The Musical Geometry of Genes

Generating Rhythms from DNA

ALVARO YANEZ

ABSTRACT

DNA encodes all sorts of information that makes us human, but, aside from encoding genes, could DNA also encode for a mapping of musical rhythms in a very abstract way? This project sought to generate rhythms out of DNA and compose a musical piece out of a gene's rhythmic sequence. Computational rules inspired by geometric analyses of rhythms guided the mapping of DNA's molecular structure into rhythmic timelines and melodic scales; these basic structures were then used to compose a song according to the sickle cell gene DNA sequence. The rhythms generated by this "genetic analysis" alternate pleasantly between even and odd time signatures.

GEOMETRIC RHYTHMS AND DNA

The discovery of deoxyribonucleic acids (DNA) and their role in encoding the information that makes us human has inspired scholars from numerous corners of academia. In musicology, DNA's unique geometric properties have typically been explored with sonification studies that primarily generate melodies out of gene sequences [1–5]. To my knowledge, however, no academic studies have attempted to generate geometric rhythms out of DNA's geometric properties, as I could not find other published descriptions of DNA-derived rhythms in literature. This is surprising given the almost natural opportunity to relate rhythms and DNA by looking at their geometry. Rhythms can be ascribed geometric properties like shapes in mathematics if understood as timelines of k sounded onsets and n total pulses that repeat over time in a circle diagram, also known as polygon notation (see Fig. 1 for an example) [6]. DNA is famous for its attractive and highly conserved molecular geometry; seeing such an opportunity, I therefore used geometry as an abstract means to generate rhythms out of DNA molecules, and I used these rhythms to compose a whole musical piece out of a DNA sequence: the sickle cell gene [7].

Alvaro Yanez (student), New York University Abu Dhabi, PO Box 129188, Abu Dhabi, U.A.E. Email: ayanez@nyu.edu. ORCID: 0000-0001-9309-1885.

See www.mitpressjournals.org/toc/lmj/29 for supplemental files associated with this issue.

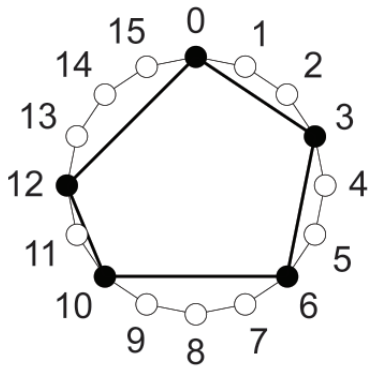

Fig. 1. The *clave son* in polygon notation ($k = 5$, $n = 16$). Geometric properties arise from the timeline (a symmetric pentagon; thick contour) and the time signature (4/4, a hexadecagon; thin contour). (© Godfried Toussaint [19])

Turning DNA Molecules into Rhythms

DNA is a long molecule made out of four basic subunits [8]. These precursors are adenosine monophosphate (AMP), thymidine monophosphate (TMP), guanosine monophosphate (GMP) and cytidine monophosphate (CMP; Fig. 2) [9]. These four basic units bond with each other in specific orders to form long sequences called genes. Molecularly, AMP, TMP, GMP and CMP have regular structures. These consist of three chemical groups: one central ribose sugar (central highlight in each panel of Fig. 2), one phosphate group (leftmost highlight in each panel of Fig. 2) and one nitrogenous base (rightmost highlight in each panel of Fig. 2). In AMP, the nitrogenous base is adenine (A); in TMP, thymine (T); in GMP, guanine (G); and in CMP, cytosine (C). The important property to note here is neither the names nor the chemistry of these subunits but rather that two out of these three chemical groups are cyclic rings—namely, the central ribose sugar and the nitrogenous base. Given that rhythms can be represented using polygon notation (i.e. essentially, a circle with enclosed polygons; see Fig. 1), the polygons formed by the central ribose sugar and the nitrogenous bases were a

