DNA 序列浮水印技術之強化設計

林琨祐　陳建銘　林韓禹　白敦文*
國立臺灣海洋大學資訊工程學系
*twp@mail.ntou.edu.tw

摘要

次世代定序儀大幅度降低定序實驗的難度與花費，隨著資料的大量產出，如何在共享研究資料的同時，保護原研發單位定序的研究成果乃是一個重要的課題。因此本研究提出一套基於比較基因體學的理論基礎且強調具有生物意義的去氧核醣核酸序列浮水印演算法，透過與模式物種基因組序列的比對，系統得以避免將數位浮水印嵌入在具有生物意義的區域上。程式並透過加密演算法、分段嵌入與反方向嵌入等技巧提升浮水應整體的強健性。實驗結果顯示本方法可以透過個人電腦有效率地完成染色體基因序列長度的浮水印加密，而且經過加密之序列具有高度的強健性及容錯能力。本數位浮水印加密技術可以提供生物學家在分享所定序之序列同時提供額外的安全保障。

關鍵詞：DNA 浮水印、Playfair Cipher、資訊安全。

Abstract

With advance development of NGS technologies, challenges and expenses of sequencing experiments are greatly reduced. In the big biological data world, one of the most important issues is how to share sequencing results with preventing the risk of plagiarism. In this research, we proposed a DNA watermarking technique by emphasizing the subject of biological significance. After comparing the target sequence with suitable genome sequences from model species, our proposed system reduced the ratio of embedding watermarks within the biological regulatory elements. The proposed method also applied several techniques including Playfair encryption, segmentation insertion, and dual-direction insertion for increasing the robustness. Experimental results have shown that our proposed methods could be efficiently applied on sequence size at a chromosome-level only by a general personal computer equipment, and the high robustness and fault tolerant ability against random-mutation attacks were also demonstrated by practical applications. The proposed digital DNA watermarking technique provides biologists an extra secure protection layer when sharing their own biological sequence data in public domain.

Keywords: DNA Watermark, Playfair Cipher, information security.

1. 前言

次世代定序儀為生物序列定序技術帶來革命性的進步，使得序列資料目前正以爆炸性的速度累積。隨著序列資料量的成長，其資料所有權的議題也日益重要。除了具有高度商業價值及專利保護的各種基因改造作物(Genetically Modified Organism, GMO)，多樣性的生物序列也因為學術單位的研究成果發表及論文出版而開始有了所有權的爭議及顧慮。原本就預計公開的序列資料，也可能因為序列遭事先的不當使用，而使得原定序者無法得到任何應有的功勞。然而序列本身並不具備原創性，縱使序列遭到剽竊使用也難以申訴。這樣的隱憂，嚴重削弱序列擁有者將所定序的序列資料上傳至公開資料庫或與他人共享資源的動機。縱使各種優秀的演算法可以用於加密序列以達到完整的保護防竊，但是這些加密過的序列可能嚴重破壞原先生物序列的特性，卻也無法再被任何生物資訊工具進行註解分析或施行具有應用價值的實驗驗證，無法發揮巨量序列資料的所有附加價值，甚為可惜。

避免數位資料被剽竊的方法之一，即是為資料加上不可見的浮水印，並透過解密程序來進行資料使用歷程的追蹤及認證。所謂的浮水印，是指透過某些方式隱含特定資訊在原始資料，乍看之下難以察覺，卻可以透過某些特定方法還原回其隱藏資訊的處理方式。相對於常見的加密方式為將原始資料完全隱藏起來的事前預防，浮水印技術則更加適合在內容無法或不願隱藏的狀態下，卻不願被剽竊盜用的事後追蹤[1]。去氧核醣核酸(DNA)浮水印，是一種將資訊隱含在 DNA 序列中，透過特殊的方式編碼並修改部分的核甘酸，藉此追蹤序列的不當使用等資訊。其序列隱含的資訊則可能包含擁有者、作者、定序時間或其他資訊編碼等可以用來驗證原始序列的所有權。DNA 浮水印常被應用於基因改造作物或是人工合成的序列中[2]。其中一個有名的例子，則是由 J. Craig Venter Institute 在 2010 所發表的人工合成基因組細菌，即在四段浮水印 DNA 中，紀錄數十位參與計畫者的姓名、三句名言、甚至包含了一個簡單的網頁。由於 DNA 序列本身的文字字串性質，因此傳統上使用空間域與頻域互相轉換的浮水印技術並不適用於此。目前已發表的浮水印技術如 DNA-Crypt，則是以透過類似傳統影像資料浮水印儲存於最低重要位元(Least

