Volume 3, Issue 9, September 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Steganography in Putative Protein Coding Regions of Eukaryotic DNA: A Novel Approach

Saritha Namboodiri*
Department of Computer Science,
Sreekrishnapuram V. T. Bhattathiripad college,
University of Calicut,
Kerala, India

Shilpa Rajan
Department of Computer Science,
Sreekrishnapuram V. T. Bhattathiripad college,
University of Calicut,
Kerala, India

Abstract*—The goal of steganography is to conceal the existence of information hidden in cover medium (most commonly digital images, text, audio, video files) during information transmission. DNA (Deoxyribo Nueclic Acid) as cover medium is one of the possible alternatives that are currently being investigated. DNA is a chain of nucleotides (Adenine (A), Guanine (G) Cytosine (C) and Thymine (T)) wherein the nucleotide taken as triplets forms a codon. Codons are translated to amino acids using the genetic code to form proteins. Mutation occurs when a change in the nucleotide of a codon affect the amino acid formed. Most of the existing DNA based steganography communication caused mutation thereby disrupting the biological function of cover DNA. In this paper, we elucidate a novel algorithm to randomly assign the nucleotides of the synthesized DNA strand of the transformed secret information in putative protein coding region of eukaryotic DNA, avoiding mutation.

Keywords*— Steganography, DNA, Eukaryotic DNA, protein coding regions, mutation, codon.

I. INTRODUCTION

The goal of steganography is to embed sensitive information into an innocuous looking cover media (most commonly digital images, text, audio, video files) with the intention of hiding its presence from unauthenticated viewers during information transmission [1-3]. In stagenography communication, the sender and the receiver agree upon a specific protocol. The secret information is first transformed into the cover media format using an encoding scheme. This transformed secret information is then hidden in the cover media by means of an embedding algorithm and the stego-key, (key to retrieve the secret information) generated. The stego-medium (cover media + secret information) together with the stego-key is transmitted over a channel to the receiver where it is processed by the extraction algorithm using the stego-key and the secret information extracted. The extracted secret information is then transformed back into its original format using the inverse encoding scheme. Fig. 1 depicts the steganography communication process.

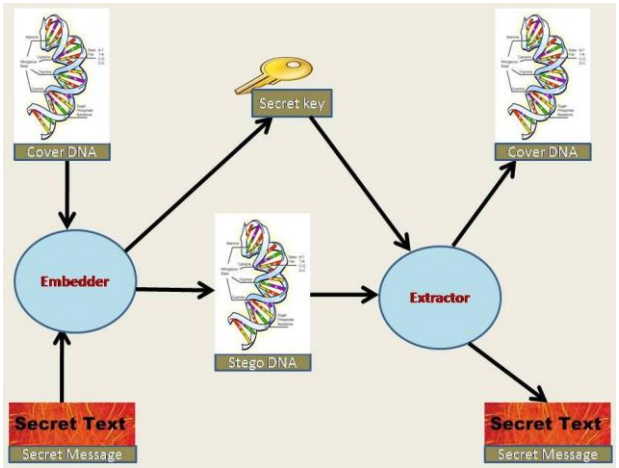


Fig. 1 Steganography communication

With the tremendous advancement in DNA computing, steganographic communication with DNA as one possible alternative over the commonly used digital image, audio and video cover media [4-11] is emerging as a new area of interest in international research.

A. DNA: Embedding the Secret of Life

DNA, embedding the secret of life, is made up of four kinds of bases, Adenine (A), Guanine (G) Cytosine (C) and Thymine (T). Each base with an attached sugar and phosphate molecule forms a nucleotide. In DNA, these

nucleotides are arranged in some predetermined order to execute its function. A unique combination of three nucleotides forms a codon [12]. Codons are translated to amino acids using the genetic code as depicted in the Genetic codon table (Fig. 2) to form a protein. Sixty four different codons (4^3) formed from the 4 nucleotides code for twenty amino acids [12]. Certain changes in the nucleotides of a codon may thereby alter the amino acid formed causing mutation whereas others may not. Except for the two of the amino acids (Met and Trp), all other amino acids can be encoded by 2 to 6 different codons.

First Nucleotide		Second Nucleotide: T		C		A		G		Third Nucleotide
		Code	Amino acid	Code	Amino acid	Code	Amino acid	Code	Amino acid	
T		TTT	phe	TCT	ser	TAT	tyr	TGT	cys	T
		TTC		TCC		TAC		TGC		C
		TTA	leu	TCA		TAA	STOP	TGA	STOP	A
		TTG		TCG		TAG	STOP	TGG	trp	G
C		CTT	leu	CCT	pro	CAT	his	CGT	arg	T
		CTC		CCC		CAC		CGC		C
		CTA		CCA		CAA	gln	CGA		A
		CTG		CCG		CAG		CGG		G
A		ATT	ile	ACT	thr	AAT	asn	AGT	ser	T
		ATC		ACC		AAC		AGC		C
		ATA		ACA		AAA	lys	AGA	arg	A
		ATG	met	ACG		AAG		AGG		G
G		GTT	val	GCT	ala	GAT	asp	GGT	gly	T
		GTC		GCC		GAC		GGC		C
		GTA		GCA		GAA	glu	GGA		A
		GTG		GCG		GAG		GGG		G

Fig 2 Genetic codon table.

