ゲノム生物学バックボーンデータベースの構築提供

国立遺伝学研究所　生命情報・DDBJ 研究センター
菅原　秀明

Development and diffusion of backbone databases for genomics

Hideaki Sugawara
Center for Information Biology and DNA Data Bank of Japan（DDBJ）
National Institute of Genetics

The International Nucleotide Sequence Databases（INSD）is the archive of all the sequence data in the public domain and also provides data services to research communities. What kinds of services are required by the user, then? Quality and coverage of the data are essential to the services. To improve the quality of annotation, we proposed the development of *Open Annotation SYStem*（OASYS）and expansion of *Genes TO Proteins*（GTOP）database. We applied OASYS and GTOP to the reevaluation of the annotation of microbial complete genome sequences and identified new ORFs. From the view point of the coverage, we explored gene expression data that is closely related with gene sequences. We have developed *MicroArray Gene expression DataBase*（MADB）and *Bio-Simulated Database*（BSD）to capture, evaluate, store, diffuse and analyze the gene expression data. Each system and their aggregation are now searchable at http://www.jst-bird.nig.ac.jp/.

1. はじめに

2006年の Nucleic Acids Research のデータベース特集号には、858の分子生物学データベースが収録されている[1]。国際塩基配列データベース（International Nucleotide Sequence Databases（INSD））はその中でも、ゲノム生物学のバックボーンとなるべきデータベースである。

INSD は、国際共同事業である International Nucleotide Sequence Database Collaboration（INSDC）[2]が構築提供する国際塩基配列データベースであり、1980年代から遺伝子の観点に立った生物学研究の基盤として機能してきた。つまり、 INSDC は、

- 実験により決定された塩基配列とその生物学的意味（アノテーション）を受付け、
- 個々のデータにアクセション番号を付与し
- データ登録者にクレジットを与え、
- 生物種や研究目的などを問わず全データを共通の形式で広く一般に公開

してきたのである。その後、1990 年代後半から急速に増加してきた各生物種のゲノムデータや大規模な cDNA データの蓄積を受けて、今度はゲノムの観点に立った生物学研究のバックボーンとして機能することが、INSDC に期待されるに至った。

ゲノム生物学バックボーンデータベースへの期待に応えるため、DDBJ[3]は INSDC の一員として（図1）、INSDC の国際実務者会議ならびに国際諮問委員会と協議しながら、データベースの質の問題に取

り組んできた。第1に、個々のアノテーションの質の向上である。追加更新が初期登録者に限られているアノテーションが徐々に陳腐かつ空疎になり、また、INSD が雑多な質のデータが混在するレポジトリーに変質してきている問題への対応である。第2に、データの多角化への対応である。新しい測定技術の進歩によって産み出される新しい型のデータへの対応である。特に、INSD の遺伝子およびゲノム配列データと直接対応関係にある遺伝子発現データベースへの展開があげられる。

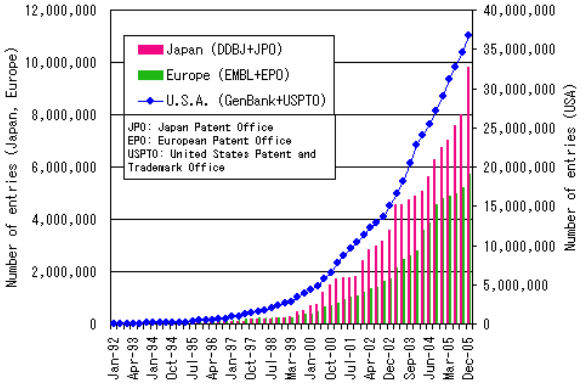

図 1. 登録エントリー件数からみた日米欧それぞれの INSD への貢献

米国の寄与を青い折れ線グラフで、日欧の寄与をそれぞれ赤と緑の棒グラフで示した。日欧の貢献はこのところ毎年それぞれ 15%前後である。2005 年 12 月のリリースでは、DDBJ 由来のデータが約 19%を占めた。

