

# DNA Sonification For Public Engagement in Bioinformatics

**Heleen Plaisier**

Royal Botanic Garden Edinburgh

**Thomas Robert Meagher**

University of St Andrews

**Daniel Barker** (✉ [Daniel.Barker@ed.ac.uk](mailto:Daniel.Barker@ed.ac.uk))

The University of Edinburgh

---

## Research note

**Keywords:** bioinformatics, sonification, DNA, sequence, public engagement, Raspberry Pi, Sonic Pi

**Posted Date:** April 30th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-471137/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Research Notes on July 15th, 2021.  
See the published version at <https://doi.org/10.1186/s13104-021-05685-7>.

# Abstract

**Objective:** Visualisation methods, primarily color-coded representation of sequence data, have been a predominant means of representation of DNA data. Algorithmic conversion of DNA sequence data to sound – sonification – represents an alternative means of representation that uses a different range of human sensory perception. We propose that sonification has value for public engagement with DNA sequence information because it has potential to be entertaining as well as informative. We conduct preliminary work to explore the potential of DNA sequence sonification in public engagement with bioinformatics. We apply a simple sonification technique for DNA, in which each DNA base is represented by a specific note. Additionally, a beat may be added to indicate codon boundaries or for musical effect. We report a brief analysis from public engagement events we conducted that featured this method of sonification.

**Results:** We report on use of DNA sequence sonification at two public events. Sonification has potential in public engagement with bioinformatics, both as a means of data representation and as a means to attract audience to a drop-in stand. We also discuss further directions for research on integration of sonification into bioinformatics public engagement and education.

## Introduction

As the field of genomics has matured, the need to provide novel tools for representing DNA sequence data has become pressing. This need ranges from comparative investigation of sequence variation across multiple sequences, searches for functional domains across extended sequences, searches for sequences of varying length within and between organisms that share homology or are otherwise similar, and capturing and comparing entire genomes. As genomics data and its potential for scientific investigation has grown, so too have methods for representing data. Visual sequence data has been a standard approach, but it has its limitations, especially when long sequences are involved. Algorithmic conversion of DNA sequences to sound – sonification – offers an alternative means of representing DNA sequences that in turn draw upon other human sensory mechanisms.

DNA sequence visualisation has a relatively long history in bioinformatics. Even the representation of the four bases as letters in the alphabet – A, C, G and T, with the sequence of one strand of the double helix represented as a string of such letters – is a simple visualisation, with a long tradition in databases and publications involving DNA sequence (e.g. [1, 2]). Visualisation of genome annotation, pairwise sequence alignment diagrams and multiple alignments take visualisation further, by adding visualisation of properties of the sequence (e.g. predicted gene structures or similarity to other sequences).

However, like any representation, visualisation has limits. In a research context, visualisation of large amounts of sequence is constrained by the size of a computer screen, the ability to convey diverse information through colours or symbols, and the attention of the researcher. In the context of public

engagement with science, DNA sequence has a monotonous visual appearance, at odds with the “whizz-bang” often expected in such public displays or shows.

Although the quantity and apparently random nature of large DNA sequences is engaging for a short time, more detailed examination of DNA sequence or annotation demands a level of focus that is atypically high for a public drop-in event or short activity. At public events we have had success with a wide range of ages using the Phylo game [3], where anyone can attempt to improve multiple alignments of regions of genes associated with human disease [4]. However, any one game or activity can only present a small aspect of bioinformatics. Although we have had success introducing bioinformatics to school pupils [5, 6] and introducing bioinformatics and computational science to the public in a science centre [4], bioinformatics as a whole is rather lacking in public engagement activities.

We aim to help fill this gap by means of sonification of DNA, where the representation of sequence is auditory rather than visual. Since music is widely understood and used for relaxation and entertainment, this may be more attractive to a public audience than a screen full of DNA sequence or associated data. By use of the Sonic Pi program [7] for sonification, we make the underlying program visible and customisable in real-time. This also provides a link between DNA and programming, which is central to bioinformatics in general.

Sonification of scientific data is not new (e.g. [8]). More specifically, the use of DNA sequence to generate music has a history in music composition for artistic purposes, and there is a developing literature on DNA sonification for research [9]. However, though DNA sonification examples for public engagement exist (e.g. [10]), they appear rare. Based on our preliminary activities, DNA sonification has potential for successful public engagement activities. Further efforts in this area are warranted. We report a brief analysis from public engagement events we have conducted that featured sonification. We also discuss some further directions for research on integration of sonification into public engagement.

## Methods

At Picademy in Glasgow on 29 November 2016, an initial sonification project was carried out. The coding sequence of GULO from mouse was converted to a Sonic Pi program, using search-and-replace to convert bases A, C, G and T to notes A, C, G and Midi 20 respectively. A filtered bass drum was added at the start of each codon (i.e. every three bases). Subsequent events used Raspberry Pi computers running the 4273 $\pi$  variant of GNU/Linux [11].

