

## 中間成果実績一覧

### 中核機関 整備実績

#### (1) 保有データ情報

##### (1-1) データの種類

①生物種	外部 DB や文献から得たデータは全生物種に及ぶ その他に、グループ内に植物関連のゲノム配列データおよびヒトの多型データを保有
②試料・ライブラリ等の種類、数	シアノバクテリア 12 種、根粒菌 11 種、ミヤコグサ 完全胞状奇胎試料、末梢血由来 DNA など
① 測定方法	ゲノム塩基の配列決定、DNA アレイ法による SNP アレルタイプ決定など
② データの内容	塩基配列、塩基配列上の予測遺伝子の番地、配列類似等の解析情報、遺伝子機能予測結果 ゲノムワイド確定ハプロタイプなど
⑤その他、特記事項	

##### (1-2) データソース

①現在のデータ量	国内 DB カタログ：国内 350、海外 50 の DB 注釈データ 横断検索：検索対象 30 サイトの検索用インデックスデータ ISND 全文検索：国際 DNA 配列バンク ISND の 800 万レコードの構造化データ 統合 TV：動画コンテンツ 59 種類 生命科学学協会検索：国内 594 の学協会情報 学会要旨統合俯瞰システム：3 学会 7400 要旨のテキストデータ 蛋白質核酸酵素全文検索：20 年間 5000 記事のテキストデータ WINGpro：国内外 420 の DB データ Web リソースポータル：解析ツール総数 456、解析ワークフロー総数 29、著名な論文 3 報のワークフロー KazusaAnnotation：05 Mbase 分のゲノム塩基配列に対する、100,680 遺伝子を対象とした遺伝子注釈情報
②データ区分	■自前 ■第三者 ■文献データ ■計算結果等の二次データ □その他
③将来の増加の見込み	開発の進行および環境の整備に伴い、随時増加の見込み
④権利関係	所有者（ 基本的に自前データないし使用許諾を受けたもの ）

	公開 ( <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ] )
⑤その他、特記事項	公開データについては、下記の(2)、(3)、(4)項を参照願います。

(1-3) データの管理状況

①更新頻度等の管理状況、体制	国内DBカタログやISND全文検索は日々更新、その他は随時更新 DB毎に担当者がアサインして管理
②その他、特記事項	

(1-4) データベース関係

①DB管理者数	中核機関グループ全体で延べ10名
②キュレータ・アナレータ数	中核機関グループ全体で延べ11名
③データ構造	関係データベース
④DB管理ソフト	MySQL、Postgres
⑤サーバのOS	Linux
⑥サーバ規模	PCクラスタ 8ノード、16CPUなど
⑦DBへのアクセス数	本プロジェクトの成果サイトへは、昨年10月の公開以来32,000件
⑧独立IP数	上と同じく、18,000件
⑨その他、特記事項	保有DBの詳細は下記[2]の項を参照ください。

(2) データ (又はDB) の連結、統合化整備

通番	データ (又はDB) の名称	公開／未公開	概要 (データの種類 (生物種)・数量 (kB等)、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述)
1	生命科学データベースカタログ	公開	生命科学分野のデータベースカタログ (注釈及びメタデータ) で、データベースサイトのトップページをサムネイルで表示。データベースの稼働状況をモニターして表示。利用者からのコメントを受け付ける。今後毎月30件のペースで掲載データベース数を増やし、今年度中には国内を完全に網羅する。記載内容の充実化を図ることが必要。課題はデータベースのIDの統一化など、他のデータベースに関する情報との対応を持たせること。 登録DB数: 400件 (H20.04現在) 開始年月: 平成20年2月~ 平均アクセス数 (月間): 3600
2	DB横断検索	限定公開	生命科学分野のデータベースや文献の横断検索サービスで、生命科学分野の国内外のデータベースと文献を網羅的に検索。検索対象とするデータベースに優先度を設け、H19年度中に掲載するデータベース、インデックス作成等掲載の準備をするデータベース、掲載が困難と予想されるデータベースに分類し計画に従って実施した。今後は、検索対象データベースをさらに広げつつ、主要なDBとそうでないアーカイブ的なDBに対する検索を分離する、またはユーザの対象とする生物種の検索を分離するなどしてユーザの利便性を図

			る。検索速度並びに検索精度の向上を図る。課題は、対象データベースが増加するに従って、維持管理のコストが高くなることが懸念される。検索対象 DB 数：30 データベース (H20.04 現在) 開始年月：平成 20 年 3 月
3	ISND 全文検索	公開	巨大な国際 DNA バンクを高速に検索し生物種、分子種、プロジェクトに分類して表示するシステム。従来巨大な計算機を必要としていたが、バンク内部を詳細に検討し 8000 万レコードを 55 万プロジェクトに縮退させ実質上、計算機一台での検索を可能にした。さらに、ある研究機関から発表されたある生物のゲノムデータを一括してダウンロードするというような研究単位でのデータ取得を実現した。BLAST による配列検索、核酸関連の特許公報へのダイレクトリンク、レコードの時系列展開も可能である。今後は、データベース受入れ、横断検索など、他のサービスとの統合を図る。検索の一層の高速化を図る。EST の臓器分類や発現データとの統合を図る。国外ユーザの獲得を図る。検索対象：8000 万レコード。開始年月：平成 19 年 3 月～平均利用者数（月間）：1487、平均アクセス数（月間）：10811、平均ダウンロード数（月間）：187
4	統合 TV	公開	データベースやウェブツールの使い方などを動画で発信するウェブサイト。平成 19 年 7 月公開。年度末の 9 ヶ月間間に 59 個の動画コンテンツを作成、訪問数で 30,423、総計 180Gbyte のダウンロードがあった。 <a href="http://togotv.dbcls.jp/">http://togotv.dbcls.jp/</a>
5	生命科学学協会検索	公開	生命科学系の学協会のデータベースと検索サービス。国内の生命科学系学協会を網羅したデータベース。各学会の詳細情報やウェブサイト、学会誌などの公開情報を掲載。H19 年度は日本学術会議の資料等をもとに生命科学系の学協会をリストアップしそれぞれの情報を調査、DB カタログと同じユーザインタフェースを利用して公開した。今後は、内容の充実化と、各学協会のウェブサイトに対するメタ検索の実施を行う。可能であれば公開している文献情報の検索も一括して行えるようにする。課題は学協会のウェブサイト等に対して検索をかける場合に許諾が必要となる場合もありうるため、各サイトのポリシーなどに配慮して進めなければならない。掲載学協会数：594 学会 開始年月：平成 20 年 4 月～
6	学会要旨統合俯瞰システム	公開	学会要旨集の検索サービス。著者名や施設名の表記揺れを吸収する辞書を搭載し、研究者の研究履歴をたどることが可能で、キーワードのトレンドを表示させることもできる。要旨集の提供を受けた遺伝学会と進化学会については全文検索全文表示を行っている。非公開の分子生物学会については、索引に対する検索のみ行っており要旨本文は表示しない。課題は、学会要旨は研究において大変有用な情報資源であり、JSTAGE など公的な仕組みを利用して電子化されているものの、会員以外非公開の場合も多く、オープン化に向けて学会への働きかけが必要であることである。検索対象：3 学会 7400 要旨。開始年月：平成 20 年 4 月
7	蛋白質核酸酵素全文検索	未公開	蛋白質核酸酵素バックナンバーの全文検索サービス。国内で出版社により発行されている総説誌、研究者にとって重要な研究発表の場であるとともに、無くてはならない良質な情報源でもある。今回、共立出版の協力により蛋白質核酸酵素のバックナンバー電子化と公開が可能となった。今後の計画は掲載論文をさらに過去にさかのぼって増やすこと。また、近刊についてはタイトルや要旨検索を可能にすること。関連論文、書籍などの検索を可能にすることが必要である。課題は著作権問題に対処可能な体制が必要ことである。そのため、著作権や知財に詳しい弁護士との協議を重ねている。収録論文：1985 年～2005 年。公開予定年月：平成 20 年 5 月

8	MotDB	公開	DBCLS を担っていく人材であるアノテータ、キュレータ、システムデータベース管理者向けの教材を提供するサイト (MotDB は Master of the database の略)。平成 19 年 3 月公開。平成 20 年 3 月リニューアルして、中核機関で独自に作成した教材 PDF ファイルもダウンロード可能とした。リニューアル後、訪問数で 1,830、計 457Mbyte のダウンロードがあった。 <a href="http://MotDB.dbcls.jp/">http://MotDB.dbcls.jp/</a>
9	ヒト遺伝子発現統合	公開	ヒト遺伝子の解剖学的な発現パターンデータの統合サイトを構築した。発現パターンは、測定法毎に異なる場合があることが知られており、ここではできるだけ客観的な発現パターンの解釈を可能にするために、5 種類の発現データ、即ち iAFLP、GeneChip、EST、NCBI の SAGEmap によるタグマップ、SAGE データの独自タグマップに基づく発現データを表示可能にした。 <a href="http://okubolab.genes.nig.ac.jp/bodymap_i/">http://okubolab.genes.nig.ac.jp/bodymap_i/</a>
10	ライフサイエンス受入れデータベース	試験的公開	cDNA 関連のデータベースとして、HGPD (Human Gene Protein Database)、FLJ Human cDNA DB (スプライシングバリエーション DB) と FLJ-DB (ヒト完全長 cDNA)、DBTSS、MiBASE (トマト EST DB) と KafTom (トマト全長 cDNA DB)、及び BodyMap、また理研補完課題に関連して KATANA (シロイヌナズナ DB) の受入と関係方法を検討し、受け入れテンプレートをを用いてその一部の受け入れの試行を行った。
11	WINGpro	公開	ライフサイエンス分野の内外のデータベースの情報を収集、整理、分類、420DB を収録。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可能。平成 19 年 3 月 30 日正式公開。平成 19 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。全文検索、分類一覧が可能。入力方式、一括データ取得、tips など記述。ユーザによる登録を促すことが課題。利用状況 (平均利用者数 1,300 (4-6 月は 300 程度であったが、その後の平均値)、アクセス数 138,208 (平成 19 年度)) <a href="http://wingpro.lifesciencedb.org/">http://wingpro.lifesciencedb.org/</a>
12	Web リソースポータルサイト for バイオインフォマティクス	公開	実験データの解析や公的データの加工に使用する解析ツールや環境を案内するサイト。ベンチメソッドオンロジーの一部を利用。解析ツール総数 456、解析ワークフロー総数 29、著名な論文 3 報のワークフローを収録、一括ダウンロードも可能。平成 19 年 3 月 30 日正式公開。教科書に近いサービスであるので、多くの研究者が行う実験フローをいち早く提供するなど工夫が必要。利用状況 (平均利用者数 100、アクセス数 2,673 (平成 19 年度))。 <a href="http://tools.lifesciencedb.jp/">http://tools.lifesciencedb.jp/</a>
13	Jabion Genome Viewer : HAL	公開	HAL (生物種 human) (ヒトゲノムに潜む遺伝子の位置と構造、その機能などの注釈情報を提供。)を既存のゲノムビューアである Jabion Genome Viewer に受け入れたテストサイト。H20.1.7 公開。ヒトゲノムに潜む遺伝子を発見する信頼性の高いコンピュータアルゴリズムにより予測した遺伝子数多数を公開。 <a href="http://www.bioportal.jp/genome/">http://www.bioportal.jp/genome/</a>
14	KazusaAnnotation	公開	高度情報集積型データベースのソーシャルブックマーク型サービスで、独自開発のソーシャルブックマークによるオープンアノテーションを可能にした遺伝子情報統合データベース 現在はβ版を運用しながらゲノムデータベースに対する情報蓄積を行い、約12万件 (5,657 エントリー、

			96,827 アノテーション、21,983 タグ) の情報を蓄積し、公開している。今後は、改善を加えながら運用し、情報の集積によるゲノムを基盤とした生物学研究情報の統合を図る。公開はH19. 10. 3で、一日平均694アクセス。 <a href="http://a.kazusa.or.jp/">http://a.kazusa.or.jp/</a>
15	KazusaNavigation	公開	高度情報集積型データベースのポータルサービスで、ソーシャルネットワーク型のElggをベースに開発。現在はβ版を運用しながら情報蓄積を続け、公開している。引き続き改善を加えながら運用する。まだ利用者数、アクセス数ともに少ないが、今後、メーリングリストのよりよい代替として、研究者コミュニティに利用をよびかけ、情報交換の促進による知識の統合を図る。公開はH19. 10. 12で、登録者数45名、一日平均27アクセス (Wikiと共通) <a href="http://navi.kazusa.or.jp/">http://navi.kazusa.or.jp/</a>
16	KazusaWiki	公開	高度情報集積型データベースのデータとりまとめサイトで、Wikiタイプのデータ共有サイトであるMediaWikiをベースに開発。現在はβ版を運用しながら情報蓄積を続け、公開している。今後は改善を加えながら運用し、研究者による情報とりまとめサイトとして発展させていき、研究情報の統合を図る。提供者側からの情報入力もすすめていく予定である。公開年月日はH19. 10. 23で、登録者数、アクセス数は上と同じ。 <a href="http://navi.kazusa.or.jp/wiki/index.php">http://navi.kazusa.or.jp/wiki/index.php</a>
17	Shared database of Applied Genomics (SJAG)	限定公開	多型情報大規模解析結果 (生データを含む) のコミュニティ内での共有を目指して開発したもので、評価を目的とした限定開示を行った。SJAG では、現在主流の SNP タイピング法であるアフィメトリックス社の DNA アレイを用いた解析データを LINUX 環境下、my. sql で管理している。

### (3) DB基盤システム、ツール等開発成果物の整備

通番	DB基盤システム、ツール等の名称	公開／未公開	概要 (主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述)
1	遺伝子名称シソーラス	公開	分子情報から医学文献までの遺伝子や蛋白質名称の正確な同定のための遺伝子固有名称 (「遺伝子名」と一般名称 (「ファミリー名」) の辞書データの構築を目的として、様々なデータベースで利用されている名称の収集と専門的キュレータによる編集を行い、遺伝子が持つ多様な名称の関係を明示した遺伝子名称シソーラスを開発した。本シソーラスは、ヒトをはじめ9種類の生物をカバーしている。 <a href="http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1">http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1</a>
2	生物学名羅日英対応辞書	公開	研究分野でよく使われる生物種の基準として、学名に日本語一般名を対応させた生物学名日本語一般名対応辞書を開発した。対応付けは、塩基配列データベース (DBJ) の登録エントリー数が多い生物種から順番に行った。また、標準和名が存在しない場合、その生物を説明する一般的な名称を用いた。登録データ数は合計14028種となっている。 <a href="http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1">http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1</a>
3	施設名称辞書	公開	日本のライフサイエンス研究を俯瞰するための重要な情報源として、各種関連学会の抄録などの報告文書が