B. DNA-> RNA->Protein

1) *Central Dogma:* DNA in the nucleus of the cell is transcribed into the mRNA. The transcribed mRNA travels to protein production sites and translates to protein [13]. Fig. 3 depicts the central dogma of molecular biology.

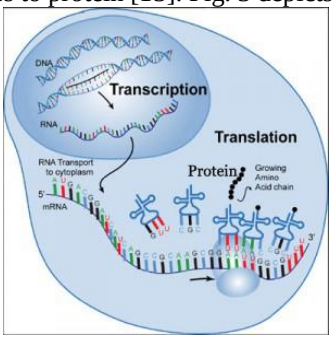


Fig. 3 Central dogma

Translation begins with the start codon, ATG which codes for the amino acid Methionine. The translation process stops when it comes across a stop codon. There are three stop codons: TAA, TAG and TGA. Any of these codons can stop the translation. The region that starts with a start codon and ends with a stop codon is the probable protein coding region or Open Reading Frame (ORF). This region is translated into amino acid using the genetic code to from a protein. Steganography using the putative coding regions of DNA to hide secret information have not been explored so far to the best of our knowledge.

Clelland et al. [6] have explored the possibility of hiding messages in DNA microdots. They transformed the secret text into a synthetic DNA sequence by replacing each character with a unique three nucleotides code and placed marker sequences at the two ends. This synthetic DNA strand was then placed into a randomly selected DNA strand and mixed with dummy DNA strands of similar length. The pool of DNA strands were dried and cut into microdots, forming DNA microdots. These microdots were then placed over the full stops in a typed letter and sent to the receiver along with the primers. From the typed dot, using the primer, the DNA containing the secret text could be located and amplified using Polymerase Chain Reaction (PCR). This DNA could then be sequenced and the secret text read.

Gehani et al. [7] mapped the plain text to cipher text using one-time pad. One-time pad is a DNA strand consisting of the plain text word and corresponding unique cipher text pair along with a stop codon. All one-time pads were assembled and shared between the sender and the receiver. Using the assembled one-time pad, the plain text was encrypted by substituting the plain text word with the corresponding cipher text and decoded by replacing cipher text with plain text.

Binary information was encoded by Leier et al. [8] by taking four short DNA sequences to represent binary 0's, binary 1's, start and end markers. These were combined with start marker in the beginning of the DNA and end marker towards the end to form the synthesized DNA strand. The esulting DNA sequence was mixed with dummy DNA strands. Analogous to Clelland et al. they can only be decoded and obtained by knowing the primer sequences.

All the above DNA steganographic algorithms used synthetic DNA sequences to store binary information. They, however, changed the target DNA sequence when introduced into living organisms thus disrupting the biological function of the cell.

Arita et al. [10] were the first to consider the biological aspect of DNA while hiding secret information. They tried to ensure that mutation does not occur while substitution, however they did not succeed entirely in conserving the biological functionality of DNA. Moreover they translated each letter of the English alphabet into six codons.

Dominik Heider et al. [11] used the DNA Crypt algorithm based on the least significant base analogous to the least significant bit in image stagenography. They had to combine binary encryption algorithms like AES, RSA or Blowfish. DNA-Crypt to correct mutations in the target DNA with several mutation correction codes such as the Hamming-code or the WDH-code.

In our work, we used a novel algorithm to hide the secret information in putative protein coding regions (ORF) of eukaryotic DNA without affecting its biological function. The secret information was mapped into a synthetic DNA strand and placed in the ORF region of the DNA. The nucleotides of the synthetic DNA strand were assigned to the rightmost nucleotide of the randomly selected cordon in cover DNA by looking up at the genetic table ensuring that the assignment caused no mutation. This algorithm keeps the DNA intact as same protein is formed thereby preserving its function.

II. METHODOLOGY

DNA to be used as cover to hide the secret text is fetched and the protein coding regions (ORF) located. Secret text consisting of alpha-numerals and special characters is mapped into synthetic DNA strand using an encoding scheme. The synthetic DNA strand is assigned to the cover DNA using the embedding algorithm and stego-key generated. The stego-medium along with the stego-key is transmitted to the receiver. At the receiver end using the stego-key, the synthetic DNA is obtained and decoded using the inverse encryption algorithm to retrieve the hidden information. The process is elaborated.

C. Locating the protein coding region (OFR) in cover DNA

DNA to be used as cover media is downloaded from the NCBI (National Center for Biotechnology Information) and its reverse complementary strand generated. The reverse complementary strand is obtained by replacing A with T and G with C in the original sequence. The original sequence is divided into 3 different reading frames (+1, +2, +3). Subsequently the reverse complementary sequence is also divided into 3 different reading frames (-1, -2 and -3) together generating 6 reading frames. The reading frame that is used determines which amino acids will be encoded by a gene. The first reading frame is obtained by considering the original DNA sequence from the first nucleotide into words of 3 nucleotides. The second reading frame is formed after leaving the first nucleotide and then grouping the original DNA nucleotide sequence into words of 3 nucleotides. The third reading frame is formed after leaving the first 2 nucleotides and then grouping the original DNA nucleotide sequence into words of 3 nucleotides. The other 3 reading frames likewise are obtained in the similar manner using the reverse complementary strand.

