

1 Sonification of genomic data to represent 2 genetic load in zoo populations

3

4 Edward J. Martin^{1,*}, Samuel A. Speak^{2a,b,c*}, Lara Urban^{3a,b,c}, Hernán E. Morales⁴, and Cock van
5 Oosterhout^{2a}

6

7 ¹ Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh,
8 Edinburgh, EH9 3FL, UK

9 ^{2a} School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich,
10 NR4 7TL, UK

11 ^{2b} Natural History Museum, London, SW7 5HD, UK

12 ^{2c} North of England Zoological Society, Chester Zoo, Chester, CH2 1LH, UK

13 ^{3a} Helmholtz AI, Helmholtz Zentrum Muenchen, Ingolstaedter Landstr. 1, 85764 Neuherberg,
14 Germany

15 ^{3b} Helmholtz Pioneer Campus, Helmholtz Zentrum Muenchen, Ingolstaedter Landstr. 1, 85764
16 Neuherberg, Germany

17 ^{3c} Technical University of Munich, School of Life Sciences, Alte Akademie 8, 85354 Freising,
18 Germany

19 ⁴ Center for Evolutionary Hologenomics, Globe Institute, Faculty of Health and Medical Sciences,
20 University of Copenhagen, DK-2200 Copenhagen, Denmark

21

22 * Joined first author

23 Corresponding authors: Edward J. Martin e.j.martin@sms.ed.ac.uk and Cock van Oosterhout [c.van-](mailto:c.van-oosterhout@uea.ac.uk)
24 oosterhout@uea.ac.uk.

25

26 Orcid ID:

27 Edward J. Martin: <https://orcid.org/0009-0000-4180-4900>

28 [Lara Urban: https://orcid.org/0000-0002-5445-9314](https://orcid.org/0000-0002-5445-9314)

29 [Cock van Oosterhout: https://orcid.org/0000-0002-5653-738X](https://orcid.org/0000-0002-5653-738X)

30 **Abstract**

31 Maintaining a diverse gene pool is important in the captive management of zoo populations,
32 especially in endangered species such as the pink pigeon (*Nesoenas mayeri*). However, due to the
33 limited number of breeding individuals and relaxed natural selection, the loss of variation and
34 accumulation of harmful variants is inevitable. Inbreeding results in a loss of fitness (i.e. inbreeding
35 depression), principally because related parents are more likely to transmit a copy of the same
36 recessive deleterious genetic variant to their offspring. Genomics-informed captive breeding can
37 manage harmful variants by artificial selection, reducing the genetic load by avoiding the inheritance
38 of two copies of the same harmful variant. To explain this concept in an interactive way to zoo
39 visitors, we developed a sonification game to represent the fitness impacts of harmful variants by
40 detuning notes in a familiar musical melody (i.e. Beethoven's *Für Elise*). Conceptually, zoo visitors
41 play a game aiming to create the most optimal pink pigeon offspring in terms of inbreeding
42 depression. They select virtual crosses between pink pigeon individuals and listen for the detuning of
43 the melody, which represents the realised load of the resultant offspring. Here we present the
44 sonification algorithm and the results of an online survey to see whether participants could identify
45 the most and least optimal offspring from three potential pink pigeon offspring. Of our 98
46 respondents, 85 (86.7%) correctly identified the least optimal offspring, 73 (74.5%) correctly
47 identified the most optimal, and 62 (63.3%) identified both the most and least optimal offspring using
48 only the sonification.

49 **Keywords**

50 Public Engagement, Genomics, Endangered Species, Pink Pigeon, Gamification

51

52 Introduction

53 Effective public engagement with research requires innovative approaches to communicate the often
54 complex concepts and vast amount of data generated in science today. In the UK, the public audience
55 tends to be disengaged with respect to science in general (Middleton et al. 2023). This is true for
56 genomics in particular, a jargon-rich and abstract branch of science which directly impacts many
57 lives: in the UK one in 17 will be affected by a rare disease, most probably with a strong genetic
58 component (HM Government 2020). Commonly, graphics are used to illustrate or conceptualise
59 patterns in big data, including genomic sequencing data. However, such figures are often created for
60 complex analytic tasks that each resist simplification (Rittenbruch et al. 2022), and can be difficult to
61 interpret for people who are not trained to read data presented in such a format. Alternatively, rather
62 than using visualisation, the information content of big data can be captured using sounds. By
63 mapping some feature of the data to a sound synthesis parameter, a parameter-mapping sonification
64 can facilitate understanding in the listener (Grond and Berger 2011). Previous work has seen success
65 in communicating genomics using sonification techniques, for analytical (Martin et al. 2021),
66 communication (Temple 2021), and public engagement purposes (Plaisier et al. 2021).

