An Investigation of DNA Architecture from a Software Engineering Perspective - the software of life

V1.0– Paul Watkinson MSc (Masters in Software Engineering from Kellogg College, Oxford University) – 18/11/2013

In nature we see at first randomness, at least from a distance. But on closer inspection we discover repeatable, consistent shapes in blades of grass, in leaves on the trees and feathers on birds. Why are they so consistent? Why do we not sometimes see oak leaves on beech trees, tulips on rose bushes, feathers on dogs? Why do leaves have a consistent identity on a tree without variation? It seems that there is a process behind this consistency and it is my hypothesis that this is "Decidable" and may indicate a hidden code in DNA. Craig Venter states "In the span of a single lifetime, we have advanced from Schrodinger's 'aperiodic crystal' to an understanding of the genetic code to the proof, through construction of a synthetic chromosome and hence a synthetic chromosome and hence a synthetic cell, that DNA is the software of life."[p7 26][p75 24][28]

In Software, it is exceedingly difficult to develop a program that has no errors, Gibbons states "We have been aware for more than twenty years that any attempt to construct programs by trial and error is doomed to failure clearly a more systematic approach than this is required." and yet nature appears to do it with ease, repeatedly in many different species and over millions of years. [29] So what is nature's secret? While working on my master's thesis at Kellogg College, Oxford University, where I investigated the Laws of Software Engineering, I noticed that there were similarities with the natural world and concluded that to make software "Decidable" was the ultimate objective in order to eliminate bugs. This could be achieved by ensuring the logic used in creating this software was "Decidable" by restricting it to natural deduction and forcing it to form a hierarchy. It should also be noted that the logic must be well formed to be correct and I have proposed an XML format for logic as it is difficult to describe requirements so they can easily be transcribed for the computer. [1][30]

Although Watson and Crick identified the structure of DNA in nature, there may be more than is currently understood about how the program it contains is made "Decidable". We also know that, in spite of decoding elements of the DNA to identify what is used to build components, the vast majority is referred to as "junk" and we do not understand its function. Watson states in his 2004 book on DNA published 50 years after he discovered the double helix with Crick: *"Fifty percent of the genome is constituted of repetitive junk like sequences of no apparent function; a full 10 percent of our DNA consists of a million scattered copies of a single sequence, called Alu:". [6 p204]* The repeated "Alu" element may have a function within a wider context. If you ask the question what is the most frequent repetitive element of a computer program, you would probably conclude that it is an element of logic such as "And". We know that all of logic can be reduced to very primitive elements, which is where Gentzen's Natural Deduction comes in, and it would appear that the "Alu" may contain elements of logic. We know that DNA is also hierarchical so there is a similarity with the Laws of Software Engineering, Watson states: *"The more universal message of their work is that genetic information is hierarchically organised".* [6 p227][11]

John Von Neumann based his architecture paper for the computer on one written by McCulloch and Pits on mapping neurons to logic, which used Turing's paper "On Computable Numbers" during the Second World War, and leads to the computer as we use it today. [4] Sir Roger Penrose explains how difficult it is to tile a surface. [2] [5] This is another way of looking at the same problem, how are we to make the pattern decidable? Stephen Wolfram has appeared to come to a similar conclusion with

his “A New Kind of Science”, where he finds that very simple programs can produce beautiful ordered patterns but also that they can generate utter randomness. There appears to be a tipping point where the program goes from being decidable to undecidable. [3] Valiant states “*These laws of computation apply to all algorithms. Because ecorithms are algorithms, though of a special kind, they too must follow the same basic laws as computation in general. This new science of the ultimate limitations on the possibility and the efficiency with which computations for learning and evolution can proceed offers a fundamental new approach to understanding these phenomena of learning and evolution, because, regardless of how they are implemented – in silicon, DNA, Neurons, or something else entirely – there are some ultimate logical laws that limit what these mechanisms can do.*”[27]

