ESCHE BACH
Paper

Bioarts, Music
Author(s):
NICOLAS REEVES
Canada, University of Quebec in Montreal, School of Design
https://design.uqam.ca/
http://www.nxigestatio.org/NXI
SÉBASTIEN RODRIGUE
Canada, University of Sherbrooke, Department of Biology
www.usherbrooke.ca
VINCENT CUSSON
Canada, University of Quebec in Montreal, School of Design
www.uqam.ca
LÉONARD SAUVÉ
Canada, University of Montreal, Department of Biophysics
www.umontreal.ca

Abstract

The name of the project comes from the homonymy between "ECCE BACH" ("Here is Bach" in Latin), and the first letters of the full name of the bacterium E. coli, escherichia coli.

The project itself consists of coding several musical pieces of J.S. Bach into DNA strands, using a four-letter coding that corresponds to the four nucleobases that enters in the composition of this molecule, forming the basic elements of the genetic code, namely adenine, guanine, cytosine and thymine, identified by their initials A, G, C T.

The DNA strands thus encoded are injected into the genome of E.coli bacteria and of yeast micro-organisms. They replace some of the silent sections of the genome. These sections, whose role is just beginning to be understood, do not normally express themselves during morphogenesis and therefore have no known impact on the appearance of the micro-organisms (phenotype) or on their vital functions. Several bacteria and yeasts are prepared, each carrying a given piece. They are released into a nutritious substrate where they are left to evolve on their own for several weeks.

During this period, they grow and reproduce a very large number of times; a single generation lasts 20 to 30 minutes. By means of natural mutations and other phenomena, all sections of the genome, including the one where the musical piece is coded, are transformed and modified. At regular intervals, samples of bacteria and yeasts are retrieved. Their DNA is extracted and decoded at the precise places where the musical pieces were recorded. The modified nucleotide sequences are then converted back to music, resulting in a set of variations from the original themes.

These variations will then transcribed on classical music scores, in anticipation of a performance / concert where a selected sample of them will be performed by a quartet of classical musicians, accompanied by a pianist, and, depending on the playability of the vocal sections, by two choristers. The première of this concert is expected April 2019.

ESCHE BACH
In vivo musical variations

Prof. N. Reeves
School of Design, University of Quebec in Montreal, Montreal, Canada
www.nxigestatio.org/NXI
e-mail: reeves.nicolas@gmail.com

Prof. Sébastien Rodrique
Department of Biology, University of Sherbrooke, Canada,
https://www.usherbrooke.ca/biologie/

Vincent Cusson
Dept. of Communications, University of Quebec in Montreal, Canada
www.uqam.ca

Léonard Sauvé
Department of Biophysics, University of Montreal, Canada,
www.umontreal.ca

Abstract

The name of the project comes from the homonymy between ECCE BACH ("Here is Bach" in Latin), and the first letters of the full name of the bacterium E. coli, or Escherichia coli. The project itself consists of coding several musical pieces of J.S. Bach into DNA strands, using a four-letter coding that corresponds to the four nucleobases that enters in the composition of this molecule, forming the basic elements of the genetic code, namely adenine, guanine, cytosine and thymine, identified by their initials A, G, C T.

The DNA strands thus encoded are injected into the genome of E. coli bacteria and of yeast micro-organisms. They replace some of the silent sections of the genome. These sections, whose role is just beginning to be understood, do not normally express themselves during morphogenesis and therefore have no known impact on the appearance of the micro-organisms (phenotype) or on their vital functions. Several bacteria and yeasts are prepared, each carrying a given piece. They are released into a nutritious substrate where they are left to evolve on their own for several weeks. During this period, they grow and reproduce a very large number of times; a single generation lasts 20 to 30 minutes. By means of natural mutations and other phenomena, all sections of the genome, including the one where the musical piece is coded, are transformed and modified.

At regular intervals, samples of bacteria and yeasts are retrieved. Their DNA is extracted and decoded at the precise places where the musical pieces were recorded. The modified nucleotide sequences are then converted back to music, resulting in a set of variations from the original themes. These variations will then be transcribed on classical music scores, in anticipation of a performance /concert where a selected sample of them will be performed by a quartet of classical musicians, accompanied by a pianist, and, depending on the playability of the vocal sections, by two choristers. The première of this concert is expected April 2019.

1 • Introduction

Esche Bach is a research-creation project that draws his name from the homonymy between "ECCE BACH" ("Here is Bach", in Latin), and the first letters of the full name of the bacterium Escherichia coli, a.k.a. E. Coli. The project itself consists in coding several musical pieces by J.S. Bach into DNA strands, using a four-letter coding that correspond to the four amino acids that constitute the alphabet of this molecule and form the basic elements of the genetic code, namely adenine, guanine, cytosine and thymine, identified by their initials A, G, C and T.

The DNA strands thus encoded are injected into the genome of E. coli bacteria and of yeast mirco-organisms. They replace some of the so-called silent sections of the genome. These sections, whose role is not yet fully understood, do not express themselves during morphogenesis and therefore have no obvious impact on the appearance of the bacterium (phenotype) or on its vital functions.

