統合データベース講習会：AJACS岐阜
2013年7月12日

DDBJを中心とした ゲノムデータベースの紹介

情報・システム研究機構（ROIS） ライフサイエンス統合データベースセンター（DBCLS）
科学技術振興機構（JST） バイオサイエンスデータベースセンター（NBDC）

河野　信

注意点

- 参加人数が多いため、サイトにつながりにくくなることが予想されます。
  - 資料を見ながら適当にタイミングをずらして実行してみてください
  - 反応が無くても、何度もクリックしない
    - ますます遅くなるだけです。おおらかな気持ちで臨みましょう
  - わからないことがあったら、講習会のスタッフに気軽に聞いてください

代表的な生命科学関係のデータベース

DNA塩基配列	GenBank/ENA (EMBL)/DDBJ, RefSeq SRA/ENA/DRA
ゲノム	Ensembl, H-InvDB, MGI, FlyBase, TAIR, SGD
タンパク質アミノ酸配列	UniProt (Swiss-Prot + TrEMBL)
タンパク質立体構造	PDB, SCOP, CATH
モチーフ	InterPro, Pfam, PROSITE, ProDom
化合物	PubChem, ChEBI, KEGG LIGAND
パスウェイ	KEGG PATHWAY, Reactome, BioCyc
遺伝子発現	GEO, ArrayExpress, BioGPS
文献	PubMed
その他	GO, NCBI Taxonomy, OMIM, GOLD

本日紹介するDB・ツール

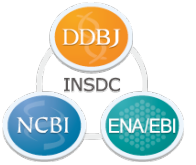
- DDBJ (DNA Data Bank of Japan)
  - DNA塩基配列を収集
- Ensembl
  - ゲノムを表示するためのブラウザ
- BioMart
  - ゲノム情報を取得するためのGUIインターフェイス
- GOLD (Genome OnLine Database)
  - ゲノムプロジェクトの情報を集めたデータベース

DDBJ

DNA Data Bank of Japan

国際塩基配列データベースの一員

- International Nucleotide Sequence Databank Collaboration (INSDC)
  - 米国：GenBank
  - 欧州：ENA
  - 日本：DDBJ



- （新型）DNAシーケンサーで解読されたDNA塩基配列を収集

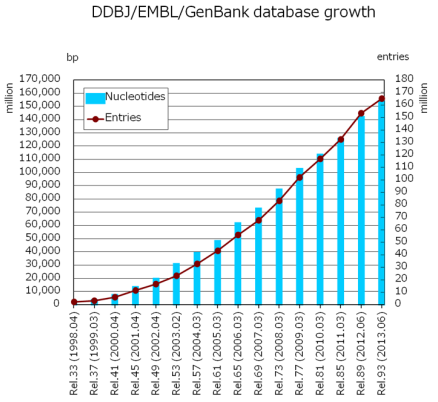
DDBJ登録ファイルの例

```
LOCUS       AB091058              2109 bp    DNA     linear   BCT 02-SEP-2003
DEFINITION  Gluconacetobacter xylinus cmcase, ccp genes for
            endo-beta-1,4-glucanase, cellulose complementing protein, complete
            cds.
ACCESSION   AB091058
VERSION     AB091058.1
KEYWORDS    .
SOURCE      Gluconacetobacter xylinus
  ORGANISM  Gluconacetobacter xylinus
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
            Acetobacteraceae; Gluconacetobacter.
REFERENCE   1  (bases 1 to 2109)
  AUTHORS   Kawano,S., Tajima,K., Uemori,Y., Yamashita,H., Erata,T.,
            Munekata,M. and Takai,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2002) to the DDBJ/EMBL/GenBank databases.
            Contact:Kenji Tajima
            Hokkaido University, Graduate School of Engineering; N13W8,
            Kita-ku, Sapporo, Hokkaido 060-8628, Japan
REFERENCE   2
  AUTHORS   Kawano,S., Tajima,K., Uemori,Y., Yamashita,H., Erata,T.,
            Munekata,M. and Takai,M.
  TITLE     Cloning of Cellulose Synthesis Related Genes from Acetobacter
            xylinum ATCC23769 and ATCC53582: Comparison of Cellulose Synthetic
            Ability Between ATCC23769 and ATCC53582
  JOURNAL   Unpublished (2002)
COMMENT
FEATURES             Location/Qualifiers
     source          1..2109
                     /db_xref="taxon:28448"
                     /mol_type="genomic DNA"
                     /note="synonym:Acetobacter xylinum"
                     /organism="Gluconacetobacter xylinus"
                     /strain="ATCC 53582"
     CDS             10..1038
                     /codon_start=1
                     /gene="cmcase"
                     /product="endo-beta-1,4-glucanase"
                     /protein_id="BAC82540.1"
                     /transl_table=11
                     /translation="MSVMAAMGGAQVLSSTGAFADTAPDAVAQQWAIFRAKYLRPSGR
                     VVDTGNGGESHSEQGGYGMLFAASAGDLASFQSMWMWARTWLQHTNDKLFSWRFLKGH
                     QPPVPDKNNATDGDLLIALALGRAGKRFQRPDYIQDAMAIYGDVLNLMTMKAGPYVVL
                     MPGAVGFTKKDSVILNLSYYVMPSLLQAFDLTADPRWRQVMEDGIRLVSAGRFGQWRL
                     PPDWLAVNRATGALSIASGWPPRFSYDAIRVPLYFYWAHMLAPNVLADFTRFWNNFGA
                     NALPGWVDLTPGARSPYNAPPGYLAVAECTGLDSAGELPTLDHAPDYYSAALTLLVYI
                     ARAETTK"
BASE COUNT          343 a          661 c          661 g          444 t
ORIGIN
        1 cgttccttta tgtcggtcat ggcggcgatg ggaggggcgc aggtgctttc atccaccggt
       61 gcgttcgcag acaccgcccc cgatgcggtc gcgcagcaat gggccatctt ccgcgccaag
      121 tatcttcgtc ccagcggacg tgtcgtggat acgggcaatg gtggcgaatc ccatagtgag
      181 gggcagggct atggcatgct ctttgccgcg tcggcggggg accttgcgtc gttccagtcg
      241 atgtggatgt gggcgcgcac caacctgcag cataccaatg acaagctgtt ttcctggcgg
      301 ttcctcaagg ggcatcagcc cccggtgccc gacaagaaca atgccacaga tggcgacctg
      361 cgatcgcgc ttgcgcttgg tcgtgcgggc aagcgtttcc agcgccccga ttacattcag
      421 gacgccatgg ccatttatgg cgatgtgctg aacctgatga cgatgaaggc gggaccgtat
      481 gtcgtcctca tgccccgtgc tgtcggcttt accaagaagg acagccgtgat cctcaacctg
      541 tcctattacg tcatgccctc gctgctgcag gcgttcgacc ttacggccga ccgcgctgg
      601 cgtcaggtga tggaagacgg gattcgcctt gtttccgccg gccgtttcgg gcagtggcgc
      661 ctgccccccg actggctggc ggtgaatcgc gccaccggtg cgctgtcgat cgcatcggga
      721 tggccgccgc gcttttccta tgatgcgatt cgggtgccgc tttatttttа ttgggcgcat
      781 atgctggcgc cgaacgtgtt ggctgatttc acccgattct ggaataattt cggggctaat
      841 gccctgccag gatgggttga tctgacaaca ggggcgcgtt cgccgtacaa cgccccgcct
      901 ggatatcttg ctgttgccga atgcacgggg cttgattctg ccggggaact cccgacactg
      961 gatcatgcgc ccgattatta ttccgcaggg ttgacgctgc tcgtttacat cgcgcgggcg
     1021 gaggagacta taaagtgagt gcttcaggct ctgatgaggt ggctgggtga gggcaggctg
     1081 gaagtccgca ggatttcag cgggtcctgc gttcttttgg tgtcgaaggt gggcagtatt
     1141 cctaccggcc gtttgttgac cgttcttttg atgtgacagg cgtgcccgag gctgttgaaa
     1201 ggcacttcga tcaggcggag catgacacgg cggttgagga gcaggtcact cccgcgccac
     1261 aaatcgcggt cgcaccgcca ccgccgccag tcgttcctga cccgccccgc atcgtgacgg
     1321 aaaccgcgcc ccgccgccct gtcgtggtca gcgctccggt cacgtatgaa ccccccggctg
     1381 ccgccgtgcc ggcagagcct cccgttcagg aagccccccgt gcaggcggcg ccggttcccc
     1441 ccgcgcctgt gcccccgatt gcggagcagg ctcctcccgc ggcgccggac ccggcatccg
     1501 tgccgtatgc gaacgtcgcg gcagcacccg ttccacctga tcccgcaccg gttacgcctg
     1561 cgccgcaggc gcgcgtgacg gggccgaaca cccgtatggt ggagccctt tcccgcccgc
     1621 aggtccgcac ggtgcaggag ggggcaaccc cgtcacgtgt accttcgcgt tcaatgaacg
     1681 ctttccccg cacatcagca tcgtccataa gtgagcgtcc ggtggacagg ggtgttgccg
     1741 atgaatggag tcctgttccg aaggcacgcc tcagcccgcg ggagcgtccg cgtccggcg
     1801 atctgagctt ttccttcag gggatgcgcg acacccgtga tgaaaagaag ttctttccg
     1861 tggcgtccac gcgatcagtt cgttctaatg tttccaggat gaccacgcatg accaagacag
     1921 acacgaattc ctctcaggct tctcgtcccg gcagcccctg cgcctcgcct gatgggtcgc
     1981 ccacaatggc cgaagtgttc atgacgctgg gtggtcgtgc gacggaactc ctcagcccc
     2041 gtccttcgct gcgggaggcg ctgttgcgtc gtcgtgaaaa cgaagaagaa tcctaaggcc
     2101 ctatattca
//
```


