma, K., Fukuchi, S., Kawabata, T., Ota, M. and Nishikawa, K. :
A systematic investigation identifies a significant number of probable
pseudogenes in the Escherichia coli genome. Gene, 294, 25-33, 2002.
- (4) * Fukuchi, S. and Nishikawa, K. (2004). Estimation of the number of
authentic orphan genes in bacterial genomes. DNA Res., 11, 219-231.
- (5) * Matsumura, Y., Shimokawa, K., Ikeo, K., Tateno, Y., Hayashizaki,
Y. and Kawai, J. Development of the reliability index for the
measurement value of each spot in a DNA microarray (RIESM) and addition
of RIESM to READ of CIBEX. Gene (accepted)

生命情報データベースの高度化・標準化 第II期

研究課題名 バイオ基幹情報資源の高準化と共用化

(a) データベース・ツール

第I期と同様。

(b) 論文、学会発表等

論文発表 国内0件、海外14件

招待、基調講演 国内2件、海外0件

口頭発表 国内10件、海外0件

ポスター発表 国内8件、海外19件

特許 国内0件、国際0件

- (1) * Leinonen R, Sugawara H and Shumway M on behalf of the International Nucleotide Sequence Database Collaboration (2011) "The Sequence Read Archive" *Nucleic Acids Res.* 39:D19-D21
- (2) * Katayama T, Arakawa K, Nakao M, Ono K, Aoki-Kinoshita KF, Yamamoto Y, Yamaguchi A, Kawashima S, Chun HW, Aerts J, Aranda B, Barboza LH, Bonnal RJ, Bruskiwich R, Bryne JC, Fernández JM, Funahashi A, Gordon PM, Goto N, Groscurth A, Gutteridge A, Holland R, Kano Y, Kawas EA, Kerhornou A, Kibukawa E, Kinjo AR, Kuhn M, Lapp H, Lehvaslaiho H, Nakamura H, Nakamura Y, Nishizawa T, Nobata C, Noguchi T, Oinn TM, Okamoto S, Owen S, Pafilis E, Pocock M, Prins P, Ranzinger R, Reisinger F, Salwinski L, Schreiber M, Senger M, Shigemoto Y, Standley DM, Sugawara H, Tashiro T, Trelles O, Vos RA, Wilkinson MD, York W, Zmasek CM, Asai K, Takagi T. (2010) "The DBCLS BioHackathon: standardization and interoperability for bioinformatics web services and workflows" The DBCLS BioHackathon Consortium. *J Biomed Semantics.* Aug 21;1(1):8.
- (3) * Kodama Y, Kaminuma E, Saruhashi S, Ikeo K, Sugawara H, Tateno Y, Nakamura Y, (2010) "Biological databases at DNA Data Bank of Japan in the era of next-generation sequencing technologies," *Adv Exp Med Biol.*, 680:125-35.
- (4) * Kwon Y., Shigemoto Y., Kuwana Y., Sugawara H., (2009) "Web API for biology with a workflow navigation system," *Nucleic Acids Res.*, Jul 1:37(Web Server issue):W11-6, Epub 2009 May 5.
- (5) * Kosuge T., Abe T., Okido T., Tanaka N., Hirahata M., Maruyama Y., Tomiki A., Kurokawa M., Himeno R., Fukuchi S., Miyazaki S., Gojobori T., Tateno Y., Sugawara H., (2006) "Exploration and grading of possible genes in 183 bacterial strains by a common fine protocol lead to new genes: Gene Trek in Prokaryote Space (GTPS)," *DNA Res.*, 13, 245-254.

3) 代表研究者：高木利久

生命情報データベースの高度化・標準化 第I期

研究会題名 シグナルオントロジーとバイオタームバンクの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
INOH	http://www.inoh.org/download.html
INOH クライアント	http://www.inoh.org/Web_Installers/install.htm
Ontology Viewer	http://www.inoh.org/ontology-viewer/
オントロジーデータ	http://www.inoh.org/download.html
Similarity Search	http://www.inoh.org/similarity-search/

(b) 論文、学会発表等

論文発表 国内11件、海外12件

口頭発表 国内11件、海外9件

ポスター発表 国内3件、海外8

特許出願 なし

- (1) * Fukuda, K. and Takagi, T., “Knowledge Representation of Signal, Transduction Pathways”, *BIOINFORMATICS*, Vol.17, pp.829-837, (22)
- (2) * Fukuda, K, Y Yamagata and Takagi, T., “FREX: a query interface for biological processes with a hierarchical and recursive structures”, *In Silico Biology*, 4(1), pp63-79, (2004)
- (3) * 福田賢一郎、山縣友紀、高木利久「パスウェイデータベース：バイオロジーにおける高次知識のデータベース化」情報処理学会シンポジウムシリーズ、情報処理学会：2003 巻、8 号、ISSN 1344 - 0640、pp. 267-272, 2003 年 11 月
- (4) * Yamamoto, S., Asanuma, T., Takagi, T. and Fukuda, K., “The Molecule, Role Ontology: an ontology for annotation of signal transduction pathway molecules in the scientific literature.”, *Comparative and Functional Genomics*, Vol.2004, 5, 528-536. (2004.12)
- (5) * Kushida, T., Takagi, T. and Fukuda, K., “Event Ontology : A Pathway-Centric Ontology for biological processes.”, *Pacific Symposium on Biocomputing 2006*, Vol.11, pp.152-163, 2006.1

生命情報データベースの高度化・標準化 第II期

研究課題名 オントロジーによるパスウェイの高度化および国際標準化 (INOH パスウェイデータベース)

(a) データベース・ツール

第I期と同様。

(b) 論文、学会発表等

論文発表 国内4件、海外5件

招待、基調講演 国内0件、海外2件

口頭発表 国内1件、海外4件

ポスター発表 国内3件、海外5件

特許 国内0件、国際0件

- (1) * Fukuda, K., Yamamoto, S., Sakai, N., Nakamura, H., Nakanishi, Y., and Takagi, T., "Graphical syntax and query for pathway database." The 10th World Multi-Conference on Systemics, Cybernetics and Informatics proceedings (WMSCI2006), vol. IV, ISBN:980-6560-65-5, pp. 7-10. (2006).
- (2) * Kushida, T., Takagi, T., Fukuda, K.: "Event Ontology: A Pathway-Centric Ontology for Biological Processes", Pacific Symposium on Biocomputing, 11, pp.152-163. (2006).
- (3) * Ken Fukuda, "INOH pathway database: Curation, Annotation, Integration" Interdisciplinary Ontology, ISBN978-4-904239-00-1, pp47-50. (2008).
- (4) * Demir, E., Cary, M.P., Paley, S., Fukuda, K., et al. The BioPAX community standard for pathway data sharing. Nat. Biotechnol., 28, 935-942. (2010)
- (5) * 福田 賢一郎: "BioPAX: パスウェイデータフォーマットの標準化とオントロジー", 生物物理, 47(3), pp.179-84. (2007).

4) 代表研究者：中村春木

生命情報データベースの高度化・標準化 第I期

研究課題名 タンパク質立体構造データベースの高度化

(a) データベース・ツール

データベース・ツール名	公開サイト等
PDBj	http://www.pdbj.org/
RCSB-PDB	http://pdb.protein.osaka-u.ac.jp/pdb/
xPSSS	http://www.pdbj.org/xpsss/
PDBj-Mine	http://www.pdbj.org/mine/
PDBjViewer	http://pdbj.protein.osaka-u.ac.jp/PDBjViewer/
jV (PDBj Viewer)	http://www.pdbj.org/jv/
PDB Archive	ftp://pdb.protein.osaka-u.ac.jp/ 、 rsync://pdb.protein.osaka-u.ac.jp/
PDBj-BMRB	http://bmrdep.protein.osaka-u.ac.jp
BMRB ミラーサーバ	http://bmrdep.protein.osaka-u.ac.jp
BMRB データ登録日本語ウェブ マニュアル	http://bmrdep.protein.osaka-u.ac.jp/manual_top.html
SeSAW	http://www.pdbj.org/sesaw/
GIRAF	http://www.pdbj.org/giraf/
MAFFTash	http://www.pdbj.org/mafftash/
eF-site	http://ef-site.hgc.jp/eF-site/
eF-surf	http://ef-site.hgc.jp/eF-surf/
eF-seek	http://ef-site.hgc.jp/eF-seek
ProMode	http://promode.socs.waseda.ac.jp/
SEALA	http://seala.cbrc.jp/~SEALA/
ASH	http://timpani.genome.ad.jp/~ash/
eProtS	http:// eprotS.protein.osaka-u.ac.jp/eProtS/Top.jsp
PROFESS	未公開
MoM	http://www.pdbj.org/mom/index.php
Protein Globe	http://www.pdbj.org/globe/
Sequence Navigator	http://www.pdbj.org/seqnavix/
Structure Navigator	http://www.pdbj.org/stnavix/
EM Navigator	http://www.pdbj.org/emnavi/
CRNPRED	http://www.pdbj.org/crnpred/
Spanner	http://www.pdbj.org/spanner/
SFAS	http://www.pdbj.org/sfas/

(b) 論文、学会発表等

論文発表 国内4件、海外26件

口頭発表 国内50件、海外15件

ポスター発表 国内36件、海外12件

ニュースレター発行 7号

- (1) * Standley DM, Toh H, Nakamura H “Detecting Local Structural Similarity in Proteins by Mixing the Number of Equivalent Residues” , Proteins: Structure, Function and Genetics, Vol. 57, pp. 381-391 (2004)
- (2) * Kinoshita K, Nakamura H ” eF-site and PDBjViewer: database and viewer for protein functional sites.” Bioinformatics Vol. 20, Iss. 8; pp. 1329-1330 (2004)
- (3) * Wako H, Kato M, Endo S “ProMode: a database of normal mode analysis on protein molecules with a full-atom model” , Bioinformatics, Vol. 20, Iss. 13; pp. 2035-2043(2004)
- (4) * Kinoshita K, Nakamura H “Identification of the ligand binding sites on the molecular surface of proteins” Protein Science, Vol. 14, pp. 711-718 (2005)
- (5) * Westbrook J., Ito, N., Nakamura, H., Henrick K., Berman H.M. “ PDBML: The representation of archival macromolecular structure data in XML.” Bioinformatics Vol. 21, Iss. 7; pp. 988-992 (2005)

生命情報データベースの高度化・標準化 第II期

研究課題名 蛋白質構造データバンクの国際的な構築と高度化 (PDBj)

(a) データベース・ツール

第I期と同様。

(b) 論文、学会発表等

論文発表 国内4件、海外55件

招待、基調講演 国内16件、海外23件

口頭発表 国内29件、海外5件

ポスター発表 国内56件、海外14件

特許 国内0件、国際0件

- (1) * Berman, H.M. , Henrick, K, Nakamura, H and Markley, J. L. “The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive

