



PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, NORTHPOLE

Deconvolution of Inherent Genomic- Melodic Linkages, or the Music of Life.

Santa Claes^{1*}, Felix Navidad², Nisse Ness E.², Jeremy Elf Tootoo¹,
Dasher Sridhar^{1,2}, Tinsel Tian³ and Laura Menorah^{2,3}

¹University of NorthPole, Dept Carolology, Arctic Lane 1, Santa's Secret Village, NorthPole

²University of NorthPole, School of Music, Arctic Lane 25, Santa's Secret Village, NorthPole

³University of NorthPole, Dept Endocrinology, Candy Cane Avenue 365, Santa's Secret Village, NorthPole

*Corresponding author: santa@DNA20.com

Received 28 June 2012, received in revised form 25 October 2012, accepted 24 December 2012

Accompanying music for this article can be found at: www.youtube.com/watch?v=snmZzGG0Ur4

Santa Claes and his favorite elves, under the stern guidance of disciplinarian Laura Menorah, have spent the last decade pursuing the musical legacy concealed in our genetic heritage. Their quest for infinite musical wisdom was originally pioneered by Prof Susumo Ohno from the City of Hope National Medical Center in whose footsteps (size 11) few elves can fit. Already in 1986, Dr. Ohno famously played the human X-linked phosphoglycerate kinase on his upright Baldwin¹ ('vaguely Satie-like' according to the LA Times). Thanks in particular to the endurance and stamina of elf Nisse-Ness, Santa is now proud to disclose the final deconvolution of the inherent musical annotation transcending our genomes. Like delicate threads of cultural beams shining through the complexities of ancestral obligations, music provides the beacon of excitation that illuminates our path towards the global manifestation of bioengineering. "If music be the food of life, play on."² Santa and his elves – all DNA2.0 employees during their off-season – have here systematized all possible variables describing the music of celestial spheres.³ Thanks to Saturn cloud computing contributed by the automatic desk,⁴ all possible enumerations of sequence-functional relationships have been bootstrapped for the evaluation of descriptor relationships.

HYPOTHESIS

In a groundbreaking effort to reach a complete understanding of melodic-genomic interaction, Santa's elves identified a total of 87 descriptors characterizing physical/cultural/psychological properties of the basic 60 musical chords (major chords, minor chords, 7th chords, minor 7th chord and major 7th chord). The elves decided against analyzing all possible chords (assuming basic music theory, given 3 or more notes per chord, 12 notes in an octave and 4 octaves on a guitar there are $\sim 450 \times 10^6$ chords. "Too many, let's round that to the nearest 60" said elf Jeremy). Descriptors included: frequency spectrum (total of 15 physical properties), global temporal and spatial musical distribution (total of 47 cultural properties) and a total of 25 psychological properties (e.g. presence of certain chords in sad love songs from the 70's). Each of the 60 basic chords were characterized for each of the 87 descriptors and the resulting values were mean centered and normalized resulting in a 60x87 matrix with the distribution within each property ranging from 0-1 (North Pole Supplemental material).

As in most multidimensional datasets there is a high degree of correlation encapsulated in such a matrix. The 87 dimensions can be condensed using Principal Component Analysis (PCA) so that Principal Component 1 (PC1) captures >50% of all information and PC2 close to 20% of all described information. Representative subsets of the 60 basic chords are distributed in accordance to their properties and denoted in red in Figure 1. PCA plots like

these are useful for clustering samples according to the combination of a large number of (often correlating) properties. As an example, chords Am6 and Am7 are co-located in the mid-left of the PCA plot due to similarity in harmony, frequency spectrum and abundance in similar musical pieces (including My Funny Valentine by Tony Bennett). Similarly, the Gmaj inversion chords in the upper left (Gmaj7 i1, -i2, -i3) all share a similar cadenza and modulation with a hint of timbre and often occur in 1967 'Summer of Love' music from San Francisco including Lovin Spoonful and The Electric Prunes (I Had Too Much to Dream Last Night).

