

第九小組 DBcheckpoint1 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

一、生物簡介

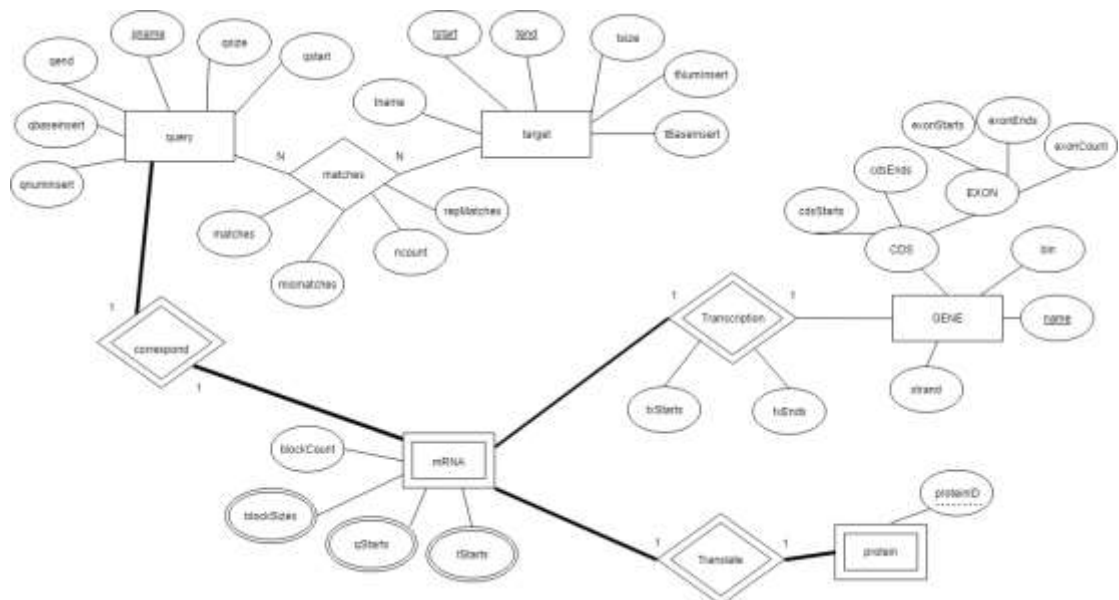
簡介：

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」。

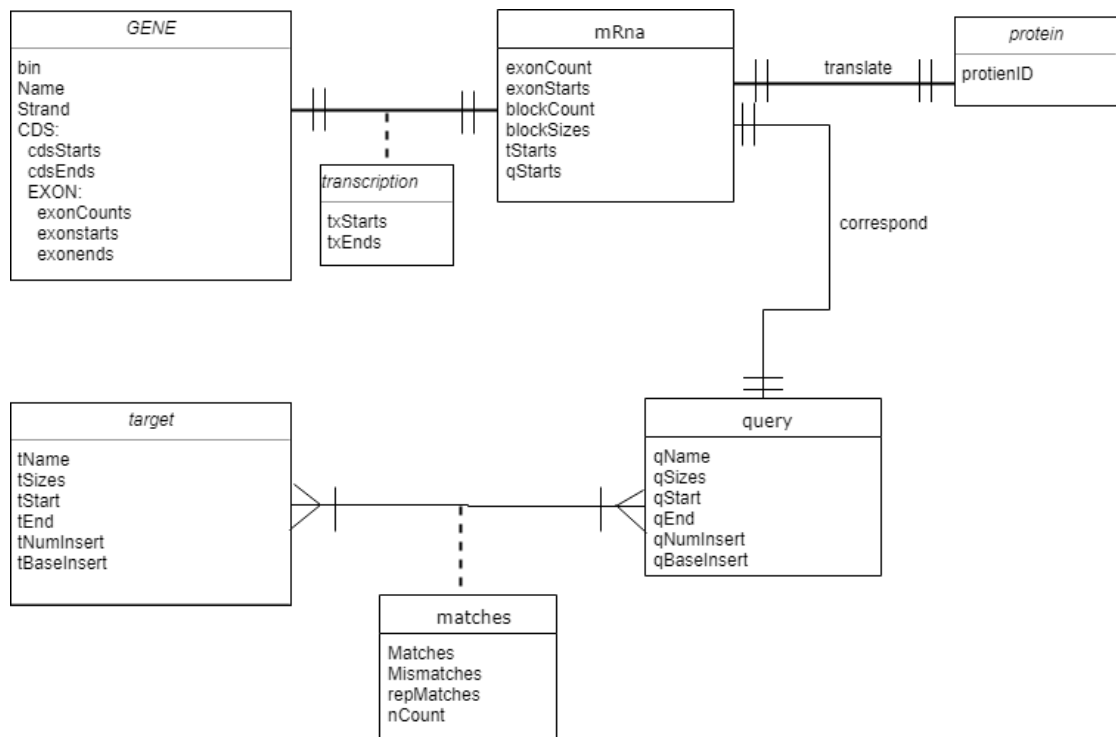
真核生物中釀酒酵母是第一個完成基因組測序，因此常被用在研究真核細胞的遺傳學上。而釀酒酵母的基因組包含大約 12,156,677 個鹼基對、16 組染色體、6275 個基因，其中約有 5800 個具有功能，並大約有 31% 的基因與人類同源。

