NorthPoleInc.

PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, NORTHPOLE

Carolome: Functional Imprints of Culture Memes in the Global Genome.

Santa Claes[1]*, Feliz Navidad[2], Tomte Ness E.[3], 10 N. Baum[3], Jeremy Elf Tootoo[1], Dasher Sridhar[1,2,3], and Laura Menorah[2,3]

[1]*University of NorthPole, Dept Bioinformatic, Arctic Lane 1, Santa's Secret Village, NorthPole*

[2]*University of NorthPole, School of Rock, Arctic Lane 25, Santa's Secret Village, NorthPole*

[3]*University of NorthPole, Department of Air Guitar, Cool Avenue 365, Santa's Secret Village, NorthPole*

**Corresponding author: santa@DNA20.com*

Received 28 June 2010, received in revised form 25 October 2010, accepted 24 December 2010

ABSTRACT

The most recent version of the public genome database – GenBank – contains as of June 11, 2010 close to 3x 10^{11} base pairs. In line with studies attempting to identify all proteins derived from the database (proteomics), all metabolites (metabolomics) and all genes (genomics), we here have made a concerted effort to systematically identify all Christmas carols deposited in the sequence data. We here name this field of research Carolomics. The most abundant entry in the Carolome is 'Deck the Halls' (*Deck the halls with boughs of holly, Fa la la la la, la la la la*). We find this carol in 21 genomes. The second most prevalent carol in the Carolome is 'I Saw Mommy Kissing Santa Claus' found in 17 genomes including that of the wine grape suggesting a genetic link between mulled wine (aka Glögg) and Christmas celebration. Third most common carol in the Carolome is Ave Maria with 12 identified locations in the GenBank genomes. These findings establish a direct role for Christmas carols in the functional imprint and transfer of genetic information. In the future it will be essential for researchers to determine the presence of carolomes in sequence data; both to increase identified database constituents as well as to more fully and completely understand the proven transference of meme data between genomes.

keywords: synthetic biology, Carolome, omics, Santa Claus, bio-meme, parrots & cheese, Norwegian blue

INTRODUCTION

Quantum Superposition states that the position of a state in any given system is the combination of all the possible states of that system. Upon measuring the state, all possible states collapse into a definite state only at the exact moment of measurement (Heinz 2002).

A meme is a unit of cultural ideas, symbols or practices, which is transmitted in a social context from one individual to another through speech, gestures, rituals or other imitable phenomena. Memes can be considered as cultural analogues to genes, in that they self-replicate and respond to selective pressures (Kay and Mart 2001).

Prior to the widespread use of the efficient DNA sequencing methods devised by Fredrick Sanger (Sanger, Nicklen et al. 1977), genomic sequences were hidden from view. We accordingly infer that the information states of the 'global' genome was a combination of all possible states prior to the emergence of high throughput sequencing (mid-1980's). Upon initiation of high throughput sequencing, an information crystallization event must have occurred consistent with the law of Quantum Superposition. During this temporal event all distributed genomic information instantaneously must have condensed into defined and finite DNA sequence (Wait and See 1972).

Inherent endogenous memes have appeared spontaneously and are different from sequences engineered into the genome. The latter are best exemplified by the poem Tomten which was synthesized and cloned by Santa's elves a few years back (Gustafsson 2009). The poem can be retrieved using GenBank accession number EU600200.1.

