

## DNA 數據存儲與科技發展的終極目標

### 一、在宇宙中留下人類的痕跡

現在開始，倘若世界像<<新石紀>>動畫一樣，遭遇了足以石化人類 3700 年的石化光線，3700 年後，人類的紀錄還剩下什麼？你可能會想說網路的數據會保留，但資料幾乎都是存放在硬碟裡的，無論 HDD 還是 SSD，甚至更高端的技術，最多也就保存 30 幾年，人類的紀錄將會歸 0，當然也可能像動畫中描述的一樣，一些金屬製成的佛像、藝術品可能還有一點痕跡可以保留，但時間若再更久，結果可能就不一樣了。

(下圖為:如果人類被石化 3700 年後，少數能保存的東西(更久之後就不一定了))




人類的科技追求除了了解自身來自何處外，也有一個終極目標，我們如何留下在宇宙中存在過的痕跡，倘若未來人類滅絕，高科技的外星文明經過地球時，也能依照我們留下的數據知道我們人類的存在，但開頭也說過，目前人類的資料存取技術，最久也只能保留幾個金屬製品，甚至數據都將消失，保守估計人類面絕後，必須 100 多年內有外星人來訪，才有可能連接上我們的數據，這是一種對於”在宇宙中留下紀錄”這件事最大的挑戰。

## 二、DNA 數據儲存

DNA 研究的歷史已過五十幾年，我能舉出來最早成功進行 DNA 資料存放的例子，2012 年美國哈佛大學 George Church 教授，他和研究團隊成功將一本 52,000 字的書籍轉換成 DNA 數據，以防有人不知道，在此介紹一下，電腦資料儲存是以 0、1 的二進位方式進行儲存，原理什麼的要講一節課的時間，在此就不多贅述了。而這群研究員們以 DNA 中的 A - 腺嘌呤 (Adenine)、T - 胸腺嘧啶 (Thymine)、C - 胞嘧啶 (Cytosine)、G - 鳥嘌呤 (Guanine) 為基底，以四進位的排序方式儲存資料。

(下圖分別為:以 ATCG 表示 0、1(左); 以防不懂如何存放數據的示意圖(右))

Previous	0	1	2
T	A	C	G
G	T	A	C
C	G	T	A
A	C	G	T



大家應該都知道基因是一段段的連接起來的雙螺旋結構(A、T 一組、C、G 一組)，每一段都能儲存一個訊息，這是一個驚人的發現，要知道，人類當前硬碟最大儲存量，是希捷科技所推出的 Seagate Exos XMozaic 3+，容量為 30TB，而假設一個 DNA 有硬碟大小，則能夠存放 21,500PB 的數據，倘若你無法體會 21,500PB 有多大，舉個例子:GPT 說一張 1920\*1080 的圖片，平均大小是 1MB，對單位不熟我可以列舉出來:

1 bit: 能表示 0、1

1 byte = 8bits

1 KB = 1024 Byte

1 MB = 1024 KB (假設一張 1920\*1080 圖片大小為此)

1 GB = 1024 MB

1 TB = 1024 GB

1 PB = 1024 TB (21,500PB 給你參考)

1 EB = 1024 PB

1 ZB = 1024 EB

1 YB = 1024 ZB

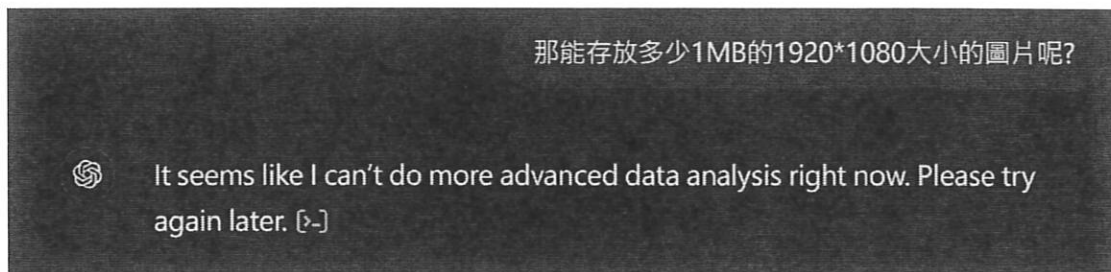
1 RB = 1024 YB

1 QB = 1024 RB

而 21,500PB 大約等於 20.9961EB 的大小，能夠存放  $2.15 \times 10^{13}$  張圖片，是否對這大小沒有概念，我給你個概念，這數字大到連 GPT 都無法具體算出來給我看的程度了。(當前日常運算最高至:  $2^{63}-1$

(9,223,372,036,854,775,807))

(下圖為:數字太大，GPT 表示無能為力)

