

# SYGD

User Manual Version 1.00

## 目次

1. 概要 .....	2
1-1. SYGDとは .....	2
1-2. SYGDの特徴 .....	2
2. SYGDチュートリアル .....	4
2-1. 解析画面共通の構成.....	4
2-2. 解析画面のトップページ（ゲノムリスト）の表示 .....	5
2-3. ゲノム情報の閲覧 .....	6
2-3-1. コンティグリストの表示 .....	10
2-3-2. コンティグマップの表示 .....	11
2-3-3. 遺伝子詳細情報の表示 .....	14
2-3-3-1. Locus（Locus 情報の表示） .....	14
2-3-3-2. Transcript（転写産物情報の表示） .....	15
2-3-3-3. CDS（CDS情報の表示） .....	16
2-3-3-4. Otholog（オーソログ情報の表示） .....	18
2-3-4. シンテニーマップの表示 .....	18
2-4. BLAST 検索 .....	22
2-5. ドットプロット解析 .....	23
2-6. 差分解析 .....	24
2-7. 基本検索 .....	25
2-7-1. コンティグ検索 .....	25
2-7-2. 遺伝子、RNA、CDS 検索 .....	27
2-7-3. オーソログ検索 .....	33
2-7-4. COG、KOG、EC 検索 .....	35

## 1. 概要

### 1-1. SYGD とは

SYGD (Sake Yeast Genome Database) は、ゲノム情報に関する膨大なデータや関連情報を容易に閲覧、解析できる統合プラットフォームです。Web ブラウザを用いて SYGD にアクセスするだけで、迅速かつ簡便にデータを閲覧、解析することが可能です。

### 1-2. SYGD の特徴

SYGD が持つ機能には、以下のようなものがあります：

#### 統計情報表示機能

総塩基数、コンティグ数、遺伝子数など、様々な統計データを表示します。この機能の詳細は、以下で説明します：

- ⇒ 2-2 ゲノムリスト（解析画面のトップページ）の表示
- ⇒ 2-3 ゲノム情報の閲覧

#### ゲノム情報の表示機能（ゲノムビューワー、アノテーションビューワー）

コンティグや染色体上に遺伝子地図を表示します。またそれぞれの遺伝子に対して、Locus情報、DNA配列、RNA配列、アミノ酸配列、エキソン・イントロン情報、EC number、オーソログ情報などの多彩な情報を表示します。

また配列情報はFASTA形式ファイルとして出力され、解析結果は Microsoft Excel 等の外部ソフトウェアに取り込んで解析を進めることも可能となります。

この機能の詳細は、以下で説明します：

- ⇒ 2-3 ゲノム情報の閲覧

#### BLAST 検索機能

任意の配列と類似性のあるものが存在するかを調べるためにBLAST検索機能を用いることができます。この機能の詳細は、以下で説明します：

- ⇒ 2-4 BLAST 検索

#### ドットプロット解析機能

ゲノム、染色体、あるいはコンティグ間の類似性を全体的にみることができます。この機能の詳細は、以下で説明します：

- ⇒ 2-5 ドットプロット解析

### 差分解析機能

複数のゲノムグループ間で、共通に存在する遺伝子、あるいは特定のグループにのみ存在する遺伝子を表示します。

2グループ間解析に加えて、3グループ間解析も実施することができます。この機能の詳細は、以下で説明します：

⇒ 2-6 差分解析

### 基本検索機能

コンティグ、遺伝子、そしてオーソログなどの様々な観点からデータを高速に検索することができます。

基本検索機能では、大きく分けて以下の検索を行なうことができます：

1. コンティグ検索 (解説：2-7-1 コンティグ検索)
2. 遺伝子, RNA, CDS検索 (解説：2-7-2 遺伝子, RNA, CDS検索)
3. オーソログ検索 (解説：2-7-3 オーソログ検索)
4. COG, KOG, EC 検索 (解説：2-7-4 COG, KOG, EC 検索)

## 2. SYGD チュートリアル

### 2-1. 解析画面共通の構成

SYGD では各画面で共通のインターフェースを採用しています。解析メニューの選択や簡易遺伝子検索が画面上部の共通インターフェースから行えます。



図 1: 各解析画面上部の共通部分

A	基本検索	基本検索画面へのリンクになっています。 詳細は、2-7.節 基本検索 で解説します。
B	ゲノム選択	「2-3. ゲノム情報の閲覧」と同様に、ここからゲノムを選択することができます。
C	BLAST 検索リンク	BLAST 検索画面へのリンクになっています。 詳細は、2-4.節 BLAST 検索 で解説します。
D	ドットプロット解析 差分解析リンク	ドットプロット解析画面・差分解析画面へのリンクになっています。 ドットプロット解析の詳細は、2-5.節 ドットプロット解析 で解説します。 差分解析の詳細は、2-6.節差分解析解析 で解説します
E	Category Browser	各カテゴリーに分類されている遺伝子の一覧表を表示することができます
F	簡易 Full text サーチ	基本検索のgeneでのFull textサーチと同様の機能となっています
G	マニュアル参照	マニュアル(PDF形式)が表示されます。

## 2-2. 解析画面のトップページ（ゲノムリスト）の表示

所定のURLにブラウザでアクセスすると、以下の解析画面のトップページが表示されます（図2）。トップページでは、登録されているゲノムのリストが表示されます。

メインメニュー →

ゲノムリスト →

Organism	Acronym	Source	Taxon. id	Mb	Chromosome	Contigs	CDSs	tRNAs	rRNAs
Saccharomyces cerevisiae S288c	S288c	SGD	559292	12.1	16	17	6008	299	14
Saccharomyces cerevisiae Kyokai no. 7	K7	NRIB_SYGD	721032	11.8	16	17	5816	296	7
All	-	-	-	-	32	34	11824	595	21

図2: トップページ画面。ゲノムリストが表示されます。

図2 のゲノムリストについて以下に説明します：

Organism	SYGD に登録されているゲノムの名称を示します。 この名称は、主にゲノムの分類学上の種 (species)・サンプル名・解析名などで示されます。
Acronym	ゲノムの略称を示します。 またゲノムトップ画面 (図3) へのリンクにもなっています。
Source	ゲノム配列のデータソースを示しています。
Taxon. id	登録されているゲノムのEntrez Taxonomy database上でのtaxonomy ID を示します。 また NCBI taxonomy browser へのリンクにもなっています。
Mb	"Organism"で示された生物種のゲノムサイズを Mega base pair 単位で示します。
Chromosome	"Organism"で示された生物種のゲノムの染色体数を示します。
Contig	それぞれのゲノムに登録されているコンティグ数を示します。 またコンティグ検索結果表示画面へのリンクにもなっています (2-7-1 節 or 図23 参照)。
CDSs	それぞれのゲノムに登録されている CDS 数を示します。 また CDS 検索結果表示画面へのリンクにもなっています (2-7-2 節 or 図27 参照)。
tRNAs	それぞれのゲノムに登録されている tRNA 数を示します。 また RNA 検索結果表示画面へのリンクにもなっています (2-7-2 節 or 図26 参照)。
rRNAs	それぞれのゲノムに登録されている rRNA 数を示します。 また RNA 検索結果表示画面へのリンクにもなっています (2-7-2 節 or 図26 参照)。

## 2-3. ゲノム情報の閲覧

トップ画面ゲノムリストの "Acronym" 欄 (図2) のリンクをクリックしますと、ゲノムトップ画面が表示されます (図3)。

ゲノムトップ画面ではSYGDに登録しているゲノムのうち、クリックした "Acronym" に対応するゲノムについて、より詳細な情報が表示されます。

The screenshot shows the genome browser for *Saccharomyces cerevisiae* S288c. The header (A) displays the genome name and taxonomy ID. The left panel (B) shows a chromosome map with labels for Chr. I through Chr. XVI and Mitochondria. The central panel (C) contains a search interface with a 'Text Search To' dropdown, a 'Full text' search box, a 'BLAST Search' section with a 'FASTA Sequence' input and 'BLAST'/'Reset' buttons, and a 'Category Browser' with options for KOG, PRINTS, EC, and COG. The right panel (D) displays genome statistics: Genome size (12.1Mb), Chromosomes (16), Contigs (17), CDSs (6008), tRNAs (299), and rRNAs (14).

図 3: ゲノムトップ画面.

ゲノムトップ画面は以下の 4 つで構成されています。

- A: 選択したゲノムの名称などの情報を示します。
- B: 選択したゲノムまたはコンティグの情報を図で示します。
- C: 検索画面。選択したゲノムに対する様々な検索を行います。
- D: 選択したゲノムに関する統計情報 (ゲノムサイズ、遺伝子数など) を示します。

### A. ゲノムの名称

選択したゲノムの名称、略称 (Acronym)、taxonomy IDなどの情報を示します。

"tax\_id"はEntrez Taxonomy database上でのtaxonomy ID を示し、NCBI taxonomy browserへのリンクにもなっています。

### B. 染色体情報

選択されたゲノムに染色体に関する情報がある場合には、染色体情報が示されます。

染色体毎に、2つのリンクがあります。Chr.番号のリンクをクリックすると、この染色体に含まれるコンティグを示すコンティグリスト画面を見ることができます。Chr.番号の下のバーをクリックすると、クリックした位置を中心としたコンティグマップ画面が現れます。

染色体に関する情報が無い場合にはコンティグ情報が表示されます。この場合には、コンティグ番号をクリックすると、コンティグ全体のコンティグマップが、下のバーをクリックするとクリックした位置を中心としたコンティグマップが現れます。

### C. 検索画面

選択されたゲノムに対する検索インターフェースを提供いたします。

より厳密な検索や横断的な検索を実施したい場合には、BLAST検索(2-4参照)や基本検索(2-7参照)をご利用してください。

検索インターフェースには Text Search, BLAST Search, Category Browser の3つがあります：

#### ● Text Search

選択されたゲノムに存在する遺伝子に対して、テキスト検索を行います。

検索結果の解説については、2-7-2 遺伝子、RNA、CDS検索 を参照してください。

検索フィールドとして、以下の7種類があります：

Full Text	登録されている遺伝子に対して全文検索を行います。
Gene ID	登録されている遺伝子に対して Gene ID で検索を行います。
Description	登録されている遺伝子に対してDescription情報で検索を行います。例えば、"enzyme" と入力すると、gene description 欄に "enzyme" という文字列が含まれる遺伝子を検索します。
COG ID	登録されている遺伝子に対して COG (オーソログ) 情報の ID で検索を行います。
KOG ID	登録されている遺伝子に対して KOG (オーソログ) 情報の ID で検索を行います。
EC Number	登録されている遺伝子に対して EC Number 情報で検索を行います。
PRINTS	登録されている遺伝子に対して PRINTS (タンパク質モチーフ) 情報で検索を行います。