Significant Bit, LSB)的特性，將其資料隱藏在編碼區的第三碼，藉由其蛋白質編碼的同義置換，來放置其隱含資訊[3]。針對透過該種編碼方式將隱含資訊置放於啟動子區域(promoter region)並進行細菌的活體實驗，結果發現加入浮水印在編碼區[4]或是在啟動子區域[5]皆不影響基因功能的正常表現。隨著生物恐怖攻擊事件危機的發生，此類技術更被建議用在具感染特性病原體的相關研究之中[6]。

隨著我國將生技產業視為行政院重點發展產業，產官學界均開始採購次世代定序儀器並投入相關定序研究的今日，可以預見各單位對序列的保全需求將日益加重。因此，本論文的目的即在於提供一套具生物意義的 DNA 序列浮水印演算法，在使用者進行定序實驗後所產出之 DNA 序列中加入隱藏浮水印資訊的同時，並保持不破壞該序列原本可能具有之基因區域、基因調控因子或是生物標記等具有生物意義重要資訊，期能不影響使用者後續進一步的分析與研究。這樣的技術應用可以做為日後序列的原始擁有權認證，杜絕蓄意剽竊他人研究成果的情形。更可以進一步鼓勵產官學界對定序資料成果進行資源共享，提升我國在生技方面的產業競爭力與學術研發能量。

2. 方法與實作

2.1 系統流程圖

本論文提出的方法可分為嵌入浮水印以及擷取浮水印兩個系統模組。在嵌入浮水印的技術，使用者需要定義要加入的浮水印訊息與一個用於加密浮水印訊息的密鑰。在嵌入浮水印的演算法部分又可細分為三個模組：首先，為了保存加入浮水印序列後原始序列具有的生物意義區域，系統針對原始的明文序列進行可能具有生物功能區域的搜尋並標記，再將可能具有功能片段以外的區域定義為容許嵌入人工浮水印序列的區域。其次，為了使隱藏的資訊加入序列中，這些資訊的內容必須以 DNA 的方式包裝。同時，為了使浮水印的內文保持隱密性，以避免被輕易背去除或竄改，系統透過改良的 Playfair 演算法，對浮水印訊息進行加密，令攻擊者在不知道密鑰的情況下，無法得知正確的浮水印內容。最後，則依照上述兩個模組的輸出，由第三個鑲嵌模組進行資訊的嵌入動作，以產生具有浮水印資訊同時又可以盡量保有原生物功能的加密程序。詳細的嵌入浮水印系統流程圖在圖 1 中呈現。

在擷取浮水印資訊的系統模組，其基本原理是加入浮水印程序的反向操作，唯一不同的是先前定義的生物功能區域可以沿用加密過程中所產生的結果，因此不需要重新進行序列搜尋。透過輸入具備生物意義區域的位置清單、使用者所設定的密鑰以及具有加密的浮水印序列，系統將會自動找出原先嵌入訊息的位置，並透過 Playfair 的解密，將原始的浮水印訊息還原，並可同時復原原始未加密的序列。詳細擷取出浮水印的系統流程圖則在圖 2 呈現。

