

## 画期的な農畜産物作出のためのゲノム情報データベースの整備

### 1. 事業概要

これまでの農畜産物ゲノム研究においては、ゲノム塩基配列の解読とともに、農業上重要な形質を担う遺伝子の単離や育種の効率化を図るためのDNAマーカーの開発等を進めてきました。これらの情報は、新たな作物や家畜の作出を効率的に進める上で、非常に有用な資産となります。

さらに、ゲノム解析機器の技術革新により迅速かつ安価なゲノム解析が可能となっていることから、これを用いて多種多様な遺伝資源のゲノム情報を解読し、塩基配列の相同性や差異、形質情報等との関連を解析することで、新規遺伝子の単離やDNAマーカーの開発の加速化、鍵となる遺伝子にターゲットを絞った品種開発など一層の育種効率向上が可能となります。

そこで、これまでに生み出されたゲノム情報等を利用しやすい形で研究者に提供するデータベースを構築、運営するとともに、超高速シーケンサー等の先端技術を効果的に活用し、大量の一次情報を高速・高精度で処理する手法の開発と、ゲノム情報を形質情報と関連づけ農業上重要な形質を担う新規遺伝子の単離やDNAマーカーの開発を支援するシステムを構築します。

### 2. 公募研究課題別の研究開発内容、目標等

「公募研究課題」ゲノム情報データベースの整備

#### (1) 研究開発の具体的内容

独立行政法人農業生物資源研究所（茨城県つくば市観音台）内に設置、運用されている「農林水産生物ゲノム情報統合データベース」（以下「旧データベース」とする。）の機能及び保存データ（資料1）<sup>(注)</sup>を継承、拡充した新規データベースを整備するため、下記①に掲げるハードウェアを設置、運用するとともに、ユーザーの利便性向上に資する機能改良を行います。また、下記②に掲げる高次解析システムを開発し、上記データベースと併せて運用します。

（注）本文中の資料1～3をご入り用の方は、下記担当までご連絡ください。

別途送付いたします。

#### ① データベースの改良、運用

これまでに蓄積されたゲノム情報データに加え、今後5年間で生み出されるデータの増大に対応できるハードウェアを整備し、ユーザーの評価等を踏まえて利便性向上に資する機能改良を行うとともに、データベースを運用します。

##### a) ハードウェアの構成

旧データベースの保存データ量並びに今後5年間で生み出される予定データ量、②で開発する解析システムを含む情報解析・提供機能が支障

なく作動する計算能力等を勘案して、想定するハードウェア能力は以下のとおりとします。

想定するハードウェア	最低限必要な水準
データベース公開用サーバ	同時に80程度のウェブサービスを提供し、年間30万訪問件数に対応できる能力を確保すること
ストレージ	旧データベースの保存データ量並びに今後5年間で生み出される予定データ量の合計(バックアップ分を含む)として、最終的に2ペタバイトの記憶容量及び大量データに効率よくアクセスできる高速・高信頼性領域を確保すること
情報解析サーバ	PCクラスター及び大容量メモリサーバを用いて、新たに開発する「超高速シーケンサーに対応したゲノム断片の整列化機能」や「新規遺伝子予測機能」を支障なく提供できる計算能力を有すること
情報収集サーバ	外部のデータベースから定期的に関連更新情報を取り込む機能を有すること
情報利用サーバ	外部ユーザーがログインして利用できるオンライン解析機能を有すること
ネットワーク装置	高速ネットワーク接続(10ギガベース以上)を確保すること
電源装置	データベース全体を最低30分間無電源下で連続稼働できること

※ 整備するハードウェアには、上記ハードウェアを制御、接続するためのシステム管理サーバやセキュリティ装置等を含む。

<留意事項>

- ・ ハードウェア構成の設計に当たっては、将来の管理、更新作業等を鑑み、「データの公開・検索」「大量情報解析」など機能目的別にハードウェアを分割運用できる形とします。
- ・ 他方、一過性の作業負荷に対してデータベース全体で負荷を分散する仕組みを導入する等、導入や運用に係るコスト低減に努めることとします。