#### ● BLAST Search

FASTA形式の配列を用いて、選択されたゲノムに対してBLAST による相同性解析を行ないます。

検索結果の解説については、2-4 BLAST検索を参照してください。

"Select BLAST DB"のセレクトボックスから任意のデータセットを選択し、"FASTA Sequence"欄へFASTA形式の配列を入力して "BLAST" ボタンをクリックすることでBLAST検索が行なえます。

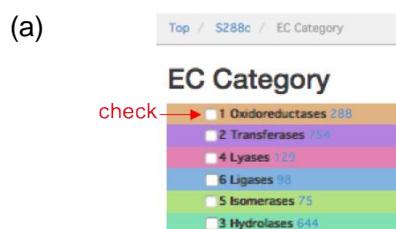
セレクトボックスから選択できるデータセットは、以下の通りです：

Genome	全コンティグ配列
All Genes	全遺伝子配列
All Transcripts	全転写産物 (RNA) 配列
All CDS	全 CDS 核酸配列
All Proteins	全アミノ酸配列
Intergenic Regions	遺伝子間領域
Upstream 2kb from all genes	全遺伝子の上流側 2kb の配列

### ● Category Browser

遺伝子には、COG, PRINTS, EC number 等でアノテーションされているものがあります。

これらでは "RNA processing and modification", "Transcription" 等のカテゴリーで遺伝子を分類しています。Category Browser では、これらカテゴリーの階層を表示するとともに、各カテゴリーに分類されている遺伝子の一覧表を表示することができます (図 4)。



(c)

Top / Search (CDS)

**Search**

Select Target: Contig Gene RNA **CDS** Ortholog COG KOG EC

Search

CDS Search :  Show Filter Option

CDS list 1 - 50 of 108 Rows: 50 Page: 1 Top Bottom Page: 1, 2, 3

No./108	Gene ID	Original description	Assigned description	Genome	Chromosome	Links	Protein sequence
1	YLR070C	D-xylulose reductase X YL2	Sorbitol dehydrogenase	Saccharomyces cerevisiae S288c (S288c)	Chr. XII	SGD:S000004060 NCBI:GeneID:850759	>YLR070C NP_013171.1 NC_001144.5:274141-275211:MTDLTTQEAIVLERPGKITLTVNSIPKISDPNEVIQIKATGICG:SSGIVLALIGENVKTLKVGDRVALEPGIPORFSPMEKGRYNDIKLPDQVDFEFGALIEPLVAIHANKLAKIKFGARCVVFGAGPIGTARQFGATHVNSGDLPHGVTVDVSVIKKAIKGGKADVVFECSCIQFPISIIPTKELTFQGCFRYQCQDYSDSIELVSSRKLKLPFIT
2	YLR100W	3-keto-steroid reductase	3-keto sterol reductase	Saccharomyces cerevisiae S288c (S288c)	Chr. XII	SGD:S000004090 NCBI:GeneID:850790	>YLR100W NP_013201.1 NC_001144.5:341810-342853:MNRKVAIVTGTNSNLGLNVFLIETEDTNVRLTVVTSRTPRVYLLVDFTNMVSVLNAVYDINKYRAINYLFVNAAGIFDGIQDW:KSKDQMGGLIFQANWFGYFYSIKLPQLTRGKAYVWISSIMSGLLHLATYKOLKLGIDINQWVQGIFFSHSFSYELNFFTYFGMLQVTRLANPNFEKQOVKYGATSARDGMPYIKTQEIDPTGMSDVF
3	YLR174W	isocitrate dehydrogenase (NADP(+)) IDP2	NADP-dependent isocitrate dehydrogenase	Saccharomyces cerevisiae S288c (S288c)	Chr. XII	SGD:S000004164 NCBI:GeneID:850871	>YLR174W NP_013275.1 NC_001144.5:504591-505829:MTKIKVANPVEVDGDEQTRIIWHLIRDKLVPLVDLKYDYDIYKVCATITPPEARVEEFLHKMWSKPNGTIRNLGGTVREPIII:ATDVIPEEGELRLVYKSKSGTHDVOLKVFYDPEHGGVAMMMTKNTLKKYDGGKFDVFEAMYARYSEKFEFSLGIWYEHRLDDI:VAQGFGLGLMTSVLITPDGKTFESEAAHGTVTRHFRQHQQG

図 4: Category Browser. ここではEC categoryの例を示します。図3のEC categoryをクリックすると、EC numberに関する最上層のカテゴリリストが表示されます(a)。EC numberでは、最初に6つのカテゴリ(Oxidoreductases, Transferases, Hydrolases, Lyases, Isomerases, Ligases)が示されています。各々のチェックボックスをチェックすると、より細かく分類されたカテゴリが表示されます(b)。各々のカテゴリの右側にあるリンク付き数値は、そのカテゴリに属するCDS数を示しています。リンクをクリックすると、そのカテゴリのCDSの一覧表を示します(c)。

#### D. 統計情報

選択されたゲノムの統計情報が示されます。以下の情報が示されます：

- Genome size
- Chromosomes
- Contigs
- CDSs
- tRNAs
- rRNAs

また Summary statistics をクリックすることで、このゲノムに対する染色体毎の統計情報が示されます (図 5)。

Top / S288c / Statistics

### Saccharomyces cerevisiae S288c

Chr.	Size(Mbp)	Length	Contigs	CDS	tRNA	rRNA	Map
Chr. I	0.2	230,218	1	94	4	-	
Chr. II	0.8	813,184	1	415	13	-	
Chr. III	0.3	316,620	1	168	10	-	
Chr. IV	1.5	1,531,933	1	767	28	-	
Chr. V	0.6	576,874	1	288	20	-	
Chr. VI	0.3	270,161	1	128	10	-	
Chr. VII	1.1	1,090,940	1	539	36	-	
Chr. VIII	0.6	562,643	1	290	11	-	
Chr. IX	0.4	439,888	1	215	10	-	
Chr. X	0.7	745,751	1	362	24	-	
Chr. XI	0.7	666,816	1	317	16	-	
Chr. XII	1.1	1,078,177	1	519	21	12	
Chr. XIII	0.9	924,431	1	469	21	-	
Chr. XIV	0.8	784,333	1	398	14	-	
Chr. XV	1.1	1,091,291	1	548	20	-	
Chr. XVI	0.9	948,066	1	472	17	-	
Mitochondria	-	85,779	1	19	24	2	
All	12.1	12,157,105	17	6008	299	14	

図 5: 統計情報。染色体毎のサイズ、コンティグ数、CDS 数などを表示。

### 2-3-1. コンティグリストの表示

図5のChr.列のリンクをクリックしますと、コンティグ配列のリストが表示されます。真核生物で選択されたゲノムに染色体に関する情報がある場合、図3-BのChr番号のリンクをクリックしますと、1つの染色体に属するコンティグ配列のリストが表示されます。表示の例を図6に示します。

Top / S288c / Contig list

### Chr. I contig list

Contig	Strand	Position	Length[bp]	CDS	tRNA	rRNA
<a href="#">NC_001133.9</a>	1	1-230218	230,218	94	4	-
All	-	-	230,218	94	4	-

図 6: コンティグリスト画面。"Contig"欄のリンクをクリックしますと、そのコンティグマップ (図7) が表示されます。

## 2-3-2. コンティグマップの表示

図6 Contigリンク部分、あるいは図3-BのChr.番号の下のバーをクリックしますと、図7のようなコンティグマップ画面が表示されます。図は4つの領域: (1). 表示コントロールパネル (2) 染色体表示領域 (3). 表示範囲コントロール領域 (4) 遺伝子表示領域 分けることができます。

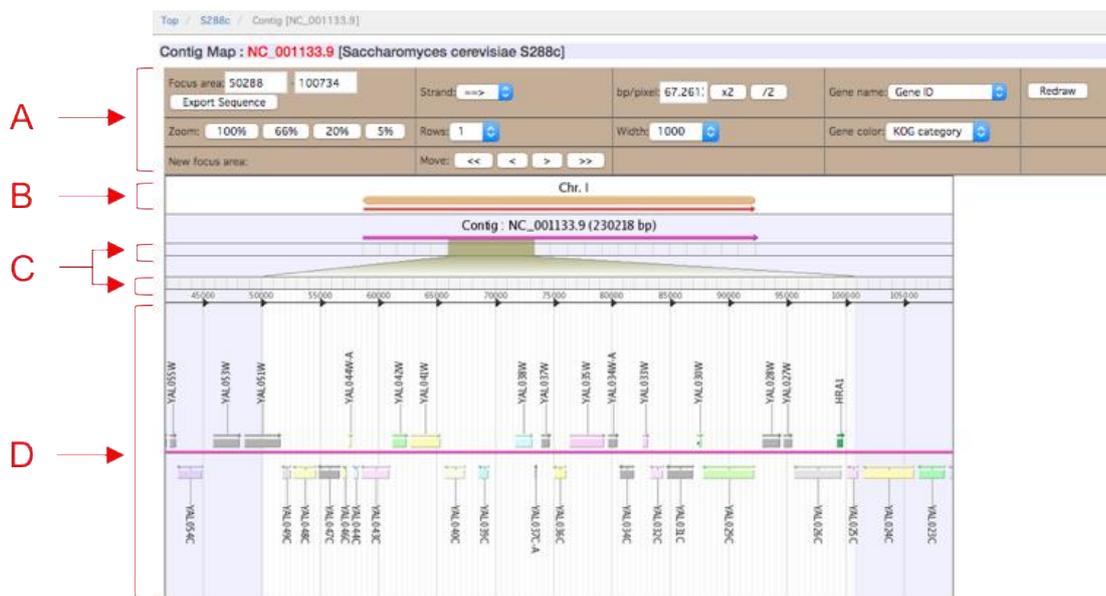


図 7: コンティグマップ画面。一つのコンティグ配列について、染色体、遺伝子へのマップ情報を図にしたもの。

コンティグマップ画面は以下の 4 つで構成されています。

- A: 表示コントロールパネル
- B: 染色体表示領域
- C: 表示範囲コントロール領域
- D: 遺伝子表示領域

### A: 表示コントロールパネル

このパネルのパラメータを変化させると、遺伝子表示領域のコンティグや遺伝子の表示を変えることができます。

以下、パネル内のパラメータを解説します：

Focus area	右にある二つのボックスにコンティグ上始点位置、終点位置を入力することで、表示範囲を指定できます。また,Focus area下にあるExport Sequenceをクリックすると、指定した範囲のコンティグ配列 (FASTA形式) がダウンロードできます。
------------	--