In the following Alu sequence there are repeated strings, given this is in base 4 (A,C,T,G), there is quite sufficient to code a logic statement. [7]

Position	DNA Sequence
1	GCCGGGCGCG GTGGCGCGTG CCTGTAGTCC CAGCTACTCG GGAGGCTGAG GCTGGAGGAT
61	CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
121	TTCGGCATCA ATATGGTGAC CTCCCGGGAG CGGGGGACCA CCAGGTTGCC TAAGGAGGGG
181	TGAACCGGCC CAGGTCGGAA ACGGAGCAGG TCAAAACTCC CGTGCTGATC AGTAGTGGGA
241	TCGCGCCTGT GAATAGCCAC TGCACTCCAG CCTGGGCAAC ATAGCGAGAC CCCGTCTCT

If we apply a=0, c=1, g=2 and t=3 to make it base 4 for the first sequence ”GCCGGGCGCG” we get “2112221212”, then if we convert it to binary we get the following “10010110101001100110”, and then we are into the world of software and computers. We also know that there are Alu sub families, which might indicate that they are the equivalent of statements in programming languages. There is a rule that arguably could be the application of “If Then”, which is as Crick states: “*Any one triplet would have only four other triplets as its neighbours on one side. For example, if the triplet in question was AAT, then the only triplets that could precede it could be TAA, CAA, AAA and GAA, while only ATT, ATC, ATA and ATG could follow it, assuming as always that the code was overlapping.”* [8 P97] There is also a pairing to consider of T=A and G=C. [8 P185] So you can see on line 241 above that the segment “GAATAGCCAC” has “GAA” preceding and “ATA” following.

Similarly a triplet “CGG” would have “TCG”,”CCG”,”ACG” and “GCG” precede it and “GGT”, “GGC”, “GGA” and “GGG” follow it. If we look at the above Alu it has a number of repeats of “CGG”, in the following table it identifies each occurrence and how it matches the possible permutations to indicate something interesting, a pattern of branches that you also see in software.

Preceding	Occurrence	Triplet	Following	Occurrence
ACG	8	CGG	GGA	4,7,8
CCG	1,4,6		GGC	3,6
GCG	5		GGG	1,2,4,5
TCG	2,3,7		GGT	

We can also look at this from another perspective, given the rule is that there are only four letters A,C,G and T. If we assume the sequence goes as follows using the alphabet for example, there is also a significance of RNA “UAG”, “UAA” and “UGA” which translate to “ATC”, “ATT” and “ACT” [10 p688]

DNA Codon	#	Base 64	DNA Codon	#	Base 64	DNA Codon	#	Base 64		#	Base 64
AAA	0	A	CAA	16	Q	GAA	32	g	TAA	48	w
AAC	1	B	CAC	17	R	GAC	33	h	TAC	49	x
AAG	2	C	CAG	18	S	GAG	34	i	TAG	50	y
AAT	3	D	CAT	19	T	GAT	35	j	TAT	51	z
ACA	4	E	CCA	20	U	GCA	36	k	TCA	52	0
ACC	5	F	CCC	21	V	GCC	37	l	TCC	53	1
ACG	6	G	CCG	22	W	GCG	38	m	TCG	54	2
ACT	7	H	CCT	23	X	GCT	39	n	TCT	55	3
AGA	8	I	CGA	24	Y	GGA	40	o	TGA	56	4
AGC	9	J	CGC	25	Z	GGC	41	p	TGC	57	5
AGG	10	K	CGG	26	a	GGG	42	q	TGG	58	6
AGT	11	L	CGT	27	b	GGT	43	r	TGT	59	7
ATA	12	M	CTA	28	c	GTA	44	s	TTA	60	8
ATC	13	N	CTC	29	d	GTC	45	t	TTC	61	9
ATG	14	O	CTG	30	e	GTG	46	u	TTG	62	+
ATT	15	P	CTT	31	f	GTT	47	v	TTT	63	/

Using the base 64 table above there are overlapping triplets called Codons so the second string on line 1 gtggcgcgtg would be "TFVMGT", not a word in English but maybe in a different language? Perhaps an acronym or even text-speak (such as 'Thanks For Verifying My Genome Transcript').