The elves, with some help from Herme the Dentist, also performed a detailed study where all 20 natural amino acids were characterized by 26 physicochemical descriptor variables. These descriptor variables include experimentally determined retention values in seven thin-layer chromatography (TLC) systems, three nuclear magnetic resonance (NMR) shift variables, and 16 calculated variables, namely six semi-empirical molecular orbital indices, total, polar, and nonpolar surface area, van der Waals volume of the side chain, log P, molecular weight, and four indicator variables describing hydrogen bond donor and acceptor properties, and side chain charge.⁵ All values were assessed in reverse sequence while listening to "The 12 Days of Christmas" (Bing Crosby version preferred, Chipmunks allowed on alternate Wednesdays) with the express intent of aligning numerical song values with physiochemical numerations (e.g. 'four calling birds' occurring during analysis of four hydrogen bonds).

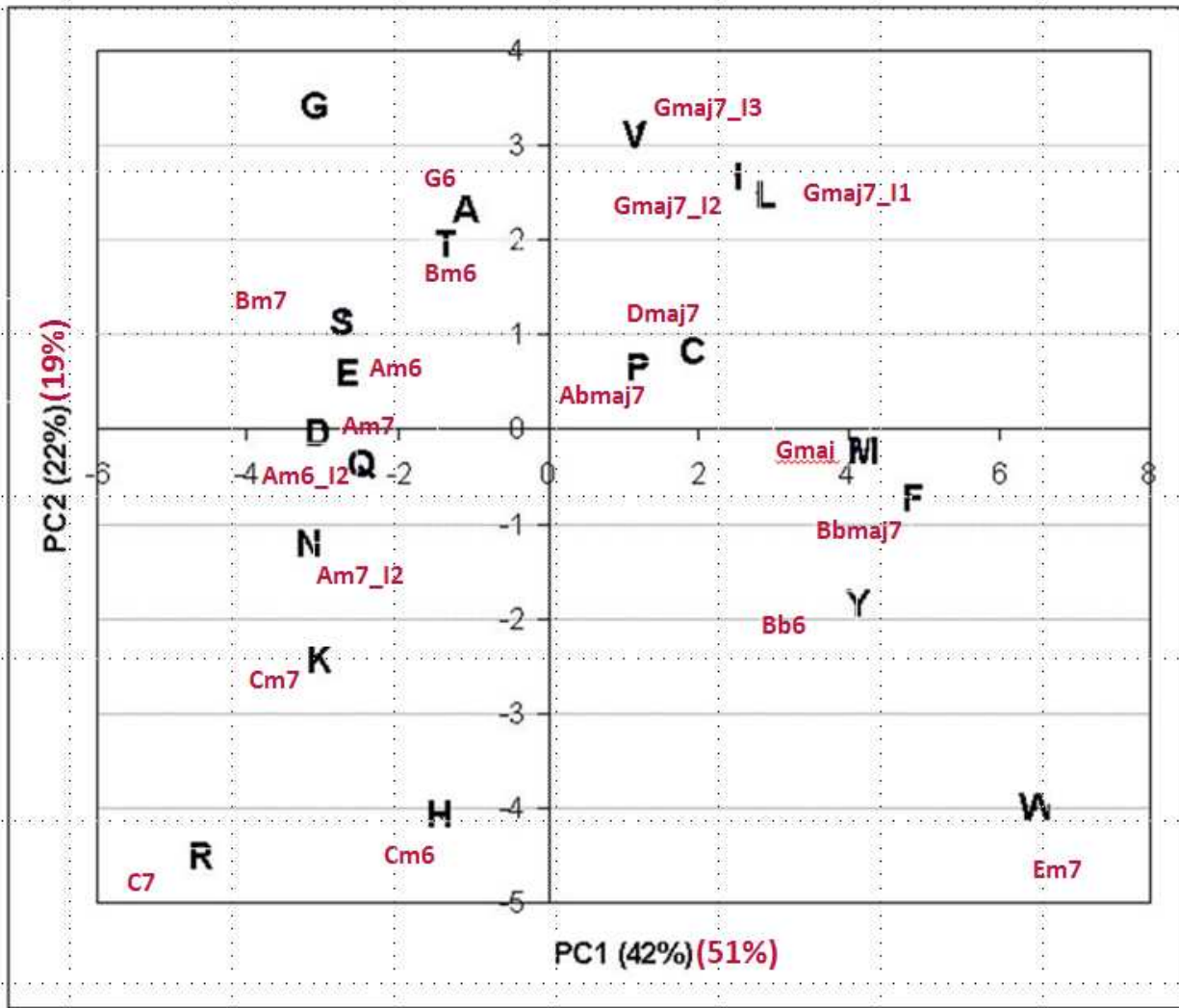


Figure 1. PCA plot of the 20 natural amino acids (in black) described in 26 dimensions and compressed into the two main components PC1 (42% of all information) and PC2 (22% of all information). Overlaid on the amino acids are a subset of the basic 60 chords (in red) described in 87 dimensions and compressed into the two main components PC1 (51% of all all information) and PC2 (19% of all information).

The quantitative relative distribution of amino acid physico-chemical properties is seen in black in Figure 1. Overlaying the musical PCA plot (red) with the amino acid PCA plot (black) reveals a high degree of correlation between certain chords and amino acids. With the striking exception of Glycine, every amino acid is nicely paired with a chord. For example,

Methionine, the starting amino acid of every protein, correlates in this compressed megadimensional space with Gmaj, a common chord for opening a chorus (e.g. 'Let it be' by the Beatles and 'Romeo and the Lonely Girl' by Thin Lizzy). Valine, Leucine and Isoleucine are very similar aliphatic amino acids all aligning well with the previously mentioned inverted

