



DBCLS Galaxy: 生物データ処理の対話的ツール組み合わせインタフェイス

○山口敦子, 全弘宇, 中尾光輝, 山本泰智, 高木利久

Database Center for Life Science (DBCLS), Japan
ライフサイエンス統合データベースセンター

DBCLS Galaxy とは？

Galaxy は生物データ, 特にゲノムデータをターゲットとして、ペンシルバニア州立大学によって作られたツール組み合わせの対話的インタフェイスである。

DBCLS Galaxy とはこの Galaxy をプラットフォームとして利用し、DBCLS 独自の機能やツールを組み込んだウェブアプリケーションである。

なぜ Galaxy か？

- ・表形式のデータ(ex BED, GFFなど)を入力とする解析ツールが揃っており、また、表形式を扱うツールならば比較的容易に組み込むことができる。
→ 汎用性の高いデータ形式であるため、DBCLS で開発した様々なツールを組み込み可能
- ・全ての計算履歴がデータ置き場に残る。
→ 任意のタイミングで、任意の途中結果のデータに処理を施すことができる。
- ・計算履歴はダウンロード可能である。
→ 計算結果をExcel など別ツールに渡すこともできる。
- ・データ置き場の各データはGalaxy利用者間で共有設定が可能である。

DBCLS Galaxy の特徴

TogoWS, テキストマイニングツールなど、DBCLS で開発したサービスを利用可能

- ・ TogoWS でサポートしている数多くのDBやサービス (NCBIのDB, EBIのDB, DDBJ, PDFj, KEGG など)を同じインタフェイス上で組み合わせ利用可能。
- ・ テキストマイニングツール(現在はMEDLINEから遺伝子名, タンパク質名, 化合物名などを抽出するツールや、関連する蛋白質核酸酵素の記事を検索するツールなどが組み込み済み)が利用できる。

認証に DBCLS OpenID を採用

- ・ LSDB の各種サービスと同一のアカウントで認証できる。

DBCLS Galaxy は現在

<http://galaxy.dbcls.jp/>
でβ公開中です。

(β公開のため、LSDBのトップからはアクセスできません)

DBCLS Galaxy の操作の流れ



データ閲覧用
ブックマークレット

操作例

例1. exon のBED形式データから

ある長さより短いexonを排除したい

i. BED形式データをアップロードするあるいは、

GetDataで取得する

→ BED形式のデータが**ヒストリー**に置かれる

ii. **Text Manipulation Compute**ツールで、

カラム間数値の簡単な演算ができるので、

3カラム目(終了位置) - 2カラム目(開始位置)を各行に対して計算する

→ i.のデータに対し exonの長さが各行の最後(7カラム目)に

付け加えられたものが新たに**ヒストリー**に置かれる

iii. **Filter & Sort** の

Filter ツールで、

7カラム目の数値が

小さい行を除去

例2. 遺伝子名からタンパク質構造を閲覧

i. **TogoWS search**で DB: KEGG-gene として、キーワード検索

→ 遺伝子の KEGG ID のリストが**ヒストリー**に置かれる

ii. **TogoWS Get Value of a Given Field** で

DB: KEGG-gene, Field: structureでのデータを選んで実行

→ タンパク質構造が既知のものに対し、PDB ID のリストが

ヒストリーに置かれる

iii. Data viewer の PDBj ブックマークレットを用いて、

PDB IDに対応するPDBjのエントリーを閲覧する

DBCLS OpenID でログイン

自動的にメールアドレスを DBCLS OpenID サーバから取得。

処理したいデータを**ヒストリー**(データ置き場)に置く

データをアップロード、あるいはGalaxy 内のデータ取得ツール(GetData)や検索ツール(TogoWSのDB検索や MEDLINE 検索など)によって、ヒストリーにデータを置くことができる。

ツールの中からデータ処理に使いたいものをひとつ選ぶ

Galaxy に元々組み込まれているゲノム解析ツールに加え、TogoWSのDB検索・Filed 取得、テキストマイニングツールを選ぶことができる。

ツール操作&データ閲覧画面で

ツールに与えるデータ(ヒストリーから選択)やパラメータを選ぶ

ツール実行

ヒストリーに計算結果が表示されるので、

ツール操作&データ閲覧画面で詳細を確認し、次の処理をする

対話的に反復

ワークフロー作成

他の利用者とデータを共有

ブックマークレットを用いて

データ提供元のサイトを閲覧

利用例・操作方法については、

TogoTV (<http://togotv.dbcls.jp>)

- Galaxyを使い倒す-特定の転写因子予測結合領域と遺伝子上流領域の「交差点」をリストアップする

・ Galaxy ScreenCasts (<http://g2.trac.bx.psu.edu/wiki/ScreenCasts>)

※英語です

- 1. Interface 1.1 - Introduction to Galaxy interface

- 2. Tool tutorials

- 3. Interval Operation Tutorial

- 4. Examples of complex analyses: Finding and analyzing Non-GenCode ESTs

- 5. Developers How To

- 6. Interactions between UCSC and Galaxy

- 7. Future

にチュートリアル動画も用意されている



DBCLS Galaxy の今後の展開

・ ユーザ各自の計算機に DBCLS Galaxy をダウンロード &

インストールできるようパッケージ化

・ 組み込む DBCLS 独自ツールの拡充

ご意見・ご要望お待ちしております。
support@dbcls.rois.ac.jp