Strand	矢印の向きを逆向きにすると、ゲノム配列を逆向き (相補的な配列) として図が再描画されます。
bp / pixel	1 ピクセル毎のbase pair数を変えることができます。例えばこの値を半分にするると、図は2倍に拡大されます。
Gene name	遺伝子表示領域に表示される遺伝子名の表記方法を変えます。
100%, 66%, 20%, 5%	現在のFocus areaを、遺伝子表示領域図全体の何%の割合で表示するかを決めます。 例えば66%とすると、遺伝子表示領域図全体の約2/3がFocus areaとなります。
Rows	遺伝子表示領域の段数を指定します。遺伝子表示領域では、存在する遺伝子を正、負それぞれの方向に対して1行で表示していますが、遺伝子密度が高い場合など見にくい場合があります。そのような場合にこの値を2以上にすることで、遺伝子を2行以上で表示することができます。
Width	画像の大きさ(幅)を指定します。この値を倍にすると、画像の幅が倍になります (表示する領域は変わらない)。
Gene Color	遺伝子表示領域に表示される遺伝子の色付け基準を切り替える。
New focus area	表示コントロール領域上にマウスを乗せると、コンティグ上の座標が表示されます。また、その右側にある > をクリックすると、Focus areaが 1/2 だけ右方向に移動します。

## B: 染色体表示領域

選択されているコンティグに対応する染色体が定義されている場合、染色体番号と共にオレンジ色の線で示されます。

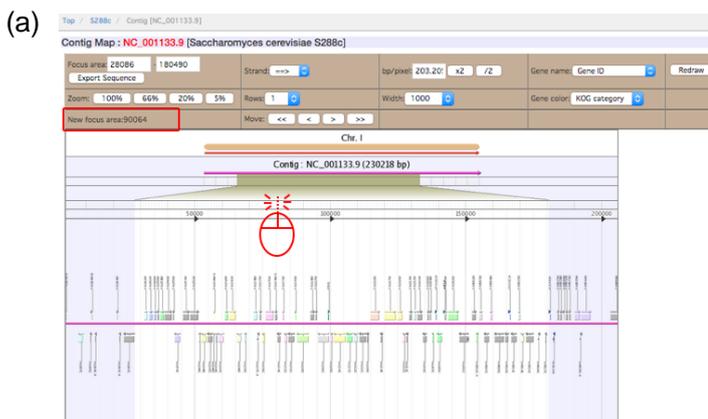
さらに、染色体に対する選択されたコンティグの部位がわかっている場合、その部位と方向が赤い線で示されます。

現在選択されていないコンティグ配列は、赤ではなく灰色の線で示されます。

染色体、コンティグを示す線はリンクになっており、クリックすることでクリックした場所周辺の表示へ移行します。

## C: 表示範囲コントロール領域

この領域でマウスクリックを2回(始点, 終点)を行うと、Focus areaが更新されます。指定領域は上下二つあり、上の領域は大まかなFocus area指定に、下の領域はより細かなFocus area指定のために用いられます(図8)。



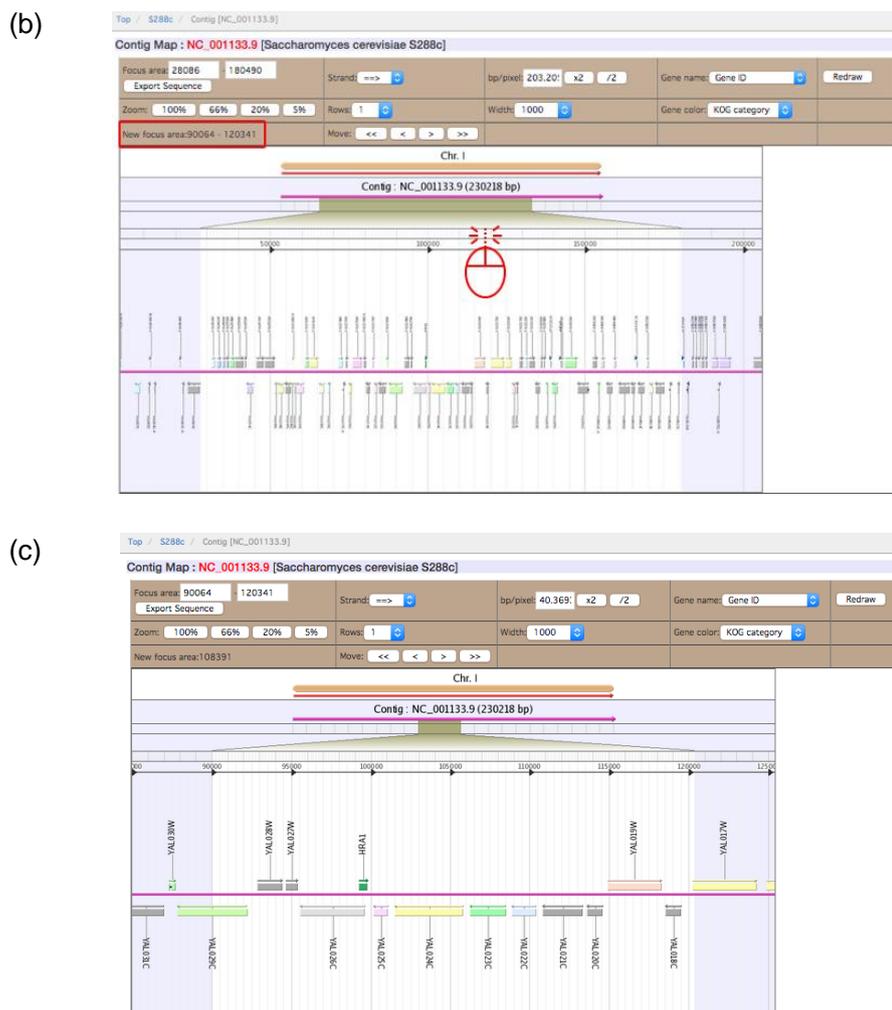


図8: 表示コントロール領域マウス操作による Focus area の変更。コントロール領域に対して、マウスを (a), (b) の順でクリックすると、その領域を新たに Focus area とした新しい図が描画されます (c)。「New focus area」(赤枠)にはマウスのある位置が表示されます。

#### D: 遺伝子表示領域

この領域には、Focus area内にある遺伝子が示されています。遺伝子名、向きが表示され、表示コントロールパネル内で指定されるGene Colorに基づいた遺伝子機能による色分けがなされています。また、各々の遺伝子はリンクとなっており、ここをクリックすると遺伝子詳細画面 (図9または 2-3-3参照) へ移動します。

### 2-3-3. 遺伝子詳細情報の表示

コンティグマップにおいて遺伝子表示領域にある遺伝子をクリックしますと、その遺伝子についての詳細情報を見ることができます(図9)。遺伝子詳細情報はタブによって区切られています：

Locus (図9)： 遺伝子名、塩基配列などのアノテーションの表示、コメントを編集

Transcript (図10)： 転写産物 (RNA) の情報を表示

CDS (図11)： タンパク質コーディング領域 (CDS) の情報を表示

Ortholog (図13)： オーソログ (パラログ) 情報を表示

#### 2-3-3-1. Locus (Locus 情報の表示)

Locus タブ (図 9) には遺伝子名、塩基配列などの各遺伝子のアノテーション情報が表示されます。Locus タブ内の情報は、A:基本情報表、B:コンティグマップ、C:遺伝子配列の3つに分かれています。

The screenshot shows the Locus browser for gene PYC2. It includes a navigation bar with tabs for Locus, Transcript, CDS, and Ortholog. The Locus tab is active, displaying a table of basic information (A), a contig map (B), and the gene sequence (C).

**A: Basic Information Table**

Gene ID	YBR218C
Original symbol	PYC2
Assigned symbol	PYC2
Original Description	pyruvate carboxylase 2 Pyruvate carboxylase isoform; cytoplasmic enzymethat converts pyruvate to oxaloacetate; differentiallyregulated than isoform Pyc1p; mutations in the humanhomolog are associated with lactic acidosis; PYC2 has aparalog, PYC1, that arose from the whole genomeduplication
Assigned Description	Pyruvate carboxylase
Length	3543 bp
GC%	43.30%
Organism	Saccharomyces cerevisiae S288c
Genome	S288c
Chromosome	Chr. II
Contig Location	NC_001134.8, 658707-662249 bp (-)
Links	NCBI_GeneID:852519
Comment	

**B: Contig Map**

The contig map shows the gene PYC2 (YBR218C) in the center, flanked by other genes: YBR223C, YBR222C, YBR223C, YBR220C, YBR219B, YBR218C, YBR216C, YBR217W, YBR215W, YBR214W, YBR213W, and YBR212W. The map includes a scale bar and a legend for strand and gene color.

**C: Gene Sequence**

The gene sequence is displayed in FASTA format, starting with the gene ID YBR218C and the length 3543 bp. The sequence is shown in a monospaced font.

図9: 遺伝子詳細画面 (Locus). 選択された遺伝子に関する詳細情報。Aには基本情報が表で示されています。Bには選択遺伝子を中心とした簡易的なコンティグマップ画面、Cには遺伝子配列が FASTA 形式で書き出されています。

## A: 基本情報表

Gene ID	遺伝子 ID を示します。NCBI 由来のゲノムデータの場合は、locus tag 名を示します。
Original symbol	データソースに記載されていた遺伝子名を示します。
Assigned symbol	SYGD データ登録時に再アノテーションされた遺伝子名を示します。
Original description	データソースに記載されていた遺伝子に対する説明を示します。
Assigned description	SYGD データ登録時に再アノテーションされた遺伝子に対する説明を示します。
Length	遺伝子の塩基長(bp)を示します。
GC%	遺伝子の GC 割合を%単位で示します。
Organism	種名 (菌株名) を示します。
Genome	ゲノム名を示します。ゲノムトップ画面(図 3)へのリンクにもなっています。
Chromosome	染色体情報のある場合に、染色体番号を示します(染色体情報が無い場合には"Unmapped"と示されます)。
Contig Location	コンティグ名を示します。コンティグ名のリンクをクリックすると、この遺伝子を中心としたコンティグマップ画面へ移動します。
Links	外部データベースへのリンクです。

## B: コンティグマップ (Contig Map)

選択された遺伝子を中心としたコンティグマップが表示されます。

## C: 遺伝子配列 (Gene Sequence)

選択された遺伝子の塩基配列が FASTA 形式で表示されます。

また遺伝子の上流と下流の塩基配列も表示できます。"with upstream", "with downstream"のテキストボックスに整数 (半角数字) を入力し"Apply"ボタンをクリックすると、入力した値の長さの配列が表示されます。

## 2-3-3-2. Transcript (Transcript 情報の表示)

Transcript タブ (図10) には転写産物 (RNA) の情報が表示されます。

Transcript タブ内の情報は、(1) 基本情報表、(2) RNA 配列情報の二つに分かれています。

The screenshot shows the transcript information for PYC2 [Saccharomyces cerevisiae S288c]. The page has tabs for Locus, Transcript, CDS, and Ortholog. The Transcript tab is active, displaying a table with the following information:

