微生物ゲノムアノテーションパイプライン MiGAP

黒川　顕[1)]，森　宙史[1)]，菅原秀明[2)]
[1]東京工業大学，[2]国立遺伝学研究所

Contact us: migap@dbcls.rois.ac.jp

◆背景

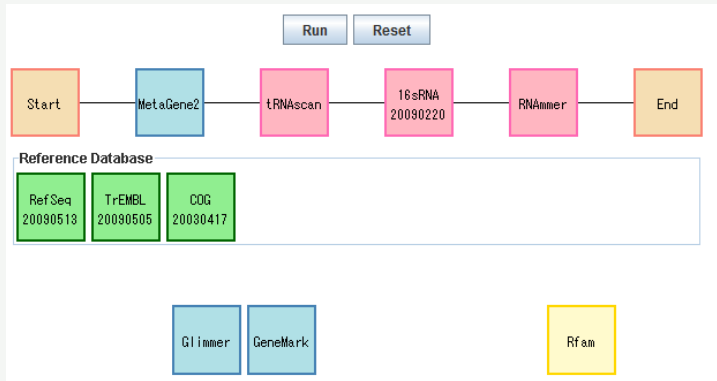
- ライフサイエンス統合データベースセンター（DBCLS）はライフサイエンス分野のデータベース（DB）に関するサービスの専門機関であると自らを位置づけ，国内のDBを中心に，DBの統合化と保全に努め，利用者の利便性を高めるための情報技術開発やポータルサイト整備を行います。
- DBCLSは2007年（H19年）から（１）戦略立案・実行評価，（２）統合データベース開発，（３）統合データベース支援 ならびに（４）事業の総合的推進 の４カテゴリーにおいて，具体的活動を始めました。
- カテゴリー「(2)統合データベース開発」においてDBCLSは，① 共通基盤技術開発，② ヒト統合DBの開発・運用 ならびに ③ モデル生物・産業応用生物統合DB の開発・運用 の課題に取り組んでいます。
- MiGAP（Microbial Genome Annotation Pipeline）は，課題「（２）ー③モデル生物・産業応用生物統合DBの開発・運用」における取り組みの成果の一つです。
- URL http://migap.lifesciencedb.jp/ を指定、または http://www.biobase.nig.ac.jp/からMiGAPを選択

◆MiGAPの課題設定

- シーケンサの技術開発は止まることを知らず，したがって，塩基配列決定の高速化も止まることを知りません。
- このため，比較的小規模で比較的シンプルな微生物ゲノム配列については，これからは多数の研究室やグループが次々に入手していくことでしょう。
- しかし，配列データの*tsunami*が知識を生み出すかについては保証の限りではありません。 ゲノム配列の「積読」に終わっては*mottainai*。
- そこで我々は，効率の良い微生物ゲノム配列アノテーションを実現する情報環境を目指しました。
- 具体には，微生物ゲノム解析において定評あるデータベースと定評あるアルゴリズムを組み合わせたパイプラインMiGAPを設計しました。
- MiGAPによって，誰でも，何時でも，どこからでも，微生物ゲノム配列の生物学的意味を入手できるようになります。
- MiGAPには，デフォルト解析の初級用のブロンズ（b-MiGAP），パラメーターを設定可能とする中級用シルバー（ｓ-MiGAP），さらにデータベースやアルゴリズムを付加することができる上級用のゴールド（g-MiGAP）の３つのレベルを用意します。
- このうち，ｂ-MiGAPは2009年度からOpenIDのもとで一般に公開しました。

◆ｂ-MiGAP（ビーマイギャップ）は何を使っているか

パイプラインに沿って見ると



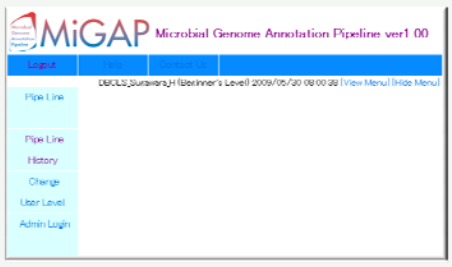
ORF予測はMetaGeneAnnotator(MGA) *を中心に； RNA領域同定はtRNAscanと高品質の独自rRNAデータベースを利用； RfamでIS領域同定； RefSeqとNRで既存アノテーションを参照； COG，KOG，ならびに eggNOGで機能の手掛かりを。上図GlimmerとGeneMarkはs-MiGAP以上で選択可能に。