67

68 The role of zoos as sites of wildlife conservation is the basis of contemporary justification for their
69 continued existence (Carr and Cohen 2011). Zoos are keen to present conservation credentials in
70 detail to their public, whilst at the same time ensuring entertainment to attract sufficient visitors to
71 maintain economic viability. Sonification has shown potential in being used to create public
72 engagement activities which can convey scientific detail, while also achieving audience experiences
73 with ‘wow-factor’ (Ballora 2014).

74

75 Genomics techniques are increasingly used for wildlife conservation and management (Hohenlohe et
76 al. 2021). However, evolutionary concepts continue to be poorly integrated into conservation policies
77 and management practices (Cook and Sgrò 2019). The mathematical complexity of quantitative
78 genetics and population genetics, as well as the rapid advances of DNA sequencing technologies and
79 bioinformatics, form real barriers that hinder the integration of conservation science into applied

80 conservation (van Oosterhout 2021). With improved understanding of the causes of inbreeding
81 depression, practitioners will be better able to manage small, captive populations in zoos. Also, as
82 zoos are often public bodies, communicating the impacts of genetic management of zoo populations
83 to visitors and other stakeholders is important.

84

85 In this proof-of-concept study, we present a sonification algorithm to help listeners to interpret the
86 information that is present in a large quantity of DNA data using sound. In particular, our approach
87 aims to explain how harmful genetic variants (e.g. deleterious mutations) reduce the fitness of
88 individuals with parents that are related. We presented our sonification approach to an audience,
89 detuning the pitch of a well-known melody to communicate how genomic data can be used in
90 conservation management to reduce inbreeding depression.

91

92 The loss of fitness due to inbreeding depression is a function of both the level of inbreeding and the
93 genetic load of the individuals in the population (Crow 1970). The *genetic load* constitutes the
94 deleterious fitness effects of harmful genetic variants that individuals carry in their genome
95 (Bertorelle et al. 2022). The genetic load can be divided into two parts emphasising the effect of the
96 load, with the *realised load* describing that which reduces the fitness in the current generation
97 (Bertorelle et al. 2022). The importance of minimising inbreeding and maintaining genetic diversity is
98 well established in the conservation community (Hoban et al. 2013; DeWoody et al. 2021). Pedigree
99 data is widely used to manage captive populations to reduce the loss of variation and avoid close
100 inbreeding (i.e. consanguineous mating) (Ballou and Lacy 1995; López-Cortegano et al. 2021). It has
101 now also become possible to quantify the genetic load of individuals using genomics data and
102 bioinformatics techniques (Bertorelle et al. 2022). The mutation-impact scores of genetic variants can
103 be estimated, for example using Combined Annotation Dependent Depletion scores (CADD scores)
104 (see below). The quantification of the genetic load allows the impact of harmful mutations to be
105 compared between all individuals with sequenced genomes. These advances considerably expand the
106 possibilities for the management of captive-bred populations, potentially reducing the severity of
107 inbreeding depression (van Oosterhout 2020). Besides minimising the relatedness of individuals using

108 pedigree-calculated inbreeding coefficients, we can now establish whether individuals share harmful
109 variants at the same genetic loci. By avoiding crossings between individuals with such a shared
110 genetic load, it has become possible to reduce the realised load of the offspring caused by
111 homozygous harmful variants. Theoretically, this could reduce inbreeding depression and improve the
112 fitness of the offspring relative to that of a randomly mating zoo population, and zoo populations that
113 are managed using pedigree-data only (Speak et al. 2024). The analysis of the genetic load is
114 increasingly performed in studies that analyse the genomes of threatened species (reviewed in
115 Bertorelle et al. 2022), and they can help guide conservation management and predict the long-term
116 extinction risk (van Oosterhout 2020).

117

118 In a recent pilot study, which we use as the source of the data for our method, Speak et al. (2024)
119 analysed the genetic load of pink pigeons (*Nesoenas mayeri*). The species recovered from a severe
120 population bottleneck, prior to which an *ex situ* population was established (Jackson et al. 2022).
121 Captive-bred individuals were released into the wild population in Mauritius to increase population
122 size and genetic variation. This conservation rescue program was very successful and resulted in
123 population recovery (Jackson et al. 2022). Consequently, the pink pigeon was twice downlisted on the
124 International Union for the Conservation of Nature (IUCN) Red List. Nevertheless, concerns remain
125 because of its high genetic load (Jackson et al. 2022).

