

第九小組 DB checkpoint3

釀酒酵母 *Saccharomyces cerevisiae* 報告

組員：A1063329 王新賦、A1063315 許雅涵、A1063306 呂承恩

指導教授：楊子賢教授

目錄

目錄	1
一、生物簡介	1
二、ER 圖和 Relational Model 說明	1
二、資料庫設定與匯入資料	4
三、網頁架設	6
四、網頁使用說明	7
五、互評表	11

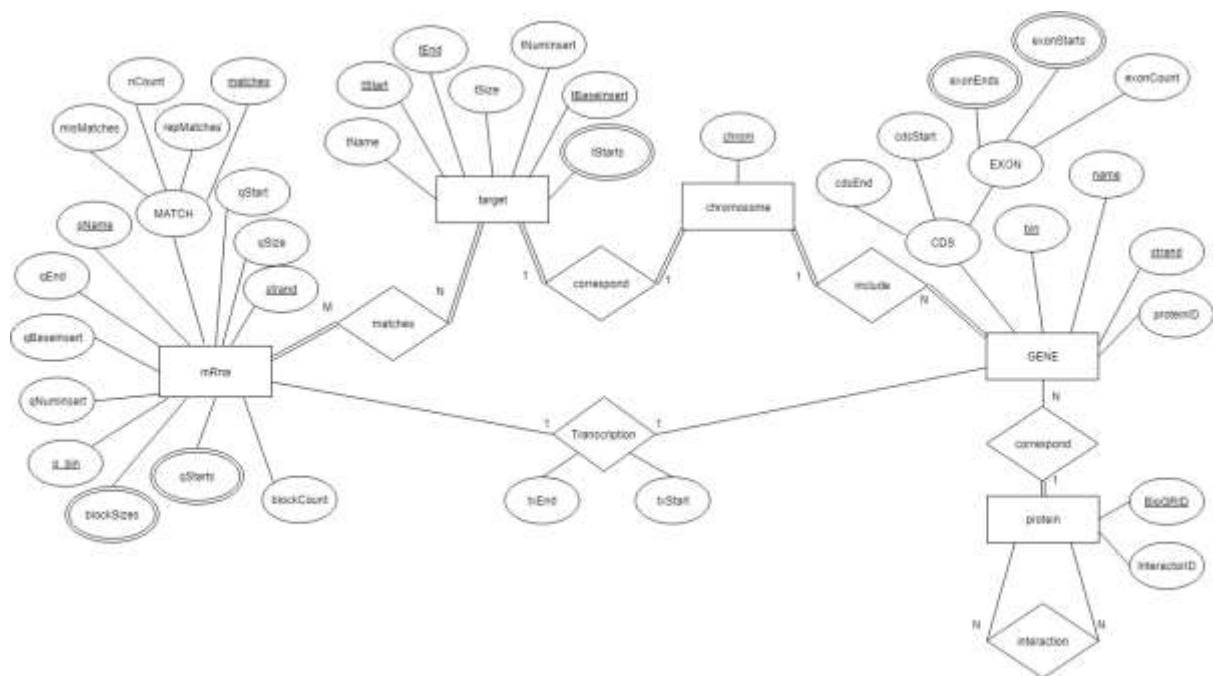
一、生物簡介

釀酒酵母的學名為 *Saccharomyces cerevisiae*，又稱麵包酵母、啤酒酵母，或者出芽酵母。它常被用在製作麵包和釀酒，而釀酒酵母因為常被營養食品廠商用來製成營養產品，所以也被稱為「營養酵母」。

真核生物中釀酒酵母是第一個完成基因組測序，因此常被用在研究真核細胞的遺傳學上。而釀酒酵母的基因組包含大約 12,156,677 個鹼基對、16 組染色體、6275 個基因，其中約有 5800 個具有功能，並大約有 31% 的基因與人類同源。