 doi:10.1162/LMJ_a_01052

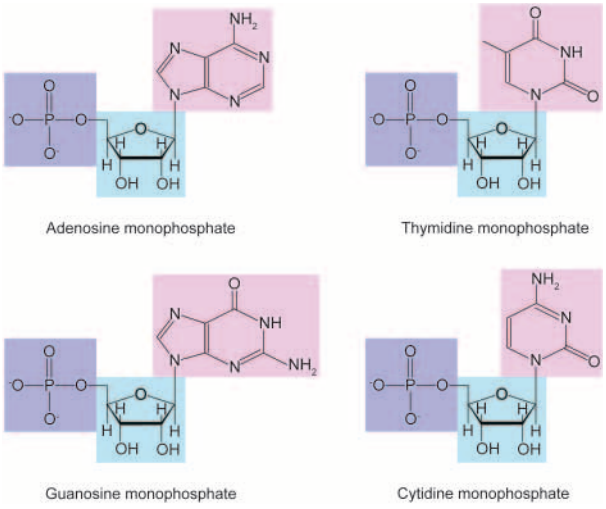

Fig. 2. Chemical structures of the four basic subunits of DNA. Central, leftmost and rightmost highlights in each panel indicate the central ribose sugar, the phosphate group and the nitrogenous base of each subunit (A in AMP, T in TMP, G in GMP and C in CMP). (© Alvaro Yanez)

natural source of inspiration for the mapping of DNA molecules into rhythms.

I created four rules to map these rhythms. These rules came about by noting that (a) chemical characteristics within the cyclic rings could be associated to (b) rhythmic outputs into a circle diagram. Namely, the size of the ring, the atoms within it and the functional groups attached to it could be associated with a circle diagram's *n* total pulses and *k* sounded onsets. For illustration purposes, all pulses in this work were written as quarter notes; however, this was a stylistic decision that could be changed. The four rules (summarized in Table 1) were:

> **Rule 1:** The number of cyclic rings in the DNA subunit will determine the number of circle diagrams to be drawn. Application: Because AMP and GMP have three cyclic rings, they yield three circle diagrams (TMP and CMP yield two).
>
> **Rule 2:** The number of atoms in each cyclic ring will determine *n*, the number of total pulses of its respective circle diagram. Application: Because AMP has a five-membered ribose ring, a six-membered first ring in the base and a five-membered second ring in the base, AMP's three circle diagrams have $n = 5$, $n = 6$ and $n = 5$ (respectively).
>
> **Rule 3:** Noncarbon atoms *within* each cyclic ring will determine the bass drum *k* polygon, i.e. the timeline of a drum kit's bass drum. Application: Because AMP's ribose ring only has one noncarbon atom (i.e. the oxygen atom, O [10]), AMP's ribose ring's circle diagram has a bass drum timeline of [x . . . .]. Similarly, AMP's base's diagrams have bass drum timelines of [. x . x . . .] and [x . x . .] [11].
>
> **Rule 4:** Functional groups *attached to* each cyclic ring will determine the snare drum *k* polygon, i.e. the timeline of a drum kit's snare drum. Application: Because AMP's ribose ring has two –OH groups attached, AMP's ribose ring's circle diagram has a snare drum timeline of [. . x x .]. Similarly, AMP's base's diagrams have snare drum timelines of [x . . . . .] and [. . . . .].

The four rules generated 10 rhythmic timelines from 10 cyclic rings in all four DNA subunits; these are shown in Fig. 3. To have one timeline per DNA subunit, I combined each subunit's timelines into one timeline [12]. This yielded four rhythms of unusual time signatures (Fig. 4a; online supplemental video S1): AMP's and GMP's timelines are in 15/4 (or 5/4, 6/4 and 5/4), and TMP's and CMP's are in 11/4 (or 5/4 and 6/4). These time signatures were a consequence of AMP and GMP both having a ribose and a double-membered base ring and of TMP and CMP both having a ribose and a single-membered base ring. The timelines generated are not *too* different from one another: AMP, TMP and GMP progress similarly, but AMP differs from TMP and GMP in a silent eighth onset, while TMP differs from GMP due to a shorter *n*. CMP is actually equal to TMP, which is a consequence of TMP and CMP having equal distributions of carbon and noncarbon atoms and functional groups in the base.

The rhythmic timeline of the first four letters of the sickle cell gene are illustrated in Fig. 4b. The time signature oddity reminded me of Dave Brubeck's *Take Five* [13], written in the unusual 5/4 (i.e. $k = 5$); the oddity in *Take Five*'s time signature has been described to both challenge the listener's

TABLE 1. Summary of the geometric rules for the rhythmic timelines.