二、ERmodel 和 relational model 的說明

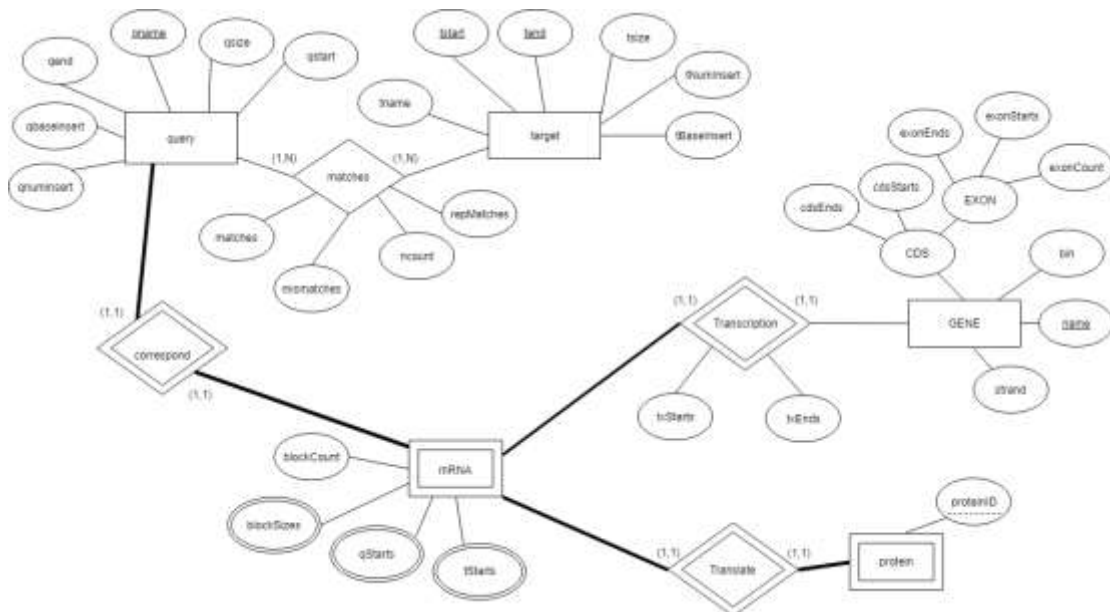
1. Chen's notation w/ cardinality notation