- of PDB data” *Nucleic Acids Research*, 35 Database issue D301–D303 (2006)
- (2) Kanamori E., Murakami, Y., Tsuchiya, Y., Standley, D.M., Nakamura, H., Kinoshita, K., “Docking of protein molecular surfaces with evolutionary trace analysis.” *PROTEINS*, 69, 832–838 (2007)
 - (3) Kinjo, A.R., Nakamura, H., Similarity search for local protein structures at atomic resolution by exploiting a database management system. *BIOPHYSICS*, 3, 75–84 (2007)
 - (4) Matsuki, Y., Akutsu, H., Fujiwara, T.,” Spectral Fitting for Signal Assignment and Structural Analysis of Uniformly ¹³C-Labeled Solid Proteins by Simulated Annealing Based on Chemical Shifts and Spin Dynamics.” *J. Biomol. NMR*, 38, 325–339 (2007)
 - (5) Henrick, K., Feng, Z., Bluhm, W.F., Dimitropoulos, D., Doreleijers, J.F., Dutta, S., Flippen-Anderson, J.L., Ionides, J., Kamada, C., Krissinel, E., Lawson, C.L., Markley, J.L., Nakamura, H., Newman, R., Shimizu, Y., Swaminathan, J., Velankar, S., Ory, J., E. Ulrich, E.L., Vranken, W., Westbrook, J., Yamashita, R., Yang, H., Young, J., Yousufuddin, M., Berman, H.M., “Remediation of the protein data bank archive.” *Nucleic Acids Res.*, 36, D426–433 (2008)
 - (6) Kinjo, A.R., Miyazawa, S., “On the optimal contact potential of proteins.” *Chemical Physics Letters*, 451, 132–135 (2008)
 - (7) * Markley, J.L., Urich, E.L., Berman, H.M., Henrick, K., Nakamura, H., Akutsu, H., “BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions.” *J. Biomol NMR*, 40, 153–155 (2008)
 - (8) * Standley DM, Kinjo A.R., Kinoshita K, Nakamura H. “Protein structure databases with new web services for structural biology and biomedical research.” *Brief Bioinform*, 9(4), 276–285 (2008)
 - (9) * Johansson, F., Toh, H. “A comparative study of conservation and variation scores.” *BMC Bioinformatics*, 11, 388 (2010)
 - (10) W Nunomura, K Kinoshita, M Parra, P Gascard, X An, N Mohandas, and Y Takakuwa, “Similarities and differences in the structure and function of 4.1G and 4.1R135, two protein 4.1 paralogues expressed in erythroid cells.” *Biochem J*, 432, 407–416(2010)
 - (11) * Kinjo A.R., Yamashita R, Nakamura H., “PDBj Mine: Design and implementation of relational database interface for Protein Data Bank Japan.” *Database*, 2010:baq021 (2010)

5) 代表研究者：稲葉一男

生命情報データベースの高度化・標準化 第II期

研究会題名 ホヤプロテイン統合データベースの構築

(a) データベース・ツール

データベース・ツール名	公開サイト等
CIPRO	http://cipro.ibio.jp/2.5
Ghost	http://hoya.zool.kyoto-u.ac.jp/cgi-bin/gbrowse/kh/
CiAID (Ciona intestinalis Adult In situ hybridization Database) ver.2 beta	http://bioinfo.s.chiba-u.jp/ciaid/
FABA1	http://chordate.bpni.bio.keio.ac.jp/faba/top.html
FABA2	http://chordate.bpni.bio.keio.ac.jp/faba2/2.0/top.html
3DPL	http://chordate.bpni.bio.keio.ac.jp/3dpl/top.html
ACBD	(2011 年中の公開を予定)

(b) 論文、学会発表等

論文発表 国内 0 件、海外 31 件

招待、基調講演 国内 12 件、海外 9 件

口頭発表 国内 30 件、海外 5 件

ポスター発表 国内 38 件、海外 11 件

特許 国内 0 件、国際 0 件

- (1) * Hotta, K, Mitsuhashi, K, Takahashi, H, Inaba, K, Gojobori, T, Oka, K. and Ikeo, K. The 3D real image resources of developmental embryos in *Ciona intestinalis*. I. From fertilized egg to hatching larva. *Dev. Dyn.* (2007) 236, 1790–1805, 2007.
- (2) * Satou Y, Mineta K, Ogasawara M, Sasakura Y, Shoguchi E, Ueno K, Yamada L, Matsumoto J, Wasserscheid J, Dewar K, Wiley GB, Macmillan SL, Roe BA, Zeller RW, Hastings KEM, Lemaire P, Lindquist E, Endo T, Hotta K, and Kazuo Inaba. Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. *Genome Biol* (2008) 9, R152.
- (3) * Nomura M, Nakajima A. and Inaba K. Proteomic profiles of embryonic development in the ascidian *Ciona intestinalis*. *Dev. Biol.* (2009) 325, 468–481.
- (4) * Endo T, Ueno K, Yonezawa K, Mineta K, Hotta K, Satou Y, Yamada L,

Ogasawara M, Takahashi H, Nakajima A, Nakachi M, Nomura M, Yaguchi J, Sasakura Y, Yamazaki C, Sera M, Yoshizawa AC, Imanishi T, Taniguchi H, Inaba K. : CIPRO 2.5: Ciona intestinalis protein database, a unique integrated repository of large-scale omics data, bioinformatic analyses and curated annotation, with user rating and reviewing functionality. Nucleic Acids Res. Suppl 1, D807-D814, 2011.

- (5) * O. Tassy, D. Dauga, F. Daian, D. Sobral, F. Robin, P. Khoeiry, D. Salgado, V. Fox, D. Caillo, M. Contensin, Renaud Schiappa, Anne Rios, Guillaume Luxardi, M. Gilchrist, K. Makabe, K. Hotta, Fujiwara, T. Kusakabe, N. Satoh, Y. Satou and P. Lemaire. The ANISEED database: digital representation, formalisation and elucidation of a chordate developmental program. Genome Research. (2010) [Epub ahead of print]

6) 代表研究者：西岡孝明

生命情報データベースの高度化・標準化 第II期

研究会題名 メタボローム MS スペクトル統合データベースの開発
(Metabolome-Mass Spectral Database)

(a) データベース・ツール

データベース・ツール名	公開サイト等
MassBank	http://www.massbank.jp
Fragmentation Library	http://metabolomics.jp/wiki/Index:MassBank (2012年12月公開予定)
ESI-MS ²	(2012年3月公開予定)
Metabolomics.JP	http://metabolomics.jp/wiki/
LipidBank	http://kanaya.naist.jp/KNApSAcK_Family/
KNApSAcK	http://kanaya.naist.jp/KNApSAcK_World/top.jsp
KNApSAcK from around the world	http://kanaya.naist.jp/KNApSAcK_World/top.jsp
Lunchbox DB	http://kanaya.naist.jp/LunchBox/top.jsp
KAMPO DB	http://kanaya.naist.jp/kampo/top.jsp
インドネシア生薬データベース	http://kanaya.naist.jp/jamu/top.jsp

(b) 論文、学会発表等

論文発表 国内 6 件、海外 16 件

招待、基調講演 国内 8 件、海外 0 件

口頭発表 国内 26 件、海外 9 件

ポスター発表 国内 21 件、海外 5 件

特許 国内 0 件、国際 0 件

- (1) * Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Yokota-Hirai, M., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K. and Nishioka, T. “MassBank: A public repository for sharing mass spectral data for life sciences”, *J. Mass Spectrometry*, 45(7), 703-714 (2010).
- (2) * Oishi, T., Tanaka, K., Hashimoto, T., Shinbo, Y., Jumtee, K., Bamba, T., Fukusaki, E., Suzuki, H., Shibata, D., Takahashi, H., Asahi, H., Kurokawa, K., Nakamura, Y., Hirai, A., Nakamura, K., Md. Altaf-Ul-Amin, Kanaya S., An approach to peak detection in GC-MS chromatograms and application of KNApSAcK database in prediction of candidate metaboltes, *Plant Biotechnol.*, 26, 167-168 (2009).
- (3) * Tanaka, K., Nakamura, K., Saito, T., Osada, H., Hirai, A., Takahashi, H., Kanaya, S., Md. Altaf-Ul-Amin, Metabolic pathway prediction based on inclusive relation between cyclic substructures, *Plant Biotechnol.*, 26, 459-468 (2009) .
- (4) * Takahashi, H., Kai, Kosuke, Shinbo, Y., Tanaka, K., Ohta, D., Oshima, T., Md. Altaf-Ul-Amin, Kurokwa, K., Ogasawara, N., Kanaya, S., Metabolomics approach for determining growth-specific metabolites based on Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Biolanal Chem.*, 391, 2769-2782 (2008)

7) 代表研究者：森下真一

生命情報データベースの高度化・標準化 第II期

研究課題名マルチモーダ統合バイオDB (Multimodal BIODB)

(a) データベース・ツール

データベース・ツール名	公開サイト等
UTGB Toolkit	http://utgenome.org/
SCMD	http://scmd.gi.k.u-tokyo.ac.jp/datamine/
siDirect/dsCheck	http://design.rnai.jp/ http://dscheck.rnai.jp/
UT Genome Browser (Medaka)	http://medaka.utgenome.org/
UT Genome Browser (Yeast)	http://yeast.utgenome.org/
PrimerStation	http://ps.cb.k.u-tokyo.ac.jp/
MachiBase	http://machibase.gi.k.u-tokyo.ac.jp/
FreClu	http://mlab.cb.k.u-tokyo.ac.jp/~quwei/DeNovoShortReadClustering/

(b) 論文、学会発表等

論文発表 国内 0 件、海外 19 件

招待、基調講演 国内 13 件、海外 8 件

口頭発表 国内 7 件、海外 2 件

ポスター発表 国内 0 件、海外 8 件

特許 国内 0 件、国際 0 件

- (1) * Miura F, Kawaguchi N, Sese J, Toyoda A, Hattori M, Morishita S, and Ito T. A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. Proc Natl Acad Sci U S A. 103(47):17846-51 (2006)
- (2) * Budrul Ahsan, Daisuke Kobayashi, Tomoyuki Yamada, Masahiro Kasahara, Shin Sasaki, Taro L. Saito, Yukinobu Nagayasu, Koichiro Doi, Yoichiro Nakatani, Wei Qu, Tomoko Jindo, Atsuko Shimada, Kiyoshi Naruse, Atsushi Toyoda, Yoko Kuroki, Asao Fujiyama, Takashi Sasaki, Atsushi Shimizu, Shuichi Asakawa, Nobuyoshi Shimizu, Shin-ichi Hashimoto, Jun Yang, Yongjun Lee, Kouji Matsushima, Sumio Sugano, Mitsuru Sakaizumi, Takanori Narita, Kazuko Ohishi, Shinobu Haga, Fumiko Ohta, Hisayo Nomoto, Keiko Nogata, Tomomi Morishita, Tomoko Endo, Tadasu Shin-I, Hiroyuki Takeda, Yuji Kohara, and Shinichi Morishita. UTGB/medaka:

- genomic resource database for medaka biology. *Nucleic Acids Research*, Vol. 36, Database issue D747-D752 (2008)
- (3) * The International Silkworm Genome Consortium (S. Morishita as one of corresponding authors). The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, Vol. 38, Issue 12, 1036-1045 (2008)
- (4) * Shin Sasaki, Cecilia C. Mello, Atsuko Shimada, Yoichiro Nakatani, Shin-ichi Hashimoto, Masako Ogawa, Kouji Matsushima, Sam Guoping Gu, Masahiro Kasahara, Budrul Ahsan, Atsushi Sasaki, Taro Saito, Yutaka Suzuki, Sumio Sugano, Yuji Kohara, Hiroyuki Takeda, Andrew Fire(#), Shinichi Morishita(#) Chromatin-Associated Periodicity in Genetic Variation Downstream of Transcriptional Start Sites. *Science*. 323(5912):401-4 (2009) # joint corresponding authors
- (5) * Budrul Ahsan, Taro L. Saito, Shin-ichi Hashimoto, Keigo Muramatsu, Manabu Tsuda, Atsushi Sasaki, Kouji Matsushima, Toshiro Aigaki, and Shinichi Morishita. MachiBase: a *Drosophila melanogaster* 5'-end mRNA transcription database. *Nucleic Acids Research*, Vol. 37, Database issue D49-D53 (2009)
- (6) * Taro L. Saito, Jun Yoshimura, Shin Sasaki, Budrul Ahsan, Atsushi Sasaki, Reginaldo Kuroshu and Shinichi Morishita. UTGB Toolkit for Personalized Genome Browsers, *Bioinformatics* 25(15):1856-1861 (2009)

創造的な生物・情報知識融合型の研究開発

平成 13 年度採択 新規研究開発課題 (研究開発期間 : 平成 13 年 10 月 ~ 16 年 9 月)

研究課題名 遺伝子破壊株イメージ・マイニング

(a) データベース・ツール

公開なし

(b) 論文、学会発表等

論文発表 国内 0 件、海外 20 件

- (1) * Miwaka Ohtani, Ayaka Saka, Fumi Sano, Yoshikazu Ohya, and Shinichi Morishita. Development of Image Processing Program for Yeast Cell Morphology. *J. Bioinfo. Comp. Biol.* 1(4), 695-709 (2004).
- (2) * Taro L. Saito, Miwaka Ohtani, Hiroshi Sawai, Fumi Sano, Ayaka Saka, Daisuke Watanabe, Masashi Yukawa, Yoshikazu Ohya, and Shinichi

Morishita. SCMD: *Saccharomyces Cerevisiae* Morphological Database. *Nucl. Acids. Res.* 32, D319-D322 (2004).

- (3) * Masaya Suzuki, Ryoji Igarashi, Mizuho Sekiya, Takahiko Utsugi, Shinichi Morishita, Masashi Yukawa, Yoshikazu Ohya. Dynactin is involved in a checkpoint to monitor cell wall synthesis in *Saccharomyces cerevisiae*. *Nature Cell Biology* 6, 861 - 871 (2004).
- (4) * Jun Sese and Shinichi Morishita. Itemset Classified Clustering. *Lecture Notes in Computer Science*, Vol. 3202, 398-409 (2004).

平成 16 年度採択 継続研究開発課題 (研究開発期間 : 平成 16 年 10 月~19 年 9 月)
研究課題名 遺伝子破壊株イメージ・マイニング

(a) データベース・ツール

公開なし

(b) 論文、学会発表等

論文発表 国内 0 件、海外 10 件

口頭発表 国内 2 件、海外 1 件

ポスター発表 国内 19 件、海外 0 件

- (1) * T.L.Sato, J. Sese, Y. Nakatani, F. Sano, M. Yukawa, Y. Ohya and S. Morishita, Data Mining Tools for the *Saccharomyces cerevisiae* Morphological Database., *Nucl. Acids. Res.*, 33, (Web Server issue): W589-W591 (2005).
- (2) * Y. Ohya, J. Sese, M. Yukawa, F. Sano, Y. Nakatani, T.L. Saito, A. Saka, T. Fukuda, S. Ishihara, S. Oka, G. Suzuki, M. Watanabe, A. Hirata, M. Ohtani, H. Sawai, N. Fraysse, J-P. Latge, J.M. Francois, M. Aebi, S. Tanaka, S. Muramatsu, H. Araki, K. Sonoike, S. Nogami and S. Morishita, High-dimensional and large-scale phenotyping of yeast mutants., *Proc. Natl. Acad. Sci. USA*, 102(52), 19015-19020 (2005)
- (3) * G. Suzuki, H. Sawai, M. Ohtani, S. Nogami, F. Sano-Kumagai, A. Saka, M. Yukawa, T.L. Saito, J. Sese, D. Hirata, S. Morishita and Y. Ohya, Evaluation of image processing programs for accurate measurement of budding and fission yeast morphology., *Curr. Genet.*, 1-22 (2006)

8) 代表研究者：伊藤啓

平成 13 年度採択 新規研究開発課題（研究開発期間：平成 13 年 10 月～16 年 9 月）

研究課題名 ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築

(a) データベース・ツール

公開なし

(b) 論文、学会発表等

論文発表 国内 0 件、海外 7 件

- (1) * Tanaka, N. K., Awasaki, T., Shimada, T. and Ito, K. Integration of chemosensory pathways in the Drosophila second-order olfactory centers. *Current Biology* 14, 449-457 (2004).
- (2) * Awasaki, T. and Ito, K. Engulfing action of glial cells is required for programmed axon pruning during Drosophila metamorphosis. *Current Biology* 14, 668-677 (2004).
- (3) * Ito, K., Okada, R., Tanaka, N. K. and Awasaki, T. Cautionary observations on preparing and interpreting brain images using molecular biology-based staining techniques. *Microsc Res Tech.* 62, 170-186 (2003).
- (4) * Kido, A. and Ito, K. Mushroom bodies are not required for courtship behavior by normal and sexually mosaic Drosophila. *J Neurobiol.* 52, 302-311 (2002).
- (5) * Hayashi, S., Ito, K., Sado, Y., Taniguchi, M., Akimoto, A., Takeuchi, H., Aigaki, T., Matsuzaki, F., Nakagoshi, H., Tanimura, T. et al. GETDB, a database compiling expression patterns and molecular locations of a collection of Gal4 enhancer traps. *Genesis* 34, 58-61(2002).

平成 16 年度採択 継続研究開発課題（研究開発期間：平成 16 年 10 月～18 年 9 月）

研究課題名 ショウジョウバエ脳神経回路の徹底解析にもとづく感覚情報処理モデルの構築

(a) データベース・ツール

公開なし

(b) 論文、学会発表等

論文発表 国内 0 件、海外 5 件

口頭発表 国内 10 件、海外 6 件

ポスター発表 国内 6 件、海外 18 件

- (1) * Kmikouchi A., Shimada T. and Ito K., Comprehensive classification of the auditory sensory projections in the brain of the fruit fly *Drosophila melanogaster.*, *Journal of Comparative Neurology* 499, 317-356 (2006).
- (2) * Otsuka H. and Ito K., Systematic analysis of the visual projection neurons of *Drosophila melanogaster*-I: Lobula-specific pathways., *Journal of Comparative Neurology* 497, 928-958 (2006).
- (3) * Liu G., Seiler H., Wen A., Zars T., Ito K., Wolf R., Heisenberg M. and Liu L., Distinct memory traces for two visual features in the *Drosophila* brain., *Nature* 439, 551-556 (2006).

9) 代表研究者：伊藤隆司

平成 13 年度採択 新規研究開発課題（研究開発期間：平成 13 年 10 月～16 年 9 月）

研究課題名 インタラクトーム解析からの生物知識獲得

(a) データベース・ツール

公開なし

(b) 論文、学会発表等

論文発表 国内 1 件、海外 16 件

- (1) * Oyama, T., Kitano, K., Satou, K. & Ito, T. Extraction of knowledge on protein-protein interaction by association rule discovery. *Bioinformatics* 18, 705-714 (2002).
- (2) * Ago, T., Kuribayashi, F., Hiroaki, H., Takeya, R., Ito, T., Kohda, D. & Sumimoto, H. Phosphorylation of p47*phox* directs PX domain from SH3 domain toward phosphoinositides, leading to activation of the phagocyte NADPH oxidase. *Proc. Natl. Acad. Sci. USA* 100, 4474-4479 (2003).
- (3) * Minakuchi, Y., Satou, K. & Konagaya, A. Prediction of Protein-Protein Interaction Site Using Support Vector Machines. The 2003 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS'03), 22-28

(2003).

- (4) * Feng, S.-Y., Ota, K., Yamada, Y., Sawabu, N. & Ito, T. A yeast one-hybrid system to detect methylation-dependent DNA-protein interactions. *Biochem. Biophys. Res. Commun.* 313, 922-925 (2004).
- (5) * Onda, M., Ota, K., Chiba, T., Sakaki, Y. & Ito, T. Analysis of gene network regulating yeast multidrug resistance by artificial activation of transcription factors: involvement of Pdr3 in salt resistance. *Gene* 332, 51-59 (2004).

平成 16 年度採択 継続研究開発課題 (研究開発期間 : 平成 16 年 10 月~18 年 9 月)
研究課題名 絶対定量オーミックスからの知識発見

(a) データベース・ツール

データベース・ツール名	公開サイト等
SDSSPrimer	http://itolab.cb.k.u-tokyo.ac.jp/GATC/SDSSPrimer.html
UTGB Yeast	http://yeast.utgenome.org/

(b) 論文、学会発表等

論文発表 国内 0 件、海外 32 件

口頭発表 国内 5 件、海外 6 件

ポスター発表 国内 3 件、海外 2 件

- (1) * Miura F., Uematsu C., Sakaki Y. & Ito T. : A novel strategy to design highly specific PCR primers based on the stability and uniqueness of 3' -end subsequences., *Bioinformatics* 21, 4363-4370 (2005).
- (2) * Pham T.H., Clemente J.C., Satou K. & Ho T.B. : Computational discovery of transcriptional regulatory rules., *Bioinformatics* 21, Suppl. 2, 101-107 (2005).
- (3) * Pham T.H., Tram D.H., Ho T.B., Satou K. & Valiente G. : Qualitatively predicting acetylation and methylation areas in DNA sequences., *Genome Inform.*, 16(2), 3-11 (2005)
- (4) * Miura F., Kawaguchi N., Sese J., Toyoda A., Hattori M., Morishita S. & Ito T. : A large-scale full-length cDNA analysis to explore the budding yeast transcriptome., *Proc. Natl. Acad. Sci. USA*, 103, 17846-17851 (2006)
- (5) * Yamaguchi Y., Ota K. & Ito T. : A novel cdc42-interacting domain of the yeast polarity establishment protein Bem 1: implications for

- modulation of mating pheromone signaling., *J. Biol. Chem.*, 282, 29-38 (2007).
- (6) * Kito, K., Ota, K., Fujita, T. & Ito, T.: A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes., *J. Proteome Res.*, 6, 792-800 (2007).
- (7) * Kito, K., Kawaguchi, N., Okada, S. & Ito, T.: Discrimination between stable and dynamic components of protein complexes by means of quantitative proteomics., *Proteomics*, 8, 2366-2370 (2008)
- (8) * Miura, F., Kawaguchi, N., Yoshida, M., Uematsu, C., Kito, K., Sakaki, Y. & Ito, T.: Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs., *BMC Genomics* 9, 574 (2008).

