

データ科学ワークショップ 2009/8/21 @ 札幌

生命情報学が直面する 大規模ゲノムデータ時代の課題

片山 俊明 <ktym@hgc.jp>

東京大学 医科学研究所
ヒトゲノム解析センター
ゲノムデータベース分野 助教（金久研究室）



自己紹介

オープンソースで学ぶ
バイオインフォマティクス
まだ売れ残ってます→



- KEGG (Kyoto Encyclopedia of Genes and Genomes)
 - 京大/東大の金久研で開発されている遺伝子とパスウェイのDB
 - オートログクラスタ KEGG OC
 - 1000生物種400万遺伝子 → 30万クラスタ (+40万シングルトン)
 - ウェブサービス KEGG API (SOAP/WSDL)
 - 分散ゲノムアノテーション KEGG DAS



<http://www.genome.jp/kegg/>

- BioRuby
 - オープンバイオ研究会
 - IPA未踏ソフトウェア創造事業採択



<http://bioruby.org/>
<http://open-bio.jp/>

- ライフサイエンス統合データベースプロジェクト
 - TogoDB, TogoWS, BioHackathon



<http://dbcls.jp/>

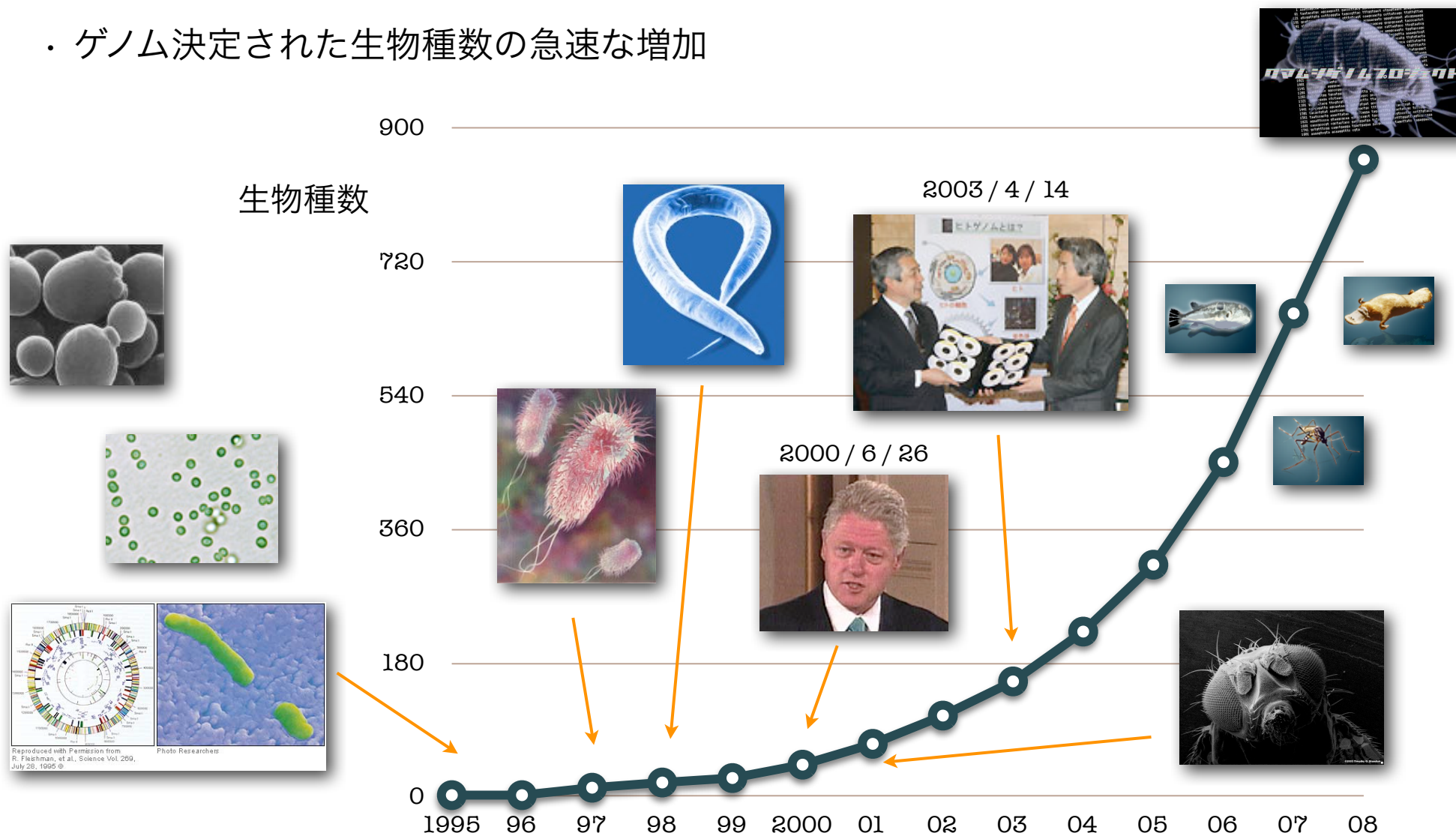
- クマムシゲノムプロジェクト

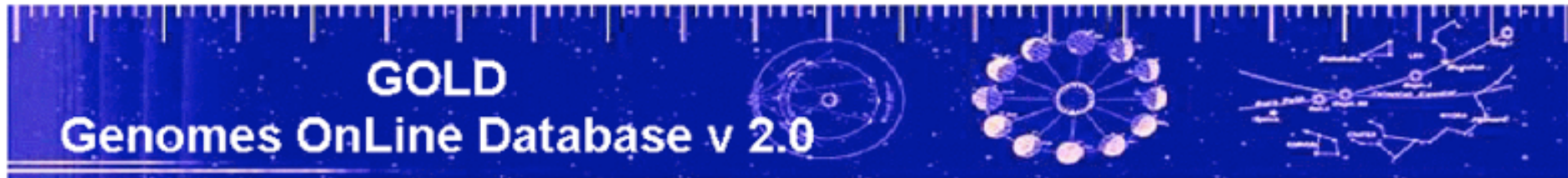


<http://kumamushi.org/>

1000生物種以上のゲノム解読が完了 (2009)

- ゲノム決定された生物種数の急速な増加





Contact: Genomesonline	Last Update: August 16, 2009	Location www.genomesonline.org
1069 Published Complete Genomes	Search GOLD: 5425 genome projects	168 Metagenomes
109 Archaeal Ongoing Genomes	2960 Bacterial Ongoing Genomes	1119 Eukaryotic Ongoing Genomes

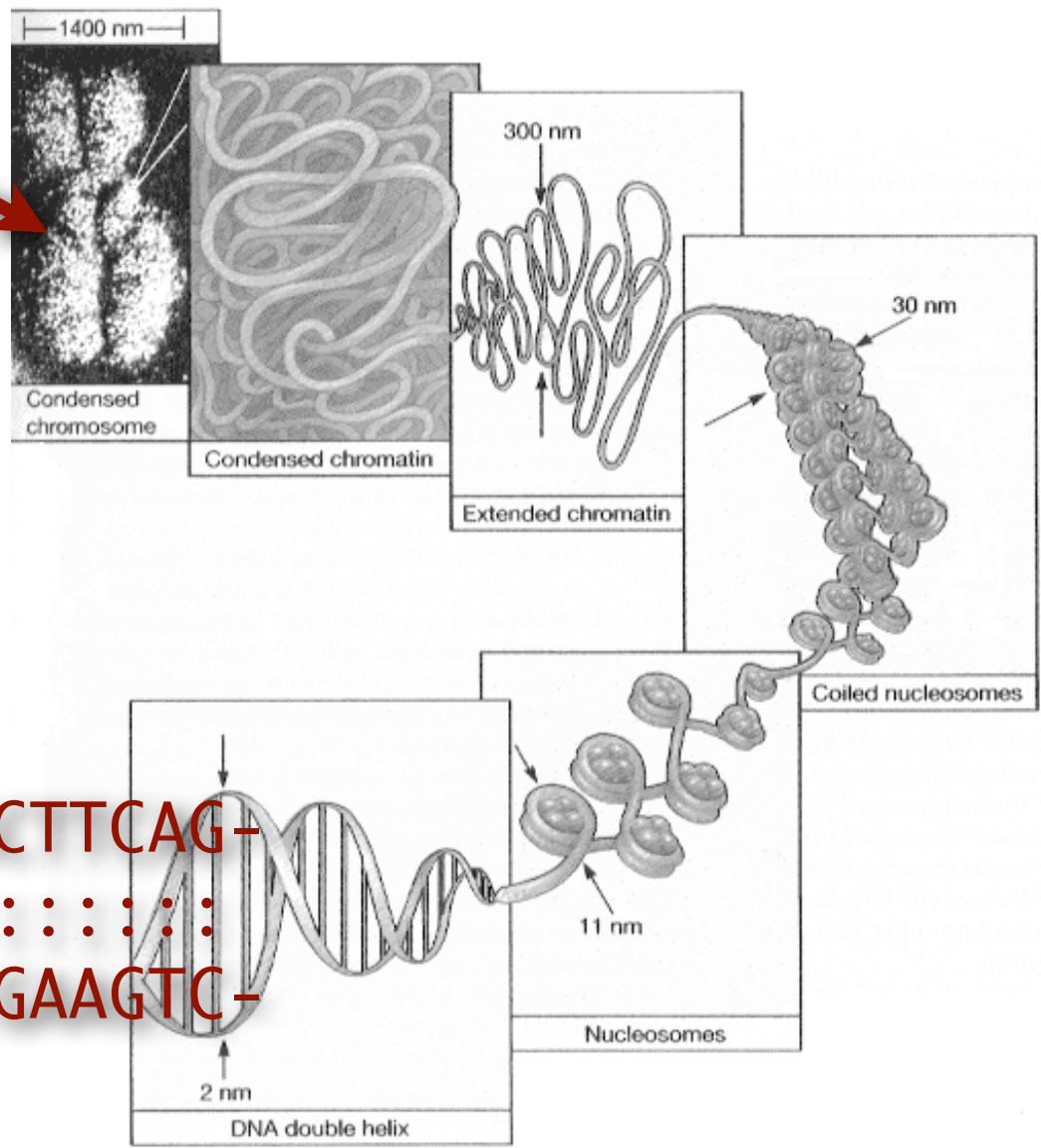
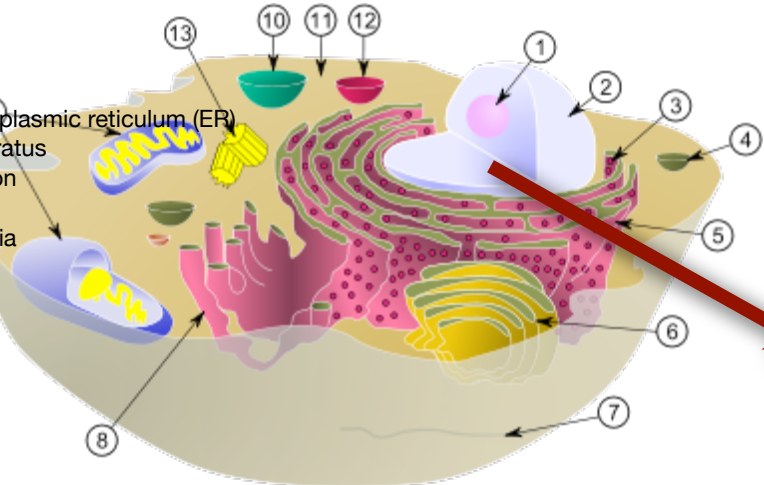
	Right-click to save all data: DOWNLOAD	
--	---	--

ゲノム解析とデータベース化

DNA sequences and annotations (~TB data)

[http://en.wikipedia.org/wiki/Cell_\(biology\)](http://en.wikipedia.org/wiki/Cell_(biology))

- (1) nucleolus
- (2) nucleus
- (3) ribosome
- (4) vesicle
- (5) rough endoplasmic reticulum (ER)
- (6) Golgi apparatus
- (7) Cytoskeleton
- (8) smooth ER
- (9) mitochondria
- (10) vacuole
- (11) cytoplasm
- (12) lysosome
- (13) centrioles



Genome browser



-ATGCTTCAG-

.....

-TACGAAGTC-

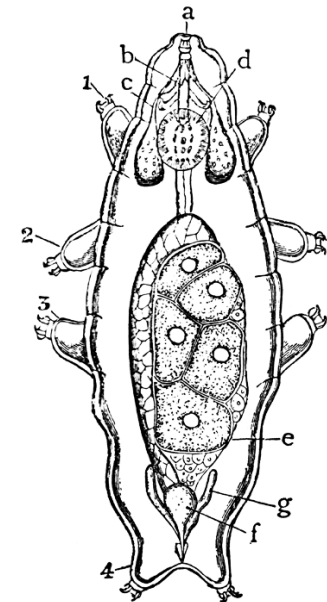


クマムシゲノムプロジェクトの実現にむけて

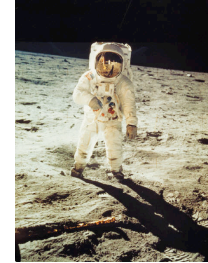
片山 俊明

東京大学医科学研究所ヒトゲノム解析センター
ゲノムデータベース分野 助教（金久研究室）

k@kumamushi.org



クマムシ



- 緩歩動物門 - 線虫と昆虫の間くらいに位置する1mm以下のかわいい動物
 - その辺のコケから深海や南極まで、地球上どこにでもいる！

- 乾燥すると究極の極限環境耐性

- 高温：+151℃ (ラーム, 1921)
- 低温：-273℃ (ベクレル, 1950)
- 高圧：7.5GPa (750,000mの水圧)
- 低圧： 5×10^{-4} Pa (ほぼ真空)
- 放射線：X線 100,000 Gy
(ヒト致死量の1,000倍)
ガンマ線 7,000Gy
Heイオン 8,000Gy
- 化学物質：アルコール (消毒剤)
臭化メチル (防虫剤)

人畜無害！



クマムシに対する社会的な注目も高い

・ テレビ

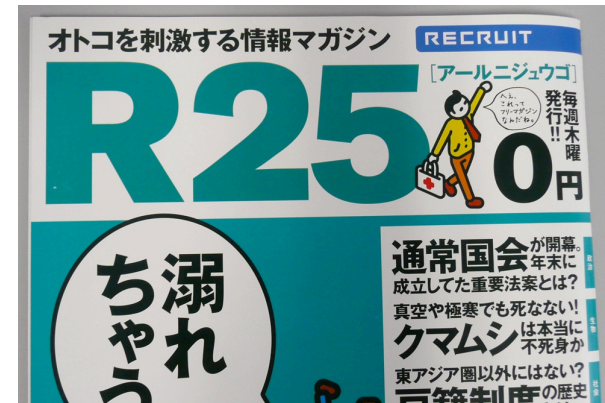
- ・ どうぶつ奇想天外 (1994)
- ・ たけしの万物創世紀 (1998)
- ・ トリビアの泉 (2003)
- ・ 木村カエラ sakusaku (2005)
- ・ むしまるQゴールド
- ・ ゆるなび (2007)
- ・ NHK科学大好き土よう塾 (2007)
- ・ NHKミクロワールド (2008)
- ・ ベストハウス123 (2008)

・ ラジオ

- ・ 中川翔子のギザサイエンス

・ 雑誌, 新聞

- ・ 遺伝
- ・ AERA
- ・ SPA!
- ・ WIRED NEWS
- ・ R25



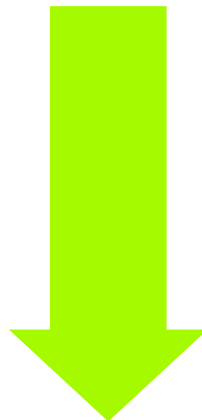
第1回クマムシ研究会

- ・ 2006年12月2日
 - ・ 国内の主要なクマムシ研究者が集結
 - ・ 参加者 - 約100人
 - ・ 東大理学部2号館講堂が満員に

・ テーマ

- ・ クマムシ分類学 (阿部)
- ・ 培養の歴史 (鈴木)
- ・ 超高压耐性 (三枝)
- ・ 極限環境耐性 (堀川)
- ・ 分子メカニズム (國枝)

マクロ (動物学)



ミクロ (分子生物学)

The 1st Domestic Workshop on Tardigrades in Japan

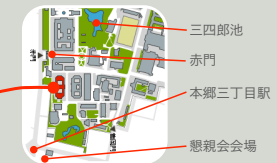
KUMAMUSHI WS#1



第1回!クマムシ研究会

日時: 2006年12月2日 (土曜日) 13:00 - 18:00

会場: 東京大学・理学部2号館・4階大講堂



最近、ひそかにクマムシに対する人気が高まっているようです。この機会に、一般向けのクマムシ研究会を開催することになりました。クマムシ専門家によるクマムシ観察の基礎から最先端のクマムシ学まで幅広い講演を聞くことができるほか、クマムシの標本や書籍など様々なクマムシ関連アイテムを一度に見られる絶好の機会です。

プログラム

- 13:00-14:00 「クマムシ培養の歴史」 (鈴木)
- 14:00-14:40 「クマムシの種を見分ける」 (阿部)
- 14:40-15:00 「クマムシ?! 本の秘密」 (塩田)
- 15:00-15:20 「クマグルミの秘密」 (会津)
- 15:20-15:50 休憩・展示・観察
- 15:50-16:30 「クマムシはどこまで耐えられるか」 (堀川)
- 16:30-17:10 「オニクマムシの超高压への耐性実験」 (三枝)
- 17:10-17:50 「クマムシの乾燥耐性の分子機構を探る」 (國枝)



入場無料!



「クマムシ?!」

2006年8月をはじめの日本語によるクマムシ本が岩波書店より出版されました。(ASIN:4000074628)

http://www.tardigrada-symposium-2009.org/

第11回 国際クマムシ学会


11th International Symposium on Tardigrada

http://www.tardigrada-symposium-2009.org/general.html

11th International Symposium on Tardigrada

General Expression of Interest Presentations Program Registration Meeting Office Travel

Welcome



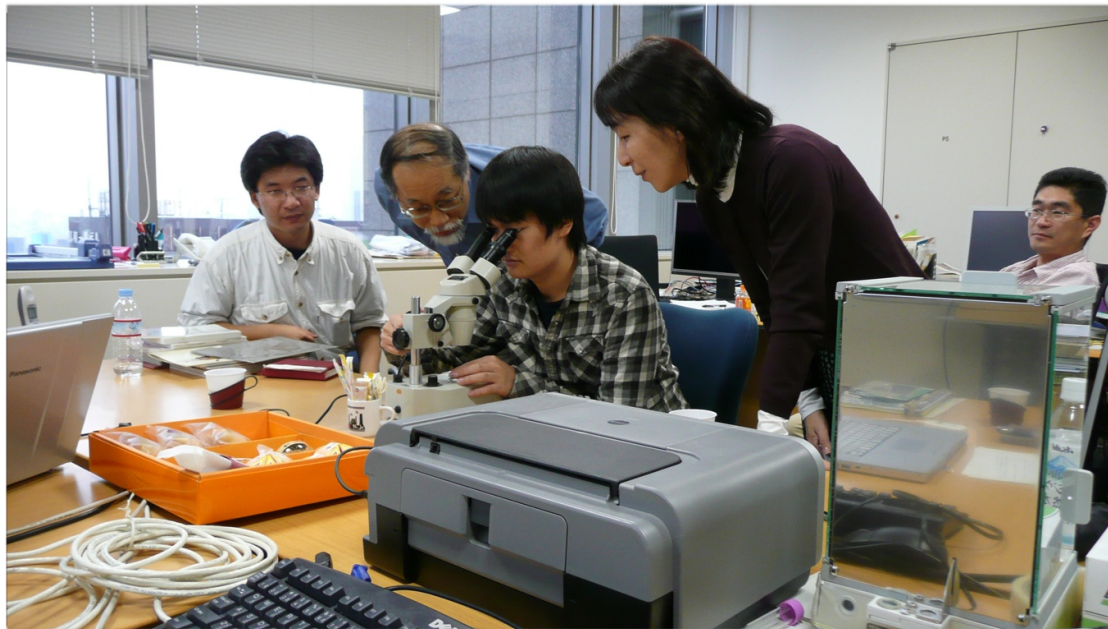
Dear Colleagues and Friends,

We are pleased to welcome you to the information pages for the 11th International Tardigrade-Symposium 2009 which will take place in Tübingen from 3. - 6. August.

The Symposium will be taking place in Germany for the third time and for good reasons: 80 years ago the German zoologist Ernst Marcus (1893-1968) discovered and described the Heterotardigrada. However, there is another anniversary which we wish to celebrate. 50 years ago, in 1959 David Keilin coined the term 'cryptobiosis' (hidden life) and defined it as "the state of an organism when it shows no visible signs of life and when its metabolic activity becomes hardly measurable, or comes reversibly to a standstill".

クマムシゲノムプロジェクト

- ・ 2000年頃 ハエからの類推で > 10 億円 (試算：森先生)
- ・ 2008年 特定領域研究「ゲノム」 比較ゲノム < 数千万
 - ・ 乾燥,放射線耐性のトランスクリプトーム,プロテオーム解析のための基盤
 - ・ 比較ゲノムによる動物進化の解析 (線形動物門と節足動物門の間)



..... **1998**

Started my personal fan site for water bears (in Japanese)
→ genome project has been my dream from the beginning !!!



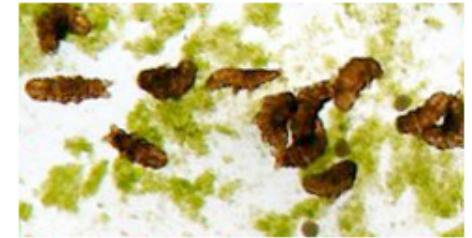
Time-line of our genome project

2006

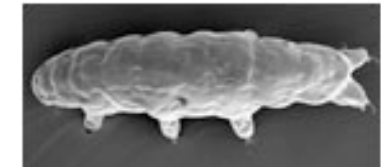
Isolation of *Ramazzottius varieronatus* in Hokkaido, Japan by Dr. Horikawa
→ development of the cultivation method

2007

The 1st domestic Tardigrada workshop in Japan
→ ~100 participants (incl. non-researchers)



Establishment of the **YOKOZUNA-1** strain
from *Ramazzottius cf. varieronatus*



2008

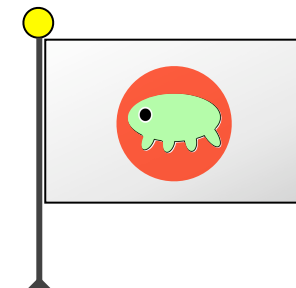
Application to MEXT (Japanese ministry) for our genome project
→ w/i Grant-in-Aid for "Genome Science"

Genome project started !!!



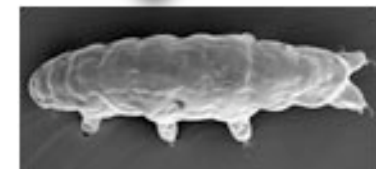
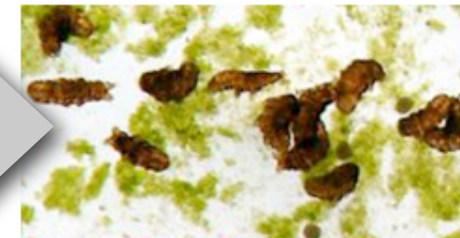
2009

The 1st draft genome sequence is assembled



The 11th International Symposium on Tardigrada

横綱クマムシ



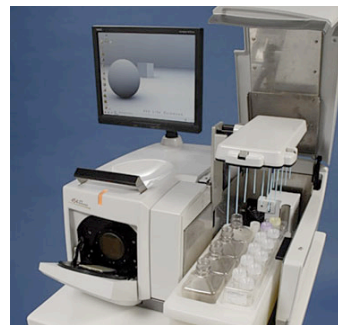
Human genome in 5min?!

Next generation DNA sequencers

ABI 3730	<i>700b x 48 / hour</i>	<i>700-1100bp</i>	<i>800Kb / day</i>	<i>2002-</i>
ABI 3730XL	<i>700b x 96 / hour</i>	<i>700-1100bp</i>	<i>1.6Mb / day</i>	<i>2002-</i>
Roche 454 GS20	<i>20Mb / 5 hours</i>	<i>100bp</i>	<i>100Mb / day</i>	<i>2005-</i>
Roche 454 FLX	<i>100Mb / 7.5 hours</i>	<i>200-500bp</i>	<i>300Mb / day</i>	<i>2006-</i>
Illumina Solexa	<i>1.5Gb / 2.5 days</i>	<i>50bp</i>	<i>600Mb / day</i>	<i>2006-</i>
ABI SOLiD	<i>4Gb / 8 days</i>	<i>35bp</i>	<i>500Mb / day</i>	<i>2007-</i>
Helicos	<i>2Gb / day</i>	<i>150bp</i>	<i>2Gb / day</i>	<i>2008?</i>
PacBio SMRT	<i>100Gb / hour</i>	<i>-25000bp</i>	<i>2.4Tb / day</i>	<i>2010?</i>



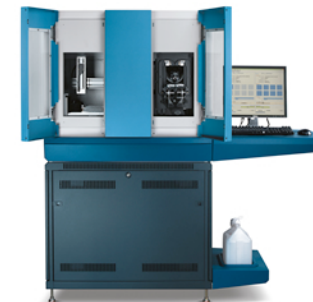
3730XL



454 Life Sciences



Solexa



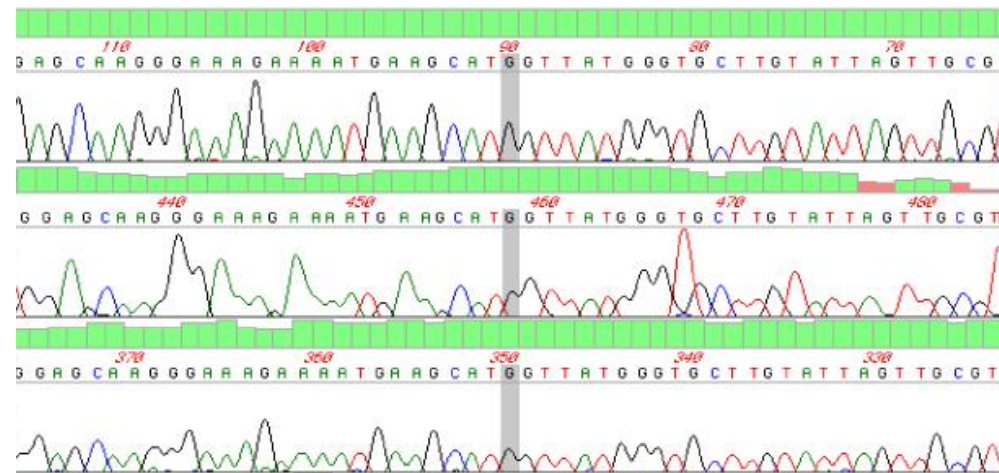
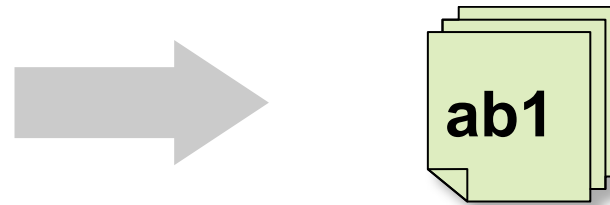
SOLiD



HeliScope



1st generation DNA sequencer (ABI3730xl - Sanger method)



ABI KB, phred etc.

```
>GKMWno465_a17.b1 CHROMAT_FILE: GKMWno465_a17.b1 PHD_FILE: GKMWno465_a17.b1.phd.1 CHEM: term DYE: big TIME: Mon
Apr 6 10:55:24 2009 TEMPLATE: GKMWno465_a17 DIRECTION: fwd
ggggccagtgaggatgatctaagatcccggaacagaagccgattcatcatcttggactgcaatgccatctgttatttggatcatcatgataatcgccacc
atcttaacggagtagttttgtacgttttcgccggcatcctcacctaaggacagctttcaacgtctacataatcaatttgcctatggctaacctgatcg
aaacattcatgatgatgccgctgaacatagctacctacacggtcacactggtggctaggttccacagcctgctctgtgttcatttatggtctgtctct
ccaagctgtcgcgatcaactcccaccttctcatcaccataaacgcatgtgggcagtggtccgctcctgtttcgtaccgaacctaccacagggaaaagggtg
tcgatctgtctttgcattttcatttggctgtacgttcacgtcgtccataccccgggatacgtcatcgacgtgtttactaccgtcctccagttgaaatcg
gctgttttgtcaacggaatgatgcagtatgggtgggtttcatgccatgcagttctgtctattcaaaattccagtcctggctataatatgctgctaccctta
```