現在の塩基配列データの量

塩基数：1,500億

登録数：1.6億



Note: CON division is not counted in statistics of DDBJ periodical releases.

DDBJデータベースを検索してみましょう

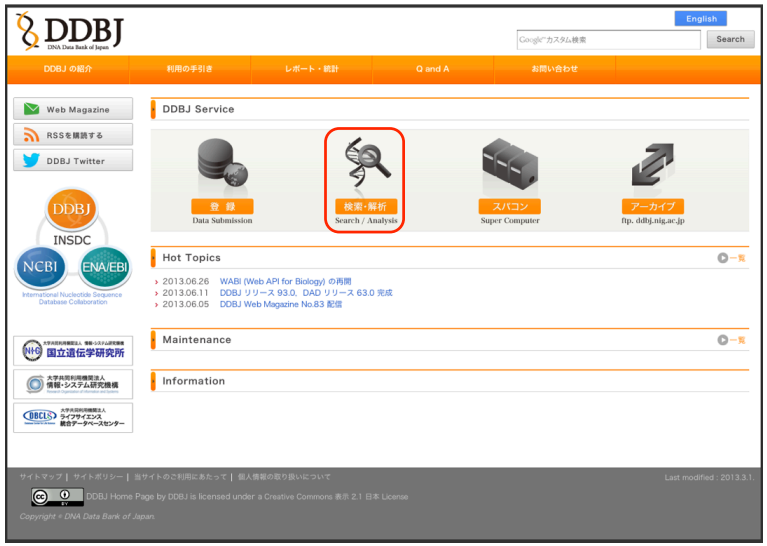
実習1-1：DDBJを検索

◇DDBJデータベースを”ARSA”という、
キーワード検索ツールで検索してみましょう

○例として大腸菌O157の全ゲノムエントリを検索してみます

DDBJにアクセスするには「DDBJ」で検索
もしくは　http://www.ddbj.nig.ac.jp/　を直接入力

DDBJトップページ



DDBJ検索・解析ページ

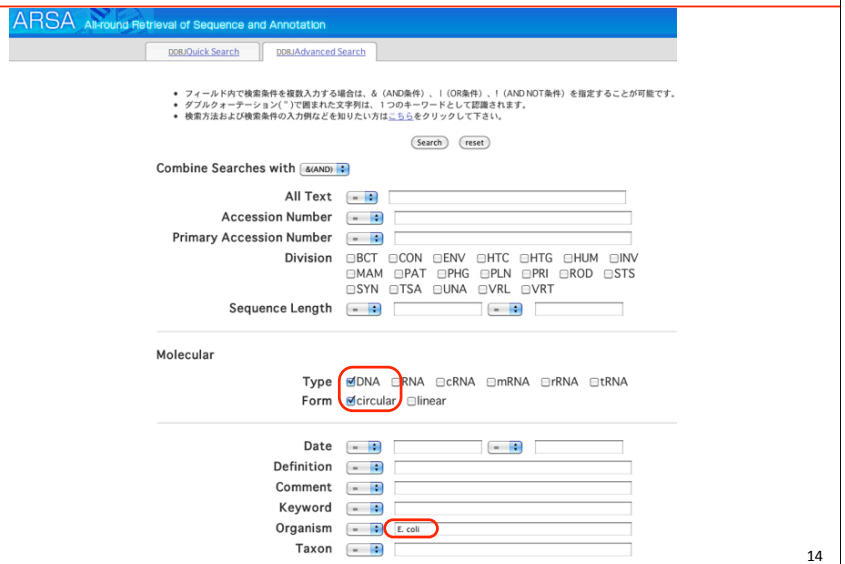
ARSA

E. coli O157 を検索

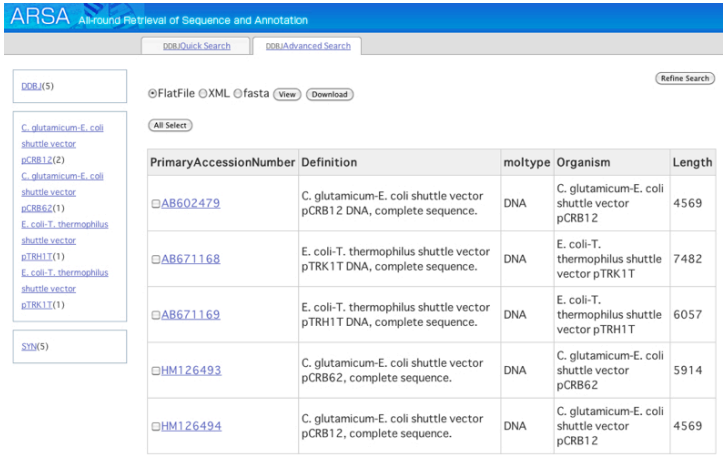
検索結果：
ヒット数が多すぎるので
絞り込みが必要

詳細検索
(Advanced Search)へ

いくつかの特徴で絞り込み

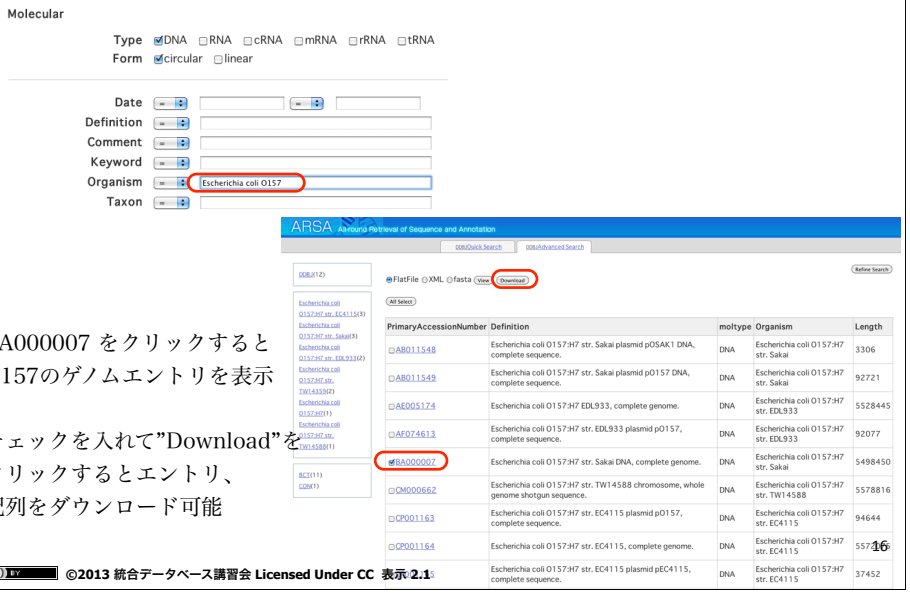


ヒットしない…



原因：Organismの項目には正式名称しか書かれていないため

Escherichia coli O157 で再検索



BA000007 をクリックすると
O157のゲノムエントリを表示

チェックを入れて”Download”を
クリックするとエントリ、
配列をダウンロード可能

実習1-2：NCBIを検索

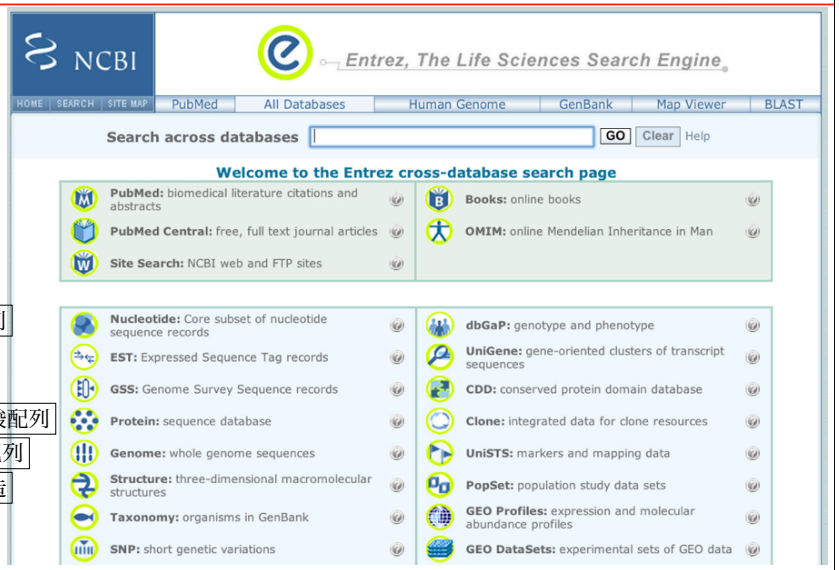
◇先ほどと同じ検索をNCBI（GenBankの提供機関）の検索システム”GQuery”で実行してみましょう

○例として大腸菌O157の全ゲノムエントリを検索してみます

NCBI GQueryにアクセスするには「gquery」で検索
もしくは　http://www.ncbi.nlm.nih.gov/sites/gquery
を直接入力

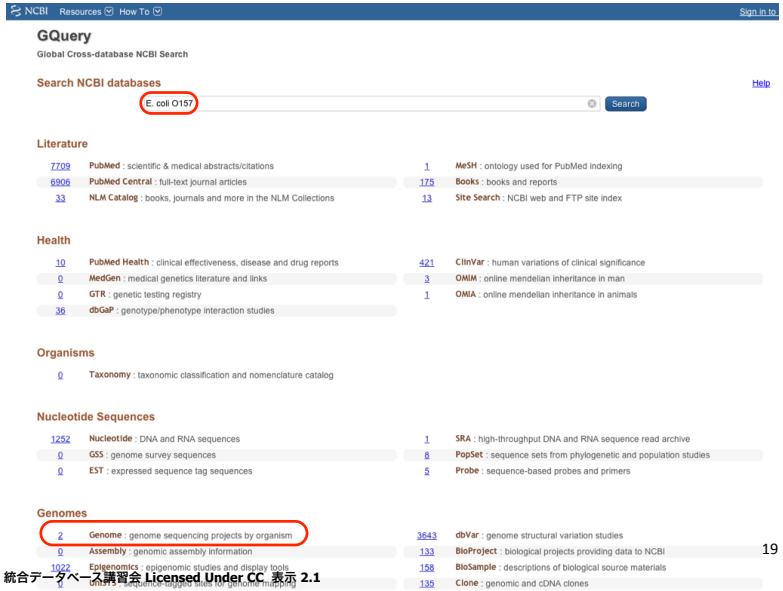
NCBI Entrez Search

文献
塩基配列
EST
アミノ酸配列
ゲノム配列
立体構造

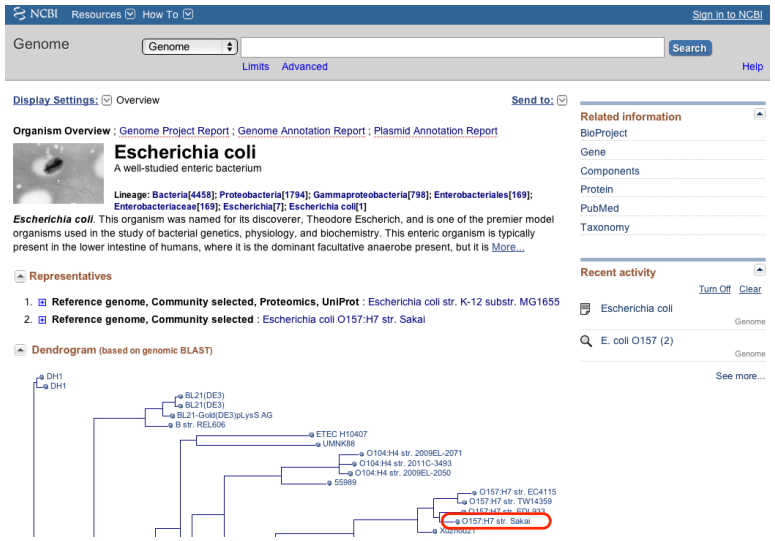


GQuery: E. coli O157 で検索