Thereafter, the start codon (ATG) and stop codons (TAA, TAG, TGA) for each of the 6 reading frame is marked. An Open Reading Frame (ORF) starts with an ATG (Met) in most species and ends with a stop codon (TAA, TAG and TGA). Most often, the longest ORF is used in translating a gene (in eukaryotes) and therefore the frame with longest ORF is selected as the protein coding region. DNA nucleotide sequence of the located ORF is translated into its corresponding amino acid sequence based on the genetic codon table as depicted in Fig. 2 to form a protein. Fig. 4 illustrates the process of locating the ORF of a toy DNA.

Fig. 4 Locating ORF of a toy DNA

D. Encoding Scheme: Transforming secret information to synthetic DNA strand

After locating the protein coding region (ORF), the secret information has to be transformed into synthetic DNA strand. For this, each alphanumeric character of the secret information is first converted into its corresponding ASCII value using the ASCII table (Supplementary file Table I). 256 unique combinations are possible using 4 nucleotides (4^4) at a time, each of which is assigned to one of the possible 256 ASCII value (Supplementary file Table II). Using this encoding scheme, the ASCII converted secret information is transformed into a synthetic DNA strand.

E. Embedding algorithm: Hiding the secret information into the protein coding region.

In order to hide the synthetic DNA strand representing secret information into the cover media, the synthetic DNA strand is scanned from left to right, one nucleotide at a time. A codon representing an amino acid is randomly selected from the protein coding region (ORF) of the cover DNA and compared to the rightmost nucleotide of the randomly selected codon.

a) If they match, the nucleotide of the synthetic DNA strand replaces the rightmost nucleotide of the randomly selected codon in the protein coding region of the cover DNA.
b) If they do not match and the substitution of the nucleotide of the synthetic DNA strand results in coding for the same amino acid i.e it does not mutate the codon in the protein coding region of the cover DNA, the nucleotide of the synthetic DNA strand and the corresponding index position in the cover DNA are noted in the stego-key.
c) If they do not match and the substitution of the nucleotide of the synthetic DNA strand in the codon of the protein coding region of the cover DNA causes mutation, the corresponding index position is ignored. The next random codon from the protein coding region (ORF) of the cover DNA which has not been selected previously is fetched and the above comparison procedure repeated.

In this manner the synthetic DNA strand is assigned to the protein coding region cover DNA and the stego-key generated.

F. Extracting algorithm: Retrieving the hidden information

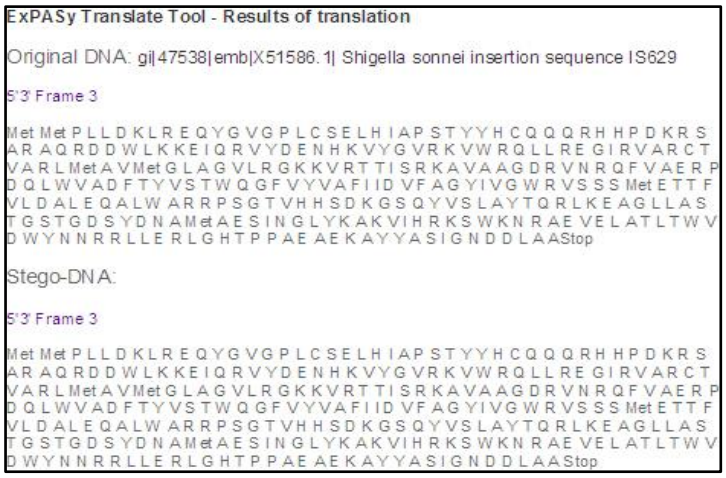
The cover DNA along with the stego-key is made available to the receiver. The stego-key contains the information about the location of each of the embedded nucleotide of the synthetic DNA representing the secret text and the nucleotides not causing mutation. Using the key, the embedded positions are located and the nucleotides extracted.

G. Inverse Encoding Scheme: Decoding the secret information

On getting all the nucleotides of the synthetic DNA strand, the nucleotides are grouped into four and the corresponding ASCII value obtained using the inverse encoding scheme (Supplementary file Table II). Subsequently, a look up at the ASCII table (Supplementary Table I) will fetch all the corresponding alphanumeric characters of the secret information retrieving the original information.

III. RESULTS AND DISCUSSION

We tested the algorithm on randomly selected nucleotides sequences (DNA) from the NCBI site. The original DNA used as cover and the stego-medium obtained was subjected to BLAST, a pairwise local alignment bioinformatics tool [14]. The result of pairwise alignment between the original cover DNA and the stego-medium DNA using BLAST revealed 100% similarity, 100% identity and 0% Gap in all the 50 test cases indicating that the DNA nucleotides are not mutated (Supplementary Table III). ExPASy [15], a tool that translates nucleotide (DNA/RNA) sequence to amino acid sequence, results showed that the protein obtained from the cover DNA and the stego-medium DNA are the same. Fig. 4 depicts the ExPASy translation results of an example cover DNA strand (gi|47538|) and stego-medium DNA strand.


