المجلة العراقية للعلوم الإحصائية (20) 2011
عدد خاص بوقائع المؤتمر العلمي الرابع لكلية علوم الحاسوب والرياضيات
ص ص [658-674]

مقارنة بين طرائق لإخفاء المعلومات ضمن سلسلة DNA

سعدون حسين عبدالله* د.احمد سامي نوري**

الخلاصة

في هذا البحث تم اقتراح ثلاث طرائق لإخفاء الرسالة السرية ضمن سلسلة DNA: الأولى أخفيت الرسالة السرية ضمن DNA للكائن البدائي النواة (*E.Coli.*) k12 <u>*Escherichia*</u> <u>*Coli*</u> بالاستفادة من إحدى خصائص جزيئة DNA وهي الطفرة الصامتة، تضمنت الطريقة الثانية صنع DNA كيميائيا بالاعتماد على الرسالة السرية النصية وبالاستفادة من جدول الأحرف الانكليزية ومايقابلها من الكودونات المبنية اعتماداً على تردد الأحرف الانكليزية القياسية وتردد الكودونات لجينات مختلفة (سلاسل مختلفة)، أما الطريقة الثالثة فقد أخفيت الرسالة السرية في جرثومة *E.Coli.* معروف تسلسل قواعدها ضمن موقع EBI أيضا لكن بالاستفادة من خاصية DNA وهي المتمم (Complementary).

الطرائق المقترحة نفذت بنجاح باستخدام برنامج Matlab ولغة #C وأظهرت النتائج تفوقا واضحا في جودة وكفاءة تسلسلات DNA في إخفاء الرسالة السرية حيث وصلت النتائج إلى 100% ونسبة الخطأ إلى 0%.

Comparison among information hiding methods in DNA Sequences

Abstract

By research, three methods for hiding secret message were proposed. In the first one the secret message was hid in a known DNA sequence belongs to prokaryotic organism <u>*Escherichia*</u> <u>*Coli*</u> k12 (*E.Coli*) DNA sequence obtained from EBI location (European Bioinformatics Institute**),** Among the DNA molecule characters, one of the mutant ability types, that is the silent mutant, was chosen to hide a secret message within the sequence. The second method, a sequence of DNA has been synthesized chemically depending on the text secret message via using tables of symbols formed from English letter, numbers and special characters, each symbols represents

*طالب ماجستير/كلية العلوم/جامعة الموصل
**استاذ مساعد/كلية علوم الحاسوب والرياضيات/جامعة الموصل

تاريخ التسلم :2011/10/1 ـــــــــ تاريخ القبول : 21 /12 /2011

codon and the symbols were arranged in the table in descending order depending on English letter frequency codon frequency (Genes). While in the third method the secret message was hid in *E.Coli* bacteria with known nitrogen base sequence in EBI using complementary character.
The proposed methods applied in Matlab program and C# . Results of study revealed that DNA sequence has a high ratio for hiding secret message. The results reached 100% and the error rate to 0%.

1- المقدمة

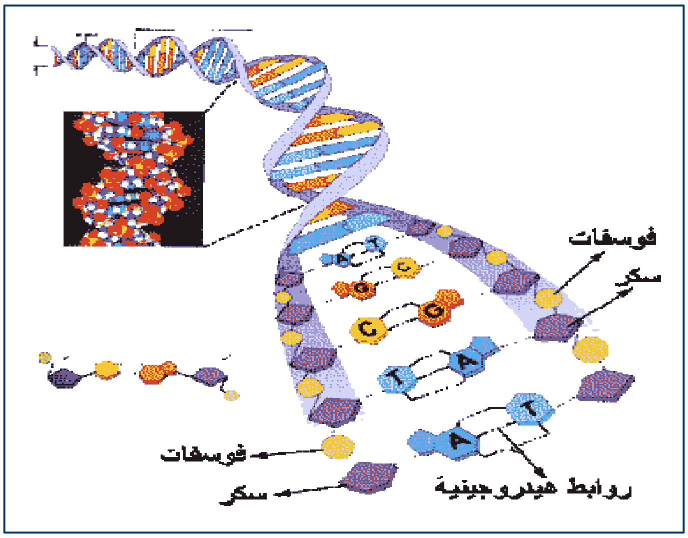
إن كثيراً من الحكومات وضعت قوانين تحد من استخدام طرائق التشفير أو منعها بصورة عامة [2،1]، لذا توجهت أنظار الكثيرين نحو استخدام تقنية (الكتابة المغطاة) للحفاظ على أمنية المعلومات المتبادلة بين طرفين، إذ أنها تقنية تخفي بيانات سرية داخل بيانات أخرى بصورة لا يمكن ملاحظتها مما يخفي حقيقة وجود بيانات سرية متناقلة [3]، في السنين الأخيرة تم استخدام سلسلة DNA (Deoxyribonucleic Acid) كوسط لإخفاء المعلومات السرية لما تتمتع به من مميزات[4]، خلال هذا البحث تم عمل مقارنة بين الطرائق الثلاثة المقترحة لإخفاء الرسالة ضمن سلسلة DNA وهذه الطرائق هي (**الاستبدال في الطفرة الصامتة،** استبدال قاعدة ضمن قاعدة ومتممه، تردد الكودونات مع تردد الأحرف).