文献
健康
生物種
塩基配列
ゲノム



Entrez: 大腸菌ゲノムページ



大腸菌O157 Sakai株のページ

豊富なリンクとツール群

配列を取得したあとの解析例

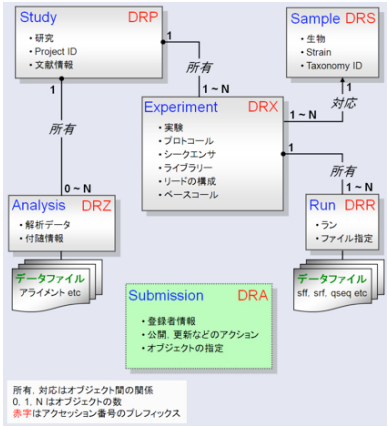
- “blast”等の配列類似性検索を実行して、　類似の配列を収集する
- ”primer3”等で配列をクローニングするためのプライマーを設計する
- “clustalW”等でマルチプルアラインメントを作成し、配列の共通部分や進化関係を調べる
- ”interproscan”等でモチーフ構造を調べる
- “swiss-model”等で立体構造を予測する

新型シーケンサからのデータ

DDBJ Sequence Read Archive (DRA)

◇新型シーケンサデータを保存・共有

新型シーケンサのデータ構造



登録されているデータ構造は少々複雑ですが、
DRAのページでは「日本語」での詳しい説明がある

登録されているデータはあまりうまく
整理されていない

SRAs: Survey of Read Archives

◇SRA/DRAに登録されているデータを
メタデータで整理

◇http://sra.dbcls.jp/

●生物種

●解析プラットフォーム

●キーワード

などで検索可能

SRAs: Survey of Read Archives

◇統計値から、分類をたどってデータにアクセス
することも可能

Search by statistics - 統計値から探す

The number of projects are indicated in "Study Types" table. The totals in "Platforms" and "Species of Samples" are larger than one of "Study Types" because a project can contain some platforms and sample species.

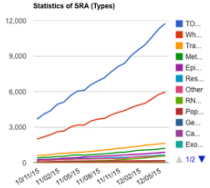
実際のプロジェクト数は"Study Types"に書かれているものです。"Platforms"や"Species of Samples"のtotalの値は、複数のプラットフォームで行われた実験が1つのプロジェクトでなされる 場合が多々あり、ダブルカウントしているので、数字が大きくなっています。

2012-06-21 updated.