ExPASy Translate Tool - Results of translation

Original DNA: gi|47538|emb|X51586.1| Shigella sonnei insertion sequence IS629

5'3' Frame 3

Met Met P L L D K L R E Q Y G V G P L C S E L H I A P S T Y Y H C Q Q Q R H H P D K R S
A R A Q R D D W L K K E I Q R V Y D E N H K V Y G V R K V W R Q L L R E G I R V A R C T
V A R L Met A V Met G L A G V L R G K K V R T T I S R K A V A A G D R V N R Q F V A E R P
D Q L W V A D F T Y V S T W Q G F V Y V A F I I D V F A G Y I V G W R V S S S Met E T T F
V L D A L E Q A L W A R R P S G T V H H S D K G S Q Y V S L A Y T Q R L K E A G L L A S
T G S T G D S Y D N A Met A E S I N G L Y K A K V I H R K S W K N R A E V E L A T L T W V
D W Y N N R R L L E R L G H T P P A E A E K A Y Y A S I G N D D L A A Stop

Stego-DNA:

5'3' Frame 3

Met Met P L L D K L R E Q Y G V G P L C S E L H I A P S T Y Y H C Q Q Q R H H P D K R S
A R A Q R D D W L K K E I Q R V Y D E N H K V Y G V R K V W R Q L L R E G I R V A R C T
V A R L Met A V Met G L A G V L R G K K V R T T I S R K A V A A G D R V N R Q F V A E R P
D Q L W V A D F T Y V S T W Q G F V Y V A F I I D V F A G Y I V G W R V S S S Met E T T F
V L D A L E Q A L W A R R P S G T V H H S D K G S Q Y V S L A Y T Q R L K E A G L L A S
T G S T G D S Y D N A Met A E S I N G L Y K A K V I H R K S W K N R A E V E L A T L T W V
D W Y N N R R L L E R L G H T P P A E A E K A Y Y A S I G N D D L A A Stop

Fig. 4 Expasy translation results of original DNA and stego_DNA

IV. CONCLUSION

In our work, we attempted to embed secret information into the putative coding regions of eukaryotic DNA taking care not to disrupt its biological function. The secret information was mapped into synthetic DNA strand and its nucleotides were randomly assigned to the putative protein coding regions of the cover DNA not allowing any mutation to seep in. BLAST results between the cover DNA and the stego-medium reveal that there is no change in the DNA. ExPASy translation of a nucleotide sequence to amino acid sequence showed that the protein obtained from the cover

DNA and the stego-medium DNA are the same thus confirming that our algorithm does not change the biological function of the DNA thereby demonstrating that DNA could be used as an alternative potential cover media in steganographic communication.

REFERENCES

[1] Arvind Kumar, "Steganography- A Data Hiding Technique", *International Journal of Computer Applications*, vol. 9, Nov. 2010.

[2] Ross J. Anderson and Fabien A.P. Petitcolas, "On The Limits of Steganography", *IEEE Journal of Selected Areas in Communications*, vol. 16 pp. 474–481, May 1998.

[3] B. Pfitzmann, "Information Hiding Terminology", *Proc.First Int'l Workshop Information Hiding, Lecture Notes in Computer Science*, *Springer-Verlag, Berlin*, pp. 347-356, 1996.

[4] H. Z. Hsu and R. C. T. Lee, "DNA Based Encryption Methods", *The 23rd Workshop on Combinatorial Mathematics and Computation Theory*, 2006.

[5] Adleman, Leonard. "Molecular computation of solutions to combinatorial problems", *Science*, vol. 266, pp. 1021-1024, Nov. 11, 1994.

[6] Clelland C., Risca V. and Bancroft C., "Hiding messages in DNA microdots", *Nature*, vol. 399, pp. 533-534, 1999.

[7] Gehani A., LaBean T. H., Reif J. H.,"DNA-based cryptography", *DimacsSeries In Discrete Mathematics and Theoretical Computer Science*, vol. 54, pp. 233-249, Sept. 2000.

[8] Leier A., Richter C., Banzhaf W. and Rauhe H., "Cryptography with DNA binary strands", *BioSystems*, vol. 55, pp. 13-22, 2000.

[9] Wong P. C., Wong K. K. and Foote H., "Organic data memory using the DNA approach", *Communications of the ACM*, 46, 2003.

[10] Arita M. and Ohashi Y., "Secret signatures inside genomic DNA", *Biotechnol Prog*., vol. 2, pp. 1605-160, 2004,.

[11] Dominik Heider and Angelika Barnekow, "DNA-based watermarks using the DNA-Crypt algorithm", *BMC Bioinformatics* , vol. 8, pp. 176, May 2007

[12] F. H. C. Crick, "The origin of the genetic code", *Journal of Molecular Biology*, vol. 38, pp. 367–379, Dec. 1968.