これら品質向上と多角化を目的として、2001年から始まった JST BIRD 事業において DDBJ は、

- Open Annotation SYStem（OASYS）によって第3者のアノテーションを情報資源として活用し、Genes TO Proteins（GTOP）によってゲノム配列からのタンパク質産物予測結果に基づいてアノテーションを評価し
- MicroArray expression DataBase（MADB）と Bio-Simulated Database（BSD）によって、遺伝子発現情報の登録・査定・蓄積・公開さらに比較解析を実現すること

を目指した。

2. 研究開発の成果

2.1 アノテーション品質向上への取組み

<u>OASYSの実証</u>

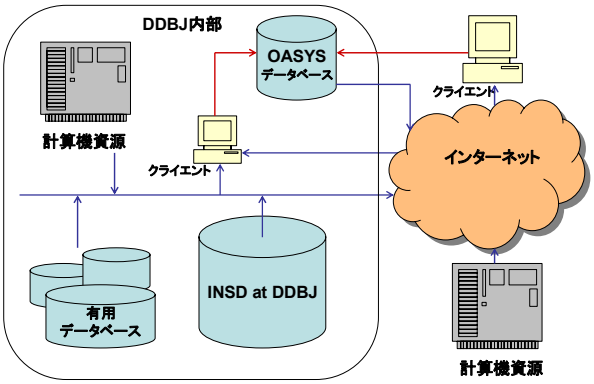

図2. OASYS の概念図

OASYS の概念図を図2に示す。図の青い矢印が示すように、INSD から対象エントリーをダウンロードし、DDBJ 内外の計算機資源とデータベースを駆使して再解析し、その結果を赤い矢印が示すように、登録者以外の第3者によるアノテーションとして DDBJ 内の OASYS データベースに登録可能とする。OASYS データベースから INSD の元データへリンクを設けるが、第3者アノテーションと元データが混同されないように留意する。この第3者いいかえるとコミュニティーによるアノテーションの概念が、JST BIRD

事業の初年度である2001年に行った大規模調査で受け入れられたため、我々は具体に OASYS システム開発を始めた。その一方で、INSDC は同年の国際実務者会議で米国 NCBI からの提案を認めて、INSD に Third Party Annotation（TPA）の枠組みを新設することとなった。OASYS は計算機解析の結果も受付けるという想定であったが、TPA は新たな実験と論文発表を登録の必要条件とした。

OASYS の開発は、INSD から公表された微生物ゲノムデータの再評価を例題として、テストを繰り返しながら進めた。以下では、OASYS を使ったこの例題プロジェクトを Gene Trek in Prokaryote Space（GTPS）と呼ぶ。

GTPS は2003年度に、2003年7月までに INSD で公開された124菌株の微生物完全長ゲノムデータを対象にして、初めて実施した。そのワークフローの概要を図3に示す。INSD のエントリー単位ではなくゲノム単位で微生物ゲノムデータを網羅した Genome Information Broker（GIB）[4]からゲノム配列データを抽出し、共通のプロトコールで網羅的に解析し、約130万の coding sequences（CDS）を推定した。次に、CDS をアミノ酸に翻訳後、相同性検索とモチーフ検索を行い、その結果を総合的に判断して絞り込んだ848,383件を対象として、CDS としての確実性の観点からランク付けをした（表1）。

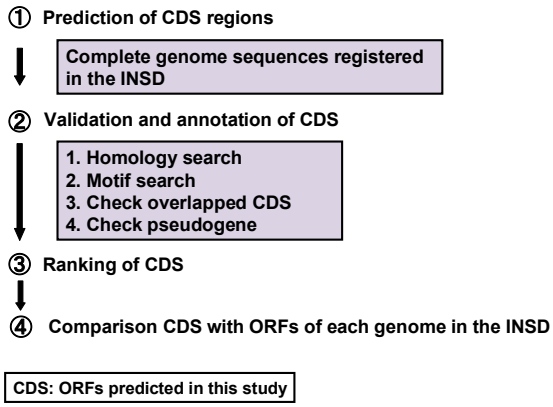

図3. GTPS のワークフロー

CDS(2003)	848,383
Rank	Fraction
A	283,247 (33.4%)
B	7,208 (0.8%)
C	4,680 (0.6%)
D	79,779 (9.4%)
Subtotal	374,914
E	6,788 (0.8%)
X	466,681 (55.0%)
Total ORFs in INSD:	362,543