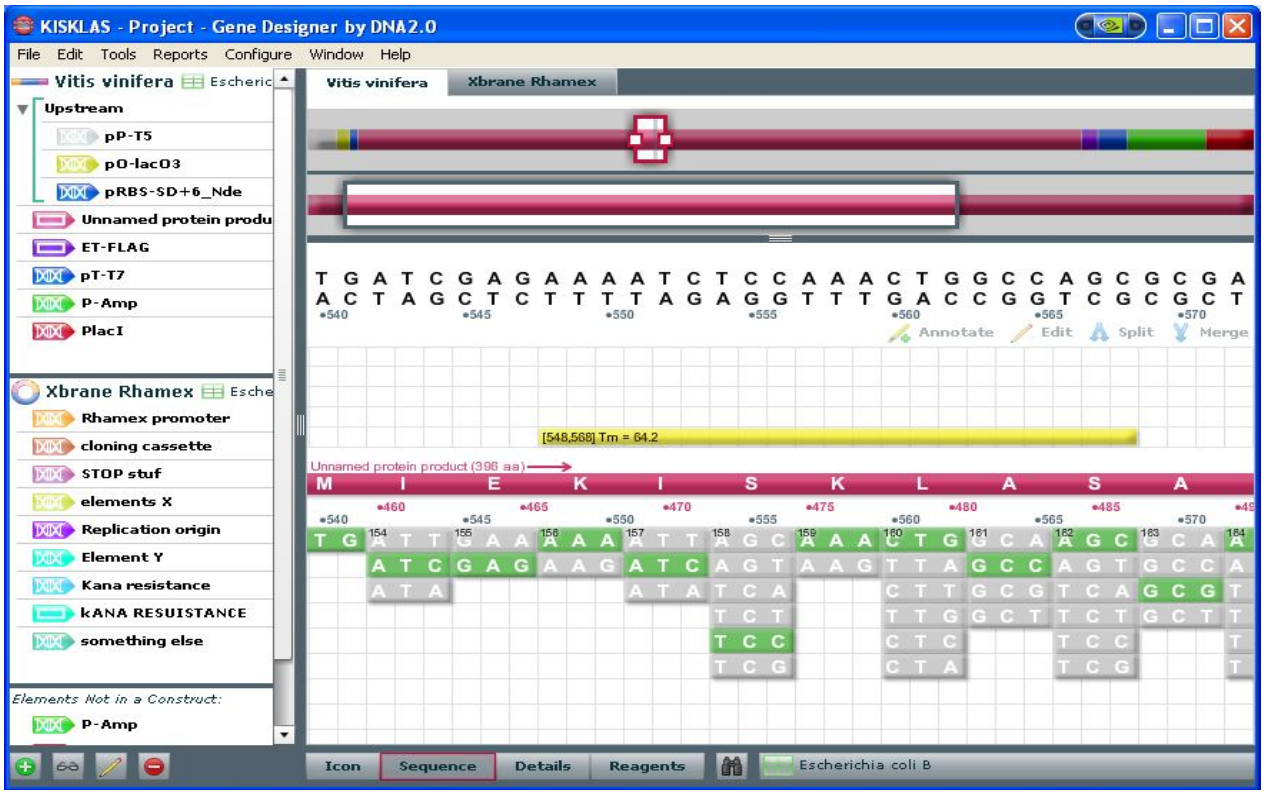

Fig. 1. Memes in the Carolome. The presence of memes located within genomic sequences is here exemplified by the KISKLAS meme within the *Vitis vinifera* genome. The DNA and protein sequence is visualized using Gene Designer 2.0 (Villalobos, Ness et al. 2006; Villalobos, Welch et al. 2011). Left side illustrates the presence of each genetic element as icons. Main window shows the DNA sequence on white background with the amino acid sequence (including meme) on red ribbon. Below each amino acid the corresponding codons are listed. The meme itself is noted using the real-time T_m calculator (T_m for the nucleotide sequence of the meme is 64.2 ºC). Gene Designer is available for free download at https://www.dna20.com/genedesigner2.

It is reasonable to hypothesize that the exact collapsed state of the genomic information was imprinted with the common cultural memes available at the spatial and temporal location where the genetic information was decoded and captured for posterity. Information condensation at sequencing laboratories that occurred during Christmas season inherently must have included Christmas carol memes. It should also be pointed out that several scientists (not only Tom Steitz at Yale) bear a striking resemblance to Father Christmas himself. Coincidental? We think not. DNA sequence is the ideal format to capture memes as the information can be coded directly into the chromosome and decoded through the subsequent reading of chromosomal DNA sequence. In our efforts to reconstitute one of the most central cultural memes of the season, we here display the functional imprints of Christmas carols. Consistent with the current naming convention of genome, proteome, transcriptome and other -omics based sciences, we define the full complement of cultural memes reflecting Christmas carols as the Carolome.