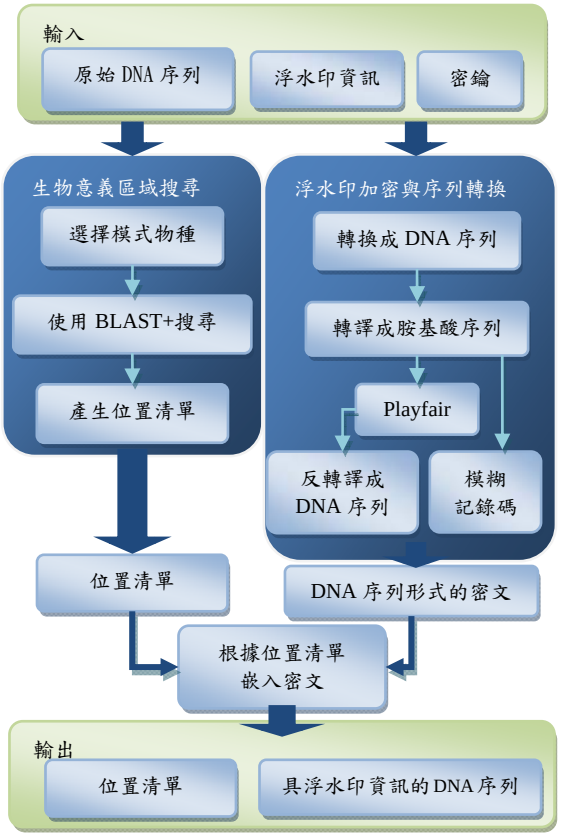

圖 1：嵌入浮水印系統流程圖。

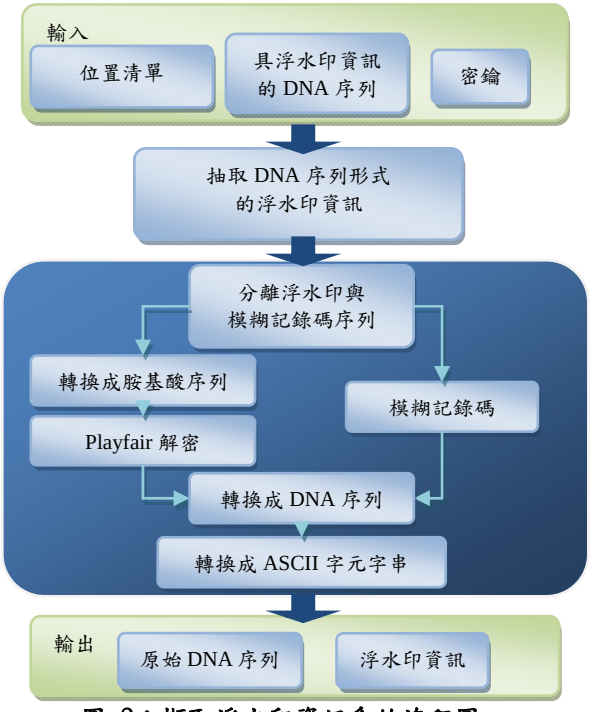

圖 2：擷取浮水印資訊系統流程圖。

2.2 生物意義區域搜尋

為了使加入浮水印的序列依然保留原先可能具備生物意義的區域，使該加密序列依然可以用於比較基因體學、轉錄因子調控分析、親緣演化分析等生物應用，本系統所提出方法的第一步，即是標示可能具有潛在生物意義之調控區域，並將這些區域在浮水印序列的鑲嵌步驟時進行排除。然而，使用者的原始序列可能來自尚未被研究過的新興物種，如何在不具備有任何標註資訊的序列上找出具有生物功能的區域將是第一個挑戰。過去的研究指出，透過跨物種比較生物序列共同保留的部分，可以有效地偵測序列中的調控因子，這也正是比較基因體學的核心概念[7]。因此，本系統將針對使用者輸入的序列，對已知完整序列的模式物種進行相似子序列的比對搜尋，並將所搜尋到的相似子序列標註為可能具有生物意義的區域。在模式物種基因體序列資料的收錄，本研究使用歐洲分子生物實驗室(Ensembl)資料庫中第 72 版的 DNA 序列，該資料庫包含 61 個不同的模式物種，使用者可以依照目標定序物種與已知模式物種的親緣演化關係進行模式物種的選擇。在選定模式物種後，系統使用 BLAST(Basic Local Alignment Search Tool)演算法，進行相似序列區域的比對搜尋。在實作方面，本研究使用由美國國家生物資訊中心 NCBI 所發表的 BLAST+工具組中的 blastn 工具，對模式物種與輸入序列的 DNA 序列進行搜尋[8]。經序列比對搜尋，系統將從 BLAST+的輸出取得一份與模式物種具備相似區域的位置清單，這份位置清單將提供後續浮水印鑲嵌的座標依據。