			あるが、これらをデータベース化する際に問題となるのが、例えば、大阪大、阪大、大阪大学大学院などといった、施設名称研究室名称の表記ゆれである。これに対応するために、施設名称辞書を開発した。 <a href="http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1">http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1</a>
4	動物解剖学自動分類タガー	公開	動物の臓器、組織を、大きく 10 のグループに分類(大分類)し、さらにそれぞれのグループを細かく分類、合計 40 の小分類グループに分類する。基本的には、与えられた解剖用語に対して、解剖用語辞書、病理関連語彙の分類辞書、形容詞の解剖用語辞書、一般的な臓器名称の分類辞書、の 4 種類の辞書を順番に検索し、上記のカテゴリーに分類する。 <a href="http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1">http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1</a>
5	植物解剖学自動分類タガー	公開	植物(維管束植物)の部位、組織を大きく 6 のグループに分類し、さらにそれぞれのグループを細かく分類、合計 11 の小分類グループに分類する。分類における検索対象としては、生物種に合わせて、種子を持たない維管束植物の解剖用語辞書、イネ科の解剖用語辞書、その他被子植物の解剖用語辞書、裸子植物の解剖用語辞書、トウモロコシ属の解剖用語辞書、フウチョウソウ目の解剖用語辞書を用いる。 <a href="http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1">http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1</a>
6	都市名国名自動検出タガー	公開	論文やデータベースレコードに見られる国の名称の未記載や国の名称にみられる表記ゆれを吸収することを目的に、国別に分類するための辞書を開発した。 <a href="http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1">http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1</a>
7	ポリゴンマン辞書	公開	解剖学用語、すなわち臓器・器官・部位の名称やそれらにくる概念をモデル人間中の 3 次元座標で定義した辞書。ポリゴンマン内の空間関係で用語関係を表現すると、ツリー型表現(いわゆる解剖オントロジー)と違い、改変にロバスタな表現が可能。今後も用語の追加を続ける。 <a href="http://lifesciencedb.jp/ag/pgm">http://lifesciencedb.jp/ag/pgm</a>
8	アナトモグラフィー	公開	ポリゴンマン辞書から部品を選択して自由に人体の部分のモデル図が書ける。視点固定の図譜や 3 D アトラスと異なり透過度や着色、視点は自由に選択可能。今後は、解剖学教育や医学用語辞書の助けになる人体版グーグルマップを目指す。 <a href="http://lifesciencedb.jp/ag/">http://lifesciencedb.jp/ag/</a>
9	受入れテンプレート	試験的 公開	データベース受け入れシステム。さまざまな形式のさまざまな種類のデータを簡単に読み込み、自動的に共通のウェブ検索インターフェイスを生成する。また、自動的にウェブサービス(SOAP/WSDL API)を生成する。これらの機能により、容易に維持困難な DB を移設できる。また、本ツールのダウンロードにより、誰でも DB を構築可能である。 <a href="http://togodb.dbcls.jp/">http://togodb.dbcls.jp/</a>
10	統合ウェブサービス	公開	国内外のウェブサービスを透過的に統合するシステム。H20. 4. 1 公開。国内各サービスの稼働状況を継続的に監視し、サービスの連携に必要なデータ形式変換機能等を提供する。これにより、ユーザは統一的命名規則と統一的データ構造で国内外のウェブサービスにアクセスすることができるため、相互運用が容易となり、ワークフローの作成・蓄積を促進する。 <a href="http://togows.dbcls.jp/">http://togows.dbcls.jp/</a>
11	DBCLS OpenID サーバ	公開	一つの ID で複数のサイトを認証できるアカウントシステム。H20. 4. 1 公開。このシステムを利用することで、ユーザはひとつの ID とひとつのパスワードを保持すればよい。これにより、各サイトで認証サービス

			を用意する必要がなく、サイト間のユーザ情報の集約が容易になる。 <a href="http://openid.dbcls.jp/">http://openid.dbcls.jp/</a>
12	OReFiL	公開	オンライン上に存在する多数の生命科学系の資源(データベースやソフトウェアなど)を効率的に見つけるための検索システム。H19.8.6公開。信頼性や概要を容易に取得するために、資源の所在だけでなく、その資源について書かれている論文情報や、他の利用者の評価を閲覧できる。アカウントを取得することで、自ら評価を行うことも可能である。また、検索に関連する機能を全てウェブサービスとして提供しているため、利用者のプログラムから OReFiL を利用できる。 <a href="http://orefil.dbcls.jp/">http://orefil.dbcls.jp/</a>
13	Allie	公開	医学生物系論文書誌情報データベース MEDLINE を対象とし、出現する略字とその正規系のペアを検索するシステム。H20.4.1公開。生命科学系の論文では非常に多くの略字が使われており、同じ表記でも全く違う意味を示していることが少なくない。そこで、利用者の興味のある略字を検索語として入力することで、その使われ方を一覧表示すると共に、論文の発表年を提示する。また、検索された各略字について、その意味で使われている論文中で共起する他の略字も提示する。検索に関連する機能を全てウェブサービスとして提供しているため、利用者のプログラムから Allie を利用できる。 <a href="http://allie.dbcls.jp/">http://allie.dbcls.jp/</a>
14	ScrapParty	公開	ブラウザのプラグインによるキュレータ支援用アプリケーション。オンラインで文献やウェブサイトを閲覧しながら、重要な記述や画像、図表を集めコメントとともに簡単に DB 化するツール。H18 年度アノテーションの調査、試験に利用した。
15	WiredMarker	公開	情報共有ツール、アノテーション支援ツール。ブラウザのプラグインとして働く。データベース構築、アノテーションを支援するブラウザ用のプラグイン。オンラインで閲覧したページ、論文のテキストや図をマークしブックマークすることが可能。問の際にも同じ箇所がハイライトされアノテーションのエビデンスとして詳細な情報が残せる。またフォルダを自由に階層化して情報を整理することが可能。 公開日：H19.12.19 総ダウンロード数：8265 件 <a href="http://www.wired-marker.org">http://www.wired-marker.org</a>
16	メタデータ要素レポジトリ (MDeR)	試験的公開	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が可能。ISO/IEC 11179 Part3(Registry metamodel and basic attributes)に準拠した形で、国際標準3種、データベース3種のメタデータ要素を収録。収録内容の精査と充実によりデータ収集・データベース構築時のデータ形式等の決定に役立つサービスを目指す。
17	グリッド環境構築	未公開	グリッドの管理サーバ、Webサーバと計算サーバからなるグリッド環境を構築し、2.のプロトタイプ(配列解析)システムを開発
18	配列解析システム(プロトタイプ)	未公開	CBRCで開発されたPOODLE-S、-L、-WおよびTMBETA-NET、TMBETADISC-COMP、GRIFFINを一度に起動し結果を得るシステム。キーワードを入力するとLSDBの生命科学データベース横断検索を行い結果を得ることも可能。
19	genoDive Pro / Eu	限定公開	DAS規格に準拠したゲノム情報を統合的に閲覧することが可能なDASサーバクライアント。現在は試験運用中、テスト終了後、ソースコードも含め完全公開し配布を行う予定。
20	遺伝子名称変換プログラム	限定公開	植物遺伝子名や遺伝子IDの名称の食い違いを解決するための名前変換サービス。現在は試験運用中、テスト終了後、完全公開の予定。

21	なんでも DAS	限定公開	任意のウェブページに含まれる遺伝子名を抽出して、GFF 形式に変換するサーバ。このサーバを活用することで、インターネット上のあらゆる遺伝子関連情報をゲノム上にマッピングし、情報の統合的な閲覧を可能とする目的で開発した。現在は試験運用中、テスト終了後、完全公開の予定。
----	----------	------	---

(4) その他の成果物 ((2)、(3) に該当しないもの)

通番	名称	公開／未公開	概要
1	統合 PJ 成果公開サイト	公開	統合 DB プロジェクトに関する中核機関の成果公開サイト。生命科学データベースカタログ、DB 横断検索、INSD 全文検索、統合 TV、OReFiL、生命科学学協会検索、学会要旨統合俯瞰システム、MotDB、LSDB ブログ等を掲載している。http://lifesciencedb.jp
2	事業広報サイト	公開	統合データベースプロジェクトの紹介サイト http://lifesciencedb.mext.go.jp/
3	ライフサイエンスの広場	公開	文科省のライフサイエンス政策等の紹介サイト http://www.lifescience.mext.go.jp/index.html
4	統合データベース間の連携調査	公開	代表的モデル研究植物である「イネ」ならびに「シロイヌナズナ」のゲノムアノテーション型公開データベース各 46 サイト、25 サイトの基本項目の調査と実験生物研究者 188 名のアンケートにより調査し、それぞれのデータベース間の連携のための現状の課題と将来の DB 統合にむけた課題の整理を行った。 http://charles.kazusa.or.jp/lifesciencedb/
5	検索アルゴリズムを含めた知識情報技術の動向調査	公開	次世代の生物情報データベース統合に必要な知識情報技術として、検索システム、データマイニング、Web 2.0 およびグリッドコンピューティングに焦点を絞り、聞き取り調査や文献調査によって動向を調べた。 http://lifesciencedb.mext.go.jp/result/strategy.html#id1-4
6	遺伝統計学分野における解析技術の基礎調査	公開	遺伝統計学分野で用いられる連鎖解析、連鎖不平衡解析)、QTL 解析等の解析手法と、それぞれの手法における代表的なアルゴリズム計 8 種類の調査を行ない、その特徴および長所・短所を評価した。併せて、各手法の代表的プログラム計 15 種類に関して、実装されているアルゴリズム、動作環境などを調査し、その評価を行った。 http://lifesciencedb.mext.go.jp/result/strategy.html#id1-4
7	臨床情報・疾患健康情報の調査	公開	わが国の代表的コホート研究に関してその研究の背景と目的、対象地域、ターゲット疾患、特徴的な検査項目、対象人数、代表研究者、研究開始時期、資金源等について調査した。 http://lifesciencedb.mext.go.jp/result/strategy.html#id1-4
8	ライフサイエンスプロジェクトの調査結果	試験的公開	この約 10 年間のライフサイエンス国家プロジェクトの事業内容や成果を明らかにする目的で、4 省庁のゲノム、ポストゲノム関連の主要プロジェクトの調査を行った。プロジェクト名、期間、予算、主要研究者、主要事業内容、公開状況、公開 DB、データダウンロードサイト、公開 HP、事業化関連、報告書、その他文献等に関する情報を情報ソースへのリンクをできるだけつけた形で整理した。
9	データベース受入に関する	試験的	受入による統合化の方針策定と受入対象の候補抽出を目的として、国内データベースを対象として、データ



	アンケート調査	公開	ベース内容と検索方法、経歴、維持・管理状況、提供可能性及び提供時の問題点などをアンケートにより調査した。送付した104件中、65件の回答が得られ、半数以上の機関が、統合DB事業に対して全データもしくは一部データを開示できると回答した。
10	BioHackathon ホームページ	公開	分散環境のまま、各種DBをアルタイムかつ統合的に利用できる環境を構築すること目的に開催した、国内外のライフサイエンスDBのウェブサービスにおけるデータ構造や命名規則を標準化するためのBioHackathon2008の活動内容と経過を紹介するサイト。2008年2月11日から15日まで都内で開催。海外参加者34名、国内参加者37名。 <a href="http://hackathon.dbcls.jp/">http://hackathon.dbcls.jp/</a>
11	データベース、ソフトウェアのSOAP化	未公開	将来のワークフローシステム構築の準備として、産総研CBRCで開発した7種類のソフトウェアのSOAP化を行った。

## 中核機関 外部発表実績一覧

### (1) セミナー、研究会等イベント開催

通番	タイトル	発表者 (代表者)	年月日	開催場所	イベント名称	概要(対象者(層、参加人数)、出席者の主な反応等)
1	ライフサイエンス分野の統合データベースプロジェクト	川本祥子	H20.1.25	長浜バイオ大学	統合データベース講演会	学部学生、他機関研究者等 60名
2	バイオデータベース構築	中谷洋一郎	同上	同上	同上	同上
3	ゲノムアノテーションとは何か：その付けられかた／活用方法の紹介	中村保一	同上	同上	同上	同上
4	バイオデータマイニング入門	瀬々 潤	同上	同上	同上	同上
5	使い倒し系バイオインフォマティクスによる遺伝子発現情報解析	坊農 秀雅	同上	同上	同上	同上
6	テキストを対象とした生命科学研究	山本 泰智	同上	同上	同上	同上
7	Objectives of this hackathon and current status of Japanese web services	T. Katayama	H20.2.11	六本木アカデミーヒルズ	BioHackathon2008 Presentation	海外研究者34名、国内研究者37名
8	From Web API for Biology (WABI) to Semantic Web API for Biology (SABI)	H. Sugawara	同上	同上	同上	同上
9	Current status of the BioMOBY project and vision for the future directions	M. Wilkinson	同上	同上	同上	同上