Rule	Chemical Input (a)*	=	Rhythmic Output (b)*
Rule 1	*N*(rings in the monomer)	=	*N*(circle diagrams)
Rule 2	*N*(atoms in the ring)	=	n
Rule 3	Noncarbon atoms in the ring	=	Bass drum onsets (k_{bass})
Rule 4	Functional groups in the ring	=	Snare drum onsets (k_{snare})

*$N(X)$ = number of X

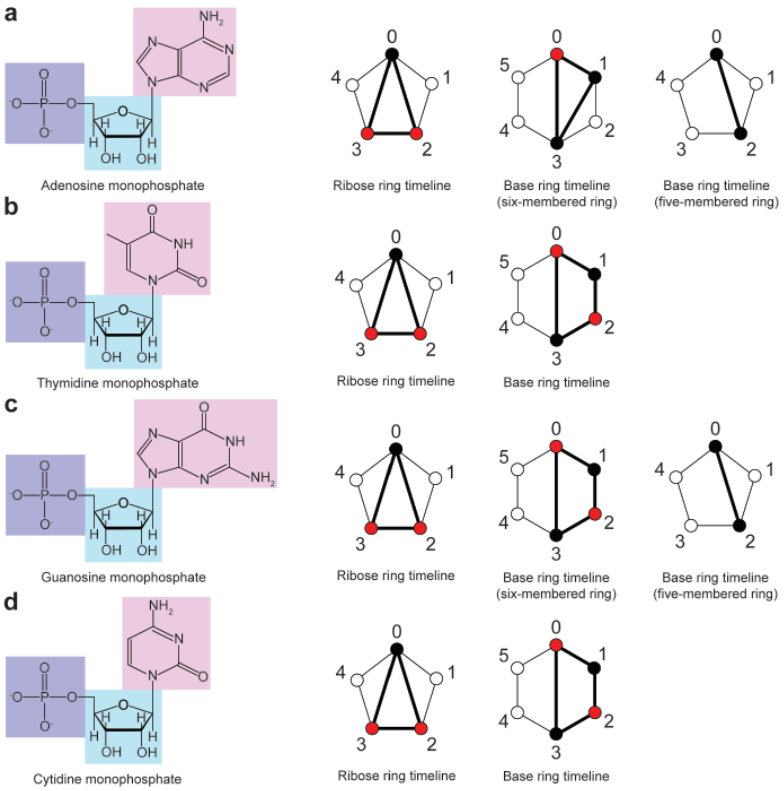

Fig. 3. Chemical structures and circle diagrams of the resulting rhythmic timelines. Chemical structures (left) and resulting rhythmic timelines (right) of AMP (a), TMP (b), GMP (c) and CMP (d). (k_{bass} indicated in black shading; k_{snare} indicated in red shading [gray in print edition].) (© Alvaro Yanez)

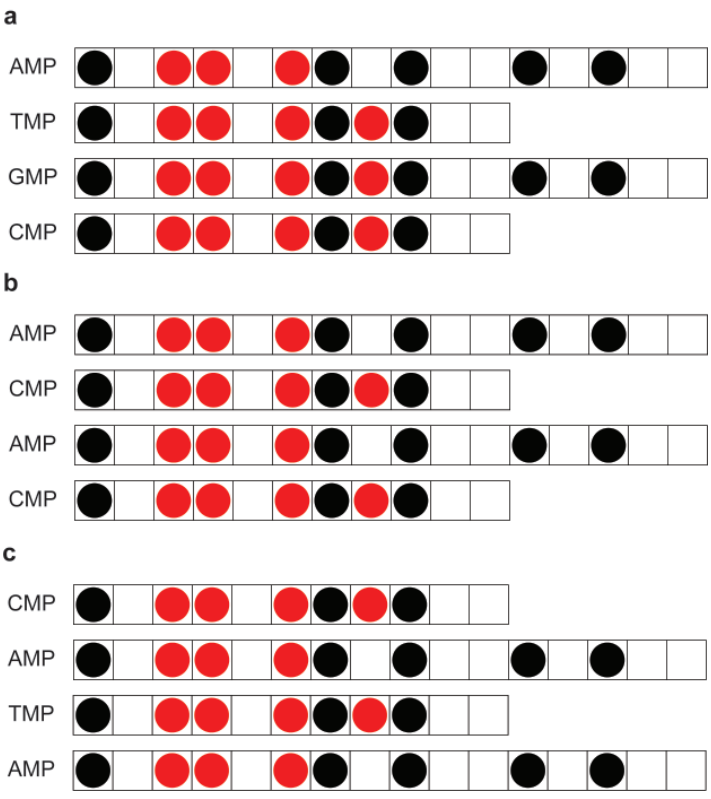

Fig. 4. (a) Full rhythmic timeline of each DNA subunit. Time signature for each full rhythmic timeline indicated by box length (k_{bass} indicated in black shading; k_{snare} indicated in red shading [gray in print edition]). (b) Rhythmic timeline of the first four letters of the sickle cell gene: ACAC [20]. (c) Rhythmic timeline of the first four letters of the MAOA gene: CATA [21]. (© Alvaro Yanez)