126

127 Our sonification approach is designed to enable zoo visitors to identify optimal mate pairs in zoo
128 populations that produce the fittest offspring. Our application is gamified in the sense that the zoo
129 visitors (i.e. players) are challenged finding the optimal crosses based on the minimal detuning in
130 Ludwig van Beethoven's composition *Für Elise*. Digital games are increasingly used in teaching the
131 value of biodiversity, and they have the potential to make a positive contribution to conservation
132 (Sandbrook et al. 2015). In our "game", deleterious genetic variants present in the genomic sequence
133 data of the offspring of two birds create detuning of the melody. The purpose of the game is to
134 minimise the detuning in the music and select the optimal breeding pairs to generate a viable
135 population.

136

137 **Material & Methods**

138 Our sonification application maps genomics data to sound in order to allow the information present in
139 the DNA to be interpreted through listening. The aspect of the genomic data that we are sonifying is
140 the *realised* load, the aspect of the genetic load that reduces fitness in the current generation. Briefly,
141 the potential fitness impacts of deleterious variants are first estimated, and these effects are then
142 separated into two components: the masked load and the realised load (Bertorelle et al. 2022). The
143 masked load is present only at heterozygous loci with harmful variants that are (partially) recessive.
144 This part of the load does not reduce the fitness of its carrier. In contrast, the realised load does reduce
145 fitness, and it is present at homozygous loci with harmful variants. In addition, some heterozygous
146 loci with partially dominant harmful variants also contribute to the realised load. Furthermore,
147 hemizygous loci with harmful variants in the sex chromosomes also add to the realised load, but only
148 in the heterogametic sex (i.e. males in mammals, and females in birds).

149

150 In small populations, the masked load becomes converted into a realised load due to inbreeding and
151 drift (Dussex et al. 2023). Genomics-informed captive breeding can help minimise the realised load
152 by crossing parents that do not share harmful variants at the same loci. To accomplish this, Speak et
153 al. (2024) calculated the expected realised load of potential offspring of all possible crossings between
154 the six pink pigeons that had been genome sequenced. Virtual crosses were produced, and the multi-
155 locus genotypes of the offspring was established assuming Mendelian segregation. The predicted
156 realised load of potential offspring is a proxy for the expected amount of inbreeding depression. This
157 approach enabled Speak et al. (2024) to identify optimal mate pairs that are expected to produce
158 offspring with the lowest realised load and the least amount of inbreeding depression.

159

160 The source data, and contextual information, for the method comes from the study by Speak et al.
161 (2024). Speak et al. (2024) used Combined Annotation Dependent Depletion (CADD) to assess the
162 genetic load in six pink pigeons. CADD ranks the severity of mutations across the genome, and the
163 CADD scores have been established for every possible nucleotide substitution in the genome of

164 model species such as humans, mice, and chicken. Speak et al. (2024) calculated the CADD scores for
165 603 genetic variants at 4929 ultra-conserved elements (UCEs) and their flanking regions in the six
166 pink pigeons. The pink pigeon reference genome consists of 2133 scaffolds, and Speak et al. (2024)
167 analysed the sequences of six birds, quantifying the genetic load in the 100 largest scaffolds (longest
168 scaffold = 38,883,656 bp and shortest = 5370 bp; median size = 5,486,524 bp). Together, these 100
169 scaffolds represent 68.96% of the pink pigeon reference genome. There are 595 UCEs across these
170 100 scaffolds that possess 658 single nucleotide polymorphisms (SNPs). These SNPs have relatively
171 high mutation-impact scores, estimated by Combined Annotation-Dependent Depletion (CADD)
172 scores. The realised load scores per scaffold, as measured by the summation of CADD scores within
173 each scaffold, were produced using R (R Core Team 2021), tidyverse (Wickham et al. 2019), and the
174 figures were produced using ggplot2 (Wickham 2016). For more detailed information about the
175 approach including visual information about the DNA sequences, see Speak et al. (2024).

176

177 In order to examine the genetic compatibility of individuals, we calculated the CADD scores of the
178 shared genetic load of deleterious variants between all different pairwise combinations of individuals
179 (Speak et al. 2024). In order to illustrate the effect of extreme inbreeding, we also included “selfing
180 crosses” where an individual was able to reproduce offspring by self-fertilisation or uniparental
181 reproduction. Offspring of parents that share many deleterious variants at the same genetic loci are
182 expected to inherit a high realised load of homozygous variants. Such offspring are likely to suffer
183 from severe inbreeding depression.