So what language has 64 letters and might fit this Alu sequence, it could be in Base 64, here I have written a program at www.formalmodel.com to convert DNA to ASCII and then to BASE 64:

Type	Code
Base 64	lWaqpmZmaru6pmZmbu5lXe7syJt1VUSJncxHd2aqoiKpne4iKpne6oiKojN2Znf+4iJt1USKoiJv93e6qpne7syJu5mZnczO5lWYjN2aqru7t1WZkRHcwCJv92apkTN0QDMzO6ru4hFXd1VWaqoiJmaqqqohFURFUSKrv+5lXcwCKoiKqqru4gBFWaplVUSKrt2aogABGaoiJkSKrt0QAABHd1VWbu5ne4jN0SJsyJu6qojN2ZmZlXe7u4gDMyJlURHe5kRHd1USJlXe6qpkQBETMyJmYilhFVVWbt3d3
Ascii	fffjffne]uUDGwf""*3vfw"mD"owUU"of3t@33;WwUVj"&f!DEQ"e]*EYeUD"&DGwUVng{"lw————————————————3"eQDGwU&Ud@3"fb"!UVn

It seems that there is some kind of code, by that we mean language (computer, human or alien) as a result of the rule. I thought initially it might be in some way used in the construction process or formed an operating system for consciousness but, on reading Crick's book, decided it could equally be a message. Venter put a "Watermark" into the DNA that he engineered with some quotations and a URL.[20] Crick stated: "… *perhaps life on earth originated on another planet* ...". He calculated that the length of time in the universe would allow for multiple human evolutions to have occurred. [8 p148]

There is another possibility and that is that we wrote the code ourselves, that it is used as the mechanism to store our memories. It has been demonstrated that it is possible to write music,

letters and video into DNA form biologically in the lab, so if we can do it there, it could be possible that we have evolved a cell that can take our memories and write these into DNA. [9] We know that DNA is capable of storing vast amounts of data, far more than we could ever generate. A storage technique applied in very heavily used web sites is to write the data into more than one place to preserve it and avoid locking, so maybe the fact that the DNA exists in every cell that we have helps to preserve the storage of our own information, which is copied to multiple DNA strings.

It is possible to get mathematically from the Trigram in the DNA which is used to express proteins to Pythagorean triples and from there to the Fibonacci sequence which has been observed in living things such as flowers.[12][23][16] There has been a recent experiment where an element of the Alu was modified and it caused the change of shape of a mouse face as if the three dimensional surface was rescaled which indicates that there might be complex number or Pythagorean triple representing a three dimensional grid, in much the same manner as computer graphics. It certainly seems that the output produced when converting DNA to ASCII could feasibly contain this sort of information. [27]

What would be really useful is to compare the DNA code for two different humans and the DNA from the same human taken at two different times. While the first DNA decoding cost $3bn it is now possible to decode in approximately an hour. [28] An experiment could be carried out on DNA in a controlled way where a primitive organism is for example exposed to a coloured light and the DNA taken before and after the experiment. This would show whether DNA changes due to external effects.

This is all a hypothesis and a different way of thinking about the problem that DNA may contain a "Decidable" program, that what was thought of as "Junk" may not be so after all, but instead a kind of language - such as a computer program. We just have to figure out the language. But one thing is certain, if it is a program then it will contain simple logic constructs, so that is a good starting point for the search and the decryption.

It is a little like the efforts of Turing and others at Bletchley Park during the second world war trying to decrypt the German secret messages. Watson and Crick started the process by identifying the message, but over the past 50 years academics all over the world have been trying to make sense of it. Watson states: *"But there is nowhere to go but back to the future, for even with the entire human genome in hand, the program and cues according to which its instructions are carried out remain a colossal mystery.*" [6 p230]

This may just be another clue which opens a door into another world of our understanding. Turing said that around about now we should be able to understand how to write software that is intelligent, maybe he had in mind that we should have decrypted the gene to work out its secret.