二、ER 圖和 Relational Model 說明

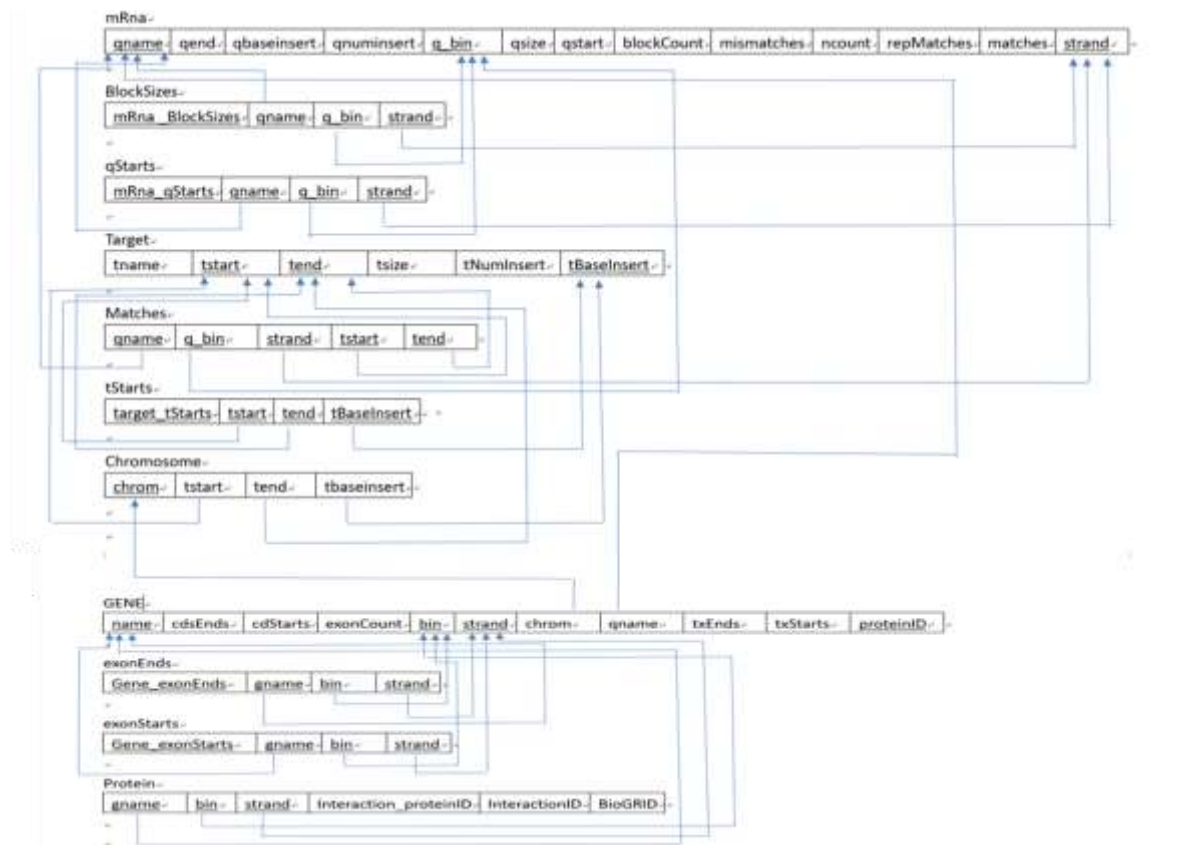
(1)cardinality notation



說明：

1. 基因裡包含 name(基因的名字)、strand(股)、bin(索引值)、CDS(編碼區)、proteinID(蛋白質)。而 CDS 包含 cdsStarts、cdsEnds、exon，分別為編碼區的開始、結束位置、外顯子。而外顯子又包含 exonCount、exonStarts、exonEnds，分別為外顯子的數量、外顯子開始和結束位置。而編碼區是由外顯子所組成，用來編碼蛋白質的部分。外顯子是基因的一部份，在經過轉錄後仍會被保留下來。
2. chromosome 的欄位包含了 chrom(染色體的名稱)，而基因是被包含在染色體之中，可以透過查找特定染色體，找到其所包含的基因。
3. 基因轉錄成 mRNA，並有記錄轉錄的 txStart(開始)和 txEnd(結束)位置。而 mRNA 欄位包含了 bin、qName、qSize、qStart、qEnd、match、qStarts、strand、qBaseInsert、qNumInser、blockCount、blockSizes，分別記錄其索引值、查詢名字、大小、查詢的開始位置及結束位置等等。
4. match 包含了 matches、misMatches、repMatches、nCount，可以用來尋找特定的 Target 值(亦可使用 mRNA 中的其他屬性尋找)。Target 會顯示找到的對應染色體，而 Target 欄位包含了 tname、tstart、tend、tsize、tstarts、tNumInsert、tBaseInsert，分別記錄其查詢結果所對應到的名字、大小、結果的開始位置及結束位置等等。
5. protein 的欄位包含了 BioGRID(實驗的編號)、InteractorID(參與蛋白質交互作用的蛋白質名字)，而 InteractorID 可以對應到基因中的 name，而在蛋白質交互作用之中 InteractorID 會和另一個 InteractorID 進行實驗。

(2)ER 圖轉 relational model



說明：

每個 entity 都獨立出來當作一個 table，而其裡面的屬性就是對應到的 table 欄位，若是屬性為複數，就也將其獨立出來當做一個 table。

1. mRNA 裡有 qname、qend、qbaseinsert、qnuminsert、q_bin、qsize、qstart、mismatches、blockCount、ncount、repMatches、matches、strand 這些欄位。
2. BlockSizes 因為是 mRNA 裡的複數屬性，因此需要獨立出來，而裡面有本身的 mRNA_BlockSizes 欄位，也因為他屬於 mRNA，因此要多加 mRNA 的主鍵 (qname、q_bin、strand)，指回去 mRNA 的 qname、q_bin、strand。
3. qStarts 同 BlockSizes 因為是 mRNA 裡的複數屬性，因此需要獨立出來，而裡面有本身的 mRNA_qStarts 欄位，也因為他屬於 mRNA，因此要多加 mRNA 的主鍵 (qname、q_bin、strand)，指回去 mRNA 的 qname、q_bin、strand。