Whole genome shotgun sequencing and assembly

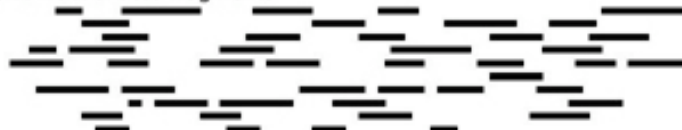
a) Multiple copies of genome



b) Sheared random fragments



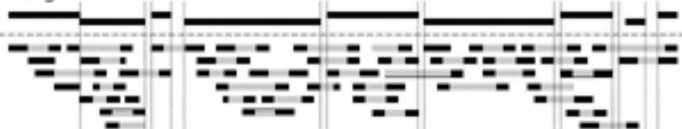
c) Size fractionated fragments



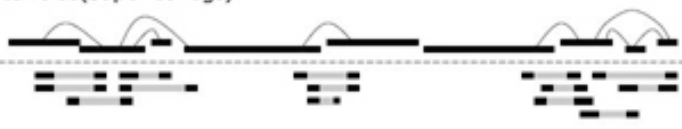
d) Reads



e) Contigs



f) Scaffolds(Super contigs)



- 57Mb of the genome

- $\approx 80\%$ of estimated genome size (70Mb)

- 383,286 reads

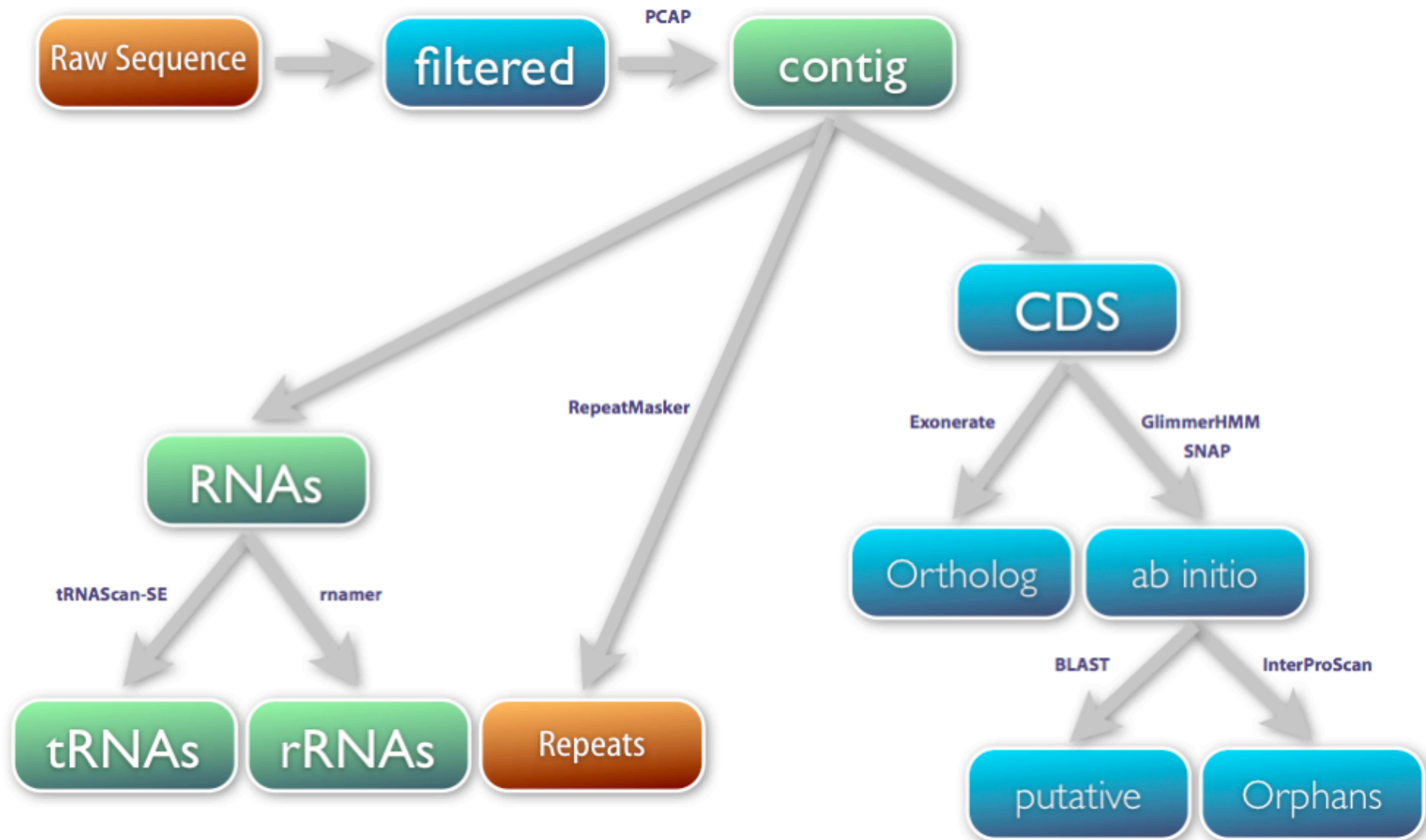
- 15,000 fosmids

- 5,671 contigs

- 1,717 supercontigs

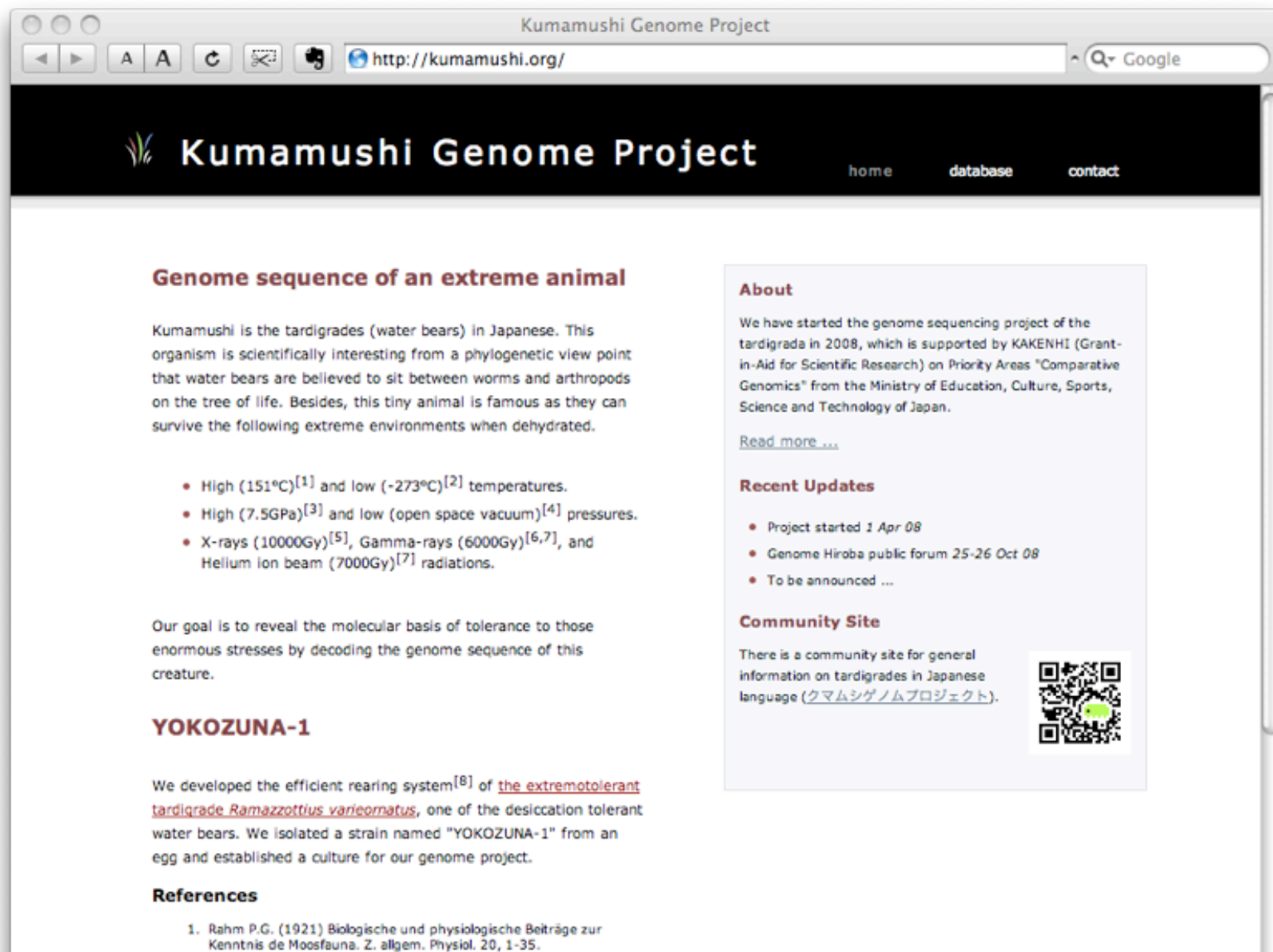


Our genome assembly/annotation pipeline



http://kumamushi.org/

クマムシゲノムプロジェクト ウェブサイト



Kumamushi Genome Project

home database contact

Genome sequence of an extreme animal

Kumamushi is the tardigrades (water bears) in Japanese. This organism is scientifically interesting from a phylogenetic view point that water bears are believed to sit between worms and arthropods on the tree of life. Besides, this tiny animal is famous as they can survive the following extreme environments when dehydrated.

- High (151°C)^[1] and low (-273°C)^[2] temperatures.
- High (7.5GPa)^[3] and low (open space vacuum)^[4] pressures.
- X-rays (10000Gy)^[5], Gamma-rays (6000Gy)^[6,7], and Hellum ion beam (7000Gy)^[7] radiations.

Our goal is to reveal the molecular basis of tolerance to those enormous stresses by decoding the genome sequence of this creature.

YOKOZUNA-1

We developed the efficient rearing system^[8] of the extremotolerant tardigrade *Ramazzottius varieornatus*, one of the desiccation tolerant water bears. We isolated a strain named "YOKOZUNA-1" from an egg and established a culture for our genome project.

References

1. Rahm P.G. (1921) Biologische und physiologische Beiträge zur Kenntnis de Moosfauna. Z. allgem. Physiol. 20, 1-35.

About

We have started the genome sequencing project of the tardigrada in 2008, which is supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas "Comparative Genomics" from the Ministry of Education, Culture, Sports, Science and Technology of Japan.


[Read more ...](#)

Recent Updates

- Project started 1 Apr 08
- Genome Hiroba public forum 25-26 Oct 08
- To be announced ...

Community Site

There is a community site for general information on tardigrades in Japanese language ([クマムシゲノムプロジェクト](#)).





Genome browser (GMOD/GBrowse)

Kumamushi Genome Project home database contact

6.001 kbp の範囲を Supercontig177 から表示、塩基番号 4,200 から 10,200

説明
配列の名前、遺伝子名、ゲノム上の位置、その他のランドマークなどを使って検索します。ワイルドカード文字として * を使うことができます。ルーラー上でクリックした位置が中心になります。スクロールとズームボタンを使って拡大率と位置を変更します。
例: Supercontig0, Supercontig1, Supercontig2, Supercontig1716.
[この表示をブックマーク][独自アニメーションをアップロード][バナーを隠す][これらの項目を共有][この画像へのリンク][高品質SVG画像][ヘルプ][リセット]

検索
ランドマークまたは領域: Supercontig177:4200..10200 出力や解析などの操作: 出力 Decorated FASTA File 設定... 実行
データソース: YOKOZUNA-1 Scaffold (20090409) - MySQL dev スクロールズーム: <<< 表示 6.001 kbp >>> 反転

オーバービュー

Supercontig177
Contig map
rRNA
tRNA
領域ビュー

Supercontig177
Contig map
Read coverage
GC content
詳細ビュー

Kumamushi scaffolds
Kumamushi contigs
Contig177.1
Kumamushi contigs (gap)
Kumamushi reads
GKMno331_i05.g1 GKMno331_i05.b1 GKMno104_e04.g1 GKMno388_k24.
GKMno322_h24.g1 GKMno20_c15.g1 GKMno301_k13.g1 GKMno301_k13.b1
GKMno334_k16.g1 GKMno418_101.b1 GKMno418_101.g1
GKMno334_k16.b1 GKMno334_k16.b1 GKMno372_k17.g1
GKMno420_i12.b1 GKMno76_c20.g1 GKMno76_c20.b1 GKMno464_n1
GKMno272_n13.b1 GKMno272_n13.g1
GKMno245_n10.b1 GKMno245_n10.g1
GKMno438_p14.g1 GKMno67_g12.g1 GKMno80_j06.g1
GKMno132_i11.g1 GKMno80_j06.b1
GKMno132_i11.b1 GKMno80_j06.b1
GKMno467_i16.g1
GKMno182_a19.g1
GKMno182_a19.b1

GlimmerHMM dme
Supercontig177.path1.gene4
GlimmerHMM cel
Supercontig177.path1.gene4
rRNA
tRNA

Supercontig177.path1.gene4

GlimmerHMM cel
Supercontig177.path1.gene4
rRNA
tRNA
Repeat
KEGG D. melanogaster (dme)
KEGG C. elegans (cel)
SwissProt
Q35117|CVB_NYCHU
Q36572|CVB_NYCHU
Q36590|CVB_NYCLA
Q34462|CVB_EUNGL
Q33452|CVB_EUNPE
Q36559|CVB_FPOCE
Q36294|CVB_TYOLE
Q33537|CVB_IDIPH
Q35805|CVB_SCICA
Q36118|CVB_TOMRA
Q36291|CVB_TORRA
Q33607|CVB_LNSNO
Q36644|CVB_PLEDA
Q35994|CVB_TQDBR
Q21425|CVB_SORSA
Q33695|CVB_ARTFU
Q34254|CVB_CHISL
Q34260|CVB_CHITR
Q31684|CVB_ANOCU
Q35467|CVB_PLNHE
Q99468|CVB_RHIMI
Q21414|CVB_SORNO
Q33377|CVB_GANSI
Q35942|CVB_STU11
Q31652|CVB_NSCE

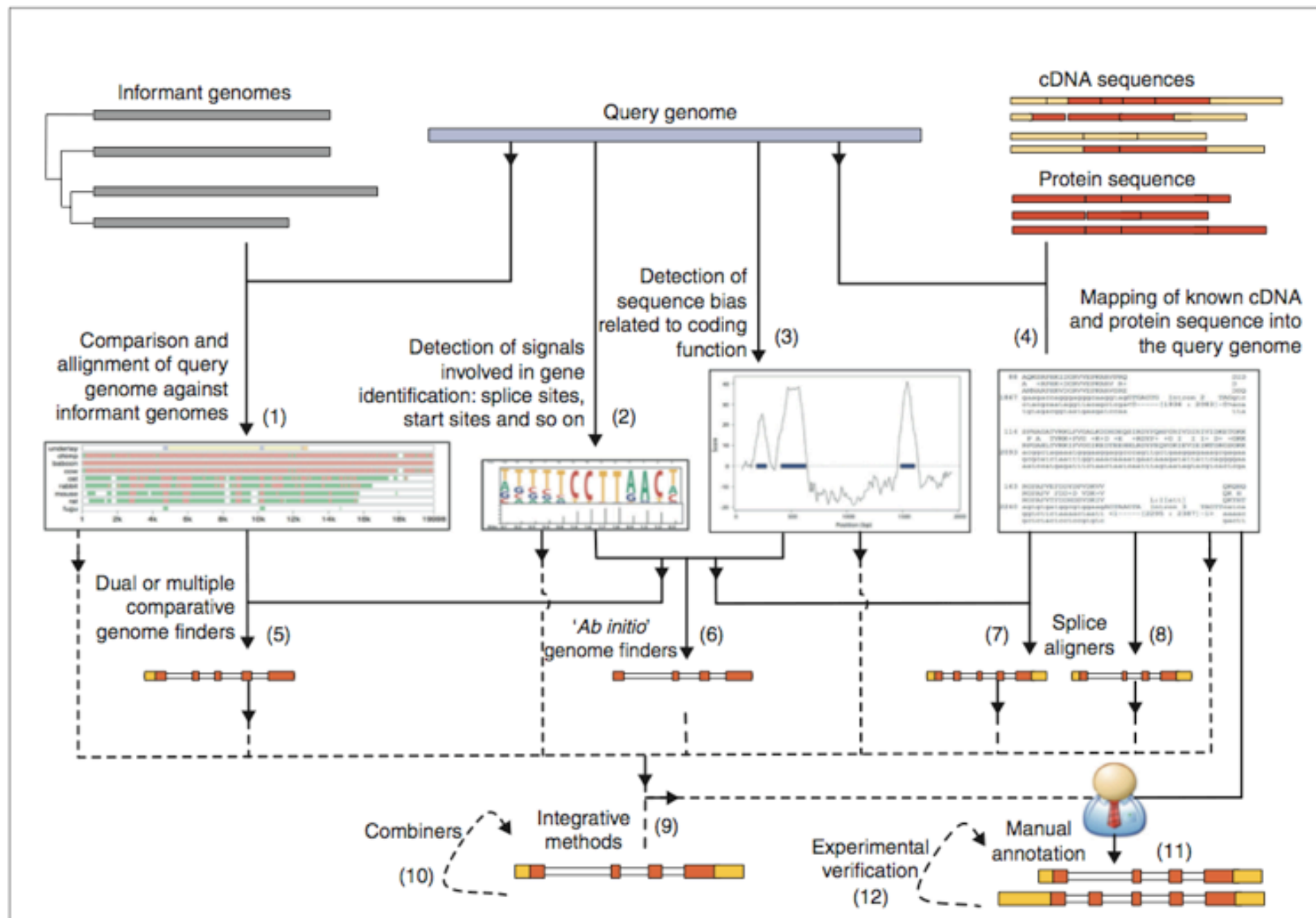
SwissProt ID30
UniRef50 ID30
UniRef50_Q15K44
UniRef50_n1BQ48
UniRef50_Q47901
UniRef50_Q59906
UniRef50_Q35271
UniRef50_Q76H05
UniRef50_P00156
UniRef50_P24964
UniRef50_Q20135
UniRef50_Q55225
UniRef50_P00158
UniRef50_Q94N89
UniRef50_Q6P777
UniRef50_Q94R91
UniRef50_Q8K370
UniRef50_Q50G55
UniRef50_B8VST0
UniRef50_Q3GKJ9
UniRef50_Q8W509

UniRef50_032905
UniRef50_B1P7R5
UniRef50_069H01
UniRef50_08W699
UniRef50_P34843
UniRef50_P18945
UniRef50_P48661
UniRef50_P00414
UniRef50_n1B75
UniRef50_B1W58

3-frame translation (forward)
DNA sequence / GC content
3-frame translation (reverse)

塩基表示を解除 画像を更新

ゲノムアノテーション: 遺伝子構造, 比較ゲノム → データベース化





tRNA : tRNAscan-SE 1.21 (Lowe TM and Eddy SR, NAR 1997)

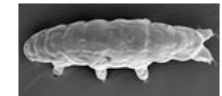
first	second				third
	T	C	A	G	
T	TTT F	TCT S 2	TAT Y	TGT C	T
	TTC F 3	TCC S	TAC Y i3	TGC C 2	C
	TTA L ii1	TCA S i1	TAA / 0	TGA / 0	A
	TTG L i2	TCG S ii1	TAG / 0	TGG W 2	G
C	CTT L 2	CCT P 2	CAT H	CGT R 3	T
	CTC L	CCC P	CAC H 2	CGC R	C
	CTA L 1	CCA P 3	CAA Q 2	CGA R 2	A
	CTG L ii1	CCG P 1	CAG Q 2	CGG R 1	G
A	ATT I 3	ACT T 3	AAT N	AGT S	T
	ATC I	ACC T	AAC N 3	AGC S 2	C
	ATA I i1	ACA T 1(2)	AAA K 2	AGA R 1	A
	ATG M 3	ACG T 1	AAG K 3	AGG R 1	G
G	GTT V 3	GCT A 2	GAT D	GGT G	T
	GTC V	GCC A	GAC D 1	GGC G 4	C
	GTA V 1	GCA A 2(1)	GAA E 3	GGA G 2	A
	GTG V 1	GCG A 1	GAG E 3	GGG G 1	G

Number of tRNA coding genes
 i = incomplete (fragmented/pseudo)
 () = tRNAs with introns

YOKOZUNA-1

91 tRNAs

8 incomplete genes
 3 tRNAs with introns



D. melanogaster (fly)

304 tRNAs

4 predicted pseudogenes
 15 tRNAs with introns



C. elegans (worm)

820 tRNAs

211 predicted pseudo genes
 31 tRNAs with introns



yellow minus charge red plus charge blue noncharge green nonpolar

<http://gtrnadb.ucsc.edu/>



rRNA : RNAmmer 1.2 (Lagesen K, et al. NAR 2007)

#	seqname	source	feature	start	end	score	+/-	frame	attribute
#	-----	-----	-----	-----	-----	-----	-----	-----	-----
Supercontig827	RNAmmer-1.2	rRNA	2	1444	309.2	+	.	28s_rRNA	
Supercontig36	RNAmmer-1.2	rRNA	281692	281806	84.4	+	.	8s_rRNA	
Supercontig2	RNAmmer-1.2	rRNA	405915	406029	80.1	-	.	8s_rRNA	
Supercontig4	RNAmmer-1.2	rRNA	1116120	1116234	83.2	+	.	8s_rRNA	
Supercontig7	RNAmmer-1.2	rRNA	226804	226918	82.9	-	.	8s_rRNA	
Supercontig2	RNAmmer-1.2	rRNA	851873	851987	82.3	+	.	8s_rRNA	
Supercontig2	RNAmmer-1.2	rRNA	94618	94735	59.1	-	.	8s_rRNA	
Supercontig7	RNAmmer-1.2	rRNA	338459	338573	84.4	+	.	8s_rRNA	
Supercontig2	RNAmmer-1.2	rRNA	906513	906627	82.3	+	.	8s_rRNA	
Supercontig36	RNAmmer-1.2	rRNA	504445	504559	84.4	-	.	8s_rRNA	
Supercontig2	RNAmmer-1.2	rRNA	844376	844490	79.3	+	.	8s_rRNA	
Supercontig4	RNAmmer-1.2	rRNA	1007680	1007794	84.4	+	.	8s_rRNA	
Supercontig6	RNAmmer-1.2	rRNA	83245	83362	78.2	+	.	8s_rRNA	
Supercontig110	RNAmmer-1.2	rRNA	103602	103716	82.0	-	.	8s_rRNA	
Supercontig36	RNAmmer-1.2	rRNA	147426	147540	46.3	+	.	8s_rRNA	
Supercontig2	RNAmmer-1.2	rRNA	731422	731536	84.4	-	.	8s_rRNA	
Supercontig5	RNAmmer-1.2	rRNA	1372783	1372897	84.4	-	.	8s_rRNA	
Supercontig36	RNAmmer-1.2	rRNA	202989	203103	81.7	-	.	8s_rRNA	
Supercontig4	RNAmmer-1.2	rRNA	1660162	1660276	75.3	+	.	8s_rRNA	
Supercontig181	RNAmmer-1.2	rRNA	12	126	84.4	-	.	8s_rRNA	
Supercontig8	RNAmmer-1.2	rRNA	146419	146533	85.5	-	.	8s_rRNA	
Supercontig2	RNAmmer-1.2	rRNA	1347335	1347455	27.1	+	.	8s_rRNA	
Supercontig101	RNAmmer-1.2	rRNA	130174	130288	84.4	+	.	8s_rRNA	
Supercontig779	RNAmmer-1.2	rRNA	46	1256	347.6	+	.	18s_rRNA	
Supercontig852	RNAmmer-1.2	rRNA	1	1085	159.9	-	.	18s_rRNA	
Supercontig1574	RNAmmer-1.2	rRNA	1	1141	30.3	+	.	18s_rRNA	
Supercontig779	RNAmmer-1.2	rRNA	46	1256	347.6	+	.	18s_rRNA	
Supercontig1266	RNAmmer-1.2	rRNA	1	1237	186.2	+	.	18s_rRNA	

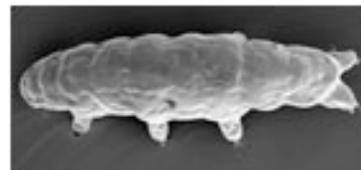


CDS : Exonerate 2.2.0 (Slater GTC *et al.*, BMC Bioinformatics 2005)



D. melanogaster
14023 genes

1760 genes
(12.6%)



1427 genes
(7.0%)



C. elegans
20185 genes

*** too few *
homologous genes !!!**



SwissProt

Percent identity > 50

20293 entries hit

1052 unique locus

4.47 introns / gene (± 3.75)

Percent identity > 30

82489 entries hit

2222 unique locus

UniRef50

Percent identity > 50

6451 entries hit

2237 unique locus

Percent identity > 30

54990 entries hit

4865 unique locus



ab initio gene finding

GlimmerHMM 3.0.1

(Majoros WH et al., Bioinformatics 2004)

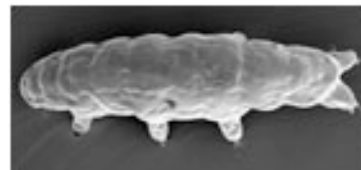
SNAP

(Korf I, BMC Bioinformatics 2004)



28666 predicted genes

Machine learning from
D. melanogaster
gene models



32545 predicted genes



Machine learning
from *C. elegans*
gene models

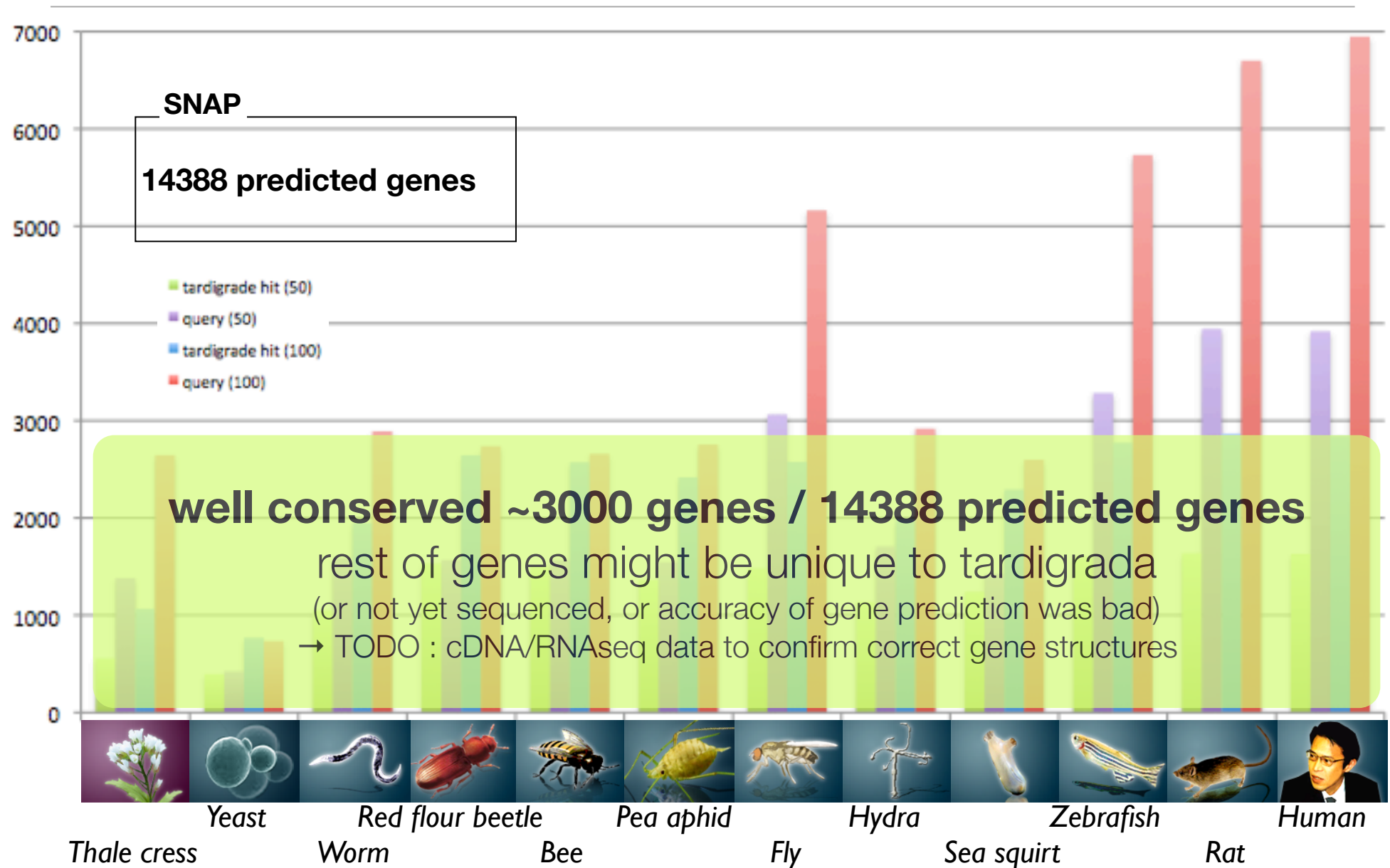
SNAP

14388 predicted genes

Accuracy looks better
(data not shown)