Gmaj chords. Tryptophan with its large indole group is demoted to the bottom right corner of the PCA plot far away from any other amino acid. This is not surprising due to the extreme properties of the amino acid side chain. Similarly the Em7 chord is often regarded as the evil step child of popular music. With the exception of Taylor Swift ('All Night Diner' and 'Stay Stay Stay'), Em7 is perceived as perhaps the most pyrrolic of chords. Not surprisingly, Em7 co-locates to the bottom right corner of the PCA plot with Tryptophan.

Extensive mapping of physico-chemical properties of amino acids overlaid on the physical/cultural /psychological properties of musical chords resulted in a surprisingly consistent pairwise correlation between amino acids and chords. The probabilistic independence of the correlation was evaluated using the Pearson Product-Moment Correlation Test⁶ and Fisher's Combined Probability Test generating P-values lower than 0.02 for the correlation occurring by chance. This data strongly substantiates the underlying finding dictating Biological-Musical correlation.

The single exception is Glycine, which does not pair with any of the standard chords. Glycine is also the only amino acid that lacks a side chain. In accordance with Shannon's information theory – void of information (i.e. side chain) is also an information descriptor on equal standing with all other information descriptors.⁷ Thus, the amino acid glycine clearly represents a musical rest.

The complete descriptor mapping of virtual essence (music) with its physical embodiment of life (amino acids) enables an exhaustive understanding of ALL information. Jimi Hendrix's Purple Haze or Victor Rydberg's Tomten⁸ can be coded into a protein and

expressed and purified for functional evaluation. Similarly the tRNA(m5U54) methyltransferase can be played and appreciated by a forgiving audience comprised primarily of tone deaf Swedish Biochemists.⁹

Amino acid	Chord
A	G6
C	Dmaj7
D	Am7
E	Am6
F	Bbmaj7
G	-
H	Cm6
I	Gmaj7_I2
K	Cm7
L	Gmaj7_I1
M	Gmaj
N	Am7_I2
P	Abmaj7
Q	Am6_I2
R	C7
S	Bm7
T	Bm6
V	Gmaj7_I3
W	Em7
Y	Bb6

Table 1. Pairwise correlation of musical chords and amino acids. Using this relationship for full instrumental transphonification leads to a complete set of arias, madrigals, sonatas, concertos, symphonies, hymns, cantatae, ballads, lullabies, waltzes and carols derived from genomic information.