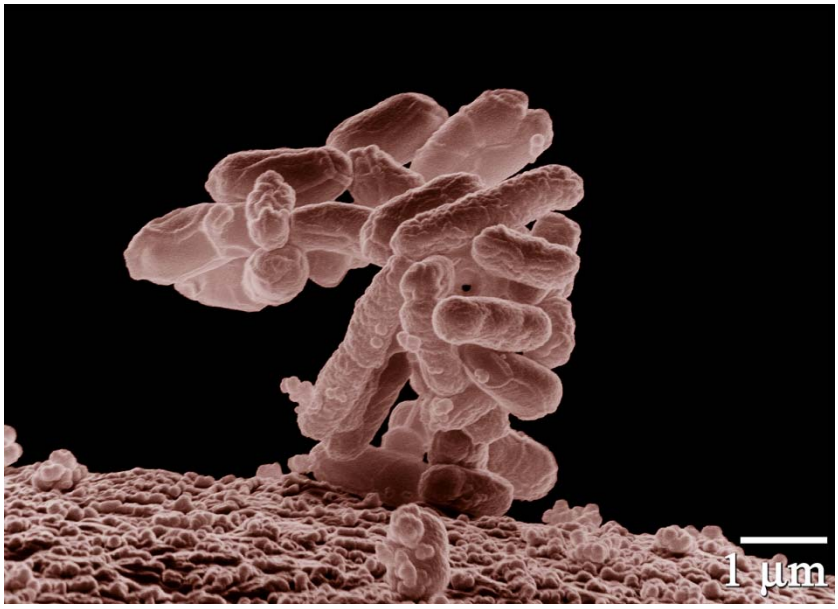

Fig. 1 • E. coli bacteria.

Several bacteria are prepared for the experiment, each carrying a given musical work. They are then released into a nutritious substrate where they are left to evolve for several weeks. Since the delay between two generations is about 20 minutes, they grow and reproduce a very large number of times. Through different modes of natural mutations, all sections of the genome, including the ones where the musical pieces are coded, are transformed and modified.

Samples of micro-organisms are collected daily. Their DNA is extracted and decoded, so the mutated pieces can be retrieved and converted back to music, resulting in a set of bio-musical variations from the original themes. The variations are then transcribed on classical music scores, in anticipation of a performance / concert where a selection of the mutated pieces will be performed by a small classical ensemble.

2 – Framework of the project

Fig. 2 • A fossil of senftenbergia plumosa, a fern that existed more than 300 million years ago.

The project was born from a questioning about the ability of living beings to preserve the information for times that are considerably longer than those typical of inert matter. When looking at a 300-million years old fossilized fern, one of the most amazing things is that it can be recognized and identified by its close similarity to the ferns we know today. This means that the information that determines the morphology of the fern has travelled during all these years, being relayed from individual to individual, without significant alterations. 300 million years is the time for a mountain range to be born, to peak in the stratosphere and to flatten into a peneplain, and even sometimes, as in the case of the Pyreneans, for a second chain to rise over its remnants. A symbol of permanence for millennia, which is demonstrated by monuments such as pyramids, which adopt the shape of mountains to carry the body of illustrious people towards eternity, the mountain has lost its status in recent decades: it is now well pale compared to the living in its ability to guarantee the permanence of information. We now begin to realize that reaching eternity, which has been one of the main mythical concerns of human beings since the dawn of time, calls for the strategies of the living, and no longer by the properties of inert matter. This question is now the object of advanced research programs : several laboratories are currently working on the design of DNA memories, which, through the replication of the information contained in these molecules, by the quasi-astronomical level of redundancy provided by the rapidity of cell replications, and thanks to all the self-repair mechanisms and control processes that exist in all living cells, could ensure the transmission of information towards the future for durations that are unimaginable today.

Although the preservation of genetic information is extremely efficient at the species level, the situation is, as we know, very different for individuals, whose chromosomes are submitted to several kinds of transformations, from mutations caused by radiations and cosmic rays to genetic cross-over, some of which being able to introduce considerable changes. They allow individuals to evolve through differentiation, in order to maintain a high-level of adaptability in the event of critical environmental variations.

Our starting hypothesis was that the transformation of existing musical pieces by these same phenomena could become a process of musical composition capable of producing unexpected propositions. Although this is outside our area of expertise and outside the specific scope of our project, the method also presents possibilities for transfer to biology and genetics : it is not unrealistic to think that this form of sonification could produce new knowledge about the precise unrolling of genetic mutations, and of the channels by which information is transmitted from one individual to its offspring.

3 • Experimental framework and methodology

The project is the concretization of an idea that I had several years ago, but which was difficult to implement at the time, due to the complexity and costs involved by the production of genes and the insemination of bacteria. Since the beginning of the current decade however, a new gene editing procedure has been developed. Called CRISPR-Cas9, it makes it possible to perform these manipulations with great efficiency and at costs considerably decreased [1]. Developed by the French biologist Emmanuelle Charpentier assisted by the American Jennifer Doudna [2], this technique is revolutionary enough to place its inventor on the list of potential Nobel prizes, which would already be done if legal challenges had not occurred in the meantime, but these considerations are also out of the scope of this paper.