4. target 裡有原本的 tName、tSize、tStart、tEnd、tBaseInsert、tNuminsert。由於 tStarts 是多值，所以將其額外建立一個表格，裡面包含了本身的 target_tStarts 欄位，以及 target 中的主鍵(tStart、tEnd、tBaseInsert)。
5. 由於 matches 是多對多的 relationship，必須將其獨立出一個 table，其可用於找到特定的 mRNA 和 target，並將 matches 加入 mRNA 的主鍵(qName、q_bin、strand)和 target 的主鍵(tStart、tEnd、tBaseInsert)。
6. chromosome 由於包含了基因且也可用 target 對應，所以將 chromosome 的主鍵 (chrom)加入到基因之中，而 chromosome 裡包含本身的 chrom，以及 target 的主鍵(tStart、tEnd、tBaseInsert)。

7. GENE 裡面有 name、cdsEnds、cdStarts、exonCount、bin、strand、chrom、qname、txEnds、txStarts、proteinID 這些欄位。其中因為 Gene 與 chromosome 為多對一的關係，若是為多對一的關係，就必須在多的那方新增對方的主鍵，因此在 Gene 裡再新增 chromosome 的主鍵(chrom)，指回去 chromosome 的 chrom。而因為 Gene 與 mRNA 為一對一關係，若是為一對一的關係，就在兩個 table 中擇一加入對方的主鍵，而我們是選擇在 Gene 裡新增 mRNA 的主鍵(qname)，指回去 mRNA 的 qname，也因為兩者的 relationship(Transcription)有額外的屬性，這些屬性也是可以擇一 table 加入，而我們也是選擇在 Gene 裡新增這些屬性(txEnd、txStart)。
8. exonStarts 因為是 Gene 裡的複數屬性，因此需要獨立出來，而裡面有本身的 Gene_exonStarts 欄位，也因為他屬於 Gene，因此要多加 Gene 的主鍵(gname、bin、strand)，指回去 Gene 的 gname、bin、strand。
9. exonEnds 同 exonStarts 因為是 Gene 裡的複數屬性，因此需要獨立出來，而裡面有本身的 Gene_exonEnds 欄位，也因為他屬於 Gene，因此要多加 Gene 的主鍵(gname、bin、strand)，指回去 Gene 的 gname、bin、strand。
10. Protein 中有 BioGRID 和 InteractorID，因為 InteractorID 會進行交互作用，所以加入了進行交互作用的兩個 InteractorID(InteractorA、InteractorB)，而 protein 要可以對應到基因，所以加入了基因的主鍵(gname、bin、strand)使其可以找到基因。