[13] Achuthsankar S. Nair, "Computational Biology & Bioinformatics: A Gentle Overview", *Communications of the Computer Society of India*, Jan. 2007.

[14] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer1, Jinghui Zhang, Zheng Zhang, Webb Miller and David J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Research, vol. 25, pp. 3389-3402, July 2007.

[15] "ExPASy: the proteomics server for in-depth protein knowledge and analysis Oxford Journals Life Sciences ", *Nucleic Acids Research*, vol. 31, pp. 3784-3788, April 2003.

STEGANOGRAPHY IN PUTATIVE PROTEIN CODING REGIONS OF EUKARYOTIC DNA: A NOVAL APPROACH

Supplementary File

Table I
ASCII Table

har	Dec	Char	Dec	Char	Dec	Char	Dec	ar	Dec	Char	Dec	Char	Dec	Char	Dec
(nul)	0	(sp)	32	@	64	`	96	Ç	128	á	160	└	192	α	224
(soh)	1	!	33	A	65	a	97	ü	129	í	161	┴	193	ß	225
(stx)	2	"	34	B	66	b	98	é	130	ó	162	┬	194	Γ	226
(etx)	3	#	35	C	67	c	99	â	131	ú	163	├	195	π	227
(eot)	4	$	36	D	68	d	100	ä	132	ñ	164	─	196	Σ	228
(enq)	5	%	37	E	69	e	101	à	133	Ñ	165	┼	197	σ	229
(ack)	6	&	38	F	70	f	102	å	134	ª	166	╞	198	µ	230
(bel)	7	'	39	G	71	g	103	ç	135	º	167	╟	199	τ	231
(bs)	8	(	40	H	72	h	104	ê	136	¿	168	╚	200	Φ	232
(ht)	9	)	41	I	73	i	105	ë	137	⌐	169	╔	201	Θ	233
(nl)	10	*	42	J	74	j	106	è	138	¬	170	╩	202	Ω	234
(vt)	11	+	43	K	75	k	107	ï	139	½	171	╦	203	δ	235
(np)	12	,	44	L	76	l	108	î	140	¼	172	╠	204	∞	236
(cr)	13	-	45	M	77	m	109	ì	141	¡	173	═	205	φ	237
(so)	14	.	46	N	78	n	110	Ä	142	«	174	╬	206	ε	238
(si)	15	/	47	O	79	o	111	Å	143	»	175	╧	027	∩	239
(dle)	16	0	48	P	80	p	112	É	144	░	176	╨	208	≡	240
(dc1)	17	1	49	Q	81	q	113	æ	145	▒	177	╤	209	±	241
(dc2)	18	2	50	R	82	r	114	Æ	146	▓	178	╥	210	≥	242
(dc3)	19	3	51	S	83	s	115	ô	147	│	179	╙	211	≤	243
(dc4)	20	4	52	T	84	t	116	ö	148	┤	180	╘	212	⌠	244
(nak)	21	5	53	U	85	u	117	ò	149	╡	181	╒	213	⌡	245
(syn)	22	6	54	V	86	v	118	û	150	╢	182	╓	214	÷	246
(etb)	23	7	55	W	87	w	119	ù	151	╖	183	╫	215	≈	247
(can)	24	8	56	X	88	x	120	ÿ	152	╕	184	╪	216	°	248
(em)	25	9	57	Y	89	y	121	Ö	153	╣	185	┘	217	∙	249
(sub)	26	:	58	Z	90	z	122	Ü	154	║	186	┌	218	•	250
(esc)	27	;	59	['	91	{	123	¢	155	╗	187	█	219	√	251
(fs)	28	<	60	\	92	|	124	£	156	╝	188	▄	220	ⁿ	252
(gs)	29	=	61	]	93	}	125	¥	157	╜	189	▌	221	²	253
(rs)	30	>	62	^	94	~	126	₧	158	╛	190	▐	222	■	254
(us)	31	?	63	_	95	(del)	127	ƒ	159	┐	191	▀	223		255