This publication represents the first unequivocal and rigorous scientific evidence linking a social cultural meme concept to a physical structure, deoxyribonucleic acid.

RESULTS AND DISCUSSION

Identification of top 15 Christmas carols. The 15 most beloved Christmas carols were identified by listening to mall music in four representative US malls (Cambridgeside Galleria, Kendall Square, Cambridge, MA. Stanford Shopping Center, Palo Alto, CA. South Park Mall, Charlotte, NC. and Westroads Mall, Omaha, NE). Each of the listed malls utilized a Santa Claus equipped with recording instruments to capture the background mall music (malls being one of the few places a Santa Claus can loiter without causing suspicion). The Santa Claus researcher utilized noise-canceling headphones (Panasonic RP HC500) and halothane vapor induced anesthesia to retain sanity during the extended recordings. No Santa Claus was physically harmed during these experiments.

Title	Meme
Ave Maria	AVMARIA
Baby, It's Cold Outside	CNTSTAY
Carol of the Bells	DINGDNG
Deck the Halls	FALALAA
Feliz Navidad	NAVIDAD
Frosty the Snowman	SNQWMAN
Holy Night	HQLYNHT
I saw Mommy Kissing Santa Claus	KISKLAS
Jingle Bell Rock	CLQKRQK
Messiah (Handel)	MESSIAH
Oh Dreidel, Dreidel, Dreidel	DREIDEL
Rudolph the Red Nosed Reindeer	REDNQSE
Silent Night	SHHNITE
The Little Drummer Boy	RAPAPQM
Twelve Days of Christmas	PEARTRE

Table 1. Complete alphabetical list of the 15 most played Christmas carols in 4 representative shopping malls and the corresponding 7 character memes (avoiding B,J,O,U,X,Z).

Meme identification. Given the enormous size of the existing Genbank database it was instrumental to use memes which length exceeded all spurious hits due solely to chance. The probability of a meme occurring by chance was calculated using the Assange principle (Assange 2010).

As negative control we used the most apparent widely played non-Christmas song available 'Hotel California' by The Eagles (Felder, Henley et al. 1977). The version of this song typically heard at shopping malls is instrumental and played by a string orchestra. We do not believe The Eagles contributed to this version. A five character meme of 'Hotel California' (HTLCA) retrieved 424,878 hits from GenBank. A six character meme of 'Hotel California' (HTLCAL) retrieved 2,433 hits from Genbank. No single instance of the seven characters 'Hotel California' meme HTLCALI or the eight characters HTLCALIF could be found throughout the entire Genbank database.

Accordingly we decided that a meme length of 7 characters was sufficient to explore memes two standard deviations away from what could be expected from the genomic background meme distribution. Furthermore, the signature memes were restricted to only encode one of the 20 letters represented in the genetic code for amino acids. Thus B,J,O,U,X,Z could not be used as characters in the meme.

The letter Q (for Glutamine) is derived from the Phoenician letter qoph and was at that time pronounced as a voiceless uvular plosive as determined from hieroglyphs of the period (Sternberg 1986). Voiceless uvular plosive (similar to how 'R' is avoided to be pronounced in Boston) is a common sound to Semitic languages (Unsofort, Tchetera et al. 1993). We propose that Q and O share a common ancestry based on historical and visual data. Furthermore, it is self-evident that any data derived from the Phoenicians is phony. Upon deriving the memes for the Christmas carols we thus conclude that Q can replace O within the textual framework that connects the bio-social implications of genomic information with that of society as a whole.

Each of the top 15 Christmas songs was evaluated by the Santa's International Helpers Union Local 1225 (as of November 1, 2010 renamed to 'Santa's Executive Program Specialist Assistants Union', Local 1225) for meme identification. The identity of each meme was established through socio-cultural analysis of the titles and content of the identified Christmas carols. Some of the memes, e.g. CNTSTAY for 'Baby, It's Cold Outside' (Louis Armstrong and Velma Middleton version) were immediate and obvious, whereas others, e.g. KISKLAS for 'I Saw Mommy Kissing Santa Claes' required further analysis of context and some endogenous craftsmanship to identify the inherent 7 character memes.