表1. 微生物ゲノム配列から予測した CDS のランキング（2003 年版）

A からDへ向かうほど機能に関する手がかりが減少する。ランク X は、機能の手がかりが全くないもの

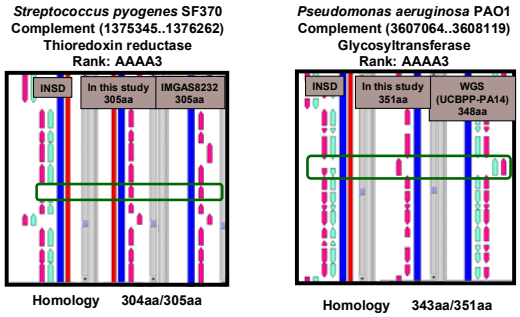

図4. GTPS で予測した新規 CDS の裏づけ

図の中の AAAA3 はランク A をさらに細分化したクラスであるが、CDS の確実性が非常に高い。ともに、INSD には存在していなかったあるいは不正確であった ORF が GTPS で特定されて、その後、実験データによって実証された例である。

この結果、機能が明確なランクAから機能について何らかの手がかりがあるランクDまでの CDS が374,914件となった。これに対して、INSD に登録されていた open reading frame（ORF）は362,543件であった。この差は GTPS によって新たに発見された ORF である可能性がある。事実、2003年版の GTPS 解析が終了後に INSD に登録された実験由来のデータと一致する事例が出てきた（図4）。また、GTPS の成果は、大腸菌ゲノムの国際協調による再アノテーションにも生かされた[5]。

微生物ゲノムの網羅的比較には膨大な計算処理が必要であったが、GRID 環境[6]や大規模な PC クラスターによって2003年、2004年そして2005年と増加し続ける微生物ゲノムアノテーションの再検証を実施することができた。2003年度の実績に基づいてDDBJ のアノテーターのノウハウを機械処理に移行しつつ、184菌株のゲノムを対象とした2004年版の GTPS を実施しその結果を公開した(図5)。さらに、2005年後半に303菌株の2005年版の解析にも着手し、現在その結果の公開準備を進めている。

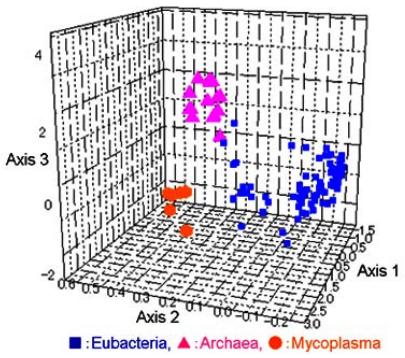

図5. 2004 年版 GTPS の提供サイトのトップページ

解析対象にした菌株を、*Escherichia coli* K12 株の遺伝子に対するホモログをもとに主成分分析してトップ頁に表示

GTOPによる評価と解析

GTOP では、微生物を含む全ゲノムの塩基配列が決定された生物種を対象に、計算機処理によってタンパク質の立体構造予測、ファミリー分類、機能モチーフ部位予測等を行った。本プロジェクトではまずGTOP が参照する多数のデータベースの自動更新も含めて GTOP 解析の自動化を進めた。特に、微生物ゲノムの場合は DDBJ が構築し OASYS でも利用した GIB と連携させることによって更新作業を大幅に簡素化することができた。これによって、GTOP 解析を4ヵ月ごとに実施することが可能となった。

さて、GTOP によって立体構造予測が可能であった ORF の割合の年変化の事例を図6に示した。年とともに予測率があがっているが、これは、主要な参照データベースである Protein Data Bank (PDB)の登録件数が増加したことに起因している。すなわち図6は、参照データベースの成長とともに、解析を更新し続けていく必要があることを示している。このことは GTOP に限らず2次データベースを公開する場合に十分に認識しておかねばならない事実である。