Table II
ASCII to DNA conversion table

ASCII	**DNA**	**ASCII**	DNA	**ASCII**	**DNA**	**ASCII**	**DNA**	**ASCII**	**DNA**	**ASCII**	**DNA**	**ASCII**	**DNA**	**ASCII**	**DNA**
0	AAAA	32	ACAA	64	GAAA	96	GCAA	128	GAAA	160	GCAA	192	TAAA	224	TCAA
1	AAAG	33	ACAG	65	GAAG	97	GCAG	129	GAAG	161	GCAG	193	TAAG	225	TCAG
2	AAAC	34	ACAC	66	GAAC	98	GCAC	130	GAAC	162	GCAC	194	TAAC	226	TCAC
3	AAAT	35	ACAT	67	GAAT	99	GCAT	131	GAAT	163	GCAT	195	TAAT	227	TCAT
4	AAGA	36	ACGA	68	GAGA	100	GCGA	132	GAGA	164	GCGA	196	TAGA	228	TCGA
5	AAGG	37	ACGG	69	GAGG	101	GCGG	133	GAGG	165	GCGG	197	TAGG	229	TCGG
6	AAGC	38	ACGC	70	GAGC	102	GCGC	134	GAGC	166	GCGC	198	TAGC	230	TCGC
7	AAGT	39	ACGT	71	GAGT	103	GCGT	135	GAGT	167	GCGT	199	TAGT	231	TCGT
8	AACA	40	ACCA	72	GACA	104	GCCA	136	GACA	168	GCCA	200	TACA	232	TCCA
9	AACG	41	ACCG	73	GACG	105	GCCG	137	GACG	169	GCCG	201	TACG	233	TCCG
10	AACC	42	ACCC	74	GACC	106	GCCC	138	GACC	170	GCCC	202	TACC	234	TCCC
11	AACT	43	ACCT	75	GACT	107	GCCT	139	GACT	171	GCCT	203	TACT	235	TCCT
12	AATA	44	ACTA	76	GATA	108	GCTA	140	GATA	172	GCTA	204	TATA	236	TCTA
13	AATG	45	ACTG	77	GATG	109	GCTG	141	GATG	173	GCTG	205	TATG	237	TCTG
14	AATC	46	ACTC	78	GATC	110	GCTC	142	GATC	174	GCTC	206	TATC	238	TCTC
15	AATT	47	ACTT	79	GATT	111	GCTT	143	GATT	175	GCTT	027	TATT	239	TCTT
16	AGAA	48	ATAA	80	GGAA	112	GTAA	144	GGAA	176	GTAA	208	TGAA	240	TTAA
17	AGAG	49	ATAG	81	GGAG	113	GTAG	145	GGAG	177	GTAG	209	TGAG	241	TTAG
18	AGAC	50	ATAC	82	GGAC	114	GTAC	146	GGAC	178	GTAC	210	TGAC	242	TTAC
19	AGAT	51	ATAT	83	GGAT	115	GTAT	147	GGAT	179	GTAT	211	TGAT	243	TTAT
20	AGGA	52	ATGA	84	GGGA	116	GTGA	148	GGGA	180	GTGA	212	TGGA	244	TTGA
21	AGGG	53	ATGG	85	GGGG	117	GTGG	149	GGGG	181	GTGG	213	TGGG	245	TTGG
22	AGGC	54	ATGC	86	GGGC	118	GTGC	150	GGGC	182	GTGC	214	TGGC	246	TTGC
23	AGGT	55	ATGT	87	GGGT	119	GTGT	151	GGGT	183	GTGT	215	TGGT	247	TTGT
24	AGCA	56	ATCA	88	GGCA	120	GTCA	152	GGCA	184	GTCA	216	TGCA	248	TTCA
25	AGCG	57	ATCG	89	GGCG	121	GTCG	153	GGCG	185	GTCG	217	TGCG	249	TTCG
26	AGCC	58	ATCC	90	GGCC	122	GTCC	154	GGCC	186	GTCC	218	TGCC	250	TTCC
27	AGCT	59	ATCT	91	GGCT	123	GTCT	155	GGCT	187	GTCT	219	TGCT	251	TTCT
28	AGTA	60	ATTA	92	GGTA	124	GTTA	156	GGTA	188	GTTA	220	TGTA	252	TTTA
29	AGTG	61	ATTG	93	GGTG	125	GTTG	157	GGTG	189	GTTG	221	TGTG	253	TTTG
30	AGTC	62	ATTC	94	GGTC	126	GTTC	158	GGTC	190	GTTC	222	TGTC	254	TTTC
31	AGTT	63	ATTT	95	GGTT	127	GTTT	159	GGTT	191	GTTT	223	TGTT	255	TTTT