The musical insemination of bacteria and yeasts can follow two paths. In the first one, the coded DNA strands is directly injected in the micro-organisms. In the second, the DNA of a virus (*phagus*) will be modified. The virus will then attack the micro-organisms and contaminate it : it will be used only as a vehicle to enter the cell. Since viruses evolve much faster than bacteria, this second method ensures faster replications and maximizes the possibilities of mutations and evolution within a given time. The modification of the genome of a phagus however is a more delicate operation. To maximize the chances of success, the two paths will be followed in parallel.

The decision to use musical pieces by J.S. Bach was taken for several reasons :

• Many pieces of Bach's repertoire are known enough to allow most audiences to easily identify the variations from the original.

• Much of Bach's repertoire has a clear and rigorous formal structure. Some pieces, like the Goldberg Variations, have been elaborated by posing it a priori, even before the composition process; not only the pieces, but also the whole set of variations, are based on a strong structural pattern [3].This also facilitates the precise identification, location and nature of the modifications, even for people that are less knowledgeable in music.

• Bach's scores have no indication of expression, giving the performer complete freedom to accentuate the play or dynamics of the different passages. This results in a vast spectrum of potential interpretations, which for the Variations range from Glenn Gould's ardency to Evgeni Koroliov's restraint and refinement. Unlike the musical notation itself, which lends itself rather easily to numerical quantization, indications of expression are qualitative: it would be very risky to attempt a formal coding, which would be necessarily blurred and biased by subjective concerns. Starting from Bach's scores, the musical notation obtained at the end of the process will remain as close as possible to the composer's initial will, leaving the performer the totality of his expressive freedom.

For these reasons, as well as for other reasons that will be outlined below, the current list of potential candidates includes the Goldberg Variations, 2- and 3-voice Inventions, as well as four-voices chorals.

4 • E. coli : in vivo considerations

In the early phases of the project, the decision was taken to start from musical pieces already coded by the MIDI protocol. Developed in the beginning of the 80's and still heavily used by contemporary musicians of about all possible styles, it includes, directly or indirectly, all the information required for the project, thus allowing to streamline the transposition of the piece into the genetic code.

Once coded, as can be seen in the picture of the program interface below (Sect. 7, Fig. 7) , the piece is first submitted to different simulations based on mutation rates whose values come directly from biological observations. Different coding methods can be tested during this phase. Hypothesis can be made about the time required by the in vivo mutations to generate perceptible transformations, and eventually about the necessity to optimize the mutation rate, either through radiations such as UV rays, or by bio-genetic methods, such as the use of phagi, as mentioned above. Last but not least, it allows to test the whole process with different musical pieces, in order to select those in which the changes will be important enough to be noticed without completely upsetting the original themes.

The first piece we considered was the Goldberg Variation No XVIII, or Canon to the Sixth. We are also experimenting with our other candidates, but this one is used as a reference piece for the present paper. Mutation rates for the simulations are set from E.coli natural rates.

E.coli are unicellular micro-organisms that enter the vast category of prokaryotes, meaning that they have no cell nucleus. Their genome, certainly the most studied of all living organisms, includes between 4,6 and 5,3 millions genes, allowing them to code between 4200 and 5300 proteins. At 37 Celsius, they reproduce at the rate of three generations per hour. Given enough nutriments, one single bacterium can thus produce a population of more than 100 million in one single night. Being deprived of nucleus, they cannot evolve through processes such as chromosome cross-over : genetic mutations are essentially caused by replication errors or environmental causes, such as mutagen substances, or natural or artificial radiations.

The central dogma of molecular biology, the one that describes the unrolling of events during cell replication and reproduction for all living beings on earth, states that the DNA encodes all the information required for the production of biological matter. This information is first transcribed into RNA strands. These very long molecules then carry the information in specific locations of the cells, where they are translated into proteins. Some of these newly created proteins are sent back to control the translation and transcription of the RNA information, but the vast majority becomes the main components of the organism's body. Three kinds of RNA actually exist; the one that is in charge of carrying the genetic code is called "messenger RNA", often abbreviated as mRNA.

ALA	Alanine	GCT	GCC	GCA	GCG		
ARG	Arginine	CGT	CGC	CGA	CGG		
ASN	Asparagine	AAT	AAC	AGA	AGG		
ASP	Aspartate	GAT	GAC				
CYS	Cystéine	TGT	TGC				
GLN	Glutamine	CAT	CAC				
GLU	Glutamate	GAA	GAG				
GLY	Glycine	GGT	GGC	GGA	GGG		
HIS	Histidine	CAT	CAC				
ILE	Isoleucine	ATT	ATC	ATA			
LEU	Leucine	TTA	TTG	CTT	CTC	CTA	CTG
LYS	Lysine	AAA	AAG				
MET	Methionine	ATG					
PHE	Phenhylalanine	TTT	TTC				
PRO	Proline	CCT	CCC	CCA	CCG		
SER	Sérine	TCT	TCC	TCA	TCG	AGT	AGC
THR	Thréonine	ACT	ACC	ACA	ACG		
TRP	Tryptophane	TGG					
TYR	Tyrosine	TAT	TAC				
VAL	Valine	GTT	GTC	GTA	GTG		
STOP		TAA	TAG	TGA			