Genbank database. The searches were performed against GenBank Release 178.0 released June 11, 2010. This release includes ~450 GB of DNA sequence corresponding to 280x 10^9 base pairs distributed over 1.8x 10^8 entries.

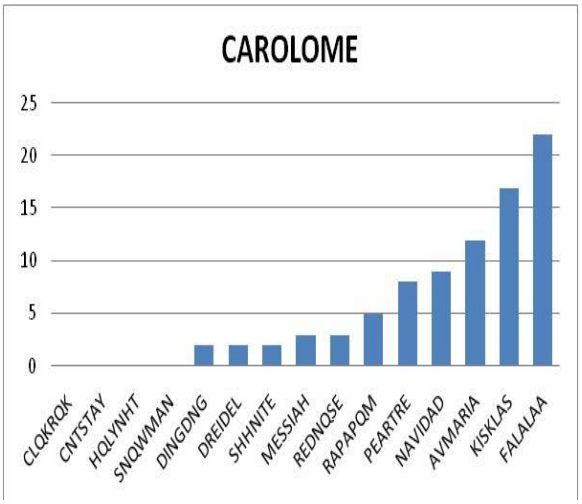

Fig. 2. Bar graph enumerating the presence of the 15 most common Christmas carol memes in the Carolome. Y axis denotes number of instances and X axis identifies the meme. Negative control, HTLCALI with 0 instances, is not present in the graph.

BLAST of memes against Genbank. Database BLASTs were performed using the BLASTP 2.2.24 (protein-protein BLAST) algorithm against the non-redundant protein sequences (nr) excluding environmental samples from WGS projects (Altschul, Madden et al. 1997). Only exact hits were retained and used in the assessment for presence in the global bio-genome.

Blast hit evaluation. The most common genome encoded meme entry of the Carolome is FALALAA corresponding to 'Deck The Halls' (*Deck the halls with boughs of holly, Fa la la la la, la la la la*). We found 21 unique instances of the FALALAA meme and attribute that to the linguistic neutrality of the text ensuring a global presence of the carol as well as the ancient origin of the song (sixteenth century Wales). The meme FALALAA is also easy to spell even for a unicellular prokaryote.

The mentioning of gay apparel in the second verse (*Don we now our gay apparel, Fa la la la la la la la la*) may reflect a similarly engrained meme referencing the recent discussion of the 'Don't ask, don't tell' policy of the US military (DeptofDefense 1993) .

The 21 hits of the FALALAA meme in GenBank include the presence in the gene encoding 23S rRNA methyltransferase of *Actinomyces vicious*.

This is apparently the most vicious of the Actinomycetes and the 23S methyltransferase gene itself most likely corresponds to an antibiotic resistance mechanism (Douthwaite, Voldborg et al. 1995).

The meme can also be found in a *Pseudomonas syringae* gene encoding a sugar efflux transporter. Considering the carbohydrate binge that occurs throughout western civilization during the lengthy holiday season (referred to as "Christmas Creep" and lasting from November 1— January 7) it is not surprising to find a meme associated with sugar transportation in the Carolome.

In support for the notion of replacing O with Q for iconizing the memes we note that the FALALAA meme is also found in the O-antigen polymerase of *Ammonifex degensii*. We conclude this to be unequivocal proof that Q replaces O in meme condensation.

Figure 3. The second most instances in the Carolome is the meme KISKLAS for 'I Saw Mommy Kissing Santa Claes'. The meme was identified in organisms ranging from roundworm (*Caenorhabditis remanei*) to dysentery-causing amoeba (*Entamoeba dispar*).

The second most common meme in the Carolome is KISKLAS corresponding to 'I Saw Mommy Kissing Santa Claes'. The University of NorthPole endorses The Ronettes version from 1963. This is the most dance-friendly version recorded (Palin 2010). KISKLAS is found in 17 genomes. It is interesting to note that the Catholic Church condemned the song for implying a link between sex and the religious holiday, and radio stations in several markets banned the song at its release in 1952 (Nelson 2009). The ban has since been lifted with a vengeance in shopping malls across USA.

