Functional interface for quick access to high-quality omics data

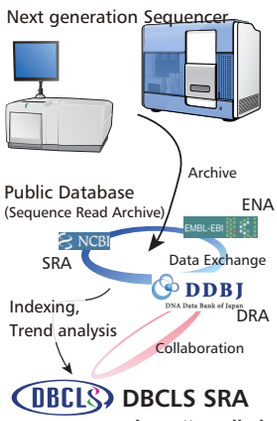
実験情報を利用した質の高い公共オミックスデータの検索と目次サイトの構築

○ Takeru Nakazato*, 仲里 猛留
Tazro Ohta, 大田 達郎
Hidemasa Bono 坊農 秀雅

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

*: nakazato@dbcls.rois.ac.jp
twitter ID: chalkless

Backgrounds



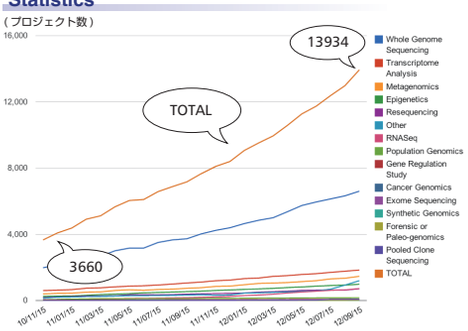
The next-generation sequencing (NGS) data is archived in public databases, SRA, and the data is collaboratively maintained by DDBJ, EBI, and NCBI. In Japan, Database Center for Life Science (DBCLS) has developed infrastructure for researchers to access and re-use these data easily by providing index and stats pages and constructing a portal site for life science databases and tools in collaboration with DDBJ.

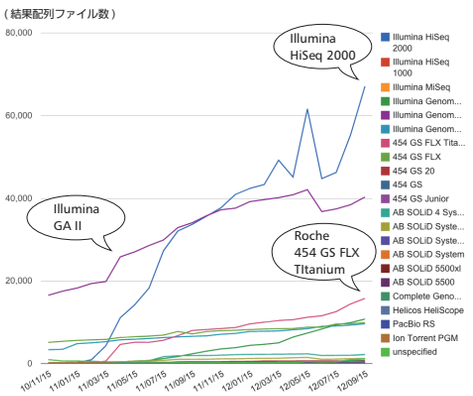
次世代シーケンサ (NGS) データも公共データベースである Sequence Read Archive (SRA) に登録され、日米欧の 3 局でデータ交換がなされている。その数は、プロジェクト数で 14000 (2012 年 12 月現在) に及んでいる。DBCLS では、登録データに対して、目次作成、データの傾向分析を行い、NGS データの検索サイトを構築、提供している。

Results and Discussions

FREE! http://sra.dbcls.jp/

Statistics

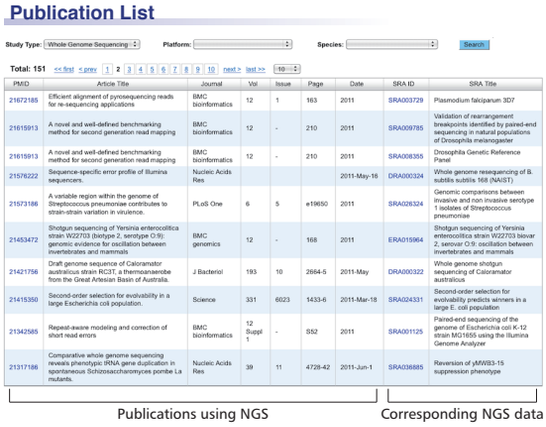




Homo sapiens	1261
unidentified	882
Mus musculus	818
Drosophila melanogaster	284
Mustela putorius furo	233
Caenorhabditis elegans	198
metagenome sequence	184
marine metagenome	165
Arabidopsis thaliana	160
unclassified phages	139
Escherichia coli str. K-12 substr. MG1655	136
soil metagenome	127
Saccharomyces cerevisiae	124
Human rhinovirus A	95
Salmo salar	84
Total	16806 (studies)

Disease List

Publication List



Publications using NGS — Corresponding NGS data

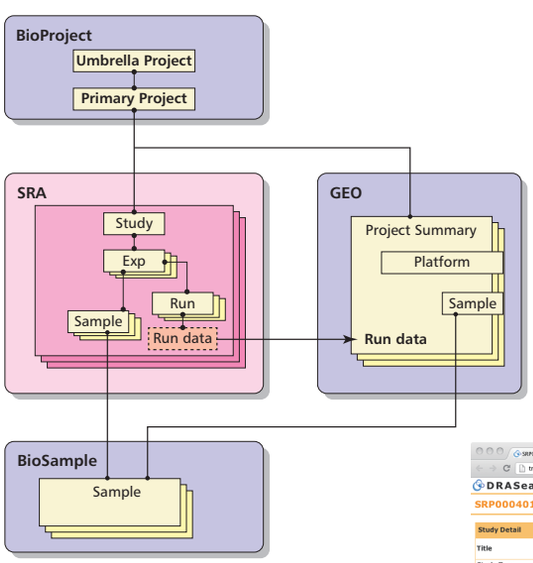
Data Quality

Please ask Tazro!