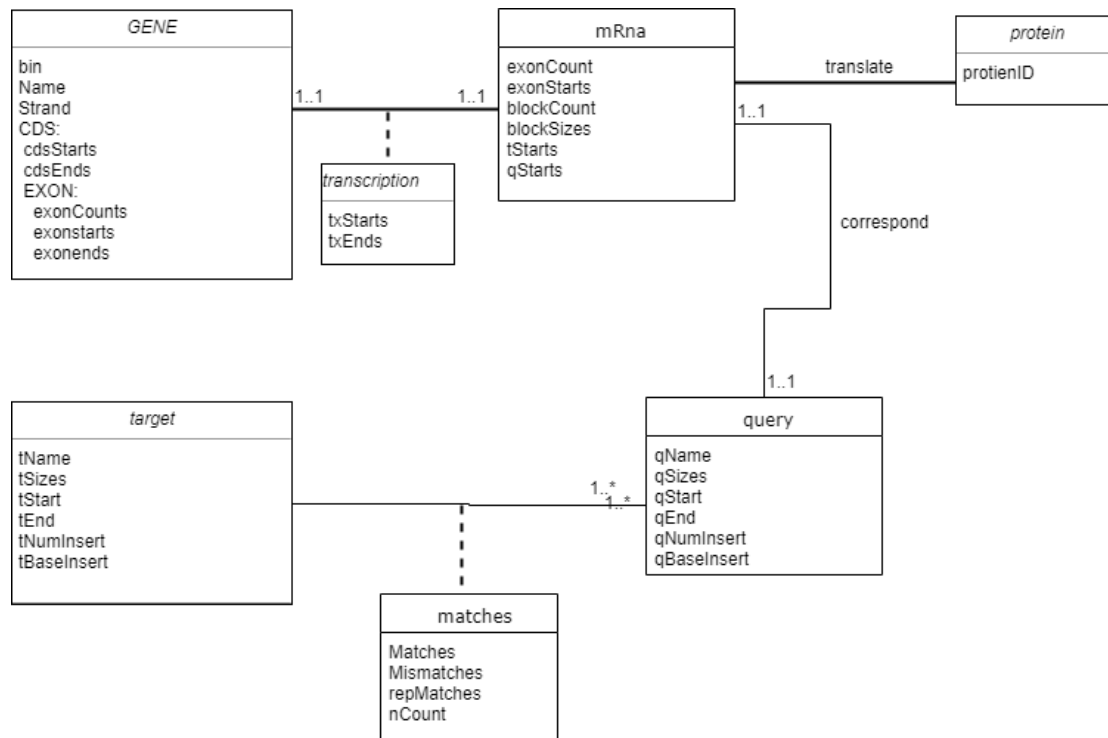
2. UML w/ Crow's notation



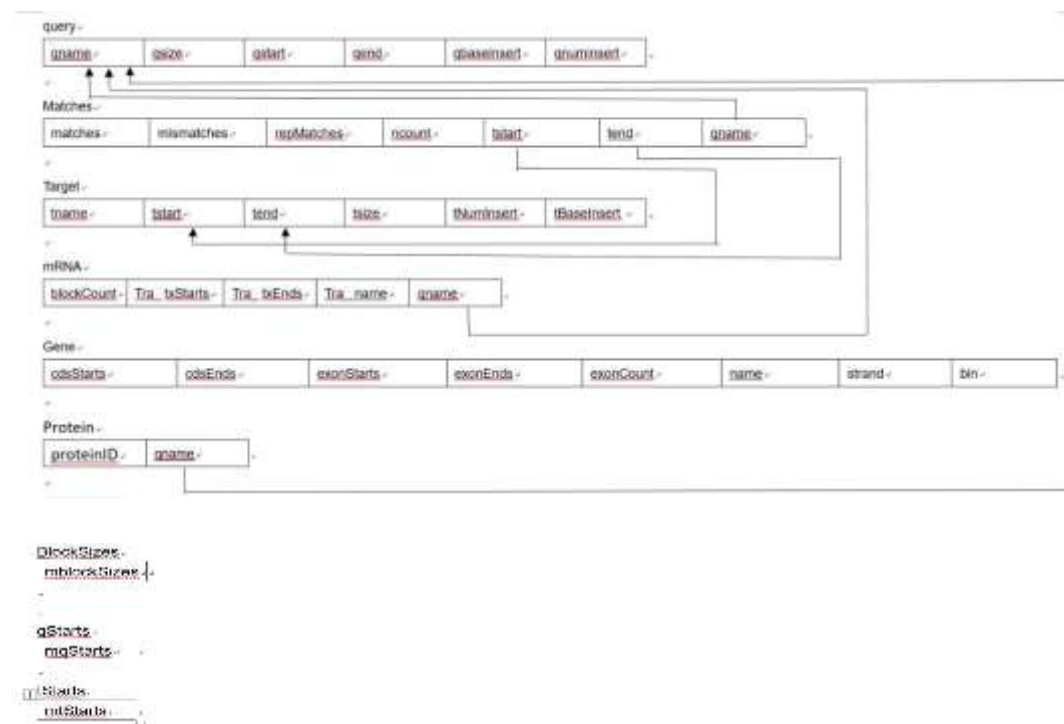
3. Chen's notation w/ (min, max) notation