VALIDATION

As in all other experimental sciences, any theory (even if the P-value of the null hypothesis is approaching zero) must be appropriately validated before it can be

accepted. Extraordinary claims require extraordinary proof, and here the proof is in the pudding.

Santa and his elves had far too much of the Christmas pudding, joined hands and whispered, "Jábba sô kalík qonay" (translates to "Santa needs insulin" in Elvish). Due to the massive amounts of fruitcake, candy canes, maple fudge and chocolate Santas (Figure 2) consumed during the holiday season, insulin production within the body is increased in a constant proportion to remove excess glucose from the blood. The insulin gene (NP_000198.1) was thus re-coded according to Table 2 resulting in the audio-visual recording found in Figure 3 and the sheet music found in Figure 4. Not surprising, given Santa's strong relationship to Insulin and genomics, the recoding resulted in a musical entity strongly related to The Carol of the Bells.¹⁰



Figure 2. Foil wrapped Chocolate Santas trio, "a holiday dream in every bite,"¹¹ leads to increased blood glucose levels and a subsequent rise in Insulin production.

Table 2. Amino acid sequence of the insulin Gene (NP_000198.1¹²):

**MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGGVELG
GGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLE
NYCN**

Figure 3. Dr. Mark Welch has created an audio-visual guitar recoding of the insulin gene:
www.youtube.com/watch?v=snmZzGG0Ur4



Figure 4. Musical recoding of Insulin using pairwise amino acid-musical chord distribution (see Table 1). Available for your listening pleasure at:

www.youtube.com/watch?v=snmZzGG0Ur4

ACKNOWLEDGEMENT: We thank the entire DNA2.0 team for reviewing audio recordings and consuming holiday treats.

FUNDING: This project was financially supported by Santa's slush-fund Grant no. 04238794.

REFERENCES:

- ¹ Baldwin Piano & Organ, Herrin, IL, www.baldwinofherrin.com
- ² Shakespeare, W. *Twelfth Night*, 1.1.1
- ³ Ibid, "The music of the spheres!", *Pericles*, 5.1.289
- ⁴ Carlos Frio, Bilbo Baggings and The Other Planet with a Ring. *Journal of Revolving Rumors*, 2009, 45:6-88.
- ⁵ U.R.Wright & I.M Ron, Identification and Characterization of Breakfast Chemicals BaNa₂ and CoFe₂. *Journal of Superfluous Redundancy*. 2005, 83:478-99.
- ⁶ Will Sliceum, 'Knife party' MD thesis, Dept Surgery, UC Wattudon, 1987.
- ⁷ Shannon, C.E. (1948), "A Mathematical Theory of Communication", *Bell System Technical Journal*, 27, pp. 379–423 & 623–656.
- ⁸ DNA2.0 Xmas Greeting 2005.
- ⁹ Particularly those whose Ph.D. theses involved said protein.
- ¹⁰ See also Claes *et al.*, PNAS North Pole (2010) 12:25-31. Carolome: Functional Imprints of Culture Memes in the Global Genome.
- ¹¹ Chocolate Santa vendor: See's Candies, www.sees.com/prod.cfm/Milk_Chocolates/Chocolate_Santa_Trio.
- ¹² Insulin preproteins [Homo sapiens], www.ncbi.nlm.nih.gov/protein/NP_000198.1.

**Merry Holidays
and
Happy New Year
2013**