二、資料庫設定與匯入資料

(一)安裝環境

安裝 appserv (參考網址 <http://joe01032002.pixnet.net/blog/post/92365869-%E3%80%90linux%E3%80%91ubuntu-install-lamp>)，根據網站上的步驟進行安裝，讓我們可以用 phpMyAdmin。

(二)如何切割表格

(1)主表格的檔案轉檔

我們採用的是用 phpMyAdmin 去建立資料庫，首先先將 all-mrna.sql 及 sdgGene.sql 轉成 excel 的 xlsx 檔，接著將 all-mrna.xlsx 及 sdgGene.xlsx 分別轉檔為 all-mrna.csv 及 sdgGene.csv，但必須將 xlsx 檔裡的第一行欄位名稱刪掉後，再儲存為 csv 檔。其中 Saccharomyces_cerevisiae 的資料因為沒有 sql 檔，所以須先手動新增所需要的欄位再轉成 csv 檔。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Seq	1738	3	0	0	0	0	0	0	0	CG67118	1738	0	1739	del	230238	18677	181418	1738	0	1739	del				
2	586	127	0	0	0	0	0	0	0	0	K00948	127	0	127	del	230238	18677	180003	127	0	127	del				
3	586	125	0	0	0	0	0	0	0	0	K00952	125	0	125	del	230238	18677	180003	125	0	125	del				
4	586	127	0	0	0	0	0	0	0	0	K00948	127	0	127	del	230238	18677	185989	127	0	127	del				
5	586	125	0	0	0	0	0	0	0	0	K00952	125	0	125	del	230238	18677	185989	125	0	125	del				
6	586	72	0	0	0	0	0	0	0	0	K00171	72	0	72	del	133593	96967	96967	72	0	72	del				
7	586	72	0	0	0	0	0	0	0	0	K00171	72	0	72	del	133593	96967	96967	72	0	72	del				
8	586	4305	0	0	0	0	0	0	0	0	K11361	4305	0	4306	del	813184	14683	28278	4305	0	4306	del				
9	586	239	0	0	0	0	0	0	0	0	K11361	239	0	239	del	813184	14683	48234	239	0	239	del				
10	586	747	0	0	0	0	0	0	0	0	K11361	747	0	747	del	813184	14683	14683	747	0	747	del				
11	586	2948	0	0	0	0	0	0	0	0	K11361	2948	0	2949	del	813184	14683	18130	2948	0	2949	del				
12	586	2948	0	0	0	0	0	0	0	0	K11361	2948	0	2949	del	813184	14683	18130	2948	0	2949	del				
13	586	183	0	0	0	0	0	0	0	0	K11361	183	0	184	del	813184	14683	38887	183	0	184	del				
14	586	183	0	0	0	0	0	0	0	0	K11361	183	0	184	del	813184	14683	38887	183	0	184	del				
15	586	936	0	0	0	0	0	0	0	0	K11361	936	0	937	del	813184	14683	415214	936	0	937	del				
16	586	228	0	0	0	0	0	0	0	0	K11361	228	0	229	del	813184	14683	43666	228	0	229	del				
17	586	71	0	0	0	0	0	0	0	0	K11361	71	0	72	del	813184	14683	43666	71	0	72	del				
18	586	20	0	0	0	0	0	0	0	0	K11361	20	0	21	del	813184	14683	43666	20	0	21	del				
19	586	1738	0	0	0	0	0	0	0	0	K11361	1738	0	1739	del	813184	14683	43666	1738	0	1739	del				
20	586	140	0	0	0	0	0	0	0	0	K11361	140	0	141	del	813184	14683	43666	140	0	141	del				
21	586	140	0	0	0	0	0	0	0	0	K11361	140	0	141	del	813184	14683	43666	140	0	141	del				
22	586	848	0	0	0	0	0	0	0	0	K11361	848	0	849	del	813184	14683	56658	848	0	849	del				
23	586	2218	0	0	0	0	0	0	0	0	K11361	2218	0	2219	del	813184	14683	56658	2218	0	2219	del				
24	586	476	0	0	0	0	0	0	0	0	K11361	476	0	477	del	813184	14683	56658	476	0	477	del				
25	586	239	0	0	0	0	0	0	0	0	K11361	239	0	240	del	813184	14683	56658	239	0	240	del				
26	586	72	0	0	0	0	0	0	0	0	K11361	72	0	73	del	813184	14683	56658	72	0	73	del				
27	586	2738	0	0	0	0	0	0	0	0	K11361	2738	0	2739	del	813184	14683	56658	2738	0	2739	del				