10) 代表研究者：大浪修一

平成 13 年度採択 新規研究開発課題（研究開発期間：平成 13 年 10 月～16 年 9 月）
研究会題名 線虫 *C. elegans* 発生過程のシステム解析

(a) データベース・ツール

データベース・ツール名	公開サイト等
DBRF-MEGN 法ソフトウェア	http://so.qbic.riken.jp/dbrf-megn/
線虫初期胚用核検出ソフトウェア	http://so.qbic.riken.jp/nuclear_detection/

(b) 論文、学会発表等

論文発表 国内 0 件、海外 2 件（投稿中 2 件）

- (1) * Kyoda, K., Baba, K., Onami, S., Kitano, H. DBRF-MEGN method: an algorithm for deducing minimum equivalent gene networks from large-scale gene expression profiles of gene deletion mutants. *Bioinformatics* 20, 2662-2675 (2004).
- (2) * Hamahashi, S., Onami, S., Kitano, H. Detection of nuclei in 4D Nomarski DIC microscope images of early *Caenorhabditis elegans* embryos using local image entropy and object tracking. *BMC Bioinformatics* 6, 125 (2005).
- (3) * Kimura, A., Onami, S. Computer simulations and image processing reveal length-dependent pulling force as the primary mechanism for *C. elegans* male pronuclear migration. *Dev. Cell* 8, 765-775 (2005).

- (4) * Hamahashi, S., Kitano, H., Onami, S. A system for measuring cell division patterns of early *Caenorhabditis elegans* embryos by using image processing and object tracking. *Systems Comput. Jpn.* 38(11), 12-24 (2007).
- (5) Kyoda, K., Baba, K., Kitano, H., Onami, S. A proof of the DBRF-MEGN method, an algorithm for deducing minimum equivalent gene networks. *Source Code Biol. Med.* 6, 12 (2011).

11) 代表研究者：岸野洋久

平成 13 年度採択 新規研究開発課題（研究開発期間：平成 13 年 10 月～16 年 9 月）

研究課題名 ゲノム進化とマッピングの階層モデルと解析アルゴリズムの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
multidivtime (分子進化速度と分岐年代の階層ベイズ推定)	http://statgen.ncsu.edu/thorne/multidivtime.html
biomc2 (ウイルスゲノム組換えの階層ベイズ推定)	http://sites.google.com/site/biomcmc/software/biomc2

(b) 論文、学会発表等

論文発表 国内 3 件、海外 35 件

- (1) *Kishino H, Thorne JL, Seo T-K, Kajitani Y (2003). Modeling of variable evolutionary rates to estimate divergence times and adaptive evolution. Invited to Proceedings of the Conference “Science of Modeling: the 30th Anniversary of Information Criterion (AIC).” Pp. 297-306.
- (2) *Kitada S and Kishino H (2004). Simultaneous detection of linkage disequilibrium and genetic differentiation of subdivided populations. *Genetics*. 167: 2003-2013.
- (3) *Kitazoe Y, Kishino H, Okabayashi T, Watabe T, Nakajima N, Okuhara Y, and Kurihara Y (2004). Multidimensional vector space representation for correlated evolution and molecular phylogeny. *Molecular Biology and Evolution*. in press.
- (4) *Seo T-K, Kishino H, and Thorne JL (2004). Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences. *Molecular Biology and Evolution*. 21: 1201-1213.

- (5) *Seo T-K, Kishino H, and Thorne JL. (2005). Incorporating gene-specific variation when inferring and evaluating optimal evolutionary tree topologies from multilocus sequence data. *Proceedings of National Academy of Sciences, USA.* 102: 4436-4441.
- (6) *Thorne JL and Kishino H (2002). Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology.* 51: 689-702.

12) 代表研究者：肥後順一

平成 13 年度採択 新規研究開発課題（研究開発期間：平成 13 年 10 月～16 年 9 月）
研究課題名 高速計算機システムによる蛋白質フォールディングの研究

(a) データベース・ツール

公開なし

(b) 論文、学会発表等

論文発表 国内 0 件、海外 17 件

- (1) * Kazuyoshi Ikeda, Oxana V. Galzitskaya, Haruki Nakamura, and Junichi Higo. β -Hairpins, α -helices, and the intermediates among the secondary structures in the energy landscape of a peptide from a distal β -hairpin of SH3 domain. *J. Comp. Chem.* 24, 310-318 (2003).
- (2) * Tsuyoshi Yokomizo, Shin Yagihara, and Junichi Higo. Rotational motions of solvent site-dipole field around a protein. *Chem. Phys. Let.* 374, 453-458 (2003).
- (3) * Kazuyoshi Ikeda and Junichi Higo. Free-energy landscape of a chameleon sequence in explicit water and its inherent α/β bifacial property. *Protein Science* 12, 2542-2548 (2003).
- (4) * Narutoshi Kamiya, Yukihiisa S. Watanabe, Satoshi Ono, Junichi Higo. AMBER-based hybrid force field for conformational sampling of polypeptides. *Chem. Phys. Let.* in press.
- (5) * Satoshi Akanuma, Hiroh Miyagawa, Kunihiro Kitamura, Akihiko Yamagishi. A detailed unfolding pathway of a $(\beta/\alpha)_8$ -barrel protein as studied by molecular dynamics simulations. *Proteins.* in press.

13) 代表研究者：矢田哲士

平成 13 年度採択 新規研究開発課題（研究開発期間：平成 13 年 10 月～16 年 9 月）

研究課題名 ヒト遺伝子の転写・発現の多様性解明を目指した基盤データベースの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
HAL	http://www.biportal.jp/genome/cgi-bin/index.cgi?org=hal

(b) 論文、学会発表等

論文発表 国内 0 件、海外 12 件

口頭発表 国内 2 件、海外 3 件

ポスター発表 国内 7 件、海外 4 件

- (1) * Sogawa, K., Numayama-Tsuruta, K., Takahashi, T., Matsushita, N., Miura, C., Nikawa, J., Gotoh, O., Kikuchi, Y., Fujii-Kuriyama, Y. A novel induction mechanism of the rat CYP1A2 gene mediated by Ahreceptor-Arnt heterodimer. *Biochem. Biophys. Res. Commun.* 318, 746-755 (2004).
- (2) * Yada, T., Totoki, Y., Takaeda, Y., Sakaki, Y., and Takagi, T. DIGIT: a novel gene finding program by combining gene-finders. *Pac. Symp. Biocomput.* 8, 375-87 (2003).
- (3) * Noguchi, H., Yada, T., and Sakaki, Y. A novel index which precisely derives protein coding regions from cross-species genome alignments. In *Proc. of Genome Informatics Workshop 2002*, pages 183-191 (2002).
- (4) * Ohshima, K., Hattori, M., Yada, T., Gojobori, T., Sakaki, Y. and Okada, N. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol.* 4, R74 (2003).
- (5) * International Human Genome Sequencing Consortium : Finishing the euchromatic sequence of the human genome, *Nature*, 431, 931-945, 2004.

平成 16 年度採択 継続研究開発課題（研究開発期間：平成 16 年 10 年～18 年 9 月）
 研究課題名 ヒトゲノムにおける広義の遺伝子発見研究

(a) データベース・ツール

データベース・ツール名	公開サイト等
HAL	http://www.biportal.jp/genome/cgi-bin/index.cgi?org=hal

(b) 論文、学会発表等

論文発表 国内 0 件、海外 10 件

口頭発表 国内 0 件、海外 2 件

ポスター発表 国内 2 件、海外 13 件

- (1) * Yamada S., Gotoh O., Yamana H., Improvement in accuracy of multiple sequence alignment using novel group-to-group sequence alignment algorithm with piecewise linear gap cost., BMC Bioinformatics, 7, 524 (2006)
- (2) * Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., Gotoh, O., Automated classification of alternative splicing and transcriptional initiation and construction of a visual database of the classified patterns, Bioinformatics, 22, 1211-1216 (2006)
- (3) * Matsui H., Sato K. and Sakakibara Y. : Pair stochastic tree adjoining grammars for aligning and predicting pseudoknot RNA structures, Bioinformatics 21, 2611-2617 (2005).
- (4) * Tabei Y., Tsuda K., Kin T. And Asai K., SCARNA: fast and accurate structural alignment of RNA sequences by matching fixed-length stem fragments., Bioinformatics 22, 1723-1729 (2006).
- (5) * Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T., Terai, G., Kusumoto, K.-I., Arima, T., Akita, O., Kashiwagi, Y., Abe, K., Gomi, K., Horiuchi, H., Kitamoto, K., Kobayashi, T., Takeuchi, M., Denning, D.W., Galagan, J.E., Nierman, W.C., Yu, J., Archer, D.B., Bennett, J.W., Bhatnagar, D., Cleveland, T.E., Fedorova, N.D., Gotoh, O., Horikawa, H., Hosoyama, H., Ichinomiya, M., Igarashi, R., Iwashita, K., Juvvadi, P.R., Kato, M., Kato, Y., Kin, T., Kokubun, A., Maeda, H., Maeyama, N., Maruyama, J.-i., Nagasaki, H., Nakajima, T., Oda, K., Okada, K., Paulsen, I., Sakamoto, K., Sawano, T., Takahashi, M., Takase, K., Terabayashi, Y., Wortman, J.R., Yamada, O., Yamagata, Y., Anazawa, H., Hata, Y., Koide, Y., Komori, T., Koyama, Y., Minetoki, T., Suharnan,

S., Tanaka, A., Isono, K., Kuhara, S., Ogasawara, N., and Kikuchi, H.
 (2005) Genome sequencing and analysis of *Aspergillus oryzae*, Nature
 438, 1157-1161

14) 代表研究者：内山郁夫

平成 17 年度採択 新規研究開発課題（研究開発期間：平成 17 年 10 月～20 年 9 月）
 研究課題名 大規模な比較ゲノム研究を展開するためのワークベンチの構築

(a) データベース・ツール

データベース・ツール名	公開サイト等
MBGD	http://mbgd.genome.ad.jp/
DomClust	http://mbgd.genome.ad.jp/domclust/
MyMBGD	http://mbgd.genome.ad.jp/MyMBGD/
CGAT	http://mbgd.genome.ad.jp/CGAT/
CoreAligner	http://mbgd.genome.ad.jp/CoreAligner/
RECOG	http://mbgd.genome.ad.jp/RECOG/

(b) 論文、学会発表等

論文発表 国内 0 件、海外 4 件

口頭発表 国内 2 件、海外 0 件

ポスター発表 国内 3 件、海外 4 件

- (1) * Uchiyama, I., Hierarchical clustering algorithm for comprehensive orthologous-domain classification in multiple genomes. Nucleic Acids Res. 34, 647-658, 2006.
- (2) * Uchiyama, I., Higuchi, T., Kobayashi, I.: CGAT: a comparative genome analysis tool for visualizing alignments in the analysis of complex evolutionary changes between closely related genomes. BMC Bioinformatics, 472, 2006.
- (3) * Uchiyama, I: MBGD: a platform for microbial comparative genomics based on the automated construction of orthologous groups. Nucleic Acids Res. 35, D343-D346, 2007.
- (4) * Uchiyama, I.: Multiple genome alignment for identifying the core structure among moderately related microbial genomes. BMC Genomics, in press.