b) 旧データベースからの移行

旧データベースをリースして、新規のデータベースを整備する間の運用を行うとともに、保存されているデータ等を新規データベースに移すための作業を行います。

<留意事項>

- ・ データ等を移動、保存するに当たって、新規データベースにおいては将来にわたって管理・増設に適したファイル・データの保存構造を設計することとします。

- c) 生命科学データベース横断検索機能の開発  
本データベースを含む国内外の主要な生命科学関連のデータベースを横断的に全文検索できる機能を、独立行政法人科学技術振興機構（以下「JST」とする。）の仕様（資料2）<sup>(注)</sup>に準じて開発、運用します。
- d) 生命科学データベースアーカイブ構築機能の開発  
本データベースを含む国内の主要な生命科学関連のデータセットを長期間安定的に維持・保管するため、統一の保存形式に変換、必要な情報を付す機能を、JSTの仕様（資料3）<sup>(注)</sup>に準じて開発、運用します。
- e) データベースの運用管理  
1年を通じて最新の情報を切れ目なくユーザーに提供するため、ハードウェア及びソフトウェアの保守管理、新規データの取り込み・公表、ユーザー利便性向上に資する機能改良等を行います。
- ＜留意事項＞
- ・ ユーザー利便性向上に資する機能改良等、本公募研究課題の実施に伴い明らかとなる取組内容は、応募要領7の②のプロジェクト研究運営委員会と相談の上実施するものとします。

## ② 高次解析システムの開発

以下に掲げる各種解析機能を開発します。開発に当たっては、試作版を限定的に運用した上でユーザーからの評価に基づき必要な改善を行うこととします。

- a) 超高速シーケンサーに対応したゲノム断片の整列化機能  
超高速シーケンサーを用いて農畜産物ゲノム解読を行うため、解読したゲノム断片を高速・高精度でつなぎ合わせ整列化する機能を開発します。
- b) ゲノム情報を活用した新規遺伝子予測機能  
表現形質や遺伝子発現情報等とゲノム情報との関連を解析し、未確認の遺伝子の存在場所を予測する機能を開発します。

## (2) 達成目標（最終目標）

- ① (1)の①のc) 生命科学データベース横断検索機能の開発及びd) 生命科学データベースアーカイブ構築機能の開発並びに②の高次解析システムの開発は、次に掲げる期限までに完了することとします（ただし、開発完了後に新規データの追加を踏まえた更新やユーザーからの評価等を踏まえた機能改良を行う場合は、この限りではありません）。
- a) 生命科学データベース横断検索機能：平成25年度末  
b) 生命科学データベースアーカイブ構築機能：平成25年度末  
c) 超高速シーケンサーに対応したゲノム断片の整列化機能：平成24年度末  
d) ゲノム情報を活用した新規遺伝子予測機能：平成26年度末

- ② データベースの運用に当たっては、24時間・365日間のデータベース公開を目指し、年間の稼働率(※)を99%以上とします(ただし、不測の停電等予期せぬ外的要因による停止時間は、稼働率算定から除外します)。

※ 年間の稼働率 = {年間予定稼働時間 - (平均停止回数 × 平均復旧時間)} / 年間予定稼働時間

- ③ 公開データの充実やユーザー利便性向上に資する機能改良等を通じ、データベースへの年間訪問件数を30万件以上(研究目的以外のロボット自動巡回等を除く)とします。

(3) 研究実施期間(予定)

平成23年度～27年度(5年間)

(4) 平成23年度の委託研究経費限度額 343,000千円

3. 委託件数

原則1件とします。

4. 問合せ先

上記の内容に関する問合せは、応募の締切までの間、下記において受け付けます。

なお、審査経過、他の提案者に関する事項、応募に当たり特定の者にのみ有利となる事項等にはお答えできません。また、これ以外の問合せについては、質問者が特定される情報等を伏せた上で、質問及び回答の内容を農林水産技術会議事務局のホームページにて公開させていただきますので、ご承知おきください。

記

○ 公募研究課題について

農林水産技術会議事務局研究開発官(食の安全、基礎・基盤)室  
担当者 宇木、財津

TEL: 03-3502-7435

FAX: 03-3593-7227

○ 契約事務について

農林水産技術会議事務局総務課契約班 担当者 藤原

TEL: 03-3502-7967

FAX: 03-5511-8622