Study Types

Whole Genome Sequencing	6024
Transcriptome Analysis	1657
Metagenomics	1270
Epigenetics	886
Resequencing	628
Other	615
RNASeq	585
Population Genomics	162
Gene Regulation Study	68
Cancer Genomics	41
Exome Sequencing	39
Pooled Clone Sequencing	21
Synthetic Genomics	6
Forensic or Paleo-genomics	6
Total	12008 (studies)

• Growth of the number of studies (including Total)



• Growth of the number of studies (without Total)

Platform

Illumina Genome Analyzer II	3629	Illumina HiSeq 2000	45470
454 GS FLX Titanium	2848	Illumina Genome Analyzer II	36940
Illumina HiSeq 2000	2328	454 GS FLX Titanium	11977
454 GS FLX	1735	Illumina Genome Analyzer	9051
Illumina Genome Analyzer	1167	454 GS FLX	8939
Illumina Genome Analyzer IIx	565	Illumina Genome Analyzer IIx	8760
454 GS 20	325	AB SOLiD 4 System	2000
AB SOLiD System 3.0	176	AB SOLiD System 3.0	1080
AB SOLiD 4 System	80	454 GS 20	686
Illumina MiSeq	64	Complete Genomics	401
AB SOLiD System 2.0	61	unspecified	394
unspecified	60	AB SOLiD System 2.0	362
Ion Torrent PGM	56	Helicos HeliScope	332
AB SOLiD System	37	PacBio RS	326

Species of samples (top 15)

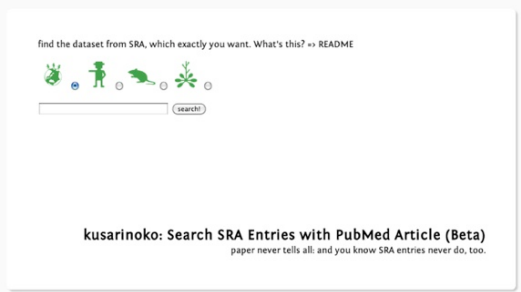
Homo sapiens	1107	*Homo sapiens*	38506
unidentified	886	unidentified	8678
Mus musculus	603	*Streptococcus pneumoniae*	7155
Drosophila melanogaster	253	*Mus musculus*	7004
Caenorhabditis elegans	181	*Drosophila melanogaster*	5523
metagenome sequence	180	human metagenome	3026
marine metagenome	160	*Danio rerio*	2940
Escherichia coli str. K-12 substr. MG1655	135	*Plasmodium falciparum*	2281
Arabidopsis thaliana	134	*Staphylococcus aureus*	1965
soil metagenome	115	*Mycobacterium tuberculosis*	1377
Saccharomyces cerevisiae	107	*Saccharomyces cerevisiae*	1325
Mustela putorius furo	99	*Oryza sativa*	1279
Human rhinovirus A	94	*Sus scrota*	1268
Salmo salar	56	marine metagenome	1217
Plasmodium falciparum	50	*Arabidopsis thaliana*	1212
Total	14554 (studies)	Total	127426 (experiments)

鎖鋸（kusarinoko）

◇論文が出ているSRA/DRAエントリのまとめ

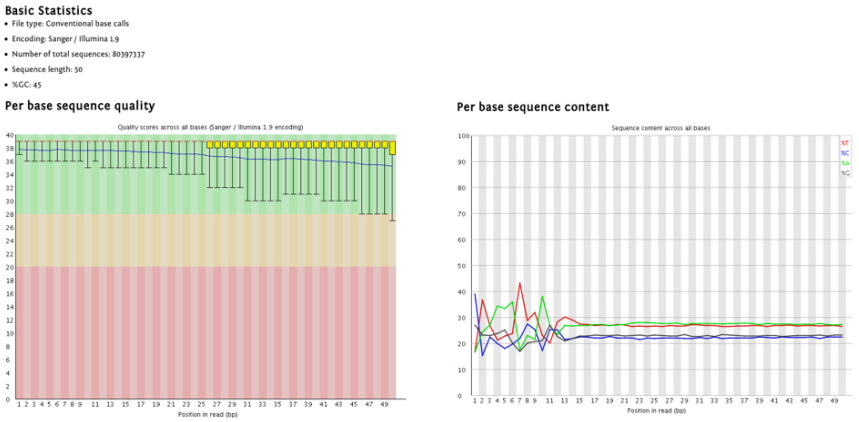
●論文が出ているということは、査読を経ているので、
一定のデータの質は担保されている（はず）

◇http://g86.dbcls.jp/kusarinoko



鎖鋸（kusarinoko）

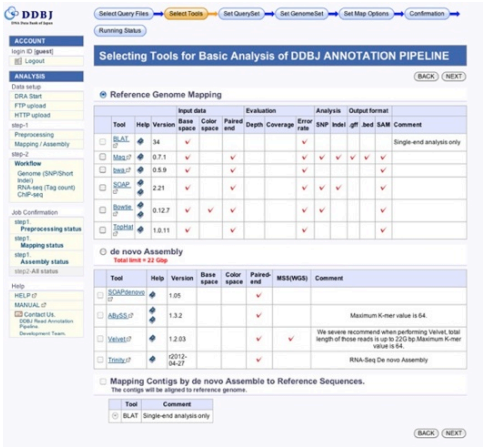
◇独自に”FastQC”をかけて、
それぞれのデータの質を評価



DDBJ Read Annotation Pipeline

◇新型シーケンサデータの解析パイプライン

○http://p.ddbj.nig.ac.jp/



Ensembl

Ensembl Genome Browser

Ensembl

http://www.ensembl.org/

◇Wellcome Trust Sanger Instituteと
European Bioinformatics Institute (EBI)が
共同開発しているゲノムブラウザ

○さまざまなアノテーション情報をゲノム上で見る
（ブラウズ）することが可能

○脊椎動物が中心

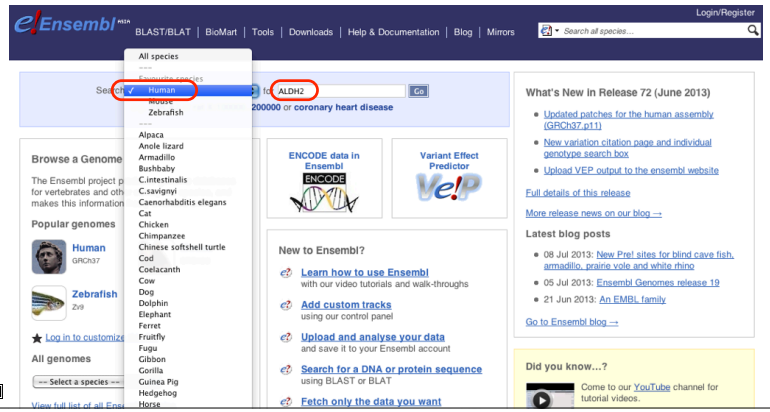
○EnsemblBacteria, EnsemblFungi, EnsemblMetazoa,
EnsemblPlants, EnsemblProtistsというのもある