平成 20 年度採択 継続研究開発課題（研究開発期間：平成 20 年 10 月～22 年 9 月）
研究課題名 大規模比較ゲノムネットワークベンチの実践的応用に向けた研究開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
RECOG	http://mbgd.genome.ad.jp/RECOG/
MBGD	http://mbgd.genome.ad.jp/

(b) 論文、学会発表等

論文発表 国内 4 件、海外 10 件

招待、基調講演 国内 6 件、海外 10 件

口頭発表 国内 16 件、海外 4 件

ポスター発表 国内 22 件、海外 4 件

特許 国内 0 件、国際 0 件

- (1) * Ikuo Uchiyama, Toshio Higuchi, Mikihiko Kawai, MBGD update 2010: toward a comprehensive resource for exploring microbial genome diversity. *Nucleic Acids Res*, 38, D361-D365 (2010)
- (2) * Furuta, Y., Kawai, M., Yahara, K., Takahashi, N., Handa, N., Tsuru, T., Oshima, K., Yoshida, M., Azuma, T., Hattori, M., Uchiyama, I., Kobayashi, I. Birth and death of genes linked to chromosomal inversion, *Proc. Natl. Acad. Sci. USA*, 108, 1501-1506, 2011.
- (3) * Kawai, M., Furuta, Y., Yahara, K., Tsuru, T., Oshima, K., Handa, N., Takahashi, N., Yoshida, M., Azuma, T., Hattori, M., Uchiyama, I., Kobayashi, I. Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genomes, *BMC Microbiology*, 11, 104, 2011
- (4) * Uchiyama, I., Functional inference in microbial genomics based on large-scale comparative analysis. in “Omics Approaches to Protein Function Prediction” (Daisuke Kihara ed.), pp.55-92, Springer, 2011.

15) 代表研究者：太田元規

平成 17 年度採択 新規研究開発課題（研究開発期間：平成 17 年 10 月～20 年 9 月）

研究課題名 タンパク質の構造・機能予測法の開発とヒトゲノム配列への適用

(a) データベース・ツール

データベース・ツール名	公開サイト等
EzCatDB	http://mbs.cbrc.jp/EzCatDB/
POODLE	http://mbs.cbrc.jp/poodle/
classPPI	http://pre-s.protein.osaka-u.ac.jp/~classppi
piSite	http://pisite.hgc.jp
PreBI	http://pre-s.protein.osaka-u.ac.jp/~prebi

(b) 論文、学会発表等

論文発表 国内 25 件、海外 29 件

口頭発表 国内 50 件、海外 3 件

ポスター発表 国内 42 件、海外 8 件

- (1) *Higurashi, M., Ishida, T. & Kinoshita, K. Identification of transient hub proteins and the possible structural basis for their multiple interactions, *Protein Sci*, 17, 72-78, 2008
- (2) *Kanamori, E., Murakami, Y., Tsuchiya, Y., Standley, D.M., Nakamura, H. & Kinoshita, K. Docking of protein molecular surfaces with evolutionary trace analysis, *Proteins*, 69, 832-838, 2007
- (3) *Shimizu, K., Muraoka, Y., Hirose, S., Tomii, K. & Noguchi, T. Predicting mostly disordered proteins by using structure-unknown protein data. *BMC Bioinformatics*, 8, 78, 2007
- (4) *Koike, R., Amemiya, T., Ota, M. & Kidera, A. Protein structural change upon ligand binding correlates with enzymatic reaction mechanism. *J. Mol. Biol.*, 379, 397-401, 2008

平成 20 年度採択 継続研究開発課題（研究開発期間：平成 20 年 10 月～22 年 9 月）

研究課題名 タンパク質の構造・機能・相互作用予測システムの開発と展開

(a) データベース・ツール

データベース・ツール名	公開サイト等
SAHG	http://bird.cbrc.jp/sahg/

(b) 論文、学会発表等

論文発表 国内 5 件、海外 23 件

招待、基調講演 国内 9 件、海外 0 件

口頭発表 国内 4 件、海外 1 件

ポスター発表 国内 25 件、海外 2 件

特許 国内 0 件、国際 0 件

- (1) * Motono, C. et al. SAHG, a comprehensive database of predicted structures of all human proteins. *Nucleic Acids Res.* in press, 2010.
- (2) * Nishi, H. and Ota, M., Amino acid substitutions at protein-protein interfaces that modulate the oligomeric state. *Proteins*, 78, 1563-1574, 2010.
- (3) * Kasahara, K., Kinoshita, K. and Takagi, T. Ligand binding site prediction of proteins based on known fragment-fragment interactions. *Bioinformatics*, in press, 2010.
- (4) * Koike, R., Kidera, A. and Ota, M. Alteration of oligomeric state and domain architecture is essential for functional transformation between transferase and hydrolase with the same scaffold. *Protein Sci.*, 18, 2060-2066, 2009.
- (5) * Kinoshita, K. and Obayashi, T. Multi-dimensional correlations for gene coexpression and application to the large-scale data of *Arabidopsis*. *Bioinformatics*, 25, 2677-2684, 2009.
- (6) * Kuwahara, Y., Unzai, S., Nagata T., Hiroaki, Y., Yokoyama H., Matsui I., Ikegami, T., Fujiyoshi, Y., and Hiroaki, H. Unusual thermal disassembly of the SPFH domain oligomer from *Pyrococcus horikoshii*. *Biophys J.*, 97, 2034-2043, 2009.

16) 代表研究者：川戸佳

平成 17 年度採択 新規研究開発課題（研究開発期間：平成 17 年 10 月～20 年 9 月）

研究課題名 脳スライス中で可視化した神経シナプスの自動解析

(a) データベース・ツール

データベース・ツール名	公開サイト等
Spiso-3D(XYZ)	http://glia.c.u-tokyo.ac.jp

(b) 論文、学会発表等

論文発表 国内 0 件、海外 69 件

口頭発表 国内 4 件、海外 8 件

ポスター発表 国内 63 件、海外 4 件

特許 国内 2 件、国際 1 件

- (1) *Mukai H., Hatanaka Y., Mitsuhashi K., Hojo Y., Komatsuzaki Y., Sato R., Murakami G., Kimoto T., and Kawato S. Automated analysis of spines from confocal laser microscopy images: application to the discrimination of androgen and estrogen effects on spinogenesis. *Cerebral Cortex*, 21:2704-11 (2011)
- (2) *Hojo Y, Murakami G, Mukai H, Higo S, Hatanaka Y, Ogiue-Ikeda M, Ishii H, Kimoto T, and Kawato S. Estrogen synthesis in the brain-Role in synaptic plasticity and memory. *Mol Cell Endocrinol.* 290, 31-43 (2008)
- (3) * Ishii H., Tsurugizawa T., Ogiue M., Asashima M., Mukai H., Murakami G., Hojo Y., Kimoto T. and Kawato S. Local Production of Sex Hormones and Their Modulation of Hippocampal Synaptic Plasticity. *The Neuroscientist*, 13, 323-334 (2007).
- (4) * Mukai H., Tsurugizawa T., Murakami G., Kominami S., Ishii H., Ogiue-Ikeda M., Takata N., Tanabe N., Furukawa A., Hojo Y., Ooishi Y., Morrison J.H., Janssen W.G.M., Rose J.A., Chambon P., Kato S., Izumi S., Yamazaki T., Kimoto T. and Kawato S. Rapid modulation of long-term depression and spinogenesis via synaptic estrogen receptors in hippocampal principal neurons. *J. Neurochem.*, 100, 950-967 (2007).
- (5) * Murakami G., Tsurugizawa T., Hatanaka Y., Komatsuzaki Y., Tanabe N., Mukai H., Hojo Y., Kominami S., Yamazaki T., Kimoto T. and Kawato S. Comparison between basal and apical dendritic spines in estrogen-induced rapid spinogenesis of CA1 principal neurons in the adult hippocampus. *Biochem. Biophys. Res. Commun.*, 351. 553-558 (2006).

17) 代表研究者：塩田浩平

平成 17 年度採択 新規研究開発課題（研究開発期間：平成 17 年 10 月～20 年 9 月）

研究課題名 ヒト胚の形態発生に関する三次元データベース

(a) データベース・ツール

データベース・ツール名	公開サイト等
ヒト胚 3 次元データベース	http://mrlab.frsc.tsukuba.ac.jp/human_embryos/

(b) 論文、学会発表等

論文発表 国内 6 件、海外 9 件

口頭発表 国内 10 件、海外 2 件

ポスター発表 国内 5 件、海外 6 件

特許 国内 1 件

- (1) * Otake Y, Kose K and Haishi T. A solution to the dynamic range problem in MRI using a parallel image acquisition. Concepts Magn Reson (Magn Reson Engineering), 29B: 161-167 (2006).
- (2) * 水田 忍, 番匠武蔵, 杜 楽宇, 松田哲也, 山田重人, 塩田浩平. 3 次元MR 顕微鏡画像群を対象とするヒト胚子標本データベース検索表示システムの構築. 生体医工学 44(4):665-673 (2006).
- (3) * Shiota K, Yamada S, Nakatsu-Komatsu T, Uwabe C, Kose K, Matsuda Y, Haishi T, Mizuta S, Matsuda T. Visualization of human prenatal development by magnetic resonance imaging (MRI). American Journal of Medical Genetics A, 143(24):3121-6 (2007).
- (4) * Matsuda Y, Ono S, Otake Y, Handa S, Kose K, Haishi T, Yamada S, Uwabe C, Shiota K. Imaging of a large collection of human embryos using a super-parallel MR microscope. Magnetic Resonance in Medical Sciences, 6(3):139-46, (2007).
- (5) * Iita N, Handa S, Tomiha, Kose K, Development of a compact MRI system for measuring the trabecular bone microstructure of the finger. Magnetic Resonance in Medicine, 57(2):272-277 (2007).