図6からは、ホモロジー検索による立体構造予測の解析に BLAST を用いた場合に対して PSI BLAST を用いると予測感度が大きく上昇することも分かる(GTOP は現在 Reverse PSI BLAST を使っている)。予測感度については、平成16年度にさらに感度が高い隠れマルコフ法(HMM)を導入した。しかし、解析に従来の10倍以上の計算時間を要しており、HMM 解析によって増加し続けるゲノムデータの検証を続けていくためには計算機資源の継続的増強が必要である。

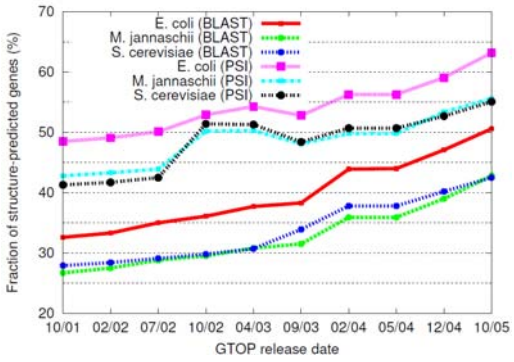

図6. GTOP による立体構造予測可能率の年変化

真正細菌、アーケアおよび真核微生物の登録 ORF のうち GTOP により立体構造予測が可能であった ORF の割合を示す。明らかに Blast よりも PSI-Blast が感度が高いこと、および、年毎に予測可能率が上がっている。

平成18年2月現在、GTOP では真核生物50、古細菌21、真正細菌203、ファージ172の合計446種の解析結果を検索可能なデータベースとして公開している。GTOP 解析およびその応用における成果を以下に示す:

- アノテーションの問題点の検出:立体構造予測率をゲノム間で比較すると異常に予測率が低いゲノムが存在することを見出した。一方で、OASYSを使ってGTPSが比較的信頼性が高いと判断したCDSを対象にすると、異常に低い予測率を持つゲノムの予測率も、他のゲノムと同程度になり、GTPSの判定の価値を裏付けることができた[7]。
- 予測されたORFが偽遺伝子か否かの判定に応用可能なことを示した[8]。
- 好熱菌・好塩菌タンパク質の組成の特徴が立体構造上タンパク質の表面に露出したアミノ酸に由来していることを明らかにした[9-11]。
- GTOPで予測した構造ドメインの組み合わせから系統樹を推定する手法を開発した。
- ヒトcDNA配列に見られる選択的スプライシングの解析を行いドメイン表面に挿入配列をもつバリアントの存在を見出した[12]。
- GTOP解析結果のグラフィック表示から、PDBやSCOPにヒットする構造ドメインと長大なdisorder領域(ループ領域)とが混在する場合があることを見出した(図7)[13]。この領域がコードするタンパク質は全てリン酸化キナーゼをはじめとする細胞内のシグナル伝達系や制御系に関与するものであり、このループ部分が生体内で何らかの機能的役割を担っていることを示唆する。

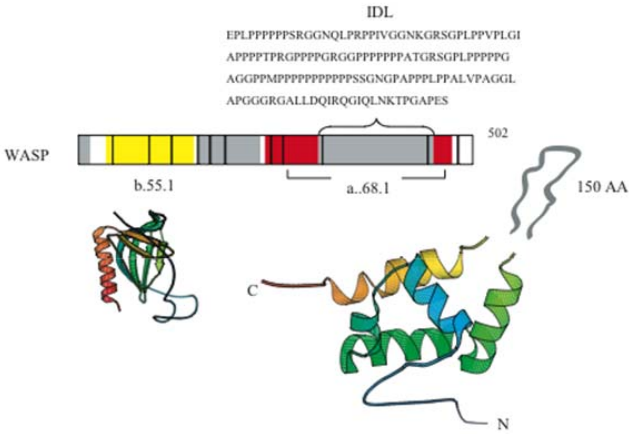

図7. 構造ドメインの途中に挿入ループ(150 残基長)をもつタンパク質の例

ヒト由来タンパク質 WASP は N 端部の通常のドメイン(黄色)の他に C 端側にもう1つのドメイン(赤)をもつ。赤いドメインは中央部に挿入配列(IDL と表記)をもち、IDL は図のようなループを形成する。灰色の領域は disorder と予測された領域であり、挿入ループも disorder 配列からなることがわかる。配列中の縦線はエキソン境界を示す。