Mutual seq similarity with model organisms





Functional prediction

SNAP

14388 predicted genes

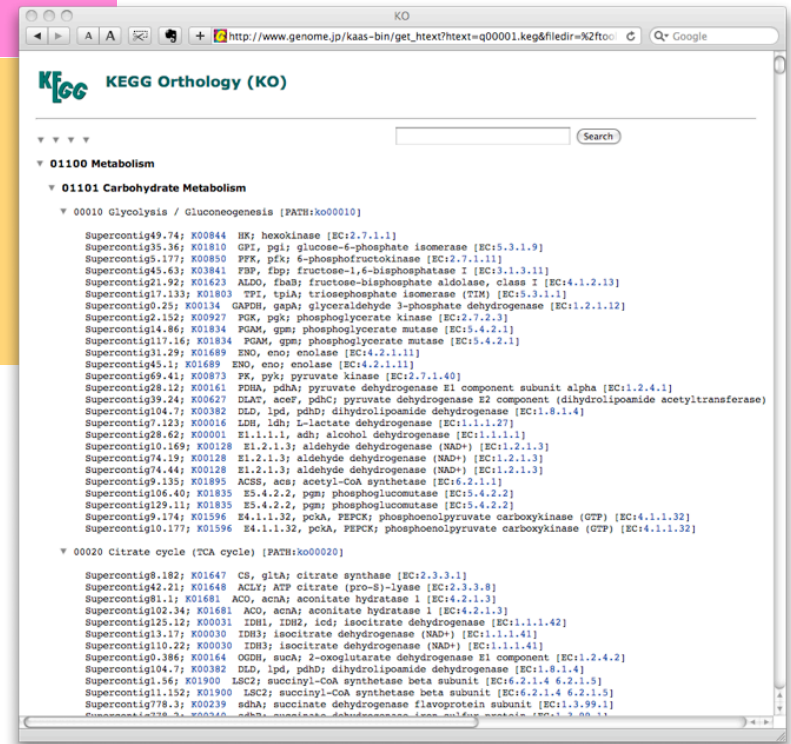


Hit with *D. melanogaster* genes (44%)

On KEGG metabolic pathway (40%)

Other genes (60%)

No hit (66%)



6362 / 14388 (44%)	<i>D. melanogaster</i>	(1e-5)
5835 / 14388 (41%)	<i>C. elegans</i>	(1e-5)
5555 / 14388 (39%)	<i>D. melanogaster</i>	(1e-10)
4884 / 14388 (34%)	<i>C. elegans</i>	(1e-10)



Functional prediction

KEGG Orthology (KO)

01100 Metabolism

01101 Carbohydrate Metabolism

00010 Glycolysis / Gluconeogenesis [PATH:ko00010]

- Supercontig49.74; K00844 HK; hexokinase [EC:2.7.1.1]
- Supercontig35.36; K01810 GPI, pgi; glucose-6-phosphate isomerase [EC:5.3.1.9]
- Supercontig5.177; K00850 PFK, pfk; 6-phosphofructokinase [EC:2.7.1.11]
- Supercontig45.63; K03841 FBP, fbp; fructose-1,6-bisphosphatase I [EC:3.1.3.11]
- Supercontig21.92; K01623 ALDO, fbaB; fructose-bisphosphate aldolase, class I [EC:4.1.2.13]
- Supercontig17.133; K01803 TPI, tpiA; triosephosphate isomerase (TIM) [EC:5.3.1.1]
- Supercontig0.25; K00134 GAPDH, gapA; glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]
- Supercontig2.152; K00927 PGK, pgk; phosphoglycerate kinase [EC:2.7.2.3]
- Supercontig4.86; K01834 PGAM, gpm; phosphoglycerate mutase [EC:5.4.2.1]
- Supercontig117.16; K01834 PGAM, gpm; phosphoglycerate mutase [EC:5.4.2.1]
- Supercontig31.29; K01689 ENO, eno; enolase [EC:4.2.1.11]
- Supercontig45.1; K01689 ENO, eno; enolase [EC:4.2.1.11]
- Supercontig69.41; K00873 PK, pyk; pyruvate kinase [EC:2.7.1.40]
- Supercontig28.12; K00161 PDHA, pdhA; pyruvate dehydrogenase E1 component subunit alpha [EC:1.2.4.1]
- Supercontig39.24; K00627 DLAT, aceF, pdhC; pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase)
- Supercontig104.7; K00382 DLD, lpd, pdhD; dihydrolipoamide dehydrogenase [EC:1.8.1.4]
- Supercontig7.123; K00016 LDH, ldh; L-lactate dehydrogenase [EC:1.1.1.27]
- Supercontig28.62; K00001 E1.1.1.1, adh; alcohol dehydrogenase [EC:1.1.1.1]
- Supercontig10.169; K00128 E1.2.1.3; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]
- Supercontig74.19; K00128 E1.2.1.3; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]
- Supercontig74.44; K00128 E1.2.1.3; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]
- Supercontig9.135; K01895 ACSS, acs; acetyl-CoA synthetase [EC:6.2.1.1]
- Supercontig106.40; K01835 E5.4.2.2, pgm; phosphoglucomutase [EC:5.4.2.2]
- Supercontig129.11; K01835 E5.4.2.2, pgm; phosphoglucomutase [EC:5.4.2.2]
- Supercontig9.174; K01596 E4.1.1.32, pcKA, PEPCK; phosphoenolpyruvate carboxykinase (GTP) [EC:4.1.1.32]
- Supercontig10.177; K01596 E4.1.1.32, pcKA, PEPCK; phosphoenolpyruvate carboxykinase (GTP) [EC:4.1.1.32]

00020 Citrate cycle (TCA cycle) [PATH:ko00020]

- Supercontig8.182; K01647 CS, gltA; citrate synthase [EC:2.3.3.1]
- Supercontig42.21; K01648 ACLY; ATP citrate (pro-S)-lyase [EC:2.3.3.8]
- Supercontig81.1; K01681 ACO, acnA; aconitate hydratase 1 [EC:4.2.1.3]
- Supercontig102.34; K01681 ACO, acnA; aconitate hydratase 1 [EC:4.2.1.3]
- Supercontig125.12; K00031 IDH1, IDH2, lcd; isocitrate dehydrogenase [EC:1.1.1.42]
- Supercontig13.17; K00030 IDH3; isocitrate dehydrogenase (NAD+) [EC:1.1.1.41]
- Supercontig110.22; K00030 IDH3; isocitrate dehydrogenase (NAD+) [EC:1.1.1.41]
- Supercontig0.386; K00164 OGDH, sucA; 2-oxoglutarate dehydrogenase E1 component [EC:1.2.4.2]
- Supercontig104.7; K00382 DLD, lpd, pdhD; dihydrolipoamide dehydrogenase [EC:1.8.1.4]
- Supercontig1.56; K01900 LSC2; succinyl-CoA synthetase beta subunit [EC:6.2.1.4 6.2.1.5]
- Supercontig11.152; K01900 LSC2; succinyl-CoA synthetase beta subunit [EC:6.2.1.4 6.2.1.5]
- Supercontig778.3; K00239 sdhA; succinate dehydrogenase flavoprotein subunit [EC:1.3.99.1]
- Supercontig778.3; K00239 sdhB; succinate dehydrogenase iron sulfur protein [EC:1.3.99.1]

Carbohydrate metabolism

259 genes

Energy metabolism

113 genes

Lipid metabolism

165 genes

Nucleotide metabolism

152 genes

Amino acid metabolism

211 genes

Other amino acid metabolism

70 genes

Glycan Biosynthesis and metabolism

95 genes

Biosynthesis of polyketides/NRP

3 genes

Cofactor and vitamin metabolism

79 genes

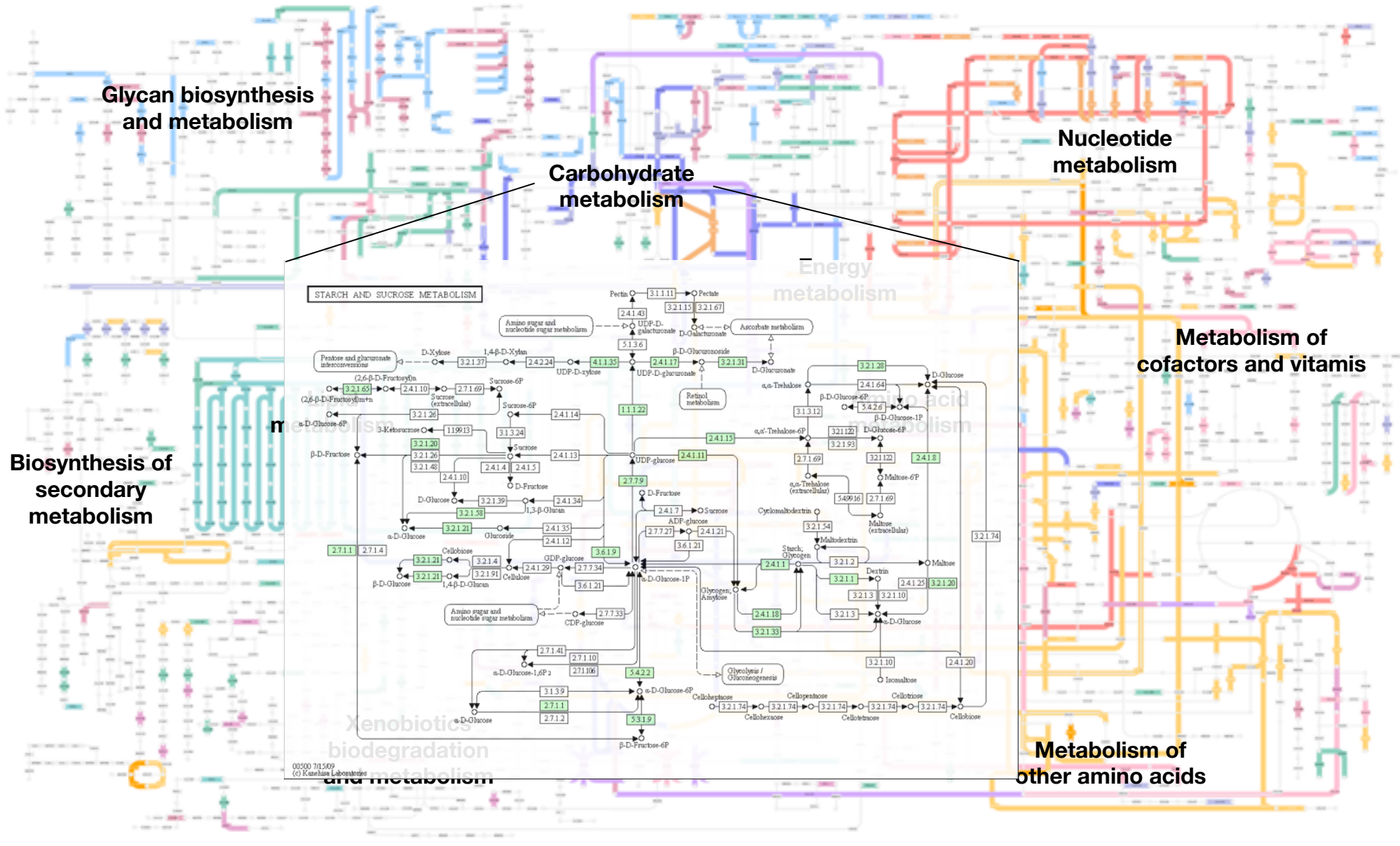
Biosynthesis of secondary metabolites

43 genes

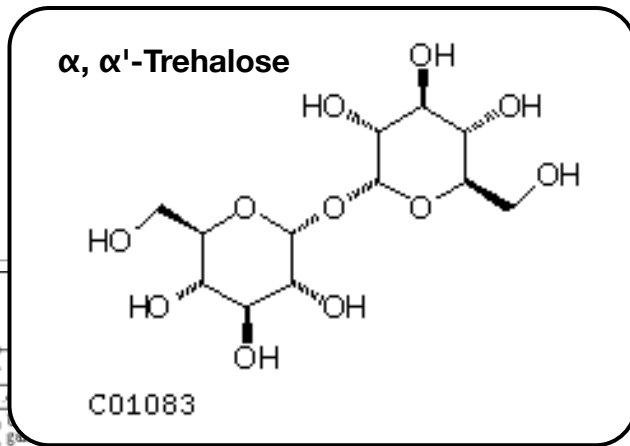
:



Pathway reconstruction



Trehalose metabolism

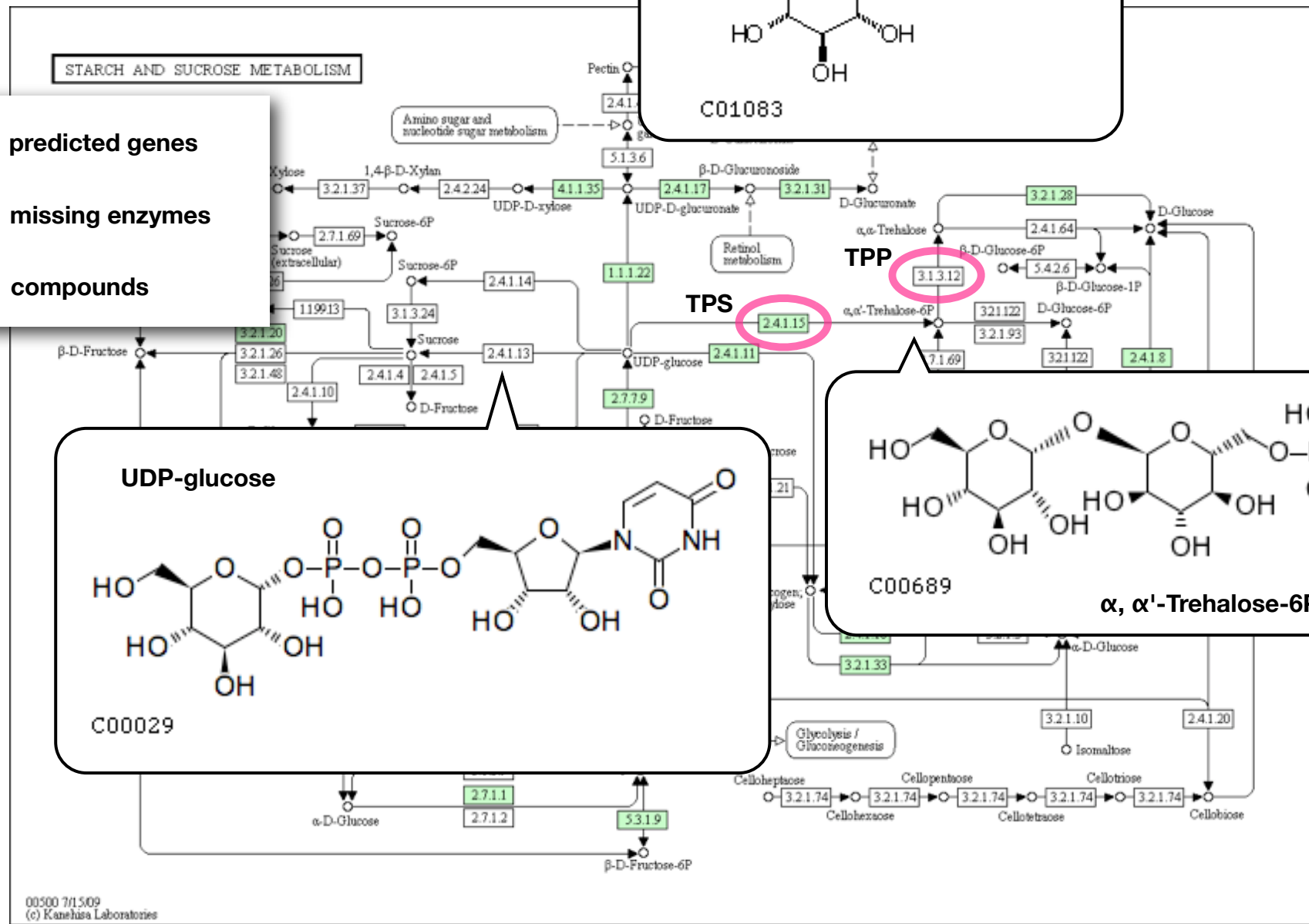
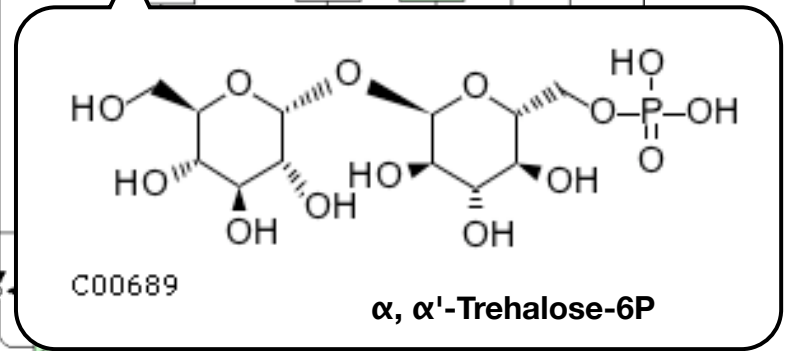
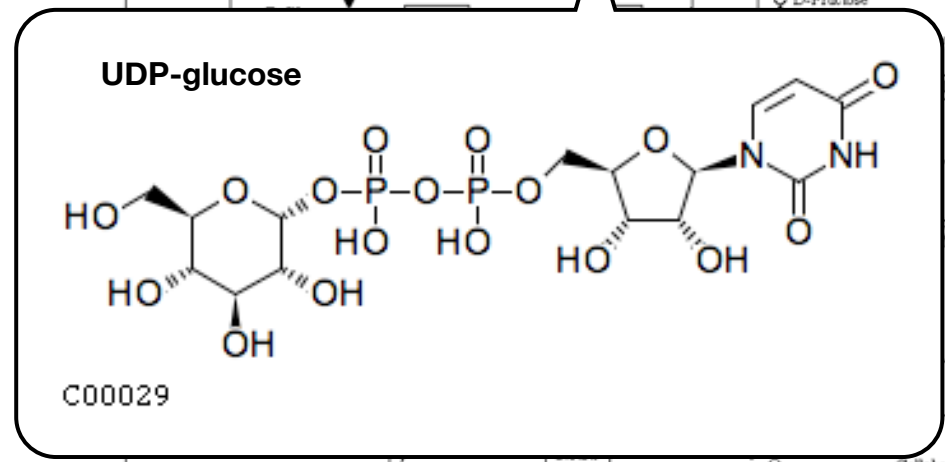


STARCH AND SUCROSE METABOLISM

2.4.1.15 predicted genes

3.1.3.12 missing enzymes

compounds



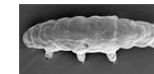
TPS (trehalose-6-phosphate syntase) / TPP (trehalose phosphatase)



Query= Supercontig4.256
(1153 letters)

Sequences producing significant alignments:

		Score (bits)	E Value
dme:Dmel_CG4104	Tps1; Trehalose-6-phosphate synthase 1 (EC:2.4...	369	e-102
dme:Dmel_CG5177	CG5177 gene product from transcript CG5177-RA (...)	56	1e-07
dme:Dmel_CG5171	CG5171 gene product from transcript CG5171-RC (...)	53	7e-07
dme:Dmel_CG13280	CG13280 gene product from transcript CG13280-R...	52	1e-06
dme:Dmel_CG3609	CG3609 gene product from transcript CG3609-RA	52	2e-06
dme:Dmel_CG3597	CG3597 gene product from transcript CG3597-RA	47	5e-05



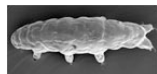
SNAP predicted gene



D. melanogaster proteins

← TPS
← TPP

>dme:Dmel_CG4104 Tps1; Trehalose-6-phosphate synthase 1
(EC:2.4.1.15); K00697 alpha,alpha-trehalose-phosphate



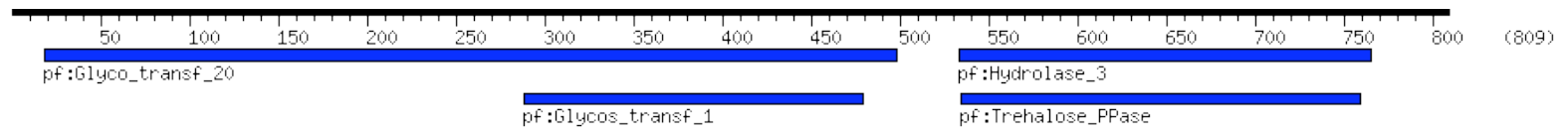
TPS1 : 1153aa (Supercontig4.256)



dme:Dmel_CG4104

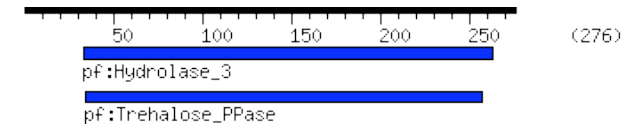


TPS1 : 809aa



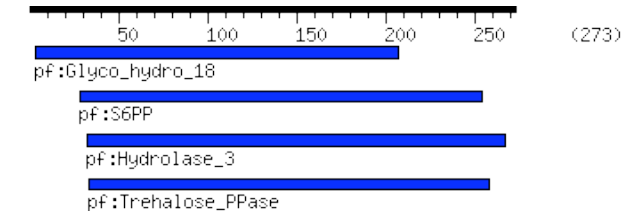
TPP : 276aa

dme:Dmel_CG5177



TPP : 273aa

dme:Dmel_CG5171



現在進行中

全データ (約2GB) の取得まで、
遺伝研 (三島) → 東大 (白金台)
転送に約1時間、
(クマムシゲノムちっこのに...)

- 昨日 (2009/8/20) 更新版の配列データが届いた!
 - 'assembly.raw.tar.gz' at 12831008 (0%) 1.89M/s eta:15m [Receiving data]

== Acknowledgements ==

- Material preparation
 - D. Horikawa (NASA)
 - T. Kunieda (U. Tokyo)
 - H. Kuwahara (U. Tokyo)
- Genome sequencing
 - A. Toyoda (NIG)
 - T. Shin-I (NIG)
 - K. Ohishi (NIG)
 - A. Motoyama (NIG)
 - T. Aizu (NIG)
 - Y. Kohara (NIG)
 - A. Fujiyama (NIG)
- Genome analysis
 - K. Arakawa (Keio U.)
 - Y. Hasebe (Keio U.)
 - N. Kido (Keio U.)
 - M. Kanehisa (Kyoto U.)



This work was supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas "Comparative Genomics" from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

ヒトゲノム解析センター



- 東大医科研内に設置、国内のゲノム研究者用にスパコンを運用



分散メモリ型サーバ

Sun Fire X4400 (16コア, 128GB RAM) x 12
Sun Blade X6250 (8コア, 32GB RAM) x 764

共有メモリ型サーバ

SGI Altix 4700 (128コア, 2TB RAM) x 1

ディスクアレイ

総容量 2PB
Lustr ファイルシステム

ジョブ管理システム

Sun Grid Engine
→ 上記の約 6000 コアで並列計算



TOP500

Rank	Site	Computer/Year Vendor	Cores	R _{max}	R _{peak}	Power
1	DOE/NNSA/LANL United States	Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell B1 3.2 Ghz / Opteron DC 1.8 Ghz, Voltaire Infiniband / 2008 IBM	129600	1105.00	1456.70	2483.47
2	Oak Ridge National Laboratory United States	Jaguar - Cray XTS QC 2.3 Ghz / 2008 Cray Inc.	150152	1059.00	1381.40	6950.60
3	Forschungszentrum Juelich (FZJ) Germany	JUGENE - Blue Gene/P Solution / 2009 IBM	294912	825.50	1002.70	2268.00
4	NASA/Ames Research Center/NAS United States	Pleiades - SGI Altix ICE 8200EX, Xeon QC 3.0/2.66 Ghz / 2008 SGI	51200	487.01	608.83	2090.00
5	DOE/NNSA/LLNL United States	BlueGene/L - eServer Blue Gene Solution / 2007 IBM	212992	478.20	596.38	2329.60
6	National Institute for Computational Sciences/University of Tennessee United States	Kraiken XTS - Cray XTS QC 2.3 Ghz / 2008 Cray Inc.	66000	463.30	607.20	
7	Argonne National Laboratory United States	Blue Gene/P Solution / 2007 IBM	163840	458.61	557.06	1260.00
8	Texas Advanced Computing Center/Univ. of Texas United States	Ranger - SunBlade x6420, Opteron QC 2.3 Ghz, Infiniband / 2008 Sun Microsystems	62976	433.20	579.38	2000.00

=== 2009年6月版 ===

1 位 Roadrunner (1456TFLOPS)

:

22 位 地球シミュレータ (131TFLOPS)

28 位 JAXA (121TFLOPS)

40 位 RIKEN (96TFLOPS)

41 位 東工大TSUBAME (163TFLOPS)

42 位 東大 IT センター T2K (113TFLOPS)

47 位 筑波大 T2K (95TFLOPS)

65 位 National Institute for Fusion Science (77TFLOPS)

69 位 東大ヒトゲノム解析センター (69TFLOPS)

78 位 京大 T2K (61TFLOPS)

93 位 National Institute for Materials Science (45TFLOPS)

順位は R max 値による (カッコ内は R peak 値)

GENOME MAP

ヒトゲノムマップ

ゲノムとはー

gen + om → genome
ゲノム (DNAの総称)

ゲノムとは、生物の体細胞に1-2組存在する遺伝子の総称。動物や植物の体細胞には、2組のゲノムが存在する。細菌やウイルスは、1組のゲノムが存在する。ゲノムには、生物の生命活動に必要な遺伝子だけでなく、生物の進化や発生に関与する遺伝子も含まれている。ゲノムマップは、ゲノム上の遺伝子の位置を特定し、その機能を明らかにするための重要なツールである。

AMY1A
唾液アミラーゼ遺伝子
唾液アミラーゼは、唾液に含まれる消化酵素で、主に澱粉を分解する役割を果たす。AMY1A遺伝子のコピー数が多い人は、消化能力が高いとされている。

TYN
テニソウ遺伝子
テニソウは、テニソウ科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

ACTA1
アクチン遺伝子
アクチンは、細胞の骨格を形成するタンパク質で、細胞の運動や分裂に重要な役割を果たす。

SET
セト遺伝子
セトは、セト科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

一冊に1冊
ヒトゲノムマップ
詳しくはWEBへ
<http://www.mext.go.jp/>
(科学技術省)