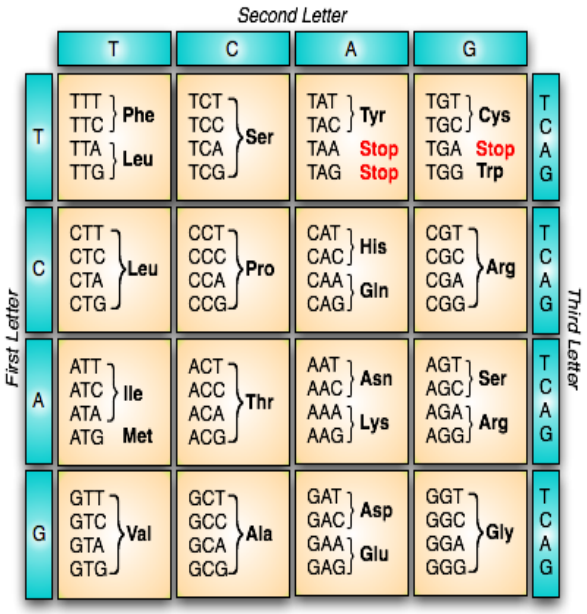

Fig. 3 • Two inverse representations of the redundancy of the genetic code. Left : the 20 main amino-acids and the STOP instruction, followed by the 3-nucleotides codons that produce them. Right : to find the amino-acid produced by a given codon, take the first nucleotide on the left of the table, the second on the top and the third on the right. The resulting nucleotide appears on the corresponding cell.

DNA and RNA molecules are made of chains of nucleotides, themselves composed of a nitrogenous base, or nucleobase, a sugar with five atoms of carbon (ribose in the case of RNA, deoxyribose in the case of DNA), and one or several phosphate groups. In both cases, there are four possible nucleobases : adenine, guanine, cytosine, uracil, designated by letters A, G, C, U. In DNA, thymine (or T) replaces uracil.

The basic elements of the coding are triplets of nucleotides, called “codons”. During the proteogenesis, each triplet produces a particular amino-acid, which is the basic component of all proteins. About 500 amino-acids are known today, but the DNA/RNA coding produces only the 20 varieties that are required for living beings[1]; it also produces a terminating instruction (a STOP codon) which specifies the moment where the protein molecule is completed and can be released within the cell cytoplasm.

Considering that each element of each triplet can take four different values (A, G, C or U), the genetic code can yield 64 different outputs (4^3), which means that there is a high level of redundancy. The two tables below (Fig. 3) show the correspondences between the triplets and the amino-acids they generate.

One can see immediately that the outputs of the code are not equally distributed : amino-acids such as leucine or serine can be produced by six different codons, when others, such as methionine or tryptophan, are generated by a unique codon.

As we will see in the following section, this repartition became a central element of the code we developed for transposing musical sequences into genetic sequences, namely the ESCHE BACH code.

5 – Musical bio-variations : towards the ESCHE BACH code.

At the coding level, there are many ways to transpose a musical score into a four-letter alphabet. Several possibilities have been explored. They are all based on the fact that the DNA sequences impose a codon vocabulary of three signs, each one taking one of the four values A, G, C or T. A codon can then take 64 different values. This is not a high number : as we will see below, it imposes important constraints on the type of coding that will be used.

We first tried to use a generic code that could be used for about all classical music pieces. Its basic principle consists in coding each note by four codons, yielding a basic word, or template, made of four triplets :

111• 222 • 333 • 444

The individual digits can take any of the four values A, G, C and T. The sequence of the letters between each triplet must be ordered so as to correspond to the sequence of the first integers; it was arbitrarily decided to define AAA as the lowest number and TTT as the last one, meaning that the ordered sequence, from the lowest to the highest value, would look like :

AAA, AAG, AAC, AAT, AGA, AGG.... (...) ... TCC, TCT, TTA, TTG, TTC, TTT

Corresponding to numbers 1 to 64.
Each triplet has a different meaning :

- Triplet 111 indicates the position of the bar containing the note in the score.
- Triplet 222 determines the position of the note within the bar.
- Triplet 333 indicates the height (or frequency) of the note, in semi-tones.
- Triplet 444 indicates the duration of the note, counted in multiple or sub-multiple of the basic note of the score, as read on the metric indication at its beginning.

As mentioned above, the MIDI protocol is used for describing each note. Some data like the height of a note, can be directly extracted from MIDI messages; it is represented by an integer value between 0 and 127. Other can be indirectly determined : the duration of a note, for instance, must be computed by comparing the moments where the note begins (“Note On”) and where it ends (“Note Off”). Since the number of possibilities for parameters such as note duration is much lower than 64, this first coding used a cycling numbering process : if we had only 8 possible durations, then the 9th element of the code would go back representing the first duration, the 10th would represent the 2nd, and so on.