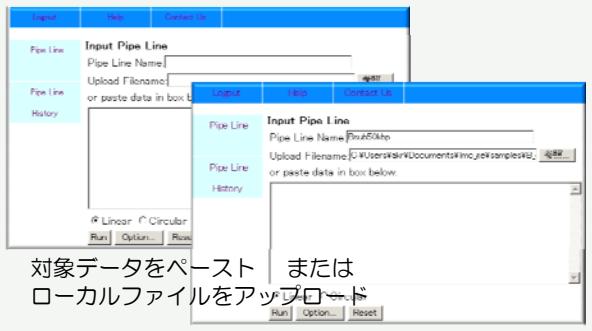
*）Noguchi H, Taniguchi T and Itoh T (2008) *DNA Res.* 15(6):387-96.

◆MiGAPを使う　本日より一般公開【解析は順次実行されます。先行解析が終了後、次の解析が始ります】

DBLCSでOpenIDを取得
☞ DBCLS OpenID

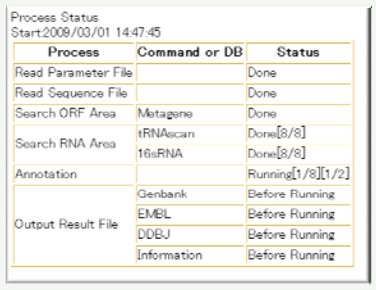
ログイン直後の画面





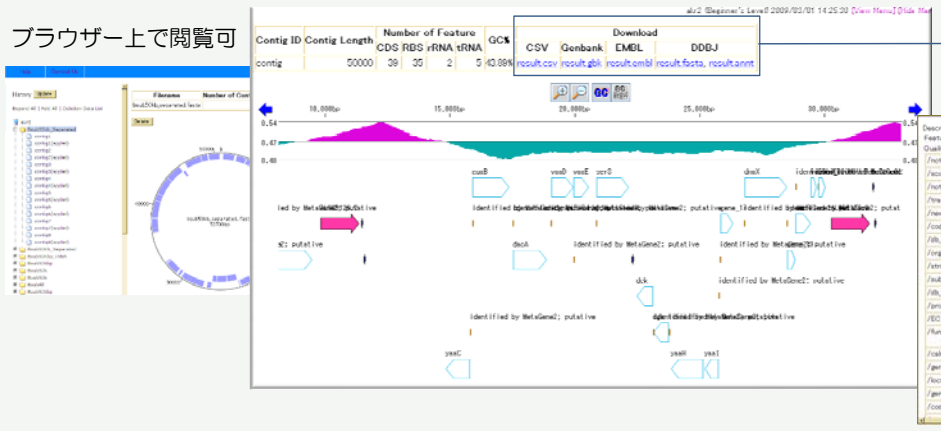
対象データをペースト　または
ローカルファイルをアップロード

進行状況をブラウザー上で確認できる



◆MiGAPから何が得られるか

ブラウザー上で閲覧可



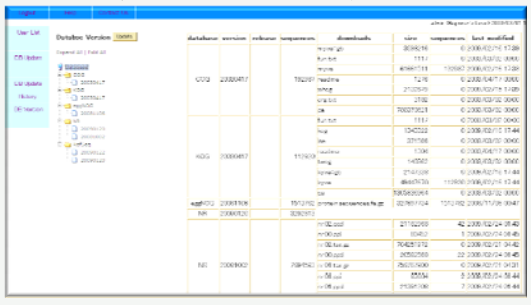
詳細情報を表示

閲覧結果をご希望のフォーマットでダウンロード可

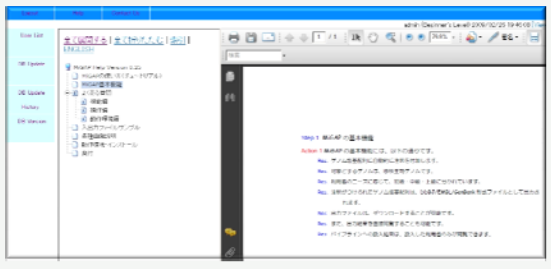
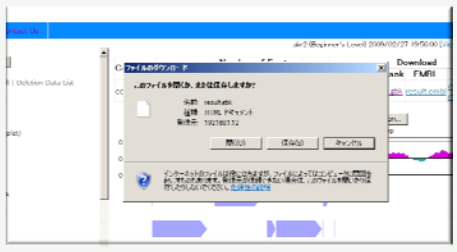
◆MiGAPの運用のために

参照データベース生成・更新実行へ

参照データベース更新情報確認



オンラインマニュアルを提供



そのままDDBJへ登録可能にも