2- وصف الـحامض النووي DNA

يعد الحامض النووي DNA المادة الوراثية لخلايا حقيقية النواة Eukaryote وبدائية النواة Prokaryote ومعظم الفيروسات عدا بعض الفيروسات تتكون مادتها الوراثية من الحامض النووي RNA، يتكون الحامض النووي DNA من شريطين ملتفين بشكل حلزون مزدوج Double helix، كل شريط مكون من وحدات مكررة للنيوكلوتيدات والمتكونة من مجموعة فوسفات وسكر خماسي 2-deoxyribose وقواعد نايتروجينة منها البيورينات Purines متمثلة بالكوانينG والأدنين A، وقواعد البريميدينات Pyrimidines المتمثلة بالثايمينT والسايتوسين C وتعد جزيئة سكر الرايبوز بمثابة العمود الفقري لشريطي الـDNA [7،6،5]

ترتبط النيوكلوتيدات فيما بينها بأواصر أسترية Phosphodiester bond، ويتم تثبيت الشريطين بأواصر هيدروجينية ما بين القواعد فيتحد البيورين بالبرمدين، إذ يرتبط الأدنين مع الثيامين بآصرتين، ويرتبط الكوانين مع السايتوسين بثلاث أواصر، مستوى هذه القواعد متعامدة مع محور ليف الـ DNA، ويكون احد الأشرطة مكملاً للآخر Complementary وعلى هذا الأساس وأثناء التضاعف ينفصل الشريطان وكل شريط يعد قالباً Template، ويكون كل شريط

بالنهاية مزدوجاً جديد[8،5] كما في الشكل(1) ، وفضلاً عن الأواصر الهيدروجينية يرتبط شريطا الـ DNA بالتفاعلات الكارهة للماء ما بين أزواج القواعد النتروجينية Hydrophobic interaction والتي تعمل على ثبات الحلزون المزدوج [6].إن القطع ثلاثية النيوكلوتيد تدعى بالشفرة Codon التي تكون مسؤولة عن إضافة حامض أميني لسلسلة متعدد الببتيد، وان العلاقة مابين الجين ومتعدد الببتيد التي تحددها الشفرة الوراثية تكون مسؤولة عن بناء بروتين معين [9] .



الشكل(1) تركيب سلسلة DNA

إن إيجاد طريقة لمعرفة تسلسل تتابعات الحامض النووي DNA أحدث ثورة حقيقية في الوراثة الجزيئية والهندسة الوراثية حيث تزودنا هذه الطريقة بمعلومات مهمة حول تركيب ووظيفة الجين ومن ثم إمكانية تحديد المنتجات التي يشفر لها الجين وتحديد المناطق الحساسة Hot spot لحدوث الطفرات وكذلك تساعد في إيجاد العلاقة التطورية للجينات المتشابهة في الكائن الحي نفسه أو في الكائنات المختلفة والمساعدة في تصنيف الأحياء والفيروسات، في عام1982 نشرت مجلة Nature الخارطة الوراثية لبعض الكائنات وفي عام2001 نشرت الخارطة الوراثية للإنسان [14،10].

2-1 الطفرات

تعرف على إنها التغيرات في النمط الجيني القابلة للتوارث. فلكل كائن حي نمط طبيعي wild type الذي يعد سلالة مرجعية وتتمثل بالإحياء عند عزلها للمرة الأولى، وأي تغيرات تحصل على هذه السلالات تعد طفرات Mutants. والطفرات بدورها تصنف إلى:

1- إضافة insertion أو حذف deletion جزء من المادة النووية وقد تكون منطقة الحذف صغيرة أو كبيرة.

2- يمكن أن تحصل إضافة أو حذف قاعدة واحدة وهذا مايطلق عليه الطفرة النقطية.

3-هنالك نمط من الطفرات يدعى الطفرات الصامتة وتحصل عند تغير الشفرة الوراثية أو جزء منها إلى شفرة وراثية مشفرة للحامض الاميني نفسه [15].

2-2 جهاز قراءة سلسلة DNA (Genetic Analyzer 3310) [11]

ظهرت فكرة بناء أجهزة دقيقة تعمل على تقنية قراءة سلسلة DNA أو تحليل الجينات مع ظهور ثورة الجينات 1981 ليسهل للباحثين تقنية العمل في مجالات البحوث المختلفة إلا أن تطور الأجهزة لم يحدث إلا في الوقت الحاضر.

إن جهاز Genetic Analyzer 3310 الموضح في الشكل(2) من الأجهزة المتطورة حيث تقوم بعدة عمليات (تطبيقات) منها عملية قراءة قطع DNA وتركيب سلسلة DNA والتطفير وعملية الترحيل الكهربائي وعمليات أخرى. إن جهاز Genetic Analyzer 3310 سهل الاستخدام وسريع العمل وذو كلفة قليلة لإظهار النتائج وذو دقة عالية في النتائج، والجهاز له القابلية على قراءة 864 قاعدة في سلسلة أحادية وقراءة 1000 قاعدة / زوج في سلسلة ثنائية.