Gene ID	YBR218C
Type	mRNA
Assigned Description	Pyruvate carboxylase
Length	3543 bp
Original Description	pyruvate carboxylase 2 Pyruvate carboxylase isoform; cytoplasmic enzymethat converts pyruvate to oxaloacetate; differentiallyregulated than isoform Pyc1p; mutations in the humanhomolog are associated with lactic acidosis; PYC2 has aparalog, PYC1, that arose from the whole genomeduplication
Exon Start-End	1-3543
Links	NCBI_GeneID:852519

Below the table is the RNA Sequence section, which displays the FASTA format sequence for PYC2:

```

>P081178566.1 [length:3543]
ATGAGCGCAGCAAAATGAGCGCTTGGGAGCAATTCAGCTTCCTCGCGCAAAAGAAATG
TGGTCCCAATGAGGTGAAATCCGATGAAATTTAGATCTGCTCAAGAGCTCTAAGAGA
GCCCCATACGCCATGAGGACCGCTTCAATGCGCAGGTGAGGCGGAGGAGGCTAGGTTA
GAGGAGGCGCCAGTATACAGCTGAGGCTTCACTGGCAGATGAGGAGATCAAGAAAATGCAAA
ATAGGTGGATTCATCCATCCAGGTATGGGTCTTCTGCTGAAAATTCGGAAATTCGCCGCAAA
GAGGCGCGGTATCACTTGGATCGGCGCTCCAGCTGAAATTAAGACTGCTGAGGAGGAGGCT
AGACACTTGGCAGGAGACTAAGCTTCCACTACCTTCCCGGACTCCAGGACTATGGAAGCTGT
AGGCACTTACTCTTATGATAGGCTACCGGTGACATGAGGCGGCTTGGGCGGCTTGGGCGG
AGGTAAGAGATCGTTAGAGAGGTGAGCAGCTGGCAGATGCCTTCAAGCTGCTACCTCCGAA

```

図 10: 遺伝子詳細画面 (Transcript)

## A: 基本情報表

Gene ID	遺伝子 ID を示します。NCBI 由来のゲノムデータの場合は、locus tag 名を示します。
Assigned description	SYGD データ登録時に再アノテーションされた遺伝子に対する説明を示します。
Type	RNA のタイプを示します。mRNA, tRNA, rRNA の 3 種類があります。
Length	遺伝子の塩基長 (bp) を示します。
Original description	データソースに記載されていた遺伝子に対する説明を示します。
Exon Start-End	遺伝子配列中、Exon の (一般には複数の) 始点-終点を示します。
Links	外部データベースへのリンクです。

## B: RNA 配列情報 (RNA Sequence)

選択された遺伝子の RNA 配列が FASTA 形式で表示されます。

## 2-3-3-3. CDS (CDS 情報の表示)

CDS タブ(図11)にはタンパク質コーディング領域(CDS)の情報が表示されます。

CDS タブ内の情報は、(1)基本情報表、(2)CDS 配列に分かれます。

The screenshot shows the CDS page for the PYC2 gene. It includes a table with the following fields:

- Original Symbol:** PYC2
- Assigned Symbol:** PYC2
- Original Description:** pyruvate carboxylase 2
- Assigned Description:** Pyruvate carboxylase
- Length:** 1188 bp
- Links:** NCBI, GeneID: 892318, S288: S000000422
- Similarity to KOG:** KOG0293 Pyruvate carboxylase eukal-e0
- Similarity to COG:** CG1339 PyxA Pyruvate carboxylase eukal-e0
- Similarity to SwissProt:** P1C1\_YEAST PYC2 Pyruvate carboxylase 2 eukal-e0, P1C1\_RIC1 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC4 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC5 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC6 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC7 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC8 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC9 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC10 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC11 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC12 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC13 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC14 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC15 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC16 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC17 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC18 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC19 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC20 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC21 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC22 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC23 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC24 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC25 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC26 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC27 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC28 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC29 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC30 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC31 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC32 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC33 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC34 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC35 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC36 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC37 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC38 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC39 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC40 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC41 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC42 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC43 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC44 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC45 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC46 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC47 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC48 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC49 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC50 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC51 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC52 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC53 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC54 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC55 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC56 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC57 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC58 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC59 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC60 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC61 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC62 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC63 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC64 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC65 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC66 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC67 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC68 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC69 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC70 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC71 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC72 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC73 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC74 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC75 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC76 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC77 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC78 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC79 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC80 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC81 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC82 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC83 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC84 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC85 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC86 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC87 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC88 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC89 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC90 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC91 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC92 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC93 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC94 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC95 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC96 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC97 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC98 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC99 PYC2 Pyruvate carboxylase eukal-e0, P1C1\_RIC100 PYC2 Pyruvate carboxylase eukal-e0
- IC number:** 6.4.1.2 pyruvate carboxylase
- Plan:** P1C1
- PRINTS:** P1C1
- Prosite:** P1C1
- TM:** 0
- Signal peptides:** 0
- GP2-anchor:** 0

The protein sequence is shown in FASTA format at the bottom of the page.

図 11: 遺伝子詳細画面 (CDS).

## A: 基本情報表

Original symbol	データソースに記載されていた遺伝子名を示します。
-----------------	--------------------------

Assigned symbol	SYGD データ登録時に再アノテーションされた遺伝子名を示します。
Original description	データソースに記載されていた遺伝子に対する説明を示します。
Assigned description	SYGD データ登録時に再アノテーションされた遺伝子に対する説明を示します。
Length	遺伝子の塩基長(bp)を示します。
Links	外部データベースへのリンクです。
similarity to KOG	KOG 配列との類似性があった場合、KOG ID とその説明, BLAST 検索によるアラインメントの e-value が示されます。e-value 値はリンクとなっており、ここをクリックするとアラインメントの詳細を見ることができます。
similarity to COG	上と同様。COG 配列との類似性を示します。
similarity to SwissProt	上と同様 SwissProt 配列との類似性を示します。
EC number	遺伝子に EC number 情報があった場合、EC number を表示します。
Pfam	Pfamドメインとの類似性があれば ID とその説明を示します。 "map" をクリックすると、類似性が認められる部分をグラフィカルに表示します
PRINTS	上と同様。PRINTS ドメインとの類似性を示します。 "map" をクリックすると、類似性が認められる部分をグラフィカルに表示します。
Prosite	上と同様。Prosite ドメインとの類似性を示します。 "map" をクリックすると、類似性が認められる部分をグラフィカルに表示します。
TMs	膜貫通領域予測計算結果を示します。 "map" をクリックすると、膜貫通領域と予測される部分をグラフィカルに表示します。
Signal peptides	signal peptide 予測計算結果を示します。
GPI-anchor	GPI アンカーとの結合が予測された部分を示します。

## B: CDS 配列 (Protein Sequence)

選択された遺伝子のアミノ酸配列が FASTA 形式で書かれます。

CDS タブ (図11) の最上部には、"Annotation Map" ボタンがあります。これをクリックすると、アノテーション情報の詳細が得られます (図12)。

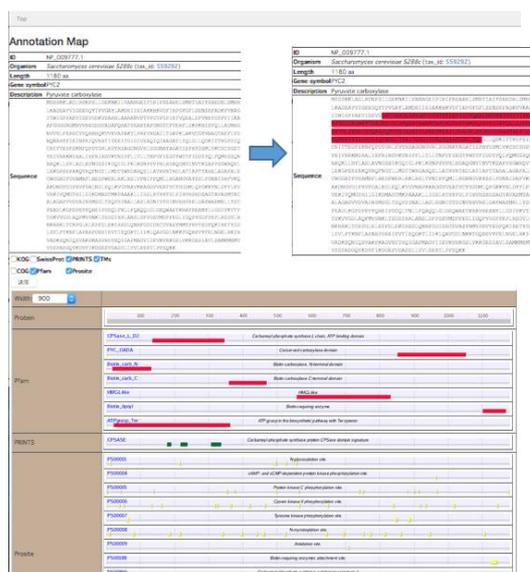


図12: Annotation Map. CDS アミノ酸配列 (上) とアノテーション情報 (下). 下部アノテーション情報では、Pfam, PRINTS等との類似性がグラフィカルに示されています。マップの対応する部分をマウスでクリックすると、上部アミノ酸配列中で該当する部分の色が変化します。SwissProt等、アノテーション計算の際の配列類似性計算にBLAST検索を実行しているものに対しては、BLAST検索結果を別ウィンドウに表示します。

#### 2-3-3-4. Ortholog (Ortholog 情報の表示)

Ortholog タブ(図 13)では、選択された遺伝子についてのオーソログ(パラログ)情報を示します。

Genome	Genes	Gene ID	Original Description	Links
Saccharomyces cerevisiae S288c	2	YGL062W	pyruvate carboxylase 1	<a href="#">NCBI_GeneID:852818</a> <a href="#">SGD:S000003030</a>
		YBR218C	pyruvate carboxylase 2	<a href="#">NCBI_GeneID:852519</a> <a href="#">SGD:S00000422</a>
Saccharomyces cerevisiae Kyokai no. 7	2	SYK7_005261	K7_Pyc2p	
		SYK7_025961	K7_Pyc1p	

図 13: 遺伝子詳細表示(Ortholog)

Orthologタブでは、ゲノムリスト(図2)のゲノムにある遺伝子配列との類似性を検索し、種を超えて相同性のある遺伝子についての一覧表が表示されます。

オーソログ計算には、類似性閾値などの違いなどいくつか異なる計算方法があります。計算方法は、左上セレクトメニューで選択できます。

また左上のセレクトメニューの下にあるリンクをクリックすると、このオーソログを用いたシンテニー解析画面を表示します(図14)。

#### 2-3-4. シンテニーマップの表示

SYGD ではオーソログ (パラログ) の関係のある遺伝子群について、シンテニー解析を行なった結果を表示する機能があります。

図 13 オーソログのリンクをクリックすると、シンテニーマップ画面を見ることができます(図 14)。

Top / Ortholog [ort1\_c70\_e100\_142]

### Ortholog : ort1\_c70\_e100\_142

No./#	Genome	Gene (Original symbol)	Location	Command
1	[S288c] Saccharomyces cerevisiae S288c	YBR218C (PYC2)	NC_001134.8:658707<-662249	Delete
2	[S288c] Saccharomyces cerevisiae S288c	YGL062W (PYC1)	NC_001139.9:385196->388732	Delete
3	[K7] Saccharomyces cerevisiae Kyokai no. 7	SYK7_005261 (K7_PYC2)	DG000038.1:636459<-640001	Delete
4	[K7] Saccharomyces cerevisiae Kyokai no. 7	SYK7_025961 (K7_PYC1)	DG000043.1:389278->392814	Delete