2.2 アノテーション多角化への取組み

遺伝子発現情報データベース MADB の構築

1995年の Pat Brown の報告[14]以来、マイクロアレイ技術は世界中に広まり、今や多くの国際雑誌にこの技術を用いた論文が多数報告されている。例えば、2006年2月現在で、PubMed 文献データベースをキーワード microarray で検索すると、13,321の論文がリストアップされる。この技術の用途は、生物学の基礎研究に留まらず、病名の迅速かつ客観的特定や育種などの応用的な分野でも盛んに用いられている。マイクロアレイ技術はこのように多用される一方で、その結果が多くの実験条件要因に依存するという問題が明らかになってきた。

この問題を具体的に検討し解決へ向けて草の根運動を開始したのが Microarray Gene Expression Data (MGED)グループである。このグループは1999年英国の European Bioinformatics Institute (EBI)が中心となり、当班員を含む世界の有志でもって始められた。この活動の成果として、Minimum Information About Microarray Experiments (MIAME) [15] がある。MIAME の骨子は、世界中

の研究室からマイクロアレイ実験によって産生されたデータを世界の研究者に提供し共有することによって、関連する分野の研究をさらに発展させる、ということである。このため、研究者からデータを提供してもらうための、データ項目や形式を制定した。この項目数を最小に押さえるということで、Minimum と冠したのである。また、この運動を世界に広めるため、少なくとも年に1回MGED国際会議を開いている。この第5回目の会議は、2002年にJST BIRD事業の当班が中心となって東京で開催された。2005年には第8回会議がノルウエーで開かれている。世界の関連研究者や技術者の賛同の輪が広まり、会議への参加数も500名以上に増加している。このように、最初は草の根運動だったが、2002年から正式に MGED Society（http://www.mged.org/）として国際学会の仲間入りしている。

MGED Society の活動は現在多岐にわたってきているが、その主要な一つに国際公共データバンクの創設がある。このことは、上記のマイクロアレイデータを世界的に共有するという趣旨の実現に外ならない。このデータバンクは EBI の ArrayExpress、米国 National Center for Biotechnology Information（NCBI）の GEO、そして当班が国立遺伝学研究所、生命情報・DDBJ 研究センターで開発公開している Center for Information Biology gene EXpression database（CIBEX）[16-18]である。MGED Society はこれら3データバンクの活動を促進するため、Nature や Science など世界の関連主要国際誌の編集者との協議により、これらの学術誌にマイクロアレイ実験に基づいた原稿を投稿するときは、先ずそのデータを上記3バンクの一つに登録するという義務を課すことにした。INSDC の登録制度を採用したことになる。

CIBEX は上記の国際協議により、開発を始め現在の活動に及んでいる。当班では、我が国の関連誌にも上記の協力をお願いしている。現在、Genes and Cells や Gene and Genetic System など数種の学術誌の協力を得ている。当班では、MIAME の内容の日本語版を作り、学術誌や国内学会などを通じて CIBEX の活動を紹介し、データの提供をお願いしてきている。データは徐々に集まってきているが、DNA 配列データと異なり、スポット情報以外に、サンプル、用いたアレイ、ハイブリッド実験、データ変換や規準化などの項目に記述することになっているので、データ登録をしようとしている研究者の一部から、データ記述が複雑であるという意見も出ている。このような意見に対応して、必要ならばデータ登録者への援助も含めてデータを収集していく。