Acknowledgements

I would like to thank the following for their wisdom and patience in discussing this matter with me, even though I am responsible for all of the errors: Professor of Software Engineering, Jeremy Gibbons, Kellogg College, Oxford University; Dr. Robert Lockhart, Director of Studies in Computing and Mathematics, Kellogg College, Oxford University; Professor Emeritus of Biochemistry, Michael

Yudkin, Kellogg College, Oxford University ; Dr. Cas Cremers, Cyber Security Centre, Kellogg College, Oxford University.

[1] An Investigation of the Laws of Software Engineering. Paul Watkinson http://fedcsis.eucip.pl/proceedings/pliks/128.pdf

[2] The Emperor's New Mind – Sir Roger Penrose.

[3] A New Kind of Science – Stephen Wolfram

[4] Collected Works of A.M.Turing – Mathematical Logic. R. O. Gandy.

[5] Lecture by Sir Roger Penrose at a meeting titled "Geometry - Euclid to Einstein" organised jointly by the Department for Continuing Education, Oxford University, and the British Society for the History of Mathematics. 22/23 June 2013

[6] DNA – The Secret of Life – 2004 – James Watson – ISBN 978-0-09-945184-6

[7] Alu – example - http://www.ncbi.nlm.nih.gov/nuccore/NR_002715.1

[8] What Mad Pursuit, Francis Crick ISBN 13-978-0-465-09138-6

[9] http://www.nature.com/nature/journal/v494/n7435/full/nature11875.html

(accessed 25/10/2013)

[10] Metamagical Themas: Questing for the Essence of Mind and Pattern, Douglas R. Hofstadter.

[11] The Blind Watchmaker, Richard Dawkins.

[12] A Guidebook to Biochemistry, Michael Yudkin, Robin Orford. Cambridge University Press.

http://books.google.co.uk/books?id=QcA5kBKwxVAC&pg=PA198&lpg=PA198&dq=Michael+Yudkin+DNA&source=bl&ots=UhClFA4bwT&sig=EHlHeD4A_iufscfxuzZjhM3JLoY&hl=en&sa=X&ei=lbZzUr25KJDo7Aa96oCICA&ved=0CCwQ6AEwAA#v=onepage&q=Michael%20Yudkin%20DNA&f=false

. . . U-U-G-A-C-C-C-C-A-A-G-G-U-A-C
. . . A-A-C-T-G-G-G-G-T-T-C-C-A-T-G
. . . T-T-G-A-C-C-C-C-A-A-G-G-T-A-C.

In this way the enzyme makes available an exact RNA transcript of the DNA. More accurately we should say that the RNA polymerase can make available an exact RNA transcript of any desired length of the DNA. The DNA contains particular sequences of bases at which the RNA polymerase binds to initiate transcription (see also p. 232), and other sequences of bases at which the RNA polymerase stops transcription. The length of DNA between a starting point and a stopping point is transcribed into a single molecule of RNA.

The great importance of this synthesis of RNA is that it enables proteins to be synthesized. In the next chapter we shall discuss in

[13] Life Rewritten into Alien Code, Linda Geddes, New Scientist, 26th October 2013.

So some of the codons encode the same amino acid – a phenomenon called redundancy. The Three combinations left over, UAG, UAA and UGA, act like a full stop or period – telling the ribosome to terminate the production process. (also 12 p170)

[14] http://bioweb.uwlax.edu/GenWeb/Molecular/Seq_Anal/Translation/translation.html

An open reading frame starts with an atg (Met) in most species and ends with a stop codon (taa, tag or tga).