平成 20 年度採択 継続研究開発課題（研究開発期間：平成 20 年 10 月～22 年 9 月）

研究会題名 ヒト胚の形態発生に関する三次元データベース

(a) データベース・ツール

データベース・ツール名	公開サイト等
Kyoto Human Embryo Visualization Project	http://bird.cac.med.kyoto-u.ac.jp/search/index.html

(b) 論文、学会発表等

論文発表 国内 2 件、海外 9 件

招待、基調講演 国内 1 件、海外 1 件

口頭発表 国内 4 件、海外 0 件

ポスター発表 国内 13 件、海外 3 件

特許 国内 0 件、国際 0 件

- (1) * Morimoto N, Ogihara N, Katayama K, Shiota K. Three-dimensional ontogenetic shape changes in the human cranium during the fetal period. *J Anat* 212:627-635, 2008
- (2) * Dhanantwari P, Lee E, Krishnan A, Samtani R, Yamada S, Anderson S, Lockett E, Donofrio M, Shiota K, Leatherbury L, Lo CW. Human cardiac development in the first trimester: a high-resolution magnetic resonance imaging and episcopic fluorescence image capture atlas. *Circulation*. 28;120(4):343-351, 2009
- (3) * Yamada S, Samtani RR, Lee ES, Lockett E, Uwabe C, Shiota K, Anderson SA, Lo CW. Developmental atlas of the early first trimester human embryo. *Dev Dyn*. 2010 Jun;239(6):1585-1595, 2010
- (4) * 塩田浩平、山田重人、森本直記、荻原直道、片山一道、巨瀬勝美 「マイクロイメージング法によるヒト胚子・胎児の骨格系および中枢神経系の観察」 *小児の脳神経* 34(1): 22-26, 2009
- (5) * 塩田浩平、山田重人、土屋真衣子、巨瀬勝美 MR顕微鏡とepiscopic fluorescence image capture (EFIC)法を用いたヒト胚子のイメージング - 中枢神経系を中心に -」 *小児の脳神経*35(1): 7-10, 2010

18) 代表研究者：白井剛

平成 17 年度採択 新規研究開発課題（研究開発期間：平成 17 年 10 月～20 年 9 月）

研究課題名 実践による超分子複合体モデリングシステムの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
SIRD	http://sird.nagahama-i-bio.ac.jp/sird/
BIOMOL	http://sird.nagahama-i-bio.ac.jp/sird/ の 3. SIRD modeling - 3. 1: Make quaternary complex より Web ツールとして利用可能
EXT	http://sird.nagahama-i-bio.ac.jp/sird/ の 3. SIRD modeling - 3. 1: Make complex より Web ツールとして利用可能
CMP	http://sird.nagahama-i-bio.ac.jp/sird/ の 3. SIRD modeling - 3. 2: Assemble complex より Web ツールとして利用可能

(b) 論文、学会発表等

論文発表 国内 5 件、海外 21 件

口頭発表 国内 15 件、海外 4 件

ポスター発表 国内 55 件、海外 7 件

- (1) * T. Shirai, V.-S. Hung, K. Morinaka, T. Kobayashi, and S. Ito, Crystal structure of GH13 α -glucosidase GSJ from one of the deepest sea bacteria. *Proteins*, 73, 126-133 (2008)
- (2) * S. Kiyonari, M. Uchimura, T. Shirai & Y. Ishino, Physical and Functional Interactions between Uracil-DNA glycosylase and proliferating cell nuclear antigen from the euryarchaeon *Pyrococcus furiosus*. *J. Biol. Chem.* 283, 24185-24193 (2008)
- (3) * T. Yoshimochi, R. Fujikane, M. Kawanami, F. Matsunaga, and Y. Ishino, The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *J. Biol. Chem.* 283, 1601-1609 (2008)
- (4) * K. Mayanagi, Y. Fujiwara, T. Miyata, and K. Morikawa, Electron microscopic single particle analysis of a tetrameric RuvA/RuvB/Holliday junction DNA complex. *BBRC* 365, 273-278, (2008)
- (5) * K. Sasaki, T. Ose, N. Okamoto, K. Maenaka, T. Tanaka, H. Masai, M. Saito, T. Shirai, D. Kohda, Structural basis of the 3' end recognition of a leading strand in stalled replication forks by PriA., *The EMBO Journal*, 26, 2584-2593 (2007)

平成 20 年度採択 継続研究開発課題（研究開発期間：平成 20 年 10 月～22 年 9 月）

研究課題名 実践による超分子ネットワークモデリングシステムの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
SIRD	http://sird.nagahama-i-bio.ac.jp/sird/ .
COMPLIG	http://sird.nagahama-i-bio.ac.jp/sird/download.php?db=080731
COMPPDB	http://sird.nagahama-i-bio.ac.jp/sird/download.php?db=080731

(b) 論文、学会発表等

論文発表 国内 7 件、海外 29 件

招待、基調講演 国内 13 件、海外 2 件

口頭発表 国内 2 件、海外 0 件

ポスター発表 国内 55 件、海外 7 件

特許 国内 1 件、国際 0 件

- (1) * T. Oyama, S. Ishino, S. Fujino, H. Ogino, T. Shirai, K. Mayanagi, M. Saito, N. Nagasawa, Y. Ishino, and K. Morikawa, Architectures of archaeal GINS complexes, essential DNA replication initiation factors. *BMC Biol.* 9:28 (2011)
- (2) * K. Mayanagi, H. Nishida, S. Kiyonari, M. Saito, D. Kohda, Y. Ishino, T. Shirai, and K. Morikawa, The architecture of the DNA polymerase-PCNA-DNA ternary complex. *Proc. Natl. Acad. Sci.* 108, 1845-1849 (2011)
- (3) * H. Nishida, K. Mayanagi, S. Kiyonari, Y. Sato, T. Oyama, Y. Ishino and K. Morikawa, Structural determinant for switching between the polymerase and exonuclease modes in the PCNA-replicative DNA polymerase complex. *Proc. Natl. Acad. Sci.* 106, 20693-20698 (2009)
- (4) * T. Oyama, H. Oka, K. Mayanagi, T. Shirai, K. Matoba, R. Fujikane, Y. Ishino and K. Morikawa, Atomic structures and functional implications of archaeal RecQ-like helicase Hjm. *BMC Struct. Biol.* 9: 2 (2009)
- (5) * S. Kiyonari, S. Tahara, T. Shirai, S. Iwai, S. Ishino, and Y. Ishino Biochemical properties and BER complex formation of AP endonuclease from *Pyrococcus furiosus* *Nucleic Acids Res.* 37, 6439-6453 (2009)

- (6) * T. Shirai, Y. Watanabe, M. Lee, T. Ogawa, and K. Muramoto Structure of rhamnose-binding lectin CSL3: Unique pseudo-tetrameric architecture of a pattern recognition protein. *J. Mol. Biol.* 391, 390-403 (2009)
- (7) * K. Mayanagi, S. Kiyonari, M. Saito, T. Shirai, Y. Ishino, and K. Morikawa, Mechanism of replication machinery assembly as revealed by the DNA ligase-PCNA-DNA complex architecture. *Proc. Natl. Acad. Sci. USA.* 106, 4647-4652 (2009)
- (8) * S. Kiyonari, S. Tahara, M. Uchimura, T. Shirai, S. Ishino, and Y. Ishino Studies on base excision repair (BER) complex in *Pyrococcus furiosus*. *Biochem. Soc. Trans.* 37, 79-82 (2009)

19) 代表研究者：宮川剛

平成 17 年度採択 新規研究開発課題（研究開発期間：平成 17 年 10 月～20 年 9 月）

研究課題名 マウスを用いた脳機能表現型データベースの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
Mouse Phenotype Database	https://behav.hmro.med.kyoto-u.ac.jp/

(b) 論文、学会発表等

論文発表 国内 15 件、海外 23 件

口頭発表 国内 17 件、海外 4 件

ポスター発表 国内 16 件、海外 21 件

- (1) * Takao, K., & Miyakawa, T., Light/dark transition test for mice, *Journal of Visualized Experiments*, 2006 (1):104
- (2) * Takao, K., Yamasaki, N., and Miyakawa, T., Impact of brain-behavior phenotyping of genetically-engineered mice on research of neuropsychiatric disorders. *Neurosci Res*, 2007. 58(2): p. 124-32
- (3) * Aiba, A., Inokuchi, K., Ishida, Y., Itohara, S., Kobayashi, K., Masu, M., Mishina, M., Miyakawa, T., Mori, H., Nakao, K., Obata, Y., Sakimura, K., Shiroishi, T., Wada, K., and Yagi, T., Mouse liaison for integrative brain research. *Neurosci Res*, 2007. 58(2): p. 103-4.
- (4) * Imayoshi, I., Sakamoto, M., Ohtsuka, T., Takao, K., Miyakawa, T., Yamaguchi, M., Mori, K., Ikeda, T., Itohara, S., Kageyama, R., Roles

- of continuous neurogenesis in the structural and functional integrity of the adult forebrain. *Nature Neuroscience* 11(10): 1153-61 (2008).
- (5) * Yamasaki, N*, Maekawa, M*, Kobayashi, K*, Kajii, Y*, Maeda, J*, Soma, M*, Takao, K*, Tanda, K., Ohira, K., Toyama, T., Kanzaki, K., Fukunaga, K., Sudo, Y., Ichinose, H., Ikeda, M., Iwata, N., Ozaki, N., Suzuki, H., Higuchi, M., Suhara, T., Yuasa, S., Miyakawa, T. (*These authors are contributed equally to this work.), Alpha-CaMKII deficiency causes immature dentate gyrus, a novel candidate endophenotype of psychiatric disorders. *Molecular Brain*, 1:6 (2008).

20) 代表研究者：石井信

平成 18 年度採択 新規研究開発課題（研究開発期間：平成 18 年 11 月～21 年 9 月）
 研究課題名 予測技術を用いた生命システムの同定手法の開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
Edge Evolution Tracking	

(b) 論文、学会発表等

論文発表 国内 0 件、海外 15 件
 招待、基調講演 国内 5 件、海外 1 件
 口頭発表 国内 10 件、海外 4 件
 ポスター発表 国内 5 件、海外 2 件
 特許出願 国内 0 件、海外 1 件
 報道発表 国内 17 件、海外 0 件

- (1) * Kim, W., Matsui, T., Yamao, M., Ishibashi, M., Tamada, K., Takumi, T., Kohno, K., Oba, S., Ishii, S., Sakumura, Y., Bessho, Y. The period of the somite segmentation clock is sensitive to Notch activity. *Molecular Biology of Cell*, 22(18), 3541-3549 (2011)
- (2) * Naoki, H., Nakamuta, S., Kaibuchi, K., & Ishii, S. Flexible search for single-axon morphology during neuronal spontaneous polarization. *PLoS ONE*, 6(4):e19034, doi:10.1371/journal.pone.0019034 (2011)
- (3) * Toriyama, M., Sakumura, Y., Shimada, T., Ishii, S., & Inagaki, N. A diffusion-based neurite length-sensing mechanism involved in neuronal symmetry breaking. *Molecular Systems Biology*, 6(394),

doi:10.1038/msb.2010.51 (2010)

- (4) * Hayashi, S., Shimoda, T., Nakajima, M., Tsukada, Y., Sakumura, Y., Dale, J.K., Maroto, M., Kohno, K., Matsui, T., Bessho, Y. Sprouty4, an FGF inhibitor, displays cyclic gene expression under the control of the notch segmentation clock in the mouse PSM. PLoS ONE, 4(5):e5603 (2009)
- (5) * Tsukada, Y., Aoki, K., Nakamura, T., Sakumura, Y., Matsuda, M., Ishii, S. Quantification of Local Morphodynamics and Local GTPase Activity by Edge Evolution Tracking, PLoS Computational Biology, 4(11), e1000223 (2008)

21) 代表研究者：黒川 顕

平成 18 年度採択 新規研究開発課題（研究開発期間：平成 18 年 11 月～21 年 9 月）
研究課題名 メタゲノムオーソログ遺伝子統合解析システムの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
Human Microbiome Body Map	http://metagenomics.jp/mg/

(b) 論文、学会発表等

論文発表 国内 5 件、海外 5 件

招待、基調講演 国内 7 件、海外 1 件

口頭発表 国内 1 件、海外 0 件

ポスター発表 国内 7 件、海外 6 件

報道発表 国内 6 件、海外 0 件

- (1) * Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V. K., Srivastava, T. P., Taylor, T. D., Noguchi, H., Mori, H., Ogura, Y., Ehrich, D. S., Itoh, K., Takagi, T., Sakaki, Y., Hayashi, T., and Hattori, M. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res., 14:169-181, 2007.
- (2) * Hattori, M., and Taylor, T.D. The human intestinal microbiome: a new frontier of human biology. DNA Res., 16:1-12, 2009.
- (3) * Noguchi, H., Taniguchi, T., and Itoh, T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for

precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.*, 15:387-396, 2008.