10	Moby, Legacy Apps and the Semantic Web	P. Gordon	同上	同上	同上	同上
11	Data, Services and Computational Resources Integration at the INB	O. Trelles	同上	同上	同上	同上
12	The EMBRACE project and WS-I standard	J.C. Bryne	同上	同上	同上	同上
13	Soaplab2 project to wrap up command line packages	M. Senger	同上	同上	同上	同上
14	Taverna (part of myGrid project)	T. Oinn	同上	同上	同上	同上
15	Generation Challenge Program effort at building interoperability	R. Bruskiwich	同上	同上	同上	同上
16	統合データベースプロジェクトとライフサイエンス統合データベースセンター	坊農秀雅、	H20. 3. 5	JST 東京本部 住宅等棟	統合データベース 講習会: AJACS 東京	学部生、大学院生 20 名
17	かずさにおける新しいアノテーションの試み	岡本忍	同上	同上	同上	同上
18	統合データベースへの期待	樋口千洋	同上	同上	同上	同上
19	ゲノム情報リテラシー入門	中村保一他	同上	同上	同上	同上
20	テキスト処理して生命科学	山本泰智	H19. 6. 20	お茶の水 女子大学	第 7 回 バイオインフォマティクスへの招待	内部・外部学生、その他 22 人
21	KEGG データベースの進化と、統合・標準化	片山俊明	H19. 6. 20	お茶の水 女子大学	第 7 回 バイオインフォマティクスへの招待	内部・外部学生、その他 22 人
22	選択的スプライシングと植物の環境適応	飯田慶	H19. 9. 3	お茶の水 女子大学	第 8 回 バイオインフォマティクスへの招待	内部・外部学生、その他 37 人
23	バイオインフォマティクスを用いた RNA 研究による生命現象の理解	程久美子	H19. 10. 2 9	お茶の水 女子大学	第 9 回 バイオインフォマティクスへの招待	内部・外部学生、その他 29 人
24	バイオリソース事業におけるインフォマティクスの役割	深海薫	H19. 12. 6	お茶の水 女子大学	第 10 回 バイオインフォマティクスへの招待	内部・外部学生、その他 16 人
25	システム進化生物学: 生命システムのランドデザインの解明を目指して	荻島創一	H20. 2. 20	お茶の水 女子大学	第 11 回 バイオインフォマティクスへの招待	内部・外部学生、その他 31 人

26	生物学者のための生命情報解析ツールの開発	田村浩一郎	H20. 3. 7	お茶の水女子大学	第12回バイオインフォマティクスへの招待	内部・外部学生、その他 29人
27	KazusaAnnotation システムによるゲノムアノテーション高度化の試み	岡本忍	H19. 12. 3	かずさアカデミアパーク	ラン藻の分子生物学2007	ラン藻研究者、約100人、KazusaAnnotation, KazusaNavigation, KazusaWikiを限定公開して実際に使っていた、ツールの有用性
28	第八回オープンバイオ研究会企画運営	中尾光輝	H19. 12. 18	科学未来館	第八回オープンバイオ研究会	JSBi 年会参加者、約20名、国内のオープンソース関連の開発者とユーザの活発な交流の場となった。

(2) プレス発表、取材対応

通番	タイトル	発表媒体	年月日	特記事項
1	統合DBプロジェクトが本格始動 仮説の枯渇と蛸壺化を改善	BTJ ジャーナル(日経BP)	H19. 8	
2	大学共同利用機関法人情報・システム研究機構、 ライフサイエンス統合データベース・ポータルサイト開設	日経バイオテク	H19. 10. 5	
3	生命科学の情報を網羅 今年度から統合データベース作成	読売新聞	H19. 11. 07	
4	「BioHackathon 2008」が六本木で開幕、13カ国から80人参加、12日から4日間お台場で「Hack、hack、hack」	日経バイオテク	H20. 2. 12	
5	生命科学DB 統合へ	朝日新聞	H20. 3. 31	

(3) 展示会等出展 (該当なし)

(4) 学会等への口頭発表

通番	タイトル	発表者	学会等の名称	年月日	特記事項
1	データベースから見たライフサイエンスプロジェクト	高木利久	日本分子生物学会 2006 フォーラムシンポジウム[プロジェクト型研究時代の生命科学の課題]	H18. 12. 8	

2	知的生産性向上のための情報処理	大久保公策	同上	同上	
3	生物学研究基盤としてのゲノムアノテーション	中村保一	同上	同上	
4	ライフサイエンスDB その歴史とわが国の現状と課題	高木利久	日本分子生物学会 2006 フォーラム バイオテクノロジーセミナー	同上	
5	オントロジーや辞書は役に立つのか	大久保公策	同上	同上	
6	使い倒し系バイオインフォマティクスによる知のめぐりのよい生物学研究のすすめ	坊農秀雅	お茶の水女子大学「魅力ある大学院教育」第6回バイオインフォマティクスへの招待	H19. 3. 16	
7	統合 DB の構築に必要な情報技術	高木利久	情報とシステム 2007	H19. 3. 1	
8	ライフサイエンスのデータベースの現状と課題	大久保公策	同上	同上	
9	ライフサイエンスにおけるゲノム情報の高度利用に向けた生命知識の構造化	大久保公策	知の構造化ワークショップー知の構造化ツールは、新しいサイエンスを開くのかー	H18. 12. 4	
10	知識発見のための癌臨床情報のデータベース化	加藤菊也	第2回 大阪大学臨床医工学融合研究教育センターシンポジウム	H18. 10. 8	
11	ゲノム研究のためのオントロジーとテキストマイニング	高木利久	第16回セマンティックウェブとオントロジー研究会	H19. 7. 23	招待講演
12	An approach to decipher gene regulatory networks from the federation of databases in life science	H. Bono, S. Kawano, S. Kawamoto and T. Takagi	FUNCTIONAL GENOMICS & SYSTEMS BIOLOGY	H19. 10. 10-13	Hinxton (UK)
13	YUZ: an environment for integration AND ANALYSIS of gene regulatory networks	H. Bono, S. Kawano, S. Kawamoto, and T. Takagi	GENOME INFORMATICS	H19. 11. 1-5	New York (USA)
14	統合 TV によるがんとハイポキシア研究の推進	坊農秀雅	第5回 がんとハイポキシア研究会	H19. 12. 1	千葉
15	MeSH terms を用いた生物学的機能付与による遺伝子群の解析手法の開発	仲里猛留、坊農秀雅 他2名	第30回日本分子生物学会年会第80回日本生化学会大会 合同大会 (BMB2007)	H19. 12. 11-15	横浜
16	ライフサイエンス統合データベースプロジェクトの課題と成果	川本祥子、坊農秀雅 他11名	同上	同上	同上
17	ライフサイエンス統合データベース基盤整備のためのデータベースポータル構築: WINGpro	黒田雅子、小池俊行 他10名	同上	同上	同上
18	ライフサイエンス統合データベースセンターに	河野信、小野	同上	同上	同上

	おけるデータベースポータル構築	浩雅 他 7 名			
19	3次元人体モデルのセグメンテーションによる解剖学用語の形式表現	三橋信孝、藤枝香 他 3 名	同上	同上	同上
20	遺伝子発現統合データベースの開発	有川浩司、飯塚高康 他 3 名	同上	同上	同上
21	分子生物学の研究動向の俯瞰を目的としたデータバンク (INSDC, GEO) 目次の開発	小笠原理、飯塚 高康 他 3 名	同上	同上	同上
22	ゲノムセントラルな遺伝子発現情報融通システム YUZ (柚子)	坊農秀雅	同上	同上	同上
23	TogoTV - a broadcast station of tutorial movies about bioinformatics resources	S. Kawano, H. Ono, H. Bono, S. Kawamoto, T. Takagi	日本バイオインフォマティクス学会年会 (JSBi2007)	H19. 12. 17-19	東京
24	ライフサイエンス統合データベースプロジェクト	高木利久	「生命をはかる」研究会	H20. 2. 18	
25	統合データベースセンターの紹介	河野信	第 37 回人工知能学会分子生物情報研究会 (SIG-MBI)・第 9 回オープンバイオ研究会	H20. 3. 7-8	能美
26	BioHackathon for Web Service の報告	片山俊明	同上	同上	同上
27	DAS workshop 2008 参加報告	中村保一	2008. 03. 08	同上	同上
28	ウェブサービスの統合に向けた BioHackathon の成果	片山俊明	国立遺伝学研究所研究会 『生物情報資源の相互運用性』	H20. 3. 25	三島
29	ライフサイエンス統合データベースプロジェクトの目標と成果	川本祥子	同上	同上	同上
30	OreFiL	Y. Yamamoto	Pacific Symposium on Biocomputing	H20. 1. 4-8	Hawaii, U. S. A.
31	Funcional analysis of groups of genes with MeSH hierarchy	T. Nakazato, H. Bono et al.	Systems Biology: Global Regulation of Gene Expression	H20. 3. 27-30	New York, USA
32	Analysis of Genes Affecting Susceptibility to Systemic Lupus Erythematosus (SLE).	T. Tahira, K. Hayashi et al.	Annual Meeting of American Society of Human Genetics	2006. 10. 9-13	New Orleans, U. S. A
33	Detection of human copy number variations	Y. Kukita, K.	The Biology of Genomes meeting	2007. 5. 6-10	Cold Spring

	using a collection of Japanese complete hydatidiform moles	Hayashi, et al.			Harbor Laboratory, U. S. A.
34	Enhanced D-HaploDB: definitive haplotypes and extended haplotype information determined by genotyping complete hydatidiform mole samples	Koichiro Higasa, Kenshi Hayashi, et al.	The Biology of Genomes meeting	同上	同上
35	D-HaploDB: A Database of Genome-Wide Definitive Haplotypes Determined using Complete Hydatidiform Moles	Y. Kukita, K. Hayashi, et al.	第5回国際バイオデータ相互運用性会議	H19. 9. 26	青海、東京
36	Evaluation of kernel-based link analysis measures on research paper recommendation	新保仁	ACM/IEEE Joint Conference Digital Libraries	H19. 6. 22	
37	A Discriminative Learning Model for Coordinate Conjunctions research paper recommendation	新保仁	Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning	H19. 6. 29	
38	A Graph-Based Approach to Named Entity Categorization in Wikipedia Using Conditional Random Fields	渡邊陽太郎	同上	同上	
39	Pivot learning for efficient similarity search	木村学	International Conference, Knowledge-based Intelligent	H19. 9. 14	
40	ストレスフリーなデータベースにむけて	中村保一	日本植物生理学会	H19. 3. 28	シンポジウム「植物データベース講習会」への招待講演。
41	KazusaAnnotation: ゲノム情報へのアノテーションを支援するシステム	岡本忍	同上	同上	
42	KazusaAnnotation: ゲノム情報への注釈付け、注釈の利用を支援するシステム	岡本忍	日本ゲノム微生物学会	H19. 3. 6	
43	System update and literature curation at CyanoBase.	岡本忍	International Biocuration meeting 2007	H19. 10. 25	
44	学部教育としての持続可能型社会への貢献	上原啓史、池	日本遺伝学会第79回大会	H19. 9. 19	

	遺伝子データベースの構築	村淑道他			
45	学部教育としての持続可能型社会への貢献 遺伝子データベースの構築	上原啓史、池 村淑道他	第30回情報化学討論会	H19.11.16	
46	持続可能型社会への貢献遺伝子データベース ～膨大な環境由来メタゲノム配列からの有用遺 伝子探索～	上原啓史、池 村淑道他	第2回日本ゲノム微生物学会年会	H20.3.6	

(5) 雑誌等への論文寄稿

通 番	タイトル	著者名	雑誌等の名称	掲載巻、号、ページ	特記事項
1	生命科学データベースの現状と課題	大久保公策	科学(岩波)	77(4), 364-369 (2007)	
2	生命科学系データベース統合化の背景	大久保公策	蛋白質核酸酵素 (共立出版)	52(9), 1027-1031 (2007)	連載企画「ライフサイエンス分野の 統合データベース」
3	統合データベースセンターがめざすもの	高木利久	同上	52(11), 1388-1393 (2007)	同上
4	ライフサイエンスにおけるデータベース構築 のための人材育成	瀬々潤、池村淑 道	同上	53(1), 87-93 (2007)	同上
5	統合データベースプロジェクトのサービスと その利用法	川本祥子、坊農 秀雅	同上	53(3), 281-287 (2007)	同上
6	利用の立場からのコメント	高木利久	同上	53(5), 686-691 (2007)	同上
7	OReFiL: an online resource finder for life sciences	Y. Yamamoto and T. Takagi	BMC Bioinformatics	8:287 (2007)	<a href="http://www.biomedcentral.com/1471-2105/8/287/">http://www.biomedcentral.com/ 1471-2105/8/287/</a>
8	BioCompass: A novel functional inference tool that utilizes MeSH hierarchy to analyze groups of genes	T. Nakazato, H. Bono et al.	<i>In Silico</i> Biology	8, 0006 (2007)	<a href="http://www.bioinfo.de/isb/2007/08/0006/">http://www.bioinfo.de/isb/ 2007/08/0006/</a>
9	Ensembl (Ensembl Genome Browser)	坊農秀雅	バイオデータベ ースとウェブツ ールの手とり足 とり活用法(羊 土社)	p175-180 (2007)	
10	Reactome	河野信	同上	p175-180 (2007)	
11	脊椎動物ゲノム進化を推定するロジック	森下 中谷	細胞工学別冊	29 - 40 (2007)	

12	Reconstruction of the Vertebrate Ancestral Genome Reveals Dynamic Genome Reorganization in Early Vertebrates.	Y. Nakatani, H. Takeda, Y. Kohara, and S. Morishita	Genome Research	17(9): 1254-1265	
13	Vertebrate genome evolution examined by comparing the human and fish genomes.	Y. Nakatani and S. Morishita	<i>Encyclopedia of Life Sciences</i> , John Wiley & Sons	印刷中	
14	D-HaploDB: a database of definitive haplotypes determined by genotyping complete hydatidiform mole samples.	K. Higasa, K. Hayashi et al.	Nucleic Acids Res.	35, D685-689 (2007)	
15	Periodicity of SNP distribution around transcription start sites.	K. Higasa, K. Hayashi	BMC Genomics	7, 66 (2006)	
16	A Large-scale Protein protein Interaction Analysis in <i>Synechocystis</i> sp. PCC6803	S. Sato, M. Nakamura et al.	DNA Research	14, 207-216 (2007)	本論文の大規模タンパク質間相互作用データを KazusaAnnotation のソーシャルブックマークとして登録し、利用に供している。
17	Complete genomic structure of the bloom-forming toxic cyanobacterium <i>Microcystis aeruginosa</i> NIES 843.	N. Nakajima, S. Okamoto et al.	同上	14, 247-256 (2007)	新規に決定したシアノバクテリアゲノム論文。本プロジェクトの成果物によるアノテーション高度化の対象である。
18	ライフサイエンス分野のデータベース：学部学生とシニア世代の共同作業としての知識発見の可能性も含めて	池村淑道	日本化学会情報化学部会誌	26(1) 1-4 (2008)	
19	エキスパートがキュレートした tRNA データベース	井口八郎、池村淑道他	同上	26(1) 11-16 (2008)	
20	持続可能型社会への貢献遺伝子データベース	上原啓史、池村淑道他	同上	26(1) 17-19 (2008)	