Table III
BLAST pairwise alignment between cover DNA and stego-medium

o	Original file	Length:	# Identity:	# Similarity:	# Gaps:
1	gi|9626685|ref|NC_001477.1| Dengue virus 1, complete genome	10735	10735/10735 (100.0%)	10735/10735 (100.0%)	0/10735 (0.0%)
2	gi|47538|emb|X51586.1| Shigella sonnei	1310	1310/1310 (100.0%)	1310/1310 (100.0%)	0/1310 (0.0%)
3	gi|22028446|dbj|E10908.1| DNA encoding Bacillus cyclic isomalto-oligosaccharide synthetase	1310	1310/1310 (100.0%)	1310/1310 (100.0%)	0/1310 (0.0%)
4	gi|2175112|dbj|E06957.1| DNA encoding a heat-resistant alkaline protease	1083	1083/1083 (100.0%)	1083/1083 (100.0%)	0/1083 (0.0%)
5	gi|2172051|dbj|E03837.1| DNA encoding rat adrenaline receptor beta-1	1398	1398/1398 (100.0%)	1398/1398 (100.0%)	0/1398 (0.0%)
6	>gi|12831192|gb|AF333324.1| Hepatitis C virus type 1b polyprotein mRNA, complete cds	9587	9587/9587 (100.0%)	9587/9587 (100.0%)	0/9587 (0.0%)
7	gi|160857876|emb|AM501482.1| Vaccinia virus Ankara strain chorioallantois vaccinia virus Ankara (CVA), complete coding genome	192353	192353/192353 (100.0%)	192353/192353 (100.0%)	0/192353 (0.0%)
8	gi|33413956|gb|AY232749.1| Hepatitis C virus clone MD2b10-2 polyprotein mRNA, complete cds	9406	9406/9406 (100.0%)	9406/9406 (100.0%)	0/9406 (0.0%)
9	gi|15422182|gb|AY051292.1| Hepatitis C virus (isolate India) polyprotein mRNA, complete cds	9441	9441/9441 (100.0%)	9441/9441 (100.0%)	0/9441 (0.0%)
10	gi|780375|gb|U18466.1|ASU18466 African swine fever virus, complete genome	170101	170101/170101 (100.0%)	170101/170101 (100.0%)	0/170101 (0.0%)
11	gi|633201|emb|X76918.1| Hepatitis C virus genes for core, envelope and NS1 proteins	9390	9390/9390 (100.0%)	9390/9390 (100.0%)	0/9390 (0.0%)
12	gi|56692997|ref|NC_006556.1| Thermoproteus tenax spherical virus 1, complete genome	20933	20933/20933 (100.0%)	20933/20933 (100.0%)	0/20933 (0.0%)
13	gi|18450236|ref|NC_001132.2| Myxoma virus, complete genome	161773	161773/161773 (100.0%)	161773/161773 (100.0%)	0/161773 (0.0%)
14	gi|119352440|gb|DQ121394.1| Vaccinia virus strain Lister clone VACV107, complete genome	189421	189421/189421 (100.0%)	189421/189421 (100.0%)	0/189421 (0.0%)
15	gi|528204214|gb|ATMS01000011.1| Paenibacillus alvei A6-6i-x PAAL66ix_50, whole genome shotgun sequence	24167	24167/24167 (100.0%)	24167/24167 (100.0%)	0/24167 (0.0%)
16	gi|2171639|dbj|E03423.1| DNA encoding STIb of entero toxigenic Escherichia coli	258	258/258 (100.0%)	258/258 (100.0%)	0/258 (100.0%)
17	gi|2171638|dbj|E03422.1| DNA encoding STIa of entero toxigenic	265	265/265 (100.0%)	265/265 (100.0%)	0/265 (0.0%)

	Escherichia coli				
18	gi|2171637|dbj|E03421.1| DNA encoding LTh of entero toxigenic Escherichia coli	1148	1148/1148 (100.0%)	1148/1148 (100.0%)	0/1148 (0.0%)
19	gi|22027811|dbj|E10718.1| DNA encoding ilvGMEDA operon from Escherichia coli	2841	2841/2841 (100.0%)	2841/2841 (100.0%)	0/2841 (0.0%)
20	gi|5711170|dbj|E16487.1| Escherichia coli dgi (gyrI) gene for DNA gyrase-inhibitory protein complete cds	1224	1224/1224 (100.0%)	1224/1224 (100.0%)	0/1224 (0.0%)
21	gi|2170415|dbj|E02177.1| DNA sequence of aminopeptidase P	1850	1850/1850 (100.0%)	1850/1850 (100.0%)	0/1850 (0.0%)
22	gi|2169705|dbj|E01449.1| Genomic DNA encoding catechol producing enzyme	5273	5273/5273 (100.0%)	5273/5273 (100.0%)	0/5273 (0.0%)
24	gi|22026028|dbj|E09401.1| DNA fragment participating cell division	3187	3187/3187 (100.0%)	3187/3187 (100.0%)	0/3187 (0.0%)
25	gi|13020764|dbj|E27911.1| Method for detecting foreign DNA fragment insert in Vero toxin gene	1369	1369/1369 (100.0%)	1369/1369 (100.0%)	0/1369 (0.0%)
26	gi|2176669|dbj|E08554.1| DNA encoding phosphotransacetylase EC 2.3.1.8	2136	2136/2136 (100.0%)	2136/2136 (100.0%)	0/2136 (0.0%)
27	gi|2169191|dbj|E00930.1| DNA encoding human IgE peptide and human IL-2 peptide	849	849 /849 (100.0%)	849 /849 (100.0%)	0 /849 (0.0%)
28	gi|2168738|dbj|E00455.1| DNA coding for hepatitis A virus antigen	2980	2980/2980 (100.0%)	2980/2980 (100.0%)	0/2980 (0.0%)
29	gi|22025020|dbj|E11386.1| DNA encoding Escherichia nitrogenase	723	723/723 (100.0%)	723/723 (100.0%)	0/723 (0.0%)
30	gi|357090133|gb|JN675621.1| Rattus rattus LIV isolate 124RrIV_61 cytochrome b (cytb) gene, partial cds; mitochondrial	945	945/945 (100.0%)	945/945 (100.0%)	0/945 (0.0%)
31	gi|2599312|gb|AF000202.1| Rattus sp. T-612 retrotransposon mlvi2-rm6 5'UTR and putative RNA binding protein gene, partial cds	1117	1117/1117 (100.0%)	1117/1117 (100.0%)	0/1117 (0.0%)
32	gi|299481206|gb|HM217751.1| Rattus rattus voucher T0827 interphotoreceptor retinoid binding protein (IRBP) gene, partial cds	1012	1012/1012 (100.0%)	1012/1012 (100.0%)	0/1012 (0.0%)
33	gi|299481196|gb|HM217746.1| Rattus rattus voucher T0814 interphotoreceptor retinoid binding protein (IRBP) gene, partial cds	1140	1140/1140 (100.0%)	1140/1140 (100.0%)	0/1140 (0.0%)