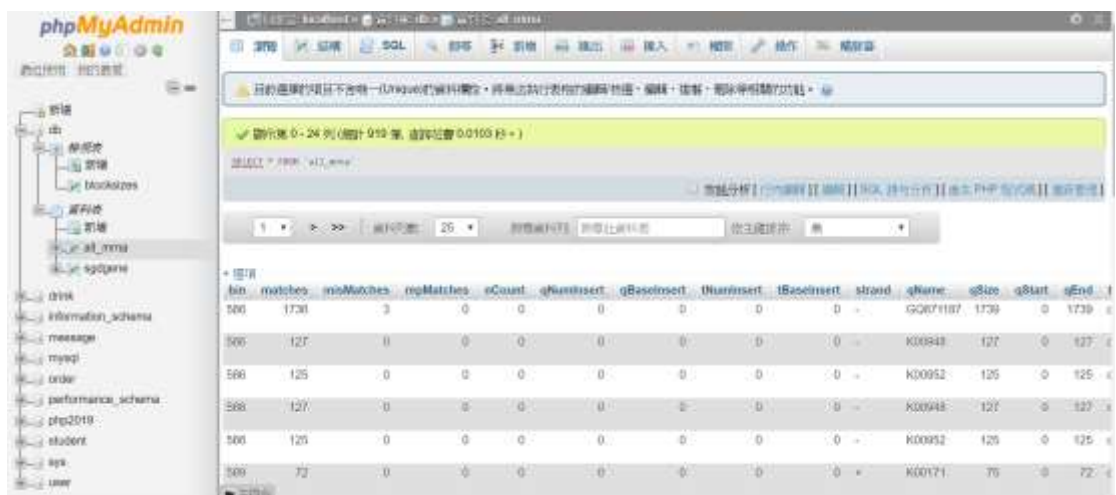
(2)主表格檔案的匯入及欄位的建立

接著在 phpMyAdmin 用匯入功能，先匯入 all-mrna.sql、sdgGene.sql、Saccharomyces_cerevisiae.sql，以便先將欄位建立起來。



(3)資料的匯入

將 all-mrna.csv、sdgGene.csv 及 Saccharomyces_cerevisiae.csv 檔分別匯入已經建立好的 all-mrna、sdgGene 及 Saccharomyces_cerevisiae 的資料表，即會將檔案內的資料值匯入進去。



(4)表格的切割

依據我們所畫的 ER 圖，並且將它轉為 Relational Model 後，可以得知要建立 mRNA、BlockSizes、qStarts、Target、Matches、tStarts、Chromosome、GENE、exonEnds、

exonStarts、Protein 這些 table。然後依據我們所觀察到的關係去下 SQL 指令將這些 table 連結。

(5) 切出來表格的檔案轉檔

先在原始資料表中下 SQL，找出我們要建立的 table 裡需要的欄位並且匯出，下面圖以建立 target 表為範例，其他資料表以此類推。然後刪除匯出後的 SQL 檔裡面多餘的值(如下圖)和修改 table 名稱(改成我們想要的名字)後，再重新匯入 phpMyAdmin 裡，即可建立。

```
1 SELECT tName,tBaseInsert,tName,tSize,tStart,tEnd
2 FROM target t ,mrna01 m
3 WHERE tend-tstart-tbaseinsert= qend-qstart-qbaseinsert
4 ;
```

