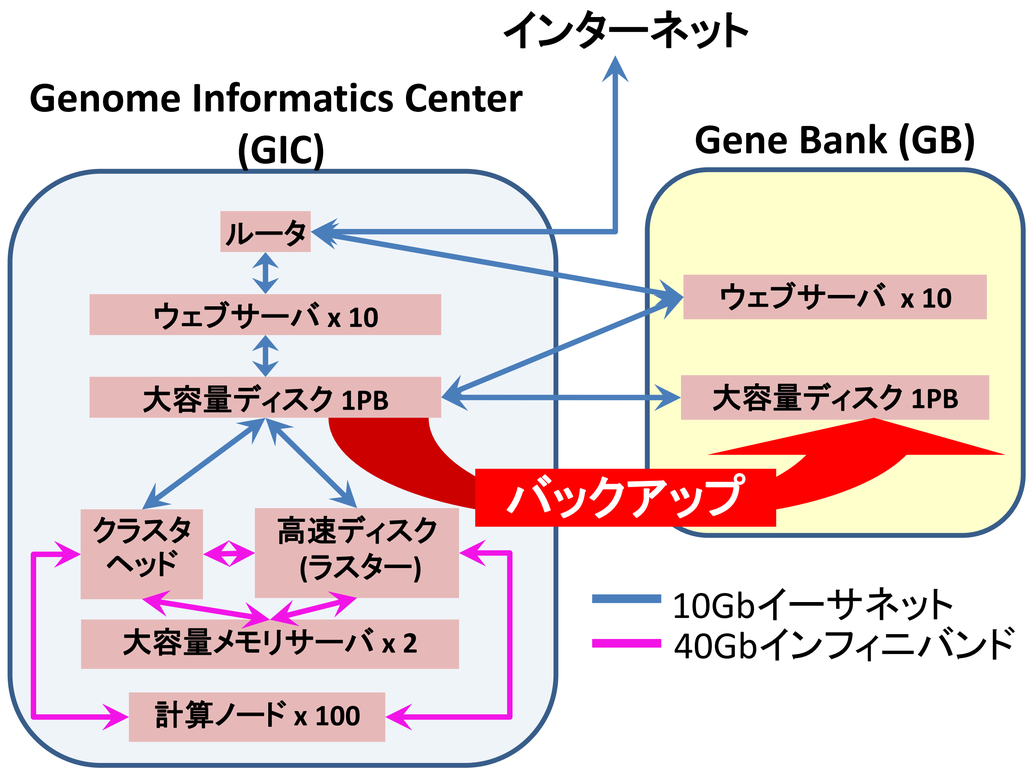
SOGO: 農畜産物ゲノム研究のための データベース統合と解析環境の整備

○土井 考爾[1]，ソロビヨワ　イェレナ[1*], 寺本 由美子[1], 石川 友美[1]，後藤 明俊[2], 宮尾 安藝雄[1]
(1)生物研・農業生物先端ゲノム研究センター (2)農研機構・作物研・稲研究領域　＊現・産総研・糖鎖創薬技術RC

新しいデータベース総合環境”SOGO”を2014年5月に正式公開した（**https://sogo.dna.affrc.go.jp/**）。
本システムは、クラウド型のミドルウェアを基盤としたデータベースと解析ツールの統合環境であり、農畜産物の研究者を対象に、データの収集・解析・共有・公開を容易に実現できるようなサービスを目指している。
SOGOのデータ・解析サービスの多くは、事前登録なしに誰でも自由に利用できる。一部のサービスは登録ユーザのみ利用できるが、登録申請はトップページ上のリンクを通して簡単に行える。

システム構成

- コンピュータに関する特別な知識がなくても、データや解析環境の安定的な維持、高速度計算などのメリットを受けられる。
- 二カ所の大容量ディスクの間で毎日データのバックアップが行われ、データ保全の確実性を高めている。



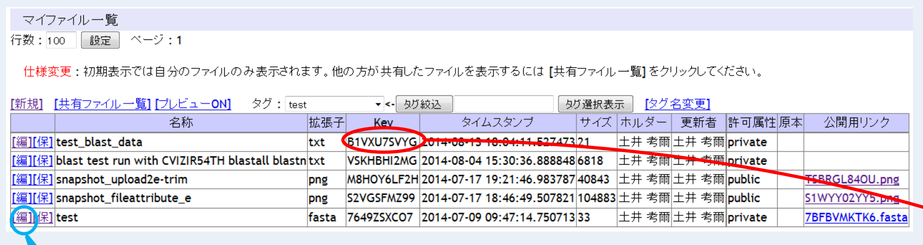
データベースの統合(AgrID)

- AgrIDは、農畜産物ゲノムに関する70以上のデータベースや解析サービスにリンクしたポータルサイトである。SOGOトップページから「データベース一覧」をクリックすると、AgrIDにアクセスできる。