## 分担機関（京都大学） 整備実績

### (1) 保有データ情報（該当なし）

### (2) データ（又はDB）の連結、統合化整備

通番	データ（又はDB）の名称	公開／未公開	概要（データの種類（生物種）・数量（kB等）、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述）
1	ゲノムネット医薬品データベース <a href="http://www.genome.jp/kusuri">http://www.genome.jp/kusuri</a>	公開	研究の最先端と医療の現場さらには一般社会をつなぐ日本語の医薬品統合データベース。JAPIC 医薬品添付文書情報（医療用医薬品 13,973 件、一般用医薬品 12,658 件、2008 年 4 月現在）を検索可能。KEGG DRUG の構造情報やターゲット情報と統合している。また、文献データベースへのリンクも付加している。医療用医薬品は 2007 年 9 月、一般用医薬品は 2008 年 1 月より公開している。アクセス数等の情報は本文の進捗状況の欄を参照のこと。
2	DBGET/LinkDB: ゲノムネット統合データベース検索システム <a href="http://www.genome.jp/ja/gn_dbget_ja.html">http://www.genome.jp/ja/gn_dbget_ja.html</a>	公開	2006 年度までに DBGET/LinkDB として開発してきたシステムを、日本語支援環境の整備、LinkDB の拡張、新たな検索システムの開発という観点から改良したもの。全データベース一括検索と外部データベースを含む LinkDB 検索を 2007 年 7 月に、日本語支援環境を 2007 年 10 月に公開した。

### (3) DB基盤システム、ツール等開発成果物の整備

通番	DB基盤システム、ツール等の名称	公開／未公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）
3	SIMPCOMP	公開	類似化合物検索システム。グラフ比較に基づいた精度の高い類似度計算を実現している。検索速度に問題があったため、平成 19 年度には高速化についての調査を行い、平成 20 年度に高速化を実現する。
4	e-zyme	公開	化学構造変化に基づく反応予測システム。基質と生成物を与えると、その間の反応パターンを予測、EC 番号との対応付けなどを行う。テンプレートとなる反応パターンの充実が課題であったため、平成 20 年度に反応パターンデータベースを整備し、化合物データとリンクさせる。また、平成 21 年度以降に、複数反応ステップの予測システムを実現する。
5	KCaM	公開	糖鎖類似構造検索システム。糖鎖に特徴的な木構造のための動的計画法を実装したシステムであり、ユーザーインタフェースを他のシステムと統一した。今後は、以下の糖鎖構造予測システムとの連携を計画している。
6	GECS	公開	遺伝子発現データから化合物構造を予測するシステム。ゲノム情報と化合物情報を結ぶためのシステムとして開発している。平成 19 年度は糖転移酵素のリストから合成可能な糖鎖構造を予測するシステムを開発し、平成 20 年 4 月に第 1 版を公開した。今後は、ユーザーインタフェースなどを改良するとともに、脂質など他の化合物のためのシステムを開発し統合する。

## 分担機関（京都大学） 外部発表実績一覧

### （1）セミナー、研究会等イベント開催

通番	タイトル	発表者 (代表者)	年月日	開催場所	イベント名称	概要（対象者（層、参加人数）、出席者の主な反応等）
1	ゲノムネットデータベース講習会	五斗進他	2008/1/30, 31	東京大学		PCを用いた実習形式での講習会。ホームページ上で一般から20名の参加者を募った。大学、公的機関の研究所、企業から幅広く集まった。

### （2）プレス発表、取材対応（該当なし）

### （3）展示会等出展（該当なし）

### （4）学会等への口頭発表

通番	タイトル	発表者	学会等の名称	年月日	特記事項
2	医薬品情報統合データベースの開発	伊藤真純他	BMB2007	2007. 12. 11-15	ポスター発表

### （5）雑誌等への論文寄稿

通番	タイトル	著者名	雑誌等の名称	掲載巻、号、ページ	特記事項
3	医薬品の統合データベース	金久實他	蛋白質核酸酵素	52, 12, 1486-1491	

## 分担機関（東京医科歯科大学グループ） 整備実績一覧

### （１）保有データ情報

#### （１－１）データの種類

①生物種	(1) 東京医科歯科大学：ヒト (2) 大阪大学：ヒト
②試料・ライブラリ 一等の種類、数	(1) 東京医科歯科大学：＜試料＞肝臓癌、大腸癌、口腔癌の手術または生検検体（約300） ＜データ＞肝臓癌、大腸癌、口腔癌の症例情報（約300）、 肝臓癌、大腸癌、口腔癌検体の遺伝子発現解析結果（約200） (2) 大阪大学：＜データ＞神経難病の症例情報（約400）
③測定方法	(1) 東京医科歯科大学：診療情報収集、DNA マイクロアレイによる遺伝子発現解析 (2) 大阪大学：診療情報収集
④データの内容	(1) 東京医科歯科大学：臨床情報（基本情報、病歴・生活歴、臨床検査、画像診断、治療、予後等） 分子情報（マイクロアレイ遺伝子発現情報） (2) 大阪大学：臨床情報（基本情報、症状、治療等）
⑤その他、特記事項	

#### （１－２）データソース

①現在のデータ量	(1) 東京医科歯科大学：試験公開可能なデータは、約100症例分の症例情報。 非公開データを含めると、約300のがん症例情報を有する。 (2) 大阪大学：試験公開可能なデータは、約100の症例分の統計情報。 非公開データを含めると、約400の神経難病症例情報を有する。
②データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input checked="" type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他
③将来の増加の見込み	倫理審査委員会が承諾する範囲内で、今後も継続的に症例情報を追加していく。
④権利関係	所有者（東京医科歯科大学、大阪大学） 公開（ <input type="checkbox"/> 可 <input type="checkbox"/> 否 <input checked="" type="checkbox"/> その他 [各機関の倫理審査委員会が承諾する範囲での公開]）
⑤その他、特記事項	個別に試験公開可能なデータの公開をしている。(http://ibmd.tmd.ac.jp) 各データベースの倫理規定に基づいての公開を行うとともに、統合医科学データベースの倫理規定案の策定を推進す

	る。
--	----

(1-3) データの管理状況

①更新頻度等の管理状況、体制	症例情報の収集及び検体からの遺伝子発現解析は継続的に実施しており、集積データをクレンジングし、公開用データベースへのデータ更新は年1回～2回を予定している。
②その他、特記事項	

(1-4) データベース関係

①DB 管理者数	(1) 東京医科歯科大学 : 3名 (2) 大阪大学 : 1名
②キュレータ・アナレータ数	(1) 東京医科歯科大学 : 3名 (CRC他) (2) 大阪大学 : 1名
③データ構造	各データベースで独自のデータ構造
④DB 管理ソフト	PostgreSQL
⑤サーバの OS	Linux
⑥サーバ規模	
⑦DB へのアクセス数	
⑧独立 IP 数	1
⑨その他、特記事項	

(2) データ (又はDB) の連結、統合化整備

通番	データ (又はDB) の名称	公開 / 未公開	概要 (データの種類 (生物種)・数量 (kB 等)、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述)
1	網羅的疾患分子病態データベース <a href="http://ibmd.tmd.ac.jp">http://ibmd.tmd.ac.jp</a>	試験公開	臨床、病理、分子情報 (遺伝子発現情報) を統合化したデータベース。疾患共通のテンプレートに情報を適応し、100 症例のデータを公開した。平成 20 年度には 200 症例、最終年度には 300 症例のデータ公開を行う予定。
2	パーキンソンデータベース <a href="http://ibmd.tmd.ac.jp">http://ibmd.tmd.ac.jp</a>	試験公開	神経難病特の、臨床情報データベース。100 症例の臨床 10 項目に対して、統計情報を取りこれを公開した。平成 20 年度には 400 症例、最終年度には 500 症例のデータ公開を行う予定。

(3) DB 基盤システム、ツール等開発成果物の整備

通	DB 基盤システム、ツール等の名称	公開 / 未	概要 (主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述)
---	-------------------	--------	--

番		公開	
1	要件定義書	プロジェクト内部公開	統合医科学データベース構築に関わる統合化技術(情報モデル、オントロジー、セマンティクス)、標準化、公開倫理等に関して、全国の全疾患DBを対象とした調査に基づき、要件定義を行った。継続的に要件定義を行い、平成20年度に要件定義書を完成する。
2	統合検索システムプロトタイプ <a href="http://ibmd.tmd.ac.jp/Prototype">http://ibmd.tmd.ac.jp/Prototype</a>	未公開	東京医科歯科大学の網羅的疾患分子病態データベース、大阪大学のパーキンソンデータベースに対しセマンティクス検索技術、ユーザインターフェースを検証するためのプロトタイプシステムを構築した。 機能限定版のプロトタイプシステムを評価用に試験公開する予定。

(4) その他の成果物 ((2)、(3) に該当しないもの) (該当なし)

### 分担機関 (東京医科歯科大学グループ) 外部発表実績一覧

(1) セミナー、研究会等イベント開催

通番	タイトル	発表者 (代表者)	年月日	開催場所	イベント名称	概要 (対象者 (層、参加人数)、出席者の主な反応等)
1	疾患データベースの統合とオミックス医療への取り組み	田中 博	2007. 9. 26	日本 IBM	製薬企業向け講演会	製薬企業、10名
2	医療データベースと医科学研究の変貌	田中 博	2007. 10. 9	新潟県湯沢町 NASPA ニューオータニ	理化学研究所 理事長フ ァンドワークショップ	

(2) プレス発表、取材対応 (該当なし)

(3) 展示会等出展 (該当なし)

(4) 学会等への口頭発表

通番	タイトル	発表者	学会等の名称	年月日	特記事項
1	文部省ライフサイエンス統合データベースプロジェクトと臨床オントロジー	田中 博	医療知識基盤データベースと用語・言語・知識処理シンポジウム	2008. 3. 13	
2	Bioinformatics and genomics for opening new perspectives for personalised care'	Hiroshi Tanaka	CeHR 2007	2007. 12. 3	

3	TMDU Clinical Omics Database Project - Integration of OMICS data and Clinical Information.	Hiroshi Mizushima, Hiroshi Tanaka	The 7th International Workshop on Advanced Genomics	2007. 11. 27	
4	網羅的臨床情報と網羅的分子情報の統合データベースの構築	水島 洋 田中 博ほか	第 27 回医療情報連合大会	2007. 11. 24	
5	Omics analysis to predict the aggressive recurrence of hepatocellular carcinoma after curative hepatectomy.	Tanaka S, Mahmut Y, Mogushi K, Aihara A, Kudo A, Nakamura N, Ito K, Imoto I, Inazawa J, Miki Y, Mizushima H, Tanaka H, Arii S.	Japan Cancer Association Annual Conference	2007. 10. 3-5	
6	TMDU Clinical Omics Database - Integrating OMICS data and Clinical Information.	Tanaka H, Arii S, Sugihara K, Miki Y, Inazawa J, Mizushima H.	Japan Cancer Association Annual Conference	2007. 10. 3-5	
7	TMDU Clinical Omics Database Project - Integration of OMICS data and Clinical Information.	Mizushima H, Mogushi K, Ohashi W, Araki E, Nishibori M, Arii S, Sugihara K, Miki Y, Inazawa J, Tanaka H.	ISMB/ECCB07	2007. 7. 20-26	
8	Development of TMDU Clinical Omics Database.	Fujisaki A, Araki E, Mizushima H, Tanaka H.	ISMB/ECCB07	2007. 7. 20-26	

(5) 雑誌等への論文寄稿

通番	タイトル	著者名	雑誌等の名称	掲載巻、号、ページ	特記事項
1	オミックス情報と医療情報の網羅的統合データベースの構築	水島 洋、田中 博	計測自動制御学会システム情報部門学術講演会 2007	2007 講演論文集 P. 39-40	
2	TMDU Clinical Omics Database System -Integrating OMICS data and Clinical Information.	Mizushima H. Tanaka H.	The7th International Workshop on Advanced Genomics Abstract book	2007, p. 82	
3	ライフサイエンス分野の統合データベース整備事業分担機関 「統合医科学データベース構築方式の開発」	田中 博	JSBi ニュースレター	Vol. 15	
4	わが国における疾患データベースの統合化について	田中 博	「蛋白質 核酸 酵素」シリーズ 『ライフサイエンス分野の統合デー	未定	

			データベース』		
5	網羅的分子病態データベースとシステム病態学	田中 博	医学の歩み	2008年5月号 掲載予定	
6	Aurora kinase B is a predictive factor for aggressive recurrence of hepatocellular carcinoma after curative hepatectomy	Tanaka S, Arii S, Yasen M, Mogusi K, Su N-T, Zhao C, Imoto I, Eishi Y, Inazawa J, Miki Y, Tanaka H	British Journal of Surgery	in press	
7	Bioinformatics and genomics for opening new perspective for personalized care,	Tanaka, H	“eHealth”, (Globel, B. ed.)	p47-58, IOS Press, 2008	

### 分担機関（東京大学グループ） 整備実績一覧

#### (1) 保有データ情報

##### (1-1) データの種類

①生物種	Homo sapiens
②試料・ライブラリ 一等の種類、数	健常者 900 検体、ナルコレプシー220 検体、多系統萎縮症 200 検体、脳動脈瘤 200 検体、パニック障害 200 検体の 50-90 万の遺伝子型データ。理化学研究所ゲノム医学研究センターが実施、保有する GWAS データを除き、国内で実施される GWAS の大部分のデータを収納する予定である。脳神経疾患関連遺伝子のリシークエンスデータ、ALS に関連する遺伝子の mutation 情報及び、臨床情報。
③測定方法	Affymetrix, Illumina の 30-90 万の SNP タイピングセット
④データの内容	検体の性別、疾患情報などの基本情報、genotype データ、genotype calling 前の画像生データ
⑤その他、特記事項	国立がんセンター吉田輝彦部長の協力を得て、同センター研究所で解析された GWAS データも登録予定

##### (1-2) データソース

①現在のデータ量	健常者のデータ 900 検体、ナルコレプシー220 検体、多系統萎縮症 200 検体、脳動脈瘤 200 検体、パニック障害 200 検体
②データ区分	■自前 ■第三者 ■文献データ ■計算結果等の二次データ □その他
③将来の増加の見込み	アルツハイマー病 (2000 検体)、肝炎 (B 型、C 型 500 検体)、1 型糖尿病 (400 検体)、変形性関節症 (350 検体) などのデータも登録予定。
④権利関係	所有者 (研究代表者に帰属)

	公開 (□可 □否 ■その他 [genotype frequency data や統計解析結果は完全公開する。個々の遺伝子型データ等は、一定の手続きを経て限定された研究者に開示する])
⑤その他、特記事項	