図8. 遺伝子発現情報データベース CIBEX の Web サイト

遺伝子発現情報の登録窓口とともに CIBEX データの検索と解析の機能も準備した。

国際公共バンクとしての CIBEX の主要な使命の一つとして、欧米の2バンクと相互にデータ交換を行い、日米欧の3バンクで同質・同量のデータを提供することが挙げられる。当班では、既に XML 形式で欧米のバンクとデータ交換を行うツールを作成し、このテストを行う段階にきている。従って、近い将来には欧米のバンクとのデータ交換が実現される予定である。CIBEX は、マイクロアレイデータの解析や表示ツールを備えているが（図8）、BSD 班と連携を計っているので、この連携の中で整備している。 また、アレイデータの信頼性についての研究も行った[19]。さらに、マイクロアレイ以外の EST や SAGE などの遺伝子発現データも収集提供活動も行っている。

BSD による遺伝子発現データの高度利用

空間的な広がりと時間的の広がりの中で遺伝子発現データを統合して表示かつ操作できるプラットフォームとして BSD を開発した。本システムでは、CIBEX に蓄積されるデータを含む様々な遺伝子発現データの中から、cDNA マイクロアレイ、DNA チップ、SAGE、EST を対象としたが、*in situ* ハイブリダイゼーションによる発現パターンの収集も念頭に開発を進めてきた。

BSD のプラットフォームによって、遺伝子発現データを3次元で表現した生物の形状データ上に表示することができ、したがって、どの遺伝子がいつ、どこで発現しているのかということを利用者は直感的に理解することが可能となる。また、時間的にどの段階から遺伝子が発現するようになるのかといった情報も視覚化される。

遺伝子発現データは、プラットフォームの舞台裏にあるデータベースに格納されているため、通常の検索はもちろん可能である。さらに特定の3次元形状に注目して、ある組織で発現している遺伝子を表示することや、複数の組織に共通に発現するものを選ぶといった検索も可能である。また、検索結果を可視化して3次元空間上に表示することで、遺伝子間の相互作用や、on/off といったネットワーク解明へとつながる情報を解析者に視覚的に認識させることができる。遺伝子発現情報は、遺伝子がいつ、どこで、どのように使われているかという情報であるが、それに加えて、BSD には例えばゲノム中での位置関係、翻訳されるタンパク質配列や、実験的に明らかになっている機能などといった付加的な情報もデータベースや文献データベースから参照する機能がある。

さらに、MGED フォーマットや CIBEX フォーマットに対応したデータを読み込むことができるので、BSD にもともと格納されているデータに利用者の独自データを加えて比較解析することができる(図9)。

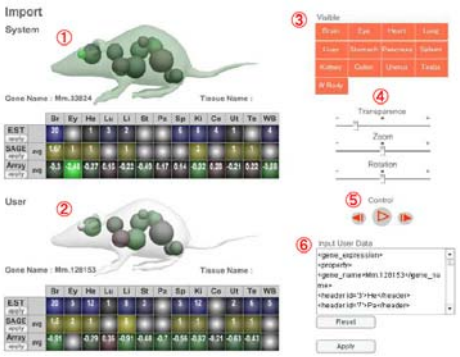

図9. BSD におけるデータ比較

BSD に登録済みのデータ(上段)と利用者のデータ(下段)の発現状況を視覚および直感で比較することができる

3. まとめ

JST BIRD 事業において DDBJ は、ゲノム生物学バックボーンデータベースを目指して、アノテーションの品質向上とデータの多角化に貢献する研究開発を進めてきた。

品質向上については、OASYS を基盤として、300種を超える微生物ゲノムを対象とした GTPS を実施した結果、計算機処理による第3者アノテーションが有効であることが分かった。また、OASYS を一般公開する準備が整った。

2005年度までの GTPS では、実施時点で最新の参照データベースを使って、INSD のゲノム・エントリーに遺伝子単位のエントリーとモチーフを投影することによって、CDS の再評価を行った。しかし、INSDC の歴史の中で、CDS だけでなく膨大な種類のアノテーションを受け入れる枠組み(features と qualifiers)が用意されていることから[20]、遺伝子単位のエントリーのアノテーションとゲノム・ワイドのアノテーションが相乗作用を及ぼしあうように、個々の遺伝子のエントリーについて features と qualifiers を