RHO
ロー遺伝子
ローは、細胞の骨格を形成するタンパク質で、細胞の運動や分裂に重要な役割を果たす。

DAR5
ダール5遺伝子
ダール5は、ダール5科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

MUC1
ムク1遺伝子
ムク1は、粘液多糖を産生するタンパク質で、細胞の表面を覆っており、細胞の保護や免疫応答に関与している。

PR1
プライ遺伝子
プライは、植物の防御反応に関与するタンパク質で、病原体の侵入に対する防御反応を誘起する。

PR1R
プライR遺伝子
プライRは、プライの阻害剤として機能するタンパク質で、植物の防御反応を抑制する。

CMAR
シマール遺伝子
シマールは、シマール科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

SHOX
ショウ遺伝子
ショウは、骨の成長に関与するタンパク質で、骨の成長障害の原因となる。

POA
ポア遺伝子
ポアは、ポア科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

MAPK1
マップキ1遺伝子
マップキ1は、細胞の増殖や分化に関与するタンパク質で、がんの発症に関与している。

MB
エムビー遺伝子
エムビーは、エムビー科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

SOD1
ソッド1遺伝子
ソッド1は、活性酸素を分解するタンパク質で、神経変性疾患の原因となる。

OPN1LW
オープン1LW遺伝子
オープン1LWは、視覚に関与するタンパク質で、視覚障害の原因となる。

PR2
プライ2遺伝子
プライ2は、植物の防御反応に関与するタンパク質で、病原体の侵入に対する防御反応を誘起する。

PR2R
プライ2R遺伝子
プライ2Rは、プライ2の阻害剤として機能するタンパク質で、植物の防御反応を抑制する。

CMAR
シマール遺伝子
シマールは、シマール科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

FOXP2
フォクプ2遺伝子
フォクプ2は、言語能力に関与するタンパク質で、言語障害の原因となる。

GULOP
グロップ遺伝子
グロップは、グロップ科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

ABD
アブド遺伝子
アブドは、アブド科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

LIPP
リップ遺伝子
リップは、リップ科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

RRB
アールビー遺伝子
アールビーは、アールビー科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

ALDH2
アルド2遺伝子
アルド2は、アルコール代謝に関与するタンパク質で、アルコール依存症の原因となる。

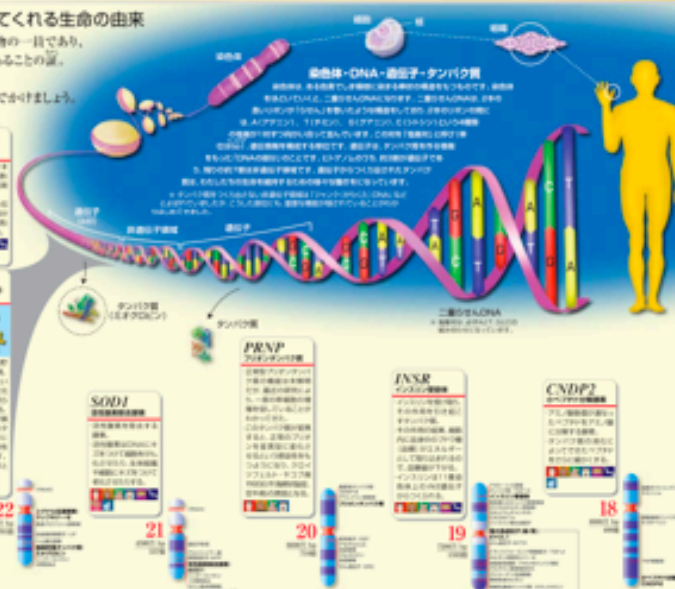
COL2A1
コラーゲン2A1遺伝子
コラーゲン2A1は、骨の形成に関与するタンパク質で、骨軟化症の原因となる。

ACTA1
アクチン遺伝子
アクチンは、細胞の骨格を形成するタンパク質で、細胞の運動や分裂に重要な役割を果たす。

SET
セト遺伝子
セトは、セト科の植物で、その根から抽出された成分は、抗がん剤として研究されている。

ここまでわかった!! ヒトゲノム

ゲノムが教えてくれる生命の由来
ゲノムとは、地球生物の1冊であり、あなたがあなたであることの証。さあ、このマップで、ヒトゲノムの解読にかけましょう。



DNAにかかれた生命の暗号集

あなたを形づくる遺伝の暗号のそれぞれが、30億文字からなるヒトゲノム(暗号集)をしています。

EGFR
エグフ遺伝子
エグフは、細胞の増殖に関与するタンパク質で、がんの発症に関与している。

EYCL&J
エック&ジェイ遺伝子
エック&ジェイは、眼の色に関与するタンパク質で、眼の色を決定する。

CDH1
シディー1遺伝子
シディー1は、細胞接着に関与するタンパク質で、がんの発症に関与している。

PER1
ペー1遺伝子
ペー1は、時計遺伝子で、生物の体内時計を調節する。

CNDP2
シディー2遺伝子
シディー2は、細胞の増殖に関与するタンパク質で、がんの発症に関与している。

INSR
インサー遺伝子
インサーは、インスリン受容体で、血糖値を調節する。

このマップの使い方

このマップは、ヒトゲノム上の遺伝子の位置を特定し、その機能を明らかにするための重要なツールである。このマップを使って、遺伝子の位置や機能を調べ、その機能を明らかにすることができる。

おいしいちゃんも、おばあちゃんも、わたしの中に

賢賢から賢賢へ伝わるゲノム

ゲノムは、生物の体細胞に1-2組存在する遺伝子の総称。動物や植物の体細胞には、2組のゲノムが存在する。細菌やウイルスは、1組のゲノムが存在する。ゲノムには、生物の生命活動に必要な遺伝子だけでなく、生物の進化や発生に関与する遺伝子も含まれている。ゲノムマップは、ゲノム上の遺伝子の位置を特定し、その機能を明らかにするための重要なツールである。

ナンバーワンよりオンリーワン

一冊に1冊だけのゲノム

ヒトゲノムは、約30億文字のDNAから構成されている。このゲノムには、約2万5千個の遺伝子が含まれている。これらの遺伝子は、生物の生命活動に必要な遺伝子だけでなく、生物の進化や発生に関与する遺伝子も含まれている。ゲノムマップは、ゲノム上の遺伝子の位置を特定し、その機能を明らかにするための重要なツールである。

ヒトにはヒトゲノム、フグにはフグゲノム

生物の進化を究む、多様なゲノム

ゲノムは、生物の体細胞に1-2組存在する遺伝子の総称。動物や植物の体細胞には、2組のゲノムが存在する。細菌やウイルスは、1組のゲノムが存在する。ゲノムには、生物の生命活動に必要な遺伝子だけでなく、生物の進化や発生に関与する遺伝子も含まれている。ゲノムマップは、ゲノム上の遺伝子の位置を特定し、その機能を明らかにするための重要なツールである。

ゲノムでできること

ゲノムが未来を導く一歩

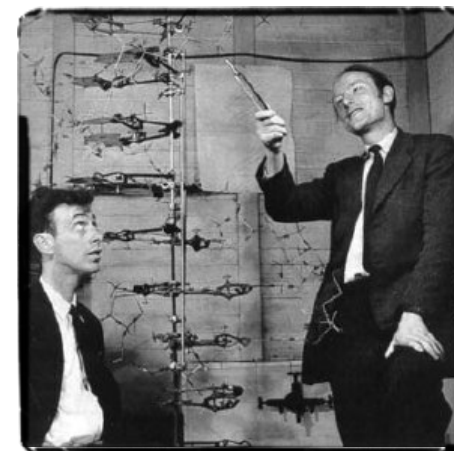
ゲノムマップは、ゲノム上の遺伝子の位置を特定し、その機能を明らかにするための重要なツールである。このマップを使って、遺伝子の位置や機能を調べ、その機能を明らかにすることができる。ゲノムマップは、生物の進化や発生に関与する遺伝子を特定し、その機能を明らかにするための重要なツールである。

ヒトゲノムマップ 科学技術省 2006/4/14

ゲノムデータベースの歴史



- 1990年 国際コンソーシアムによるヒトゲノムプロジェクト開始 (HUGO)
 - Johns Hopkins 大学に Genome Database (GDB) 開設
 - モデル生物のゲノムデータベースの整備も平行して進展
 - マウス Mouse Genome Database (MGD)
 - ショウジョウバエ FlyBase
 - 酵母 Saccharomyces Genome Database (SGD)
- 1990年代半ば INSD に自動シーケンシングのデータ公開
 - Expressed Sequence Tag (EST) - cDNA 断片の大量登録
 - Genome Survey Sequence (GSS) - シーケンサーの出力データ蓄積
 - High Throughput Genome Sequence (HTG) - アセンブリ配列の蓄積
- 1998年 セレラ社 Craig Venter の参入
 - HUGO が GDB を打ち切り UCSC の Golden path が後を継ぐ
 - NCBI が HUGO のデータと公共データをマージし Build を公開



ヒトゲノムの解読へ

- 1990年 HUGO 開始
- 1998/5 セレラ社設立
- 1999/3 セレラ社ヒトゲノム解読開始
- 1999/12 HUGO (日,米,英) 22番染色体解読完了
- 2000/5 HUGO (日,独) 21番染色体解読完了
- 2000/6/26 ドラフト解読完了の宣言
- 2002年 HapMap プロジェクト開始
- 2003/4/14 ヒトゲノム解読完了
- 2007年 ワトソン博士のゲノム解読 (2ヶ月で完了)
- 2008年 1000人ゲノムプロジェクト



ヒトゲノムの構成：Ensemblの場合 - 2009

- Ensembl v.53 (Mar 2009)

Statistics	
Known genes	21,343
Novel genes	73
Pseudogenes	9,899
RNA genes	5,732
Genscan predictions	49,796
Gene exons	297,252
Gene transcripts	62,877
Base pairs	3,253,037,807
Golden path length	3,093,120,360
SNPs	15,040,632

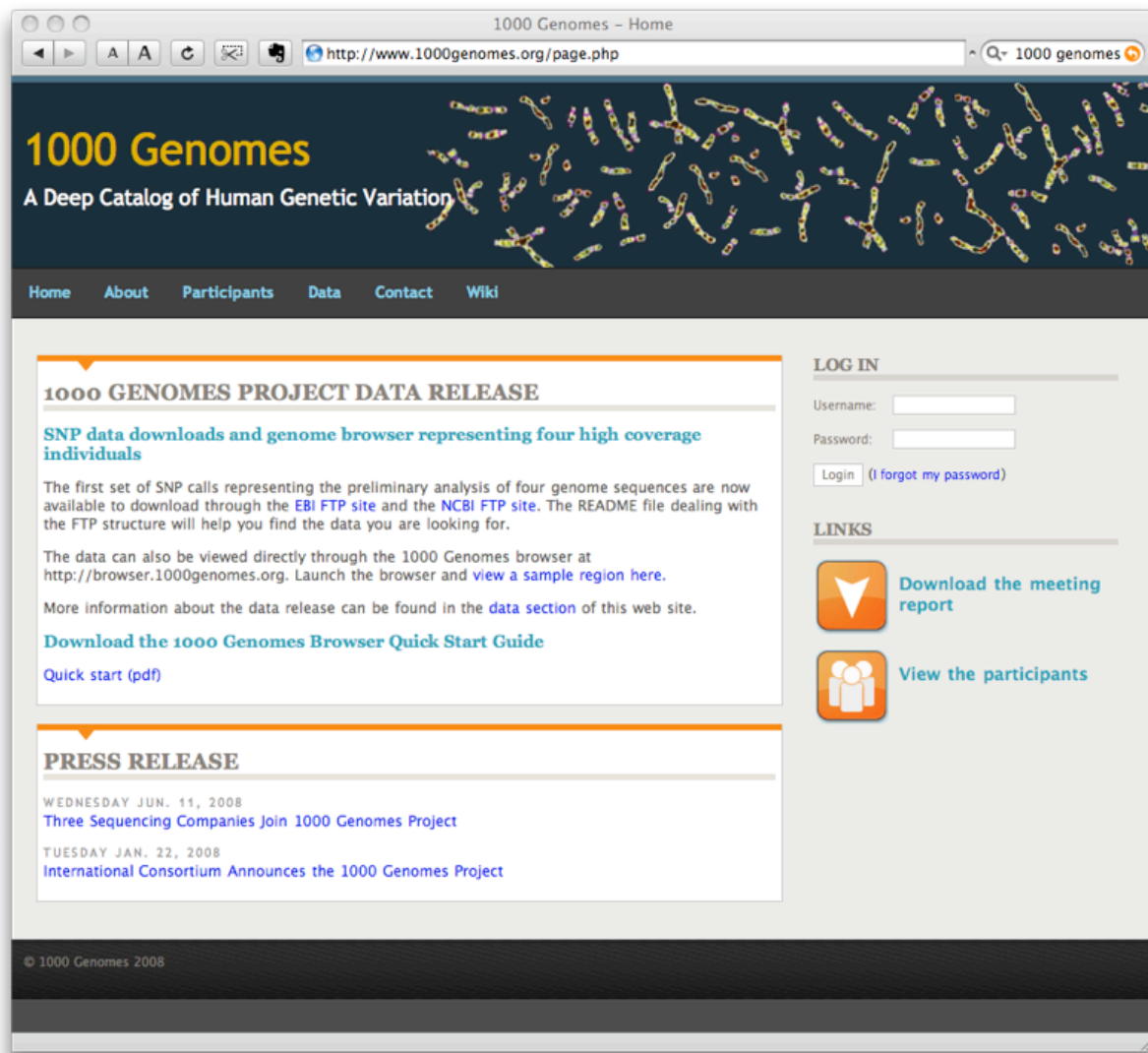
The screenshot displays the Ensembl Human website interface. At the top, there's a search bar and navigation links. The main content area is divided into several sections:

- Use Ensembl to...:** Includes links for BLAST search, BioMart, and data upload/download.
- Docs and downloads:** Links to information, news, and software.
- Select a species:** Options for Mammals, Other chordates, and Other eukaryotes.
- Other links:** Home, Sitemap, Vega, and Pre Ensembl.
- Explore the Homo sapiens genome:** Features a karyotype visualization and a search tool for jumping to sequence positions.
- About the Human genome:** Provides information on the assembly and annotation.
- Statistics:** A table summarizing key genomic metrics.
- What's New in Ensembl 38:** A news section detailing updates to the database.

The **Statistics** section in the screenshot matches the table provided in the previous block:

Category	Count
Known genes	21,343
Novel genes	73
Pseudogenes	9,899
RNA genes	5,732
Genscan predictions	49,796
Gene exons	297,252
Gene transcripts	62,877
Base Pairs	3,253,037,807
Golden Path Length	3,093,120,360
SNPs	15,040,632

1000人ゲノムプロジェクト



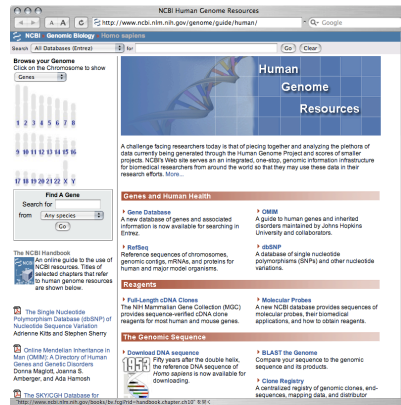
次世代シーケンサーの登場を受けて 2008 年より開始

ヒトの遺伝的多型を1000人のゲノムを決めることで詳細に調べ、医学的に重要な変異を探索するプロジェクト

<http://1000genomes.org/>

ゲノムデータベースの多様化と大容量化

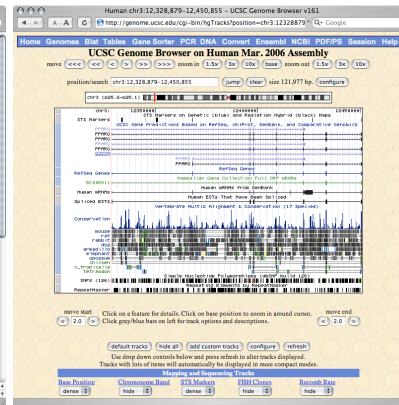
ヒトゲノム



NCBI Map Viewer



Ensembl



UCSC Golden Path

微生物ゲノム



モデル生物

TAIR	<i>Arabidopsis thaliana</i> (thale cress)
DictyBase	<i>Dictyostelium discoideum</i>
FlyBase	<i>Drosophila</i> (fly)
MGI	<i>Mus musculus</i> (mouse)
RGD	<i>Rattus norvegicus</i> (rat)
PomBase	<i>Schizosaccharomyces pombe</i> (f. yeast)
SGD	<i>Saccharomyces cerevisiae</i> (b. yeast)
WormBase	<i>Caenorhabditis elegans</i> (nematode)
ZFIN	<i>Danio rerio</i> (zebrafish)



サードパーティ アノテーション

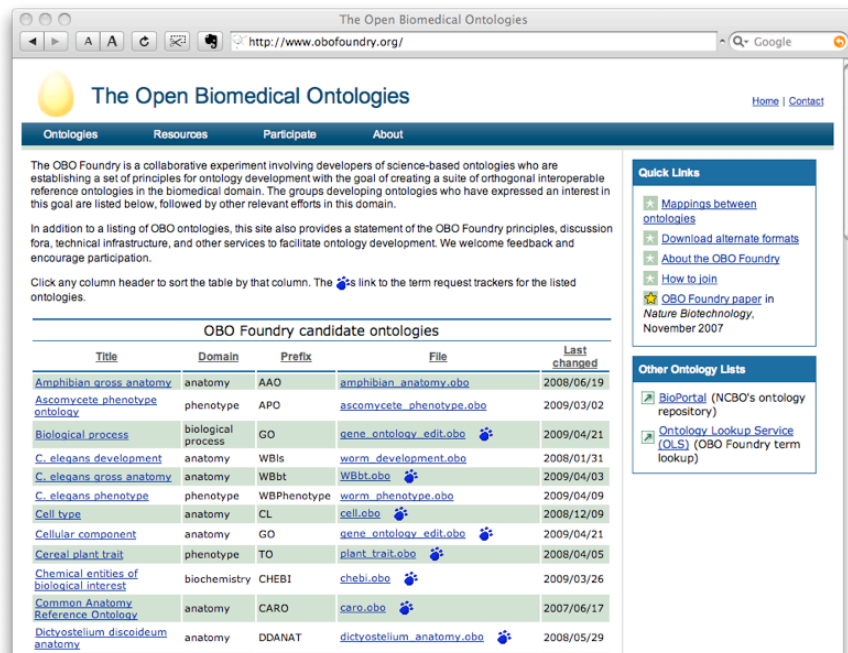
- RefSeq
- Vega
- GenomeReviews
- H-inv
- HapMap
- JSNP
- DBTSS
- KEGG
- GeneCards
- ...

ゲノムデータベースに使われるシステム

データ共有のための標準化 大規模なデータのハンドリング

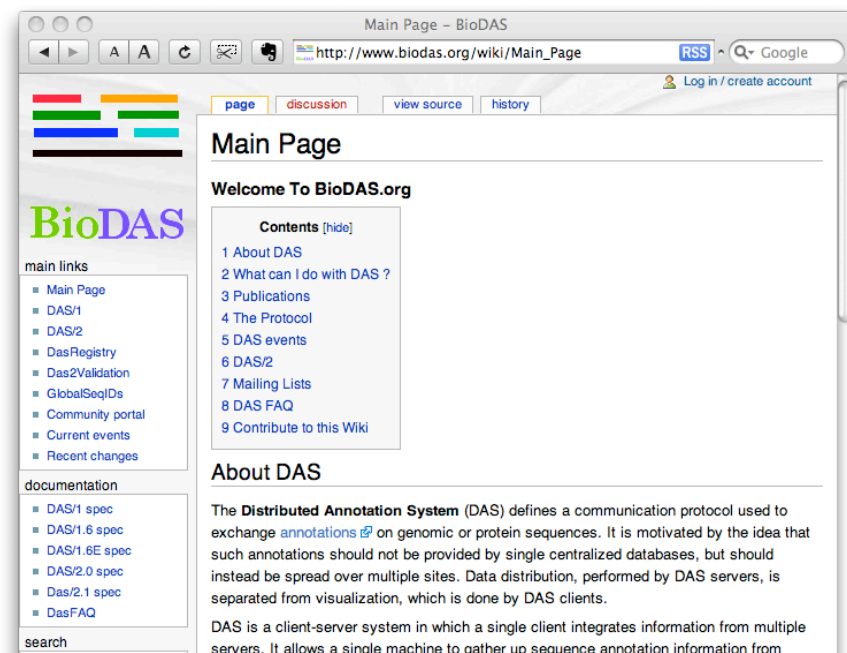
用語の統一：Gene Ontology や
Sequence Ontology など

ゲノムデータの相互運用：
BioDAS や GMOD など



The screenshot shows the website 'The Open Biomedical Ontologies' with a table titled 'OBO Foundry candidate ontologies'. The table has columns for Title, Domain, Prefix, File, and Last changed. It lists various ontologies such as Amphibian gross anatomy, Ascomycete phenotype ontology, Biological process, C. elegans development, C. elegans gross anatomy, C. elegans phenotype, Cell type, Cellular component, Cerebral plant trait, Chemical entities of biological interest, Common Anatomy Reference Ontology, and Dictyostelium discoideum anatomy.

Title	Domain	Prefix	File	Last changed
Amphibian gross anatomy	anatomy	AAO	amphibian_anatomy.obo	2008/06/19
Ascomycete phenotype ontology	phenotype	AP0	ascomycete_phenotype.obo	2009/03/02
Biological process	biological process	GO	gene_ontology_edit.obo	2009/04/21
C. elegans development	anatomy	WBls	worm_development.obo	2008/01/31
C. elegans gross anatomy	anatomy	WBbt	WBbt.obo	2009/04/03
C. elegans phenotype	phenotype	WBPhenotype	worm_phenotype.obo	2009/04/09
Cell type	anatomy	CL	cell.obo	2008/12/09
Cellular component	anatomy	GO	gene_ontology_edit.obo	2009/04/21
Cerebral plant trait	phenotype	TO	plant_trait.obo	2008/04/05
Chemical entities of biological interest	biochemistry	CHEBI	chebi.obo	2009/03/26
Common Anatomy Reference Ontology	anatomy	CARO	caro.obo	2007/06/17
Dictyostelium discoideum anatomy	anatomy	DDANAT	dictyostelium_anatomy.obo	2008/05/29



The screenshot shows the 'Main Page - BioDAS' website. It features a navigation bar with 'page', 'discussion', 'view source', and 'history' tabs. The main content area includes a 'Welcome To BioDAS.org' message, a 'Contents [hide]' section with a numbered list of links (1 About DAS, 2 What can I do with DAS?, 3 Publications, 4 The Protocol, 5 DAS events, 6 DAS/2, 7 Mailing Lists, 8 DAS FAQ, 9 Contribute to this Wiki), and an 'About DAS' section. The 'About DAS' section explains that the Distributed Annotation System (DAS) defines a communication protocol for exchanging annotations on genomic or protein sequences, and that it is a client-server system.

GMOD/GBrowse : ゲノムデータベースの標準化

- Generic Model Organism Database

- <http://www.gmod.org/>

- ゲノムブラウザの標準化 - GBrowse

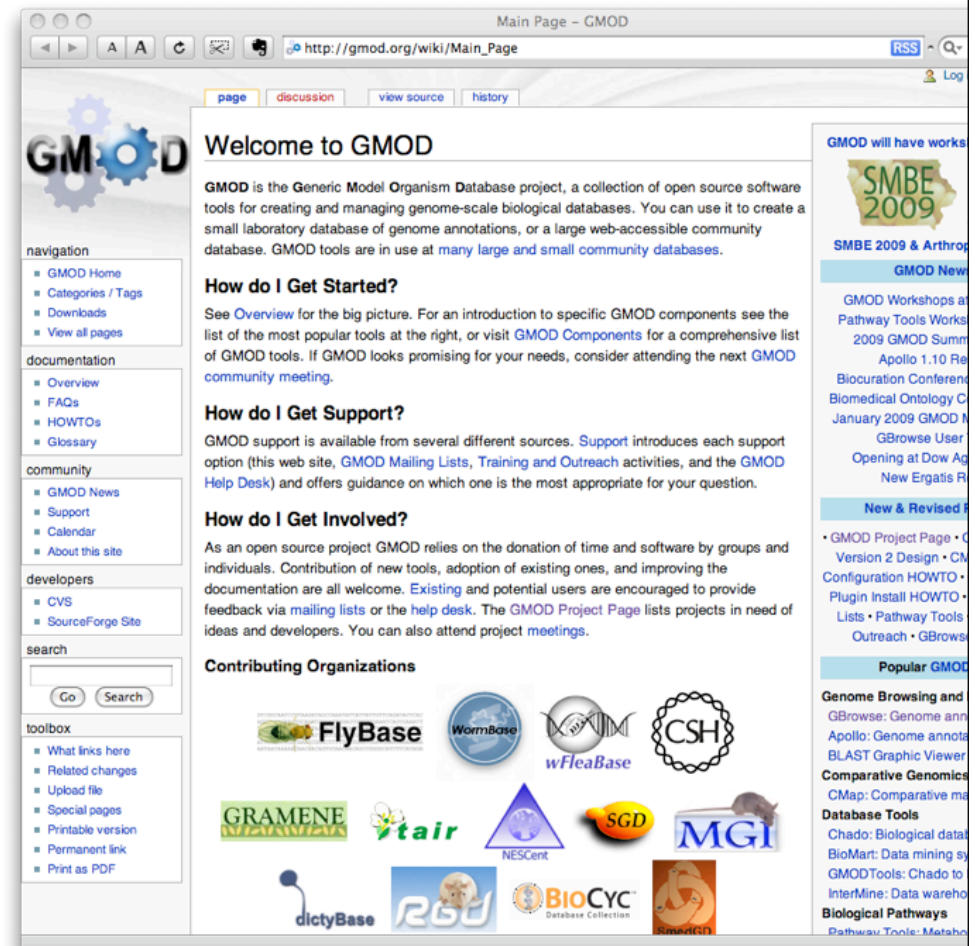
- アノテーションツール

- オントロジー

- データベーススキーマ

- パイプライン

- WormBase, FlyBase, HapMap, KEGG DAS など採用



OBO : アノテーションに使う用語を標準化

- OBO - Open Biomedical Ontologies

- <http://www.obofoundry.org/>

- GO - Gene Ontology

- <http://www.geneontology.org/>

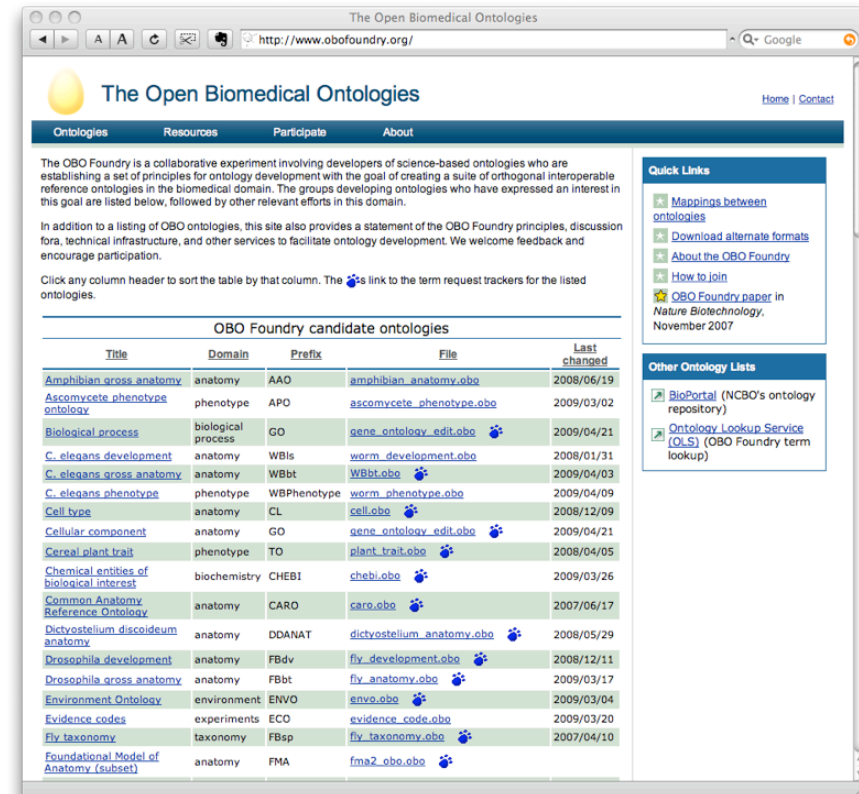
- molecular function

- biological process

- cellular component

- SO - Sequence Ontology

- <http://www.sequenceontology.org/>



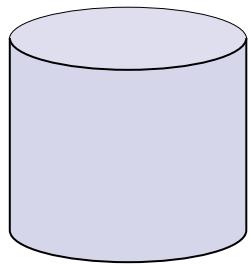
The screenshot shows the website 'The Open Biomedical Ontologies' with a table titled 'OBO Foundry candidate ontologies'. The table has columns for Title, Domain, Prefix, File, and Last changed. The table lists various ontologies such as Amphibian gross anatomy, Ascomycete phenotype ontology, Biological process, C. elegans development, C. elegans gross anatomy, C. elegans phenotype, Cell type, Cellular component, Cereal plant trait, Chemical entities of biological interest, Common Anatomy Reference Ontology, Dictyostelium discoideum anatomy, Drosophila development, Drosophila gross anatomy, Environment Ontology, Evidence codes, Fly taxonomy, and Foundational Model of Anatomy (subset).