Redraw

---

Width: 600    bp/pixel: 30    x4    x2    /2    /4    Gene color: KOG category    Gene name: Gene ID    リセット    送信

#### Gene info.

Gene ID	YBR218C
Original symbol	PYC2
Assigned symbol	PYC2
Original Description	pyruvate carboxylase
Length	3543 bp (1180 aa)
GC%	43.30%
Genome	[S288c] Saccharomyces cerevisiae S288c
Chromosome	Chr. II
Contig	NC_001134.8 [658707 - 662249 bp]
BLAST	

#### Synten map

---

Gene ID	YGL062W
Original symbol	PYC1
Assigned symbol	PYC1
Original Description	pyruvate carboxylase
Length	3537 bp (1178 aa)
GC%	43.43%
Genome	[S288c] Saccharomyces cerevisiae S288c
Chromosome	Chr. VII
Contig	NC_001139.9 [385196 - 388732 bp]
BLAST	

図14: シンテニーマップ画面。オーソログ関係 (同じ種であればパラログ関係) にある遺伝子群の情報を示します。画面上部にオーソログとして選択された遺伝子の一覧表および、表示をコントロールするパネルがあります (図15)。画面下部に、各遺伝子の基本情報 (左, Gene info) と遺伝子マップ (右, Synten map) を示します。左の遺伝子の基本情報では、Gene IDやDescription などの情報、およびBLAST検索の結果が表示されます。右の遺伝子マップではマップの中央に注目している遺伝子が配置され、その周辺の情報が表示されます。遺伝子マップは、"Width", "bp/pixel", "Gene color", "Gene Name" のセレクトボックスで設定を変更することにより、表示を変更させることができます。

シンテニーマップ (図14) の上には、オーソログ (パラログ) として選択された遺伝子の一覧表があります (図15)。

これは、シンテニーマップの表示順序を変えたり、遺伝子を非表示にしたり (Delete) するコントロールパネルとなります。

#### Ortholog : ort1\_c70\_e100\_142

No./#	Genome	Gene (Original symbol)	Location	Command
1	[S288c] Saccharomyces cerevisiae S288c	YBR218C (PYC2)	NC_001134.8:658707<-662249	Delete
2	[S288c] Saccharomyces cerevisiae S288c	YGL062W (PYC1)	NC_001139.9:385196->388732	Delete
3	[K7] Saccharomyces cerevisiae Kyokai no. 7	SYK7_005261 (K7_PYC2)	DG000038.1:636459<-640001	Delete
4	[K7] Saccharomyces cerevisiae Kyokai no. 7	SYK7_025961 (K7_PYC1)	DG000043.1:389278->392814	Delete

Redraw

図15: シンテニーマップのコントロールパネル。Commandにあるボタンで表示を操作する。"Delete"ボタン を押すと該当する遺伝子がシンテニーマップ表示から削除されます。また、矢印ボタン (↑↓) を押すと表示順序を変えることができる。設定の後 Redraw ボタンを押すと、設定に基づき再描画される。

## 2-4. BLAST 検索

SYGDでは登録されている全ゲノムに対してBLAST検索を行なうことができます。メインメニューの"BLAST"をクリックすると、BLAST 検索画面に移動します (図 16)。

図 16: BLAST 検索画面。

BLAST 検索画面の説明を以下に記します：

Target DB	検索対象とする配列を選択します。 Genome：全ゲノム配列 All Genes：遺伝子配列 All Transcripts：RNA配列 All CDSs：CDS 配列 All Proteins：アミノ酸配列 Intergenic Regions：遺伝子間領域 Upstream 2kb from all genes：遺伝子上流の 2kb の配列
Query Sequence/ File Upload	検索したい配列を Query Sequence 内に入力するか、あるいは File Upload から配列情報の入ったファイルをアップロードします。 対応フォーマット：配列のみ、FASTA 形式、GenBank 形式
Target Genome	ゲノムリストにあるゲノムのうち、どのゲノムを検索対象にするかを指定します。
E-value	BLAST 検索 e-value の閾値を決めます。 セレクトメニューまたは、テキストボックスにて閾値を入力してください。
Additional position	他のオプションを用いたい場合には、オプションと値をここに直接書き込むことができます。

BLAST 検索結果の例を図 17 に示します。

```

>NC_001134.8 NC_001134.8 ([S288c] Saccharomyces cerevisiae S288c)
      Length=813184
      Score      = 7002.17 (3532), Expect      = 0,
      Identities = 3532/3532 (100%), Positives  = 3532/3532 (100%), Gaps      = 0/3532 ( 0%)
      Strand = Plus / Minus

Query 1      ATGAGCAGTAGCAAGAAATTGGCCGGTCTTAGGGACAATTTAGTTTGGCTCGGCGAAAAG 60
             |||
Sbjct 662249 ATGAGCAGTAGCAAGAAATTGGCCGGTCTTAGGGACAATTTAGTTTGGCTCGGCGAAAAG 662190

Query 61     AATAAGATCTTGGTCGCCAATAGAGGTGAAATTCGGATTAGAATTTTATAGATCTGCTCAT 120
             |||
Sbjct 662189 AATAAGATCTTGGTCGCCAATAGAGGTGAAATTCGGATTAGAATTTTATAGATCTGCTCAT 662130

Query 121    GAGCTGCTATGAGAACCATCGCCATATACTCCATGAGGACCGTCTTTCAATGCACAGG 180
             |||
Sbjct 662129 GAGCTGCTATGAGAACCATCGCCATATACTCCATGAGGACCGTCTTTCAATGCACAGG 662070

Query 181    TTGAAGGCGGACGAAGCGTATGTTATCGGGGAGGAGGGCCAGTATACACCTGTGGGTGCT 240
             |||
Sbjct 662069 TTGAAGGCGGACGAAGCGTATGTTATCGGGGAGGAGGGCCAGTATACACCTGTGGGTGCT 662010

Query 241    TACTTGGCAATGGACGAGATCATCGAAATTGCAAAGAAGCATAAAGGTGATTTCATCCAT 300
             |||
Sbjct 662009 TACTTGGCAATGGACGAGATCATCGAAATTGCAAAGAAGCATAAAGGTGATTTCATCCAT 661950

Query 301    CCAGGTTATGGGTTCTTGTCTGAAAATTCGGAATTTGCCGACAAAGTAGTGAAGGCCGGT 360
             |||
Sbjct 661949 CCAGGTTATGGGTTCTTGTCTGAAAATTCGGAATTTGCCGACAAAGTAGTGAAGGCCGGT 661890

Query 361    ATCACTTGGATCGGCCCTCCAGCTGAAGTTATTGACTCTGTGGGTGACAAAGTCTCTGCC 420
             |||
Sbjct 661889 ATCACTTGGATCGGCCCTCCAGCTGAAGTTATTGACTCTGTGGGTGACAAAGTCTCTGCC 661830

Query 421    AGACACTTGGCAGCAAGAGCTAACGTTCTACCGTTCCCGTACTCCAGGACCTATCGAA 480
             |||
Sbjct 661829 AGACACTTGGCAGCAAGAGCTAACGTTCTACCGTTCCCGTACTCCAGGACCTATCGAA 661770

Query 481    ACTGTGCAAGAGGCACTTGACTTCGTTAATGAATACGGCTACCCGGTGATCATTAAAGGCC 540
             |||
Sbjct 661769 ACTGTGCAAGAGGCACTTGACTTCGTTAATGAATACGGCTACCCGGTGATCATTAAAGGCC 661710

Query 541    GCCTTTGGTGGTGGTAGAGGTATGAGAGTCGTTAGAGAAGGTGACGACGTGGCAGAT 600
             |||
Sbjct 661709 GCCTTTGGTGGTGGTAGAGGTATGAGAGTCGTTAGAGAAGGTGACGACGTGGCAGAT 661650

```

図 17: BLAST 検索結果の例

## 2-5. ドットプロット解析

ドットプロット解析を行なうことで2種類のゲノム、染色体、あるいはコンティグ間の類似性を全体的に眺めることができます。

メインメニューの "Dot plot analysis" のリンクをクリックすることで、ドットプロット解析画面に移動します(図 18)。

The screenshot shows a navigation menu with 'Dot plot analysis' highlighted. Below it is a table of genome data:

Organism	S288c	SGD	Taxon. id	Mb	Chromosome	Contigs	CDSs	tRNAs	rRNAs
Saccharomyces cerevisiae S288c	S288c	SGD	559292	12.1	16	17	6008	299	14
Saccharomyces cerevisiae Kyokai no. 7	K7	NRIB_SYGD	721032	11.8	16	17	5816	296	7
All	-	-	-	-	32	34	11824	595	21

Below the table, a blue arrow points to the 'Dotplot analysis' configuration page, which includes a title 'Dotplot analysis [region based](count based)', a 'Genome vs Genome' dropdown, a resolution input field set to '50000 [bp/pixel]', an 'e-value' input field set to '1e-10', and two 'Query' dropdowns for 'X axis' and 'Y axis'. An 'Apply' button is at the bottom.

図18: ドットプロット解析画面。メインメニューの "Dot plot analysis" をクリックすることで、ドットプロット解析画面へ移動します。

左上のセレクトメニューで、ゲノム同士の比較、染色体同士の比較、コンティグ同士の比較を選択できます。中上のメニューでは、図の解像度 (図19) を決めます。数値が小さくなるほど大きな図となります。右上メニューは検索(BLAST 検索)の e-value 閾値となります。一般に、ゲノム同士の比較を行うと非常に時間がかかる場合があります。region based はゲノム上の塩基単位で、count basedはCDS単位で図を作成します。

ゲノム同士のドットプロットから、見たい染色体、コンティグ、コンティグ内の領域における具体的な遺伝子を見ることができます (図19)。

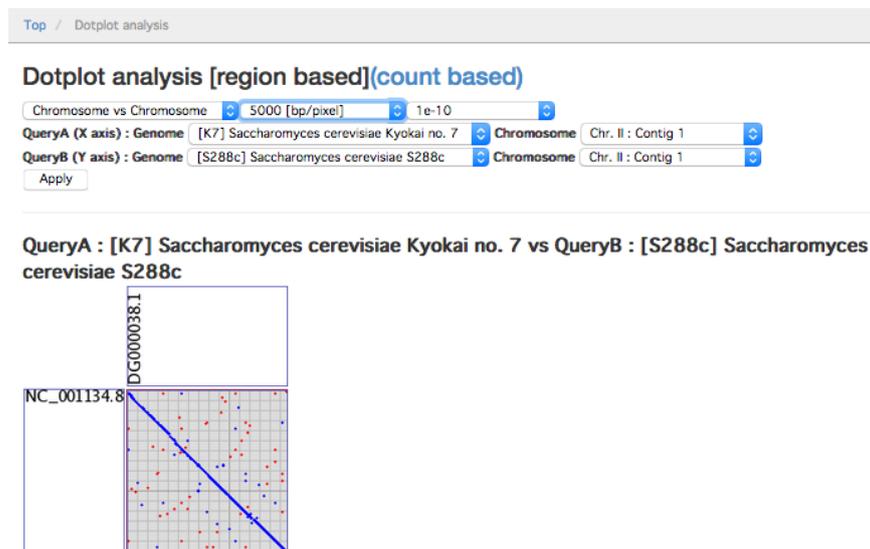


図19: ドットプロット解析画面.ゲノム同士のドットプロット解析結果を示します。

## 2-6.差分解析

差分解析を行なうことにより、複数のゲノムグループ間で共通に存在する遺伝子、あるいは特定のグループにのみ存在する遺伝子を検索することができます。

メインメニューの "Differential analysis" のリンクをクリックすることで、差分解析画面に行くことができます(図 20)。

The screenshot shows the 'Differential analysis' interface. At the top, there is a navigation menu with 'Differential analysis' highlighted. Below it is a table of genome data for 'Saccharomyces cerevisiae'. A red arrow points to the 'Differential analysis' link. A blue arrow points down to the main interface. The main interface shows a 'Differential analysis' page with a 'Group' list, a 'Venn's diagram' showing two overlapping circles labeled 'A' and 'B', and a 'Search' button.