丁寧に充実していくことが考えられる。この目的のために、DDBJ を含むコミュニティーが結集した open annotation が望まれるところである。なお、INSDC の TPA については、2005年末までに4,560エントリーが登録されたが、国内からの登録は55件に留まっている。OASYS と TPA の定義が少し異なっているが、OASYS への共感をこれから国内で広げていくことによって、遺伝子配列とゲノム配列を問わずアノテーションに衆知を結集したい。

ところで、INSD は近年 WGS という枠組みを作った。Whole Genome Shotgun 法を使いながらも完全ゲノム決定は目指さずに、一定の長さのコンティグまでに結合した結果を、プロジェクト単位で格納する枠組みである。この件数が図10に示すように急速に増大している。そこで、OASYS に基づいた GTPS の成果を活かして、DDBJ 自ら WGS のデータにアノテーションを付与する試みも始めた。

多角化については、塩基配列を主対象としたデータベース INSD をモデルとして、遺伝子発現情報を主対象とするデータベースの可能性を実証するとともに、グラフィックを駆使した情報抽出の可能性を示すことができた。BSD のグラフィックは3次元表示と時系列表示が可能なことから、生物実験では不可能な角度からの観察や時間を圧縮した観察が、ディスプレイ上で可能になることを期待できる。

GTPS、GTOP、MGED および BSD の成果はそれぞれ、当所の JST BIRD プロジェクト用 Webs サイト http://www.jst-bird.nig.ac.jp/から利用可能である(図11)。加えて、遺伝子記号、記載内容あるいは配列データによって全ての成果を一括検索可能としてある。また、KEGG もこの一括検索に加えてあり、PDBj についてもこの対象に加えようと準備を進めている。

最後に、今回のシステム開発にあたっては、XML 技術[21]と Web サービス技術[22]を一部で採用し、その経験をもとに、従来行ってきた DDBJ のデータサービスにプログラムインターフェースを整備するという波及効果があったことを特記しておきたい。Web サービスは、かってホームページが在って当然となったと同様に、主要なサイトが備えておかなければならないサービスとなっている。

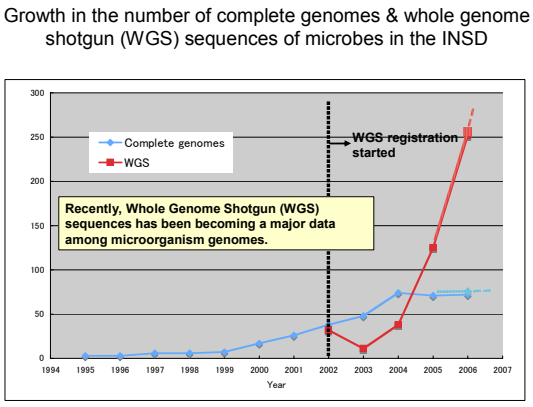

図10. Whole Genome Shotgun (WGS)の範疇の登録件数と完全ゲノムの登録件数の対比(年単位)

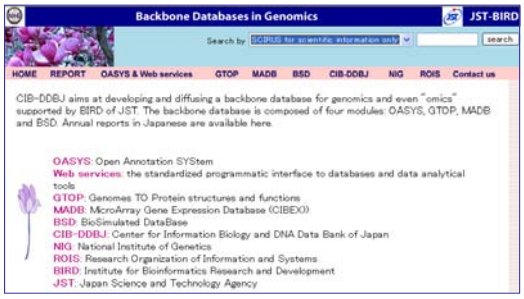

図11. DDBJ における JST BIRD 事業のホームページ

4. 研究開発実施体制

代表研究者　菅原秀明(国立遺伝学研究所　生命情報・DDBJ研究センター)

研究開発題目

(1) Open annotation system（OASYS）

グループリーダー　菅原秀明(国立遺伝学研究所生命情報・DDBJ研究センター)

(2) Genomes to proteins（GTOP）

グループリーダー　西川　建(国立遺伝学研究所生命情報・DDBJ研究センター)

(3) Microarray gene expression database（MADB）

グループリーダー　舘野義男(国立遺伝学研究所生命情報・DDBJ研究センター)

