Let The DNA Speak
Project Proposal

Stefan Appelhoff, Kim Philipp Jablonski,
Nina Krüger, Sourabh Lal,
Tom Wiesing & Mengyuan Zhang

March 17, 2014

Contents

1 Introduction

One of the most obvious reasons to use different representations of DNA is to compare it. Ranging from graphical devices like codon atlas or chromatograms, over more data centered possibilities like simple tables, they offer a wide variety of learning and perceiving the structure of DNA through many different senses. However, we think that one promising opportunity has been wantonly neglected, namely the way of sonification. In our opinion this offers a whole new spectrum of interaction and understanding of the DNA you

1

probably only know from images or graphs. In order to demonstrate the applicability of our approach, we are going to compare several DNA strands. The main difficulty then was to find DNA sequences which are similar enough to be comparable, but different enough to have a noticable auditory difference. In order to achieve that goal we chose genes which encode for the same function but exist in different organisms. Another approach would be to compare mutated and healthy genes and thus deduce their possible severity. Hereby we hope to observe a general similarity of the compared strands, but with significant differences in distinct positions, which will then lead to a new auditory result.

2   Purpose

Our program mainly allows for fast comparison of very large DNA sequences. For instance, imagine you are in the laboratory replicating a huge amount of genes by cloning. Of course you always want to clone the exact same gene, meaning they should all have the exact same DNA sequence. But how do you want to make sure, that this is the case? DNA sequencing is relatively easy. However, you still have to compare every single nucleotide of your original sequence to every single nucleotide of your cloned gene. Naturally, this would be a huge waste of time. Now further imagine you could listen to music while you work. Then, suddenly the music sounds a little off and exactly that tells you that there is a wrong triplet in your cloned DNA sequence. In exactly the same way you could compare several sequences at the same time and as soon as you hear something weird, you can just check the corresponding nucleotide triplets, which are depicted on the screen and easily identify the faulty one.

3   Methods

The required features will be using JavaScript HTML and CSS. We are going to create a website where can interact with the sonified DNA. We will provide a lot of sample DNA as well as allow the end user to input their own DNA sequences. The website will be available at [12]. We are using two different libraries for implmenting the sonfication, timbre.js from [13] and MIDI.js from [14]. The libraries allow us to both play sine waves and

thereby construct all kinds of sound manually as well as to play notes on different instruments directly.

On the website there will be different modes to sonify DNA. One mode is to group the DNA into triplets and then assign a note to each triplet. Another mode is to play DNA on a per-base basis, i. e. assign each "letter" in the input string a note and then play that. Apart from having different modes, it will also be possible to "play" DNA using different instruments such as piano, xylophone, violin and trumpet. When playing DNA it is also possible to play different DNA at the same time. When using different instruments for different DNA, it will be possible to easily compare DNA by hearing if the instruments play in sync.

We will get our data from the European Nucleotide Archive. The European Nucleotide Archive is a publicly available database of Nucleotide sequences. Individual sequences have a length of around 300000 characters. Sonifying these long datasets will be very difficult because we will have to select only small fractions of them to use. We expect to find large parts of the sequences to be random because not all information contained in them is actually required by the organism the sequences come from.

4   Background & Motivation

TBD Stefan

5   Who does what

TBD Mengyuan

6   References

1. http://www.ebi.ac.uk/ena/

2. http://sciencecentres.org.uk/projects/handsondna/4.8%20-%20Amazing%20facts%20and%20quiz%20questions.pdf

3. http://holykaw.alltop.com/17-interesting-facts-about-dna

4. http://chemistry.about.com/od/lecturenoteslab1/a/10-Interesting-Dna-Facts.htm

5. http://www.express.co.uk/fun/top10facts/394517/Top-10-facts-about-DNA

6. http://www.bbc.com/news/magazine-25975712

7. http://sonification.de/handbook/

8. http://datasonification.tumblr.com/

9. http://lhcsound.hep.ucl.ac.uk/page_sonification/Sonification.html

10. http://sonify.psych.gatech.edu/research/

11. http://lhcsound.wordpress.com/2011/01/03/how-we-sonify-atlas-data-technical-

12. http://www.letthednaspeak.tk

13. http://mohayonao.github.io/timbre.js

14. https://github.com/mudcube/MIDI.js

15. http://blog.gleitzman.com/post/63283830260/midi-js-playing-audio-in-the-brows