34	gi|388240422|dbj|AP012462.1| Rattus norvegicus DNA, BAC clone: RNB1-434D18, strain: F344/Stm, complete sequence	104609	104609/104609 (100.0%)	104609/104609 (100.0%)	0/104609 (0.0%)
35	gi|388240417|dbj|AP012457.1| Rattus norvegicus DNA, BAC clone: RNB1-392D09, strain: F344/Stm, complete sequence	93805	93805/93805 (100.0%)	93805/93805 (100.0%)	0/93805 (0.0%)
36	gi|3452547|emb|AJ232238.1| Yersinia pestis 16S rRNA gene, isolate: SS-Yp-116	1469	1469/1469 (100.0%)	1469/1469 (100.0%)	0/1469 (0.0%)
37	gi|391703034|gb|AKTK01000001.1| Yersinia pestis PY-95 PY_95.contig.0_1, whole genome shotgun sequence	7767	7767/7767 (100.0%)	7767/7767 (100.0%)	0/7767 (0.0%)
38	gi|391702979|gb|AKTK01000003.1| Yersinia pestis PY-95 PY_95.contig.3_1, whole genome shotgun sequence	3020	3020/3020 (100.0%)	3020/3020 (100.0%)	0/3020 (0.0%)
39	gi|391702960|gb|AKTK01000004.1| Yersinia pestis PY-95 PY_95.contig.3_2, whole genome	17953	17953/17953 (100.0%)	17953/17953 (100.0%)	0/17953 (0.0%)
39	gi|391702836|gb|AKTK01000007.1| Yersinia pestis PY-95 PY_95.contig.6_1, whole genome shotgun sequence	7375	7375/7375 (100.0%)	7375/7375 (100.0%)	0/7375 (0.0%)
40	gi|391702807|gb|AKTK01000008.1| Yersinia pestis PY-95 PY_95.contig.7_1, whole genome shotgun sequence	2593	2593/2593 (100.0%)	2593/2593 (100.0%)	0/2593 (0.0%)
41	gi|391702780|gb|AKTK01000009.1| Yersinia pestis PY-95 PY_95.contig.8_1, whole genome shotgun sequence	15042	15042/15042 (100.0%)	15042/15042 (100.0%)	0/15042 (0.0%)
42	gi|391702771|gb|AKTK01000010.1| Yersinia pestis PY-95 PY_95.contig.8_2, whole genome shotgun sequence	5505	5505/5505 (100.0%)	5505/5505 (100.0%)	0/5505 (0.0%)
43	gi|391702759|gb|AKTK01000011.1| Yersinia pestis PY-95 PY_95.contig.9_1, whole genome shotgun sequence	6950	6950/6950 (100.0%)	6950/6950 (100.0%)	0/6950 (0.0%)
44	gi|391702710|gb|AKTK01000013.1| Yersinia pestis PY-95 PY_95.contig.12_1, whole genome shotgun sequence	3528	3528/3528 (100.0%)	3528/3528 (100.0%)	0/3528 (0.0%)
45	gi|391702660|gb|AKTK01000015.1| Yersinia pestis PY-95 PY_95.contig.14_1, whole genome shotgun sequence	28625	28625/28625 (100.0%)	28625/28625 (100.0%)	0/28625 (0.0%)
46	gi|391702638|gb|AKTK01000016.1|	4879	4879/4879	4879/4879	0/4879

	Yersinia pestis PY-95 PY_95.contig.17_1, whole genome shotgun sequence		(100.0%)	(100.0%)	(0.0%)
47	gi|391702594|gb|AKTK01000018.1| Yersinia pestis PY-95 PY_95.contig.19_1, whole genome shotgun sequence	15274	15274/15274 (100.0%)	15274/15274 (100.0%)	0/15274 (0.0%)
48	gi|391702547|gb|AKTK01000019.1| Yersinia pestis PY-95 PY_95.contig.20_1, whole genome shotgun sequence	9177	9177/9177 (100.0%)	9177/9177 (100.0%)	0/9177 (0.0%)
49	gi|391702513|gb|AKTK01000021.1| Yersinia pestis PY-95 PY_95.contig.20_3, whole genome	960	960/960 (100.0%)	960/960 (100.0%)	0/960 (0.0%)
50	gi|391702379|gb|AKTK01000025.1| Yersinia pestis PY-95 PY_95.contig.26_1, whole genome shotgun sequence	41438	41438/41438 (100.0%)	41438/41438 (100.0%)	0/41438 (0.0%)

.

