

地球上最早的文字

李家同

暨南國際大學資訊管理學系

E-mail:rctlee@ncnu.edu.tw

很多人會以為，地球上的生物中，只有人類有文字，人類雖然存在很久了，但是有文字，只有五千年的歷史，前些日子在歷史博物館中所展覽的美索不達尼亞的文明古物中，可以看到五千年前的古老文字。

人類有文字，是一件非常特殊的事，別的動物，最多只有一種語言，這些語言也非常簡單，比方說，狗會搖尾巴，猩猩會做一些手勢，牠們都沒有文字的，其實人類中也有很多民族是只有語言，沒有文字的。

至於植物，我們都有一個認識：植物是沒有語言的。其他更低等的生物，像細菌，我們更加不能想像它們有語言了。

其實，地球上的所有生物，從細菌到人類，都有文字的，大家都知道，所有生物都會傳宗接代，而且生物的下一代都要和上一代相似。牛的後代必須是牛，西瓜的種子生出來的應該是西瓜。可是我們如何能告訴我們的下一代我們的特徵，顯然我們必須有一種語言，甚至必須有一種毫不含糊的文字，使我們的下一代能夠明確地知道他們的行為應該是如何。

地球上的生命，起始於三十億年前，因此我們生物的文字也有三十億年之久，生物中關於遺傳所用的文字有四個字母，這四個字母就是 A、C、G、T，所有有關遺傳的指令都是用這四個字組成的，下面就是一段如此的文字：

AAAGAAACTGCTGCTGCTAAATTTGAACGT.

任何文字都需要一種工具，才能將文字表現出來，古時候人類大概用刀來雕刻文字，除此以外，任何文字必須用一種媒體將文字紀錄在上面，我們生物的 A、C、G、T，這四個字母，表現的方法是最高級的，它們是用四種化學合成物來表示的，這四種合成物就是 Adenine(A)，Guanine(G)，Cytosine(C)和 Thymine(T)，在化學界，這四種合成物都是所謂的 nucleic acid。

什麼意思呢？當自然界要寫一個 A 的話，他就去用一個 Adenine

來表示 A，如果要寫一個 C，自然界就去拿一個 Cytosine。

在我們中學裡教化學的時候，老師會告訴我們某某化學合成物的功能，而這種功能都是與化學有關的，我們從來沒有想到一個化合物可以用作一種符號。

我們人類最近常將文字寫在紙上，忽然存到電腦的記憶體裡，自然界呢？自然界是將文字存進了生物的細胞裡，以我們人類為例，我們細胞裡的遺傳文字長達三億個字，大多數放在細胞核內的染色體裡，也就是說，如果我們打開染色體，會發現三億多個 A、C、G、T，每一個 A、C、G、T 都對應一個 nucleic acid。這一連串的 A、C、G、T 就是 DNA。

我必須在這裡強調，不論動物、植物，甚至細菌或微生物，都用同樣的四個字母，也都用同樣的化合物。

最後，我要解釋的是：究竟 A、C、G、T 代表什麼？我這樣說吧！A、C、G、T 主要的任務是要細胞製造不同的蛋白質，蛋白質決定我們是何種生物，我們的一切生命現象都由蛋白質來表現的。

蛋白質由什麼組成的呢？蛋白質由二十種 amino acid（氨基酸）組成的，因此所有 A、C、G、T 就是要表示這些不同的氨基酸。自然界有一種編碼的方法，舉例來說，TTT 表示一種叫做 phenylalanine 的氨基酸，而 GAG 代表一種 glutamine 的氨基酸，假設我們看到 TTTGAG，我們就知道我們所要製造的蛋白質裡有 phenylalanine 氨基酸，後面跟著 glutamine 的氨基酸。

以我們現代電腦來說，個人有個人的編碼，以中文來說，我們常用 Big Five 編碼（繁體中文），也有大陸常用的簡體中文編碼，還有一種 Unicode 的中文編碼，我們常常收到一封電子郵件，因為弄不清楚編碼方式而煩惱不已，但是自然界沒有這種問題，自然界在三十億年來，統一使用同樣的編碼方式，打開任何一種生物的細胞，看到 GAG，永遠知道這代表 glutamine。

一連串的 A、C、G、T 因此就是代表一連串的氨基酸，每一連串的 A、C、G、T 叫做基因，每一個基因對應一連串的氨基酸，而每一連串的氨基酸對應一個蛋白質。

每一個細胞都要讀取若干基因來製造若干蛋白質，問題是：如何知道基因的結束呢？大自然的解決方法是用 TAA，TGA 和 TAG 來代表基因的結束。

自從人類有了程式以後，就有 end of program 的觀念，看到 end of program，就知道程式在此結束，其實，在三十億年前，大自然早就懂得這個觀念了。當細胞讀到 TAA，TGA 或 TAG，就知道基因就此結束了。

每一個細胞，只要製造它所需要的蛋白質，因此它不能讀錯了基

因，一旦讀錯了基因，就會製造無用甚至有害的蛋白質，大自然對每一個基因，都加了鎖，每一個細胞只有正確的鑰匙，凡是它不該讀取的基因，它就無法打開。

我們常會感到驕傲，但很多偉大的科學家都很謙虛，我想這是因為知道大自然的奧秘以後，一定覺得自己是很沒有學問的。順便提一句，DNA 的編碼還是用的 error correcting code 呢。