In addition to the link between the KISKLAS meme and sex, it is noteworthy that it links to intoxicating agents as well since one of the identified memes in the Carolome for KISKLAS is found in the *Vitis vinifera* genome (commonly known as wine grape), the source for much of the alcohol consumed during Christmas. The encoded gene is unknown, or alternatively identification was inhibited due to libation of the genome host.

The favorite *Clostridium difficile* strain of the Sheriff of Nottingham encodes the KISKLAS meme within a transcription anti-terminator region. We have initiated a corollary project to see if the anti-terminator gene can be invalidated. The project is funded by the California Governor Arnold Schwarzenegger History Revision Foundation (Sturm and Drang 1973).

The KISKLAS meme is also found in the CPA2 transporter of *Prevotella bivia* (the renamed order of the oral plaque bacteria that was formerly called *Bacteroides melaninogenicus*). The University of North Pole research team is aware of at least one CPA with poor dental hygiene due to excessive consumption of Sugar Plums during the holiday period. Future work will focus on identifying the CPA in question.

The third most common meme in the Carolome is the theist meme AVMARIA corresponding to Ave Maria closely followed by NAVIDAD (Feliz Navidad), PEARTRE (Twelve Days of Christmas) and RAPAPQM (The Little Drummer Boy). For the complete list see Figure 2.

To our great surprise, neither CLQKRQK (Jingle Bell Rock), CNTSTAY (Baby, Its's Cold Outside), HQLYNHT (Holy Night) or SNQWMAN (Frosty The Snowman) could be located within the Carolome. This could be due to sequencing errors during the characterization of the global genome, or it could be that we misinterpreted the memes for those songs (less likely), or that the sequences are only available within Craig Venter's privately held DNA sequence databases (Davidsen, Beck et al. 2010).

We hereby urge colleagues across the field of Carolomics to join us in the quest for identifying and validating more and merrier entries.

