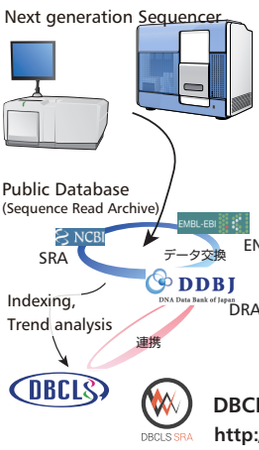
Database Center for Life Science (DBCLS)
Research Organization of Information and Systems (ROIS)

公共 NGS データから非モデル生物のデータをより簡単に得るための検索

仲里 猛留 (Takeru Nakazato)
nakazato@dbcls.rois.ac.jp
@chalkless
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センター (DBCLS)

Backgrounds and motivations

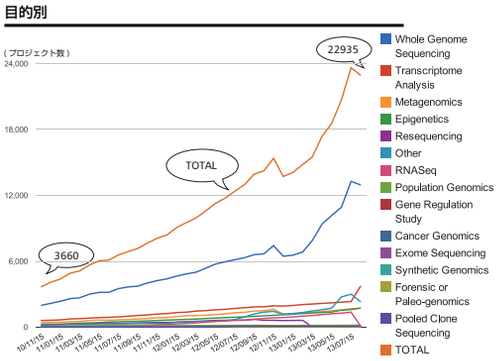


ここ最近、次世代シーケンサ (NGS) による成果が次々に発表されている。マイクロアレイのデータが GEO に登録されるのと同様に、NGS データも公共データベースである Sequence Read Archive (SRA) に登録され、日米欧の 3 局でデータ交換がなされている。その数は、プロジェクト数で 23000 (2013 年 9 月現在) と前年の倍近くに及んでいる。DBCLS では、DDBJ と連携し、登録データに対して、目次作成、データの傾向分析を行い、NGS データの検索サイトである DBCLS SRA を構築、提供している。

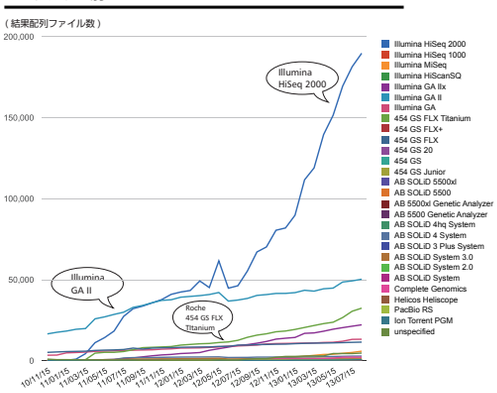
DBCLS SRA
http://sra.dbcls.jp/

Statistics (as of Sep 2, 2013)

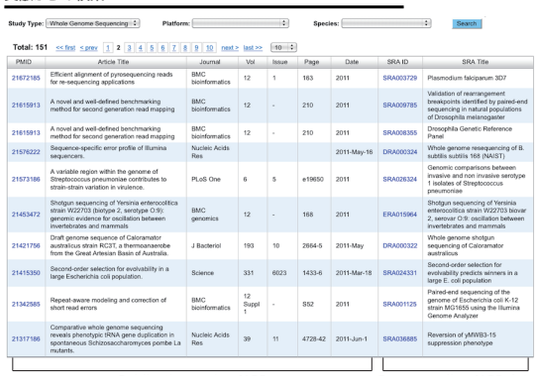
目的別



シーケンサー別



文献からの検索



Results and Discussions

生物種別 (Top 15)

Homo sapiens	2019
Mus musculus	1325
unidentified	867
Drosophila melanogaster	507
Caenorhabditis elegans	282
soil metagenome	271
[TaxonID]	248
Arabidopsis thaliana	243
marine metagenome	197
Saccharomyces cerevisiae	191
Escherichia coli str. K-12 substr. MG1655	174
Bacteria	106
human gut metagenome	93
Danio rerio	88
Zea mays	83
Total	25701 (studies)

現場の声

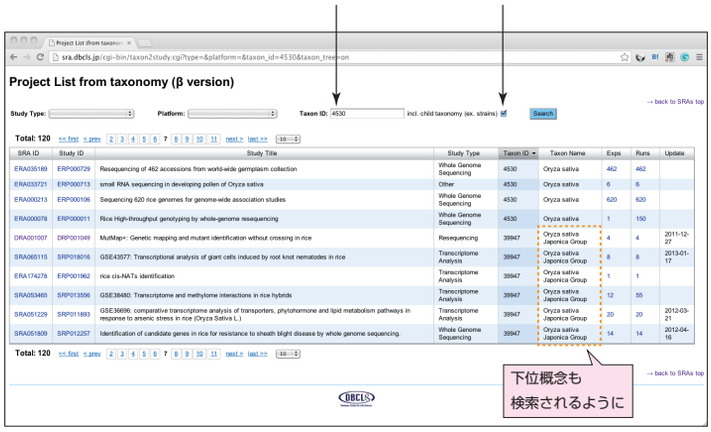
イネなんですけど、
japonica とか *indica* とかもあって
探すの大変なんですよね。

カイコをやっているんですけど、
昆虫全体とかで見たいんですが。

作ってみました
(ざっくりですけど)

Data Visualization

Taxonomy ID 入力　　下位概念も検索するときはチェック



※ 随時 アップデートするので、機能追加+インターフェースの変更もある予定

イネ (とその亜種) の例

- 4530 Oryza sativa — 64
  - 39947 Oryza sativa japonica group — 35
  - 120 39946 Oryza sativa indica group — 19
  - 1050722 Oryza sativa Indica Group x Oryza sativa Japonica Group — 1
  - 1080340 Oryza sativa Japonica Group x Oryza sativa Indica Group — 1

イネそのものだけだと 64 件

indica/japonica など入れると 120 件

カイコ (とその上位、近縁) の例

- 7088 Lepidoptera (チョウ目)
  - 37569 Bombycoidea (カイコガ上科)
    - 153 26 …
      - 7091 Bombyx mori (カイコ) 18
      - 7092 Bombyx mandarina (クワコ) 3
  - 37572 Papilionoidea (アゲハチョウ上科)
    - 91 7143 Papilionidae (アゲハチョウ科) … 6
    - …
    - 40037 Heliconiiae (ドクチョウ亜科) 76

カイコだけだと 18 件

近縁も入れて 26 件 (タバコスズメガ が 6 件)

チョウ目全体だと 153 件

Conclusions

- 公共データベース SRA に登録された次世代シーケンサによるデータを生物種の切り口から整理した。
- 非モデル生物について、上位概念、下位概念を検索ができるようにした。
- 整理したデータは DBCLS SRA より閲覧可能。 http://sra.dbcls.jp/

NGS 現場の会 第 3 回研究会
神戸国際会議場
平成 25 年 9 月 4 日~5 日