4. UML w/ (min, max) notation



5. relational model



說明：

1. 基因裡面包含 name(DNA 的名字)、strand(股)、bin、CDS(編碼區)。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. 基因轉錄成 mRNA，並有記錄轉錄的開始和起始位置。而 mRNA(弱屬性)必須有基因轉錄後才會存在。mRNA 裡面有 blockCount、blockSizes、qStarts、tStarts，分別為對應中的 block 數量、block 的大小、block 在 query 裡起始位置、block 在 target 裡的起始位置。
3. mRNA 透過轉譯，才能生成蛋白質，而蛋白質有 proteinID 記錄其蛋白質 ID 名稱，且蛋白質(弱屬性)必須有 mRNA 轉譯後才會存在。
4. 用 Query 來查詢對應的 mRNA 序列，而 Query 欄位包含了 qname、qsize、qstart、qend，分別記錄其查詢名字、大小、查詢的開始位置及結束位置。Target 會顯示找到的對應染色體，而 Target 欄位包含了 tname、tstart、tend、tsize，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。Query 和 Target 中間的關聯以 match 表示，且其對應與 matches、mismatches、ncount、qbaseinsert 有關。

三、問題討論

1. 一開始對於各欄位的作用不是很了解，必須上網查詢相關的生物知識才能比較了解。
2. 在研究 all_mrna 的資料時，發現在 query 查詢這部分， $qend - qstart = matches + mismatches + ncount + qbaseinsert$ ，但 $qend - qstart$ 不一定等於 $qsize$ 。
3. 而在查詢到後 target 顯示的結果中，若是 tsize 一樣，則 tname 也會一樣，同樣的在 query 的部分，Qname 若是一樣，其 qsize 也會一樣。
4. 在研究 sgdGene 時，外顯子 exonend-exonstarts 值若相同，則蛋白質 ID 會一樣。
5. 若 blockCount 為 1 的時候，則 tStarts 加上 blockCount 會等於 tend。若 blockCount 有複數值，則 tStarts 也會有複數值，假設 blockCount 為 2，那就另設 blockSizes1 和 blockSizes2，還有 tStarts1 和 tStarts2，則 tStarts1 加上 tBaseInsert 和 blockSizes1 會等於 tStarts2，tStarts2 再加上 blockSizes2 會等於 tend。
6. $Tend - tstart - tbaseinsert = matches + mismatches + ncount$
7. $Tend - tstart - tbaseinsert = qend - qstart - qbaseinsert$
8. 在第一次時的 ER 圖時，我們為關係 matches 建立表格，但是第二次 ER 圖時，我們把表格拿掉，因為我們發現原本在 matches 上的屬性可以設給實體 mRNA，這樣同時也能節省空間。

9. 我們把原本的實體 mRNA 刪掉，把原本的實體 query 改名 mRNA，因為

四、互評表

✓ 王新賦

	評分	理由
王新賦	5	認真討論和畫圖
呂承恩	5	認真討論和畫圖
許雅涵	5	認真討論和畫圖

✓ 呂承恩

	評分	理由
王新賦	5	積極參與、畫圖
呂承恩	5	積極參與、畫圖
許雅涵	5	積極參與、畫圖

✓ 許雅涵

	評分	理由
--	----	----

王新賦	5	努力了解生物觀念和畫圖
呂承恩	5	努力了解生物觀念和畫圖
許雅涵	5	努力了解生物觀念和畫圖