الشكل(2) جهاز قراءة سلسلة DNA (Genetic Analysis) DNA Sequencer3310

هنالك مجموعة من البرامج في الحاسوب المرتبط مع الجهاز منها مايستخدم لقراءة سلسلة DNA وبرنامج لعملية الطفرة وغيرها.أن سلسلة DNA عبارة عن مجموعة من المنحنيات كل منحن يمثل قاعدة وكل قاعدة تكون بلون محدد (A- لون أخضر،C- لون أزرق، G- لون أسود، T- لون أحمر) فضلا عن إعطاء ملف نصي يمثل سلسلة DNA.

3-2 التفاعل التضاعفي لسلسلة الـ DNA Polymerase Chain Reaction(PCR)

PCR من التقنيات المهمة في البيولوجي الجزيئي، وتستخدم لزيادة كمية DNA ويقصد بها التضاعف الإنزيمي لقطعة DNA معينة ملايين المرات في أنبوبة الاختبار *in vitro* بوجود البادئات التي ترتبط مع التتابع المكمل لها على الشريط الأصلي [14,5].

3- الطرائق المقترحة

3-1الطريقة الأولى: الاستبدال في الطفرة الصامتة

تستند هذه الطريقة إلى فكرة الطفرة الصامتة التي تحدث لسلسلة DNA داخل جسم الكائن والتي تتضمن حدوث طفرة في القاعدة الأخيرة من الكودون لإنتاج نفس الحامض الاميني. والجدول(1) يوضح بان هنالك 64 كودوناً والذي يشفر 20 حامضاً أمينياً بمعنى أن هنالك أكثر من كودون ينتج نفس الحامض ألأميني. تم استغلال هذه الفكرة بالعمل على تمثيل الرسالة السرية بقواعد ثلاثية إذا كانت الرسالة نصية، وبقواعد رباعية في حالة الرسالة الصورية (لاحظ الملحق(1))، ومن ثم استبدالها في الكودونات القابلة للتغيير والمؤشرة ضمن الجدول(1) [16،15].

الجدول (1) الكودونات (المناطق المظللة تمثل الكودونات التي تنتج نفس الحامض الأميني)[15 ،16].

القاعدة الأولى					
G	A	C	T		القاعدة الثانية
TGT Cys[C] TGC Cys[C] TGA Stop[O] TGG Try[W]	TAT Tyr[Y] TAC Tyr[Y] TAA Stop[O] TAG Stop[O]	TCT Ser[S1] TCC Ser[S2] TCA Ser[S3] TCG Ser[S4]	TTT Phe[F] TTC Phe[F] TTA Leu[L4] TTG Leu[L5]	T	
CGT Arg[R1] CGC Arg[R2] CGA Arg[R3] CGG Arg[R4]	CAT His[H] CAC His[H] CAA Gln[Q] CAG Gln[Q]	CCT Pro[P1] CCC Pro[P2] CCA Pro[P3] CCG Pro[P4]	CTT Leu[L1] CTC Leu[L2] CTA Leu[L3] CTG Leu[L4]	C	
AGT Ser[S5] AGC Ser[S6] AGA Arg[R5] AGG Arg[R5]	AAT Asn[N] AAC Asn[N] AAA Lys[K] AAG Lys[K]	ACT Thr[T1] ACC Thr[T2] AGA Thr[T3] ACG Thr[T4]	ATT lle[I] ATC lle[I] ATA lle[I] ATG Met[M]	A	
GGT Gly[G1] GGC Gly[G2] GGA Gly[G3] GGG Gly[G4]	GAT Asp[D] GAC Asp[D] GAA Glu[E] GAG Glu[E]	GCT Ala[A1] GCC Ala[A2] GCA Ala[A3] GCG Ala[A4]	GTT Val[V1] GTC Val[V2] GTA Val[V3] GTG Val[V4]	G	

مثال توضيحي لطريقة الاستبدال في الطفرة الصامتة

للتوضيح يمكن اختيار سلسلة DNAبدائية النواة لبكتيريا(إيكولاي *E.Coli*) معروفة التسلسل من الموقع EBI على إن يتم الاتفاق عليها بين الطرفين.

S_i=ATGAACGGCTCGCCCGGTCTGGTCTACATGGAGTCGGTGGCC...

بعد تقطيع سلسلة S_i إلى كودونات تنتج سلسلة جديدة S_c.

S_c=ATG AAC GGC **TCG CCC GGT CTG GTC** TAC ATG GAG **TCG** GTG **GCC** ……………………..

لغرض تحديد طول الرسالة السرية(النصية) (M_k) تمت إضافة ! إلى نهاية الرسالة.

M_k= Silent Mutation Method!

بعد ترميز كل حرف وبما يقابلها من الكودون حسب الملحق(1) تنتج :

M_k'' = CAG AGA AGT ACA ATC CAT ………………………….TAT.

عملية إخفاء M_k'' ضمن الكودونات القابلة للطفرة

S_{new}= ATG AAC **GG<u>C</u> TC<u>A</u> CC<u>G</u>**…………………………………….

بالإمكان استخدام الجدول(1) لتحويل سلسلة S_{new} إلى سلسلة من الأحماض الامينية S_{Acid}

S_{Acid}=MNG2S3P4……………………

تم إخفاء السلسلة الناتجةS_{Acid} ضمن صورة C_k (رقمية) وبأنواع مختلفة منها (PNG , BMP) باستغلال تقنية البت الأقل أهمية (LSB) وبتقنية عشوائية.