◇他にUCSC Genome Browserが有名

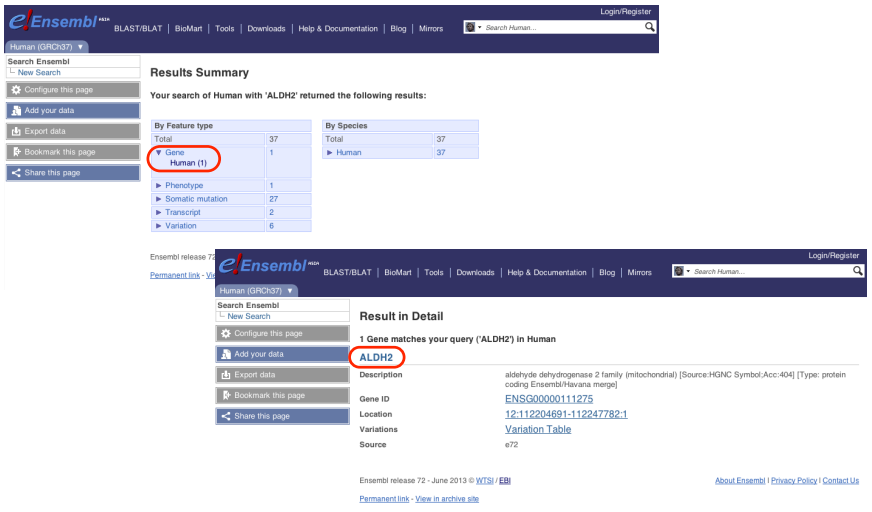
実習2：Ensemblを使ってみる

◇ヒトアセトアルデヒドデヒドロゲナーゼ2（ALDH2）周辺のアノテーション情報を眺める

生物種を選択して ALDH2 で検索

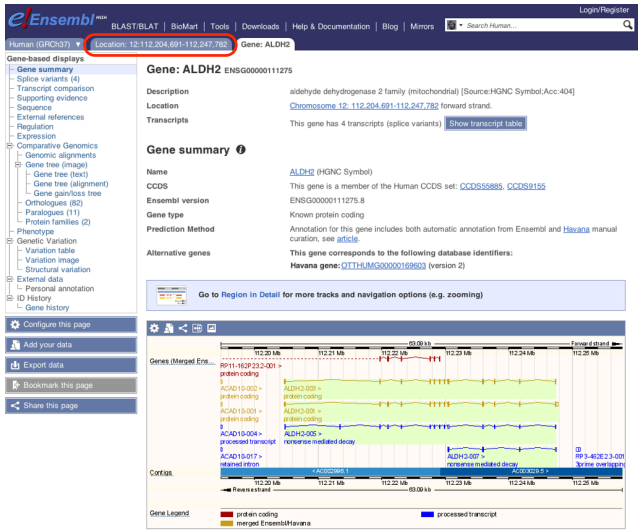


対象遺伝子の選択

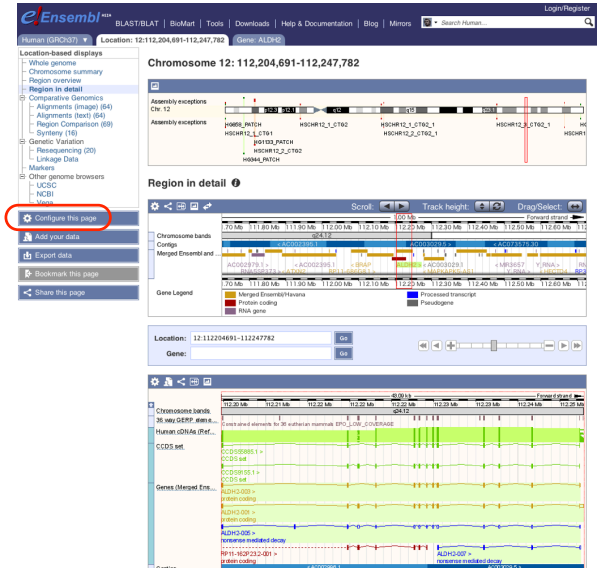


サマリページの表示

ゲノムブラウザで表示

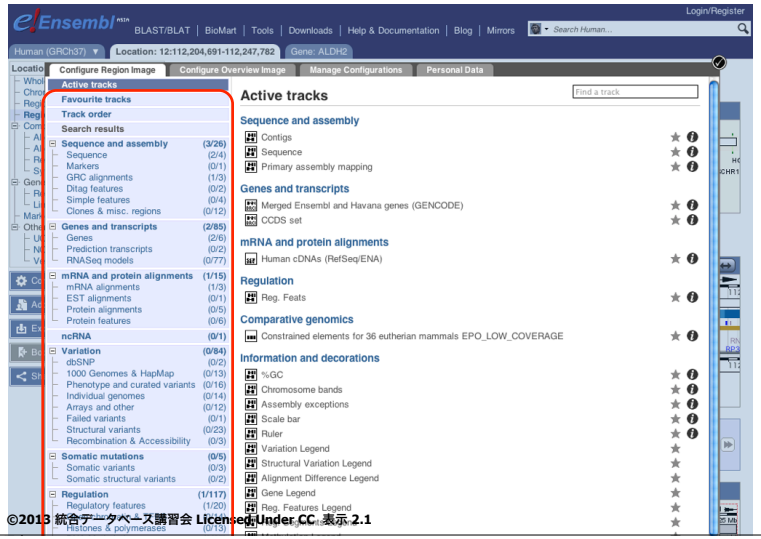


ゲノムブラウザでの表示



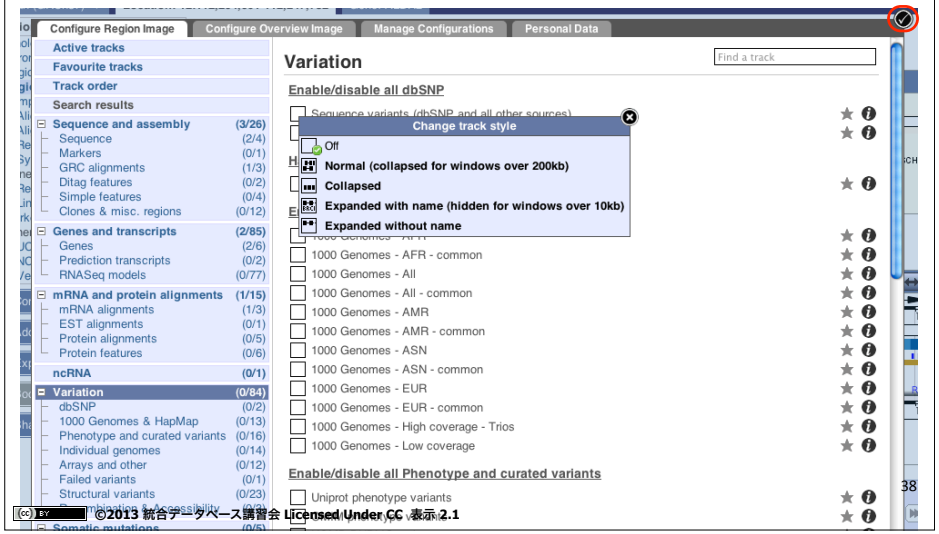
表示するアノテーションの指定

表示させる項目を選択

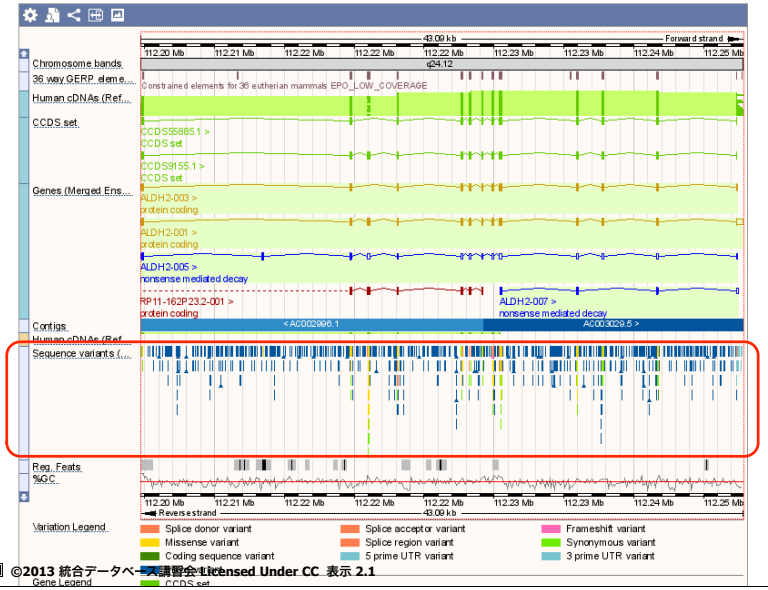


表示するアノテーションの指定

設定後、✓マークをクリックすると自動更新される



選択したアノテーション情報が表示される