2.3 浮水印加密與序列轉換

由於 DNA 序列的性質，浮水印資訊需經過一連串的轉換與加密才能嵌入到 DNA 序列。本論文將使用由 Mona Sabry 在 2010 年所發表的基於 Playfair 加密演算法的 DNA 序列訊息轉換方式進行實作[9]。以下先針對該演算法進行精簡的描述。DNA 序列是由 A、T、C、G 四種不同鹼基所構成的字串，可以知道該類字串資料的儲存，每一個鹼基至少需使用兩個位元的資料進行儲存。該演算法首先將浮水印訊息的每個字元根據其 ASCII 編碼，轉換成對應的位元流(bitstream)，再將該位元流中以每兩個位元為基本單位進行取出，並根據表 1 轉換成 DNA 的序列。由表 1 知道前兩個位元的不同組合可以直接代表 A、T、C、G 四種不同鹼基的組合，而其中模糊（Ambiguity）記錄則是為了提供胺基酸序列轉譯回 DNA 序列時、面對一對多情形提供精準反轉譯的依據，模糊記錄詳細的使用時機將在下一節說明。

表 1、位元與模糊記錄碼轉 DNA 表。

第一位元	第二位元	模糊記錄	DNA
0	0	0	A
0	1	1	C
1	0	2	G
1	1	3	T

表 2、修改過具唯一對應特性的 Codon Table。

Codon	模糊記錄			
	0	1	2	3
A	GCT	GCC	GCA	GCG
R	CGT	CGC	CGA	CGG
N	AAT	AAC		
D	GAT	GAC		
C	TGT	TGC		
Q	CAA	CAG		
E	GAA	GAG		
G	GGT	GGC	GGA	GGG
H	CAT	CAC		
I	ATT	ATC	ATA	
B	TAA	TGA	TAG	
U	AGA	AGG		
Z	TAC			
L	CTT	CTC	CTA	CTG
K	AAA	AAG		
M	ATG			
F	TTT	TTC		
P	CCT	CCC	CCA	CCG
S	TCT	TCC	TCA	TCG
T	ACT	ACC	ACA	ACG
W	TGG			
Y	TAT			
V	GTT	GTC	GTA	GTG
O	TTA	TTG		
X	AGT	AGC		

為了使浮水印序列的內容保持隱密性，以避免被惡意使用者可以直接偵測、甚至進一步的竄改所有權資訊，因此對序列使用隱藏式的加密技術是一項必要的步驟。基於生物序列的特性，演算法使用 Playfair 加密演算法作為核心。Playfair 是一種對稱式加密法，其金鑰為 25 個英文字母所構成一個 5*5 的置換矩陣，該矩陣通常可透過任意使用者指定的片語所產生。在加密過程中，將預計要加密的訊息以兩個字元為一組的方式，根據此置換矩陣進行替換。該加密方式除了具有高效率特性外，由於人體內的蛋白質約由 20 餘種不同的胺基酸所組成，因此該加密技術亦十分適合以胺基酸為主的蛋白質序列使用。在該演算法中，所有欲加密的 DNA 序列將會先轉譯成對應的蛋白質序列，並透過 Playfair 加密後，再轉換回原始對應的 DNA 序列。在標準的生物轉譯過程中，DNA 轉譯成蛋白質胺基酸會由三個 DNA 核苷酸轉換為一個胺基酸的規則進行，而不同的核苷酸組合有可能會轉譯到同一種蛋白質胺基酸，因此在進行蛋白質反轉譯過程會發生一對多的模糊(Ambiguity)轉換的問題。為了克服此一問題，演算法透過記錄轉譯蛋白質時所對應的形式

編號進行克服。但是，在自然的脊椎動物蛋白質轉譯表中，可能有多達六個不同 DNA 核苷酸的組合可以轉譯到同一個胺基酸，為了節省浮水印所需要的序列空間，演算法使用一個經過人工修改的特殊編碼表來進行轉譯與反轉譯的唯一對應機制，使其每一個胺基酸對應機制的可能性皆不超過四種組合，如此即可以透過單一一個 DNA 字元進行儲存。表 1 記錄模糊對應編號的轉換位元標示，表 2 則為修改過的編碼表。