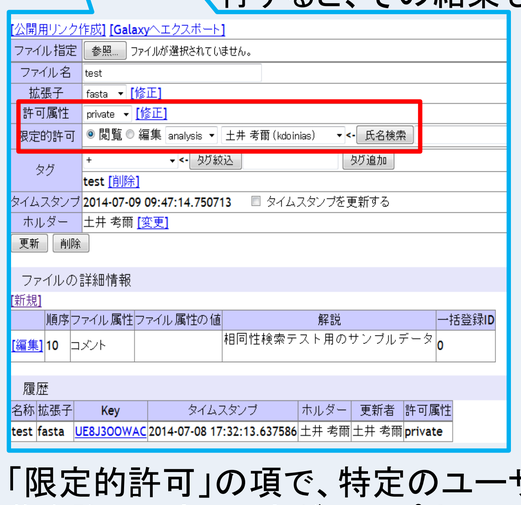
データの解析と管理

- SOGOには、相同性検索やPCRプライマー設計・NGS多型検出などを目的としたゲノム解析用ツールを現時点で約20種類組み込んでおり、今後も随時追加する予定である。SOGO上の解析ツールは、スケジュール・データ管理システムとの連携で効率的に解析作業を進められるように工夫されている。
- ここでは、BLASTN検索を例にとってSOGOでのデータ解析サービスの特徴を説明する。

ファイル管理



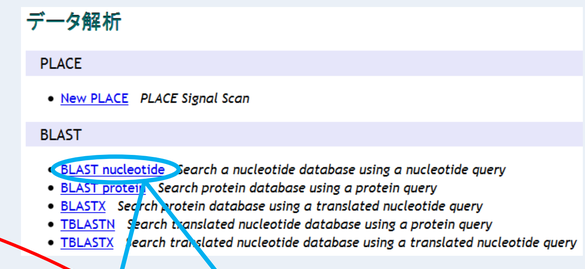
ログイン後、トップページ上部のメニューバーからanalysis→ファイルと進むと、サーバ上に各ユーザが保持しているデータの一覧が現れる。「新規」リンクを通して、手元のパソコンから、解析対象とするデータのアップロードもできる。解析ツールを実行すると、その結果も一覧に追加される。



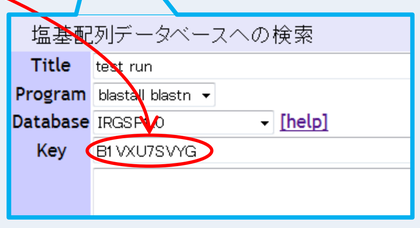
表中の[編]リンクから「ファイルの更新」画面に移ると、当該ファイルの名称変更やコメントの付加ができ、ファイルごとの公開の可・不可の設定も可能である（左図赤枠）。許可属性をpublicに設定すると当該データは一般公開される。

「限定的許可」の項で、特定のユーザに対してデータの閲覧・編集を許可でき、研究グループ内のみでデータ共有できる。

データ解析



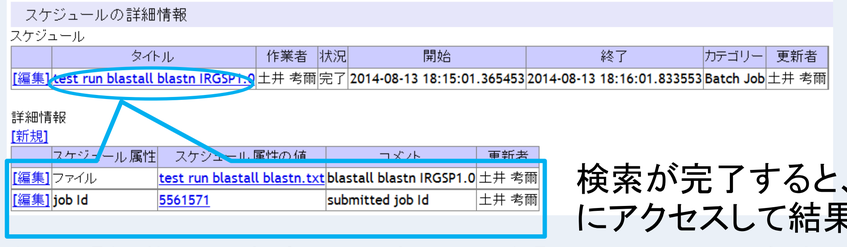
ログイン後、トップページ上部のメニューバーからanalysis→解析ツールと進むと、解析用ウェブツールの一覧が現れる。さらに“BLAST nucleotide”をクリックすると、下図のようなフォームが現れる。



解析を実行する前に、相同性検索のクエリーとして使う塩基配列データを記したファイルを作成し、「ファイル管理」画面で事前にアップロードしておく。マイファイル一覧で該当ファイルのkeyを確認し、これをkey欄に入力する。

SOGO上のBLAST系ツールでは、NCBI提供のdatabaseに加え、主要作物やカイコのゲノム配列・cDNA配列・アミノ酸配列などをプルダウンメニューで指定してdatabaseとして利用できる。

スケジュール管理



解析開始後、トップページ上部のメニューバーからanalysis→「スケジュール」と進むと、左図のようなリストが現れ、進捗状況が確認できる。

検索が完了すると、「タイトル」から「スケジュール属性の値」にアクセスして結果を確認できる。

また、「編集」で、解析に関するメモを「コメント」として自由に追加できる。これらの履歴は後からいつでも辿ることができ、ラボノートのように利用できる。

Copyright (C) 2014 独立行政法人農業生物資源研究所