3P-0047: Advanced interface of public NGSeq database for efficient data search
Tazro Ohta, Takeru Nakazato, Hidemasa Bono

公共の次世代シーケンスデータを効率良く検索するためのインターフェース開発
大田 達郎、仲里 猛留、坊農 秀雅

Future works



3 局では、複数のデータベースを俯瞰するため、プロジェクトのデータベースである BioProject と、サンプルのデータベースである BioSample を立ち上げており、今後、これらによる統合的な活用ができるよう、開発を行っている。

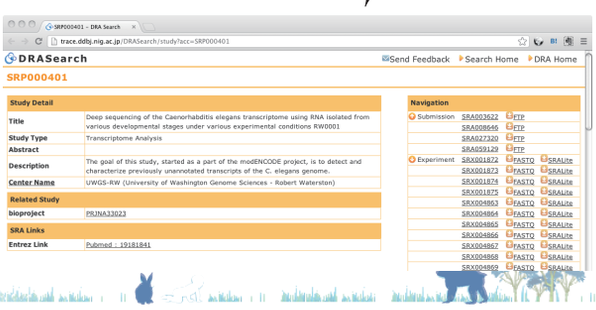
Project List

No.	SRA id	Study id	Title	Study Type	Exp	Run	Update Date
1	SRA051306	SRP011970	Genetic Dissection of Model Complex Trait Using the Drosophila Synthetic Population Resource	Other	2640	2640	2012-03-27
2	ERA007057	ERP000190	Plasmodium_falciparum_natural_genome_variation	Whole Genome Sequencing	2447	2532	2012-05-29
3	ERA015449	ERP000426	ZF_MrSoi	Other	2440	2536	2012-04-25
4	SRA026060	SRP002163	Human Microbiome Project Metagenomes Production Phase	Metagenomics	2306	6017	2010-06-09
5	SRA023558	SRP003279	ARRA Autism Sequencing Collaboration	Resequencing	2257	1385	2010-08-03
6	SRA056418	SRP014601	Jackson Heart Study Allelic Spectrum Sequencing Discovery	Other	2030	1948	
7	ERA071297	ERP001040	High_resolution_QTL_mapping_in_an_outbred_population_of_mice_using_low_coverage_next_generation_sequencing	Whole Genome Sequencing	1767	1767	2012-04-03
8	SRA026080	SRP003341	Women's Health Initiative Clinical Trial and Observational Study Sequencing Project	Resequencing	1741	1805	2010-08-03
9	SRA050250	SRP011021	Prostate Cancer Genome Sequencing Project	Resequencing	1640	328	2012-02-23
10	SRA025956	SRP004051	Framingham Heart Study Allelic Spectrum Sequencing Discovery	Resequencing	1638	456	2010-11-02
11	SRA026060	SRP002860	Human Microbiome Project 16S rRNA 454 Clinical Production Phase II	Metagenomics	1517	1823	2010-03-17
12	ERA122942	ERP001384	Defining the core Arabidopsis thaliana root microbiome	Metagenomics	1277	1277	2012-06-04
13	SRA050859	SRP011540	Melanoma Genome Sequencing Project	Resequencing	1170	290	
14	ERA005746	ERP000129	The Genomics of Speciation in Campylobacter	Population Genomics	1089	1213	2012-03-20
15	SRA051296	SRP011949	Swedish Schizophrenia Population-Based Case-control Exome Sequencing	Other	1067	1032	2012-03-26
16	SRA023556	SRP003277	Ciliopathies Exome Sequencing Initiative	Resequencing	1062	1066	2010-08-16
17	ERA095466	ERP001277	Pig_X_Y_clone_sequencing	Other	1033	1033	2012-03-09
18	ERA068391	ERP001012	Staphylococcus_aureus__BSAC__study	Whole Genome Sequencing	1030	1030	2012-02-03
19	SRA052357	SRP012694	Genetic Analysis of Hirschsprung Disease	Resequencing	981	344	
20	ERA011044	ERP000270	Investigating the diversity of Vibrio cholerae isolates.	Whole Genome Sequencing	981	1167	2012-05-08

Coming soon...

8102 projects — BioProject

1912 projects — GEO Gene Expression Omnibus



DRA (DDBJ SRA)
http://trace.ddbj.nig.ac.jp/DRASearch/

第 35 回 日本分子生物学会年会
福岡国際会議場、マリンメッセ福岡
平成 24 年 12 月 11 日~14 日