自分のデータをゲノムブラウザ上に表示

既存のアノテーションと
自分のデータをゲノム上で
重ね合わせて表示することが可能

BioMart

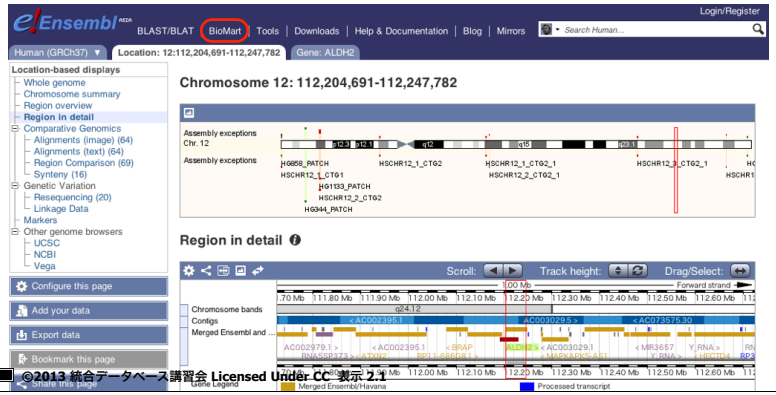
BioMart

http://www.biomart.org/

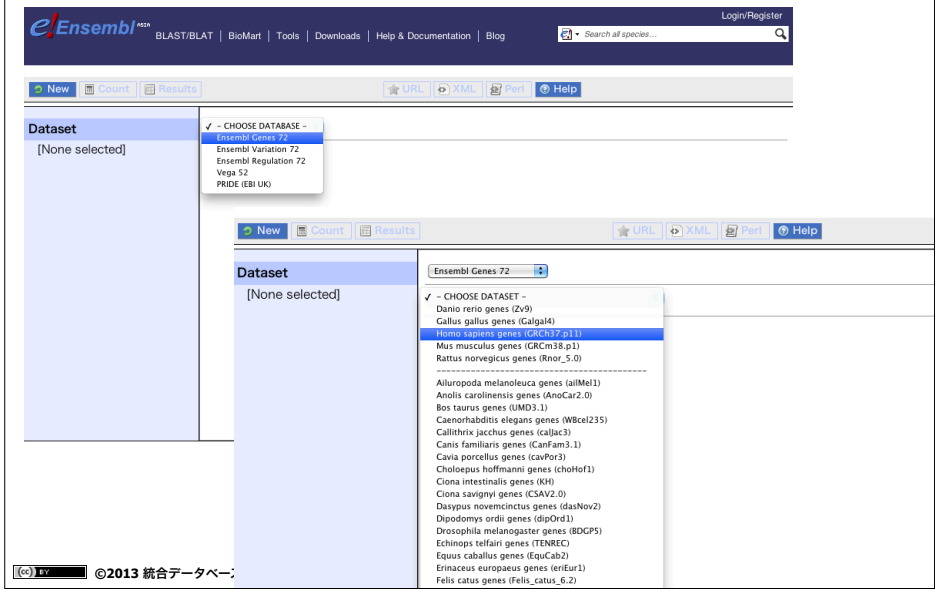
- Ontario Institute for Cancer Research (OiCR)とEuropean Bioinformatics Institute (EBI)が共同開発しているデータ管理システム
- さまざまなデータベースのGUIデータ取得インターフェイスとして利用されている
  - Ensembl, UniProt, WormBase, KazusaMartなど

実習3：BioMartを使ってEnsmblからデータ取得

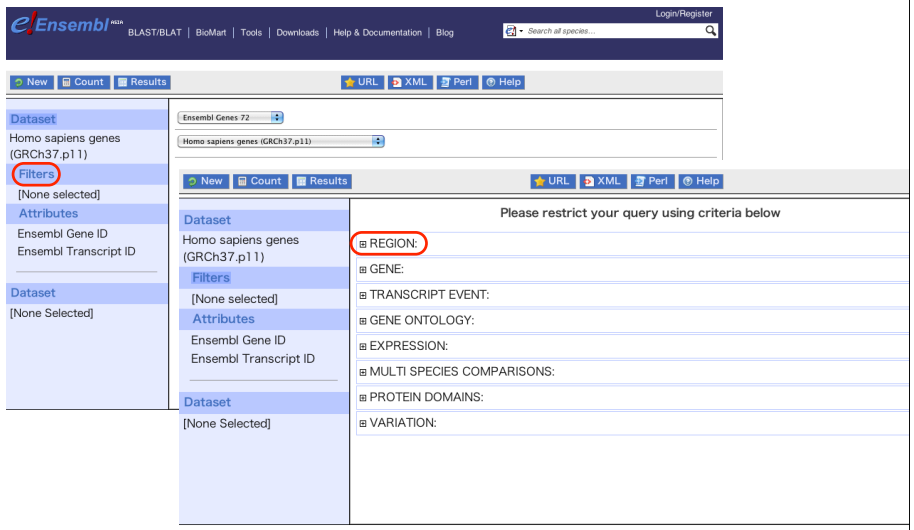
- ヒトゲノムの22番染色体にある遺伝子について、転写開始点から上流の配列1kbをFASTA形式で取得する