Title	Domain	Prefix	File	Last changed
Amphibian gross anatomy	anatomy	AAO	amphibian_anatomy.obo	2008/06/19
Ascomycete phenotype ontology	phenotype	APO	ascomycete_phenotype.obo	2009/03/02
Biological process	biological process	GO	gene_ontology_edit.obo	2009/04/21
C. elegans development	anatomy	WBIs	worm_development.obo	2008/01/31
C. elegans gross anatomy	anatomy	WBbt	WBbt.obo	2009/04/03
C. elegans phenotype	phenotype	WBPhenotype	worm_phenotype.obo	2009/04/09
Cell type	anatomy	CL	cell.obo	2008/12/09
Cellular component	anatomy	GO	gene_ontology_edit.obo	2009/04/21
Cereal plant trait	phenotype	TO	plant_trait.obo	2008/04/05
Chemical entities of biological interest	biochemistry	CHEBI	chebi.obo	2009/03/26
Common Anatomy Reference Ontology	anatomy	CARO	caro.obo	2007/06/17
Dictyostelium discoideum anatomy	anatomy	DDANAT	dictyostelium_anatomy.obo	2008/05/29
Drosophila development	anatomy	FBdv	fly_development.obo	2008/12/11
Drosophila gross anatomy	anatomy	FBbt	fly_anatomy.obo	2009/03/17
Environment Ontology	environment	ENVO	envo.obo	2009/03/04
Evidence codes	experiments	ECO	evidence_code.obo	2009/03/20
Fly taxonomy	taxonomy	FBsp	fly_taxonomy.obo	2007/04/10
Foundational Model of Anatomy (subset)	anatomy	FMA	fma2_obo.obo	

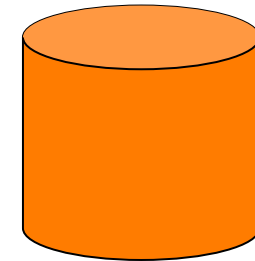
BioDAS : アノテーション情報の交換方法を標準化

- DAS - Distributed Annotation System (分散アノテーションシステム)
- <http://www.biodas.org/>

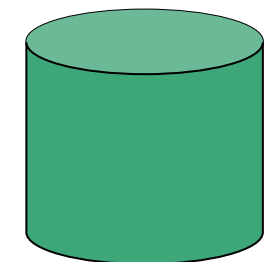
🇯🇵 (日)
プロモーターDB



🇬🇧 (英)
遺伝子予測DB



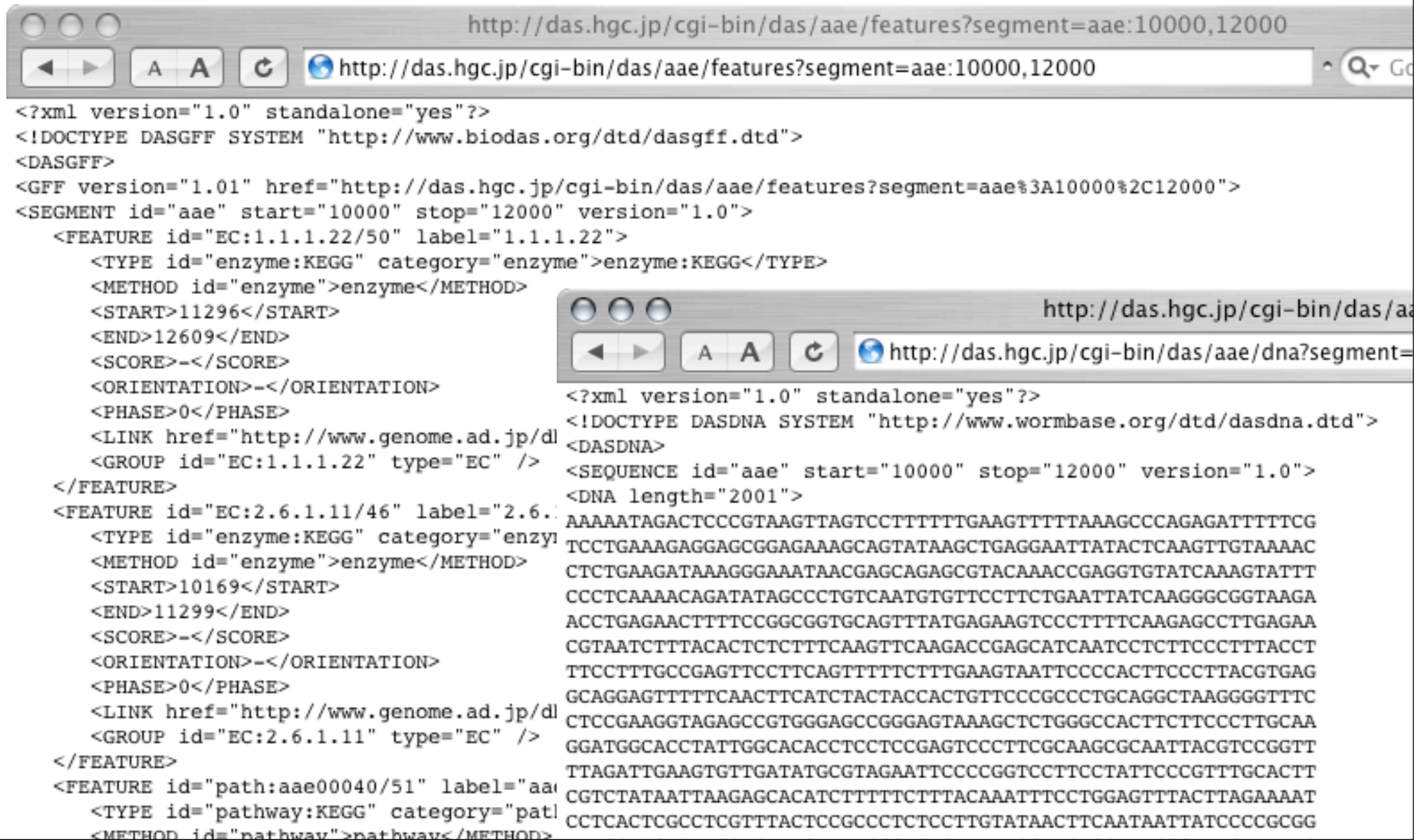
🇺🇸 (米)
遺伝子発現DB



ユーザ独自のデータ

The screenshot shows the BioDAS web interface for H. Sapiens (via NCBI-annotation April 2002) at chromosome position Chr1:45397191..45497190. The interface displays various genomic tracks including Overview of Chr1, Contigs, Transcripts, LocusLink genes, RefSeq Transcripts, and refSNPs. Arrows from the database icons point to specific data elements in the tracks.

BioDAS : DAS の XML データ



```
<?xml version="1.0" standalone="yes"?>
<!DOCTYPE DASGFF SYSTEM "http://www.biodas.org/dtd/dasgff.dtd">
<DASGFF>
<GFF version="1.01" href="http://das.hgc.jp/cgi-bin/das/aae/features?segment=aae%3A10000%2C12000">
<SEGMENT id="aae" start="10000" stop="12000" version="1.0">
  <FEATURE id="EC:1.1.1.22/50" label="1.1.1.22">
    <TYPE id="enzyme:KEGG" category="enzyme">enzyme:KEGG</TYPE>
    <METHOD id="enzyme">enzyme</METHOD>
    <START>11296</START>
    <END>12609</END>
    <SCORE>-</SCORE>
    <ORIENTATION>-</ORIENTATION>
    <PHASE>0</PHASE>
    <LINK href="http://www.genome.ad.jp/dl" />
    <GROUP id="EC:1.1.1.22" type="EC" />
  </FEATURE>
  <FEATURE id="EC:2.6.1.11/46" label="2.6.1.11">
    <TYPE id="enzyme:KEGG" category="enzyme">enzyme:KEGG</TYPE>
    <METHOD id="enzyme">enzyme</METHOD>
    <START>10169</START>
    <END>11299</END>
    <SCORE>-</SCORE>
    <ORIENTATION>-</ORIENTATION>
    <PHASE>0</PHASE>
    <LINK href="http://www.genome.ad.jp/dl" />
    <GROUP id="EC:2.6.1.11" type="EC" />
  </FEATURE>
  <FEATURE id="path:aae00040/51" label="aae00040">
    <TYPE id="pathway:KEGG" category="pathway">pathway:KEGG</TYPE>
    <METHOD id="pathway">pathway</METHOD>
```

```
<?xml version="1.0" standalone="yes"?>
<!DOCTYPE DASDNA SYSTEM "http://www.wormbase.org/dtd/dasdna.dtd">
<DASDNA>
<SEQUENCE id="aae" start="10000" stop="12000" version="1.0">
<DNA length="2001">
AAAAATAGACTCCCGTAAGTTAGTCCTTTTTTTGAAGTTTTTAAAGCCCAGAGATTTTTTCG
TCCTGAAAGAGGAGCGGAGAAAGCAGTATAAGCTGAGGAATTATACTCAAGTTGTAAGAAC
CTCTGAAGATAAAGGGAAATAACGAGCAGAGCGTACAAACCGAGGTGTATCAAAGTATTT
CCCTCAAACAGATATAGCCCTGTCAATGTGTTCTTCTGAATTATCAAGGGCGGTAAGA
ACCTGAGAACTTTCCGGCGGTGCAGTTTATGAGAAGTCCCTTTTCAAGAGCCTTGAGAA
CGTAATCTTTACACTCTCTTTCAAGTTCAAGACCGAGCATCAATCCTCTTCCCTTTACCT
TTCTTTTGCCGAGTTCTTTCAGTTTTTCTTTGAAGTAATCCCCACTTCCCTTACGTGAG
GCAGGAGTTTTTCAACTTCATCTACTACCCTGTTCCCGCCCTGCAGGCTAAGGGGTTTC
CTCCGAAGGTAGAGCCGTGGGAGCCGGGAGTAAAGCTCTGGGCCACTTCTTCCCTTGCAA
GGATGGCACCTATTTGGCACACCTCCTCCGAGTCCCTTCGCAAGCGCAATTACGTCCGTT
TTAGATTGAAGTGTGATATGCGTAGAATCCCCGGTCCCTTCTATTCCCGTTTGCACCT
CGTCTATAATTAAGAGCACATCTTTTTCTTTACAAATTTCCCTGGAGTTACTTAGAAAAT
CCTCACTCGCCTCGTTTACTCCGCCCTCTCCTTGATAACTTCAATAATTATCCCCGCGG
```

BioMart, Galaxy : 大規模データの取得と解析

BioMart - MartView
http://www.biomart.org/biomart/martview/73d83d4774884156f3e62818cf720574

bio::mart

HOME MARTVIEW MARTSERVICE DOCS CONTACT NEWS CREDITS

New Count Results URL XML Perl Help

Dataset: Homo sapiens genes (NCBI36)

Filters: Gene type: protein_coding, Chromosome: 4, Encode region: 4:118466104:118966103

Attributes: VEGA gene ID, VEGA transcript ID, Description, Gene Start (bp), Gene End (bp)

Dataset: [None Selected]

biomart version 0.7

BioMart - <http://biomart.org/>

ゲノムデータベースから必要な情報を絞り込んで取得

Galaxy - <http://main.g2.bx.psu.edu/>

取得したゲノムデータをウェブ上で解析するプラットフォーム

Galaxy
http://main.g2.bx.psu.edu/root

Analyze Data Workflow Libraries Help User

Tools: Get Data, Send Data, ENCODE Tools, Lift-Over, Text Manipulation, Convert Formats, FASTA manipulation, Filter and Sort, Join, Subtract and Group, Extract Features, Fetch Sequences, Fetch Alignments, Get Genomic Scores, Operate on Genomic Intervals, Statistics, Graph/Display Data, Regional Variation, Multiple regression, Evolution: HyPhy, Metagenomic analyses, Short Read Analysis, EMBOSS, Workflows

Build distribution of base quality

Quality score file:

No dataset? Read tip below

Execute

To use this tool your dataset needs to be in Quality Score format. Click pencil icon next to your dataset to set datatype to Quality Score (see below for examples of quality scores).

What it does

This tool takes Quality Files generated by Roche (454), Illumina (Solexa), or ABI SOLID machines and builds a graph showing score distribution like the one below. Such graph allows you to perform initial evaluation of data quality in a single pass.

Examples of Quality Data

Roche (454) or ABI SOLID data:

```
>seq1
23 33 34 25 28 28 28 32 23 34 27 4 28 28 31 21 28
```

Illumina (Solexa) data:

```
-40 -40 40 -40 -40 -40 -40 40
```

Output example

Quality scores are summarized as boxplot (Roche 454 FLX data):

boxplot of quality scores

History: refresh | collapse_all, Unnamed history, Your history is empty. Click on the left pane to start a new history.

- ・ 文部科学省委託研究開発事業
 - ・ ライフサイエンス分野におけるデータベース統合化の拠点形成を目指す
- ・ 中核機関
 - ・ ライフサイエンス統合データベースセンター
 - ・ 2007/4/1 情報・システム研究機構内に設立
- ・ 研究参画機関
 - ・ JST, CBRC, かずさDNA研, 九大, 奈良先端, 長浜バイオ, 東大, お茶女
- ・ 研究分担機関
 - ・ 京大, 東京医科歯科大, 阪大, 東大, 東海大, 日立


統合DBのミッション

- ・ 所在や利用法がわからない
- ・ 日本語化されていない
- ・ データベースを使いこなすための技術開発の不足
- ・ データベース構築、維持を担う人材の不足
- ・ 各機関やプロジェクトで構築されたデータベースがなかなか公開されない
- ・ プロジェクト終了後はデータベースが維持されない
- ・ 実験系と情報系の研究者間の相乗効果が小さい
- ・ 実験データと文献の知識との連携が弱い
- ・ データベース整備の中長期的戦略を担う体制が未整備



- ・ デー
- ・ 日本
- ・ デー
- ・ ライ
- ・ 人材
- ・ デー
- ・ アノ
- ・ 文南
- ・ 戦略

統合DBのミッション

- 
- ・データベース利用窓口であるポータルサイトの構築
 - ・日本語による研究情報流通の促進
 - ・データベース利用のための基盤技術開発
 - ・ライフサイエンスデータベース構築を担う人材の育成
 - ・データベース受入 / 維持のための体制構築
 - ・アノテーションによるデータの価値向上
 - ・文献データの活用
 - ・戦略の立案および体制の構築

情報・システム研究機構 (ROIS)



ライフサイエンス統合DBセンター (DBCLS)

ライフサイエンス統合データベースセンター

Database Center for Life Science

ライフサイエンス統合データベースセンター

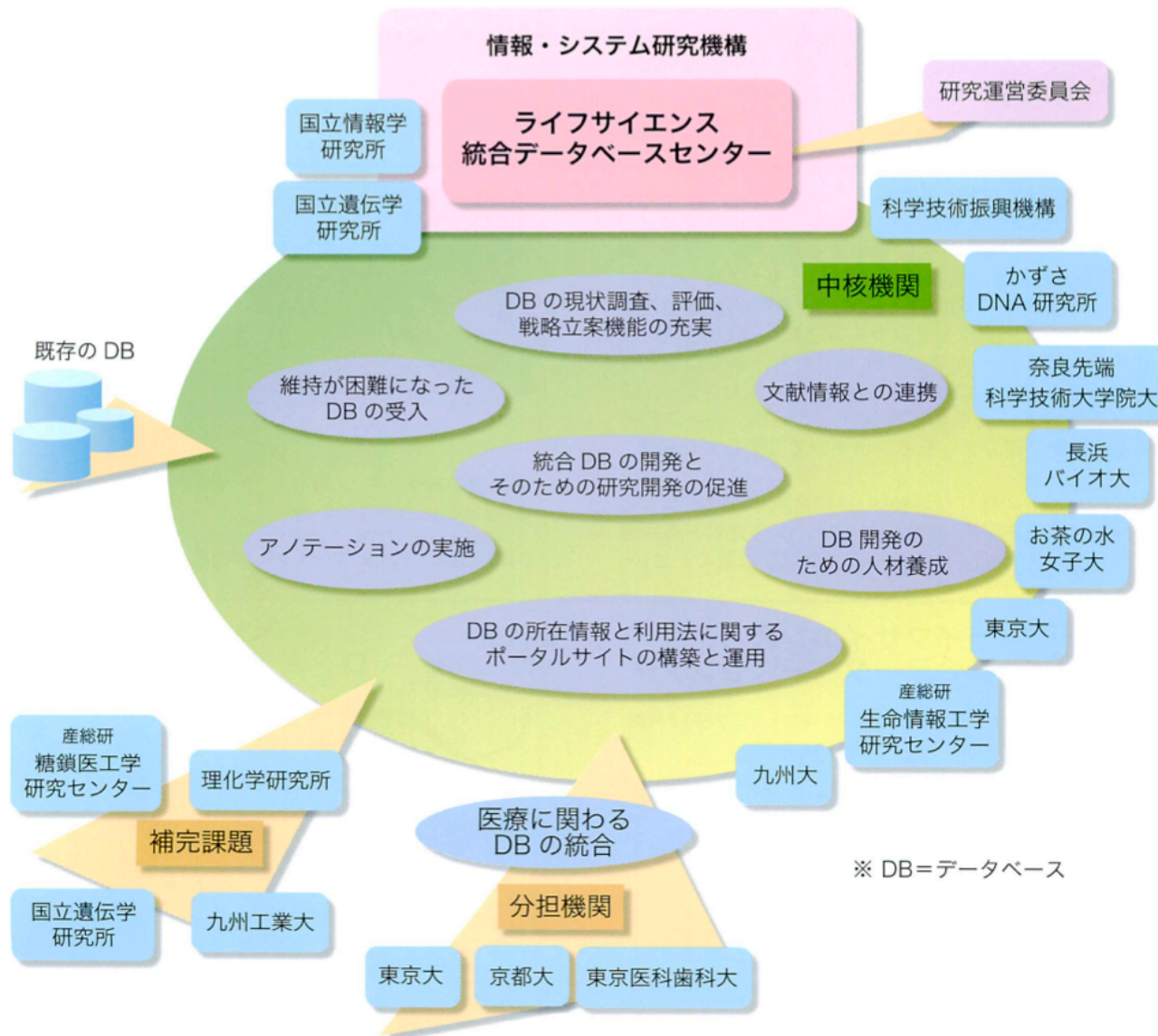
インフォメーション

- 2009/07/08 「生命科学系データベースアーカイブサービス」アップデートのお知らせ
2009年7月3日(金)より「[生命科学系データベースアーカイブサービス](#)」へ、シアノバクテリアの変異株ごとに表現型としてクロロフィル蛍光強度の時間変化を取得したデータベース「[FluoromeJ](#)」(東京大学准教授 園池公毅)を追加しました。
- 2009/07/06 新サービス公開のお知らせ
2009年6月12日(金)より文献中に用いられている英語表現の用例を高速に検索できる新サービス「[FinMeXes](#)」を公開し、本日「[統合ホームページ](#)」からのリンクを開設しました。
- 2009/06/25 「OpenID」と「OReFiL」停止のお知らせ
2009年6月26日(金)の17:00~17:30までメンテナンスのため「[OpenID](#)」と「[OReFiL](#)」のサービスを停止します。ご迷惑をおかけしますが、よろしくお願いたします。
- 2009/06/22 第8回国際バイオEXPOへ出展のお知らせ
2009年7月1日(水)~3日(金)に「[第8回国際バイオEXPO](#)(会場:東京ビッグサイト)」にて「統合データベースプロジェクト」を紹介するブースを出展いたします。
- 2009/06/12 新サービス公開のお知らせ
2009年6月12日(金)より微生物ゲノム配列のアノテーションを効率よく実現する情報環境、微生物ゲノムアノテーションパイプライン(Microbial Genome Annotation Pipeline)「[JMGAP](#)」を公開いたしました。
- 2009/06/05 「生命科学データベース検索」アップデートのお知らせ
2009年6月05日(金)より「[生命科学系データベース検索](#)」を用いて「[生命科学系データベースアーカイブサービス](#)」および「[ゲノム・ポストゲノム主要プロジェクト一覧](#)」のコンテンツが検索できるようになりました。
- 2009/05/22 「生命科学系データベースアーカイブサービス」アップデートのお知らせ
2009年5月22日(金)より「[生命科学系データベースアーカイブサービス](#)」の登録データベースが7件増えました。

- ・ 中核機関
 - ・ センター長 高木利久 教授
 - ・ 実は東大内に存在 (浅野)

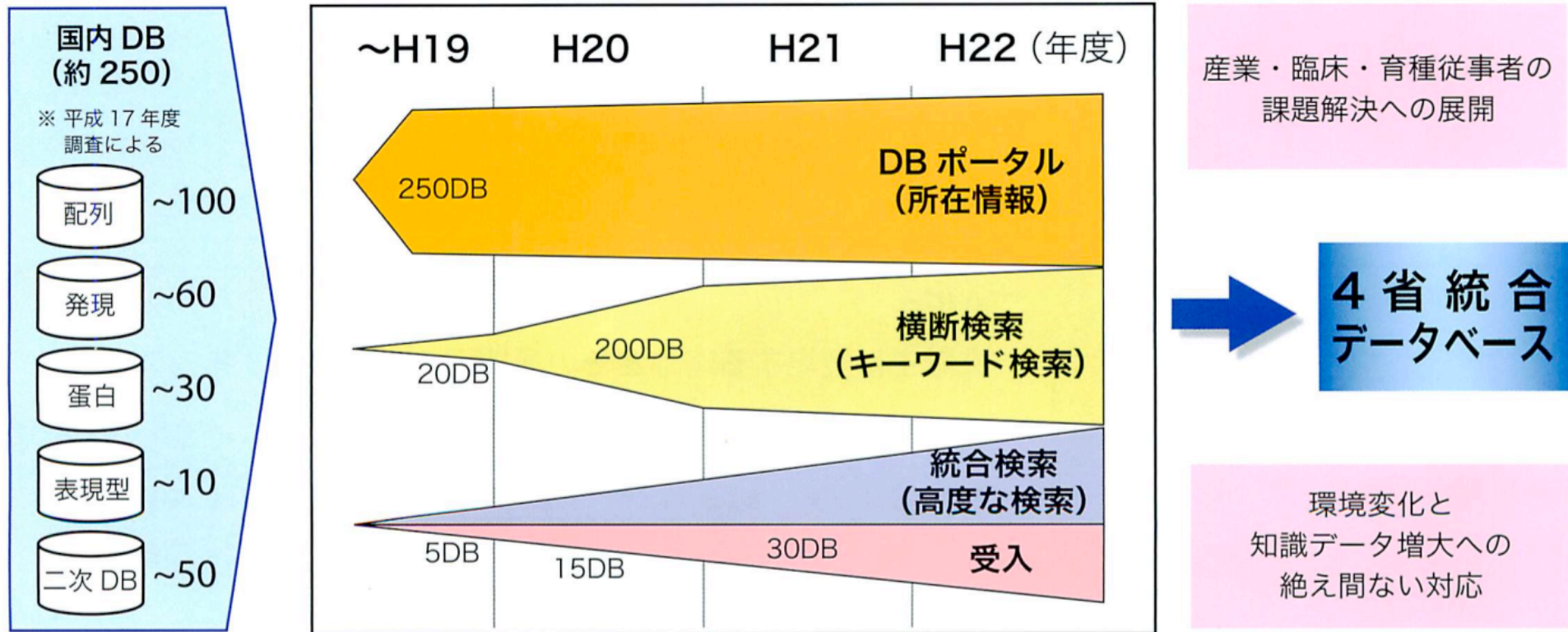


統合データベース整備事業の概要



統合化の年次計画

統合化の年次計画



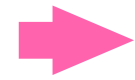
統合ホームページ



- ・ 横断検索 - 文献, 特許, 用語含む
- ・ 全文検索
 - ・ 蛋白質核酸酵素 (PNE)
 - ・ 文科省「ゲノム特定」報告書
 - ・ 学会要旨統合検索
- ・ DBカタログ, 学協会カタログ
- ・ アナトモグラフィー, 生物アイコン
- ・ 統合TV, MotDB
- ・ OReFiL, Allie, inMeXes
- ・ OpenID, TogoDB, TogoWS

http://lifesciencedb.jp/

データベースカタログ



[データベースカタログ] ライフサイエンス統合データベースプロジェクト-統合ホームページ-LSDB

http://lifesciencedb.jp/lpdb.cgi?pg=1

データベース

ホーム データベースカタログ 検索 ツール ダウンロード About us

Search: [検索欄]

Tree List

タイプ: 分類

トップ(612)

- DB型(343)
 - カタログ(10)
 - データバンク(20)
 - プログラム(67)
 - プロジェクト(134)
 - リソース(9)
 - 注釈(22)
 - 知識モデル(27)
 - 解析サービス(15)
 - 辞典(39)
- 対象(318)
 - 生物種(226)
 - ウイルス(3)
 - ミジンコ(1)
 - 全生物(2)
 - 微生物, シアノバクテリア
 - 微生物, 原生動物(3)
 - 微生物, 古細菌(5)
 - 微生物, 大菌菌(5)
 - 微生物, 枯草菌(4)
 - 微生物, 細菌 Bacillus(1)
 - 微生物, 酵母(7)
 - 植物, アサゴ(1)
 - 植物, イネ(16)
 - 植物, オオムギ, コムギ
 - 植物, シロイヌナズナ(1)
 - 植物, タバコ(1)
 - 植物, トウモロコシ(1)
 - 植物, トマト(1)
 - 植物, ヒメツリガネゴケム
 - 植物, ヒヤクニチソウ(1)
 - 対象(318)
 - アノテーション(24)
 - タンパク質, アミノ酸
 - タンパク質, プロテオ...
 - タンパク質, モチーフ(5)
 - タンパク質, 変異(1)
 - タンパク質, 局在(1)
 - タンパク質, 抗体(2)
 - タンパク質, 機能(9)
 - タンパク質, 立体構造(2)
 - タンパク質, 配列(10)
 - マーカー分子(1)
 - 代謝産物(2)
 - 化合物(1)
 - 医学(2)
 - 形態(8)
 - ヤシ(1)

データベース カatalogの一括ダウンロード

DBCLS
Database Center for Life Sciences

©2008 DBCLS | サイトポリシー | お問い合わせ

システム情報

トップページ

データベースカタログは世界に散在する生命科学系データベースの情報(URL・稼働状況・提供機関・生物種・データ種)を対象や生物種などの項目別に関連検索できるカタログです。現在の収録数593件。

新着情報

- PrognScan 2009/07/08 10:31:53
- ICOD 2009/07/07 14:12:13
- IBMD 2009/07/07 14:10:49
- 2009/0...
- KOMUGI 2009/06/15 20:19:50
- BMRB - Bl... 2009/06/03 19:44:15
- MachBase 2009/06/03 15:53:32
- 2009/0...
- CyanoClus... 2009/06/03 15:47:54
- WorTS 2009/06/03 14:41:26

参照履歴 コメント(0)

DaphniaBASE
ミジンコデータベース

画像を選択してください

表示形式: サムネイル ソート: 無し

- ABA
ホヤ発生過程の形態
- Brain Atl...
ヒト、ニホンザル、...
- C. elegan...
線虫RNA遺伝子破壊株
- MCTDB
Medaka Cr...
- Phenome A...
シロイヌナズナトラ...
- Rat Brain...
ラットの脳の断面画像

横断検索



生命科学データベース横断検索

国内外の生命科学データベースや文献を横断的に検索できるサービス

検索したいキーワードを一度入力するだけで、NCBIをはじめとした国内外のデータベースや特許、「蛋白質核酸酵素」、新聞記事、ウィキペディアなどを検索した結果を取得、閲覧できます。検索対象データベース数：>230(2009年4月現在)



共立出版 「蛋白質核酸酵素」 全文検索



総説誌である「蛋白質核酸酵素」のバックナンバーについて全文検索し、～2006年分は全文PDFファイルを、～2008年分は要旨を閲覧できます。
検索対象：1985～2008年
(2009年5月現在)

検索対象データの例

特許全文検索

分子データベースと特許データが一度にキーワード検索できる、国内初のサービスです。
検索対象：2004～2008年
(2009年5月現在)



その他、
・主要な分子データベース
・文科省「ゲノム」研究報告書
・Wikipedia(日・英)
など、多様な情報が対象です。

学会要旨統合検索



著者名、施設名のばらつきを吸収する辞書を搭載。著者名、所属先、発表タイトルや要旨中のキーワードで検索できます。結果は一覧表示のほか、キーワードの出現頻度(トレンド)がグラフで表示されます。

横断検索 - PNE



生命科学データベース横断検索

http://lifesciencedb.jp/dbsearch/?phrase=p53&button=検索

生命科学データベース横断検索 p53 データベース一覧 | HELP | 検索結果URL表示

データベース

- 全てのデータベース(20886)
- 総合DBプロジェクト(0)
- 文献(108)
 - 蛋白質核酸酵素(83)
 - 蛋白質核酸酵素 (近刊) (3)
 - 文科省「ゲノム特定領域」
 - 文科省「旧ゲノム特定領域」
 - 日本生物物理学会大会発表
 - BIOPHYS: 生物物理(2)
 - 科学新聞記事-(0)
- 特許関連文書(2217)
- 日本国特許(2217)
- 用語解説(122)
- 日本語版ウィキペディア(39)
- Wikipedia(English)(83)
- ゲノム・遺伝子・RNA(4918)
- 遺伝子発現・転写制御(3433)
- タンパク(1461)
- パスウェイ・相互作用・生体反応(243)
- 糖・脂質(3)
- 細胞・組織(5)
- 発生(0)
- 医療・薬(3217)
- 食品・栄養(1)
- 農学・環境(0)
- 生物圏・分類(83)
- 海外のデータベース(5075)

検索結果 ALL

検索結果 蛋白質核酸酵素

蛋白質核酸酵素 がん抑制遺伝子p53ファミリー研究の最近の進展 [one]

蛋白質核酸酵素 44 2 1999 101-111 総説 がん抑制遺伝子 p53 ファミリー研究の最近の進展 p53 Family Gen arch Division, Tokyo Medical and Dental University p53 は初めてウイルスによる実験的腫瘍発生で見いだされ、その重要な制御因子をコードする。一昨年、昨年と相対して p53 に類似のp73, p51が見いだされた。それぞれ神経系、筋、

http://lifesciencedb.jp/dbsearch/Literature/get_pne_cgpdf.php?year=1999&number=4402&file=p630Gz5V3VWvXAPgRvFrA==&search="p53"

191170 TUMOR PROTEIN 53; T responds to div

http://lifesciencedb.jp/dbsearch/Literature/get_pne_cgpdf.php?year=1999&number=4402&file=p630Gz5V3VWvXAPgRvFrA==&search="p53"

共立出版 「蛋白質核酸酵素」 全文検索

PNE 蛋白質核酸酵素

総説誌である「蛋白質核酸酵素」のバックナンバーについて全文検索し、～2006年分は全文PDFファイルを、～2008年分は要旨を閲覧できます。検索対象:1985～2008年(2009年5月現在)

学会要旨統合俯瞰システム



The Title Processing System

学会要旨統合俯瞰システム

表示: エントリー 学会: ANY ISSU: ANY LANG: 日本語 収録要旨

Author: Org: Keyword: 癌 Search

Items: 1-20 of 382 Previous Page: 1 of 20 Next Show: 20 Send to:

ISSU: 2007 ID: 1A08
AUTHOR: 田村 勝1; 田中 成和1; 藤井 智明1; 加藤 依子1; 城石 俊彦1
ORG: 1 遺伝研・哺乳動物遺伝
TITLE: Gasdermin family遺伝子群とヒト疾患
ABSTRACT: 我々は、これまでに皮膚・消化管上皮特異的に発現する新規遺伝子遺伝子がファミリーメンバーの一つであるGsdma3であること、2) Gsdmファミリー。しかしながら、未だGsdmファミリーの機能については、不明な点が多く残に、癌とGsdm遺伝子発現レベルの相関、ならびに遺伝子発現制御のメカニズム

ISSU: 2007 ID: 1W12-1
AUTHOR: 石井 直明; 石井 恭正; 宮沢 正樹
ORG: 東海大・医・分子生命科学
TITLE: 細胞傷害におけるミトコンドリアからの酸化ストレスの影響
ABSTRACT: ...