Organism	S288c	SGD	Taxid: 4914	Chr	Chromosome	Contigs	CDSs	tRNAs	rRNAs
Saccharomyces cerevisiae S288c	S288c	SGD	559292	12.1	16	17	6008	299	14
Saccharomyces cerevisiae Kyokai no. 7	K7	NRIB_SYGD	721032	11.8	16	17	5816	296	7
All	-	-	-	-	32	34	11824	595	21

図20: 差分解析画面。メインメニューの "Differential analysis" をクリックすることで、差分解析画面へ移動します。

用いる ortholog set を選択し、作成するベン図 (No. of group. 1, or 2, or 3) を選択、さらに group に属する種 (図では A, B に属する種) を選択する。

Search ボタンを押すと ortholog の検索画面 (図 32, 2-7-3 オーソログ検索参照) へ移動します。

## 2-7. 基本検索

メインメニューの "Search" のリンクをクリックすることで、基本検索画面に行くことができます (図 21)。

Acronym	Source	Taxon.id	Mb	Chromosome	Contigs	CDSs	tRNAs	rRNAs
S288c	SGD	559292	12.1	16	17	6008	299	14
K7	NRIB_SYGD	721032	11.8	16	17	5816	296	7

図 21: 基本検索画面. メインメニューの "Search" をクリックすることで、基本検索画面へ移動します。

基本検索では、以下に示すものをキーワードや条件で検索することができます。

- (1). コンティグ
- (2). 遺伝子、RNA、CDS
- (3). オーソログ (パラログ)
- (4). COG、KOG、EC アノテーション

### 2-7-1. コンティグ検索

SYGD に登録されているゲノム情報の中から、コンティグを検索することができます。

コンティグ検索画面を図22、コンティグ検索結果表示画面を図23 に示します。

コンティグ検索画面は、基本検索画面 (図 21) の "Contig" リンクをクリックすることで表

示されます。

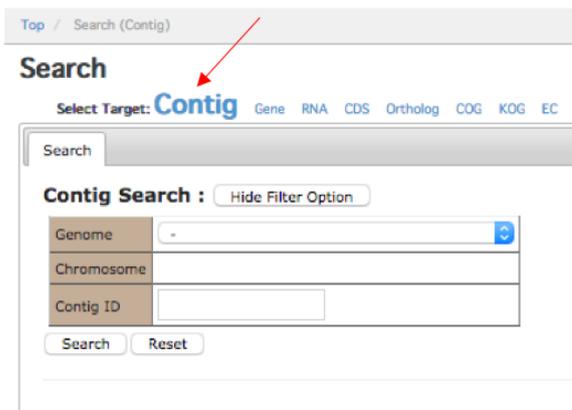


図22: コンティグ検索画面。基本検索画面の "Contig" をクリックする (図の矢印) ことで、コンティグ検索画面を表示させることができます。基本検索画面の表示方法については、2-7 基本けんさく参照。表中にある項目に検索条件を入れて、"Search" ボタンを押すと、条件に合致したコンティグのリストが表示されます。

コンティグ検索画面では、検索したいゲノム、染色体をメニューから選択し、ID等を入力することでコンティグの検索を行うことができます。コンティグ検索画面の詳細を以下に記します：

Genome	検索したいゲノムを選択することができます。セレクトメニューから適当なゲノム名を選択してください。
Chromosome	"Genome"のセレクトメニューで染色体情報を持っているゲノムを選択した場合、検索したい染色体を選択することができます。セレクトメニューから適当な染色体を選択してください。
Contig ID	登録されているコンティグに対して、Contig ID で検索します。

コンティグ検索結果表示画面は以下の 3 つで構成されています (図23)。

また、以降説明する遺伝子、RNA、CDS、オーソログ、COG、KOG、EC アノテーションの検索結果表示画面も同様な構成になっています。

- (1). フィルターオプションの表示: フィルターオプションの表示／非表示
- (2). コンティグリスト: コンティグの一覧表が表示されます
- (3). 表示コントロール部: 一覧表の並び替え (ソート) を実行、一覧表のカラムの表示／非表示、データのダウンロードを行なうことができます。

Top / Search (Contig)

**Search**

Select Target: **Contig** Gene RNA CDS Ortholog COG KOG EC

Search

Contig Search :

Genome:

Chromosome:

Contig ID:

Contig list 1 - 1 of 1 Rows:  Page:    Top Bottom Page: 1

No./1	Contig ID	Organism	Chromosome	Length	GC%	No. of Genes
1	NC_001144.5	Saccharomyces cerevisiae S288c	Chr. XII	1078177 bp	38.48%	572

Contig list 1 - 1 of 1 Rows:  Page:    Top Bottom Page: 1

Sort Columns

First Key:   Ascend  Second Key:   Ascend  Third Key:   Ascend

Show / Hide Columns

Contig ID  Organism  Genome  Chromosome  Length  GC%  No. of Genes

Download Data

Nucleotide Sequences  Tab Delimited File

図 23: コンテック検索結果表示画面.

コンテック検索結果表示画面のカラム (図 23) の説明を以下に記します：

Contig ID	登録されているコンテックのContig ID を示します。 またコンテックマップ画面へのリンクにもなっています。
Organism	分類学上の種 (species) を示します。
Genome	SYGDに登録されているゲノムの名称を示します。 この名称は、主にゲノムの分類学上の種 (species) ・サンプル名・解析名・略称などで示されます。 また略称はゲノムトップ画面 (図3) へのリンクにもなっています。 初期状態 (デフォルト) では表示されません。表示コントロールで"Genome" がチェックすることで表示されます。
Chromosome	Chromosome情報がある場合には、属しているChromosome名が表示されます。 Chromosome情報がない場合には "Unmapped" と表示されます。 コンテックリスト画面へのリンクにもなっています。クリックすることで属しているChromosome のコンテックリスト一覧が表示されます ("Unmapped"である場合には、"Unmapped"のコンテック一覧が表示されます) 。
Length	コンテックの長さを示します。
GC%	コンテックの GC% を示します。
No. of Genes	コンテック上の遺伝子数を示します。 また遺伝子検索結果表示画面 (図 25) へのリンクにもなっています。

## 2-7-2. 遺伝子、RNA、CDS 検索

SYGD に登録されているゲノム情報の中から、遺伝子・RNA・CDS を検索することができます。

遺伝子検索画面を図24 に示します (RNAやCDS検索画面も、ほぼ同じ仕様となっています)。また遺伝子検索結果表示画面を図25、RNA検索結果を図26、CDS検索結果を図27 に示します。

遺伝子検索画面は、基本検索画面 (図21) の "Gene" リンクをクリックすることで表示されます。同様に、RNA検索画面は "RNA"、CDS検索画面は "CDS" をクリックすることで表示されます。

Top / Search (Gene)

**Search**

Select Target: Contig **Gene** RNA CDS Ortholog COG KOG EC

Search

Gene Search :

Full Text	<input type="text"/>
Gene ID	<input type="text"/>
Gene Symbol	<input type="text"/>
Description	<input type="text"/>
Genome	<input type="text"/> <input type="button" value="v"/>
Chromosome	<input type="text"/>
Contig	<input type="text"/>
Contig Location	From <input type="text"/> To <input type="text"/>
Comment	<input type="text"/>
Ortholog	<input type="text"/>
COG ID	<input type="text"/>

図24: 遺伝子検索画面。基本検索画面の "Gene" をクリックする (図の矢印) ことで、遺伝子検索画面を表示させることができます。同様にRNA検索画面は "RNA" を、CDS検索画面は "CDS" をクリックすることで、それぞれの検索画面を表示させることができます。基本検索画面の表示方法については、2-7 基本検索参照。表中にある項目に検索条件を入れて、"Search" ボタンを押すと、条件に合致した遺伝子 (RNA, CDS) のリストが表示されます。

遺伝子検索画面の各項目の詳細を以下に記します：

Full Text	登録されている遺伝子に対して全文検索を行います。
Gene ID	登録されている遺伝子に対して Gene ID で検索を行います。
Gene Symbol	登録されている遺伝子に対して遺伝子の symbol 名で検索を行います。 スペースを空けて複数指定できます (or 検索)。
Description	登録されている遺伝子に対してDescription情報で検索を行います。例えば、 "enzyme" と入力すると、gene description 欄に "enzyme" という文字列が含まれる 遺伝子を検索します。
Genome	検索したいゲノムを選択することができます。セレクトメニューから適当なゲノム名 を選択してください。
Chromosome	"Genome"のセレクトメニューで染色体情報を持っているゲノムを選択した場合、 検索したい染色体を選択することができます。セレクトメニューから適当な染色体 を選択してください。
Contig	"Genome"のセレクトメニューでコンティグ情報を持っているゲノムを選択した場 合、検索したいコンティグを選択することができます。セレクトメニューから適当 なコンティグを選択してください。
Contig Location	コンティグ中の検索したい領域を指定できます。テキストボックスに検索したい領 域を数値 (単位bp) で入力してください。
Comment	登録されている遺伝子に対してComment情報で検索を行います。例えば、ある遺伝 子のComment欄に"enzyme"という単語を保存していた場合、"enzyme"と入力する ことでComment欄に"enzyme"と記載した遺伝子を検索します。
Ortholog	登録されている遺伝子に対して Ortholog ID で検索を行います。
COG ID	登録されている遺伝子に対して COG (オーソログ) 情報の ID で検索を行います。
COG Functions	登録されている遺伝子に対して COG (オーソログ) 情報の functional categories で 検索を行います。
KOG ID	登録されている遺伝子に対して KOG (オーソログ) 情報の ID 情報で検索を行います。