ability to keep track of the beat and "keep the brain more active while listening and performing" [14]. These "genetic" rhythms are not only oddly timed, however: They also alternate with even time signatures. This reminds me of the work of Dream Theater, a progressive rock band famous for virtuosic alternation between time signatures. In *The Count of Tuscany* [15], for instance, the time signature starts in 3/4 (0:00–1:02) and begins alternating in 1:03–2:36 with a repeated sequence of 6/8, 9/8, 6/8 and 12/8 to form the core of the piece. A second instance of this unusual alternation can be found 9:33–10:12, where the piece follows an ABAB structure in which A = 7/8, 15/16, 7/8 and 4/4 and B = 7/8, 15/16, 7/8 and 19/16. Here, the rhythms generated by DNA resemble Dream Theater's style more than that of Dave Brubeck because we obtained oddly timed rhythms whose time signatures alternate. These mixed time signatures could also be written in an arbitrary number of different ways; nonetheless, regardless of how you choose to write them, the underlying rhythmic timelines will never be regular when compared to each other. An additional illustration of how DNA sequences can be mapped into rhythmic timelines using this method is shown in Fig. 4c, where the "rhythm" of the first four letters of the MAOA gene are depicted.

TURNING DNA MOLECULES INTO MELODIC SCALES

Having generated four rhythms from DNA, I then generated melodic scales that would help me constrain the compositional process to a narrower range of notes from which to pick. I took advantage of the geometric isomorphism between rhythms and scales described by Toussaint in *The Geometry of Musical Rhythm* [16]. In short, this isomorphism converts rhythms into melodic scales (and vice versa) by using a rhythm's circle diagram polygon. If we overlay a timeline's polygon into an $n = 12$ circle diagram, where each "onset" represents a note of the chromatic scale (with C as the zeroth onset), then the k timeline resulting from this overlay indicates which notes should be played. For example, AMP's ribose timeline (Fig. 3a; a triangle with corners at onsets 0, 2 and 3) transposes into the scale C-F-G (Fig. 5a; a triangle with corners at onsets 0, 5 and 7).

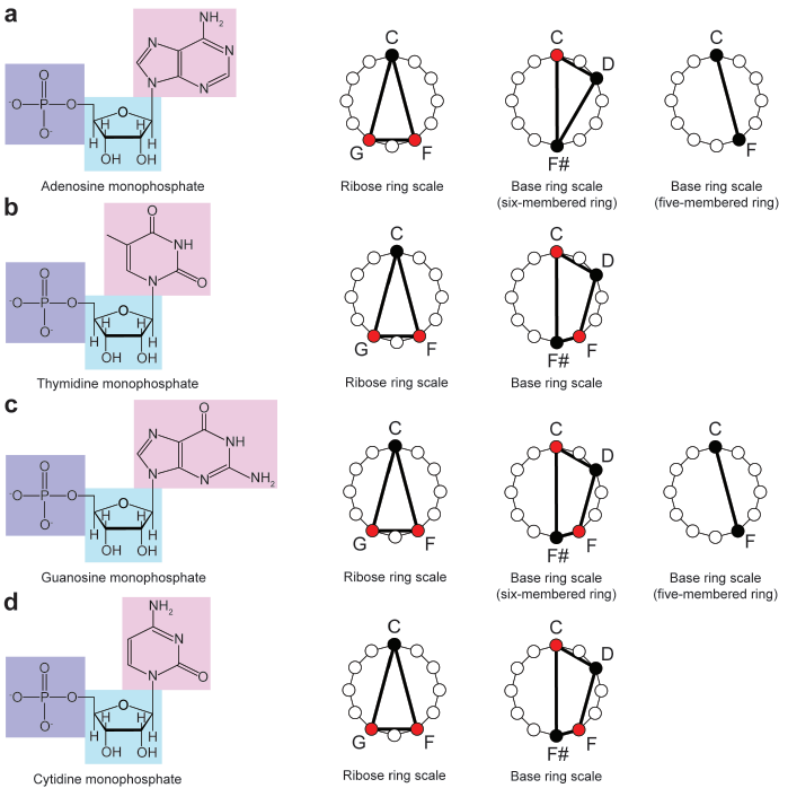

Fig. 5. Chemical structures and circle diagrams of the resulting melodic scales. Chemical structures (left) and resulting melodic scales (right) of AMP (a), TMP (b), GMP (c) and CMP (d). (Sounded onsets indicated in black and red shading [gray in print edition]). (© Alvaro Yanez)

