Carex Siderosticta Plastid - Photosystem II

By Tomasz Neugebauer and Nicolas Royer-Artuso

This composition is based on musical scores generated by software we developed that maps DNA sequences into musical notation. This particular example converts genes responsible for photosynthesis (photosystem II) found on the plastid of a carex siderosticta plant ([1]). We then had that score performed on two violins. We focused on the coding of the photosystem genes. However, the development of the software means that one could quickly convert any of the over 100 million individual sequences in GenBank into a musical score.

The software we developed parses FASTA nucleotide coding sequence files, and maps these into a musical composition. The algorithm maps each of the 20 amino acids onto specific pitches and each codon synonym onto duration for those pitches. The mapping is listed in Table 1. We used quartertones to be able to keep the results inside an octave. Rests at the end of a bar are added to create an 8/4 time signature – each amino acid note duration is based on which codon synonym appears in the sequence. The rests are added to avoid having to split notes across a bar. Our program writes out the results into two musical score files, one for the genes on each of the two strands of DNA. The resulting files use Lilypond ([2]) format to express a musical score for the genes located on each strand of DNA. Finally, the open source Lilypond program is used to generate the PDF and MIDI files of the score. The two scores are played simultaneously. An example transformation is illustrated in Figure 1.

The sonification of DNA data by conversion to a musical score through mapping has been done by a number of research scientists and musicians over the last thirty years ([3], [4], [5]). As far as we are aware, this is the first time that the *two* strand locations of genes are used together in building the compositional structure, i.e. DNA is a two-stranded molecule and a gene can be on the coding or the complementary strand. Thus, we generated two musical tracks, one for each strand, played simultaneously and this creates the specific polyphonic/contrapuntal textures we hear on this recording. Furthermore, even though we do create MIDI files, opening up the path towards electronic music composition, our approach is human centered in the sense that we insist on interpretation and performance by live interacting musicians.

Amino Acid	Pitch	Codon synonym and corresponding duration
Ala/A	C	GCT GCC GCA GCG 1 ¾ ½ ¼
Arg/R	C half sharp	CGT CGC CGA CGG AGA AGG 1 ¾ ½ ¼ 1/8 1/16
Asn/N	C sharp	AAT AAC 1 ¾
Asp/D	D	GAT GAC 1 ¾
Cys/C	D half sharp	TGT TGC 1 ¾

Gln/Q	D sharp	CAA CAG 1 ¾
Glu/E	E	GAA GAG 1 ¾
Gly/G	E half sharp	GGT GGC GGA GGG 1 ¾ ½ ¼
His/H	F	CAT CAC 1 ¾
Ile/I	F sharp	ATT ATC ATA 1 ¾ ½
Leu/L	F half sharp	TTA TTG CTT CTC CTA CTG 1 ¾ ½ ¼ 1/8 1/16
Lys/K	G	AAA AAG 1 ¾
Met/M START	rest	ATG 1
Phe/F	G half sharp	TTT TTC 1 ¾
Pro/P	G sharp	CCT CCC CCA CCG 1 ¾ ½ ¼
Ser/S	A	TCT TCC TCA TCG AGT AGC 1 ¾ ½ ¼ 1/8 1/16
Thr/T	A half sharp	ACT ACC ACA ACG 1 ¾ ½ ¼
Trp/W	A sharp	TGG 1
Tyr/Y	B	TAT TAC 1 ¾
Val/V	B half sharp	GTT GTC GTA GTG 1 ¾ ½ ¼
STOP	rest	TAA TGA TAG 1 ¾ ½

Table 1. Mapping of 20 amino acids to pitches and duration based on the codon synonym