- (4) * Goto, N., Kurokawa, K., and Yasunaga, T. Analysis of invariant sequences in 266 complete genomes. *Gene*, 401:172-180, 2007.

平成 21 年度採択 継続研究開発課題 (研究開発期間 : 平成 21 年 10 月 ~ 23 年 9 月)
研究課題名 メタゲノム統合解析システムの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
MetaGeneAnnotator	http://metagene.cb.k.u-tokyo.ac.jp/
Visualization tool for taxonomic Compositions of microbial community (VITCOMIC)	http://mg.bio.titech.ac.jp/vitcomic/
Human Meta BodyMap	http://metagenomics.jp/mg/

(b) 論文、学会発表等

論文発表 国内 2 件、海外 11 件

招待、基調講演 国内 8 件、海外 0 件

口頭発表 国内 7 件、海外 0 件

ポスター発表 国内 1 件、海外 2 件

- (1) * Arumugam M, Raes J, Pelletier E, Paslier D L, Yamada T, Mendel D R, Fernandes G R, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen H B, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal E G, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, MetaHIT Consortium, Weissenbach J, Ehrlich S D, Bork P, Enterotypes of the human gut microbiome. *Nature*, 473:174-180, 2011.
- (2) * Kuwahara T, Ogura Y, Oshima K, Kurokawa K, Ooka T, Hirakawa H, Itoh T, Nakayama-Imaohji H, Ichimura M, Itoh K, Ishifune C, Maekawa Y, Yasutomo K, Hattori M, Hayashi T, The Lifestyle of the Segmented Filamentous Bacterium: A Non-Culturable Gut-Associated Immunostimulating Microbe Inferred by Whole-Genome Sequencing. *DNA Res.*, 18:291-303, 2011.

- (3) * Mori, H., Maruyama, F., and Kurokawa, K. VITCOMIC: visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences., *BMC Bioinfo.*, 11:332, 2010.
- (4) * 森宙史, 丸山史人, 黒川 顕メタゲノムインフォマティクス, 難培養微生物研究の最新技術II, pp82-91, 2010, シーエムシー出版, 東京
- (5) * 森宙史, 丸山史人, 黒川 顕メタゲノムデータベース, メタゲノム解析技術の最前線, pp42-53, 2010, シーエムシー出版, 東京

22) 代表研究者：榊原康文

平成 18 年度採択 新規研究開発課題（研究開発期間：平成 18 年 11 月～21 年 9 月）
 研究課題名 タンパク質化合物相互作用の網羅的予測手法とデータベースの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
COPICAT	http://copicat.dna.bio.keio.ac.jp/

(b) 論文、学会発表等

論文発表 国内 3 件、海外 7 件
 招待、基調講演 国内 0 件、海外 1 件
 口頭発表 国内 4 件、海外 2 件
 ポスター発表 国内 1 件、海外 0 件
 特許出願 国内 1 件、海外 1 件

- (1) * N. Nagamine, T. Shirakawa, Y. Minato, K. Torii, H. Kobayashi, M. Imoto, Y. Sakakibara : Integrating Statistical Predictions and Experimental Verifications for Enhancing Protein-Chemical Interaction Predictions in Virtual Screening, *PLoS Computational Biology*, 5(6), e1000397, 2009.
- (2) * N. Nagamine and Y. Sakakibara : Statistical prediction of protein-chemical interactions based on chemical structure and mass spectrometry data, *Bioinformatics*, 23(15), 2004-2012, 2007.
- (3) * K. Sasaki, N Nagamine, and Y. Sakakibara : Support vector machine prediction of N- and O-glycosylation sites using whole sequence information and subcellular localization, *IPSJ Transactions on Bioinformatics*, 2, 25-35, 2009.

平成 21 年度採択 継続研究開発課題（研究開発期間：平成 21 年 10 月～23 年 9 月）
研究課題名 タンパク質化合物相互作用の網羅的予測手法とデータベースの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
COPICAT	http://copicat.dna.bio.keio.ac.jp/

(b) 論文、学会発表等

論文発表 国内 2 件、海外 8 件

招待、基調講演 国内 0 件、海外 1 件

口頭発表 国内 1 件、海外 0 件

ポスター発表 国内 1 件、海外 0 件

- (1) *H. Jo, K. Yugi, S. Ogawa, Y. Suzuki, Y. Sakakibara. Molecular basis of chemical chaperone effects of N-octyl- β -valienamine on human β -glucosidase in low/neutral pH conditions. J Proteomics Bioinform, 3(4), 104-112, 2010.
- (2) *T. Kawamura, T. Fujimaki, N. Hamanaka, K. Torii, H. Kobayashi, Y. Takahashi, M. Igarashi, N. Kinoshita, Y. Nishimura, E. Tashiro and M. Imoto. Isolation and structure elucidation of a novel androgen antagonist, arabilin, produced by Streptomyces sp. MK756-CF1. J Antibiot, 63, 601-605, 2010.
- (3) *K. Sato, Y. Saito, Y. Sakakibara. Gradient-based optimization of hyperparameters for base-pairing profile local alignment kernels. Genome Inform, 23, 128-138, 2009.
- (4) *T. Kawamura, T. Fujimaki, N. Hamanaka, K. Torii, H. Kobayashi, Y. Takahashi, M. Igarashi, N. Kinoshita, Y. Nishimura, E. Tashiro and M. Imoto (2010). Isolation and structure elucidation of a novel androgen antagonist, arabilin, produced by Streptomyces sp. MK756-CF1. J. Antibiot., 63, 601-605.

23) 代表研究者：清水謙多郎

平成 18 年度採択 新規研究開発課題（研究開発期間：平成 18 年 11 月～21 年 9 月）

研究課題名 高精度タンパク質間相互作用予測システムの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
タンパク質-リガンド結合状態／非結合状態ペアデータベース	http://www.bi.a.u-tokyo.ac.jp/services/buddy/current/
糖鎖結合タンパク質予測システム	http://bolero.bi.a.u-tokyo.ac.jp:8201/Lectin-Predictor/
GPI アンカー型タンパク質予測システム	http://bolero.bi.a.u-tokyo.ac.jp:8201/GPI-Predictor/
タンパク質-タンパク質ドッキング予測システム	公開していない

(b) 論文、学会発表等

論文発表 国内 2 件、海外 21 件

招待、基調講演 国内 8 件、海外 0 件

口頭発表 国内 10 件、海外 0 件

ポスター発表 国内 33 件、海外 0 件

- (1) * Y. Sakakibara, T. Hachiya, M. Uchida, N. Nagamine, Y. Sugawara, M. Yokota, M. Nakamura, K. Popendorf, T. Komori, and K. Sato :
COPICAT: A software system for predicting interactions between proteins and chemical compounds. *Bioinformatics*, accepted for publication, 2012.
- (2) * S. Yamasaki, S. Nakamura, T. Terada, K. Shimizu: Mechanism of the difference in the binding affinity of E. coli tRNA^{Gln} to glutaminyl-tRNA synthetase caused by non-interface nucleotides in variable loop, *Biophysical Journal*, 92, 192-200 (2007).
- (3) * M. Morita, S. Nakamura, K. Shimizu: Highly accurate method for ligand-binding site prediction in unbound state (apo) protein structures, *PROTEINS: Structure, Function, and Bioinformatics*, 73, 2, 468-479, (2008).
- (4) * R. Ishitani, T. Terada, K. Shimizu: Refinement of comparative models of protein structure by using multicanonical molecular dynamics simulations, *Molecular Simulation*, 34, 327-336, (2008).
- (5) * M. Kakuta, S. Nakamura, K. Shimizu: Prediction of protein-protein interaction sites using only sequence information and using both

sequence and structural information, IPSJ Transactions on Bioinformatics, 49, 25-35 (2008).

- (6) * S. Yamasaki, T. Terada, K. Shimizu, H. Kono, A. Sarai: A Generalized Conformational Energy Function of DNA Derived from Molecular Dynamics Simulations, Nucleic Acids Research, accepted.

平成 21 年度採択 継続研究開発課題 (研究開発期間 : 平成 21 年 10 月~23 年 9 月)
研究課題名 高精度タンパク質間相互作用予測システムの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
タンパク質-リガンド結合状態/非結合状態ペアデータベース	http://www.bi.a.u-tokyo.ac.jp/services/buddy/current/
糖鎖結合タンパク質予測システム	http://bolero.bi.a.u-tokyo.ac.jp:8201/Lectin-Predictor/
GPI アンカー型タンパク質予測システム	http://bolero.bi.a.u-tokyo.ac.jp:8201/GPI-Predictor/

(b) 論文、学会発表等

- (1) * W. Cao, J. Maruyama, K. Kitamoto, K. Sumikoshi, T. Terada, S. Nakamura, K. Shimizu. : Using a new GPI-anchored-protein identification system to mine the protein databases of *Aspergillus fumigatus*, *Aspergillus nidulans*, and *Aspergillus oryzae*, J. Gen. Appl. Microbiol. 55, 5, 381-393 (2009).
- (2) *R. Jono, Y. Watanabe, K. Shimizu, T. Terada: Multicanonical ab initio QM/MM Molecular Dynamics Simulation of a Peptide in an Aqueous Environment, J. Comput. Chem., 31, 6, 1168-1175 (2009).
- (3) *S. Yamasaki, T. Terada, K. Shimizu, H. Kono, A. Sarai: A Generalized Conformational Energy Function of DNA Derived from Molecular Dynamics Simulations, Nucleic Acids Res., 37, e135 (2009).
- (4) * S. Someya, M. Kakuta, M. Morita, K. Sumikoshi, W. Cao, Z. Ge, O. Hirose, S. Nakamura, T. Terada, K. Shimizu: Prediction of carbohydrate-binding proteins from sequences using support vector machines, Advances in Bioinformatics 2010, 289301 (2010).
- (5) *O. Hirose, K. Shimizu: A self-organizing state space approach to infer time-varying causalities between regulatory proteins. Proceedings of International Conference on Information Technology in Bio- and Medical Informatics, Lecture Notes in Computer Science, 6266, 158-171 (2010).