データセットを選択



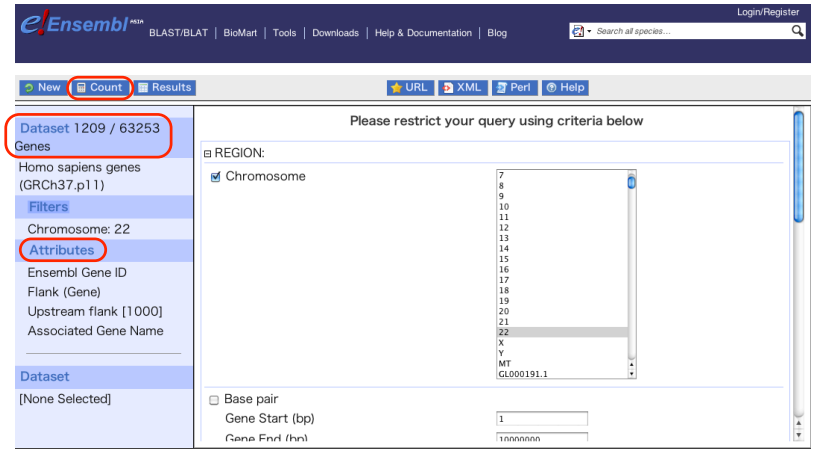
取得したい遺伝子セットを指定



フィルタ機能で遺伝子を絞り込む

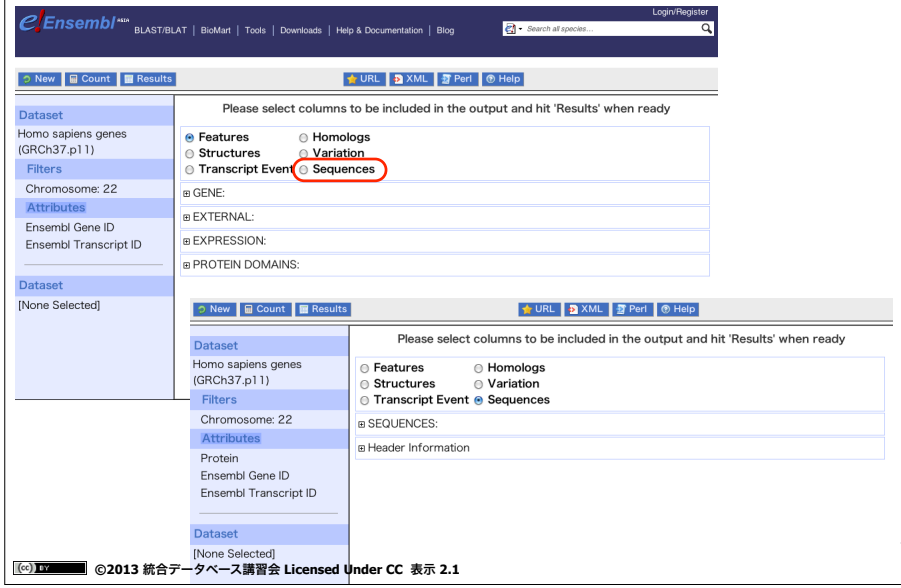
取得したい遺伝子セットを指定

22番染色体を指定（複数の条件を組み合わせることも可能）
Countで絞りこまれた遺伝子数を表示できる



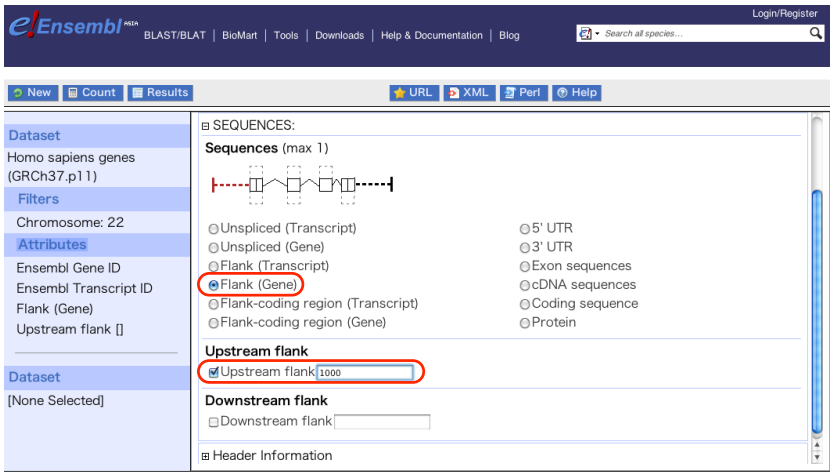
Attributesで取得する情報を指定する

取得したいデータの種類を指定



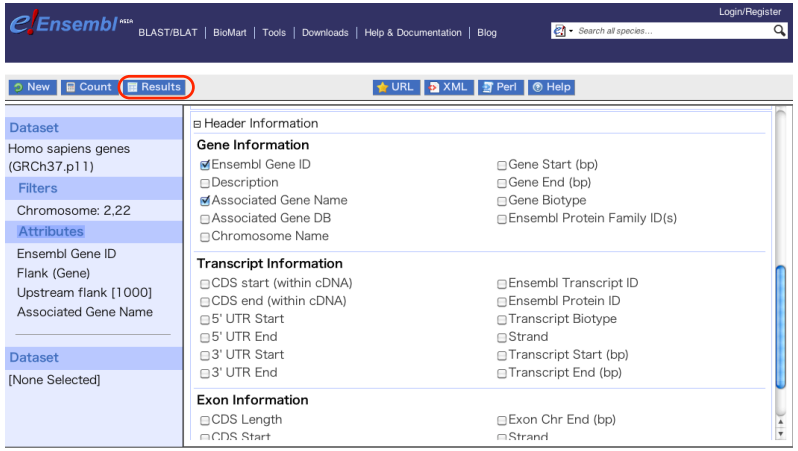
取得したいデータの種類を指定

遺伝子の上流域1000bpを指定



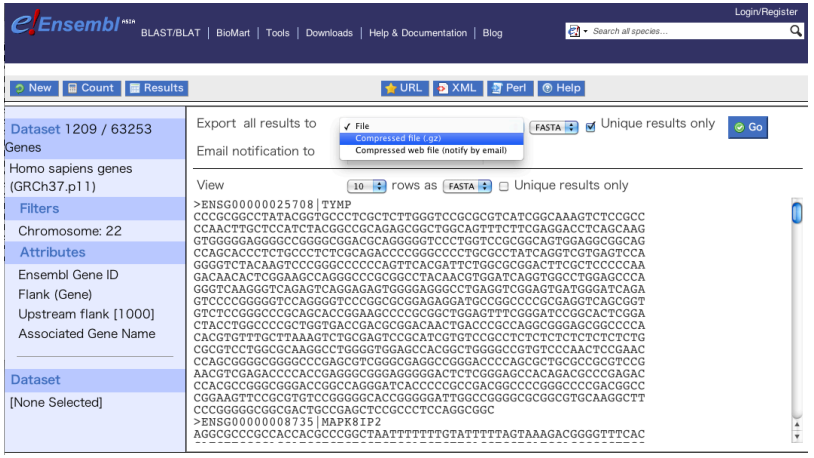
取得したいデータの種類を指定

ヘッダ情報（FASTAのコメント行）の指定



Resultsで取得データを表示

データのダウンロード



GOLD

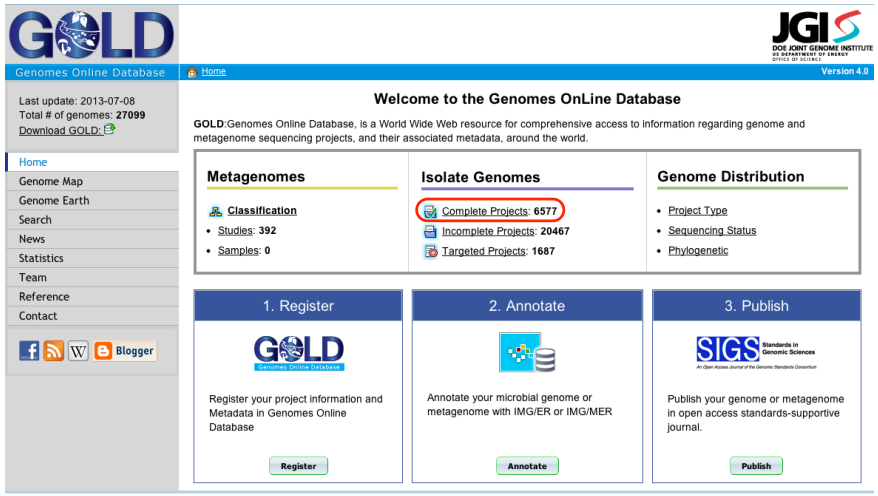
Genome OnLine Database

GOLD: Genome OnLine Database

http://www.genomesonline.org/

- ◇1995年のHaemophilus influenzaeゲノムの解読以来、現在までにゲノムが解読された／解読中の生物を集めたDB
- ◇日本で最初に解読されたゲノムは光合成細菌であるSynechocystisのゲノムで、1996年（かずさDNA研究所）
- ◇真核生物最初のゲノムは1997年のSaccharomyces cerevisiae（出芽酵母）
- ◇メタゲノムプロジェクト（ある環境にいる微生物のゲノムをまとめて読む）もある
  - ○ 腸内細菌、皮膚常在菌、海水、排水、etc...

実習4-1：ゲノム解読済みの生物を眺める

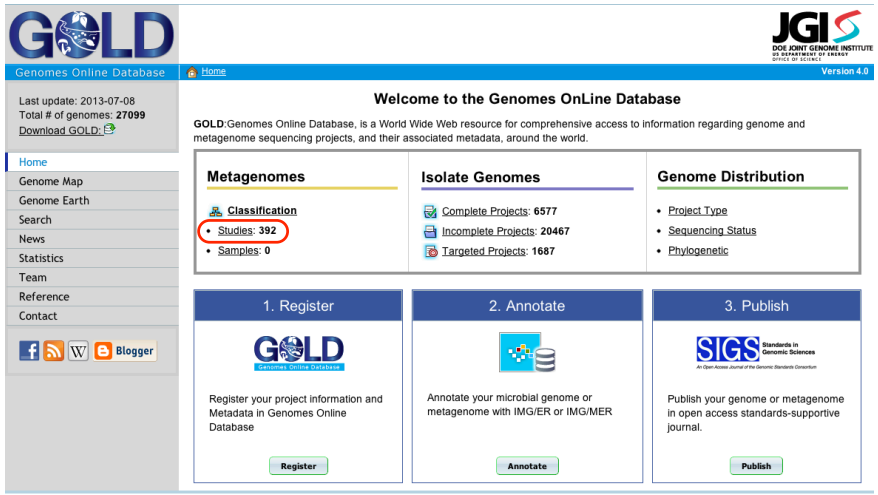


実習4-1：ゲノム解読済みの生物を眺める

ID	生物種名	分類	ドメイン	ゲノムサイズ	染色体数	GC含量	データ 実施機関	データベース 論文	終了日	連絡先
Gc00008	Escherichia coli K-12, MG1655	PROTEOBACTERIA-GAMMA	B	4640 Kb 4497 orfs	1		51%	U00096 U00096 Univ of Wisconsin-Madison	Science 277:1453-1474	1997-09-05	Blattner FR
Gc00007	Helicobacter pylori 26695	PROTEOBACTERIA-EPSILON	B	1668 Kb 1624 orfs	1		39%	AE000511 AE000511 J. Craig Venter Institute	Nature 388:539-547	1997-08-07	Tomb JF
Gc00006	Saccharomyces cerevisiae S288C	FUNGI-ASCOMYCOTA	E	12157 Kb 6273 orfs	16		38%	A1011856 D50617 X59720 International Collaboration	Nature 387.5-105	1997-05-29	
Gc00005	Mycoplasma pneumoniae M129	TENERICUTES	B	816 Kb 733 orfs	1		40%	U00089 Univ of Heidelberg	Nucleic Acids Research 24:4430-4449	1996-11-15	Herrmann R
Gc00004	Methanocaldococcus jannaschii DSM 2661	EURYARCHAEOTA-METHANOCOCCI	A	1740 Kb 1844 orfs	1	2	31%	L77117 J. Craig Venter Institute Univ of Illinois at Urbana-Champaign	Science 273:1058-1073	1996-09-28	Venter CJ
Gc00003	Synechocystis sp. PCC 6803	CYANOBACTERIA	B	3947 Kb	1	4		BA000022 Kazusa DNA Research Institute	DNA Research 3:109-136	1996-06-30	Tabata S
Gc00002	Mycoplasma genitalium G37	TENERICUTES	B	580 Kb 525 orfs	1		32%	L43967 J. Craig Venter Institute	Science 270:397-403	1995-10-20	Fraser CM
Gc00001	Haemophilus influenzae Rd (KW20)	PROTEOBACTERIA-GAMMA	B	1830 Kb 1762 orfs	1		38%	L42023 L42023 J. Craig Venter Institute	Science 269:496-512	1995-07-28	Venter CJ