all_mrna (8).sql - 記事本
檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

-- 點: ?轉碼?或?target?

```
CREATE TABLE `target` (
  `strand` char(2) NOT NULL,
  `qName` varchar(255) NOT NULL,
  `qSize` int(10) UNSIGNED NOT NULL,
  `qStart` int(10) UNSIGNED NOT NULL,
  `qEnd` int(10) UNSIGNED NOT NULL,
  `blockCount` int(10) UNSIGNED NOT NULL,
  `blockSizes` longblob NOT NULL,
  `qStarts` longblob NOT NULL,
  `bin` smallint(5) UNSIGNED NOT NULL,
  `matches` int(10) UNSIGNED NOT NULL,
  `misMatches` int(10) UNSIGNED NOT NULL,
  `repMatches` int(10) UNSIGNED NOT NULL,
  `nCount` int(10) UNSIGNED NOT NULL,
  `qNumInsert` int(10) UNSIGNED NOT NULL,
  `qBaseInsert` int(10) UNSIGNED NOT NULL,
  `tNumInsert` int(10) UNSIGNED NOT NULL,
  `tBaseInsert` int(10) UNSIGNED NOT NULL,
  `tName` varchar(255) NOT NULL,
  `tSize` int(10) UNSIGNED NOT NULL,
  `tStart` int(10) UNSIGNED NOT NULL,
  `tEnd` int(10) UNSIGNED NOT NULL,
  `tStarts` longblob NOT NULL,
) ENGINE=MyISAM DEFAULT CHARSET=latin1;
```

刪除部分

三、網頁架設

(1) 前置作業

輸入 `sudo apt-get install taskel`，安裝完成後再輸入 `sudo taskel` 開啟套裝軟體安裝程式，選則安裝 LAMP server，安裝途中會需要輸入設定 mysql 的密碼



完成安裝後再輸入 `sudo apt-get install phpmyadmin` 安裝 phpmyadmin 的 MySQL 網頁管理套件，過程中需設定 phpadmin 的 root 密碼，完成後再瀏覽器輸入 `140.127.220.175/phpmyadmin`，有顯示下面畫面代表成功，之後建立個人使用者。