The situation was different for the height of the note, since the number of possibilities (128) is twice the possible number of “words” of the corresponding triplet. We then had to consider a first adjustment to our code : since our reference piece had an ambitus of 52, we considered only the 64 values starting 6 notes under the lowest note, and ending 6 notes over the highest. The ambitus was then precisely centred within the 64 possible values.

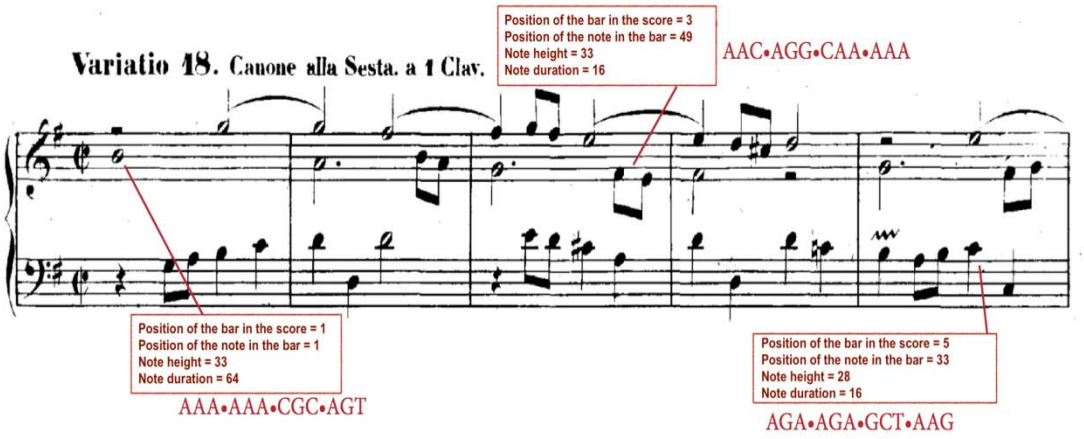

Fig. 4 • The first bars of J.S. Bach's Goldberg Variation No XVIII, or Canon to the Sixth, with three notes represented by the corresponding quadruplets of codons in the ESCH BACH code.

Any classical score can be seen as 2-D coordinates frame, where time lies along the horizontal axis and frequency along the vertical one. Since frequency is the inverse of a time (1/t), any score is the spatial representation, or recording, of a strictly temporal phenomenon. The main interest of our basic coding is the fact that it determines an absolute time-positioning system in this frame. Each note is precisely defined by its height (frequency), coded in the third triplet, the very moment it occurs within the piece, coded by the two first ones, and its duration, coded in the last one. Each quadruplet then carries all the information required to completely define the corresponding note: even a random list of quadruplets will allow a precise reconstruction of the whole score.

6 • The ESCHE BACH coding

Unfortunately, despite the simplicity of this process, and as opposed to what we had initially hoped, several attempts led us to conclude that, even by proceeding to specific adjustments for height or other parameters, no straightforward unique code could be established for all possible pieces. We could have bypassed this problem by using longer codons, but we would have lost the correspondence with the genetic coding; the resulting chains of codons would have become useless in a real biological environment. Sure, we have no idea of the potential impact of musically crypted strands of nucleotides in a living cell, but we wanted from the beginning to preserve a maximal level of similarity between the genetic and the musical codes : we wanted our coded data to be theoretically able to generate proteins when reinjected into a bacterium - even if no one can predict which kind of proteins would be produced, or if they would be viable.

In order to work within these constraints, we decided to do our first experiments with scores including a maximum of 64 bars. Each bar is divided in 64 time intervals, with notes mapped to a 64-steps frequency scale, and a maximum duration of 64 times the smallest time increment of the piece.

These factors could have been seen as strongly limiting. They actually proved quite relevant for our experiments. Using a biological analogy, a musical piece, when played, can be seen as the phenotype produced by the score, which plays the role of the underlying genotype. If we want the evolution of this individual of a particular species not to diverge towards quasi-random sequences, and thus to lead to an entirely different musical species, it becomes essential to start from the

main musical features of the original piece and to ensure that the coding will be able to preserve them. It was then decided to submit the piece to a pre-processing analysis during which a unique coding key is generated. This key will be required for the inverse coding and will always travel with the piece. It includes all the information specific to the piece, such as its ambitus, or the list of all note types and heights.

As seen above, the ambitus of our reference piece, or the interval between its lowest and highest note, is 52 : the lowest and highest notes are respectively A1 and C6, which corresponds to MIDI notes 33 to 84. Eight different durations and thirty-seven different tones are encountered. The coding key will insure that the evolution occurs between the ambitus interval, using only durations and tones existing in the original piece. These constraints will preserve the kinship with the original musical phenotype, and also, on a more pragmatic level, make sure that the resulting score will not incorporate sequences that would be impossible to play with classical instruments.
The coding of note durations has been the object of several discussions and attempts. The final decision was to base it on the metric of the score, which, for our piece, is 2/2, as indicated by the C-bar at the beginning. The duration unit is defined by the denominator of this fraction, corresponding here to a half note. To this duration, we give the value 1. All other durations take their values from there : a whole note will take the value 2; a dotted half will be 1,5; a quarter note will be 0,5; an 8th note will be 0,25, a 16th note will be 0,125, and so on. One half of the value of a given note is added to generate its dotted version. Each note of a triplet will use the value 2/3 of the two equivalent notes in the score. The absolute duration of a note, measured in milliseconds, will thus depends both on the metrics of the piece and of its indicated tempo.