Position	DNA Sequence
1	GCCGGGCGCG GTGGCGCGTG CCTGTAGTCC CAGCTACTCG GGAGGCTGAG GCTGGAGGAT
61	CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
121	TTCGGCATCA ATATGGTGAC CTCCCGGGAG CGGGGGACCA CCAGGTTGCC TAAGGAGGGG
181	TGAACCGGCC CAGGTCGGAA ACGGAGCAGG TCAAAACTCC CGTGCTGATC AGTAGTGGGA
241	TCGCGCCTGT GAATAGCCAC TGCACTCCAG CCTGGGCAAC ATAGCGAGAC CCCGTCTCT

[15] http://ds9a.nl/amazing-dna/

These comments are fascinating in their own right. Like C comments they have a start marker, like /*, and a stop marker, like */. But they have some more structure. Remember that DNA is like a tape - the comments need to be snipped out physically! The start of a comment is almost always indicated by the letters 'GT', which thus corresponds to /*, the end is signalled by 'AG', which is then like */.

However because of the snipping, some glue is needed to connect the code before the comment to the code after, which makes the comments more like html comments, which are longer: '<!--' signifies the start, '-->' the end.

DNA
GCCGGGCGCG GTGGCGCGTG CCTGTAGTCC CAGCTACTCG GGAGGCTGAG GCTGGAGGAT
1111111111 1011111101 1101001011 1011001011 1101110101 1101101100
CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
1110010101 1011010010 1111010010 1111000111 1001111010 1111010001
TTCGGCATCA ATATGGTGAC CTCCCGGGAG CGGGGGACCA CCAGGTTGCC TAAGGAGGGG
0011110010 0000110101 1011111101 1111110110 1101100100 0001101111
TGAACCGGCC CAGGTCGGAA ACGGAGCAGG TCAAAACTCC CGTGCTGATC AGTAGTGGGA
0100111111 1011011100 0111011011 0100001011 1101101001 0100101110
TCGCGCCTGT GAATAGCCAC TGCACTCCAG CCTGGGCAAC ATAGCGAGAC CCCGTCTCT
0111111010 1000011101 0110101101 1101111001 0101110101 111101010

GCCGGGCGCG	GTGGCGCGTG	CCTGTAGTCC	CAGCTACTCG	GGAGGCTGAG	GCTGGAGGAT

1111111111	1011111101	1101001011	1011001011	1101110101	1101101100
CGCTTGAGTC	CAGGAGTTCT	GGGCTGTAGT	GCGCTATGCC	GATCGGGTGT	CCGCACTAAG
1110010101	1011010010	1111010010	1111000111	1001111010	1111010001
TTCGGCATCA	ATATGGTGAC	CTCCCGGGAG	CGGGGGACCA	CCAGGTTGCC	TAAGGAGGGG
0011110010	0000110101	1011111101	1111110110	1101100100	0001101111
TGAACCGGCC	CAGGTCGGAA	ACGGAGCAGG	TCAAAACTCC	CGTGCTGATC	AGTAGTGGGA
0100111111	1011011100	0111011011	0100001011	1101101001	0100101110
TCGCGCCTGT	GAATAGCCAC	TGCACTCCAG	CCTGGGCAAC	ATAGCGAGAC	CCCGTCTCT
0111111010	1000011101	0110101101	1101111001	0101110101	111101010

[16] Numbers Matters, Lecture Notes for a Summer School Course at the Department for Continuing Education, Rewley House, Oxford. July 17 – July 24, 2010. Dr Robert Lockhart

Trigram -> Pythagorean triples -> Fibonacci

[17] Turing's "on computable"

[18] Godel's "undecidable"

[19] Human Molecular Genetics 3 Tom Strachan & Andrew P. Read

GT to AG is an intron according to [19]