E 真核生物　B 原核生物　A アーキア

実習4-2：メタゲノムプロジェクトを眺める



実習4-2：メタゲノムプロジェクトを眺める

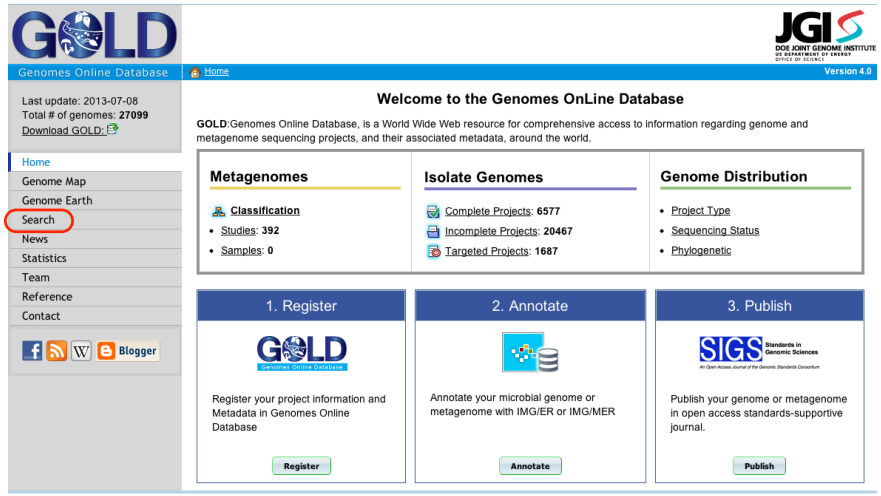
Metagenome Studies:392

Engineered: 24　Environmental: 230　Host-associated: 128

<< first　< prev　1 2 3 4　next >　last >>　100

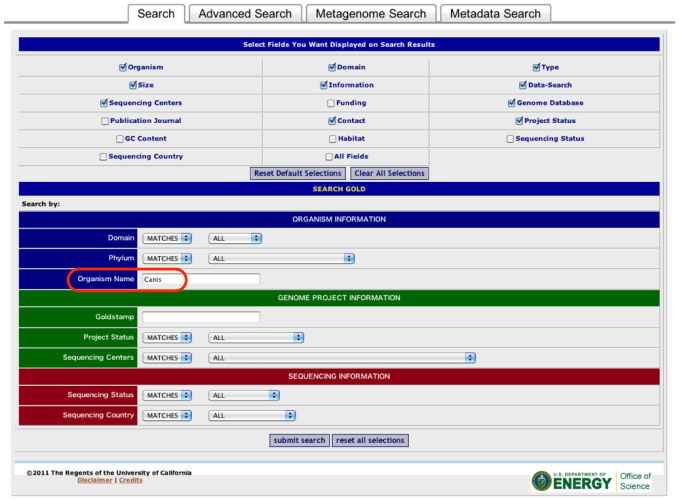
GOLD ID	METAGENOME STUDY	SAMPLE COUNT	ECOSYSTEM	ECOSYSTEM CATEGORY	ECOSYSTEM TYPE	ECOSYSTEM SUBTYPE	SPECIFIC ECOSYSTEM	SEQUENCING CENTER	PUBLICATION	CONTACT
Gm00100	Human fecal microbiome from healthy Japanese infants and adults, Univ of Tokyo		Host-associated	Human	Digestive system	Large intestine	Fecal	Univ of Tokyo	DNA Research 14, 169-81 2007-12-05	Hattori M
Gm00013	Nasutitermes gut microbiome from Costa Rica		Host-associated	Arthropoda	Digestive system	Hindgut	Unclassified	DOE Joint Genome Institute	Nature 450, 560-565 2007-11-26	Hugenholtz P
Gm00012	Wastewater bioreactor microbial communities from Singapore and Univ of Illinois at Urbana, that are terephthalate-degrading		Engineered	Wastewater	Industrial wastewater	Unclassified	Unclassified	DOE Joint Genome Institute National University of Singapore		Liu Wen-Tso
Gm00011	Endophytic crenarchaeal communities of Plant roots		Host-associated	Plants	Rhizoplane	Endophytes	Unclassified	DOE Joint Genome Institute		Tringe SG
Gm00053	Marine viral communities from Chesapeake Bay		Environmental	Aquatic	Marine	Intertidal zone	Unclassified	Univ of Delaware		Wommack E
Gm00073	Soil Fungal communities from the Arctic		Environmental	Terrestrial	Soil	Unclassified	Unclassified	Univ of Alaska Fairbanks		Taylor D
Gm00072	Marine viroplankton communities from surface and estuarine waters		Environmental	Aquatic	Marine	Neritic zone	Unclassified	J. Craig Venter Institute		Williamson S
Gm00070	Elephant Grass decomposer microbial community		Engineered	Solid waste	Composting	Grass	Bioreactor	DOE Joint Genome Institute		Tringe SG
Gm00069	Gut microbiome of Tipula abdominalis		Host-associated	Arthropoda	Digestive system	Unclassified	Unclassified	DOE Joint Genome Institute		Tringe SG
Gm00068	Anoplophora glabripennis gut microbiome from Worchester, MA		Host-associated	Arthropoda	Digestive system	Unclassified	Unclassified	DOE Joint Genome Institute		Tringe SG
Gm00067	Poplar biomass bioreactor microbial communities from Brookhaven National Lab, NY		Engineered	Solid waste	Composting	Wood	Bioreactor	DOE Joint Genome Institute		Daniel van der Lelie

実習4-3：あの生物のゲノムは読まれているのか調べてみる

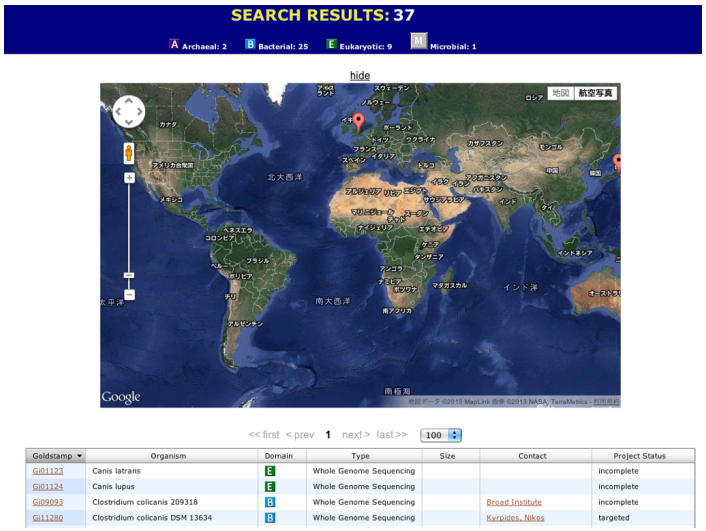


実習4-3：あの生物のゲノムは読まれているのか調べてみる

Organism Nameに学名を入れて検索



実習4-3：あの生物のゲノムは読まれているのか調べてみる



モデル生物のゲノムデータベース

代表的なモデル生物のゲノムデータベース

◇ヒト H-InvDB Annotated Human Gene Database
- H-InvDB: H-Invitational Database
- http://www.h-invitational.jp/

◇ゼブラフィッシュ ZFIN
- ZFIN: The Zebrafish Model Organism Database
- http://zfin.org/

◇マウス MGI
- MGI: Mouse Genome Informatics
- http://www.informatics.jax.org/

◇シロイヌナズナ tair
- TAIR: The Arabidopsis Information Resource
- http://www.arabidopsis.org

◇ショウジョウバエ
- FlyBase: FlyBase (A Database of Drosophila Genes & Genomes)
- http://flybase.org/

◇酵母 SGD Saccharomyces GENOME DATABASE
- SGD: Saccharomyces Genome Database
- http://www.yeastgenome.org/

◇線虫
- WormBase
- http://www.wormbase.org

◇シアノバクテリア CyanoBase
- CyanoBase: Genome database for Cyanobacteria
- http://genome.microbedb.jp/cyanobase

おわりに

◇DDBJ、Ensembl、BioMart、GOLD共に、時間の都合で紹介できなかった機能がたくさんありますので、統合TVなどを参考にしながらぜひ使い倒してみてください