Christmas Carol	**Meme**	**Protein**	**Organism**
Deck the Halls	FALALAA	Putative major facilitator superfamily (MFS) transporter	*Pseudomonas aeruginosa*
		Serine carboxypeptidase	*Glomerella graminicola*
		ABC transporter family protein 10	*Achromobacter xylosoxidans*
		Hypothetical protein STAUR_1180	*Stigmatella aurantiaca*
		Peptidase, M16 family protein	*Ketogulonicigenium vulgare*
		Hypothetical protein CRE_26972	*Caenorhabditis remanei*
		Periplasmic binding protein	*Oscillochloris trichoides*
		Tripartite ATP-independent periplasmic transporter DctQ	*Sinorhizobium meliloti*
		Enoyl-CoA hydratase/isomerase	*Sphingobium chlorophenolicum*
		23S rRNA methyluridine methyltransferase	*Actinomyces viscosus*
		Extracellular solute-binding protein	*Ahrensia sp. R2A130*
		Sugar efflux transporter	*Pseudomonas syringae*
		Hypothetical protein SCHCODRAFT_67855	*Schizophyllum commune*
		Conserved hypothetical protein	*Acidobacterium*
		Major facilitator superfamily protein	*Nitrosococcus watsoni*
		Hypothetical protein HacjB3_00160	*Halalkalicoccus jeotgali*
		Protein of unknown function DUF1800	*Acidobacterium*
		Diguanylate cyclase	*Alicycliphilus denitrificans*
		ABC transporter-related protein	*Haliangium ochraceum*
		O-antigen polymerase	*Ammonifex degensii*
		NRPS-type-I PKS fusion protein	*Streptomyces hygroscopicus*
		chromate transporter	*Thiobacillus intermedius*
I saw Mommy Kissing Santa Claus	KISKLAS	Argininosuccinate lyase	*Methanothermus fervidus*
		CRE-EPG-5 protein	*Caenorhabditis remanei*
		Recombination protein	*Mycoplasma hyorhinis HUB-1*
		Histidinol-phosphate aminotransferase	*Bacteriovorax marinus*
		UDP-N-acetylmuramoylalanine/D-glutamate ligase	*Arcobacter nitrofigilis*
		Transporter, CPA2 family	*Prevotella bivia*
		Unnamed protein product	*Vitis vinifera*
		Riboflavin synthase, alpha subunit	*Campylobacter gracilis*
		Hypothetical protein NECHADRAFT_84677	*Nectria haematococca*
		Putative transcription antiterminator	*Clostridium difficile*
		Hypothetical protein	*Populus trichocarpa*
		Recombination protein RecR	*Orientia tsutsugamushi*
		Hypothetical protein	*Entamoeba dispar*
		Hypothetical protein LELG_01367	*Lodderomyces elongisporus*
		Hypothetical protein MGG_07070	*Magnaporthe oryzae*
		DNA mismatch repair protein MutS	*Microscilla marina*
		Sensory box/GGDEF/EAL domain protein	*Congregibacter litoralis*
Ave Maria	AVMARIA	Acetyl-CoA acyltransferase	*Sus scrofa*
		CRE-CYP-13A8 protein	*Caenorhabditis remanei*
		Hypothetical protein AZL_022360	*Azospirillum sp*
		Hypothetical protein SSPB78_15942	*Streptomyces sp*
		OmpA/MotB domain protein	*Rhodobacter sphaeroides*
		Predicted protein	*Sclerotinia sclerotiorum*
		Predicted protein	*Botryotinia fuckelian*
		isoform CRA_a	*Rattus norvegicus*
		Hypothetical protein	*Monodelphis domestic*
		Flagellar MotB protein	*Rhodobacter sphaeroides*
		Aldehyde dehydrogenase	*Bordetella parapertussis*
		Protein-L-isoaspartate O-methyltransferase	*Paracoccus denitrificans*
Feliz Navidad	NAVIDAD	Possible nuclease	*Mobiluncus curtisii*
		2-isopropylmalate synthase	*Candidatus Puniceispirillum marinum*
		Homoserine dehydrogenase	*Haloterrigena turkmenica DSM*
		ATP-dependent dsDNA exonuclease SbcC	*Mobiluncus curtisii*
		Hypothetical protein Arad_2939	*Agrobacterium radiobacter*
		Hypothetical protein PROVALCAL_01930	*Providencia alcalifaciens*
		Serine protease 72	*Nasonia vitripennis*
		Hypothetical protein Mjls_4020	*Mycobacterium sp.*
		Hypothetical protein	*Gibberella zeae*
Twelve Days of Christmas	PEARTRE	Hypothetical protein VOLCADRAFT_86001	*Volvox carteri f. nagariensis*
		Protein of unknown function DUF945	*Desulfovibrio sp. FW1012B*
		Regulatory protein TetR	*Frankia sp. EuI1c*
		Choline dehydrogenase	*Streptomyces hygroscopicus*
		ThiE, Thiamine monophosphate synthase	*Rhizobium sp. NGR234*
		Thiamine monophosphate synthase/TENI subfamily	*Labrenzia alexandrii*
		Thiamine biosynthesis lipoprotein ApbE	*Roseibium sp. TrichSKD4*
		Thiamine-phosphate pyrophosphorylase	*Mesorhizobium sp. BNC1*
The Little Drummer Boy	RAPAPQM	Hypothetical protein	*Volvox carteri f. nagariensis*
		Glucose-methanol-choline oxidoreductase	*Burkholderia graminis*
		Glutathione S-transferase	*Hyphomonas neptunium*
		NAD-dependent epimerase/dehydratase	*Chloroflexus aurantiacus*