Position	DNA Sequence
1	GCCGGGCGCG GTGGCGCGTG CCTGTAGTCC CAGCTACTCG GGAGGCTGAG GCTGGAGGAT
1	GCCGGGCGCG GTGGCGCGTG CCTGTAGTCC CAGCTACTCG GGAGGCTGAG GCTGGAGGAT
1	GCCGGGCGCG GTGGCGCGTG CCTGTAGTCC CAGCTACTCG GGAGGCTGAG GCTGGAGGAT
1	GCCGGGCGCG GTGGCGCGTG CCTGTAGTCC CAGCTACTCG GGAGGCTGAG GCTGGAGGAT
1	GCCGGGCGCG GTGGCGCGTG CCTGTAGTCC CAGCTACTCG GGAGGCTGAG GCTGGAGGAT
1	GCCGGGCGCG GTGGCGCGTG CCTGTAGTCC CAGCTACTCG GGAGGCTGAG GCTGGAGGAT
61	CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
61	CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
61	CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
61	CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
61	CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
61	CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
61	CGCTTGAGTC CAGGAGTTCT GGGCTGTAGT GCGCTATGCC GATCGGGTGT CCGCACTAAG
121	TTCGGCATCA ATATGGTGAC CTCCCGGGAG CGGGGGACCA CCAGGTTGCC TAAGGAGGGG
121	TTCGGCATCA ATATGGTGAC CTCCCGGGAG CGGGGGACCA CCAGGTTGCC TAAGGAGGGG
121	TTCGGCATCA ATATGGTGAC CTCCCGGGAG CGGGGGACCA CCAGGTTGCC TAAGGAGGGG
181	TGAACCGGCC CAGGTCGGAA ACGGAGCAGG TCAAAACTCC CGTGCTGATC AGTAGTGGGA
181	TGAACCGGCC CAGGTCGGAA ACGGAGCAGG TCAAAACTCC CGTGCTGATC AGTAGTGGGA
181	TGAACCGGCC CAGGTCGGAA ACGGAGCAGG TCAAAACTCC CGTGCTGATC AGTAGTGGGA
181	TGAACCGGCC CAGGTCGGAA ACGGAGCAGG TCAAAACTCC CGTGCTGATC AGTAGTGGGA
241	TCGCGCCTGT GAATAGCCAC TGCACTCCAG CCTGGGCAAC ATAGCGAGAC CCCGTCTCT
241	TCGCGCCTGT GAATAGCCAC TGCACTCCAG CCTGGGCAAC ATAGCGAGAC CCCGTCTCT

Formalmodel -> Number -> Base 4 to -> Ascii

Looks a little like a formula

Û]vÛmv×m·Ûmv×mö×]ößM·×]tÛ]ô×}vÛm6Û]öÓmµßm´Ûm7×mwßm6ß]tÛm6ß}wÛmµßmôÛ}µÛ]ôß muÛMõÛm·Û}uÛ]5ßM6ß}vÛ]7×M7Ó}¶ßm5×}u×m¶ÓmvÛm¶Ó]t×]6Û}ö×]ôÓm´Ûm¶ßm4×]¶×]tÛmõ Ûm4Ó]¶ÓmtÛmõÓM4×}u×mö×}´ß]6ßM·Ûm´ß]µÛ]wÛ}´Ó}6×]5ßmt×}uÓmußm¶×M5Ó}6×m6Ó]u×mõ

1) Using FormalModel:

Input: GTGGCGCGTG CCTGTAG

Out: 232212123211323

Out: u6pmZmbu5lXe7sy

2) To Base 10: 781826427

http://www.unitconversion.org/numbers/base-4-to-base-10-conversion.html

3) From Base 64 to ascii
   Input: u6pmZmbu5lXe7sy
   Output: »ªfffîæUÞÎÌ

http://home.paulschou.net/tools/xlate/

4) From Base 64

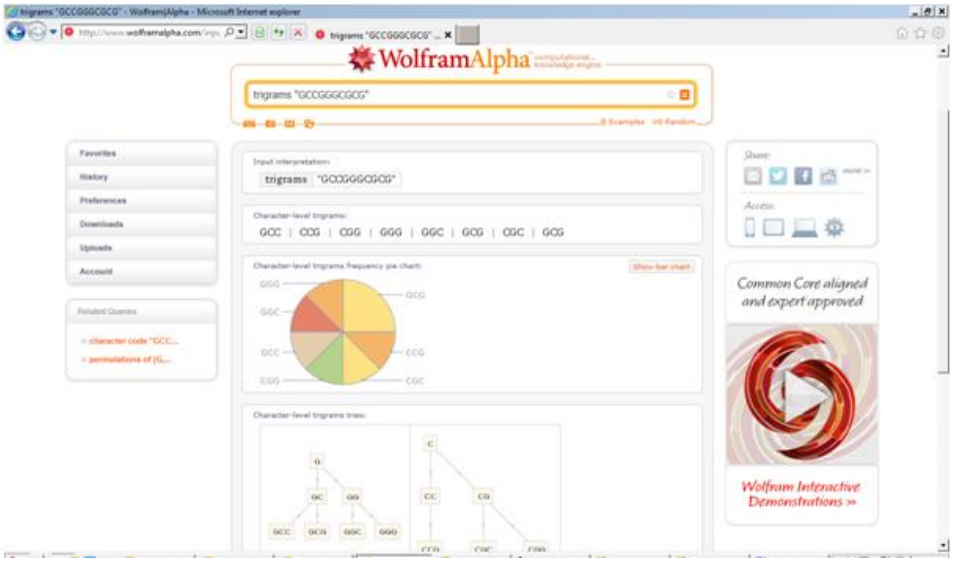
   Input: 232212123211323
   Output: Û}¶×mvßmußm

http://home.paulschou.net/tools/xlate/

[20] http://www.sciencemag.org/content/329/5987/52.long

We report the design, synthesis, and assembly of the 1.08–mega–base pair *Mycoplasma mycoides* JCVI-syn1.0 genome starting from digitized genome sequence information and its transplantation into a *M. capricolum* recipient cell to create new *M. mycoides* cells that are controlled only by the synthetic chromosome. The only DNA in the cells is the designed synthetic DNA sequence, including “watermark” sequences and other designed gene deletions and polymorphisms, and mutations acquired during the building process. The new cells have expected phenotypic properties and are capable of continuous self-replication.

[21] http://www.wolframalpha.com/input/?i=trigrams+%22GCCGGGCGCG%22



[23] Cryptological Mathematics, Robert Edward Lewland, The Mathematical Association of America

P37 The 15 Most Common Trigraphs in the English Language (1 = the, 2= and, 3=tha)

[24] What is Life, Edwin Schrodinger, A Double Day Anchor Book.[p 75]

“An organism’s astonishing gift of concentrating a ‘stream of order’ on itself and thus escaping the decay into atomic chaos – of ‘drinking orderliness’ from a suitable environment – seems to be connected with the presence of the ‘aperiodic solids’, the chromosome molecules, which doubtless represent the highest degree of well-ordered atomic association we know of – much higher than the ordinary periodic crystal – in virtue of the individual role every atom and every radical is playing here. To put it briefly, we witness the event that existing order displays the power of maintaining itself and of producing orderly events. That sounds plausible enough, though in finding it plausible we, no doubt, draw on experience concerning social organisation and other events which involve the activity of organisms. And so it might seem that something like a vicious circle is implied.”

[25] Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Daniel G. Gibson, J. Craig Venter, et al Science 52 (2010)

http://www.ncbi.nlm.nih.gov/pubmed/20488990?dopt=Citation

[26] Life at the speed of light. J. Craig Venter.

[27] Probably Approximately Correct, Leslie Valiant. Professor of Computer Science and Applied Mathematics at Harvard University.

[28] What is Life? Craig Venter. http://www.youtube.com/watch?v=qi2MhsUSu0U

[29] Algebras for Tree Algorithms, Jeremy Gibbons. http://www.cs.ox.ac.uk/files/3422/PRG94.pdf

[30] http://www.google.com/patents/US20130173522