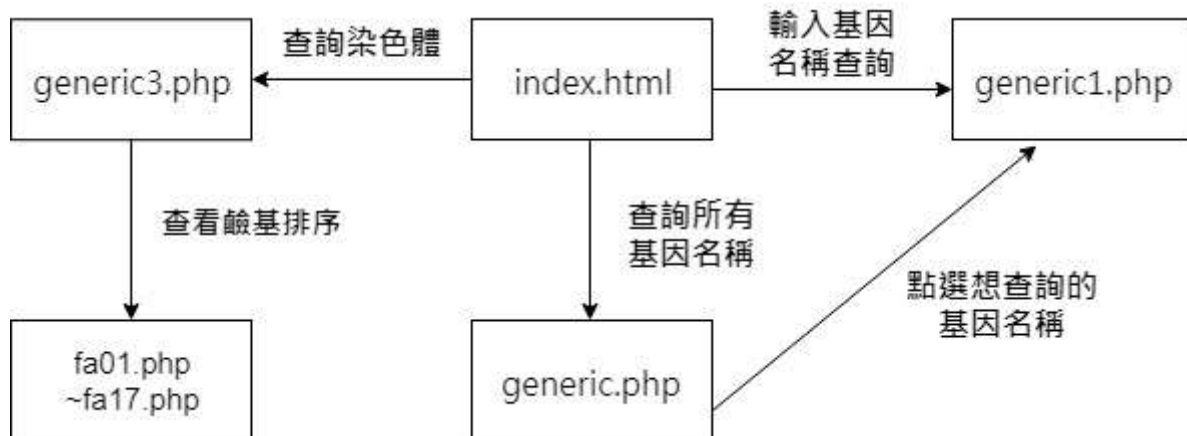


(2) 網頁編寫

進入主機的目錄到 `/var/www/html` 下編寫我們所要呈現的網頁，由於會有權限上的限制，所以在編寫時要加上 `sudo` 以取得權限，使用到的程式語言有 `php`、`css`、`mysql`，由於是編寫在 `index.html` 所以只需要在網址處輸入 `140.127.220.175` 即可到我們的主頁。因為為了編寫上的方便我們額外用 `sublimetext` 來編輯程式，由於我們組內沒有人修前端，所以使用網路上的模板來輔助我們網頁的呈現。

四、網頁使用說明

網頁流程架構圖：



網頁介紹：

(1) 首頁

右上角分別為: 首頁、基因查詢、欄位說明、染色體查詢的連結。首頁可以直接利用基因名字去查詢相關資訊且點選右上角的首頁連結就會回到此頁面。



(2) 基因查詢

點選右上角的基因查詢連結會列出釀酒酵母的所有基因名字，讓使用者可以直接點選欲查詢的基因名字。點選後會顯示該基因的相關資訊。

釀酒酵母的所有基因

共有6887個基因

Q0001	Q0002	Q0003	Q0004	Q0005	Q0006	Q0007	Q0008	Q0009	Q0010
Q0011	Q0012	Q0013	Q0014	Q0015	Q0016	Q0017	Q0018	Q0019	Q0020
Q0021	Q0022	Q0023	Q0024	Q0025	Q0026	Q0027	Q0028	YAL001C	YAL002W
YAL003W	YAL004W	YAL005C	YAL006C	YAL007W	YAL008W	YAL009C	YAL010W	YAL011W	YAL012W
YAL013C	YAL014C	YAL015C-A	YAL016C-B	YAL017W	YAL018W	YAL019C	YAL020W	YAL021W-A	YAL022C

Q0010的相關資訊

Gene											
bin	gene-name	chrom	strand	hstart	hvmid	cdstart	cdend	proteinid	exoncount	exonStarts	exonEnds
585	Q0010	chr12	+	3981	4338	3951	4338	his	1	3951,	4338,

Target							
qName	pName	tSize	tStart	tEnd	tBaseInset	tMutInset	tSigns
K00948	chr1	230218	160274	160401	0	0	*160274,
K00952	chr1	230218	160278	160401	0	0	*160278,
K00948	chr1	230218	165862	165889	0	0	*165862,

(3)欄位說明

點選右上角的欄位說明連結會列出相關資訊會看到的所有欄位，其所對應的中文說明，包含：Gene、mRNA、Target、Protein 的相關欄位。

Gene

be: 是
 gene-race: 基因有種
 stream: 溪流
 strand: 和江面正相反向
 lecter: 轉錄的粒體
 bend: 轉錄的結束位置
 robust: 強健有活力的
 robust: 強健的結晶性
 primed: 進行轉錄
 increment: 增加數目
 wordList: 外顯子開始位置
 wordInfo: 外顯子結束位置

mRNA

點選右上角的染色體查詢連結會列出所有染色體其對應的鹼基序列，點選後會顯示完整的鹼基序列排序。

共有17種序列

[illegible]

五、互評表

王新賦

	評分	理由
王新賦	5	編寫網頁的前端、整理寫 word
呂承恩	5	處理資料庫如何連結主機、如何下 sql 來選出我們要的資料
許雅涵	5	編寫前端網頁和後端，和修改 word

呂承恩

	評分	理由
王新賦	5	建立前端網頁、word 以及蒐集情報和拋頭露面
呂承恩	5	編寫資料庫連結及其他的 php 檔和安裝 apache
許雅涵	5	編寫前端網頁，處理 word 中有修改的細節

許雅涵

	評分	理由
王新賦	5	前端的網頁製作、word
呂承恩	5	完成資料庫連結的 php 檔、word
許雅涵	5	前端的網頁製作、word