		Myocardin-like	*Danio rerio*
Rudolph the Red Nosed Reindeer	REDNQSE	Predicted protein Hypothetical protein Hypothetical protein	*Nematostella vectensis* *Paramecium tetraurelia* *Aspergillus oryzae*
Silent Night	SHHNITE	DEHA2D09130p Putative yir1 protein	*Debaryomyces hansenii* *Plasmodium yoelii yoelii*
Oh Dreidel, Dreidel, Dreidel	DREIDEL	Fructose-1,6-bisphosphatase class 3 Coiled-coil protein	*Bacillus atrophaeus 1942* *Giardia lamblia P15*
Carol of the Bells	DINGDNG	DNA topoisomerase IV subunit A	*Vibrio salmonicida*
Holy Night	HQLYNHT	None	
Frosty the Snowman	SNQWMAN	None	
Baby, It's Cold Outside	CNTSTAY	None	
Jingle Bell Rock	CLQKRQK	None	
NEGATIVE CTRL			
The Eagles – Hotel California	HTLCALI	None	

Table 2. The Global Carolome has been set up so that scientists involved in carolomics can use the data as a reference point for future studies and provide a common platform for testing and validating carolomics results.

ACKNOWLEDGEMENT

We thank Alan Villalobos for providing Gene Designer 2, Elaina 11Bosch for happily singing Jingle Bell Rock extraordinarily off-key, and the DNA IT-team for supporting the bioinformatics efforts.

Previous publications by the Santa Xmas Symposium:

Christmas Greeting 2008 about synthetic biology. https://www.dna20.com/files/PDF/Christmas2008.pdf

Christmas Greeting 2006 about the Hoho2 gene. https://www.dna20.com/files/PDF/Christmas2006.pdf

Christmas Greeting 2005. Print and fold for full viewing pleasure. Original included gene on filter. https://www.dna20.com/files/PDF/Christmas2005.pdf

Competing Interests: The authors declare competing financial interests: DNA2.0 performs gene design optimization and gene synthesis to a global customer base. Better, smarter and faster than anybody. And better looking too.

Funding: The project was financially supported by Santa's slush-fund Grant no. 0638333.

REFERENCES

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. **25**: 3389-3402.

Assange, J. (2010). "More is better than less...or at least more." Wikileaks **454**: 34487.

Davidsen, T., E. Beck, et al. (2010). "The comprehensive microbial resource." Nucleic Acids Res. **38**: D340-345.

DeptofDefense (1993). "Qualification Standards for Enlistment, Appointment and Induction." US Department of Defense Directive **1304**(26).

Douthwaite, S., B. Voldborg, et al. (1995). "Recognition determinants for proteins and antibiotics within 23S rRNA." Biochem Cell Biol. **73**: 1179-1185.

Felder, D., D. Henley, et al. (1977). "Hotel California." The Eagles **Hotel California**.

Gustafsson, C. (2009). "For anyone who ever said there's no such thing as a poetic gene." Nature **458**: 703.

Heinz, D. (2002). "Biological effects of infinite ketchup splatching." J. Food Cosmet. Ind. **72**: 42-62.

Kay and Mart (2001). "I'm gonna let ya' in on a little secret, Ray. K-Mart sucks." Hum. Genet. **32**(33): 33-44.

Nelson, V. J. (2009). "Jimmy Boyd dies at 70; singer of 'I Saw Mommy Kissing Santa Claus'." Los Angeles Times **March**(10).

Palin, B. (2010). "Dancing with the stars." ABC TV **11**(8): 32.

Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci USA. **74**(12): 5463-5467.

Sternberg, L. (1986). "Walk like an Egyptian." The Bangles **Different Light**(1): 4.

Sturm, U. and F. Drang (1973). Musikalische Katastrophe. Berlin, W. de Gruyter.

Unsofort, H., K. G. B. Tchetera, et al. (1993). "Shout and Yell." Yale J. Med. **9**(9-19).

Villalobos, A., J. E. Ness, et al. (2006). "Gene Designer: a synthetic biology tool for constructing artificial DNA segments." BMC Bioinformatics **7**: 285.

Villalobos, A., M. Welch, et al., Eds. (2011). In Silico Design of Functional DNA constructs. Methods Mol Biol. Clifton, NJ.

Wait, H. and C. See (1972). "Ballistic requirements set forth in the Big Bang." Nasa Rep. Astrophys. **F22** (1-472).

Merry Holidays and Happy New 2011