(1-3) データの管理状況

①更新頻度等の管理状況、体制	サーバー管理は日立製作所がおこなっているが、常駐SEがいるわけではない。更新は新たなデータが提供された都度行う。
②その他、特記事項	

(1-4) データベース関係

①DB 管理者数	1
②キュレータ・アナレータ数	0
③データ構造	Relational DB
④DB 管理ソフト	mySQL
⑤サーバーの OS	Redhat enterprise linux ES v4
⑥サーバー規模	Poweredge2900
⑦DB へのアクセス数	未公開
⑧独立 IP 数	1
⑨その他、特記事項	

(2) データ (又はDB) の連結、統合化整備

通番	データ (又はDB) の名称	未公開/未公開	概要 (データの種類 (生物種)・数量 (kB 等)、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述)
1	標準 DB <a href="http://133.11.184.23/snpdbnew/snp_top.php">http://133.11.184.23/snpdbnew/snp_top.php</a>	未公開 (テストサイト有)	日本人健常者の 30 万 SNP 約 200 検体、50 万 SNP 約 500 検体、90 万 SNP 約 200 検体の genotype frequency, allele frequency, Hardy-Weinberg 平衡検定値、ハプロタイプ頻度など。 進捗状況：システムは完成しており、論文が accept され次第公開する。 今後の計画：データを随時登録していく。 (画面の snapshot については別紙参照)
2	GWAS-DB <a href="http://133.11.184.23/cgi-bin/gwasdbnew/gwas_top.cgi">http://133.11.184.23/cgi-bin/gwasdbnew/gwas_top.cgi</a>	未公開 (テストサイト有)	SNP ごとの genotype frequency, allele frequency, call rate、Hardy-Weinberg 平衡検定値、genotypic model, allelic model, additive risk model, recessive model, dominant model など各種遺伝統計値を



			<p>登録している。copy number variation, OMIMなどの他の情報と共に上記計算結果をグラフ表示することが可能である。</p> <p>進捗：システムを構築し、ナルコレプシー、脳動脈瘤、多系統萎縮症などを登録してある。論文がacceptされ次第公開。</p> <p>今後の計画：ユーザーフレンドリーになるように、インターフェース周りの改良を行うとともに、2次スクリーニング結果を登録できるようにデータベースの拡張を行う。また、臨床情報および、臨床情報による層別化の解析なども登録していく。</p> <p>さらに、論文発表、学会発表等により、データのsubmissionを広く呼びかけていく。</p> <p>(画面のsnapshotについては別紙参照)</p>
3	ALS リシーケンス <a href="https://133.11.102.101/resequence/SearchDisease.do?targetId=1">https://133.11.102.101/resequence/SearchDisease.do?targetId=1</a>	未公開（テストサイト有り）	ALS（筋萎縮性側索硬化症）に関するリシーケンスデータベースであり、東京大医学部附属病院で産出したALS関連遺伝子のリシーケンスデータ及び臨床データのほか、論文から抽出したmutationの位置、頻度、家系情報と共に、発症してから何年で人工呼吸器をつけたか、どのような症状か等の臨床情報、更には、蛋白質の2次構造、3次構造などのデータも登録している。 <p>進捗：システムはほぼ完成している。ユーザのフィードバックで改良を加えた後、8月ごろ公開予定。</p> <p>(画面のsnapshotについては別紙参照)</p>

(3) DB基盤システム、ツール等開発成果物の整備

通番	DB基盤システム、ツール等の名称	公開／未公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）
1	Path consistency アルゴリズムによる遺伝子型と表現型のネットワーク解析	未公開	多変量の関連構造をネットワークグラフによって表現型と遺伝子型の関係を表示するソフト

(4) その他の成果物 ((2)、(3) に該当しないもの) (該当なし)

## 分担機関（東京大学グループ） 外部発表実績一覧

### (1) セミナー、研究会等イベント開催

通番	タイトル	発表者 (代表者)	年月日	開催場所	イベント名称	概要（対象者（層、参加人数）、出席者の主な反応等）
1	Path consistency アルゴリズムを用いた相互作用解析	成田 暁	平成 19 年 11 月 22 日～23 日	東京大学 柏 キャンパス	第 2 回インフォマティクス 研究者と医学研究者の交 流会	
2	Affymetrix Genome-Wide Human SNP Nsp/Sty 6.0 による タイピングプラットフォームの構築	西田 奈央	平成 19 年 11 月 22 日～23 日	東京大学 柏 キャンパス	第 2 回インフォマティクス 研究者と医学研究者の交 流会	

### (2) プレス発表、取材対応

通番	タイトル	発表媒体	年月日	特記事項
1	統合データベース	日本バイオインフォマティクス学会ニュースレター	平成 19 年 9 月 1 日	

### (3) 展示会等出展（該当なし）

### (4) 学会等への口頭発表

通番	タイトル	発表者	学会等の名称	年月日	特記事項
1	A search for genetic variants attributing to the risk of formation of intracranial aneurysms.	安野勝史	米国人類遺伝学会第 57 回大会	2007 年 10 月 23～27 日	

### (5) 雑誌等への論文寄稿

通番	タイトル	著者名	雑誌等の名称	掲載巻、号、ページ	特記事項
1	疾患感受性遺伝子の探査の実際：ゲノムワイド関連解析を中心として	宮川 卓、徳永 勝士	最新医学	62 巻、9 月増刊号、150-157	
2	第 4 章 SNP による連鎖解析	成田 暁	BIOWEB 電子出版	in press	

3	第6章 多重検定についての考え方および解決策	成田 暁	BIOWEB 電子出版	in press	
4	第7章 遺伝子間相互作用の検出法	成田 暁	BIOWEB 電子出版	in press	
5	ゲノムワイド関連解析データベースの開発	小池 麻子、西田 奈央 徳永 勝士	蛋白質核酸酵素	in press	
6	ゲノムワイド SNP タイピング技術の現状と将来	西田 奈央、徳永 勝士	医学のあゆみ	in press	

### 補完課題実施機関（理化学研究所）整備実績一覧

#### (1) 保有データ情報

[1 件目]

##### (1-1) データの種類

① 生物種	シロイヌナズナ
② 試料・ライブラリ 一等の種類、数	シロイヌナズナ完全長 (RIKEN Arabidopsis Full-Length, RAFL) cDNA クローン : 245,946 個 (18,090 グループ、全遺伝子の約 60% に相当。世界最大のシロイヌナズナ完全長 cDNA のリソース)
③ 測定方法	
④ データの内容	全長配列データ : 21,130 個 3' 末端側からの端読み配列データ = 151,574 個 5' 末端側からの端読み配列データ = 82,766 個
⑤ その他、特記事項	

##### (1-2) データソース

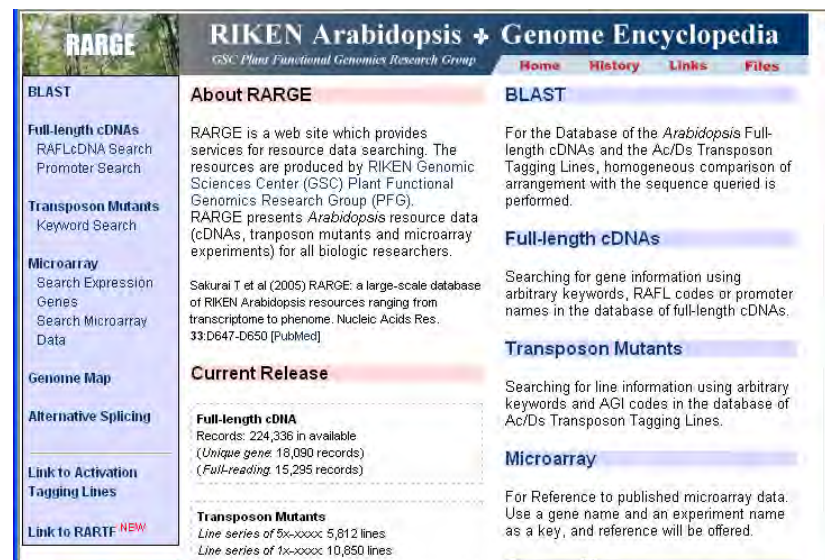
① 現在のデータ量	
② データ区分	■自前 □第三者 □文献データ □計算結果等の二次データ □その他
③ 将来の増加の見込み	なし
④ 権利関係	所有者 ( ) 公開 (■可 □否 □その他 [ ])
⑤ その他、特記事項	

(1-3) データの管理状況

①更新頻度等の管理状況、体制	不定期にアノテーション情報を更新。
②その他、特記事項	

(1-4) データベース関係

(1)DB 管理者数	2 名
(2)キュレータ・アノテータ数	1 名
(3)データ構造	リレーショナルデータベース
(4)DB 管理ソフト	PostgreSQL
(5)サーバの OS	Red Hat Enterprise Linux ES
(6)サーバ規模	1 基
(7)DB へのアクセス数	データベース RARGE 総計 約 5000 ページ/月
(8)独立 IP 数	1
(9)その他、特記事項	画面の様子 RARGE の画面



[2 件目] トランスクリプトーム (タイリングアレイデータ)

## (1-1). データの種類

(1) 生物種	シロイヌナズナ
(2) 試料・ライブラリ 一等の種類、数	シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データ 19種類(各々FおよびRアレイを用いた計6回のハイブリ実験を行う) GEOデータベースに登録されている、シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データの内、3回以上繰り返し実験を行ったものは、わずか4種類のみである。
(3) 測定方法	
(4) データの内容	シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データ 19種類(各々FおよびRアレイを用いた計6回のハイブリ実験を行う) 実験内容：播種後2週間目の植物体を用いた乾燥、低温、塩ストレス、ABA処理、再吸水処理による乾燥ストレスからの回復過程など
(5) その他、特記事項	

## (1-2). データソース

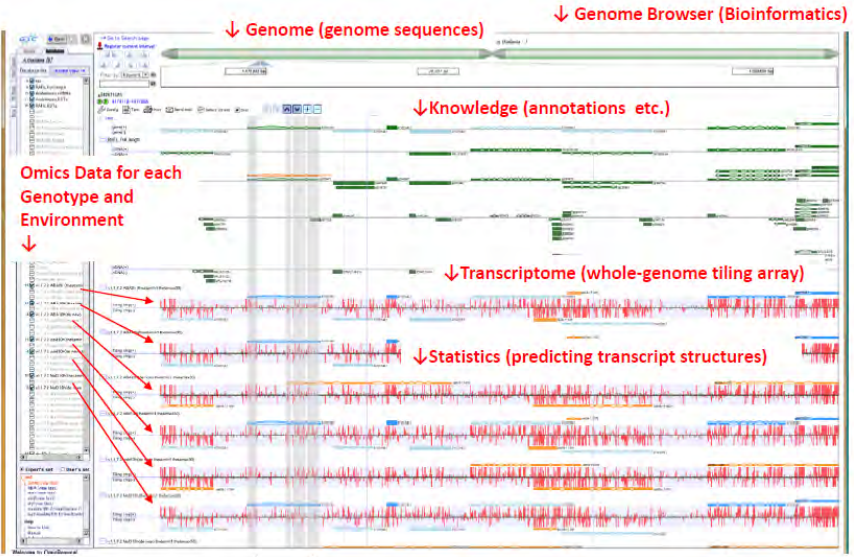
(1) 現在のデータ量	シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データ 19種類
(2) データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input checked="" type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他
(3) 将来の増加の見込み	22年までには、シロイヌナズナ全ゲノムタイリングアレイを用いた発現解析実験を少なくとも100種類以上行う予定である。発現解析だけでなく、今後シロイヌナズナ全ゲノムタイリングアレイを用いてChIP-chipおよびmCIP-chip解析も行い、種々の条件におけるヒストン修飾やメチル化パターンも解析する予定である。また、Solexaシーケンサーを用いたヒストン修飾パターンの解析も行う予定である。
(4) 権利関係	所有者（理研PSC、植物ゲノム機能研究グループ） 公開（ <input type="checkbox"/> 可 <input type="checkbox"/> 否 <input checked="" type="checkbox"/> その他〔論文発表後データを公開する予定である。〕
(5) その他、特記事項	

## (1-3). データの管理状況

(1) 更新頻度等の管理状況、体制	実験データの取得に際して随時追加
(2) その他、特記事項	

## (1-4). データベース関係

(1) DB管理者数	2名
------------	----

(2)キュレータ・アナ データ数	1名
(3)データ構造	ファイル
(4)DB管理ソフト	ファイルシステム
(5)サーバのOS	CentOS
(6)サーバ規模	2ノード
(7)DBへのアクセス数	未公開
(8)独立IP数	
(9)その他、特記事項	画面の様子 OmicBrowse の画面 

[3 件目]

(1-1). データの種類

(1) 生物種	シロイヌナズナ
(2) 試料・ライブラリー等の種類、数	454 シーケンサーを用いた small RNA の大量解析データ 11 種類 454 シーケンサーを用いたシロイヌナズナの small RNA データは、これまでに 6 種類報告されている。
(3) 測定方法	
(4) データの内容	454 シーケンサーを用いた small RNA の大量解析データ 11 種類 用いた植物材料：播種後 2 週間目の植物体を用いた乾燥、低温、塩ストレス、ABA 処理、無処理の植物体等
(5) その他、特記事項	

(1-2). データソース

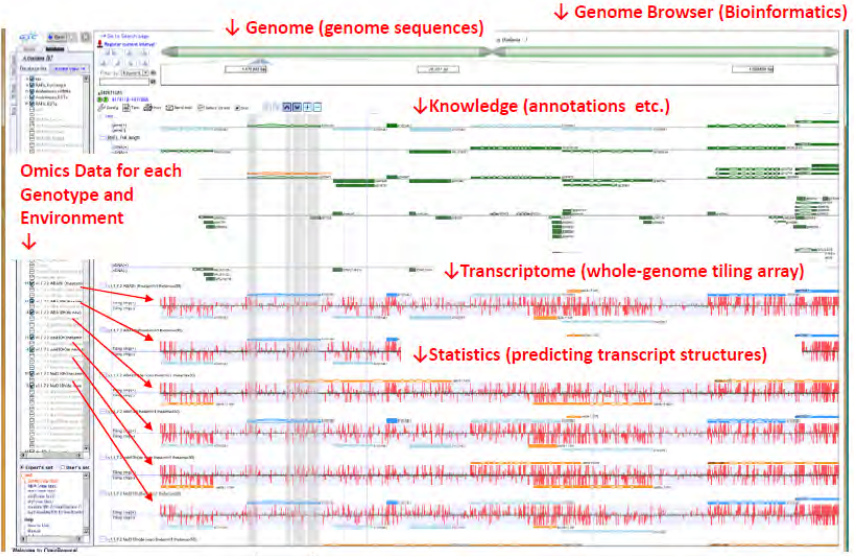
(1) 現在のデータ量	454 シーケンサーを用いた small RNA の大量解析データ 11 種類
(2) データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input checked="" type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他
(3) 将来の増加の見込み	今後、454シーケンサーよりも大量の配列解析が可能なSolexaシーケンサーを用いてsmall RNAの大量解析を行う予定である。
(4) 権利関係	所有者（理研 PSC、植物ゲノム機能研究グループ） 公開（ <input type="checkbox"/> 可 <input type="checkbox"/> 否 <input checked="" type="checkbox"/> その他 [論文発表後データを公開する予定である。]
(5) その他、特記事項	