ISSU: 2007 ID: 1W12-7
AUTHOR: 大澤 郁朗; 石川 正洋; 高橋 久美子; 福田 慶二; 麻生 洋一
ORG: 日本医大・医・加齢科学・細胞生物
TITLE: 新しい抗酸化治療法: 細胞傷害性酸素ラジカルの水素による還元
ABSTRACT: ...

ISSU: 2007 ID: 1W18-1
AUTHOR: 伊藤 昭博1; 島津 忠広1; 前田 里子1; 吉田 稔1,2
ORG: 1 理研・吉田化学遺伝学; 2 JST・CREST
TITLE: HDAC分類とHDAC6の新規基質の同定
ABSTRACT: ...

ISSU: 2007 ID: 1W19-5
AUTHOR: 大貫 (長崎) 玲子1; 長崎 昇1; 藤田 聡史1,2; 増田 隆子1
ORG: 1 産総研・セル; 2 東大院・工学
TITLE: トランスフェクションアレイ技術を用いた細胞運動性評価チップの開発
ABSTRACT: ...

ISSU: 2007 ID: 1W19-6
AUTHOR: 安井 明
ORG: 東北大・加齢研・遺伝子機能
TITLE: 蛋白の可視化とプロテオミクスを融合した網羅的および手づる式ゲノム安定性ネットワーク解析法

試してみよう: 癌, siRNA

学会要旨統合俯瞰システム

表示: 学会俯瞰 学会: ANY ISSU: ANY LANG: 日本語 Fixed 収録要旨

Author: Org: Keyword: 癌 Search

ISSU ENTRY

Keyword: 癌

1984
1994
1997
1998

The Title Processing System

学会要旨統合俯瞰システム

表示: グラフ 学会: ANY ISSU: ANY LANG: 日本語 Fixed Fill 収録要旨

Author: Org: Keyword: 癌 Search

ISSU GRAPH

Keyword: 癌

試してみよう: エントリー 学会俯瞰 グラフ

1994 1997 1990 1999 2000 2001 2002 2003 2004 2005 2006 2007

学協会カタログ



[学協会カタログ] ライフサイエンス統合データベースプロジェクト-統合ホームページ-LSDB

http://lifesciencedb.jp/lsdb.cgi?gg=tool_academy

データベース | Guest Account | アカウント | ログイン | English

ホーム | 学協会カタログ | 検索 | ツール | ダウンロード | About us

Search:

Tree | List

タイプ: 分類

- トップ(801)
 - 理学(98)
 - 化学(25)
 - 家政(2)
 - 数学(8)
 - 生物(61)
 - 地学(2)
 - 複合(5)
 - 医学(336)
 - 基礎(68)
 - 保健(17)
 - 臨床(251)
 - 歯学(28)
 - 農学(108)
 - 工学(9)
 - 薬学(16)

画像を選択してください

表示形式: サムネイル | ソート: 無し

マリンバイオテクノロジー学会

応用動物行動学会

The Society of Population Ecology

マリンバイオテクノロジー学会

応用動物行動学会

個体群生態学会

The Society for the Study of the History of Regenerative Biology

種生学会

種生学会

情報計算化学生物学会 (CBI)

種生学会

種生学会

日本RNA学会

SGMV 日本ゲノム微生物学会 (SGMJ)

種生学会

日本バイオイメージング学会

日本バイオインフォマティクス学会 (JSBI)

日本バイオフィードバック学会

学協会 カタログの一括ダウンロード

DBCLS
Database Center for Life Science

©2008 DBCLS | サイトポリシー | お問い合わせ

情報 | コメント(0)

学協会概要

URL:
<http://www.jspp.org/>

名称:
日本植物生理学会

略称:
JSPP

分類:
理学/生物

関連:
生命科学

会員数:
(2006年5月18日現在)
3,200名 UMIN

会費:
通常会員 (国内); 12,000円、通常
会員 (国外); 70ドル (学生は35ド
ル)、団体会員 (国内); 40,000
円、賛助会員; 1口60,000円

学会備考:
1959年設立 UMIN

DNAバンク (INSDC) 目次



DNAデータベース (DDBJ/EMBL/GenBank=INSD) 総覧と検索, トップ

http://lifesciencedb.jp/ddbj/?country=all&graph=1&group=encode&id_start=1&id_width=20&order=count_ff&sample=

DNAデータベース (DDBJ/EMBL/GenBank=INSD) 総覧と検索

バージョン: DDBJ リリース 73 (2008-03-27)

DNAデータベースをプロジェクト単位で閲覧、検索、分析、取得 [詳細>>](#)

トップ プロジェクト検索: [実行&二元分類] レコード検索: [実行&二元分類] 配列検索(blastn): [実行&二元分類] [登録データ数の全容](#) [ダウンロード](#) [HELP](#)

すべてのレコード

生物群区分 (研究プロジェクト数)

ヒト (69803)	霊長 (6648)	霊長 (56832)	哺乳 (20770)	脊椎 (30005)	無脊椎 (50597)	植物 (93522)	バクテリア (68578)	ウイルス (30909)	ファージ (1535)	合成 (4352)	環境 (8006)	特許 (101517)	未分類 (72)	すべて (534483)
------------	-----------	------------	------------	------------	-------------	------------	---------------	--------------	-------------	-----------	-----------	-------------	----------	--------------

研究の型別分類 (研究プロジェクト数)

mRNA (26484)	機能RNA・RNAゲノム (175)	免疫 (3644)	嗅覚 (39)	マーカー (13985)	エクソン構造 (11419)	集団 (562)	オルガネラ (108)	EST (1005)	GSS (431)	STS (654)	HTC (446)	HTG (10778)	TPA (142)	UNA (0)	CON (173)	すべて (69803)
--------------	--------------------	-----------	---------	--------------	----------------	----------	-------------	------------	-----------	-----------	-----------	-------------	-----------	---------	-----------	-------------

ヒト x ゲノム(マーカー)型 [詳細>>](#)

レコード数: 98,078 表示切替: 登録国単位
 研究プロジェクト数: 13,085

研究プロジェクトの一覧
 サイズ順 / 登録日順

研究プロジェクトのサイズ(レコード数)の分布

研究プロジェクト名	例	サイズ (レコード数)	比 (%)	バンク登録日
1 NotI flanking sequences: a tool for gene discovery and verification of the human genome 詳細	AJ322533	21,361	217.80	2001-05-16
2 Purification of CpG islands using a methylated DNA binding column 詳細	X76662	12,265	125.05	1993-11-03
3 アフリカ人内の遺伝的差異はユーラシアとアフリカの差より大きい 詳細	AY286547	1,500	15.29	2003-04-29
4 Homo sapiens chromosome 19 clone CTD-2547N9, *** SEQUENCING IN PROGRESS *** 詳細	AC010322	1,415	14.43	1999-08-03
5 The 1993-94 Genethon human genetic linkage map 詳細	Z16742	1,267	12.92	1992-10-05
6 レトロウイルスプラスミドライブラリによるレポーター遺伝子アッセイでヒトゲノムのプロモーターを同定 詳細	AY270202	1,051	10.72	2003-04-07
7 Transcription start regions in the human genome are favored targets for MLV integration, Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses 詳細	AY515855	1,026	10.46	2004-01-02
8 ソータで回収した6番染色体のTaqI断片 詳細	Z77857	1,005	10.25	1996-08-01
9 Homo sapiens genomic DNA 詳細	AP000485	967	9.86	1999-05-10
10 Foamy ウイルスの挿入部位のゲノムワイドマップ 詳細	DQ192669	847	8.64	2005-09-06

試してみよう: ヒト x mRNA
植物 x オルガネラ

試してみよう: 詳細

http://lifesciencedb.jp/ddbj/

遺伝子発現バンク (GEO) 目次



試してみよう：国別分布

試してみよう：cell cycle, circadian

遺伝子発現バンク(GEO)目次

http://lifesciencedb.jp/geo/

遺伝子発現バンク (GEO) 目次 バージョン：2007-06-07
GEO に登録されているデータを、測定技術と材料の属性に基づいて整理しました。

登録データリスト | **国別登録データ分布** | 登録データ推移 | 登録データ全容

データ単位：[データセット / サンプル / プラットフォーム] 単位の説明>>
各タブ内に表示される数値は、そのタブ分類に属するデータ数です。

ヒト (3,177)	霊長 (22)	霊長 (2,272)	哺乳 (85)	脊椎 (122)	無脊椎 (441)	植物 (1,459)	バクテリア (497)	ウイルス (7)	ファージ (4)	未分類 (35)	すべて (8,121)		
SAGE NlaIII (127)	SAGE RsaI (1)	SAGE Sam3A (10)	MPSS (16)	GeneChip (3,190)	タイリングアレイ (725)	cDNAアレイ (2,716)	オリゴアレイ (1,259)	ピースアレイ (31)	タンパク質アレイ (3)	抗体アレイ (10)	RT-PCR (4)	その他 (29)	すべて (8,121)

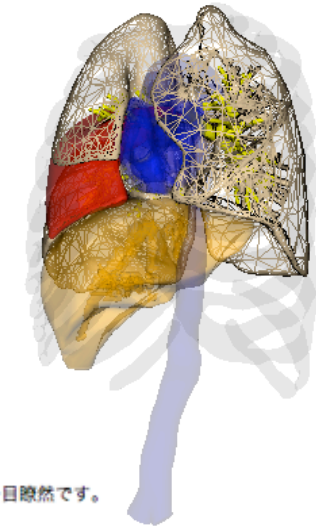
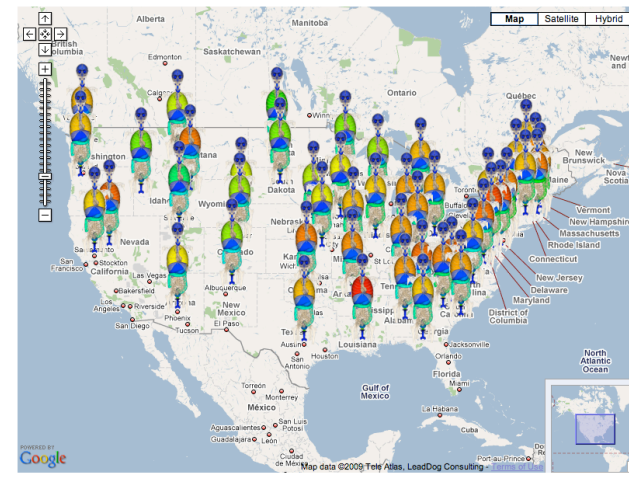
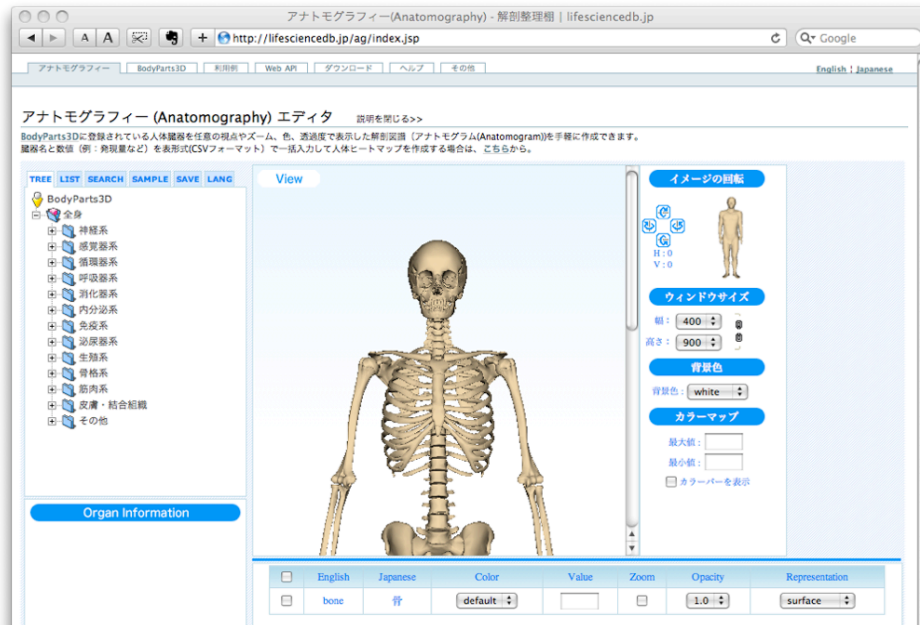
1 | 2 | 3 | 4 | 5 | > [407]

タイトル	プラットフォーム	登録機関	生物種	データサイズ (プローブ数×サンプル数)	測定サンプル														NCBI FTP サイト
					脳	血	結合	生殖	筋	消化	肝	肺	腎	分泌	混合	胎児	分類不能		
1 Mouse Atlas of Gene Expression Project (GSE4726)	[SAGE NlaIII] SAGE:17:NlaIII:Mus musculus (GPL1349)	Canada's Michael Smith Genome Sciences Centre	ハツカネズミ	374,564,124 (1,891,738×198)	70	9	5	14	11	12	3	7	9	22	17	19	by_platform by_series		
2 High Resolution Mapping and Functional Analysis of the Methylome in Arabidopsis (mCIP, MBD) (GSE5094)	[タイリングアレイ] A1Tile1F to Arabidopsis Tiling 1.0F (GPL1979)	University of California, Los Angeles	シロイヌナズナ	148,433,280 (6,184,720×24)													24 by_platform by_series		
3 CGAP SAGE (GSE14)	[SAGE NlaIII] SAGE:10:NlaIII:Homo sapiens (GPL4)	National Cancer Institute	ヒト	145,848,360 (694,516×210)	67	13	11	31	1	14	2	4	3	36	25	1	2 by_platform by_series		
4 Genes regulating B cell development are mutated in acute lymphoid leukaemia (GSE5511)	[GeneChip] Affymetrix GeneChip Mapping 500K Set Array (250K_Sty2_SNP) (GPL3720)	St Jude Children's Research Hospital	ヒト	87,934,176 (238,304×369)		369											by_platform by_series		
5 Expression Project for Oncology (expO) (GSE2109)	[GeneChip] Affymetrix GeneChip Human Genome U133 Plus 2.0 Array (GPL570)	IGC	ヒト	87,808,050 (54,675×1,606)	8	27	92	396		303	38	108	242	350	12	30	by_platform by_series		
6 Global variation of copy number in the human genome_EA (GSE5013)	[GeneChip] Affymetrix GeneChip Mapping 500K Early Access Array (250K_Sty_SNP) (GPL3812)	Affymetrix, Inc	ヒト	81,784,314 (267,269×306)													306 by_platform by_series		
7 Global variation of copy number in the human genome_EA (GSE5013)	[GeneChip] Affymetrix GeneChip Early Access Mapping 500K Set Array (250K_Nsp_SNP) (GPL3811)	Affymetrix, Inc	ヒト	81,772,686 (267,231×306)													306 by_platform by_series		
8 Global variation of copy number in the human genome_COMM (GSE5173)	[GeneChip] Affymetrix GeneChip Mapping 500K Set Array (250K_Nsp_SNP) (GPL3718)	Affymetrix, Inc	ヒト	72,405,288 (262,338×276)													276 by_platform by_series		

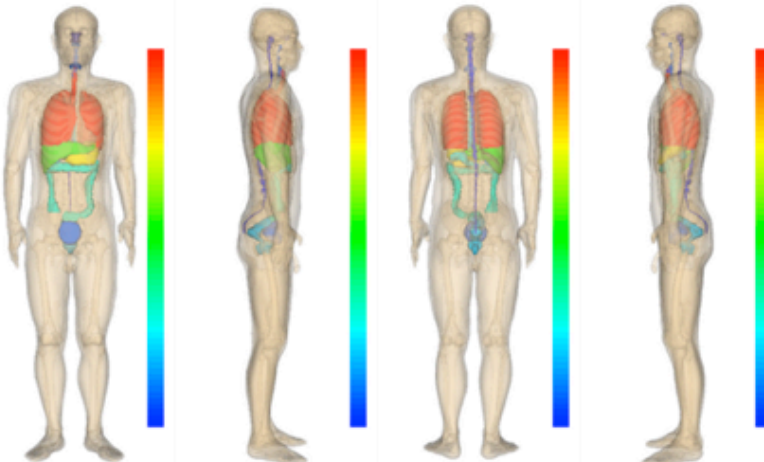
試してみよう：サンプルのサマリー

http://lifesciencedb.jp/geo/

解剖学用語 アナトモグラフィー ポリゴンマン



以下の例は、がんの部位別の死亡数統計（国民衛生の動向 下表赤枠内、2006年）をアナトモグラフィー表示したものです。肺がん（赤）が最も多く、胃がん（黄）、肝臓がん（緑）の順に多いことが一目瞭然です。



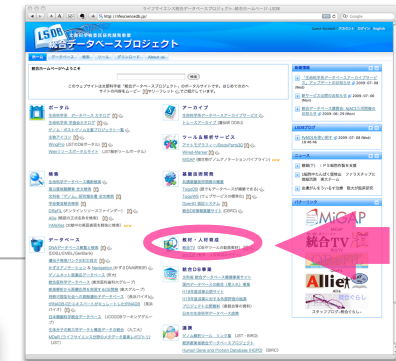
2006年「国民衛生の動向」

第14表 (3-1) 部位別男性死 死亡数の推移

死因国際分類コード	死 因	昭和30年 (75)	35 (70)	40 (85)	45 (90)	50 (95)	55 (00)	60 (05)
02300	悪性新生物	76 922	93 594	110 660	130 285	153 623	179 343	193 696
02302	口、咽、喉、気管支、肺	943	1 233	1 462	1 886	2 360	3 613	4 834
02303	胃	3 862	4 436	5 046	6 064	7 233	8 396	9 459
02304	肝臓	2 862	3 842	5 146	7 791	10 420	12 139	13 305
02305	胆嚢、胆管、膵臓、膵臓及び膵臓	3 337	3 882	4 300	5 495	6 852	7 723	8 537
02306	胃及び胃以外の消化器	6 477	9 741	13 780	17 796	22 773	23 602	23 417
02307	膵臓及び膵臓以外の消化器	1 905	2 791	3 949	5 069	6 189	6 913	7 471
02308	膵臓	3 156	4 483	5 963	7 317	8 965	10 340	11 931
02309	膵臓	738	736	747	776	872	958	984
02310	気管支、気管支及び肺	10 711	15 438	20 837	26 872	33 389	39 953	43 821
02311	肺	371	327	367	381	451	502	555
02312	気管支	27	44	34	34	34	37	65
02313	肺	-	-	-	-	-	-	-
02314	肺	-	-	-	-	-	-	-
02315	肺	1 267	1 736	2 443	3 460	5 398	7 544	8 849
02316	肺	1 174	1 606	1 706	2 110	2 700	3 184	3 711
02317	気管支、気管支及び肺	253	509	993	896	906	869	676
02318	気管支、気管支及び肺	-	-	-	-	-	-	-
02319	気管支、気管支及び肺	45	2 624	2 983	3 225	3 645	3 970	4 133
02320	その他の呼吸器、消化器及び泌尿器	-	-	-	-	-	-	-
02321	大	5 799	7 724	10 112	13 286	17 312	19 888	21 825

がんの部位別の死亡数統計
- 国民衛生の動向 2006年 -
肺がん（赤）
胃がん（黄）
肝臓がん（緑）

統合TV 様々なDBやツールの使い方をムービーで解説



統合TV (togotv) <http://togotv.dbcls.jp/>

使い直し系チャンネル
統合TV
統合TV (「とうごうていびい」もしくは「とうごうてれび」と発音します) は、ライフサイエンス統合データベースセンター(DBCLS)が発信する動画によるデータベースやツールの使い方のコンテンツ (β版) です。ご意見は各コンテンツへのツッコみや (YouTubeやニコニコ動画でも承ります)、ページ最下部のメールアドレス (イメージファイル) へ電子メールでどうぞ。

統合TVの使い方 番組を検索する:

番組カテゴリから探す: DBCLS | EMBOSS | English | Firefox | IE6 | IE7 | IE8 | Firefox | 辞書 | 設計ツール | 多型情報 | 二次構造 | 配列解析 | 発現情報 | 文献検索

1. NCBI BLASTを使って機能未知塩基配列の機能を特定する

2009年7月10日(金)に統合データベース講習会 AJACS

2009-07-08

[macosx][firefox][塩基配列][配列解析] BLASTを使って類似配列

NCBI Basic | ここで行い、環境D shock | 微生物する、画像を

Flashよりも高画質なQuickTimeムービー版はこちら。YouTube

[ツッコミを入れる]

統合TV - PyMOL2 http://togotv.dbcls.jp/movie/090707PyMOL2_f.html

PyMOLを使い倒す

```
MacPyMOL File Edit Build Movie Display Setting Scene Mouse Wizard Help
MacPyMOL
Recently: selection "sele" defined with 1940 atoms.
PyMOL> select from sele
Selector: selection "sele" defined with 354 atoms.
PyMOL> select (resi 100-200) and (chain A)
Selector: selection "sele" defined with 170 atoms.
PyMOL> select (resi 100-200) and (ss h)
Selector: selection "sele" defined with 140 atoms.
PyMOL> select a, chain A
Selector: selection "a" defined with 4692 atoms.
```

この方法を使うと、特定の残基を強調するなど、複雑な分子イメージを作成することができます。

Mouse Mode 3-Button Viewing
Buttons L R Wheel
s Key Rotate Move Move Slab
Shift -Box -Box Clip Move
Ctrl -/- PRAT PRT Move
Ctrl -/s PRAT PRT Move
Space -/- Move Move Move
Click Menu PRAT
Selecting Residues
Scale 1.0 0.5sec

作成日: 2009年7月7日

Generated by DesktopToMovie

さらなるデータベース統合に向けて

- ・ 標準化と相互運用性
 - ・ ドメイン毎に異なるデータフォーマット (XML, GFF etc.)
 - ・ ウェブサービスの連携 (SOAP/REST etc.)
- ・ 高度な検索技術開発
 - ・ セマンティックウェブ

→ 国際開発者会議 BioHackathon の開催



そもそもBioHackathonとは



- Hackathon = ハック + マラソン = 開発合宿
 - 約1週間、20~40人くらいの開発者を缶詰にしプログラム開発
- BioHackathon
 - バイオインフォのツール、ライブラリに様々な不整合や課題があった
 - 世界中に分散している BioPerl など Open Bio* の開発者を集めて改善
 - バイオインフォに興味をもっていた O'Reilly などがスポンサーに
 - オープンソースによる開発の利点とコミュニティの醸成
 - 最先端、最高レベルの開発者を1週間タダ働きさせられる
 - 開発者間での情報流通によるお互いのフィードバックと効率
 - 成果物は誰でも利用できるように公開される
 - 結果的にバイオインフォの環境が改善され市場も開拓される？！

BioHackathonという文化

- 2002 第1回, 第2回 BioHackathon
 - Open Bio* の各プロジェクトからコアディベロッパーが集結
 - 1月 アリゾナ州 Tucson - スポンサー O'Reilly
 - 2月 南アフリカ Cape Town - スポンサー Electric Genetics
- 2003 第3回 BioHackathon
 - 2月 シンガポール - スポンサー Apple Asia
 - 主に OBDA の整備が完了
 - BioSQL, Chado, GBrowse, BioMoby など
- BioPAX, GMOD, SBML などさまざまな Hackathon
- 2006 Phyloinformatics Hackathon
 - 12月 ノースキャロライナ州 - スポンサー NESCent, Duke大学



BIOHACKATHON 26-28 JAN 2002, TUCSON, ARIZONA, USA





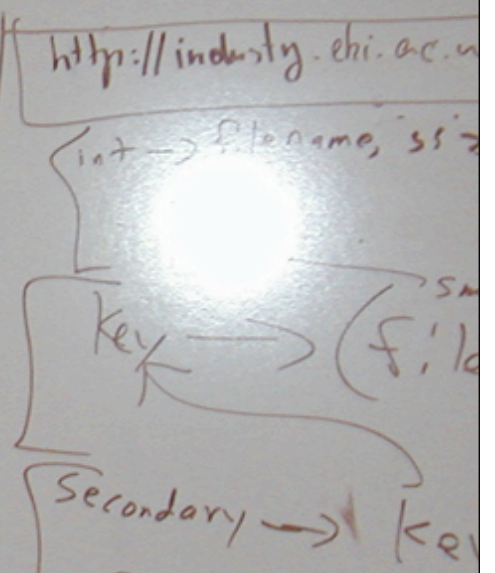
SAVE MASTER Module List

	Spec	Java	Perl	Python	Ruby
Registry	✓	Brian	2/1 ✓ 2/14 ✓	Jeff 2/1 (✓)	
Index - BDB	✓		Lincoln	Andrew	
Index - Flat	✓	Matt	Nichole	Andrew	
Bio Fetch	✓	Dyfad	Kristen Heike	Andrew	server ✓ Antoine
CORBA	✓	ch ✓ sue ✓	ch Enan ✓ sue ✓	ch ✓ sue ✓	
BESQL	✓	Paul ✓ (cust) write ✓	Paul ✓ Birk Mark ✓ Wick ✓	Rand ✓ Vi ✓	
Xenobl	✓	Cyl SOAP	cyi *? SOAP ✓	cyi ✓ SOAP X	
UPDT		Matt Brian			

REGISTRY mysql -u root

92/68 .0.95

Super Store



Gentoo k/day

P1234 100
P1234 10

37 minutes -> DB
14 + 1:24 minutes -> Plat

BioHackathon 2008 開催の意図

- ・ ライフサイエンス統合データベースセンターのミッション
 - ・ 国内のライフサイエンスDBを統合する
 - ・ 現実には（特にDDBJやKEGGなど大規模なDBの統合は）困難
 - ・ バーチャルに統合するのが現実的
 - ・ 現状ではおそらくウェブサービスが最適
- ・ どのように統合するのがベストか
 - ・ BioMoby?
 - ・ WS-I?
 - ・ セマンティックウェブ?
 - ・ 既存のライブラリ?
 - ・ 瀧内メソッドの出番



新しい技術を習得することは難しい

まず最初に直面する状態・・・自分の知らない領域で身動きが取れない状態・・・

ここから抜け出す事が技術自体の学習よりも遥に困難なのだ

その分野に精通した人が適切な入門書(ドキュメント)を提示してくれる・・・

さらに言えば、出向いてもらい、手取り足取りボランティアで教えてもらう・・・

私はそんな願望一杯の状況を自分で簡単に作り出す恐るべき方法を思いついた

「勉強会」にして人を集めればいいんだ・・・

「序章：頼むより巻き込め！！」(瀧内メソッド入門)より抜粋

BioHackathon 2008 for Web Services

Dates & Venue

Feb 11-15 @ Roppongi Academy Hills & Computational Biology Research Center, Tokyo, Japan

Objectives

Increase interoperability of bioinformatics web services through

- Standardize data types among web services
- Provide stable and functional services by cross platform validation

To accomplish this, gather developers from various domains

- Major web service providers (DDBJ, KEGG, PDBj, CBRC, EBI, INB, BioMoby, ...)
- Open Bio* library projects (BioPerl, BioJava, BioRuby, BioSQL, ...)
- Workflow & client software projects (Taverna, G-language, Cytoscape, ...)
- Standard data exchange format providers (PSI-MI, NEXML, GLYDE-II, ...)