2.4 浮水印序列的鑲嵌

在經過前面所述兩個子模組運算之後，進行浮水印加密的最後一步即為加密浮水印序列與原始序列的鑲嵌組合。透過生物意義區域的比對搜尋結果，程式自動避開可能具有生物功能或影響生物功能的區域進行嵌入作業。透過排除由 BLAST 所確認與模式物種高度相似的序列區域，我們可以得到許多分散在序列中與模式物種無法匹配的區段，這些區段被定義為可寫入浮水印資訊的區域。在統計所有可寫入區段的座標位置之後，系統可依寫入區域註解資訊及區域數量平均分為前後兩個部分，如果可寫入區域的總數為奇數，則依照其座標忽略最中間可寫入區段。再針對目標序列的前後兩部份依相反方向及相對應的可寫入區段，進行相同加密內容但方向相反的浮水印密文嵌入動作，如此的設計可以提升加密資訊在目標序列的均勻分布特性並加強序列浮水印的強健性與隱密性。

為了避免惡意使用者透過剪切攻擊，亦即捨棄某一段連續序列的方式來刪除浮水印資訊，本系統對欲加入的浮水印密文進行分段嵌入[10]。首先程式依照序列前半部可寫入區域的正向區段進行統計分析，將浮水印密文序列依照長度平均切割，切割數量選擇與前半部分可寫入區域的數量相同，唯為了避免雜訊攻擊，每一段嵌入的浮水印密文片段不得少於 3 個鹼基，若是無法達到此一標準則減少分段的數量，並放棄部分可寫入區域進行嵌入的過程。當切割完畢後，系統在每一個可寫入區域的起始位置進行密文嵌入。最後針對序列後半部的可寫入區域，以反方向的順序重複進行一遍浮水印密文片段的嵌入，以完成序列加密的所有步驟。最後，系統記錄在嵌入過程所有隱藏在原始序列的座標資訊並輸出至序列擁有者，以提供未來解密或確認之用。

2.5 浮水印序列的擷取與解密

若要針對已加入浮水印的密文序列進行浮水印的擷取與解密，並進一步還原最原始的生物序列，則使用者需要同時輸入密文序列、加密程式所產生嵌入位置清單、以及加密時所設定的密鑰。其解密流程基本上就是加密流程的反向操作，唯解密過程並不需要再透過 BLAST 工具進行生物意義區域搜尋，因此速度可以更有效率。在輸入嵌入位置清單後，系統自動將加密過的密文擷取，並合成一段完整的密文 DNA 序列，此時的序列即為原始不含任何浮水印資訊的明文序列。從密文 DNA 序列中，程式將自動分離反轉譯所需要的模糊序列與用來加密的胺基酸序列編碼；此時，使用者所指定的密鑰可以產生與加密相同的矩陣，並可以透過反向運算進行 Playfair 解密。經 Playfair 解密後，所得到的胺基酸序列可以依照模糊資訊進行 DNA 序列的反轉譯，最後再轉回 ASCII 序列，即可以得到原始的浮水印資訊。最後，上述所有步驟需要針對正向與反向可寫入區域的解碼，同時在不同方向進行兩次解密動作，如此即可得到兩組同步解出的浮水印資訊。

3. 結果與討論

3.1 浮水印序列加密範例展示

表 3、浮水印訊息內容、序列轉換及 DNA 密文展示。

	內容	長度
密鑰	Hello World	10
浮水印資訊	Kyl Good Day 2013-08-1323:00:00	30
DNA 序列	CAGTCTGCCGTACACTCG TTCGTTCGCAAGAACACA CGACCTGCATAGATAAAT ACATATAGTCATAAATGA AGTCATACATATAGAAAT AGATATATGGATAAATAA ATGGATAAATAA	120
密文 DNA 序列	TCTCACTAGATCAGTGTT ACAATTGGTCAGAATAAC TTTCCCTCAATGGATGAA TATAAGTGTGTTTGCAAA TCTGTACAAGAACTTATC TAAAACAGAACTTATCTG TACACATAAGAAGTTATA CAGGTATGTATCTAAAAA GGTATGTATCTAAAAA	160