(4) Bio-simulated database（BSD）

グループリーダー　五條堀　孝(国立遺伝学研究所生命情報・DDBJ研究センター)

5. 参考文献

[1] Galperin MY. The Molecular Biology Database Collection: 2006 update. *Nucl. Acids Res.* 34（Database issue）: D3-5 （2006）

[2] http://www.insdc.org/

[3] Okubo K, Sugawara H, Gojobori T and Tateno Y. DDBJ in preparation for overview of research activities behind data submissions. *Nucl. Acids Res.* 34（1）: D6-D9 （2006）

[4] Fumoto M, Miyazaki S and Sugawara H. Genome Information Broker （GIB）: data retrieval and comparative analysis system for completed microbial genomes and more. *Nucl. Acids Res.* 30（1）:6-68 （2002）

[5] "ytjA gene （4616790..4616969）" in: Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett G 3rd, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL. *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005.*Nucl. Acids Res.* 34（1）:1-9 （2005）

[6] Sugawara H. Gene Trek in Prokaryote Space powered by a GRID environment, *Proceedings of the First International Workshop on Life Science Grid （LSGRID2004）*, May3 （2004）

[7] Fukuchi S and Nishikawa K. Estimation of the number of authentic orphan genes in bacterial genomes. *DNA Res.* Aug 31;11（4）:219-31, 311-313 （2004）.

[8] Homma K, Fukuchi S, Kawabata T, Ota M and Nishikawa K. A systematic investigation identifies a significant number of probable pseudo genes in the Escherichia coli genome. *Gene* 294: 25-33（2002）

[9] Fukuchi S and Nishikawa K. Protein surface amino-acid composition distinctively differ between thermophilic and mesophilic bacteria. *J. Mol. Biol.* 309: 835-843 （2001）

[10] Nakashima H, Fukuchi S and Nishikawa K. Compositional changes in RNA, DNA and

proteins for bacterial adaptation to higher and lower temperatures. *J. Biochem.* 133: 507-513 (2003)

[11] Fukuchi S, Yoshimune K, Wakayama M, Moriguchi M and Nishikawa, K. Unique amino acid composition of proteins in halophilic bacteria. *J. Mol. Biol.* 327: 347-357 (2003)

[12] Homma K, Kikuno RF, Nagase T, Ohara O and Nishikawa K. Alternative splice variants encoding unstable protein domains exist in the human brain. *J. Mol. Biol.* 343: 1207-1220 (2004)

[13] Fukuchi S, Homma K, Minezaki Y and Nishikawa K. Intrinsically disordered loops inserted into the structural domains of human proteins. *J. Mol. Biol.* 355: 845-857 (2006)

[14] Schena M, Shalon D, Davis R and Brown P. *Science* 270: 467-470 (1995)

[15] Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J and Vingron M. *Nature Genet.* 29: 365-371 (2001)

[16] Ikeo K, Ishi-i J, Tamura T, Gojobori T and Tateno, Y. *C. R. Biol.* 326:1079-1082 (2003)

[17]Brazma A, 池尾一穂, 舘野義男. マイクロアレイデータの標準化. *蛋白質核酸酵素* 48: 280-285, (2003)

[18] 舘野義男, 池尾一穂. 国際公共遺伝子発現データベース(CIBEX)とデータの登録. *蛋白質核酸酵素* 49: 2678-2683 (2004)

[19] Matsumura Y, Shimokawa K, Hayashizaki Y, Ikeo K, Tateno Y and Kawai J. Development of a spot reliability evaluation score for DNA microarrays, *Gene* 350: 149-160(2005)

[20] 舘野義男. 国際塩基配列データファイルの構造、*DDBJの利用法(五條堀・菅原編著)*、共立出版(東京): 21-31(2005)

[21] Miyazaki S, Sugawara H, Gojobori T and Tateno Y. DNA Data Bank of Japan (DDBJ) in XML, *Nucl. Acids Res.* 31 (1) :3-16 (2003)

[22] Sugawara H and Miyazaki S. Biological SOAP servers and web services provided by the public sequence data bank, *Nucl. Acids Res.* 31 (13) : 3836-3839 (2003)