By analogy with in vitro mutations, during the equivalent musical mutation, existing notes can be deleted, and new notes can be added, at different locations in the mutated score. To deal with this phenomenon, the precise structure of the MIDI code allowed us to explore two alternatives. The first one is based on the absolute positioning of each note in the score : a suppressed note will generate a silence; if a note is added, or if the duration of a note is increased, the tones will superimpose and the listener will hear a short polyphonic sound. The second one is based on a relative positioning : the position of each note is determined relatively to the previous one. This means that the introduction or deletion of a note will result in a temporal shift for the remaining of the code. Our simulations have shown that the second possibility opens on a very rich landscape of unexpected rhythmic variations. Examples of both possibilities have been computed. They will be presented during the lecture, and are available on the project web site.

The question of the coding redundancy has been addressed by using a ponderation table for the evolution of duration and height parameters. The genetic coding redundancy will be represented here by a wheel diagram (Fig. 4) :

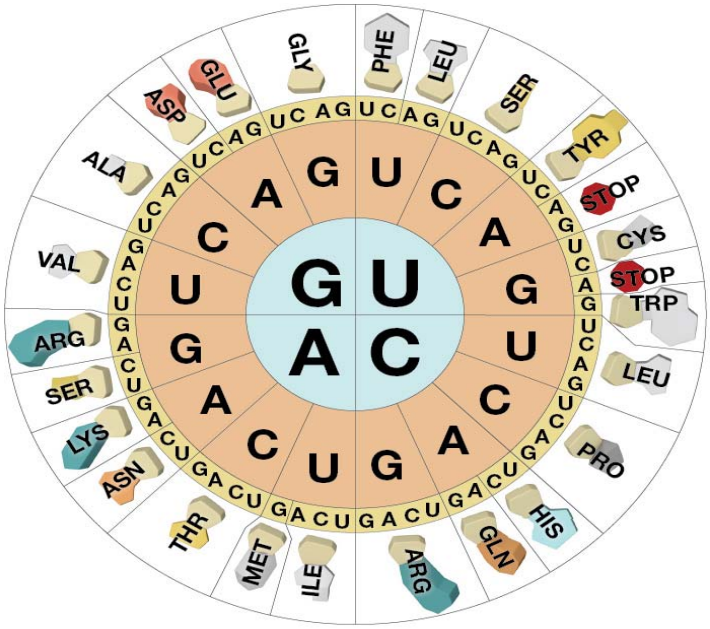

Fig. 4 • Another representation, in form of a wheel diagram, of the amino-acids that are produced by the different codons, in the case of mRNA.

To know the correspondence between a nucleotide triplet and an amino-acid, we start by choosing a letter in the centre of the wheel (in blue), for instance G. Then a second letter is selected among the four corresponding possibilities in the light brown section, for instance C. A final choice is made in the yellow section, for instance C. The amino-acid that will be produced can be read in the external ring : we can see that the GCC triplet produces alanine (abr. ALA). We see also that the last letter does not really matter in this case, since all triplets beginning by GC will lead to the same amino-acid. On another hand, the AUG triplet is the only one that will yield methionine (or MET).

Our ponderation table is established by a direct mapping between the frequencies of each musical parameter in the piece and the occurrence frequency of each amino-acid : the notes that appear more often in the piece will get a ponderation similar to the amino-acids that are produced by the largest numbers of codons. Since the number of values for each parameter differs most of the time from the number of possible amino-acids, a scaling is required : the frequency of occurrence of the 8 possible notes durations, and of the 37 possible heights or tones, is computed; the most frequently encountered values are associated with the amino-acids that are produced by the most triplets.

For the duration and height parameters of the reference piece, the process yields wheel diagrams similar to the one above :