為了驗證本論文所提出方法的可行性，我們透過 Java 語言進行觀念驗證，該加解密系統是使用單機程式的設計開發。為示範本研究將明文的密鑰與浮水印資訊轉換成 DNA 序列的過程，表 3 顯示一個簡單的密鑰與浮水印資訊，透過前面所述之 Playfair 演算法進行加密的運算。在本範例中，使用的密鑰字串為“Hello World”，而使用的浮水印資訊為“Kyl Good Day 2013-08-1323:00:00”，長度為 30 個字元。表格的第三行為直接透過 ASCII 轉換的 DNA 序列。因為 ASCII 字元是由 8 個位元進行一個字元的表示，而 DNA 每一個鹼基僅有兩個位元的資訊含量，因此從 ASCII 字串轉換成 DNA 序列的話，其長度可預期會成長 4 倍，亦即本範例 30 個字元的浮水印資訊需 120 個 DNA 鹼基字元。表格的第四行為經過 Playfair 加密後的 DNA 序列。經過 Playfair 的方式加密後，因為每 3 個 DNA 就會額

外儲存一個模糊位元的數字，加上 Playfair 演算法中所可能產生的額外空間消耗，因此序列會比單純轉換為 ASCII 字串要來的更長。以本範例來說，密文與明文的長度比率約為 5.3 倍。其實際倍率可能因序列與密鑰內容不同而有所浮動，但是可預見差距不會有過大的情形。

3.2 實際生物序列浮水印應用測試

除了上述所展現演算法細節的範例外，本論文為了展現所提出的方法可以實際應用在真正的生物序列上，實際使用常被生物學家作為研究樣板的模式物種序列進行測試。在本項測試中，我們分別使用稻田魚(Oryzias latipes)與金娃娃(Tetraodon nigrovirdis)的第一條染色體序列，作為輸入的明文序列。此範例是選擇以斑馬魚(Danio rerio)做為比對生物意義區域的模式物種。所有上述的序列皆從 Ensembl FTP 第 72 版下載。

針對這兩條明文序列，我們先與斑馬魚的全基因組序列進行比對，先定義所有可寫入區域的座標資訊。在確認所有可寫入區域後，分別使用兩個長短不同的浮水印訊息進行加解密測試，訊息內容分別是(1) “Tun-Wen Pai”及(2) “Tun-Wen Pai Department of Computer Science and Engineering National Taiwan Ocean University”。密鑰則為 Hello World。表 4 為詳細測試所需的執行時間與加密序列長度等結果。其中時間的單位一律為秒。本研究所使用的設備為一般個人電腦，詳細的規格：Intel i7-920 中央處理器、6G DDR3 記憶體、 7200 轉 SATA2 硬碟機、及 Eclipse 開發環境進行範例測試。測試結果顯示主要的時間瓶頸仍在於 BLAST 比對過程，但是整體執行時間已證明本方法僅依靠個人電腦等級的硬體，即能夠有效率套用於實際的生物基因組規模序列中。此外，對執行時間的主要影響因素主要是取決於生物序列的總長度，浮水印序列的長度對執行時間僅有小部分的影響。

表 4、實際生物序列長度、可加密區塊數量、及加解密所需電腦計算時間。

		稻田魚	金娃娃
	序列總長度 (bps)	39973033	22983436
	BLAST 所需時間	728.443s	56.170s
	可寫入區域個數	2846	1195
密文 1	**浮水印嵌入長度**	60	60
	浮水印嵌入時間	4.995s	1.421s
	浮水印解密時間	9.609s	5.647s
密文 2	**浮水印嵌入長度**	516	516
	浮水印嵌入時間	4.937s	5.026s
	浮水印解密時間	10.374s	6.146s

3.3 強健性測試

為證明本方法所加入之數位浮水印具備高強度及抗雜訊的特性，本研究更針對前節所述之兩條序列進行強健性測試。在強健性測試中，我們使用前節所使用之序列、浮水印資訊、及相同的密鑰。但是在進行解密前，我們將加密過的序列，以千分之一的機率進行隨機 DNA 字元替換，再測試由前後兩端是否還可以解出浮水印資訊，若是任何一個字元與原來的浮水印資訊不同即為失敗。圖 3 為四組資料的強健性測試結果。從結果可以發現當浮水印序列越長時，由於嵌入序列的長度等比例增加的關係，較容易受到雜訊攻擊的影響。但是即使是在長度達 516 個字元的浮水印序列下，在稻田魚的序列亦保有 90%的召回率，在金娃娃魚的序列亦有 93%的召回率。針對短浮水印資訊的測試狀況下，都可以達到 100%的召回率。從測試結果可以觀察到經由其中一端所提供正確召回的比率極高，證實由本論文所提出之兩端反向進行嵌入加密訊息的方式確實可以有效的提升整體強健性。