Using this methodology, I transposed each of the rhythmic timelines generated in Fig. 3 into their respective melodic scales (Fig. 5; online supplemental video S2). (When a polygon's corner landed between onsets rather than on an onset, the nearest clockwise onset was chosen to be sounded.) Having generated both rhythmic timelines and melodic scales out of DNA, I then proceeded to compose a piece out of a well-studied gene sequence: the sickle cell gene.

COMPOSING A SONG OUT OF A GENE

The composition of a musical piece out of the sickle cell gene followed a similar methodology to that of eighteenth- and nineteenth-century artists playing composing games. These games challenged artists to compose waltzes on the spot using two dice. The number sequence drawn by the dice decided a priori the meter, scale and chords that the artist should use. Artists would draw on a mental "library" that associated (also a priori) every dice combination to precomposed chord progressions in the meter and scale indicated [17].

Inspired by the concept of using dice's randomness to compose a piece's "backbone," I used the first 52 letters of the sickle cell gene sequence to accomplish the same purpose. These letters are ACACTCGCTT CTGGAACGTC TGAGGTTATC AATAAGCTCC TAGTCCAGAC GC [18], where A stands for AMP, C for CMP, T for TMP and G for GMP. I arrived at the number 52 by experimenting with several values first. I thought 52 letters would be significant enough to compose a two-minute long piece with verses, choruses, a bridge and a solo; using all 586 letters in the sequence, on the other hand, would have given a very long piece.

The song's rhythmic "backbone" was generated by substituting each letter with its respective timeline (according to the "rhythmic library" in Fig. 4). The accompanying chord progression was generated by trying out progressions from each timeline's transposed scale (i.e. the "melodic library" in Fig. 5) and by modifying these chords at will to obtain a pleasurable progression, as the scales generated did not inherently sound pleasing. This method gave me both a rhythm and a chord progression from which I could start improvising creatively, like artists playing composing games. Adding a creative component to the composition was important to me because I wanted the final composition to be perceived as a fluid song, rather than the rigid outcome of a geometric "code."

To facilitate the creative process, I subdivided the 52-lettered sequence into smaller sequences that I could compose piece-wise. I decided that these smaller structures would be multiples of four, based on *The Count of Tuscany*'s alternating four-tuple main theme (1:03–2:36) and ABAB structure (9:33–10:12). Subdividing the sequence eased the compositional process because the piece's time signature sequence was as random as the gene sequence itself. Giving me a localized structural context to work with facilitated the creative aspect of this exercise.

The final product of this project, a two-minute-long piece that I titled *E8Q*, can be found in the online supplemental files (Audio S1). Like the gene sequence's letters, the rhythm of the piece does not repeat itself, but this lack of repetition does not seem to sacrifice musicality. Astoundingly, the song's rhythm holds a somewhat regular and periodic structure. This is entirely due to the choice of DNA as a source of musical inspiration. Because DNA consists of a sequential combination of four subunits from which two are bigger than the other two (i.e. AMP and GMP versus TMP and CMP), the resulting rhythmic sequence (i.e. the two different 15/4 timelines and the two equal 11/4 timelines) produce a structure that sounds odd and unusual, but not *too* unusual when permuted. Consequently, the chemical similarity between DNA subunits, and its ability to combine into a sequence, allows DNA to produce unusual rhythms that are musically attractive despite being irregular.

CONCLUSION

A deep analysis of geometry in the context of rhythms and biochemistry enabled me to map DNA sequences to rhythms and scales. Using rules inspired in DNA's cyclical rings, I

generated geometric polygons that could be used both as rhythmic timelines and melodic scales. These timelines and scales, when ordered according to the gene sequence of the sickle cell gene, created a "backbone" from which I composed a two-minute song that alternated randomly but pleasantly between two unusual time signatures: 11/4 and 15/4. I know of no such prior compositional exercise in musicology. Therefore, the main lesson of this project—that chemical structures can be related to rhythmic structures to create a musical "backbone" over which one can improvise and compose—has the potential to serve as an original and abstract compositional technique for artists seeking to relate biology and music even more intimately.