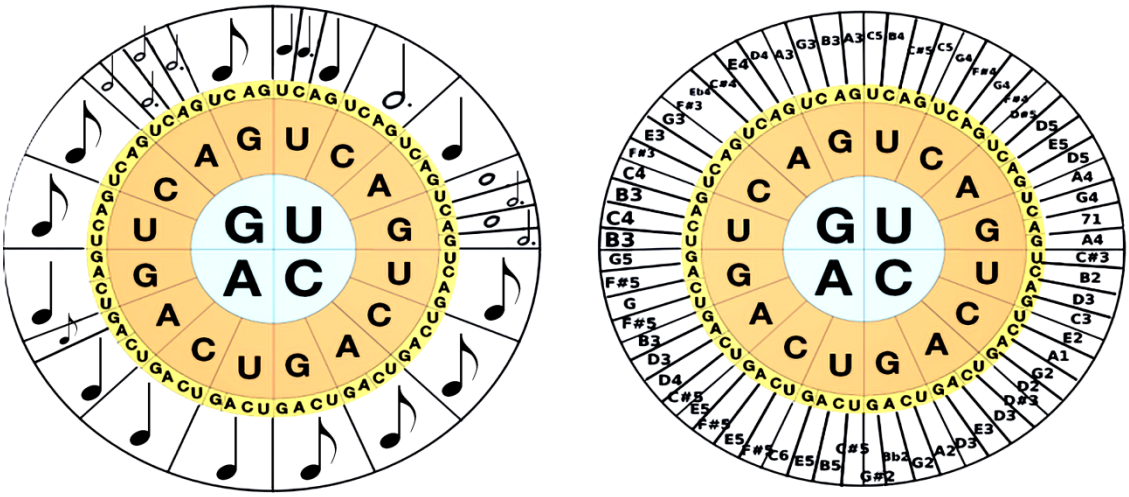

Fig. 5 • Wheel diagrams showing the pondered correspondences between codons and note durations (left), and between codons and notes (right), for the final ESCHE BACH coding. The durations and the heights include only those that are encountered into the original piece.

The evolution of the piece will thus generate musical mutations in which the relative frequency of these two parameters will be mapped to the relative frequency of occurrence of the 20 main amino-acids, so as to maintain a kind of bio-correspondence with their frequencies on the original piece. Since this ponderation varies for each musical piece, it is recorded in the coding key.

7 – The ESCHE BACH software interface

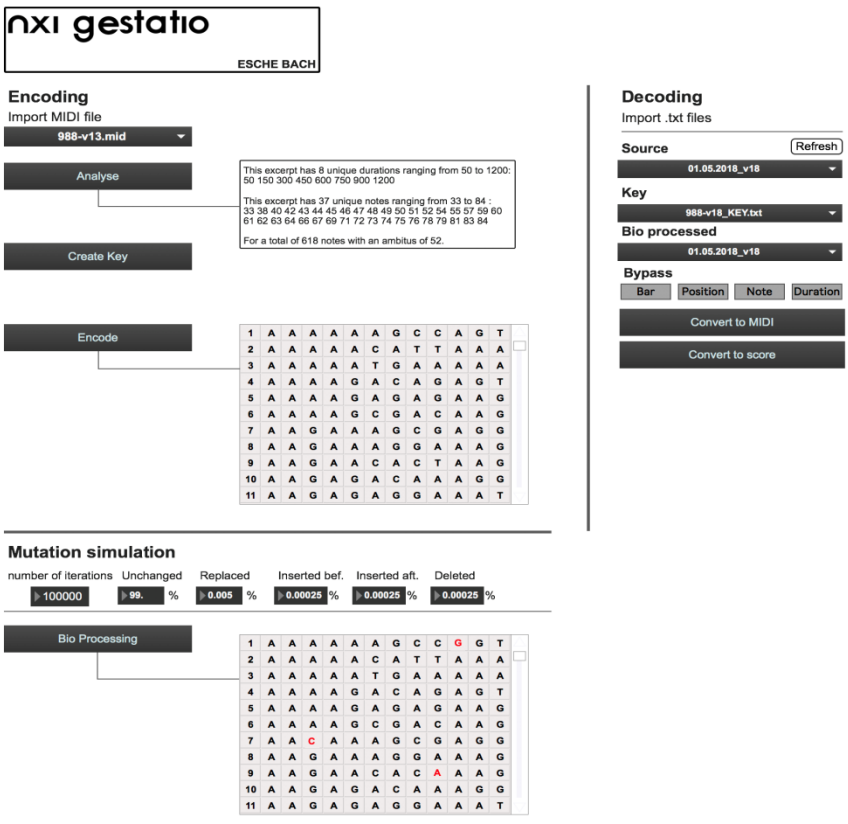

Fig. 7 • The ESCHE BACH interface.

Using the MAX/MSP platform, we developed a program that allows to go through all the steps of the encoding-decoding process, for simulations (*in silico*) as well as for real (*in vivo)* explorations. To start the work, the user first selects a MIDI file with the upper left box (see Fig. 7). By clicking on the "Analyse" button, he extracts the required musical parameters : ambitus, list of durations and heights, total number of notes... The "Create Key" generates the coding key specific to the piece. The "Encode" button converts each note in 3-nucleotides quadruplets, generating, for Variation Goldberg No 18, a 618-quadruplets list that can be either directly sent to the genetic engineering firm in charge of producing the real DNA strings, or used for computer simulation. The list itself appears in the top scrolling window.

The bottom section is meant to explore the effect of different parameters on the mutation. The composer can enter numbers derived from real biological data, in order to predict the modifications that will occur after a certain amount of generations, but he can also enter numbers completely different, so as to use the simulation module for direct musical explorations unrelated to genetic evolution, thus skipping the *in vitro* phase. The bottom scrolling window shows the mutated nucleotides. They are displayed in red, so that the composer can see immediately observe the rhythm of changes along successive generations.