(1-3). データの管理状況

(1) 更新頻度等の管理状況、体制	実験データの取得に際して随時追加
(2) その他、特記事項	

(1-4). データベース関係

(1) DB 管理者数	1 名
(2) キュレータ・アナデータ数	1 名
(3) データ構造	ファイル
(4) DB 管理ソフト	ファイルシステム
(5) サーバの OS	CentOS

(6) サーバ規模	1 ノード
(7) DB へのアクセス数	未公開
(8) 独立 IP 数	
(9) その他、特記事項	画面の様子 OmicBrowse の画面 

[4 件目]メタボローム（代謝物質質量プロファイル）

(1-1). データの種類

(1) 生物種	シロイヌナズナ
(2) 試料・ライブラリー等の種類、数	<ul style="list-style-type: none"> <li>◇ 野生型および単一遺伝子欠損変異体およそ 50 サンプルの網羅的な代謝物質質量プロファイル</li> <li>◇ 物質の同定に用いる、標準物質のマススペクトルデータ 10,000 スペクトル（約 1000 物質）</li> </ul>
(3) 測定方法	ガスクロマトグラフィー質量分析計 液体クロマトグラフィー質量分析計 キャピラリー電気泳動質量分析計
(4) データの内容	質量分析計より出力されるマススペクトル
(5) その他、特記事	計測方法が確立しているため、シロイヌナズナの完全長 cDNA の過剰発現体または遺伝子欠損変異体約数百種類のデータ



項	を取得することも可能である。また各植物体について出来るだけ多く（数個体以上）のサンプルを計測することが望ましい。
---	--

(1-2). データソース

(1)現在のデータ量	1サンプルあたり数百メガバイト
(2)データ区分	■自前 □第三者 □文献データ □計算結果等の二次データ □その他
(3)将来の増加の見込み	データ数、量を増やすことは容易だが、効率よくアノテーションする作業が困難
(4)権利関係	所有者（理研） 公開（ <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ]）
(5)その他、特記事項	全て公開できるが、アノテーション作業を解さない限り、データだけ外に出しても価値が低い

(1-3). データの管理状況

(1)更新頻度等の管理状況、体制	データは新規代謝物が同定される度にアップデートする必要がある。現在は、計測グループが自前で維持管理
(2)その他、特記事項	管理するためのデータベースを現在構築中

[5件目]フェノーム（トランスポゾン・タグライン）

(1-1). データの種類

(1)生物種	シロイヌナズナ ( <i>Arabidopsis thaliana</i> )
(2)試料・ライブラリー等の種類、数	シロイヌナズナのトランスポゾン・タグライン 18,000 系統と、全てのラインに関するトランスポゾン挿入位置情報。シロイヌナズナ 26,000 遺伝子のうち 5,000 以上の遺伝子に関する変異を含んでいると推測される。
(3)測定方法	トランスポゾン挿入部位近傍の塩基配列の決定
(4)データの内容	変異体番号、トランスポゾン挿入位置情報、近傍遺伝子情報
(5)その他、特記事項	cDNA 情報や他の公共データベースとリンクさせている。

(1-2). データソース

(1)現在のデータ量	トランスポゾン・タグライン 18,000 系統に関する情報
(2)データ区分	■自前 □第三者 □文献データ □計算結果等の二次データ □その他
(3)将来の増加の見込み	現在のところ予定はありません。

(4) 権利関係	所有者（理化学研究所植物科学研究センター） 公開（ <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ]）
(5) その他、特記事項	http://rarge.gsc.riken.go.jp/

(1-3). データの管理状況

(1) 更新頻度等の管理状況、体制	理化学研究所植物科学研究センターのサーバーで維持されている。
(2) その他、特記事項	

(1-4). データベース関係

(1) DB 管理者数	2 名
(2) キュレータ・アナ データ数	1 名
(3) データ構造	リレーショナルデータベース
(4) DB 管理ソフト	PostgreSQL
(5) サーバの OS	Red Hat Enterprise Linux ES
(6) サーバ規模	1 基
(7) DB へのアクセス数	データベース RARGE 総計 約 5000 ページ/月
(8) 独立 IP 数	1

(9) その他、特記事項	画面の様子
--------------	-------

**RARGE** RIKEN Arabidopsis Genome Encyclopedia  
GSC Plant Functional Genomics Research Group

Home History Links Files

**Transposon Mutants:**

"RIKEN Arabidopsis Transposon mutants" is a series of mutant lines which have a Ds transposon in the genome of *Arabidopsis thaliana* Nössen ecotype (background by Fedoroff and Smith). This web page provides information on the mutants produced in our laboratory. Each mutant line is assigned by stipulated line codes (ex. 13-4480-1). We determined the flanking sequences of Ds insertion for each independent line. Transposon insertion sites of mutants were estimated by a BLASTN homology against the genome sequence database of *Arabidopsis thaliana* Columbia ecotype. The closest genes (predicted by AGI) to the transposon insertion sites were picked up. The results of the BLASTP homology search against the nr database of NCBI for the closest genes have been collected for keyword searches.

**Structure of the transposon**

Ds11; 51; 52	325 bp	360 bp	466 bp
Ds12; 54; 15; 53	378 bp	281 bp	466 bp
Ds13; 16	512 bp	541 bp	466 bp

G-edge → GUS → 19S-hygro → H-edge  
↑ 35S core in Ds15, 16, 52, 53

[6 件目]フェノーム（ノックアウト・表現型）

(1-1). データの種類

(1) 生物種	シロイヌナズナ ( <i>Arabidopsis thaliana</i> )
(2) 試料・ライブラリー等の種類、数	シロイヌナズナの 4000 遺伝子の変異体に関する表現型情報。シロイヌナズナ 26,000 遺伝子のうち 4,000 遺伝子に関する変異体を調べている。
(3) 測定方法	系統的な表現型解析
(4) データの内容	変異体番号、トランスポゾン挿入位置情報、挿入変異遺伝子情報、表現型の画像データ
(5) その他、特記事項	他の公共データベースとリンクさせている。

(1-2). データソース

(1) 現在のデータ量	トランスポゾン・タグライン 4000 系統 (4000 遺伝子) の表現型に関する情報
(2) データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他
(3) 将来の増加の見込み	現在は地上部形態異常に関して載せているが、他の表現型項目についてもデータを収集中である。
(4) 権利関係	所有者 (理化学研究所植物科学研究センター) 公開 ( <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ])
(5) その他、特記事項	<a href="http://range.gsc.riken.jp/phenome/">http://range.gsc.riken.jp/phenome/</a>

(1-3). データの管理状況

(1) 更新頻度等の管理状況、体制	理化学研究所植物科学研究センターのサーバで維持されている。
(2) その他、特記事項	

(1-4). データベース関係

(1) DB 管理者数	2 名
(2) キュレータ・アナデータ数	1 名
(3) データ構造	リレーショナルデータベース、ファイル
(4) DB 管理ソフト	PostgreSQL、ファイルシステム
(5) サーバの OS	Red Hat Enterprise Linux ES
(6) サーバ規模	1 基

(7) DB へのアクセス数	データベース RARGE 総計 約 5000 ページ/月
(8) 独立 IP 数	1
(9) その他、特記事項	画面の様子

### RIKEN Arabidopsis Phenome Information Database

[Home](#) [Search](#) [Line list](#)

#### About this database

RIKEN Arabidopsis Phenome Information Database (RAPID) is a searchable site of phenotypic data in transposon-insertional mutants.



We selected about 4,000 transposon insertion lines which have the Ds transposon in gene coding region, and observed visible phenotypes systematically depend on growth stage. Phenotypic

## [7 件目]フェノーム (アクティベーション・タグライン)

### (1-1). データの種類

(1) 生物種	シロイヌナズナ
(2) 試料・ライブラリ一等の種類、数	シロイヌナズナ完全長 cDNA 遺伝子高発現型変異体は、理研オリジナルの変異体であり、約 1 万の遺伝子リソースを網羅する。これは、現在報告されている遺伝子の 40%にあたる。 シロイヌナズナ Activation tagging 変異体系統は、7 万系統あり、シロイヌナズナのほぼすべての遺伝子の活性化をしている数と考えられる。
(3) 測定方法	塩基配列決定による遺伝子情報 目視及び計測機器 (光合成、色素吸収) による変異形質の情報
(4) データの内容	種子番号 遺伝子番号 遺伝子アノテーション情報 形質情報 (光合成、色素、形態) 変異体画像情報
(5) その他、特記事項	完全長 cDNA 高発現変異体系統は、作成した 1 万系統についてすべて遺伝子情報を記載している。

(1-2). データソース

(1)現在のデータ量	シロイヌナズナ完全長 cDNA 高発現型変異体 1 万系統 Activation tagging 変異体 7 万系統
(2)データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input checked="" type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他
(3)将来の増加の見込み	種々の形質計測により情報増加の可能性がある。
(4)権利関係	所有者（理研・NEC ソフト ） 公開（ <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ]）
(5)その他、特記事項	<a href="http://amber.gsc.riken.jp/act/top.php">http://amber.gsc.riken.jp/act/top.php</a>

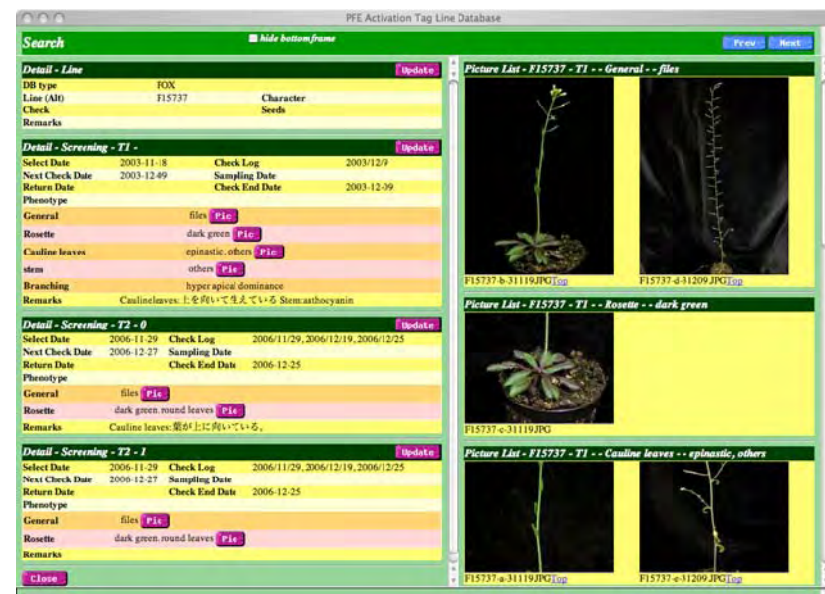
(1-3). データの管理状況

(1)更新頻度等の管理状況、体制	1 週間に 1 回アノテーション情報の更新を行っている。
(2)その他、特記事項	

(1-4). データベース関係

(1)DB 管理者数	1 名
(2)キュレータ・アノデータ数	1 名
(3)データ構造	RDB
(4)DB 管理ソフト	PostgreSQL
(5)サーバの OS	Red Hat Enterprise Linux ES (外部公開用) Red Hat Enterprise Linux AS (内部用)
(6)サーバ規模	2CPU (Pentium Xeon) (外部公開用) 4CPU (Pentium Xeon) (内部用)
(7)DB へのアクセス数	登録外部ユーザ数 884 (2007 年登録 82、2006 年登録 142、2005 年登録 141、2004 年登録 327、2003 年登録 192)
(8)独立 IP 数	1 (外部公開用)、1 (内部用)
(9)その他、特記事項	画面の様子

--	--



[8 件目]リソース情報

(1-1). データの種類

(1) 生物種	シロイヌナズナ
(2) 試料・ライブラリー等の種類、数	完全長 cDNA クローン (RAFL clone) シロイヌナズナ (エコタイプ: Columbia) のほぼ全ての転写領域をカバー
(3) 測定方法	cDNA 塩基配列の全長もしくは両端を決定
(4) データの内容	※記録しているデータ項目 (例えば、試料番号、遺伝子名、発現データ (画像) 等) リソース番号、クローン番号、塩基配列のアクセッション番号、遺伝子コード領域の AGI 番号、塩基配列
(5) その他、特記事項	

(1-2). データソース

(1) 現在のデータ量	246, 203 エントリー
(2) データ区分	<input type="checkbox"/> 自前 <input checked="" type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他
(3) 将来の増加の見込み	あり
(4) 権利関係	所有者 (理研ゲノム科学総合研究センター)

	公開 ( <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ] )
(5) その他、特記事項	http://www.brc.riken.jp/lab/epd/catalog/cdnac1one.html http://saber.epd.brc.riken.jp/sabre7/SABRE0101.cgi

(1-3). データの管理状況

(1) 更新頻度等の管理状況、体制	データの増加に応じて、年に数回までの更新あり 外部業者に運用を業務委託している
(2) その他、特記事項	

(1-4). データベース関係

(1) DB 管理者数	理研側担当者2名、業者側数名
(2) キュレータ・アナテータ数	なし
(3) データ構造	テキストファイル、リレーショナルデータベース
(4) DB 管理ソフト	PostgreSQL
(5) サーバの OS	Linux
(6) サーバ規模	パーソナルコンピュータ
(7) DB へのアクセス数	計測値なし
(8) 独立 IP 数	2
(9) その他、特記事項	画面の様子



[9 件目]ストラクチュローム (高等動植物等由来)