24) 代表研究者：中井謙太

平成 18 年度採択 新規研究開発課題（研究開発期間：平成 18 年 11 月～21 年 9 月）

研究課題名 転写制御領域の構築原理解明

(a) データベース・ツール

データベース・ツール名	公開サイト等
Melina II	http://mellina.hgc.jp/
DBTBS(rel.5)	http://dbtbs.hgc.jp/
DBTGR(ver.2.0)	http://dbtgr.hgc.jp/
DBTSS(Rel.7)	http://dbtss.hgc.jp/

(b) 論文、学会発表等

論文発表 国内 0 件、海外 19 件

招待、基調講演 国内 0 件、海外 3 件

口頭発表 国内 9 件、海外 6 件

ポスター発表 国内 22 件、海外 4 件

- (1) * Tsuritani, K., Irie, T., Yamashita, R., Wakaguri, H., Kanai, A., Mizushima-Sugano, J., Sugano, S., Nakai, K., and Suzuki, Y., Distinct class of putative "non-conserved" promoters in humans: comparative studies of alternative promoters of human and mouse genes, *Genome Res.*, 17(7), 1005-1014 (2007)
- (2) * Kohji Okamura and Kenta Nakai, Retrotransposition as a source of new promoters, *Mol. Biol. Evol.*, 25(6):1231-1238, 2008
- (3) * Alexis Vandenbon and Kenta Nakai, Using simple rules on presence and positioning of motifs for promoter structure modeling and tissue-specific expression prediction, *Proceedings of the 19th International Conference (Genome Informatics Series Vol. 21)*, 189-199, 2008
- (4) * Alexis Vandenbon and Kenta Nakai, Modeling tissue-specific structural patterns in human and mouse promoters, *Nucl. Acids Res.*, in press.

25) 代表研究者：長野希美

平成 19 年度採択 新規研究開発課題（研究開発期間：平成 19 年 10 月～22 年 9 月）

研究課題名 酵素反応分類に基づく酵素反応予測システムの開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
EzCat-BLAST	現在（2010 年 12 月時点）は限定公開であるが、近日一般公開する予定。
EzMetAct	現在（2010 年 12 月時点）は未公開であるが、近日一般公開する予定。
Overview of EzCatDB	http://mbs.cbrc.jp/EzCatDB/overview/index.html

(b) 論文、学会発表等

論文発表 国内 2 件、海外 7 件

口頭発表 国内 3 件、海外 1 件

ポスター発表 国内 13 件、海外 2 件

特許 国内 0 件、国際 0 件

- (1) * T. Kato, N. Nagano: Metric Learning for Enzyme Active-Site Search. *Bioinformatics*, 26, 2698-2704 (2010).
- (2) * T. Kato, K. Suwa, N. Nagano: Parametric Templates: A New Enzyme Active-Site Prediction Algorithm. In conjunction with the 10th IEEE International Conference on Data Mining (ICDM 2010) Biological Data Mining and its Applications in Healthcare, accepted.
- (3) * T. Kato, N. Nagano: Discriminative Structural Approaches for Enzyme Active-Site Prediction. *BMC Bioinformatics*, presented at The Ninth Asia Pacific Bioinformatics Conference (APBC2011), accepted.
- (4) * T. Kato, H. Kashima, M. Sugiyama, K. Asai: Conic Programming for Multi-Task Learning, *IEEE Transactions on Knowledge and Data Engineering*, 22(7), 957-968 (2010).

26) 代表研究者：馳澤盛一郎

平成 19 年度採択 新規研究開発課題（研究開発期間：平成 19 年 10 月～22 年 9 月）

研究課題名 進化型計算と自己組織化による適応的画像分類法の開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
KBI ImageJ Plugins, revision 882b (for Scala 2.8)	http://hasezawa.ib.k.u-tokyo.ac.jp/zip/Kbi/ImageJKbiPlugins
Modrec, revision 98	http://hasezawa.ib.k.u-tokyo.ac.jp/zip/Kbi/ModrecProj
opencv_scala, revision 16	http://hasezawa.ib.k.u-tokyo.ac.jp/zip/Kbi/OpencvScala
LIPS	http://hasezawa.ib.k.u-tokyo.ac.jp/lips/

(b) 論文、学会発表等

論文発表 国内 6 件、海外 35 件

招待、基調講演 国内 4 件、海外 2 件

口頭発表 国内 40 件、海外 4 件

ポスター発表 国内 49 件、海外 10 件

特許 国内 1 件、国際 1 件

- (1) * Inoue K, Sato T, Kitamura H, Ito M, Tsunoda Y, Hirayama A, Kurosawa H, Tanaka T, Fukushi M, Moriyama N and Fujii H (2008) Diagnosis supporting algorithm for lymph node metastases from colorectal carcinoma on (18)F-FDG PET/CT. *Ann Nucl Med* 22: 41-8.
- (2) * Higaki T, Kutsuna N, Sano T, Kondo N, Hasezawa S (2010) Quantification and cluster analysis of actin cytoskeletal structures in plant cells: role of actin bundling in stomatal movement during diurnal cycles in Arabidopsis guard cells. *Plant J* 61: 156-165.
- (3) * Nagano AJ, Maekawa A, Nakano RT, Miyahara M, Higaki T, Kutsuna N, Hasezawa S, Hara-Nishimura I (2009) Quantitative Analysis of ER Body Morphology in an Arabidopsis Mutant. *Plant Cell Physiol* 50: 2015-2022.
- (4) * Gambe AE, Matsunaga S, Takata H, Ono-Maniwa R, Baba A, Uchiyama S, Fukui K. (2009) A nucleolar protein RRS1 contributes to chromosome congression. *FEBS Lett.* 583(12):1951-6.
- (5) * Ueda, H., Yokota, E., Kutsuna, N., Shimada, T., Shimmen, T., Hasezawa, S., Dolja, V.V. and Hara-Nishimura, I. (2010) Myosin-dependent endoplasmic reticulum motility and F-actin organization in plant cells. *Proc. Natl. Acad. Sci. USA* 107: 6894-6899.

27) 代表研究者：馬見塚拓

平成 19 年度採択 新規研究開発課題（研究開発期間：平成 19 年 10 月～22 年 9 月）

研究課題名 生命科学上の非構造化データの統合マイニング

(a) データベース・ツール

データベース・ツール名	公開サイト等
FTGI (Fast finding Three-way Gene Interactions)	http://www.bic.kyoto-u.ac.jp/pathway/kayano/bioinfo_three-way.html
PathRanker	http://www.bic.kyoto-u.ac.jp/pathway/timhancock
MetaMHC	http://www.biokdd.fudan.edu.cn/Service/MetaMHC.html
ROS-DET (Robust Detector of Switching Mechanisms)	http://www.bic.kyoto-u.ac.jp/pathway/kayano/ros-det.htm
GRASP	http://www.bic.kyoto-u.ac.jp/pathway/grasp

(b) 論文、学会発表等

論文発表 国内 1 件、海外 25 件

招待、基調講演 国内 4 件、海外 9 件

口頭発表 国内 6 件、海外 16 件

ポスター発表 国内 17 件、海外 0 件

特許 国内 0 件、国際 0 件

- (1) * Hashimoto, K., Takigawa, I., Shiga, M., Kanehisa, M. and Mamitsuka, H., Mining Significant Tree Patterns in Carbohydrate Sugar Chains. *Bioinformatics*, 24 (16) (Proceedings of the Seventh European Conference on Computational Biology (ECCB 2008), Cagliari, Sardinia-Italy, September, 2008), i167-i173, 2008.
- (2) * Shiga, M., Takigawa, I. and Mamitsuka, H., A Spectral Approach to Clustering Numerical Vectors as Nodes in a Network. *Pattern Recognition*, 44 (2), 236-251, 2011.
- (3) * Takigawa, I. and Mamitsuka, H., Efficiently Mining δ -Tolerance Closed Frequent Subgraphs. To appear in *Machine Learning*, 2011. doi:10.1007/s10994-010-5215-6
- (4) * Hancock, T., Takigawa, I. and Mamitsuka, H., Mining Metabolic Pathways through Gene Expression. *Bioinformatics*, 26 (17), 2128-2135,

2010.

- (5) * Hancock, T. and Mamitsuka, H., Boosted Optimization for Network Classification. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010) (JMLR: Workshop and Conference Proceedings, Vol. 9), pp. 305-312, Sardinia, Italy, May 2010, MIT Press.
- (6) * Kayano, M., Takigawa, I., Shiga, M., Tsuda, K. and Mamitsuka, H., ROS-DET: Robust Detector of Switching Mechanisms in Gene Expression. Nucleic Acids Research, 39 (11), e74, 2011.

28) 代表研究者：水口賢司

平成 19 年度採択 新規研究開発課題（研究開発期間：平成 19 年 10 月～22 年 9 月）
研究課題名 ダイナミクスを考慮した膜蛋白質の構造モデリング法の開発

(a) データベース・ツール

データベース・ツール名	公開サイト等
HTM-ONE	http://tardis.nibio.go.jp/netasa/htmone/
MultiCoMP	http://multicomp.nibio.go.jp/

(b) 論文、学会発表等

論文発表 国内 6 件、海外 15 件

招待、基調講演 国内 22 件、海外 14 件

口頭発表 国内 7 件、海外 1 件

ポスター発表 国内 18 件、海外 4 件

特許 国内 0 件、国際 0 件

- (1) * Tsukazaki, T. †, Mori, H. †, Fukai, S., Ishitani, R., Mori, T., Dohmae, N., Perederina, A., Sugita, Y., Vassylyev, D. G., Ito, K. and Nureki O. Conformational transition of Sec machinery inferred from bacterial SecYE structures. Nature, 455, 988-991 (2008)
† These authors contributed equally to this work.
- (2) * Kokubo, H. and Okamoto, Y. Analysis of helix-helix interactions of bacteriorhodopsin by replica-exchange simulations, Biophysical Journal, 96, 765-776 (2009)
- (3) * Mori, T., Ishitani, R., Tsukazaki, T., Nureki, O. and Sugita, Y. Molecular mechanisms underlying the early stage of protein

translocation through the Sec translocon, *Biochemistry*, 49, 945-950 (2010)

- (4) * Ahmad, S., Singh, Y.H., Paudel, Y., Mori, T., Sugita, Y. and Mizuguchi, K. Integrated prediction of one-dimensional structural features and their relationships with conformational flexibility in helical membrane proteins, *BMC Bioinformatics*, 11:533 (2010)
- (5) * Morita, M., Katta, A. V. S. K. Mohan, Ahmad, S., Mori, T., Sugita, Y., Mizuguchi, K. Lipid recognition propensities of amino acids in membrane proteins from atomic resolution data, *BMC Biophys.*, 4, 21 (2011)
- (6) * Mori, T., Ogushi, F., Sugita, Y. Analysis of lipid surface area in protein-membrane systems combining voronoi tessellation and monte carlo integration methods, *J. Comput. Chem.*, 33, 286-293 (2012)