<u>ولاسترجاع الرسالة السرية من سلسلة DNA S_{stego}</u>

S_{stego} = ATG AAC **GGC TCA CCG**…………………

نأخذ فقط أخر قاعدة من الكودونات القابلة للطفرة ونتوقف عند S=TAT

S= **CAG** ……………………….........................

بالاعتماد على الملحق(1) تكون النتيجة النهائية

M= S………………………………………………………

2-3 الطريقة الثانية:طريقة تردد الكودونات مع تردد الأحرف

من خصائص كل لغة وجود خاصية تردد أو نسبة تكرار كل حرف من الأحرف ضمن جملة معينة. وقد أثبتت الإحصائيات أن لكل لغة حية حرفاً له تكرار يختلف عن غيره، وعلى ذلك تم بناء جداول كما هو في اللغة الانكليزية الجدول (2)[12] كذلك الحال بالنسبة لكل جين (بروتين) حيث تتواجد الحوامض الامينية (الكودونات) التي تتكرر أكثر من غيرها في سلسلة الـ DNA

الجدول (2) نسبة تردد الأحرف في اللغة الانكليزية[12]

Letter	Frequency	Letter	Frequency	Letter	Frequency
E	12.702%	R	5.987%	P	1.929%
T	9.056%	D	4.253%	B	1.492%
A	8.167%	L	4.025%	V	0.978%
O	7.507%	C	2.782%	K	0.772%
I	6.966%	U	2.758%	J	0.153%
N	6.749%	M	2.406%	X	0.150%
Space	6.378%	W	2.360%	Q	0.095%
S	6.327%	F	2.228%	Y	1.974%
H	6.094%	G	2.015%	Z	0.074%

تمت الاستفادة من هذه الخاصية لغرض إعداد هذه الطريقة والمتمثلة بما يأتي: حساب تكرار الأحرف الأبجدية المذكورة سابقا مع تكرار الكودونات في جينات مختلفة يتم بناء جدول لإعطاء كل حرف وما يقابلها من الكودونات والاستفادة منها في بناء سلسلة DNA اصطناعية والتي تمثل الرسالة النصية السرية.

<u>مثال توضيحي لطريقة تردد الكودونات مع تردد الأحرف</u>

وللتوضيح تم اختيار سلاسل DNA لـ G لبكتيريا (أيكولاي *E.coli*) لجينات مختلفة وبأطوال مختلفة مأخوذة من موقع EBI(European Bioinformatics Institute)، بترتيب نسبة تواجد كل حامض أميني (codon) في السلاسل المختلفة في الجدولT1ثم إيجاد وبترتيب الجدولT1 حسب تكرار الأعلى إلى الأدنى ينتج الجدولT2 (لاحظ الملحق(2)) تتم الاستفادة من الجدول T2 مع الجدول (2) بحيث تتم مقابلة حقول الجدولين وإعطاء كل رمز في الجدول(2) مايقابلها من الكودون وحسب الترتيب وتكوين الجدول(3).

الجدول (3) ناتج تقابل كل كودون مع حروف وأرقام ورموز

S	ATT	J	TTG	+	CAC	,	CTT
H	ATC	X	CCG	>	GGG	;	TGA
R	TTC	Q	TAT	<	TAC	“	ATA
D	ACC	Z	AAC	=	CCC	:	AGG
L	GCA	0	GTT	.	CAT	‘	CTA
C	GGT	1	TGG	#	TCT	~	TGT
U	CAG	2	GAC	%	CGG	]	AGA
M	GAT	3	GCT	(	CCA	[	TAA
W	ACG	4	CGT	)	TCA	$	TAG

ولتكن الرسالة المرسلة M_i باستخدام الجدول (3) يتم تصنيع سلسلة S_{new}

M_i =Genetic code is a DNA base sequence

S_{new} = GTCCTGGCGCTGGAAGGCGGTATG GGTGCCACCCTGATG
GGCATTATG GTGACCGCGGTGATG TTTGTGATTCTGATG
ATTCTGTATCAGCTGGCGGGTCTG

ارسلت سلسلة S_{new} إلى دولة كورية الجنوبية حيث وصنعتها شركةBIONEER [13] وتمت قراءة سلسلة DNA من قبل شركة BIONEER وتطابقها مع سلسلة الرسالة السرية في المثال كذلك تم نقلها إلى(جامعة فاتح بمدينة أنقرة/تركيا) وقراءتها باستخدام جهاز DNA Sequencer(genetic analyzer 3130).

<u>لاسترجاع الرسالة السرية</u> يقوم المتسلم أيضا بخطوات تكوين الجداول ((T1)،(T2)(لاحظ الملحق(2)،(3)) بعدها يتم استرجاع الرسالة السرية من سلسلةS_{stego} وبالشكل الآتي:

S_{stego} =GTC CTG GCG CTG GAA GGC GGT ATG GGT GCC
ACC CTG ATG GGC ATT ATG GTG
ACC GCG GTG ATG TTT GTG ATT
CTG ATG ATT CTG TAT CAGCTG GCG
GGT CTG

M_i= g e n ....

