黑色字體是同學的問題；綠色字體是我的答覆。
答覆劉同學：

1. 為什麼 suffix array 乍看之下像是 quadratic space?

每個 entry 只要記住每個 suffix 的起始位置即可，不必記錄整個 suffix。

2. BLAST takes into account only those w-mers, whose number is roughly mn/4^w for DNA sequences 這個值是怎麼來的？

因為 DNA 序列有 4 個符號，每個點是個 match 的機率約 1/4，每個點開始的斜對線長度 w 都是 match 的機率約 1/4^w，共有 mn 個點，所以 w-mers 個數約 mn/4^w。

3. 從 seed 開始，找尋失分在 Xg 以內的 gapped-BLAST 是說從 seed 開始，畫出 Xg 的範圍，如果在這個範圍內有另外一段 HSP 的端點，就把 HSP 跟原本的 seed 連在一起嗎？

這是好問題，若有另一段 HSP，會連在一起。被連進來的 HSP，會記錄起來，下次它就不必重複處理了。

4. BLAT 把 database 裡面的 sequence 分割成不重疊的 k-mers，這樣跟 query sequence 比較的時候不就有可能漏掉一些了嗎？

是的，所以它又加入 seed 允許一個 mismatch 的寬鬆方式。

答覆陳同學：

在關於 find mininum set of haplotypes 的部分
最後面有教到，把 problem 變為 linear 的方法
summantion Wjrt = 1
這邊不需要>=1 原因是因為只要有一組符合就可以
但是如果回到
summation Xr*Xt >=1
這個式子一樣是 = 1 就可以了。
為什麼必須要 >=1

仔細想了一下似乎是因為如果用

summation Xr*Xt >=1 在解的時候必須跑過所有的可能
所以會找出所有解，而 summation Xr*Xt >=1 則是有一組解就停了。
請問我這樣理解是否正確?
因為剛剛仔細想了一下還是有點 confused 為什麼前面必須>=1
後面只要=1 就可以了

請老師指教了

這是很好的問題。若限定 Xr*Xt=1，則當有兩組的 Xr Xt 值都是 1 時就爆掉了。
但這情況在 Wjrt = 1 是 OK 的，因為其中一組的 Wjrt 設為 0 亦無礙它們的 Xr 及 Xt 可以為 1。