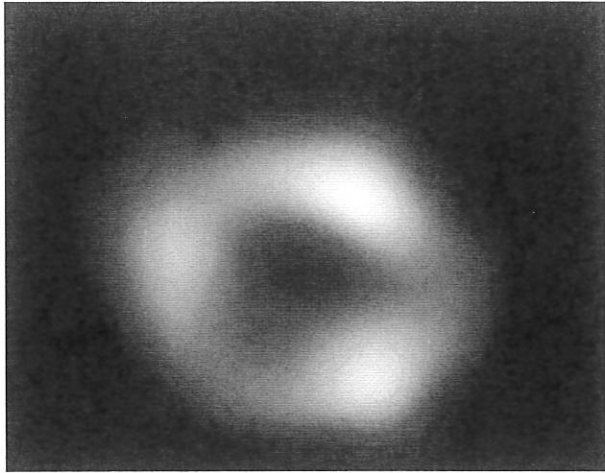


### 三、為何選擇用 DNA 進行資料存儲

除了那龐大的數據存量，我已經用非常直觀的方式呈現出來，現在來述說，選擇 DNA 的最主要原因，就如開頭所說:人類的其中一個終極目標是”在宇宙中留下自己的痕跡”，題外話:數據保存時限最大期望值是能保存至太陽系撞上銀河系中心的黑洞之前，這是數據能保存的最久時間，無論如何，最久就這樣了，當然，這時間久到人類可以忽略不計。我們的目標是能夠存放越久越好，目前 Seagate Exos X Mozaic 3+能保存 7 年，保守估計給他延長個 100 年，在宇宙中的時間有如滄海一粟，而科學家發現，良好的 DNA 存放，最少至 1000 年~最多至 999 萬年左右都是有可能的，這讓外星高科技文明發現我們的可能性上升了幾百萬倍，同時也讓<<新石紀>>的男主角不用再為了數據遺失而苦惱沒有資料保存，

當然男主角也得先再次創造出 DNA 讀取技術才行，呵呵。

(下圖為:銀河系的中心，名為人馬座 A 星的黑洞，太陽系未來”可能”會撞上此黑洞，但時間是天文數字，不是人類應該擔心的問題)



#### 四、當前實驗與應用

現在知道了人類有能力存放數據至 DNA，以及選擇使用 DNA 存放數據的理由，現在來看看目前的研究概況：

2012 年-George Church 將 52000 字的書籍轉換為 DNA 數據

2013 年-歐洲生物資訊研究所(EBI) 的研究人員與 Church，詳細介紹了超過 500 萬位資料的儲存、檢索和複製。所有 DNA 文件都以 99.99% 到 100% 的準確度再現了資訊。

2013 年-Gupta 和同事開發了一款名為 DNACloud 的軟體，用於將電腦檔案編碼為其 DNA 表示形式。

2015 年-蘇黎世聯邦理工學院的研究人員在一篇文章中報告了 DNA 編碼資料的長期穩定性。

2016 年，Church 和 TRI 發表研究，儲存了 22 MB 的壓縮影片序列並從 DNA 中恢復，發現序列的恢復具有零誤差。

2017~2020 年略過

2021 年，Krasnogor 的大學研究團隊使用 DNA 實現了堆疊資料結構，他們

使用雜交和鏈置換來記錄 DNA 中的訊號，然後以相反的順序釋放訊號。研究表明，類似資料結構的操作在分子領域是可能的，強調了基於 DNA 的計算系統的潛力。

## 五、一位修女的疑問

1970 年-尚比亞的一位修女(瑪莉·尤肯達)詢問了 NASA 太空航太中心的科學副總監一個問題:「目前地球上還有那麼多孩子吃不上飯，你怎麼捨得為遠在火星的項目，花費數十億美元。」，雖然在此處該修女是針對宇宙天文的研究進行質疑，但我個人認為地球上的每個普通人對於科技發展都會有一種疏離感，認為與其把項目資金投入於科技研究，不如全部拿來維持生計，這疑問也能拿來套用在人類花費資金研究 DNA 資料存儲技術上。而當時的 NASA 副總監給出了以下回答:「我不是不知道有很多小孩吃不上飯，無論哪個國家，政府都有撥出預算用在救助項目上，意味著:國家會救助，但不會把所有的錢全部投入救助項目上。」

救助只救急不救窮，很多窮人你給他錢，沒多久他花完了，還是一樣窮，這彷彿是一個無底洞，把錢全部投入無底洞中，絕對不是一個英明的國家領導人該做的事，因為國家的錢還得用在醫療、教育、城市建設、交通運輸、國防、環保、農業等各方面，科技發展雖然不能直接提供食物解決飢荒問題，但它所帶來的大量新技術可以被民用化，經不完全統計，每年有大約 1,000 項新技術被民用化，譬如:醫院的核磁共振技術、電腦 CT 斷層，嬰兒用的尿布濕...，都是在宇宙探索時所得來的新技術。(by. youtuber 司文)

NASA 副總監以及該名 youtuber 的理解也讓我進行了深度思考，倘若幾百年來，世界都只拿錢去進行救助項目，現在的人類科技八成還是跟百年前沒什麼區別，甚至退步，DNA 存儲技術不只是為了留下人類存在過的痕跡，也能提供人們數據上傳再也不用擔心空間太大問題為其一目標，像我自身電腦有好幾 TB 的資料想進行存儲，但無奈於硬碟大小限制，只能進行刪除或上傳至付費平台後刪

除，DNA 存儲的大容量對人類是有益的，即使不是全部。

最後講一句我想述說的話：「人類追求科技就是為了追求人類的便利，可以讓人類不用再為一些重複性過高的瑣事，而拖住人類發展、追求自我，甚至讓人有更多的時間做自己想做的事。」

-The End-

六、資料來源：

- 1.維基百科(英文版): DNA digital data storage
- 2.youtube 百強排行科學精選頻道:未來 DNA 將成為數據存儲工具！10 億部高清電影只占橡皮擦大的空間
- 3.youtube TED-Ed: Is DNA the future of data storage? - Leo Bear-McGuinness
- 4.youtube 老鳴 TV: DNA 本質是儲存器，人類是行走的 U 盤，我們只是宇宙的工具 1 老鳴 TV
- 5.youtube 司文的不思議世界:【天文】登月無用嗎？不如把錢拿來改善民生？
- 6.GPT 幫忙翻譯與運算