في بعض الأحيان لايمكن قراءة سلسلة DNA لذا نحتاج إلى مضاعفتها باستخدام تقنية PCR. تم وضع السلسلة المصنعة ضمن الجهاز المذكور وتشغيل البرنامج الخاص لتقرأ السلسلة حيث تكون النتيجة عبارة عن سلسلة من المنحنيات وبألوان مختلفة حسب قواعد سلسلة DNA.

3-3 الطريقة الثالثة: استبدال قاعدة ضمن قاعدة ومتممه

من الخصائص الموجودة في سلسلة DNA هي المتمم (Complementary) وعلى هذا الأساس تم اقتراح هذه الطريقة: يتم العمل على سلسلة DNA معروفة التسلسل من موقع EBI بحيث يتم البحث داخل السلسلة عن الصيغة ($BAB^{\backslash}$) ومعناه انه تم الحصول على ثلاث قواعد متتالية ضمن السلسلة بحيث تكون القاعدة الثالثة هي المتمم للقاعدة الأولى أما القاعدة التي في الوسط فتشمل أية قاعدة من القواعد الأربعة المعروفة، بعد ذلك تتم عملية الإخفاء لمفردات الرسالة السرية وفق العلاقات الموضحة بين القواعد الوسطية في الجدول(4).

إن الحالة الطبيعية للمتمم في سلسلة DNA كما مذكور سابقا هي بالشكل الآتي: (A متممه T) والعكس بالعكس و(C متممه G) والعكس بالعكس، لكن بالإمكان وضع استراتجيات متعددة لعملية المتمم ليشمل احتمالات أخرى كما في الجدول(4).

الجدول(4) يمثل احتمالات المتمم

ت	قاعدة	متممه	قاعدة	متممه	قاعدة	متممه	قاعدة	متممه
1	A	C	C	G	G	T	T	A
2	A	C	C	T	T	G	G	A
3	A	G	G	C	C	T	T	A
4	A	G	G	T	T	C	C	A
5	A	T	T	C	C	G	G	A
6	A	T	T	G	G	C	C	A
7	C	A	A	G	G	T	T	C
8	C	A	A	T	T	G	G	C
9	C	G	G	A	A	T	T	C
10	C	G	G	T	T	A	A	C
11	C	T	T	A	A	G	G	C
12	C	T	T	G	G	A	A	C
13	G	A	A	C	C	T	T	G
14	G	A	A	T	T	C	C	G
15	G	C	C	A	A	T	T	G
16	G	C	C	T	T	A	A	G
17	G	T	T	A	A	C	C	G
18	G	T	T	C	C	A	A	G
19	T	A	A	C	C	G	G	T
20	T	A	A	G	G	C	C	T
21	T	C	C	A	A	G	G	T
22	T	C	C	G	G	A	A	T
23	T	G	G	A	A	C	C	T
24	T	G	G	C	C	A	A	T

وتتمثل هذه الطريقة بثلاث خطوات:

1- قراءة ثلاث قواعد متتالية من سلسلة DNA الأصلية S، $S(_{j,j+1,j+2})$.

2- تمر هذه الخطوة بحالتين:

- إذا كانت القواعد الثلاثة السابقة غير محققة لصيغة $BAB^{\backslash}$ تتم إضافة S_j إلى S_{new} ، ++j ، العودة إلى الخطوة 1.
- إذا كانت القواعد الثلاثة محققة لصيغة $BAB^{\backslash}$ فيكون الاحتمال وحسب الرسالة:

أ- إذا كانت الرسالة $M_{bin}=0$ تتم إضافة S_j إلى S_{new} ، j++ ، العودة إلى الخطوة 1.

ب- إذا كانت الرسالة$M_{bin}=1$ فيؤخذ متمم لـ------ S_{j+1} ونضعه في السلسلة الأصلية (S)، تتم إضافة S_j إلى S_{new} ، j++ ، العودة إلى الخطوة 1.

3- نستمر بالعمل لحين الانتهاء من الرسالة السرية الثنائية.

مثال توضيحي لطريقة استبدال قاعدة ضمن قاعدة ومتممه

لتوضيح عملية الإخفاء ليكن المثال الآتي: تم اختيار سلسلة DNA قصيرة (السلسلة الأصلية) لغرض تطبيق الإخفاء.

S=GATGCAGGCTG السلسلة الأصلية

M_i=110101 الرسالة السرية

وكان الاتفاق على اختيار الاحتمالية (1) من الجدول(4) فتكون خطوات الإخفاء كما موضح:

الخطوات	المدخلات	الرسالة السرية	المخرجات
	S_j	M_i	S_{new}
1	**G A T** G C A G G C T G	1	G
2	G **C T G** C A G G C T G	1	GC
3	G **C A G** C A G G C T G	0	GCA
4	G C A **G C A** G G C T G		GCAG
5	G C A G **C A G** G C T G	1	GCAGC
6	G C A G C **C G G** C T G	0	GCAGCC
7	G C A G C C **G G C** T G		GCAGCCG
8	G C A G C C G **G C T** G	1	GCAGCCGG
9	G C A G C C G G G T G		GCAGCCGGGTG

إن السلسلة الجديدة الحاملة للرسالة السرية هي S_{new}

S_{new} = GCAGCCGGGTG

تم إرسال السلسلة الناتجة S_{new} بإخفائها ضمن صورة C_k (رقمي) وبأنواع مختلفة منها (PNG , BMP) باستغلال تقنية البت الأقل أهمية (LSB).