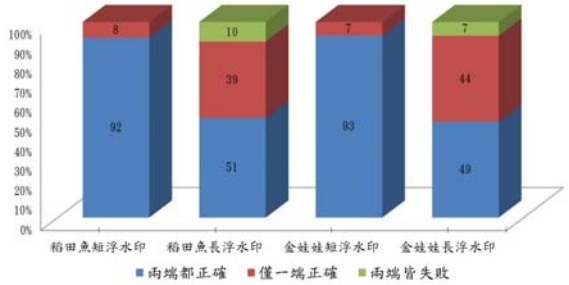

圖 3：強健性測試結果。使用兩組測試序列及兩組不同長度的浮水印訊息內容進行加密測試，對已經加入浮水印的序列中分別隨機插入原長度 1/1000 的隨機突變雜訊。將含有隨機雜訊序列進行解密還原後，以短浮水印訊息加密的序列可以完全被解密及辨識，但嵌入長浮水印訊息及含隨機雜訊的序列則分別有 10%及 7%的辨識失敗率。

4. 結論

本研究提出一個針對 DNA 序列加入浮水印資訊的強化技術。除了應用前人所提出以 DNA 與胺基酸作為基礎的 Playfair Cipher 加密法外，本研究更透過比較基因體學的方式，辨識目標序列與模式物種序列相似區域，進行嵌入浮水印資訊的排除區端參考，便免可能具備生物意義的區域受到浮水印的破壞。此外，本研究更進一步的改良前人的浮水印嵌入技術，透過頭尾兩端反向進行平均分散式密碼嵌入的設計，提升整體加密與解密的強健性及容錯力。透過實驗結果顯示本方法不僅可以在有限的運算資源下以合理的時間套用至全基因組等級的序列，其強健性及容錯能力測試更顯示出該方法可以有效的抵抗隨機突變或是區域整段刪除的攻擊。此方法將可以作為生物學家未來在開放共享自

己研究所產生序列時，可以提供額外的保障，更有機會進一步成為生技產業中用以保護智慧財產的有效依據，以保障相關學術與產業應用的發展。

誌謝

本論文承蒙國立台灣海洋大學海洋中心與行政院國家科學委員會之計畫經費贊助，計畫編號為 NSC 102-2321-B-019 -001，僅此誌謝。

參考文獻

[1] M. Stamp, *Information Security: Principles and Practic*: wiley, 2006.

[2] D. Heider, *et al.*, "Watermarking sexually reproducing diploid organisms," *Bioinformatics*, vol. 24, pp. 1961-2, 2008.

[3] D. Heider and A. Barnekow, "DNA-based watermarks using the DNA-Crypt algorithm," *BMC Bioinformatics*, vol. 8, p. 176, 2007.

[4] D. Heider and A. Barnekow, "DNA watermarks: a proof of concept," *BMC Mol Biol*, vol. 9, p. 40, 2008.

[5] D. Heider, *et al.*, "DNA watermarks in non-coding regulatory sequences," *BMC Res Notes*, vol. 2, p. 125, 2009.

[6] D. C. Jupiter, *et al.*, "DNA watermarking of infectious agents: progress and prospects," *PLoS Pathog*, vol. 6, p. e1000950, 2010.

[7] E. H. Margulies and E. Birney, "Approaches to comparative sequence analysis: towards a functional view of vertebrate genomes," *Nat Rev Genet*, vol. 9, pp. 303-13, Apr 2008.

[8] C. Camacho, *et al.*, "BLAST+: architecture and applications," *BMC Bioinformatics*, vol. 10, p. 421, 2009.

[9] M. Sabry, "A DNA and Amino Acids-Based Implementation of Playfair Cipher," *International Journal of Computer Science and Information Security*, vol. 8, pp. 129-136, 2010.

[10] A. K. Ahmed Atito, S. Z. Rida, "DNA-Based Data Encryption and Hiding Using Playfair and Insertion Techniques," *Journal of Communications and Computer Engineering* vol. 2, pp. 44-49, 2012.