Acknowledgment

I deeply thank Godfried Toussaint (New York University Abu Dhabi) for his active support in the development of this project.

References and Notes

1 R. Takahashi and J.H. Miller, "Conversion of Amino-Acid Sequence in Proteins to Classical Music: Search for Auditory Patterns," *Genome Biology* **8**, No. 5 (2007) p. 405.

2 S. Ohno and M. Ohno, "The All Pervasive Principle of Repetitious Recurrence Governs Not Only Coding Sequence Construction But Also Human Endeavor in Musical Composition," *Immunogenetics* **24**, No. 2, 71–78 (1986).

3 J. Dunn and M.A. Clark, "Life Music: The Sonification of Proteins," *Leonardo* **32**, No. 1, 23–32 (1999).

4 S. Alexjander and D. Deamer, "The Infrared Frequencies of DNA Bases: Science and Art," *IEEE Engineering in Medicine and Biology Magazine* **18**, No. 2, 74–79 (1999).

5 X.J. Shi, Y.Y. Cai and C.W. Chan, "Electronic Music for Bio-Molecules Using Short Music Phrases," *Leonardo* **40**, No. 2, 137–141 (2007).

6 Godfried Toussaint, *The Geometry of Musical Rhythm: What Makes a "Good" Rhythm Good?* (New York: CRC Press, 2013) pp. 13–14.

7 The sickle cell gene is a gene that encodes for a protein that is important for the proper function of red blood cells. The gene's technical name is *HBG2*, and mutations in *HBG2* cause sickle cell anemia. Sickle cell anemia is a condition in which red blood cells no longer look smoothly biconcave but like a sharp sickle and hence cannot transport oxygen around the body as effectively.

8 J.D. Watson and F.H.C. Crick, "Molecular Structure of Nucleic Acids," *Nature* **171** (1953) pp. 737–738.

9 S.J. Elledge, Z. Zhou and J.B. Allen, "Ribonucleotide Reductase: Regulation, Regulation, Regulation," *Trends in Biochemical Sciences* **17**, No. 3, 110–123 (1992).

10 In the chemical structures given in Figs 2 and 3, carbon atoms are represented by corners and noncarbon atoms are represented by their atomic symbols (as indicated in the Periodic Table of Elements). H: hydrogen; O: oxygen; N: nitrogen; and P: phosphorus.

11 Both carbon and hydrogen atoms were intentionally left out of the mapping methodology. In organic chemistry, hydrocarbons build the simplest skeletal structures of all organic compounds. Thus, I determined that their fundamental and ubiquitous presence in all organic compounds would not add the variety that a rhythmic mapping procedure based on atomic differences would require.

12 I avoided arbitrariness by putting the ribose timeline first, followed by the nitrogenous base timeline, because, during the chemical synthesis of DNA, the ribose ring comes first, and the bases attach to it. For bases that generated two timelines, the six-membered timeline went first because the scientific community defines the starting point of the base as the six-membered ring.

13 Dave Brubeck, *Take Five*, sound recording (1959).

14 Evelyn Lamb, "Uncommon Time: What Makes Dave Brubeck's Unorthodox Jazz Stylings So Appealing?" *Scientific American* (11 December 2012): www.scientificamerican.com/article/uncommon-time-dave-brubeck (accessed 10 December 2017).

15 Dream Theater, *The Count of Tuscany*, sound recording (2009).

16 Toussaint [6] pp. 51–55.

17 Gerhard Nierhaus, *Algorithmic Composition: Paradigms of Automated Music Generation* (Mörlenbach: SpringerWienNewYork, 2009) p. 36.

18 Nucleotide, "Homo sapiens monoamine oxidase A (MAOA) gene, complete cds," *National Center for Biotechnology Information*: www.ncbi.nlm.nih.gov/nuccore/AH002871.2 (accessed 2 December 2017).

19 Toussaint [6] p. 25.

20 Nucleotide, "Homo sapiens hemoglobin subunit gamma 2 (HBG2), mRNA," *National Center for Biotechnology Information*: www.ncbi.nlm.nih.gov/nuccore/NM_000184.2 (accessed 2 December 2017).

21 See [19].

Manuscript received 26 July 2018.

ALVARO YANEZ *was a student at New York University Abu Dhabi, where he studied biology with a specialization in brain and cognitive science, and premedical and health studies.*