ولتوضيح عملية الاسترجاع نتبع مايأتي:

S = GATGCAGGCTG سلسلة DNA الأصلية

S_{stego} = GCAGCCGGGTG السلسلة المحتوية على الرسالة السرية

وكان الاتفاق على اختيار الاحتمالية (1) من الجدول(4)فتكون خطوات الاسترجاع كما موضح:

الخطوات	المدخلات S	المدخلات S_{stego}	المخرجات R_i
1	GATGCAGGCTG	GCAGCCGGGTG	1
2	GCTGCAGGCTG	GCAGCCGGGTG	1
3	GCAGCAGGCTG	GCAGCCGGGTG	0
4	GCAGCAGGCTG	GCAGCCGGGTG	
5	GCAGCAGGCTG	GCAGCCGGGTG	1
6	GCAGCCGGCTG	GCAGCCGGGTG	0
7	GCAGCCGGCTG	GCAGCCGGGTG	
8	GCAGCCGGCTG	GCAGCCGGGTG	1
9	GCAGCCGGGTG	GCAGCCGGGTG	

الرسالة السرية R=110101

4- المناقشة

من خلال تطبيق الطرائق الثلاثة المقترحة تم التوصل إلى أن هنالك اختلافاً في عدة مميزات، بحيث ميزت كل طريقة عن الأخرى كما موضح في الجدول(5).

الجدول(5) فروقات بين الطرائق الثلاثة المقترحة

ت	الخاصية	الطريقة الأولى الاستبدال في الطفرة الصامتة	الطريقة الثانية تردد الكودونات مع تردد الأحرف	الطريقة الثالثة استبدال القاعدة ضمن قاعدة ومتممه
1-	طول سلسلة DNA الناتجة	الأطول	الأقصر	طويل
2-	سرية نقل سلسلة DNA	Insilico(صورة)	Invitro (مادة كيميائية)	Insilico(صورة)
3-	زمن الإخفاء والاسترجاع للرسالة (ضمن السلسلة)	الأكثر	قليل	كثير
4-	مستويات الإخفاء	ثلاثة مستويات	مستوى واحد	مستويان
5-	قوة الإخفاء ضمن السلسلة	الأقل	الأقل	أقوى
6-	متطلبات بين الطرفين	جدول الترميز	جدول الترميز	لايحتاج

		مع وجود مفاتيح		
-7	المتغيرات على السلسلة	احتمالية عدم حصول التغيير للسلسلة	------	احتمالية عدم حصول التغيير للسلسلة
-8	PSNR تتناسب طردياً مع حجم الصورة مع ثبوت حجم الرسالة (طول سلسلة DNA)	الأقل	------	أفضل
-9	BER%	0	0	0
-10	المعدل	اقل أو مساو للطريقة الثالثة	ألأكثر	اقل أو مساو للطريقة الأولى
-11	كلفة الإخفاء ضمن الوسط	لايوجد	تحتاج إلى كلفة مالية	لايوجد
-12	زمن الإخفاء والاسترجاع للسلسلة (ضمن الوسط)	الأقل	ألأكثر	الأقل
-13	نوع الترميز للرسالةالسرية	كودون	كودون	ثنائي

5- الاستنتاجات

1- تعد سلسلة DNA وسطاً جيداً جداً لإخفاء الرسالة السرية.

2- من الصعوبة كشف المحال لسلسلةDNA الحاملة للرسالة السرية والمتفق عليها بين الطرفين والمأخوذة من قاعدة البيانات EBI بسرعة لانها تحتوي على مايقارب من 163 مليون سلسلة DNA لكائنات مختلفة.

3- من نقاط القوة للطرائق الجديدة هو ثبوت طول سلسلة الـ ـ ـ ـ ـ DNA وأشكالها وترتيبها قبل الإخفاء وبعده.

4- بالإمكان نقل سلسلة DNA الحاملة للرسالة السرية ضمن وسط من الوسائط المتعددة مما أعطى أمنية أعلى للطرائق.

5- إضافة درجة أمنية أعلى من خلال افتراضها من مواقع الكودونات للسلسلة نفسها.

6- في الطريقة الثانية صنعت سلسلة DNA مشابهة بالصفات لسلسلة DNA الحقيقية وذلك بالاستفادة من فكرة تردد الكودونات لكل جين أو سلسلة DNA.

7- بالإمكان بناء سلسلة DNA صناعياً بالاعتماد على الرسالة السرية.

8- بالإمكان نقل سلسلة DNA الحاملة للرسالة السرية بالطريقة البيولوجية أو الحاسوبية.

9- إن العمل على سلسلة DNA حقيقي النواة (الإنسان) تحتاج إلى جهد أكثر من العمل على سلسلة DNA لكائن بدائي النواة مما تعطي سرية عالية وجهد اكبر نتيجة الطبيعية البيولوجية لكل منهم.