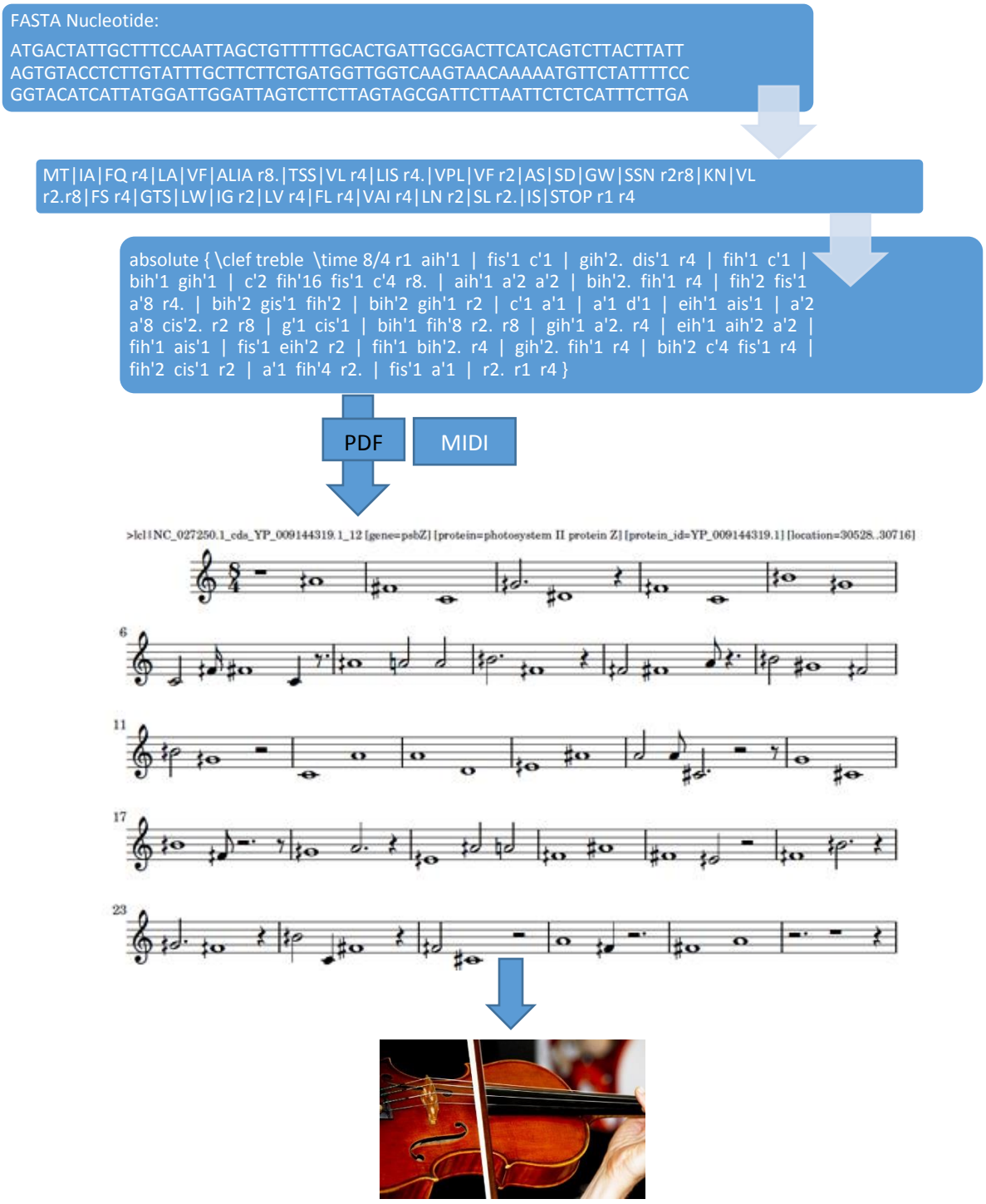

Figure 1. Transformation Process Example. The NC_027250.1_cds_YP_009144319.1_12 (photosystem II) gene FASTA nucleotide sequence is parsed for amino acids. Amino acids are mapped to pitch while the duration of each note is mapped by the codon synonym. Rests at the end of a bar are added when required to create 8/4 time signature. The software generates results in LilyPond format so that Lilypond can be used to generate the musical score as PDF and MIDI files. Finally, musicians play the score.

Biography of the Authors:

Tomasz Neugebauer is an Associate Librarian, Digital Projects & Systems Development at Concordia University (Montreal, Canada). His research interests include digital research data, information visualization, bioinformatics and open source software development. He has published in various scholarly journals, including Information Technology and Libraries, International Journal on Digital Libraries, International Journal of Digital Curation, Art Libraries Journal, and The Indexer. His most recent publication co-authored in the journal PLoS One describes the development of his DNA data visualization software.

Nicolas Royer-Artuso holds diplomas in music, composition, cognitive sciences and linguistics and is currently doing a PHD in linguistics at Université Laval. He is a proficient musician specialized in Ottoman music and related traditions (Iraqi,Syrian, Egyptian, Rembetiko). He has published and presented works on heterophony, Ottoman music theory, Turkish phonology and morphology, metrics, linguistic theory and language contact. He has released an album of his compositions based on cognitively realistic algorithms that generate heterophonic textures. He is currently working on the grammar of rhythm-free poetic vocal improvisation (gazel) and textsetting in the Ottoman era.

References:

[1] Carex siderosticta plastid, complete genome. NCBI Reference Sequence: NC_027250.1. National Center for Biotechnology Information, U.S. National Library of Medicine. http://www.ncbi.nlm.nih.gov/nuccore/NC_027250.1

[2] LilyPond... music notation for everyone. http://lilypond.org/

[3] Susumu Ohno and Midori Ohno. The All Pervasive Principle of Repetitious Recurrence Governs Not Only Coding Sequence Construction But Also Human Endeavor in Musical Composition. Immunogenetics 24: 71-78, 1986. http://www.ncbi.nlm.nih.gov/pubmed/3744439

[4] John Dunn and Mary Anne Clark. Life Music: The Sonification of Proteins. Leonardo. February 1999, Vol. 32, No. 1, Pages 25-32. http://www.mitpressjournals.org/doi/abs/10.1162/002409499552966#.Ve3ukNNViko doi:10.1162/002409499552966

[5] Rie Takahashi and Jeffrey H Miller. Conversion of amino-acid sequence in proteins to classical music: search for auditory patterns. Genome Biology. 2007, 8:405 doi:10.1186/gb-2007-8-5-405 http://genomebiology.com/2007/8/5/405