Subsequently we incorporated sonification into a series of public events as follows.

In preparation for Doors Open Day, a public drop-in event at the Ashworth Laboratories, King’s Buildings, University of Edinburgh on 23 September 2017 from 10 AM to 4 PM [12], further sonifications were prepared, using a Perl script we wrote to convert DNA sequence in Fasta format to a Sonic Pi program followed by manual editing of the result. The event brought in a range of different people, some potential new students and parents who were interested in biology, but also families with young children and other

local people. For this event, the sonification of “T” was changed, from the rather low Midi 20 to “E”. These additional sonifications consisted of: (1) a codon-free version of GULO; (2) a rave version of the GULO sonification, in which the 1:3 beat was switched with a 4:4 beat for musical effect; and (3) a sonification of part of an intron of the CFTR gene in human, which includes a dinucleotide microsatellite (AT), to highlight the potential research benefit of sonification. Additionally, a brief tutorial to Sonic Pi including DNA sonification exercises was written and printed out in laminated form, along with the GULO coding sequence and the start of the intron sequence from CFTR, with the microsatellites highlighted. Doors Open Day 2017 at the Ashworth Laboratories was attended by at least 258 people. During the event, we engaged ~30 visitors in a discussion about DNA, mainly by approaching them with the question “Have you ever wondered what DNA would sound like?”. The event was evaluated by means of direct experience and diaries of H.P. and D.B.

As part of Ada Lovelace Day, a free but ticketed public event at the James Clerk Maxwell Building, King’s Buildings, University of Edinburgh on 10 October 2017 [13], we gave an introductory presentation for ~50 participants, and offered three workshops later in the day, which were attended by a total of 13 people. The workshops lasted approximately 30-40 minutes each and consisted of a short DNA investigation using BLAST [14] with the GULO gene (as in [15]), followed by a guided exploration of prepared DNA sonifications, and the opportunity to go through the printed sonification tutorial. The event was evaluated by means of a short questionnaire and diaries of H.P. and D.B.

## Results And Discussion

Our preliminary study indicates that DNA sonification is highly suitable for public engagement activities, both for short drop-in events and for more focused workshops.

Doors Open Day was a large public event, during which people engaged well with the sonification activity. The question “Have you ever wondered what DNA would sound like” was a successful way of starting a discussion about DNA. Everyone who was asked this question stayed to listen to a sample of our sonifications and engaged in a short discussion about DNA. Visitors were guided through prepared sonifications, and given the opportunity to go through the printed sonification tutorial. Visitors found the event creative and inspiring and were either happy to see it as a light distraction, or to discuss scientific applications. Relatively little use was made of the Sonic Pi tutorial. During the event, Minecraft proved a distraction, so we discretely removed it from the main menu of each Raspberry Pi. Later, Scratch was also some distraction. For the future, we would remove these in advance because they are not part of our display. URL-based restrictions on Web browser access may also have some value. The Phylo game ran slowly at times on the Raspberry Pi computers (a problem that would be reduced or removed with newer models of the hardware).

The sonification workshops during Ada Lovelace Day were well-attended, and the audience appeared engaged. Six participants completed questionnaires. Others did not, due to leaving early for lunch (1 participant) and leaving at short notice to collect their bags from the main activity room (5 participants).

When asked what the best thing about the workshop was, participants stated “The opportunity to engage with DNA or Biology from a different perspective. Hearing a sequence made me think of it as an object in a different way.”; “Sonic Pi (amazing!)”; and “Updating me on sequence investigation”. When asked what the worst thing about the workshop was, participants did not have many comments, i.e. “None”, “Nothing”, “not enough time!”. We had difficulty connecting to the University WiFi network and used a 4G phone as a WiFi hotspot instead. However, this was resolved before participants arrived and was not noticeable to them. The event was a success. However, H.P. and D.B. thought a little more time to focus on sonification would be helpful for such events in the future.

It appears that DNA sonification engages both researchers and the public in thinking about DNA from a different perspective. Further development of DNA sonification for public engagement activities is warranted.

## LIMITATIONS

Although our preliminary study is sufficient to suggest DNA sonification has potential in public engagement activities, a larger study would be required to discover the full limits of its successful application, for example in terms of audience age, gender, ethnicity, academic background and in terms of event type. Our audience for evaluation consisted of adults, mostly staff and students at Edinburgh University. We recommend a larger program of qualitative and quantitative analysis, with a diverse range of audiences including young people.