6- المصادر

1. Al-Husainy M. A. F., (2009), “Image Steganography by Mapping Pixels to Letters”, J. Computer Sci.,Vol. 5, No. 1, pp. 33-38.
2. Prasad M. S. and etal., (2009), “A Novel Information Hiding Technique For Security By Using Image Steganography”, Journal of Theoretical and Applied Information Technology, Vol. 8, No. 1, pp.35-39.
3. Motameni H.and etal, (2007), “Labeling Method in Steganography”, World Acad. Sci. Eng. Technol., Issue 30, pp. 349-354.
4. feng,B.S.J. ,potkonjak,M. , (2002) , "Data Hiding in DNA” ,international workshop on information Hiding , pp (373-386)
5. http://www.rit.edu/~vxr8205/crypto2/cryptopaper.html. Tamari , R.H. ,(1996), “principles of genetics” , Sth ed , Mc Graw –Hill companies , USA.
6. Dale , J. and Schants , M. V. (2002), “From genes to Genomes” JohnWiley & sons , Inc . New York .
7. Nelson , D. L. and cox , M. M. (2005), “Lehn : nger principles of Biochemistry” W. H. freeman and company , New York .
8. Watson, J.D. and Crick, F.H.,(1953), Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature, 171: 737-738.
9. Paolella, P. ,(1998), “Introduction to Molecular Biology”. McGraw-Hill Companies, Inc., New York.
10. http://www.ebi.ac.uk ,(2011),European Bioinformatics Institute.
11. http://www.appliedbiosystems.com,(2011),AppliedBiosystems 3130 and 3130xl Genetic Analyzers, USA,08/2006publication106BR 1002.Sophisticate,dautomation,superiorperformance,streamlined setup, and 24-hour unattended operation.
12. http://www.Wikipedia.com,(2011),Frequency of Letters of the alphabet in English.
13. http://www.bioneer.com,Bioneer’s,(2008) novel Nano-Technology, (Korea patent no.10-0850430,pct pending wo/072865).

14.العباسي ، رغد رياض شفيق، (2008)،"التغيرات الحيوية في خصائص الـ DNA لبكتريا Escherichia coli بعد تكسيره واستخدامه في الكلونة"،بحث ماجستير، قسم علوم الحياة،كلية العلوم، جامعة الموصل، العراق.

15. البكري، غالب حمزة ، (1991) ، مبادئ الهندسة الوراثية ، كلية العلوم ، جامعة البصرة، العراق .

16. .الخفاجي، زهرة محمود، (2008),التقنية الحيوية الميكروبية (توجيهات جزيئية)، جامعة بغداد ،العراق .

الملحق (1)

نموذج يوضح ترميز الرسالة الصورية وفق طريقة الاستبدال في الطفرة الصامتة

الملحق(1)

نموذج يوضح ترميز الرسالة الصورية وفق طريقة الاستبدال في الطفرة الصامتة

Pixel value	Sequence	Pixel value	Sequence	Pixel value	Sequence	Pixel value	Sequence	Pixel value	Sequence
0	AAAA	51	ATAT	102	CGCG	153	GCGC	204	TATA
1	AAAC	52	ATCA	103	CGCT	154	GCGG	205	TATC
2	AAAG	53	ATCC	104	CGGA	155	GCGT	206	TATG
3	AAAT	54	ATCG	105	CGGC	156	GCTA	207	TATT
4	AACA	55	ATCT	106	CGGG	157	GCTC	208	TCAA
5	AACC	56	ATGA	107	CGGT	158	GCTG	209	TCAC
6	AACG	57	ATGC	108	CGTA	159	GCTT	210	TCAG
7	AACT	58	ATGG	109	CGTC	160	GGAA	211	TCAT
8	AAGA	59	ATGT	110	CGTG	161	GGAC	212	TCCA
9	AAGC	60	ATTA	111	CGTT	162	GGAG	213	TCCC
10	AAGG	61	ATTC	112	CTAA	163	GGAT	214	TCCG
11	AAGT	62	ATTG	113	CTAC	164	GGCA	215	TCCT
12	AATA	63	ATTT	114	CTAG	165	GGCC	216	TCGA
13	AATC	64	CAAA	115	CTAT	166	GGCG	217	TCGC
14	AATG	65	CAAC	116	CTCA	167	GGCT	218	TCGG
15	AATT	66	CAAG	117	CTCC	168	GGGA	219	TCGT
16	ACAA	67	CAAT	118	CTCG	169	GGGC	220	TCTA
17	ACAC	68	CACA	119	CTCT	170	GGGG	221	TCTC
18	ACAG	69	CACC	120	CTGA	171	GGGT	222	TCTG
19	ACAT	70	CACG	121	CTGC	172	GGTA	223	TCTT
20	ACCA	71	CACT	122	CTGG	173	GGTC	224	TCAA
21	ACCC	72	CAGA	123	CTGT	174	GGTG	225	TGAC
22	ACCG	73	CAGC	124	CTTA	175	GGTT	226	TGAG
23	ACCT	74	CAGG	125	CTTC	176	GTAA	227	TGAT
24	ACGA	75	CAGT	126	CTTG	177	GTAC	228	TGCA
25	ACGC	76	CATA	127	CTTT	178	GTAG	229	TGCC
26	ACGG	77	CATC	128	GAAA	179	GTAT	230	TGCG
27	ACGT	78	CATG	129	GAAC	180	GTCA	231	TGCT
28	ACTA	79	CATT	130	GAAG	181	GTCC	232	TGGA
29	ACTC	80	CCAA	131	GAAT	182	GTCG	233	TGGC
30	ACTG	81	CCAC	132	GACA	183	GTCT	234	TGGG
31	ACTT	82	CCAG	133	GACC	184	GTGA	235	TGGT
32	ACAA	83	CCAT	134	GACG	185	GTGC	236	TGTA
33	AGAC	84	CAAA	135	GACT	186	GTGG	237	TGTC
34	AGAG	85	CCCA	136	GAGA	187	GTGT	238	TGTG
35	AGAT	86	CCCC	137	GAGC	188	GTTA	239	TGTT
36	AGCA	87	CCCG	138	GAGG	189	GTTC	240	TTAA
37	AGCC	88	CCCT	139	GAGT	190	GTTG	241	TTAC
38	AGCG	89	CCGA	140	GATA	191	GTTT	242	TTAG
39	AGCT	90	CCGC	141	GATC	192	TAAA	243	TTAT
40	AGGA	91	CCGG	142	GATG	193	TAAC	244	TTCA
41	AGGC	92	CCGT	143	GATT	194	TAAG	245	TTCC
42	AGGG	93	CCTA	144	GCAA	195	TAAT	246	TTCG
43	AGGT	94	CCTC	145	GCAC	196	TACA	247	TTCT
44	AGTA	95	CCTT	146	GCAG	197	TACC	248	TTGA
45	AGTC	96	CGAA	147	GCAT	198	TACG	249	TTGC
46	AGTG	97	CGAC	148	GCCA	199	TACT	250	TTGG
47	AGTT	98	CGAG	149	GCCC	200	TAGA	251	TTGT
48	ATAA	99	CGAT	150	GCCG	201	TAGC	252	TTTA
49	ATAC	100	CGCA	151	GCCT	202	TAGG	253	TTTC
50	ATAG	101	CGCC	152	GCGA	203	TAGT	254	TTTG
255	TTTT								