(1-1). データの種類

(1) 生物種	動物、植物、微生物
(2) 試料・ライブラリ一等の種類、数	シロイヌナズナ：40 程度、その他：2500 程度
(3) 測定方法	NMR (核磁気共鳴) や、X 線による構造解析等
(4) データの内容	試料番号、タンパク質名、ドメイン名、PDBID、生物種、解析実験装置、発現系、試料の詳細 (全長タンパク質の生物学的意味、ドメインの機能、構造上の特性、基質結合ポケット・相互作用部位)、PDB に登録した蛋白質立体構造データ、構造決定のもととなった測定データ
(5) その他、特記事項	日本語記載

(1-2). データソース

(1) 現在のデータ量	シロイヌナズナ：40 程度、その他：2500 程度
(2) データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他
(3) 将来の増加の見込み	レコードは随時アップデートで、レコード数は随時追加される。
(4) 権利関係	所有者 (独立行政法人理化学研究所) 公開 ( <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ])
(5) その他、特記事項	


(1-3). データの管理状況

(1) 更新頻度等の管理状況、体制	常時管理。
(2) その他、特記事項	

(1-4). データベース関係

(1) DB 管理者数	
(2) キュレータ・アナテータ数	
(3) データ構造	
(4) DB 管理ソフト	ファイルメーカーPro



(5) サーバの OS	
(6) サーバ規模	
(7) DB へのアクセス数	
(8) 独立 IP 数	
(9) その他、特記事項	画面の様子 

[10 件目] ストラクチュローム (微生物由来)

(1-1). データの種類

(1) 生物種	(試料調整・結晶化・回折実験データベース) 9 種類 ( <i>Thermus thermophilus</i> HB8, <i>Pyrococcus horikoshii</i> OT3, <i>Escherichia coli</i> K-12, <i>Aeropyrum pernix</i> K1, <i>Sulfolobus tokodaii</i> strain7, <i>Aquifex aeolicus</i> VF5, <i>Geobacillus kaustophilus</i> HTA426, <i>Thermotoga maritima</i> MSB8, <i>Methanocaldococcus jannaschii</i> DSM 2661) (変異体構造解析データベース) 2 種類 ( <i>Thermus thermophilus</i> HB8, <i>Pyrococcus horikoshii</i> OT3) (重原子データベース) 不明だが多数
(2) 試料・ライブラリ等の種類、数	(試料調整・結晶化・回折実験データベース) 総レコード数 11190 のうち発現プラスミドがあるもの 11021 (98%)。回折画像数 300 (変異体構造解析データベース) 変異体総数 241 種類のうち、99 種類を構造決定済み (41%)。 <i>Pyrococcus horikoshii</i> OT3 由来 PH0725 蛋白質 (265 残基) については、変異体 179 種類をプラスミド構築し 79 種類の結晶構造を決定済み。 <i>Thermus thermophilus</i> HB8 由来 TTHB049 蛋白質 (177 残基) については、変異体 62 種類をプラスミド構築し 20 種類を解析済み。両蛋白質を通じ、セレノメチオニン化のための Leu-Met 変異はほぼ網羅し

	<p>ている。</p> <p>(重原子データベース) 重原子を結合した蛋白質の情報 784 件を収録する。特に水銀に関しては 351 件と豊富なデータを有する。現在のところ 22 種類の重原子をカバーしている。</p>
(3) 測定方法	<p>(試料調整・結晶化・回折実験データベース) 実験データベースであるため多種多様である。特に発現精製は多様な機器の出力データ等を含む。結晶化は主に結晶化ロボット TERA の出力データである。</p> <p>(変異体構造解析データベース) X線結晶構造解析により変異体蛋白質の結晶構造を決定。</p> <p>(重原子データベース) 重原子を結合した蛋白質の結晶構造を X線結晶構造解析により決定。</p>
(4) データの内容	<p>(試料調整・結晶化・回折実験データベース)</p> <p>(基礎情報)</p> <ol style="list-style-type: none"> <li>1. 試料蛋白質の由来生物種名、遺伝子名、吸光係数、分子量、等電点など、いずれも計算結果等の二次データ。</li> <li>2. 構築プラスミドデータ (自前のデータ)。ホスト、ベクター、予備発現結果 (5 段階の発現ランク)。</li> </ol> <p>(発現精製)</p> <ol style="list-style-type: none"> <li>1. 培養情報 (自前のデータ)。発現データ (画像と 5 段階の発現ランク)、発現の諸条件 (誘導状態、培養時間、培養温度、培養量溶液量等)。</li> <li>2. 精製情報 (自前のデータ)。各タンパク質について精製の諸条件 (懸濁方法、超音波破碎方法、遠心時間、各タンパク質に最適なカラム情報) 加えて精製タンパク質の吸収スペクトルの画像、精製蛋白質の SDS および Native ページ画像、精製蛋白質の収量等。精製タンパク質の DLS 測定結果 (画像および数値)。選択カラム情報の詳細について: カラム名、緩衝液名、フラクションサイズ、フロー速度、グラジエント方法、溶出濃度、カラムチャート (A280、イオン強度等) の画像、各フラクションの SDS ページ画像。</li> </ol> <p>(結晶化)</p> <p>結晶化ロボット TERA の出力 (自前のデータ)。精製タンパク質の結晶化スクリーニング情報 (結晶化条件、観察画像、10 段階のスコア)。</p> <p>(回折実験)</p> <p>回折実験データ (自前のデータ)。回折画像データとその計算処理のための各種パラメータ。</p> <p>(変異体構造解析データベース) 二種のタンパク質について網羅的な変異体構造解析データベース。結晶化データ、回折実験データ (画像データ、分解能、測定条件等)、回折画像処理データ、精密化データ、構造座標データ (PDB)。</p> <p>(重原子データベース) 自前の構造解析から得たデータ。さらに、文献データおよび登録されている PDB からの計算等による二次データ。内容は重原子実験データで、タンパク質名、重原子名、重原子試薬名、実験方法、沈澱剤名、緩衝液名、pH、文献名、重原子結合サイトの二次構造等。インターフェースとして重原子選択予測機能等。</p>
(5) その他、特記事項	

## (1-2). データソース

(1)現在のデータ量	(試料調整・結晶化・回折実験データベース) 基礎データ収録数 11190 件。構築プラスミドデータ 11021 件 (変異体を含む)。培養情報 5878 件 (重複あり)。精製情報 4401 件 (重複あり)。結晶化情報 1300752 件、回折画像 300 件。容量は、培養精製情報 120GB、結晶化情報 1.3TB。回折実験情報 1.5TB。 (変異体構造解析データベース) サンプル数 241 件。回折データ 262 件。容量は 1.1TB。 (重原子データベース) 784 件。
(2)データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input checked="" type="checkbox"/> 文献データ <input checked="" type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他 (下欄に詳細を記述)
(3)将来の増加の見込み	3 データベースとも増加の見込みあり。
(4)権利関係	所有者 (独立行政法人理化学研究所) 公開 ( <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ])
(5)その他、特記事項	(試料調整・結晶化・回折実験データベース) 内部利用で公開はしていない。 (変異体構造解析データベース) 内部利用で公開はしていない。 (重原子データベース) HATODAS Ver1 は <a href="http://hatodas.harima.riken.go.jp/">http://hatodas.harima.riken.go.jp/</a> で公開。Ver 2 は内部利用。

## (1-3). データの管理状況

(1)更新頻度等の管理状況、体制	(試料調整・結晶化・回折実験データベース) 実験生データのため毎日。 (変異体構造解析データベース) 実験生データのため毎日。 (重原子データベース) 不定期。
(2)その他、特記事項	

## (1-4). データベース関係

(1)DB 管理者数	1 人 (浅田)
(2)キュレータ・アナレータ数	キュレータ 1 人 (浅田) アナレータ 2 人 (国島・菅原) その他オペレーターとして実験補助 2 人程度を採用する予定
(3)データ構造	いずれもリレーショナルデータベース + Web ユーザーインターフェース (試料調整・結晶化・回折実験データベース) サーバ: Apache, Tomcat (JAVA)。バイナリーデータの場合、リンクのみデータベースに格納実体はディスクサーバーに格納 (変異体構造解析データベース) サーバ: Microsoft IIS (ASP.NET)。 (重原子データベース) サーバ: Apache (Ver 1:Perl CGI, Ver 2: Tomcat)。

(4) DB 管理ソフト	(試料調整・結晶化・回折実験データベース) Oracle (変異体構造解析データベース) SQL Server (重原子データベース) Postgres
(5) サーバの OS	(試料調整・結晶化・回折実験データベース) Linux (変異体構造解析データベース) Windows Server (重原子データベース) Linux
(6) サーバ規模	(試料調整・結晶化・回折実験データベース) PC 4 台 (変異体構造解析データベース) PC 1 台 (重原子データベース) PC 1 台
(7) DB へのアクセス数	(試料調整・結晶化・回折実験データベース) 内部利用のため実験の進捗に依存 (変異体構造解析データベース) 内部利用のため実験の進捗に依存 (重原子データベース) 1100 検索/3 年
(8) 独立 IP 数	(試料調整・結晶化・回折実験データベース) 内部利用のためグローバル IP は取得していない。 (変異体構造解析データベース) 内部利用のためグローバル IP は取得していない。 (重原子データベース) 1
(9) その他、特記事項	

## (2) データ (又はDB) の連結、統合化整備

通番	データ (又はDB) の名称	公開／未公開	概要 (データの種類 (生物種)・数量 (kB 等)、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述)
1	トランスクリプトーム (タイリングアレイデータ)	未公開	公開整備は次年度につき該当なし
2	454 シーケンサーを用いた small RNA 解析データ	未公開	公開整備は次年度につき該当なし

## (3) DB 基盤システム、ツール等開発成果物の整備

通番	DB 基盤システム、ツール等の名称	公開／未公開	概要 (主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述)
1	植物統合 DB 計算機環境	未公開	<b>主機能・特徴点</b> =データデポジット用のハードウェア環境の整備・4 年間のデータ編纂作業に十分な容量のデータストレージの構築。 <b>進捗状況</b> =アノテーション等のデータを安全に保持できるデータストレージの準備を行い、システム全体の運用体制を整えた。データストレージは4年間のデータ編纂作業に十分な容量を準備した。

		またサービス品質を上げるための工夫として災害時の運用維持対策を考慮し、理研の横浜研究所および和光研究所で同じ内容のストレージを設置し同期をとるようにシステムを構築した。 今後の計画＝アノテーションサーバとなる中心の計算機のソフトウェア環境を整備していく予定である。
--	--	---

(4) その他の成果物 ((2)、(3) に該当しないもの) (該当なし)

### 補完課題実施機関 (理化学研究所) 外部発表実績一覧

(1) セミナー、研究会等イベント開催

通番	タイトル	発表者 (代表者)	年月日	開催場所	イベント名称	概要 (対象者 (層、参加人数)、出席者の主な反応等)
1	データベースの中に築く生物像 —生命現象を読み解くためのデータベースとWEBリソース—	豊田哲郎、 中村保一、 青木考	2008年3月 20日	札幌コンベンションセンター	第49回日本植物生理学会年会シンポジウム	対象層＝植物研究者、 参加人数＝約80名 出席者の主な反応＝総合討論にて著作権など実務問題が議題に上がった。

(2) プレス発表、取材対応 (該当なし)

(3) 展示会等出展 (該当なし)

(4) 学会等への口頭発表

通番	タイトル	発表者	学会等の名称	年月日	特記事項
1	シロイヌナズナのおミックス統合解析とデータベース	豊田哲郎	理研シンポジウム「植物トランスクリプトーム解析の新展開：シロイヌナズナワークショップ2007	2007年12月10日	
2	シロイヌナズナのおミックス統合解析とデータベース	豊田哲郎	第30回日本分子生物学会年会・第80回日本生化学会大会 (BMB2007)	2007年12月11-14日	
3	シロイヌナズナ miRNA の in silico ターゲッ	神沼 英里 松井 章浩,	第30回日本分子生物学会年	2007年12月11-14日	

	ト探索における特徴パラメータの決定木分析	栗原 志夫, 諸澤 妙子, 関 原明, 豊田 哲郎	会・第 80 回日本生化学会大会 (BMB2007)		
4	シロイヌナズナ LucTag ラインを用いた遺伝子発現定量解析	神沼 英里、吉積 毅、栗山 朋子、越 智子、武藤 周、松井 南、豊田哲郎	人工知能学会第 36 回分子生物情報研究会 (SIG-MBI)	2008 年 1 月 11 日	
5	セマンティックウェブ技術による理研データベースの統合化	豊田哲郎	「生命をはかる」研究会第 23 回研究会	2008 年 2 月 18 日	
6	デジタル whole mount in situ hybridization に向けたシロイヌナズナ LucTag ラインの遺伝子発現時系列画像の定量化	神沼 英里、吉積 毅、栗山 朋子、越 智子、武藤 周、豊田哲郎、松井 南、	第 49 回日本植物生理学会年会 (2008)	2008 年 3 月 20-22 日	
7	シロイヌナズナのオミックス統合データベース：オミックス進化論とデータベースの役割	豊田哲郎	第 49 回日本植物生理学会年会 (2008)	2008 年 3 月 20-22 日	

(5) 雑誌等への論文寄稿

通番	タイトル	著者名	雑誌等の名称	掲載巻、号、ページ	特記事項
1	コンピュータの中の脳：情報基盤の進化論	豊田 哲郎	生体の科学	4 巻 1 号、pp.20-32	

補完課題実施機関（産業技術総合研究所） 整備実績一覧

(1) 保有データ情報

(1-1) データの種類

①生物種	細菌、ウイルス、植物、線虫、ハエ、マウス、ラット、ヒト
②試料・ライブラリ等の種類、数	<ul style="list-style-type: none"> <li>・産総研・糖鎖関連データベース (ヒト、マウス、ラット、ショウジョウバエの、線虫、植物(シロイネナズナ)、酵母糖鎖合成に関する 163 反応、文献で報告された基質の情報を 260 の反応、ヒトの糖鎖関連遺伝子：189 種類、マウスの糖鎖関連遺伝子：182 種類、ラットの糖鎖関連遺伝子：172 種類、ショウジョウバエの糖鎖関連遺伝子：74 種類、線虫の糖鎖関連遺伝子：63 種類、植物(シロイネナズナ)の糖鎖関連遺伝子：13 種類、酵母の糖鎖関連遺伝子：18 種類)</li> <li>・産総研・レクチンデータベース (哺乳類、両生類、魚類、植物、細菌、ウイルスなど計 200 種類ほどの生物、レクチン：266 種、相互作用データ：131 種)</li> <li>・産総研・糖タンパク質データベース (線虫：プロテアーゼで消化したペプチドを 3 種のレクチン (conA、小麦胚芽アグ</li> </ul>