In addition, our sonification has been relatively simple. Future directions involve codon or protein sonification. However, for public engagement, it will be important to avoid strong reliance on biological knowledge which the audience may lack. For example, multiple alignment sonification may be particularly valuable for research (Martin et al, in prep.). But multiple alignment depends on other biological concepts such as homology and cross-sequence comparison, unlikely to be rapidly comprehensible to the general public. The Phylo multiple alignment game [3] is a counter-example. However, as well as an exercise in multiple alignment, Phylo may be understood as a logic puzzle – with the benefit of a link to real data and disease – potentially bypassing these biological concepts for many users while leaving them as useful discussion points for those with expertise in the area.

One promising future direction for public engagement and education may be to highlight the effect of frameshift mutations. We have developed and apply a workshop for biology students in secondary education [15] (open educational resources available at [16]), centred on the GULO gene (e.g. [17]), which in humans is disrupted by frameshift mutations. Visually, a frameshift mutation is difficult to notice on-screen in BLASTN results comparing the GULO coding sequence and the human pseudogene. A frameshift is indicated by a single “-” among DNA symbols.

In conclusion, sonification has a demonstrated potential for public outreach and engagement. Our preliminary efforts to use this methodology were well-received. It is our aspiration to build on these

preliminary efforts to use sonification to make DNA sequence information entertaining as well as informative and to increase the nuance and complexity that we convey in public engagement.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

Sound files, a Perl script and Sonic Pi scripts supporting the conclusions of this article are available in the University of Edinburgh's DataShare repository, <https://hdl.handle.net/10283/3868>.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This work was funded by the Science and Technology Facilities Council (STFC) grant numbers ST/R000328/1 and ST/T000872/1. The funding body had no role in the design of the study or collection, analysis, or interpretation of data or in writing the manuscript.

### Authors' contributions

H.P. and D.B. designed and carried out the research and analysed the data. All authors wrote the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

We thank Craig Steele for work on the project at the Glasgow PiCademy event. We thank Edward Martin for comments on an earlier version of this manuscript.

## References

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the USA. 1977;12:5463-5467.
2. Andresen JS, Anderson J, Atencio E, Bergman BE, Bilofsky HS, Brown L, et al. Nucleotide sequences 1984: A compilation from the GenBank and EMBL data libraries. Parts 1 and 2. Oxford: IRL Press;

1984.

3. Kawrykow A, Roumanis G, Kam A, Kwak D, Leung C, Wu C, Zarour E, Phylo players, Sarmenta L, Blanchette M, Waldispühl J. Phylo: a citizen science approach for improving multiple sequence alignment. *PLoS One*. 2012;7:e31362.
4. McDonagh JL, Barker D, Alderson RG. Bringing computational science to the public. *SpringerPlus*. 2016;5:259.
5. Barker D, Alderson RG, McDonagh JL, Plaisier H, Comrie MM, Duncan L, Muirhead GTP, Sweeney SD. University-level practical activities in bioinformatics benefit voluntary groups of pupils in the last 2 years of school. *International Journal of STEM Education*. 2015;2:17.
6. Bain SA, Meagher TR, Barker D. Design, delivery and evaluation of a bioinformatics education workshop for 13-16-year-olds. *Journal of Biological Education*. 2020;doi:10.1080/00219266.2020.1858932
7. Aaron S. Sonic Pi – performance in education, technology and art. *International Journal of Performance Arts and Digital Media*. 2016;12:171-178.
8. Barrett N, Mair K. Aftershock: a science-art collaboration through sonification. *Organised Sound*. 2014;19:4-16.
9. Temple MD. An auditory display tool for DNA sequence analysis. *BMC Bioinformatics*. 2017;18:221.
10. Temple MD. Real-time audio and visual display of the Coronavirus genome. *BMC Bioinformatics*. 2020;21:431.
11. Barker D, Ferrier DEK, Holland PWH, Mitchell JBO, Plaisier H, Ritchie MG, Smart SD. 4273π: bioinformatics education on low cost ARM hardware. *BMC Bioinformatics*. 2013;14:243.
12. Doors Open Day 2017, The University of Edinburgh College of Science and Engineering. <https://www.ed.ac.uk/science-engineering/news-events/archive/2017/doors-open-day-2017> . Accessed 31 March 2021.
13. Frantzana A. Ada Lovelace Day 2017 – Celebrating women in STEM. <http://eusci.org.uk/2017/10/26/ada-lovelace-day-2017-celebrating-women-in-stem>. Accessed 31 March 2021.
14. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997;25:3389-3402.

15. Barker D, Plaisier H, Bain SA, Attwood TK. Bioinformatics: the Power of Computers in Biology – a Practical Guide. F1000Research. 2019;8:813
16. 4273p. <https://4273pi.org>. Accessed 31 March 2021.
17. Yang H. Conserved or lost: molecular evolution of the key gene *GULO* in vertebrate vitamin C biosynthesis. Biochemical Genetics. 2013;51:413-425.