Conference web site

<http://hackathon.dbcls.jp/>

Sponsored by

Database Center for Life Science (DBCLS) & Computational Biology Research Center (CBRC)



分野間の連携

ウェブサービスのプロ
バイダ

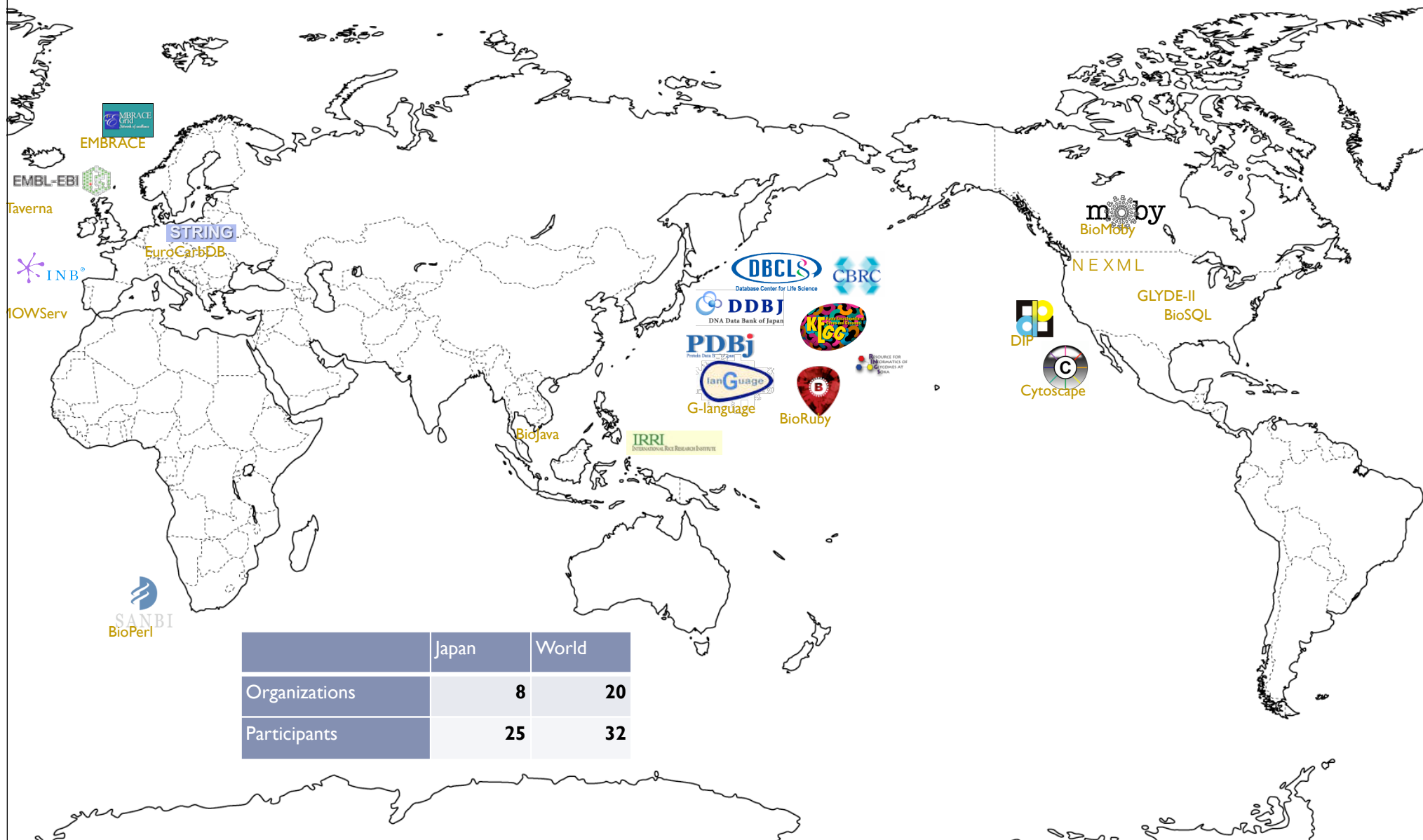
Open Bio* の
開発者

ウェブサービスのク
ライアント開発者

BioMoby

交換データ
フォーマットの
標準化

BioHackathon 2008 participants



	Japan	World
Organizations	8	20
Participants	25	32

BioHackathon 2008 - program

• 2/11 morning session - presentations

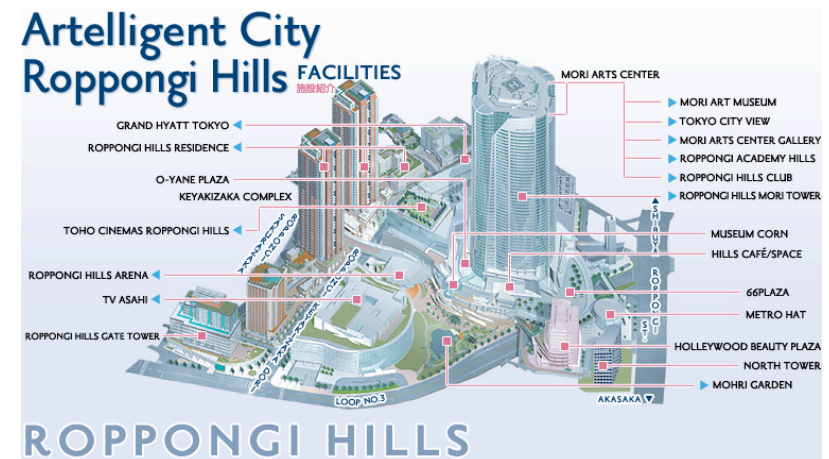
- From Web API for Biology (WABI) to Semantic Web API for Biology (SABI) - Hideaki Sugawara
- Current status of the BioMOBY project and vision for the future directions - Mark Wilkinson
- The EMBRACE project and WS-I standard - Jan Christian Bryne
- Soaplab2 project to wrap up command line packages - Martin Senger
- Proxying legacy applications and CGIs into BioMOBY - Paul Gordon
- Generation Challenge Program effort at building interoperability - Richard Bruskiwich

• 2/11 afternoon session - open space

- ServiceProvider_workgroup
- OpenBio_workgroup
- Workflow_workgroup
- BioMOBY_workgroup
- Exchangeformat_workgroup
- Semanticweb_workgroup
- PhyloWS_workgroup
- Distributed storage - initiative (BIO.SLURP)
- Distributed services - initiative (BIO.ORG.ANISM)

• 2/12-15 hack, hack, hack

- We will have banquet on 2/12 18:30- @CBRC



オントロジーと交換データ型の標準化

- Open Bio Semantics 共通データ型
 - バイオインフォの WS で交換される全データ型の標準化が望ましい
 - 少なくともドメイン毎のデータモデルは必須
- 個別ドメイン
 - Glycoinformatics (糖鎖)
 - 糖鎖データ解析フローの BioMoby/Taverna による実現
 - 相互作用ネットワーク
 - PSI-MI を採用し IntAct などで共通の WS を構築、Cytoscape で利用
 - Phyloinformatics (系統解析)
 - NEXML の利用促進、入出力型の整理と対応する解析 WS のデザイン
 - テキストマイニング
 - UIMA とバイオインフォの融合、iHOP は BioMoby 上に載っている

交換データ型の標準化の流れ

- BioMoby のオブジェクトオントロジー
 - 似たようなオブジェクトが大量に登録されている (>500)
 - BLAST の出力など主要なデータ型はフラットに formatted-text 下
 - BioMoby サーバの構築が Perl/Java 以外では困難
- SOAP/WSDL (complexType) - XML schema
 - 標準データ型で定義し .xsd ファイルを共有 (open-bio.org でホスト?)
 - 同じオブジェクトモデルを Open Bio* とウェブサービス間で利用
- UML (BioUML?)
 - BioPerl, BioRuby, BioPython, BioJava で共通のモデルを設計
 - 相互運用性のあるオブジェクトモデルからコードを生成
 - もしくは Bio* は共通の C ライブラリに対するバインディングを使用

非同期通信 (Async service)

- ・ バイオインフォの解析フローでは時間のかかるステップも多い
 - ・ サーバにリクエスト→ジョブIDを発行→終わっているかポーリング
 - ・ 双方に WS サーバ機能があれば、WS-Addressing などでの通知
- ・ BioMoby は WSRF を採用 (時代遅れになりつつある技術とも)
 - ・ WSRF は OGSA でも利用されている
 - ・ 実装が無く Ruby では利用できない
- ・ Soaplab にも非同期実装がある
- ・ → WS-I に準拠してほしい (WS-* などのスタックが必要)
 - ・ WSO2 (<http://wso2.org/>) の実装が良くなれば?

大規模データ (Large data) とセキュリティ

- ・ SOAP では MIME アタッチメントで送られがち
 - ・ Base64 エンコードされ一段と巨大化、全部届くまで解凍できない
- ・ データに対するリファレンスを渡して DL は REST や bittorrent 等で？
 - ・ リファレンスにはデータ型の情報がつけられない
 - ・ クライアントが解析のためのデータを一時的における場所の提供
 - ・ データの生存期間の指定
 - ・ 解析ワークフローではクライアントを介さずサーバ間で受け渡す
- ・ セキュリティ
 - ・ X509 電子署名
 - ・ ユーザのデータ保護、サーバのデータ秘匿、計算機資源 の割当などを管理

ワークグループごとの成果

The image displays two overlapping browser windows from the BioHackathon website. The top window shows the main page with the BioHackathon logo and navigation tabs. The bottom window shows a list of meeting topics for BioHackathon 2008.

まとめサイト
<http://hackathon.dbcls.jp/>

BioHackathon 2008
Towards integrated web service in life science with Open Bio* libraries.

Summary
The number of web service providers in year. In theory, these services are inter languages. However, each service uses naming conventions. Moreover, these se (partly, due to the lack of compliance of library).

This lack of interoperability prevents the users and developers alike and they are research.

The ⇒ **BioMOBY** project has tried to over data types and methods used in the ser providers which are not yet participating are several web service providers in Jap but none of these are integrated in any

For this reason, at ⇒ **DBCLS**, we started service providers. If all [⇒ <http://www.o> interoperability among these services ar of web services by every web service pr

BioHackathon 2008 meeting topics

New version

- **OpenBio* & Web Service Semantics** - Richard Bruskiwich
 - domain models & ontology: data types, service types and model API
- standardization
 - **Glycoinformatics** - Will York
 - **Interaction networks, DAS** - Bruno Aranda
 - **Text Mining, Text mining** - Kano
 - **Phyloinformatics** - Hilmar Lapp
- **Async services** - José M^a Fernández
 - WSRF?
- **Large data** - Tom Oinn



Roll call at Villa Fontaine Shiodome

2008 Feb 11

Photo by Dr. Heikki Lehvaslaiho

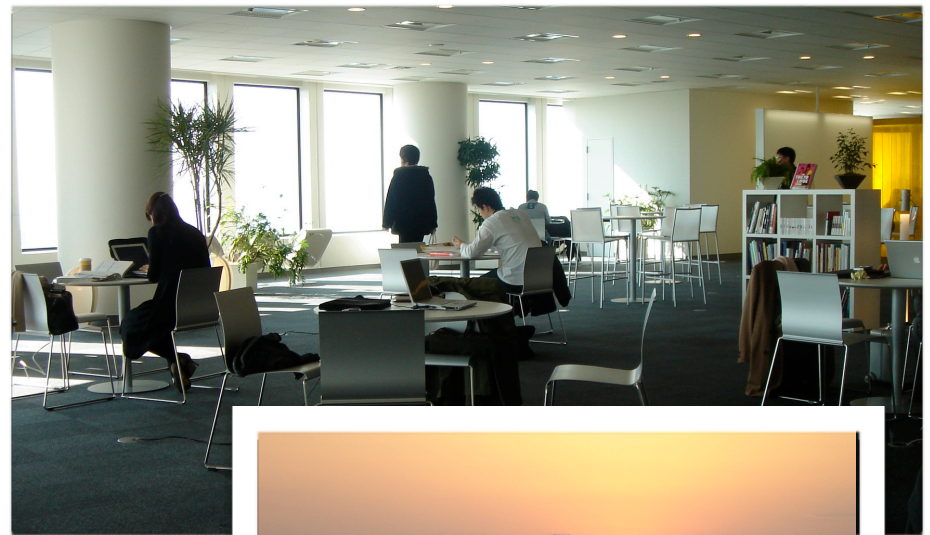
academyhills 40F
 Quasres C4
 キャラント04
 10:30~
 17:00
 Bio Hackathon



BioHackathon
 2008
 for Web Services
 Feb 11-15 @ Tokyo, Japan
 #channel/biohackathon-2008.tokyo
 #channel/biohackathon2008
 @biods.jp

BioHackathon
 2008
 for Web Services
 Feb 11-15 @ Tokyo, Japan
 #channel/biohackathon-2008.tokyo
 #channel/biohackathon2008
 @biods.jp

Open Space





Hack
Hack
Hack

&

Banquet

@

CBRC

2/12-15

2008

主要センターのウェブサービスと BioMoby

- NCBI SOAP
- EBI Web Service
- DDBJ XML (WABI)
- KEGG API
- PDBj SOAP
 - 全て独自に開発実装、特に標準化はされていない
 - 使われているデータ型もバラバラ
 - サービスの命名規則もバラバラ
 - 非同期通信などの仕様もバラバラ
- BioMoby
 - 植物ゲノム関係などを中心に数十サイトが対応
 - 共通のデータ型（ただし独自XML）
 - スペインの MOWServ などグリッド化の実例



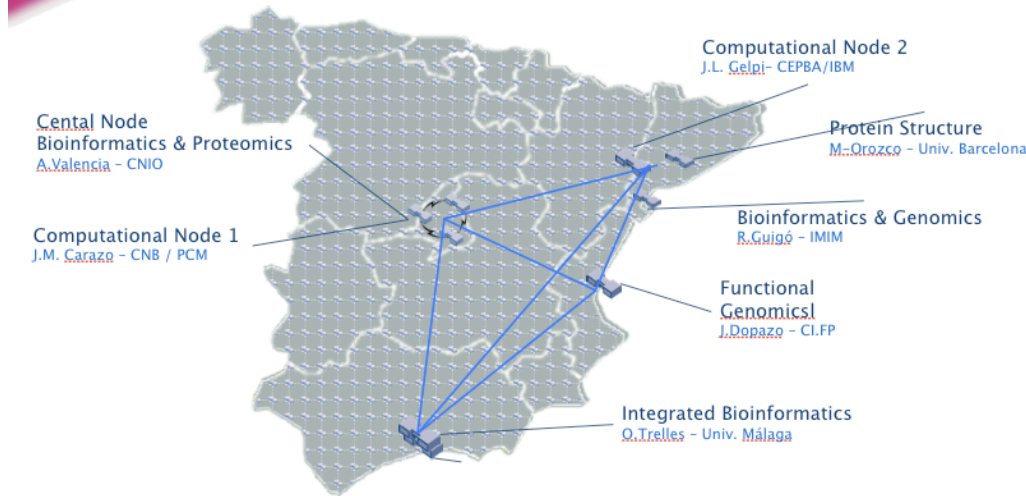
```
<Object namespace="GenBank/AC"
ID="AY070397.1"/>
```

```
<Sequence namespace="GenBank/AC"
ID="AY070397.1"/>
<Length>960</Length>
<SequenceString>
  acaaaaagattaaacaagagag...
</SequenceString>
</Sequence>
```

スペインの MOWServ による統合例



枠組みとして BioMoby を採用
 スペイン国内の BI センターをグリッド化
 データ型とサービスのオントロジー
 ワークフロー作成
 ジョブ管理



Scalable & expandable “Virtual” Organization

Integrated Bioinformatics, INB-UMA

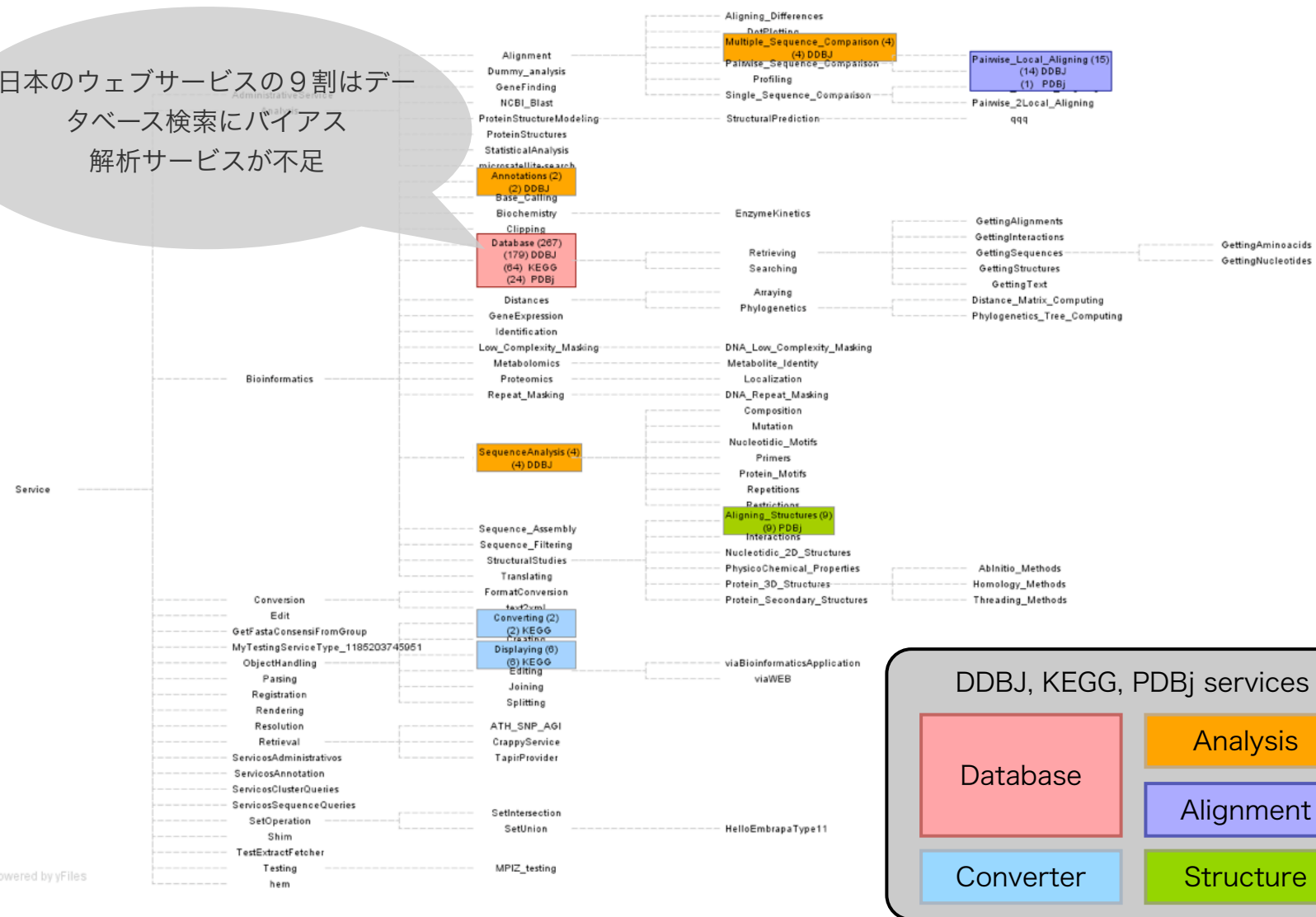
O.Trelles, BioHackaton-Jap

Slides by Prof. Trelles (INB/MOWServ)

Intelligent client for integrating bioinformatics services”, Navas-Delgado, et al. Bioinformatics, vol.22 no.1 2006 pages 106-111


BioMoby service ontology

日本のウェブサービスの9割はデータベース検索にバイアス
解析サービスが不足



BioHackathon 2008 で残された課題

- 現状の様々な問題点が提示された
 - Standard data types and workflow
 - Large data
 - Async service
 - Security
 - Error handling
 -
- Publication
 - Manifesto 標準化の指標
 - Journal article 成果のレビュー
- 継続的な会議の開催ができれば
 - BOSC session for web service
 - annual BioHackathon

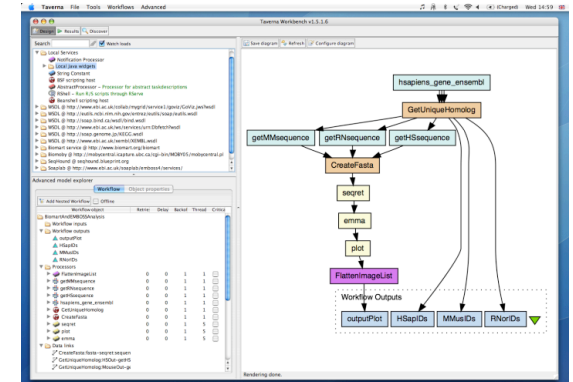


Open questions
Short term

- **Asynchronous services (& progress status notification)**
- **Error handling**
- **Replication (mirroring) system**
- **Daily Service quality control**
- **Service's help system**
- **Advanced service discovering**
- **Collections**
- **Large data sets (network overloading)**
- **Indirect (by reference) pass of data (locality of D & S)**
- **Workflows (storage in the repository, WWE-jmf)**
- **Semantic annotations (datatypes & Services)**
- **Define services categories**
- **New datatypes: Gene expression; images; ...**

相互運用性のあるワークフローの構築

- DDBJ/PDBj/KEGG を利用するワークフローの構築
 - Taverna を利用
 - 条件分岐が出来ない
 - 連携するためのプログラミングが必要
 - 現状の WS では解析のための機能が不足
- それぞれのサービスで何が可能か、何が不足しているか
 - 必要とされているワークフローの洗い出しと実装
 - 定期的なミーティングによる統合のための協力
 - 各サービスの品質向上
 - DDBJing, PDBjing, KEGGing などの講習会



統合にはほど遠く

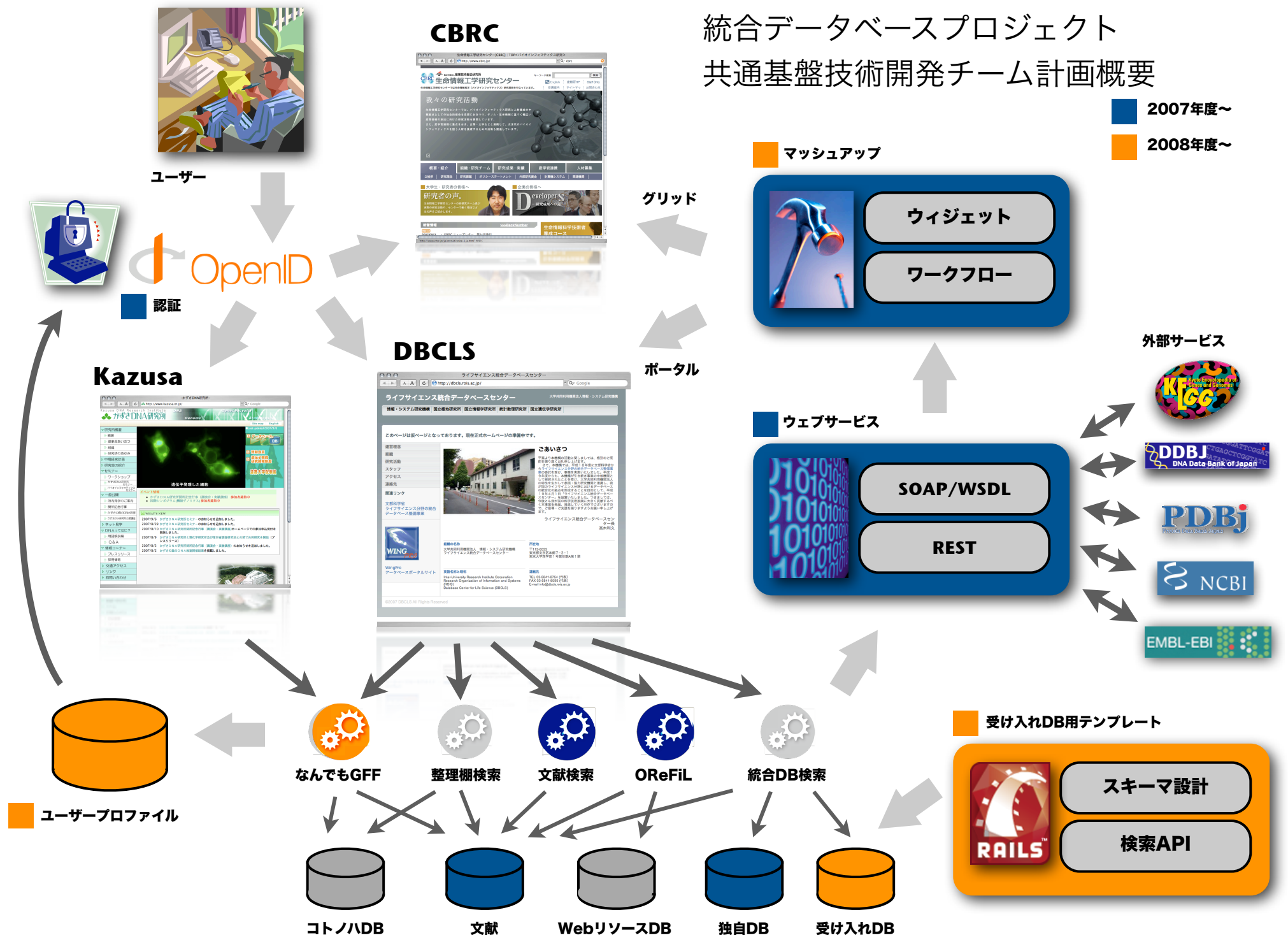
- ・ データ型と解析ソフトウェアの数は増加し続けている
 - ・ それらに我々の標準化提案を強制することは現実的には困難
- ・ ウェブサービスにはスケールしない部分がある
 - ・ Large data - 生物学データの量は指数関数的に増加中
 - ・ Async service - 計算機資源はサーバに依存し限られている
 - ・ Workflow - プログラミングなしでパイプライン化することは困難
- ・ 反省と課題
 - ・ 個別技術要素の問題点は洗い出せたが、グローバルな結論はでていない
 - ・ 個々の技術についてはグリッドなどでもすでに議論されていた点が多い
 - ・ データやサービスの統一化には強力なイニシアチブと人員が必要
 - ・ 全体の分かる人が long term な目的を持って方向性を決めるべき
 - ・ 今のインフラは圧倒的に力不足→サイエンスのできるウェブサービスへ

今後の統合にむけて何が可能か

- ・ 相互運用性の実現には継続的な国際コミュニティの努力が必要
 - ・ 主要なサービスプロバイダは相互運用性の高い API を提供すべき
 - ・ EBI, NCBI, DDBJ, KEGG, PDBj, CBRC, ...
 - ・ BioMoby
- ・ 非標準のサービスもクライアント側のソフトウェアで統合可能
 - ・ Open Bio* ライブラリ (BioPerl, BioRuby, BioJava, BioPython)
 - ・ Taverna, Seahawk, MOWServ, jORCA, ...
- ・ 日本では DBCLS において国内のサービスを統合する TogoWS を始動
 - ・ データベースやソフトウェア出力のパパーザやコンバータなどを提供
 - ・ BioMoby に対して WS-I 準拠のプロキシサーバを提供
 - ・ 国内外のサービスを統合したワークフローを集積、ドキュメント化

統合データベースプロジェクト 共通基盤技術開発チーム計画概要

■ 2007年度～
■ 2008年度～



TogoDB - 誰でもデータベースを作製・公開できるサイトとツール



- ・ 表形式のデータを登録
 - ・ 1分で公開完了
 - ・ 並べ替え
 - ・ 絞り込み検索
 - ・ ウェブ上で編集
 - ・ ダウンロード可能
- ・ QuickLook 機能搭載
- ・ OpenID 対応
- ・ ウェブサービス対応
- ・ オープンソース
- ・ フリーで安定運用

<http://togodb.dbcls.jp/>

TogoWS - 国内外のウェブサービスを統合し透過的に利用



• SOAP/WSDL 統合

- DDBJ
- KEGG
- PDBj
- CBRC
- 統合DB
- BioMOBY



• REST 統合

- NCBI
- EBI
- BioRuby内蔵



• ワークフロー化



<http://togows.dbcls.jp/>

TogoDB and TogoWS

How to make your DB accessible via WS?