	<p>ルチニン (WGA)、線虫ガレクチン 6) で独立に捕集した糖ペプチドを試料とした。同定された糖タンパク質数は 829 種、糖鎖付加部位は 1465 カ所。ゲノム解析から予測される糖タンパク質のポテンシャル数は約 6,000 種なので、カバー率はおよそ 14%。他の研究者による報告は 100 種に満たない。マウス：脳、肝臓、肺、腎臓、精巣から変性条件下で抽出したタンパク質をトリプシン消化して得られた可溶性ペプチドを出発材料とした。3 ないし 5 種 (conA、RCA120、AAL の 3 種は共通に、肝臓については WGA と SSA も含めた) のレクチンカラムで捕集した糖ペプチドを試料とした。同定されたタンパク質は約 2,300 種、糖鎖付加部位は約 4,500 箇所。ポテンシャル糖タンパク質は約 10,000 種なので、カバー率は約 23%。実験データに基づく糖タンパク質としては約 200 種が既報。糖鎖付加部位が判明しているものは 100 種に満たない。)</p> <p>・産総研・糖鎖スペクトルデータベース (質量分析スペクトル 2756 スペクトル。糖鎖の種類としては、300 種類程度。)</p>
③測定方法	<p>・産総研・糖鎖関連データベース クローニングと酵素活性測定と独自に糖鎖を合成した反応経路 (<i>in vitro</i>)</p> <p>・産総研・レクチンデータベース フロンタル・アフィニティークロマトグラフィーと呼ばれるレクチンをカラムに固定化し糖鎖とレクチンが相互作用すると遅れて溶出される現象を数値化</p> <p>・産総研・糖タンパク質データベース 捕集した糖ペプチドを、安定同位体で標識した水 (<math>H_2^{18}O</math>) 中で N-グリカナーゼ処理することによって、糖鎖切除と同時に糖鎖付加位置を特異的に標識 (IGOT) した後、LC-MS (質量分析) 法で分析、同定した。</p> <p>・産総研・糖鎖スペクトルデータベース リコンビナント酵素を用いて糖鎖を合成し、多段階タンデム質量分析 (MALDI-QIT-TOF MS) で計測</p>
④データの内容	<p>・産総研・糖鎖関連データベース ドナーとアクセプター構造情報、酵素の情報、アクセプターにしなかった構造情報</p> <p>・産総研・レクチンデータベース レクチン名、由来生物種名、由来器官、タンパク質ファミリー名、一次構造、立体構造、GenBank アクセッション番号、レクチン-糖鎖間相互作用情報</p> <p>・産総研・糖タンパク質データベース 遺伝子名 (NCBI など公共データベースの ID)、タンパク質名、糖鎖付加位置、部位ごとの結合したレクチンの種類および検出された組織、タンパク質の特性 (分子量、等電点；共に計算値) と構造モチーフ (シグナル配列、膜貫通領域)。遺伝子オントロジー情報。</p> <p>・産総研・糖鎖スペクトルデータベース 糖鎖 ID、糖鎖構造、質量分析スペクトル (画像イメージ)、プリカーサーイオンの <math>m/z</math> 値</p>
⑤その他、特記事項	

(1-2) データソース

①現在のデータ量	「1-1) データの種類 ②試料・ライブラリー等の種類、数」を参照のこと
----------	--------------------------------------

②データ区分	<ul style="list-style-type: none"> <li>・産総研・糖鎖関連データベース</li> <li>■自前 □第三者 ■文献データ □計算結果等の二次データ □その他</li> <li>・産総研・レクチンデータベース</li> <li>■自前 □第三者 ■文献データ □計算結果等の二次データ □その他</li> <li>・産総研・糖タンパク質データベース</li> <li>■自前 ■第三者 □文献データ □計算結果等の二次データ □その他</li> <li>・産総研・糖鎖スペクトルデータベース</li> <li>■自前 □第三者 □文献データ □計算結果等の二次データ □その他</li> </ul>
③将来の増加の見込み	論文等で公表したデータを優先的に公開する。サプリメントデータも公開する。
④権利関係	所有者（ 産業技術総合研究所と共同研究先の企業等 ） 公開（■可 □否 ■その他 [ 整備終了後、順次公開 ]） <a href="http://jcgddb.jp">http://jcgddb.jp</a> （糖鎖ポータルサイト） <a href="http://riodb.ibase.aist.go.jp/rcmg/ggdb/">http://riodb.ibase.aist.go.jp/rcmg/ggdb/</a> （糖鎖関連遺伝子データベース） 他準備が整い次第公開。
⑤その他、特記事項	-

(1-3) データの管理状況

①更新頻度等の管理状況、体制	3～4ヶ月に1度更新する。運営事務局が責任を持って更新する。
②その他、特記事項	-

(1-4) データベース関係

①DB 管理者数	プロジェクト専従雇用3名+派遣1名、他産総研雇2名
②キュレータ・アナテータ数	10名
③データ構造	RDBMS、XMLDB
④DB 管理ソフト	Oracle 11, eXist
⑤サーバの OS	Solaris 10
⑥サーバ規模	Web サーバ Fujitsu 社製 T5120
⑦DB へのアクセス数	公開している糖鎖関連遺伝子データベースに関して2008年1月～3月、約3000アクセス
⑧独立 IP 数	不明
⑨その他、特記事項	-

(2) データ（又はDB）の連結、統合化整備



通番	データ（又はDB）の名称	公開／未公開	概要（データの種別（生物種）・数量（kB等）、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述）
1	<u>産総研・糖鎖関連データベース</u> <a href="http://riodb.ibase.aist.go.jp/rcmg/ggdb/">http://riodb.ibase.aist.go.jp/rcmg/ggdb/</a>	公開	2007年11月23日 GGDB version2をリニューアル。酵素の基質特異性の情報を配列データベースに統合。定期的に基質特異性の情報を追加。新しく同定された遺伝子情報を追加。2008年1月～3月、約3000アクセス。
2	<u>産総研・レクチンデータベース</u> <a href="http://jcgdb.jp">http://jcgdb.jp</a> からアクセス可能にする	未公開	構造同定のツールとして平成20年度外部機関のDBと統合を行う。セキュリティチェック並びに対策を終えた後に公開。
3	<u>産総研・糖タンパク質データベース</u> <a href="http://jcgdb.jp">http://jcgdb.jp</a> からアクセス可能にする	未公開	NEDOプロジェクトでヒトのデータを採取中のためプロジェクト終了までに、ヒトのグライコプロテオミクスの糖鎖の付加位置の情報を公開する予定。セキュリティチェック並びに対策を終えた後に線虫とマウスの情報を公開。
4	<u>産総研・糖鎖スペクトルデータベース</u> <a href="http://jcgdb.jp">http://jcgdb.jp</a> からアクセス可能にする	未公開	平成20年度には構造同定方法論を統合しどの技術であればある糖鎖を同定できるか方法の選択と同定をサポートするシステムを構築する。セキュリティチェック並びに対策を終えた後に公開。

(3) DB基盤システム、ツール等開発成果物の整備

通番	DB基盤システム、ツール等の名称	公開／未公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）
1	DBのスキーマ定義	未公開	DBを連携するにあたり共通する項目の標準化 ※開発途中
2	構造の定義・描画ツール	未公開	糖鎖構造IDの共通化（XML, 画像） ※開発途中
3	構造の呼び名の統一	未公開	Sialyl Lewis x等の糖鎖構造に関する有名な名称と構造を辞書化 ※開発途中

(3) その他の成果物（(2)、(3)に該当しないもの）（該当なし）

補完課題実施機関（産業技術総合研究所） 外部発表実績一覧

(1) セミナー、研究会等イベント開催（該当なし）

(2) プレス発表、取材対応

通番	タイトル	発表媒体	年月日	特記事項
1	特集レポート「日本の強み糖鎖科学オールジャパン体制へ」	日経BP社BTJジャーナル	2008年3月号	

2	タンパク質の機能左右、「糖鎖」総合データベース、産総研、今春ネット公開	日本経済新聞 朝刊、21 ページ	2008年1月21日
---	-------------------------------------	------------------	------------

(3) 展示会等出展 (該当なし)

(4) 学会等への口頭発表

通番	タイトル	発表者	学会等の名称	年月日	特記事項
1	日本糖鎖科学のポータルサイト	鹿内俊秀	第5回 糖鎖科学コンソーシアムシンポジウム(JCGG)	2008年11月27日	
2	「糖鎖データベースの紹介」と「糖鎖産業技術フォーラムの具体的取組」	新聞陽一	糖鎖産業技術フォーラム(GLIT)設立総会 & 第1回 糖鎖産業技術フォーラム	2008年1月23日	

(5) 雑誌等への論文寄稿 (該当なし)

### 補完課題実施機関 (国立遺伝学研究所) 整備実績一覧

(1) 保有データ情報

(1-1) データの種類

①生物種	多数
②試料・ライブラリー等の種類、数	多数
③測定方法	従来型シーケンサーからの出力、一部、第2世代シーケンサーからのもの少数。
④データの内容	塩基配列決定時の波形データ、結果の配列データ、その生物種、目的分類 (WGS、ゲノム、等)、他
⑤その他、特記事項	

(1-2) データソース

①現在のデータ量	2 生物種、2,542,476 件
②データ区分	<input type="checkbox"/> 自前 <input checked="" type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他
③将来の増加の見込み	サービス開始後、急速に増加の見通し。サービス後1年程度で1TB以上を見込み
④権利関係	所有者 (それぞれのデータ登録者に帰属)

	公開 ( <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ] )
⑤その他、特記事項	

(1-3) データの管理状況

①更新頻度等の管理状況、体制	現在は開発・試験フェーズのため必要に応じて更新。運用開始後は、登録要求に応じて随時更新。
②その他、特記事項	

(1-4) データベース関係 (該当なし)

(2) データ (又はDB) の連結、統合化整備

通番	データ (又はDB) の名称	公開 / 未公開	概要 (データの種別 (生物種)・数量 (kB 等)、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述)
1	トレースデータベース	未公開	塩基配列決定の1次データとして産生されるトレースデータのデータベース。ウェブサイト、またはウェブAPIからの検索サービスを予定。現在、データベース自体は最初の試験実装段階にあるが、ウェブインタフェースは未実装。
2	トレース付随データ	未公開	FTPによる公開を予定。トレースデータ自体をこれに含めるかに議論があり検討中。

(3) DB基盤システム、ツール等開発成果物の整備

通番	DB基盤システム、ツール等の名称	公開 / 未公開	概要 (主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述)
1	トレース登録データチェックツール	未公開	トレースアーカイブサービスへのデータ登録に際して、データの文法的・意味的なチェックを行うためのツール。現在は内部試験使用中。将来的には登録希望者への提供を予定。

(4) その他の成果物 ((2)、(3)に該当しないもの) (該当なし)

補完課題実施機関 (国立遺伝学研究所) 外部発表実績 (該当なし)

補完課題実施機関（九州工業大学） 整備実績一覧

(1) 保有データ情報

(1-1) データの種類

①生物種	複数
②試料・ライブラリ 一等の種類、数	①蛋白質およびその変異体の構造安定性に関する熱力学データ約 22,000 件 ②蛋白質と核酸の相互作用の定量的な熱力学実験データ約 8,000 件
③測定方法	熱測定、分光測定など
④データの内容	書誌情報、数値情報、画像情報。内容の詳細は別紙参考資料参照。
⑤その他、特記事項	なし。

(1-2) データソース

①現在のデータ量	①蛋白質およびその変異体の構造安定性に関する熱力学データは、22,000 件以上を文献から収集。 ②蛋白質と核酸の相互作用に関する熱力学データは、8,000 件余りを収集。
②データ区分	■自前 ■第三者 ■文献データ □計算結果等の二次データ ■その他（下欄に詳細を記述）
③将来の増加の見込み	蛋白質の安定性および蛋白質・核酸相互作用の熱力学データともに、年間 2,000～3,000 件程度の新規データが発生。
④権利関係	所有者（データベース作成者） 公開（ <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 [ ]）
⑤その他、特記事項	①ProTherm: <a href="http://gibk26.bse.kyutech.ac.jp/jouhou/protherm/protherm.html">http://gibk26.bse.kyutech.ac.jp/jouhou/protherm/protherm.html</a> ②ProNIT: <a href="http://gibk26.bse.kyutech.ac.jp/jouhou/pronit/pronit.html">http://gibk26.bse.kyutech.ac.jp/jouhou/pronit/pronit.html</a>

(1-3) データの管理状況

①更新頻度等の管理状況、体制	月に 1 回程度更新。データベースサーバーは無休でアクセス可能。
②その他、特記事項	なし。

(1-4) データベース関係

①DB 管理者数	1 名
②キュレータ・アナレータ数	2 名
③データ構造	リレーショナル

④DB 管理ソフト	SYBASE
⑤サーバの OS	Linux
⑥サーバ規模	ワークステーション
⑦DB へのアクセス数	年間約 10 万件
⑧独立 IP 数	約 1 万個
⑨その他、特記事項	DB の検索メニューの画面コピーは別紙参考資料添付。 オントロジーは今後整備。

- (2) データ (又はDB) の連結、統合化整備 (該当なし)
- (3) DB 基盤システム、ツール等開発成果物の整備 (該当なし)
- (4) その他の成果物 ((2)、(3) に該当しないもの) (該当なし)

### 補完課題実施機関 (九州工業大学) 外部発表実績一覧

- (1) セミナー、研究会等イベント開催 (該当なし)
- (2) プレス発表、取材対応 (該当なし)
- (3) 展示会等出展 (該当なし)

#### (4) 学会等への口頭発表

通番	タイトル	発表者	学会等の名称	年月日	特記事項
1	生体分子間相互作用の熱力学データベースと解析	皿井明倫	生物物理学会	2007 年 12 月 21 日	

#### (5) 雑誌等への論文寄稿

通番	タイトル	著者名	雑誌等の名称	掲載巻、号、ページ	特記事項
1	Thermodynamic Database for Proteins: Features and Applications	M. Michael Gromiha and Akinori Sarai	Methods in Molecular Biology	印刷中	