KOD Functions	登録されている遺伝子に対して KOG (オーソログ) 情報の functional categories で検索を行います。
PRINTS Category	登録されている遺伝子に対して PRINTS (タンパク質モチーフ) 情報で検索を行います。
EC Number	登録されている遺伝子に対して EC Number 情報で検索を行います。
Pfam	登録されている遺伝子に対して Pfam (タンパク質モチーフ) 情報で検索を行います。
Prosite	登録されている遺伝子に対して Prosite (タンパク質モチーフ) 情報で検索を行います。
No. of TMs	登録されている遺伝子に対して膜貫通領域と予測されたドメインについてその数をもとに検索を行います。 右のテキストボックスに適当なドメイン数を半角文字で入力し、左のセレクトメニューから、以上(>=), 同じ(=), 以下(<=) を選択してください。
GPI-anchor	GPI アンカーとの結合が予測された遺伝子について検索を行います。
Signal Peptide	シグナルペプチドと予測された遺伝子について検索を行います。

Gene list 1 - 2 of 2      Rows: 50    Page: 1    Top Bottom    Page: 1

No./2	Gene ID	Original symbol	Assigned symbol	Original description	Assigned description	Links	Organism	Chromosome	Comment
1	YBR218C	PYC2	PYC2	pyruvate carboxylase 2	Pyruvate carboxylase	<a href="#">NCBI_GeneID:852519</a>	Saccharomyces cerevisiae S288c	Chr. II	-
2	<a href="#">SYK7_005261</a>	K7_PYC2	PYC2	K7_Pyc2p	Pyruvate carboxylase		Saccharomyces cerevisiae K yokai no. 7	Chr. II	-

Gene list 1 - 2 of 2      Rows: 50    Page: 1    Top Bottom    Page: 1

図 25: 遺伝子検索結果表示画面.

RNA list 1 - 50 of 11883      Rows: 50    Page: 1    Top Bottom    Page: 1, 2, 3, 4, 5, 6

No./11883	Gene ID	Original symbol	Assigned symbol	Links	Description	RNA-type	Organism	Chromosome	Comment
1	YLL067C		YRF1-1	<a href="#">NCBI_GeneID:850659</a> <a href="#">NCBI_Rna:NM_001181887.1</a>	Y' element ATP-dependent helicase Putative Y' element ATP-dependent helicase	mRNA	Saccharomyces cerevisiae S288c	Chr. XII	-
2	YLL066W-B			<a href="#">NCBI_GeneID:1466404</a> <a href="#">NCBI_Rna:NM_001184605.1</a>	hypothetical protein hypothetical protein; overexpression causes a cellcycle delay or arrest	mRNA	Saccharomyces cerevisiae S288c	Chr. XII	-
3	YLL066C		YRF1-4	<a href="#">NCBI_GeneID:850660</a> <a href="#">NCBI_Rna:NM_001181886.1</a>	Y' element ATP-dependent helicase Putative Y' element ATP-dependent helicase; YLL066C is not an essential gene	mRNA	Saccharomyces cerevisiae S288c	Chr. XII	-

図 26: RNA 検索結果表示画面.

CDS list 1 - 50 of 6008      Rows: 50    Page: 1    Top Bottom    Page: 1, 2, 3, 4, 5, 6

No./6008	Gene ID	Original description	Assigned description	Genome	Chromosome	Links	Protein sequence
1	YLL067C	Y' element ATP-dependent helicase	Y' element ATP-dependent helicase YLL067C	Saccharomyces cerevisiae S288c (S288c)	Chr. XII	SGD:S000003990 NCBI_GeneID:850659	>YLL067C NP_013033.1 NC_001144.5:585-4202:- 120: MRTFTDFVSGAPIVRSLSQKSTIRRYGYNLAPHMFLLLHVDLSI KPIETKYFSQICNDMMNKDRLDGLVATAQRIRRRYKNGSSSE FSSTRNAEDVAGEAASSDHDQKISRVTQRKRPREPKNSTNDLIV DTPSVAVQAPPYGGKTELFHLLPLIALASKGDVKKYVFLFVPTV GCDGVTDLVGIYDDLASTNFTDRIAAWENIVECTFRTNMVKL
2	YLL066W-B	hypothetical protein	Uncharacterized protein YLL066W-B	Saccharomyces cerevisiae S288c (S288c)	Chr. XII	SGD:S000028672 NCBI_GeneID:1466404	>YLL066W-B NP_878115.1 NC_001144.5:5605-5775:+ 56 MSLRPCLTPSSMQYSDIYHHTHTPHPHHTHTHTHTPHSLTSTF

図 27: CDS 検索結果表示画面。

遺伝子検索結果 (図 25) で初期状態 (デフォルト) で表示されるカラムの説明を以下にいたします。

Gene ID	遺伝子の ID を示します。NCBI 由来のゲノムデータの場合は、locus tag 名を示します。
Original symbol	データソースに記載されていた遺伝子名を示します。
Assigned symbol	SYGD データ登録時に再アノテーションされた遺伝子名を示します。
Original description	データソースに記載されていた遺伝子に対する説明を示します。
Assigned description	SYGD データ登録時に再アノテーションされた遺伝子に対する説明を示します。
Links	外部データベースへのリンクです。
Organism	種名 (菌株名) を示します。
Chromosome	染色体情報のある場合に、染色体番号を示します (染色体情報が無い場合には "Unmapped" と示されます)。
Comment	記録したコメントを示します。
Homologous genes	SYGDに登録された他のゲノムと相同性のある遺伝子を表示することができます。セレクトボックスから任意のゲノムを選択し、"Apply"ボタンを押すことで相同性のある遺伝子が表示されます。

遺伝子検索結果 (図 25) で選択することにより表示されるカラムの説明を以下にいたします。

Description	
Genome	ゲノム名を示します。リンクをクリックすると、ゲノムトップ画面(図 3)へ移動します。
Chromosomal location	染色体情報のある場合に、染色体上の位置情報が示されます。
Contig	この遺伝子のあるコンティグ名が示されます。
Contig location	この遺伝子のあるコンティグの、コンティグ上の位置情報が示されます。
Map	この遺伝子を中心としたコンティグマップが表示されます。
Gene Sequence	遺伝子配列が FASTA 形式で表示されます。
Upstream sequence	この遺伝子の上流の配列を表示します。テキストボックスに数値 (半角数字) を入力し "Apply" ボタンを押してください。
Downstream sequence	この遺伝子の下流の配列を表示します。テキストボックスに数値 (半角数字) を入力し "Apply" ボタンを押してください。

Upstream variations	変異解析(M-option)を実施していた場合に利用できます。セレクトボックスからゲノムを選択し"Apply"ボタンを押すことで、遺伝子上流の変異の数が表示されます。この数値は変異リストへのリンクにもなっています (詳細は3-1節に記載)。
Ortholog	この遺伝子のオーソログ (パラログ) 情報がある場合、シンテニー解析画面 (図 14) へのリンクが表示されます。
COG	この遺伝子が COG (オーソログ) 情報を含む場合、COG ID を表示します。
COG Function	この遺伝子が COG (オーソログ) 情報を含む場合、COG functional categories を表示します。
KOG	この遺伝子が KOG (オーソログ) 情報を含む場合、KOG ID を表示します。
KOG Function	この遺伝子が KOG (オーソログ) 情報を含む場合、KOG functional categories を表示します。
PRINTS Category	この遺伝子が PRINTS ドメインと類似性がある場合に、PRINTS Category を表示します。
PRINTS	この遺伝子が PRINTS ドメインと類似性がある場合に、PRINTS 情報を表示します。
EC number	この遺伝子が EC number 情報がある場合に、EC number 情報を表示します。
Pfam	この遺伝子が Pfam ドメインと類似性がある場合に、PRINTS 情報を表示します。
Prosite	この遺伝子が Prosite ドメインと類似性がある場合に、PRINTS 情報を表示します。
TMs	この遺伝子が膜貫通領域と予測されたドメインがある場合に、そのドメイン数を示します。
GPI-anchor	この遺伝子が GPI アンカーとの結合が予測された配列を含む場合に、その配列の長さを表示します。
Signal Peptide	この遺伝子がシグナルペプチドと予測された配列を含む場合に、その配列の長さを表示します。

RNA 検索結果(図 26)で初期状態(デフォルト)で表示されるカラムの説明を以下にいたしません。

Gene ID	遺伝子の ID を示します。NCBI 由来のゲノムデータの場合は、locus tag 名を示します。
Original symbol	データソースに記載されていた遺伝子名を示します。
Assigned symbol	SYGD データ登録時に再アノテーションされた遺伝子名を示します。
Links	外部データベースへのリンクです。
Description	データソースに記載されていた遺伝子に対する説明を示します。
RNA-type	RNA のタイプを示します。mRNA, tRNA, rRNA の種類があります。
Chromosome	染色体情報のある場合に、染色体番号を示します (染色体情報が無い場合には "Unmapped" と示されます)。
Organism	種名 (菌株名) を示します。
Comment	記録したコメントを示します。
Homologous genes	SYGDに登録された他のゲノムと相同性のある遺伝子を表示することができます。セレクトボックスから任意のゲノムを選択し、"Apply"ボタンを押すことで相同性のある遺伝子が表示されます。

RNA 検索結果 (図 26) で選択することにより表示されるカラムの説明を以下にいたします。

Genome	ゲノム名を示します。リンクをクリックすると、ゲノムトップ画面 (図 3) へ移動します。
--------	--

RNA Sequence	RNA 配列が FASTA 形式で表示されます。
Upstream sequence	この遺伝子上流の配列を表示します。テキストボックスに数値 (半角数字) を入力し "Apply" ボタンを押してください。
Downstream sequence	この遺伝子下流の配列を表示します。テキストボックスに数値 (半角数字) を入力し "Apply" ボタンを押してください。
Ortholog	この遺伝子のオーソログ (パラログ) 情報がある場合、シンテニー解析画面 (図 14) へのリンクが表示されます。
COG	この遺伝子が COG (オーソログ) 情報を含む場合、COG ID を表示します。
COG Function	この遺伝子が COG (オーソログ) 情報を含む場合、COG functional categories を表示します。
KOG	この遺伝子が KOG (オーソログ) 情報を含む場合、KOG ID を表示します。
KOG Function	この遺伝子が KOG (オーソログ) 情報を含む場合、KOG functional categories を表示します。
PRINTS Category	この遺伝子が PRINTS ドメインと類似性がある場合に、PRINTS Category を表示します。
PRINTS	この遺伝子が PRINTS ドメインと類似性がある場合に、PRINTS 情報を表示します。
EC number	この遺伝子が EC number 情報がある場合に、EC number 情報を表示します。
Pfam	この遺伝子が Pfam ドメインと類似性がある場合に、PRINTS 情報を表示します。
Prosite	この遺伝子が Prosite ドメインと類似性がある場合に、PRINTS 情報を表示します。
TMs	この遺伝子が膜貫通領域と予測されたドメインがある場合に、そのドメイン数を示します。
GPI-anchor	この遺伝子が GPI アンカーとの結合が予測された配列を含む場合に、その配列の長さを表示します。
Signal Peptide	この遺伝子がシグナルペプチドと予測された配列を含む場合に、その配列の長さを表示します。