The screenshot shows the TogoDB web interface. At the top, there is a search bar and a 'Search' button. Below it is a table with columns: SEQID, Frame, ORF, Length, TargetID, AIN_Length, E-value, Score, and Accession. The table contains several rows of data. Below the table, there is a detailed view of a specific entry (FC02BD09) showing its sequence, domain structure, and various annotations.

SEQID	Frame	ORF	Length	TargetID	AIN_Length	E-value	Score	Accession	
FC02BD09	1	ORF2	147	chr08_finished_C08HBa0197317.1.v1.seq_C08HBa0197317.1	416	0	825.0	AK224609	Show
FC02BD09	1	ORF2	147	chr08_finished_C08HBa0197317.1.v1.seq_C08HBa0197317.1	229	2E-127	454.0	AK224609	Show
FC02BD09	1	ORF2	147	chr08_finished_C08HBa0197317.1.v1.seq_C08HBa0197317.1	140	1E-39	163.0	AK224609	Show
FC02BH09	1	ORF1	358	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	438	5E-93	341.0	AK224612	Show
FC02BH09	1	ORF1	358	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	343	6E-37	155.0	AK224612	Show
FC02BH09	1	ORF1	358	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	343	6E-37	155.0	AK224612	Show
FC02BH09	1	ORF1	358	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	343	6E-37	155.0	AK224612	Show
FC03AD05	1	ORF1	102	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	102	1E-39	163.0	AK224609	Show
FC03AD05	1	ORF1	102	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	102	1E-39	163.0	AK224609	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-39	163.0	AK224609	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-39	163.0	AK224609	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-39	163.0	AK224609	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-39	163.0	AK224609	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-39	163.0	AK224609	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-39	163.0	AK224609	Show

<http://togodb.dbcls.jp/>

The screenshot shows the TogoWS web interface. At the top, there is a navigation bar with 'Home', 'REST', 'SOAP', and 'Status' links. Below it is a banner with the text 'Integration of the bioinformatics web services' and a URL. The main content area is divided into two columns: 'Related resources' and 'Summary of TogoWS service'. The 'Related resources' section lists several databases and services. The 'Summary of TogoWS service' section provides a detailed overview of the service and its capabilities.

Related resources

- Status log of Japanese web services
- Database of sample codes for Japanese web services (for Perl, Python, Ruby and Java)
- Database of methods provided by EBI, NCBI, and Japanese web services (in Japanese)
- Database of documentations provided by EBI, NCBI, and Japanese web services (in Japanese)

Summary of TogoWS service

Towards an integrated web service in life science

To date, various databases have been made available through web services provided by major bioinformatics centers including NCBI, EBI, DDBJ, KEGG, PDBJ, and CBRC. These web services are helpful especially for the high throughput analysis, however, there are no consistent way to utilize them uniformly. Our new web service, TogoWS, is designed to address several major issues for the improved usability and seamless integration.

- (1) SOAP based web services are ideally follow the open standard and they should be independent from computer languages, however, many services still require language specific hacks in practice. TogoWS proxies these services to make them available in any programming language without difficulty.
- (2) Query mechanism and syntax differs service to service that requires user to learn each usage beforehand. TogoWS provides a simple REST interface to query and retrieve in a unified manner.
- (3) Entries obtained from databases are needed to be parsed by the client program to be fully utilized. TogoWS embeds open source bioinformatics libraries like BioPerl and BioRuby (which we also developed) to provide functionalities for parsing and conversion of various biological data formats without any installation on user side.

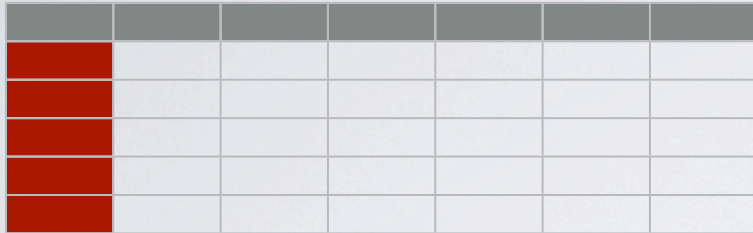
We, Database Center for Life Science (DBCLS), add a practical value which enable users to access and utilize any major biological databases without any additional software to create truly interoperable workflows with web services.

- REST service: Interoperable REST interface
- SOAP service: Integrated Japanese SOAP web services

Part of the concept of this work was emerged from the BioHackathon 2008, the International workshop entitled "Towards integrated web service in life science with Open Bio* libraries."

<http://togows.dbcls.jp/>

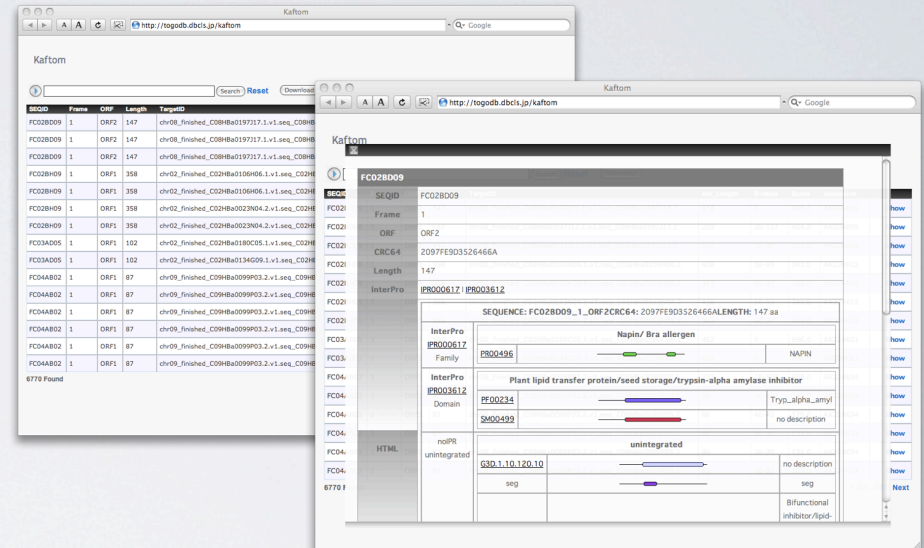
TogoDB: deploy your DB in 5min



Your DB in CSV format



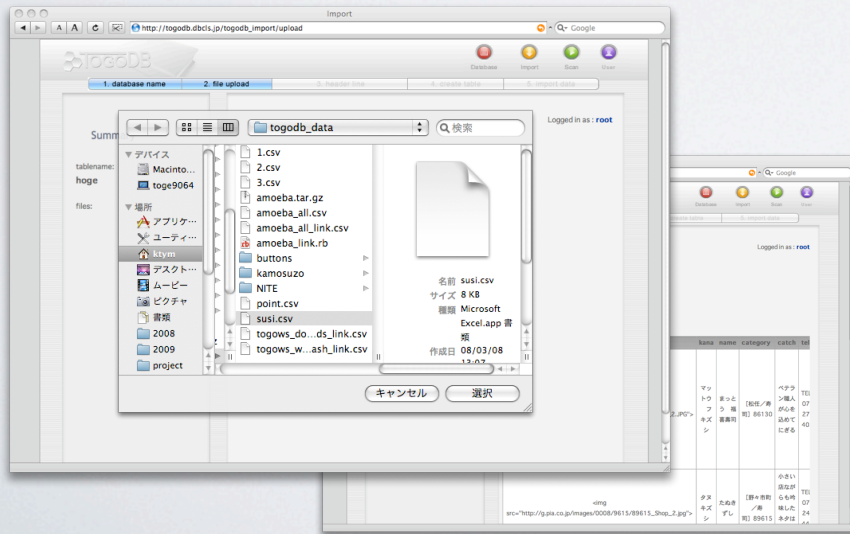
upload



The top screenshot shows a table of records with columns: SEQID, Frame, ORF, Length, TargetID. The bottom screenshot shows a detailed view for record FC02BD09, including fields like Frame (1), ORF (ORF2), Length (147), and InterPro annotations such as Napin/ Bra allergen and Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor.



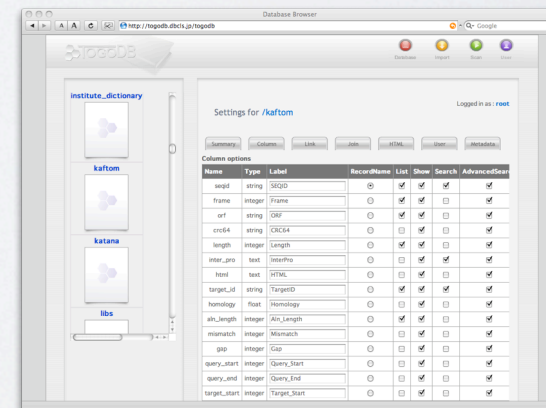
all set



The 'Import' screen shows a file upload interface with a file browser displaying various CSV files. The 'Database Browser' screen shows a table with columns: kuka_name, category, catch, etc.

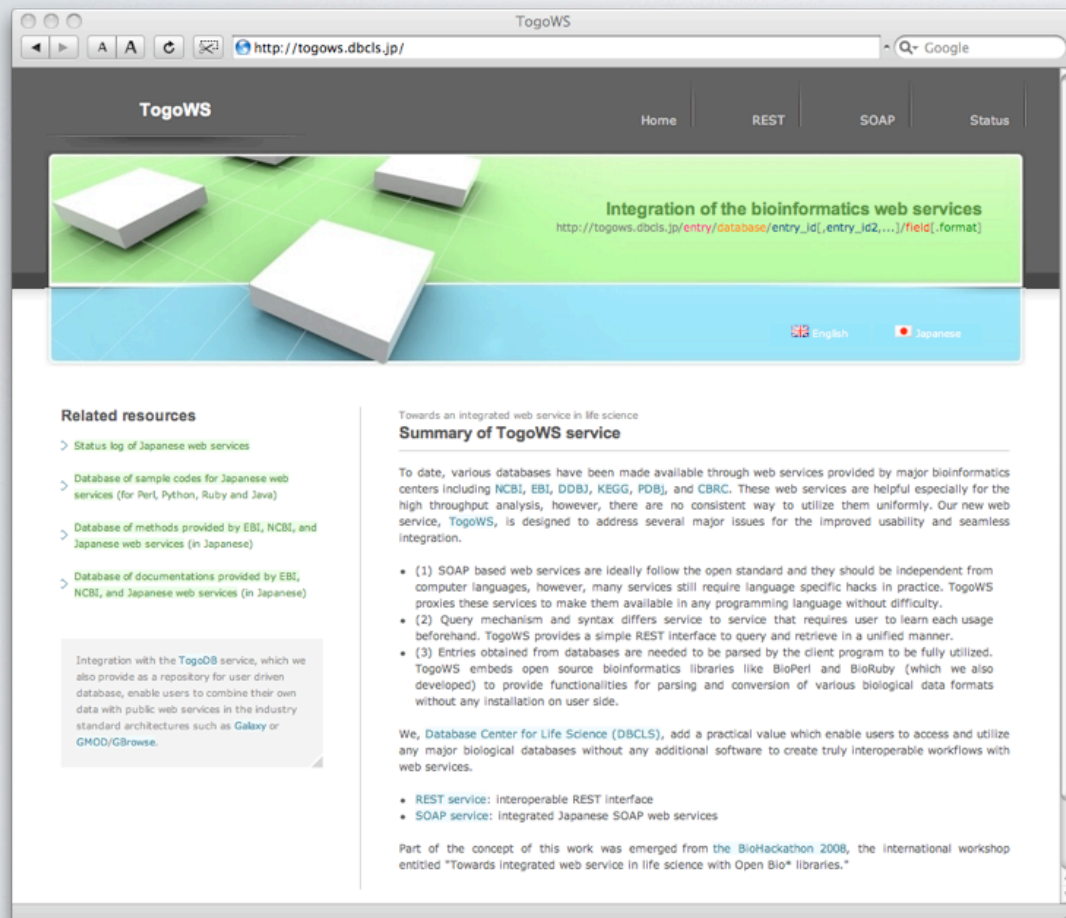


configure



The 'Settings for /kaftom' page shows a table of column options with columns: Name, Type, Label, RecordName, List, Show, Search, AdvancedSearch. The table lists various fields like seqid, frame, orf, length, etc., with checkboxes for each option.

TogoWS: integrated access for WSs



- SOAP/WSDL

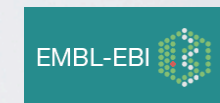
- DDBJ
- KEGG
- PDBj
- CBRC
- BioMOBY

- REST

- NCBI
- EBI
- TogoDB

- Embedded Parsers

- BioRuby
- BioPerl



<http://togows.dbcls.jp/entry/ddbj/AB000100>
<http://togows.dbcls.jp/entry/ddbj/AB000100.xml>
<http://togows.dbcls.jp/entry/ddbj/AB000100.gff>
<http://togows.dbcls.jp/entry/ddbj/AB000100/definition>
<http://togows.dbcls.jp/entry/ddbj/AB000100/organism>



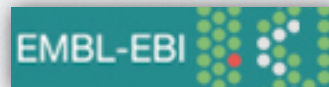
REST



TogoWS + BioRuby



SOAP/WSDL OR REST



Standardization of the REST API

```
/entry/database/entry_id[,entry_id2,...][[/field][.format]  
http://togows.dbcls.jp/entry/uniprot/A1AG1_HUMAN,A1AG1_MOUSE  
http://togows.dbcls.jp/entry/uniprot/A1AG1_HUMAN,A1AG1_MOUSE.fasta  
http://togows.dbcls.jp/entry/uniprot/A1AG1_HUMAN,A1AG1_MOUSE.gff  
http://togows.dbcls.jp/entry/pubmed/16381885  
http://togows.dbcls.jp/entry/pubmed/16381885/authors  
http://togows.dbcls.jp/entry/pubmed/16381885/authors.json  
database:  
  KEGG: gene, orthology, enzyme, compound, drug, glycan, reaction  
  DDBJ: ddbj, dad  
  PDBj: pdb  
  NCBI: gene, genome, genomeprj, geo, journals, mesh, nucleotide, omim, pmc,  
        protein, pubmed, taxonomy, cdd, popset, snp, unigene, homologene,  
        nuccore, nucest, nucgss, unists  
  EBI: biomodels, chebi, ensembl, go, interpro, reactome, uniprot, uniparc,  
        uniref100, uniref90, uniref50, msdchem, msdpdb  
format:  
  xml, json, gff, fasta etc.  
/search/database/query+string[/offset,limit][.format]  
http://togows.dbcls.jp/search/uniprot/lung+cancer/count  
http://togows.dbcls.jp/search/uniprot/lung+cancer/1,10  
/convert/source.format  
http://togows.dbcls.jp/convert/blast.gff  
source: genbank, blast, fasta, psl, sim4, hmmer, exonerate, wise etc.
```

TogoDB and TogoWS

Instantly publish your DB in a standardized way!

Kaftom

Search

SEQID	Frame	ORF	Length	TargetID	Alt_Length	E-value	Score	Accession	
FC02BD09	1	ORF2	147	chr08_finished_C08HBa0197317.1.v1.seq_C08HBa0197317.1	416	0	825.0	AK224609	Show
FC02BD09	1	ORF2	147	chr08_finished_C08HBa0197317.1.v1.seq_C08HBa0197317.1	228	2E-127	454.8	AK224609	Show
FC02BD09	1	ORF2	147	chr08_finished_C08HBa0197317.1.v1.seq_C08HBa0197317.1	140	1E-59	163.0	AK224609	Show
FC02BH09	1	ORF1	358	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	438	5E-93	341.0	AK224612	Show
FC02BH09	1	ORF1	358	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	343	6E-37	155.0	AK224612	Show
FC02BH09	1	ORF1	358	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	358	1E-100	163.0	AK224612	Show
FC02BH09	1	ORF1	358	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	358	1E-100	163.0	AK224612	Show
FC03AD05	1	ORF1	102	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	102	1E-100	163.0	AK224612	Show
FC03AD05	1	ORF1	102	chr02_finished_C02HBa0106H06.1.v1.seq_C02HBa0106H06.1	102	1E-100	163.0	AK224612	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-100	163.0	AK224612	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-100	163.0	AK224612	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-100	163.0	AK224612	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-100	163.0	AK224612	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-100	163.0	AK224612	Show
FC04AB02	1	ORF1	87	chr09_finished_C09HBa0106H06.1.v1.seq_C09HBa0106H06.1	87	1E-100	163.0	AK224612	Show

6770 Found

Any tables by the user

TogoWS

Home REST SOAP Status

Integration of the bioinformatics web services

Can be accessed with RESTful APIs

Related resources

- Status log
- Database of sample codes for Japanese web services (for Perl, Python, Ruby and Java)
- Database of methods provided by EBI, NCBI, and Japanese web services (in Japanese)
- Database of documentations provided by EBI, NCBI, and Japanese web services (in Japanese)

Integration with the TogoWS service, which we also provide as a repository for our shared database, enable users to connect their own data with public web services in the existing standard architectures such as Galaxy or GMOD/GBrowse.

We, Database Center for Life Science (DBCLS), add a practical value which enable users to access and utilize any major biological databases without any additional software to create truly interoperable workflows with web services.

- REST service: Interoperable REST interface
- SOAP service: Integrated Japanese SOAP web services

Part of the concept of this work was emerged from the BioHackathon 2008, the International workshop entitled "Towards integrated web service in life science with Open Bio* libraries."

<http://togodb.dbcls.jp/foobardb/>

<http://togows.dbcls.jp/entry/togodb-foobardb/123.json>

Develop an integrated and reliable bioinformatics infrastructure:
Interoperable web services and database resources
Workflow platforms for genome annotation
Visualization of the large scale data sets

BIO HACKATHON 09



15-21 Mar. 2009
DBCLS Tokyo / OIST Okinawa, Japan

Welcome !!



<http://hackathon2.dbcls.jp>



15-21 Mar 2009 @ DBCLS (Tokyo), OIST (Okinawa)

BioHackathon 2009 Sponsors



Database Center for Life Science
Ministry of Education, ..., Japan (MEXT)



Marine Genomics Unit
Okinawa Institute of Science and Technology
Cabinet Office, Government of Japan

Primary Interests



Prof. Yoshiki Hotta (President of ROIS)



Prof. Sydney Brenner (President of OIST)



Infrastructure Unit
Takagi group

- Integration of DB resources
- Standardization
- Sustainability
- Interoperability of WSs



Marine Genomics Unit
Satoh group

- Genome sequencing
- Meta genomics
- Evolutionary development
- Preservation of coral seas

Topics of this year (BioHackathon 2009)

- New paradigm for the vast amount of data
 - Next generation sequencers
 - 1000 genomes
- Standardization and Interoperability
 - Open Bio* libraries and web services to **handle** ↑
- Platforms
 - Galaxy, ANNOTATOR, RIKEN LSA, BioMart, GMOD, BioMoby etc. to **integrate** ↑
- Visualization
 - Cytoscape (networks), Comparative/Mata genome, transcriptomes etc. to **understand** ↑





Symposium

9:00 Toshiaki Katayama

Introduction to the **BioHackathon** 2009 and integration efforts in DBCLS

9:20 James Taylor

Integrated reproducible analysis with **Galaxy**

9:40 Tobias Gattermayer

Biomolecular Function Discovery with the **ANNOTATOR**

10:00 Jessica Severin

Accelerating experimentation and biological discovery within **RIKEN** OSC LifeScienceAccelerator -- eeDB gLyphs

10:20 Arek Kasprzyk

BioMart and third party applications: Working towards better interoperability

10:40 Tom Oinn

The **Taverna 2** Workflow Engine API : Extensible, High Performance and Easy to Integrate

11:00 Oswaldo Trelles

Multi-repository architecture for **service integration**

11:20 Alberto Labarga

Integrating literature and annotation resources: the **Elsevier Grand Challenge** experience

11:40 Hideaki Sugawara

Looking back on a **WABI** life for tomorrow



Satellites

Use Cases (from biologist side)

Sequence analysis, Transcription regulation, Glycobiology, DDBJ-KEGG-PDBj workflow

Large Data Handling

How to handle large amount of data (SRA) produced by **NG sequencers**

Semantic Annotation

How to create semantically annotated Web Services for service discovery (**BioMoby**)

Galaxy, ANNOTATOR, Taverna

Clients for annotation, analysis and workflow

Visualization

Multi-repository architecture for **service integration**

Literature (text mining)

Text mining tools and development of web services.

Open Bio* software development

BioRuby, G-language etc.





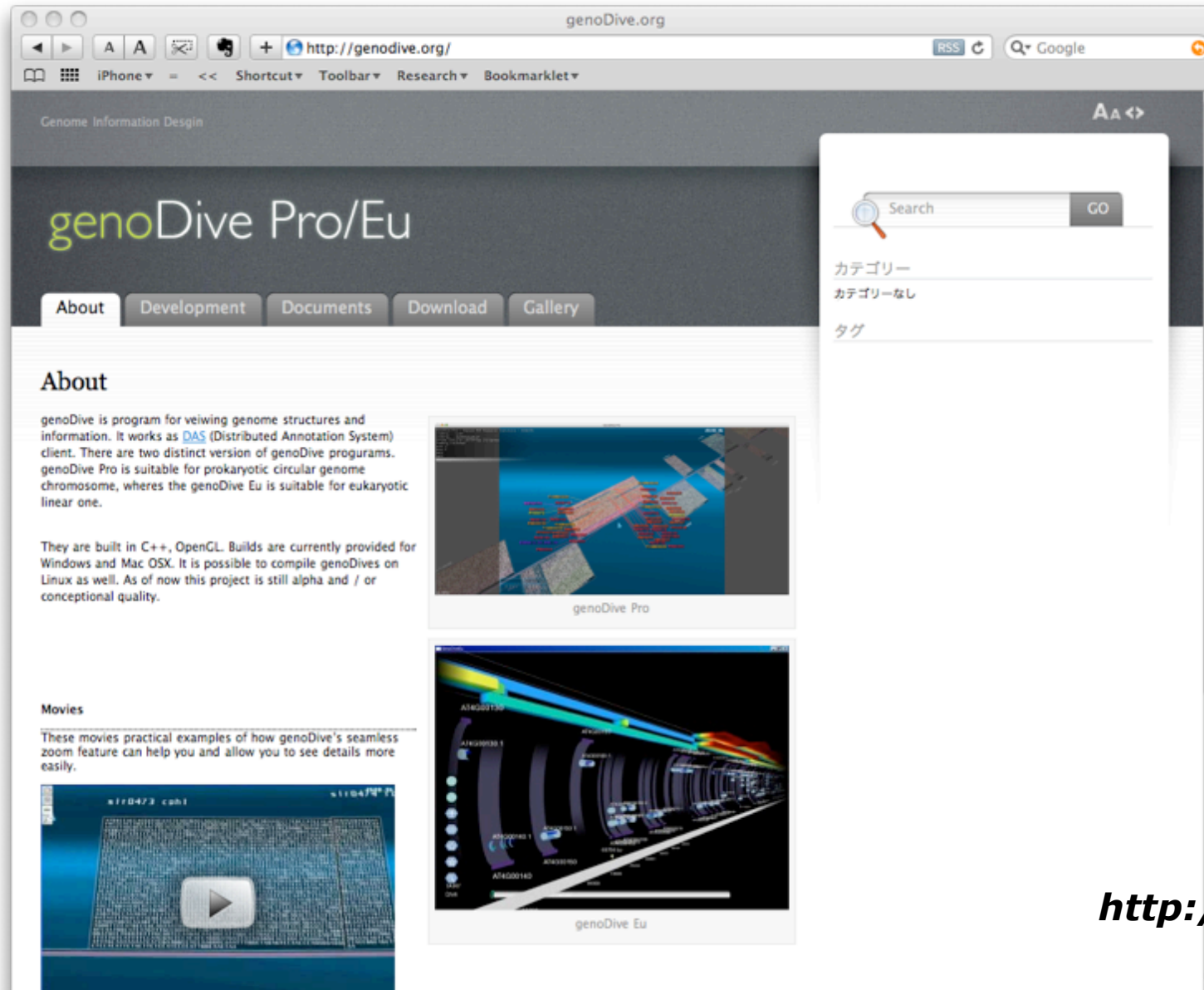


R SKC
 multi p. ...
 SUBJECT
 PREDICATE
 OBJECT
 Why:
 Technically: get other specialists' info, need
 not NLP, computational meaning
 Philosophically
 Pipelines
 AI, smart agents, grants, in science you need to clearly c
 (flexibility for future) collaboration, reproducibility, pre
 usage: less hardcoding
 State of the practice - ambiguity of data structures → ontologies (e.g. NESCENT)

Visualization
 Network
 - cytoscape - eeOB3
 - ecell 3D
 - Graphviz
 - Biocichlid
 - interaction with data
 - 1, 2, > 2
 Expression
 - cytoscape - bluejay
 - GeneSpring - Biocichlid

"Semantics" / Web Services & Interoperability
 Why:
 Technically: get other specialists' info, need
 not NLP, computational meaning
 Philosophically
 Pipelines
 AI, smart agents, grants, in science you need to clearly c
 (flexibility for future) collaboration, reproducibility, pre
 usage: less hardcoding
 State of the practice - ambiguity of data structures → ontologies (e.g. NESCENT)
 HTTP, XML, learn their protocol, OBO, Web Services & Pipelines
 State of the art
 OWL, RDF Reasoners, Rules, Triple Stores
 Who:
 RDF schema not recommended

GenoDive - 3Dゲノムブラウザ



Genome Information Desgin

genoDive Pro/Eu

About Development Documents Download Gallery

About

genoDive is program for veiwing genome structures and information. It works as [DAS](#) (Distributed Annotation System) client. There are two distinct version of genoDive programs. genoDive Pro is suitable for prokaryotic circular genome chromosome, wheres the genoDive Eu is suitable for eukaryotic linear one.

They are built in C++, OpenGL. Builds are currently provided for Windows and Mac OSX. It is possible to compile genoDives on Linux as well. As of now this project is still alpha and / or conceptional quality.

Movies

These movies practical examples of how genoDive's seamless zoom feature can help you and allow you to see details more easily.

genoDive Pro

genoDive Eu

Search GO

カテゴリー
カテゴリーなし

タグ

<http://genodive.org/>

Prokaryote list

Anabaena	TAIR7_das
BTA11	TAIR7m
Bradyrhizobium	Thermo
CC9311	WHB102
Chlorobium	
Et11	
Gloeobacter	
Leptominosarum	
MED4	
MIT9313	
Mesorhizobium	
NGR234	
ORS278	
ORS571	
PCC6301	
R7A	
Rhodopseudomonas	
55120	
Sinorhizobium	
Synechocystis	



情報処理学会誌 9月号 (バイオ小特集)

小特集

生命情報学が直面する 大規模ゲノムデータ 時代の課題

- 1 分散データの統合とセマンティック Web
- 2 巨大データの扱いと解析
- 3 生命科学分野におけるテキストマイニング
- 4 生命科学分野・ゲノムデータの可視化技術

情報処理 Vol.50 No.9 Sep. 2009 833

小特集 生命情報学が直面する
大規模ゲノムデータ時代の課題

1 分散データの統合と セマンティック Web



中尾 光輝
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

片山 俊明
東京大学医科学研究所
ヒトゲノム解析センター

小特集 生命情報学が直面する
大規模ゲノムデータ時代の課題

2 巨大データの扱いと解析



小西 史一
東京工業大学

小特集 生命情報学が直面する
大規模ゲノムデータ時代の課題

3 生命科学分野における テキストマイニング



山本 泰智
ライフサイエンス統合データベースセンター

小特集 生命情報学が直面する
大規模ゲノムデータ時代の課題

4 生命科学分野・ゲノムデータの 可視化技術



西村 邦裕
東京大学大学院情報理工学系研究科