Symbol	Codon	Symbol	Codon	Symbol	Codon	Symbol	Codon
A	AAA	Q	CAA	6	GAA	-	TAA
B	AAC	R	CAC	7	GAC	/	TAC
C	AAG	S	CAG	8	GAG	@	TAG
D	AAT	T	CAT	9	GAT	!	TAT
E	ACA	U	CCA	*	GCA	#	TCA
F	ACC	V	CCC	&	GCC	^	TCC
G	ACG	W	CCG	+	GCG	,	TCG
H	ACT	X	CCT	Space	GCT	;	TCT
I	AGA	Y	CGA	>=	GGA	“	TGA
J	AGC	Z	CGC	<=	GGC	:	TGC
K	AGG	0	CGG	=	GGG	‘	TGG
L	AGT	1	CGT	.	GGT	~	TGT
M	ATA	2	CTA	#	GTA	]	TTA
N	ATC	3	CTC	%	GTC	[	TTC
O	ATG	4	CTG	(	GTG	‘	TTG
P	ATT	5	CTT	)	GTT	~	TTT

الملحق(2)

جدول T1 يبين نسبة تكرار كل كودون في سلاسل مختلفة

Codon	Frequency	Codon	Frequency	Codon	Frequency	Codon	Frequency
AAA	52	CAA	36	GAA	150	TAA	4
AAC	45	CAC	29	GAC	43	TAC	28
AAG	47	CAG	60	GAG	51	TAG	3
AAT	37	CAT	27	GAT	60	TAT	45
ACA	22	CCA	24	GCA	65	TCA	24
ACC	65	CCC	27	GCC	90	TCC	30
ACG	59	CCG	45	GCG	98	TCG	37
ACT	23	CCT	18	GCT	43	TCT	27
AGA	7	CGA	17	GGA	20	TGA	14
AGC	49	CGC	59	GGC	98	TGC	30
AGG	12	CGG	25	GGG	28	TGG	44
AGT	17	CGT	41	GGT	64	TGT	10
ATA	12	CTA	10	GTA	23	TTA	41
ATC	76	CTC	32	GTC	54	TTC	68
ATG	85	CTG	160	GTG	103	TTG	46
ATT	79	CTT	14	GTT	44	TTT	49

جدول T2 يبين ترتيب الكودونات حسب التكرار من الأعلى إلى الأدنى

Codon	Frequency	Codon	Frequency	Codon	Frequency	Codon	Frequency
CTG	160	CGC	59	TTA	41	GTA	23
GAA	150	GTC	54	AAT	37	ACT	23
GTG	103	AAA	52	TCG	37	ACA	22
GCG	98	GAG	51	CAA	36	GGA	20
GGC	98	TTT	49	CTC	32	CCT	18
GCC	90	AGC	49	TCC	30	CGA	17
ATG	85	AAG	47	TGC	30	AGT	17
ATT	79	TTG	46	CAC	29	CTT	14
ATC	76	CCG	45	GGG	28	TGA	14
TTC	68	TAT	45	TAC	28	ATA	12
ACC	65	AAC	45	CCC	27	AGG	12
GCA	65	GTT	44	CAT	27	CTA	10
GGT	64	TGG	44	TCT	27	TGT	10
CAG	60	GAC	43	CGG	25	AGA	7
GAT	60	GCT	43	CCA	24	TAA	4
ACG	59	CGT	41	TCA	24	TAG	3