CDS 検索結果 (図 27) で初期状態 (デフォルト) で表示されるカラムの説明を以下にいたします。

Gene ID	遺伝子の ID を示します。NCBI 由来のゲノムデータの場合は、locus tag 名を示します。
Original description	データソースに記載されていた遺伝子に対する説明を示します。
Assigned description	SYGD データ登録時に再アノテーションされた遺伝子に対する説明を示します。
Genome	ゲノム名を示します。リンクをクリックすると、ゲノムトップ画面 (図 3) へ移動します。
Chromosome	染色体情報がある場合に、染色体番号を示します (染色体情報が無い場合には "Unmapped" と示されます)。
Links	外部データベースへのリンクです。
Protein Sequence	アミノ酸配列が FASTA 形式で表示されます。
Homologous genes	SYGD に登録された他のゲノムと相同性のある遺伝子を表示することができます。セレクトボックスから任意のゲノムを選択し、"Apply" ボタンを押すことで相同性のある遺伝子が表示されます。

CDS 検索結果 (図 24) で選択することにより表示されるカラムの説明を以下にいたします。

Description	
StartCodon	この遺伝子のスタートコドン、そのコドンから翻訳されるアミノ酸、遺伝子産物のアミノ酸配列長が表示されます。

CDS length	この遺伝子の CDS の配列長が表示されます。
CDS sequence	この遺伝子の塩基配列が FASTA 形式で表示されます。
Upstream sequence	この遺伝子の上流の配列を表示します。テキストボックスに数値 (半角数字) を入力し "Apply" ボタンを押してください。
Downstream sequence	この遺伝子の下流の配列を表示します。テキストボックスに数値 (半角数字) を入力し "Apply" ボタンを押してください。
Ortholog	この遺伝子のオーソログ (パラログ) 情報がある場合、シンテニー解析画面 (図 14) へのリンクが表示されます。
COG	この遺伝子が COG (オーソログ) 情報を含む場合、COG ID を表示します。
COG funtion	この遺伝子が COG (オーソログ) 情報を含む場合、COG functional categories を表示します。
KOG	この遺伝子が KOG (オーソログ) 情報を含む場合、KOG ID を表示します。
KOG Function	この遺伝子が KOG (オーソログ) 情報を含む場合、KOG functional categories を表示します。
PRINTS Category	この遺伝子が PRINTS ドメインと類似性がある場合に、PRINTS Category を表示します。
PRINTS	この遺伝子が PRINTS ドメインと類似性がある場合に、PRINTS 情報を表示します。
EC number	この遺伝子が EC number 情報がある場合に、EC number 情報を表示します。
Pfam	この遺伝子が Pfam ドメインと類似性がある場合に、PRINTS 情報を表示します。
Prosite	この遺伝子が Prosite ドメインと類似性がある場合に、PRINTS 情報を表示します。
TMs	この遺伝子が膜貫通領域と予測されたドメインがある場合に、そのドメイン数を示します。
GPI-anchor	この遺伝子が GPI アンカーとの結合が予測された配列を含む場合に、その配列の長さを表示します。
Signal Peptide	この遺伝子がシグナルペプチドと予測された配列を含む場合に、その配列の長さを表示します。

### 2-7-3. オーソログ検索

SYGD に登録されているゲノム情報の中から、オーソログ (パラログ) である遺伝子群 (オーソログクラスター) を検索することができます。

オーソログ検索画面を図28、コンティグ検索結果表示画面を図29 に示します。

オーソログ検索画面は、基本検索画面 (図 21) の "Ortholog" リンクをクリックすることで表示されます。

図28: オーソログ検索画面。基本検索画面の "Ortholog" をクリックする (図の矢印) ことで、オーソログ検索画面を表示させることができます。基本検索画面の表示方法については、2-7 基本検索参照。表中にある項目に検索条件を入れて、"Search" ボタンを押すと、条件に合致したオーソログクラスターのリストが表示されます。

図25 のオーソログ検索画面には、4 つの項目 (ID, Ortholog Set, Genome, Differential) があります。それぞれについて以下に説明します。これらは、組み合わせて検索することも可能です。

ID	オーソログ内部ID を入力して検索します。 (図 13 における ort1_c70_e100_7960 等の ID を入力します)
Ortholog Set	以下の Genome, Differential で用いるオーソログ計算方法をメニューから選択します。
Genome	ゲノムリスト画面 (図2) にあるゲノムが表示されています。それぞれのゲノムについて、P(positive)/N(negative)/-(don't care) が選択できるようになっています。
Differential	差分解析を行います。No. of group のメニューを選択 (1, 2 or 3) すると、ベン図が現れます (図32) ので、グループA, B (group=3 を選択したときは C も) に、同じグループに入れたいゲノムを選択します。Area では、右ベン図の領域を選択します。

Top / Search (Ortholog)

### Search

Select Target: Contig Gene RNA CDS **Ortholog** COG KOG EC

Search

Ortholog Search :

ID

Ortholog Set

Genome

All Positive

All Negative

All Ignore

Genome	P	N	-	Genome	P	N	-
S288c	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	K7	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Differential

Num. of group:

Ortholog list 1 - 50 of 5513 Rows: 50 Page: 1 Top Bottom Page: 1, 2, 3, 4, 5, 6

No./5513	ID	Ortholog Set	Genome count	Gene count	Annotation
1	<a href="#">ort1_1_4725</a>	ort1_1 (Bi-directional best hit)	2	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Repression factor of MSEs protein 1
2	<a href="#">ort1_1_5363</a>	ort1_1 (Bi-directional best hit)	2	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Uncharacterized protein YFR016C

図29: オーソログ検索画面。ID にあるリンクをクリックするとシンテニーマップ (図14, 詳細は 2-3-4 シンテニーマップの表示を参照) が示されます。Gene countの数値のリンクをクリックすると、選択した種でこのオーソログID を持つ遺伝子を検索した結果が表示されます。

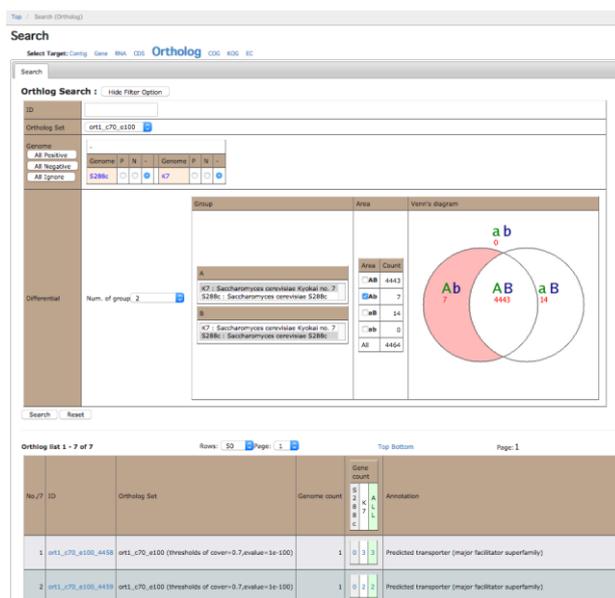


図 32: オーソログ検索画面 (Differential. 抜粋). Differential の項目で Num. of group を選択すると、上のようなベン図が現れます (この図ではgroup=2と選択したため、ベン図は図のようになる)。Group A, B としてゲノムを複数選択 (Ctrl キーを押しながらマウスクリック, で複数選択できる) します。Area 項目で Ab を選択し検索を実行しますと、図のように選択した領域 (ここでは Ab) が色つきで示されます。Area, ベン図に記してある数値はそれぞれの条件に合致したオーソログクラスターの数です。

注: 各領域に表示されているラベル以下のような意味である:

大文字: そのグループに存在している。

小文字: そのグループに存在していない。

例)

AB = (グループA に存在) かつ (グループ B に存在)

Ab = (グループA に存在) かつ (グループ B に存在しない)

aB = (グループA に存在しない) かつ (グループ B に存在)

ab = (グループA に存在しない) かつ (グループ B に存在しない)

#### 2-7-4. COG、KOG、EC 検索

SYGD に登録されているゲノム情報の中から、COG, KOG, EC アノテーションについての検索も可能です。

COG検索画面を図31に示します。KOGやEC検索画面も、ほぼ同じ仕様となっています。COG検索画面は、基本検索画面 (図21) の "COG" リンクをクリックすることで表示されます。同様に、KOG検索画面は "KOG", EC検索画面は "EC" をクリックすることで表示されます。

The screenshot shows a web interface for searching COG clusters. At the top, there's a navigation bar with 'Top / Search (COG)'. Below it, a 'Search' section has a 'Select Target:' menu with options: Contig, Gene, RNA, CDS, Ortholog, **COG**, KOG, EC. A red arrow points to the 'COG' option. The main search area is titled 'COG Search' and has a 'Hide Filter Option' button. It contains three main sections: 'ID' with a text input field; 'COG Functions' with a grid of checkboxes for categories A through W; and 'Genome' with a grid of checkboxes for genome names (e.g., S288c, K7) and P/N options. At the bottom, there are 'Search' and 'Reset' buttons.

図31: COG検索画面. 基本検索画面の "COG" をクリックする (図の矢印) ことで、COG検索画面を表示させることができます。同様にKOG検索画面は "KOG" を、EC検索画面は "EC" をクリックすることで、それぞれの検索画面を表示させることができます。基本検索画面の表示方法については、2-7 基本検索参照。表中にある項目に検索条件を入れて、"Search" ボタンを押すと、条件に合致したオーソログクラスターのリストが表示されます。

図31 のCOG検索画面には、3つの項目(ID, COG Functions, Genome) があります。それぞれについて以下に説明します。これらは、組み合わせて検索することも可能です。

ID	COG ID を入力して検索します。
COG Functions	COG functional categories で検索を行います。
Genome	ゲノムリスト画面(図2)にあるゲノムが表示されています。それぞれのゲノムについて、P(positive)/N(negative)/-(don't care) が選択できるようになっています。例えばゲノムeco1, eco2には存在し、cgl1には存在しない、bsu1には存在していても存在していなくとも良いといったオーソログクラスターを検索して表示します。

これらの検索画面は、オーソログ検索画面(図 29)と同様ですので、オーソログ検索(2-7-3 オーソログ検索) を参照してください。