For *in vitro* mutations, the DNA strands, once received from the firm that assembled them, are injected into bacteria and yeasts, which are then let to evolve for a certain period. Even if the simulation phases allow a certain level of predictions about the results, it remains difficult to precisely determine the optimal evolution time after which potentially promising results can be obtained. From the mutation parameters, as well as from our discussions with bio-geneticists, it appears that several weeks are needed. We decided to start with a one-month experiment.

During that time, samples are collected every day and put in a freezer. At the end of the experiment, the genome from the last generation is sequenced to see the level of modifications introduced by the mutations. If it proves too important, or unsatisfying for any reason, samples from previous generations, taken at various intervals, are thawed. Their DNA is sampled and analysed, and the most promising sets of variations is selected. The corresponding lists of codons can then go through the following phases.

In the Decoding section, the composer must input three files : the original piece, the corresponding coding key, and the processed piece; the last one can indifferently result from in vitro or in silico mutations. The Bypass section allows to select the variations introduced by one or several of the four parameters : he can for instance limit the variations to height or durations, for instance, leaving the two addressing parameters (address of the bar and address of the note in the bar) unchanged. The conversion of the results to MIDI files ("Convert to MIDI") allow to listen to the different bio-modified pieces, to submit them to a preliminary evaluation, and to select those that will be played. Finally, the "Convert to score" button generates the scores for the small classical ensemble that will interpret them during the final performance.

8 • Conclusion : arts, cures and disasters, or the ambiguous potential of genetic engineering.

Since the beginning of the 90's, innumerable attempts have been made to analyse, generate, compose or simulate music through computer algorithms inspired from biological phenomena :

neural networks, cellular automata, genetic algorithms... Genetic algorithms of several kinds, all inspired by the ways chromosomes and genes evolve in natural cells, remain especially popular [4]. Looking at the amount of experiments and attempts in this field, one could easily suppose that this fascination for biological and genetics simulations would have quickly resulted in the implementation of *in vitro / in vivo* experiments, using real chromosomes and micro-organisms. Such attempts are surprisingly rare. They did not, to our knowledge, produce any valuable result. One of the reasons for this is probably, as mentioned above, the complexity and costs of the equipment and resources required for such attempts. Since the invention of the CIRSPR-Cas9 method however, things have changed, and it is now possible to foresee the moment where biologists and geneticists, professional or amateurs, will be able to afford and use desktop genetic engineering devices – a potentially frightening situation.

ESCHE BACH is an art project that addresses directly the complex problematics created by the possibility of cheap, efficient and easy genetic manipulations for everyone. It comes at a time when critical questions arise in terms of the risks and potential of genetic manipulation in particular, and in genomics in general. Some of these issues are worrisome. Apart from the obvious and dreadful risks of genetic terrorism, how can we ensure that methods that efficient and that accessible will not be used to modify the human genome, with potentially devastating consequences for the entire species? How can we prevent genomic techniques to drift towards eugenics, or to lead to situations where only a few companies, thanks to patenting strategies, will determine future researches on living beings? How can we prevent potential employers to use people's gene pool as a criterion for hiring, by government agencies as an immigration criterion, or by insurance companies as eligibility criteria? How to balance all these risks with the huge potential benefits of these same techniques for the treatment of rare diseases, or for food production?

We do not have answers for these questions. Since our project proposes a new method for musical composition, aiming at the creation of new musical pieces, its objectives are essentially artistic. However, as it is the case for many works from bio-arts, the use of genomic methods as part of a creative process is both a way of staging these techniques and to make the general public aware of their ever-growing presence in our daily life. By revealing their power and their stakes, not only do we try to demonstrate to new audiences their potential for explorations in research-creation, but we also hope to stimulate discussions, and even controversies, over their potential uses in all fields and disciplines.

NOTE

1 • Living matter is actually made of 22 amino-acids, but we concentrate here on the 20 that are directly generated by the genetic code.

REFERENCES

[1] - Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014 Jun 5;157(6):1262-78. doi:10.1016/j.cell.2014.05.010. Review.

[2] - Doudna JR, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science, Vol. 346, Issue 6213, 28 Nov 2014.

[3] - http://www.pianosociety.com/pages/bach_goldberg_variations/. Last consulted 23/11/18.

[4] - The page https://link.springer.com/article/10.1007/s00500-012-0875-8 includes a list of references that shows the level of activity in the field of genetic algorithms applied to music.

ILLUSTRATIONS

All illustrations by Nicolas Reeves & NXI Gestatio, except :

Fig. 1 • E. Coli bacteria. Photo by Eric Erbe, digital colorization by Christopher Pooley, both of USDA, ARS, EMU. - ARS Image Gallery Image Number K11077-1 Public Domain, https://commons.wikimedia.org/w/index.php?curid=130129.
Last opened Sept. 3rd, 2018.

Fig. 2• Fossil leaves of the plant species Senftenbergia plumosa, Upper-Carboniferous. Specimen ca 40 cm in width, collection of the Universiteit Utrecht. Photo by Woudloper, Bern (Switzerland), 2008.
Public domain, https://commons.wikimedia.org/wiki/File:Senftenbergia_plumosa.jpg

Last opened Sept. 3rd, 2018.