184

185 In order to sonify the load, we have developed an audio analogy relating the ‘wrongness’ of a familiar
186 melody with the risk of deleterious mutations in the genome of a potential pink pigeon offspring. The
187 realised load scores of an individual potential offspring, calculated by the summation of the CADD
188 scores across each scaffold, detune a well-known melody. The detuning of the pitch of a note
189 represents the realised load score of an individual potential offspring at a corresponding locus, as
190 calculated by the summation of CADD scores across a scaffold. This communicates the predicted
191 relative fitness loss of the offspring. An alphabetical ordering of the scaffolds was used and is an

192 artefact of the analysis. This ordering is consistent between different outputs and facilitates
193 comparison, with size or scaffold location unused alternatives. In order to detune a familiar western
194 classical melody, we used the MIDI (Musical Instrument Digital Interface) protocol as a method for
195 manipulating sound. MIDI is a language based on western classical music tonality that allows the
196 real-time control of sound synthesis and remains an industry standard. To create the effect of detuning
197 individual notes, we made use of the pitch-bend functionality of MIDI. However, rather than varying
198 the pitch of a note dynamically while it is played as in *glissando*, *portamento*, or the use of pitch-bend
199 wheels, we used the functionality to give detuned notes which were off-key but unvarying as the note
200 sounded, closer to the idea of *blue notes* in jazz or folk music. MIDI represents pitch-bend as a 14-bit
201 integer, allowing for 16,384 values with -8,192 representing maximum downward bend, 8,192
202 representing maximum upwards bend, and 0 representing no bend. We used the default maximum
203 bend as ± 2 semitones.

204

205 The realised load scores of our data range between 0 and 376. We made two transformations to these
206 to give MIDI pitch-bend values. First, we rounded them to the nearest factor of five to obtain integer
207 values and to not detune the note in the case of low genetic loads. Secondly, we multiplied these
208 values by -20 to obtain only negative values, representing downward bend or ‘detuning’ in our
209 analogy, and to render the detuning more distinguishable. The range of values between 0 to -7520
210 resulted in an appropriate level of ‘detuning’. We then used Python version 3.7.4 (van Rossum and
211 Drake 2009) and Jupyter Notebook version 6.0.1 (Kluyver et al. 2016) for data analysis. We used the
212 mido package (v1.2.6) to read and write MIDI data. We used the pandas package (v0.25.1) to read
213 our csv-formatted genomic data (sum of CADD scores of 100 largest scaffolds of 36 resultant crosses
214 from six birds), and also to create a list of strings to manipulate the MIDI messages. We used pygame
215 (v2.0.1) to play the output MIDI file in the Jupyter Notebook. We created wave file (.wav) examples
216 of the sonifications using Ableton 10 Live Suite.

217

218 The sonification algorithm requires a single-channel MIDI track as the basis for detuned sonification
219 outputs. We used MuseScore 3 (v3.6.2) to transcribe a version of the well-known Ludwig van

220 Beethoven composition *Für Elise* (see Supplementary Materials). The version we created includes
221 104 notes, of which we used 100 notes to be detuned according to the realised load scores of the 100
222 largest pink pigeon genome scaffolds. We opted to allow the first four notes of the melody to have no
223 detuning – this gives the listener a brief frame of reference to aid in the perception of detuning
224 throughout the rest of the sonification.

225

226 Programmatically, the algorithm uses `mido` to read the MIDI files. We then parse them into a list of
227 strings so that we can use string methods to edit the MIDI messages. A loop, indexed through the list,
228 alters the data by adding a MIDI *pitchbend* message preceding each *note_on* message which indicates
229 the onset of a note. This MIDI *pitchbend* parameter is set to the value of the i^{th} scaffold realised load
230 score of the parental selection. It is worth noting that *note_on* messages are also used to end notes by
231 setting the velocity parameter to 0, so our algorithm checks for a non-zero velocity of the *note_on*
232 message. The algorithm also ignores the first four note onset messages, leaving an initial run of four
233 unchanged notes at the start of each output sonification. After the completion of this loop, the list of
234 strings is parsed by means of another loop so `mido` can save it as a new MIDI file. The algorithm is
235 available as a Jupyter notebook at <https://github.com/sonified/pinkpigeons>.

236

237 To test whether our sonification algorithm was successful in communicating genetic load, we created
238 a short experiment via a JISC online survey (see supplementary materials). Participants were given
239 access to audio recordings of the sonification representations of three pink pigeon potential offspring:
240 ‘Offspring A’ (<https://soundcloud.com/sonified/pink-pigeons-offspring-a>), ‘Offspring B’
241 (<https://soundcloud.com/sonified/pink-pigeons-offspring-b>), and ‘Offspring C’
242 (<https://soundcloud.com/sonified/pink-pigeons-offspring-c>). Participants were also given access to a
243 version of *Für Elise* without any detuning for comparison (<https://soundcloud.com/sonified/fur-elise>).

244

245 After a short introduction, providing some context about the pink pigeon and some explanation of
246 how to listen to the sonifications, participants were asked *By listening to the sound, can you determine*
247 *which is the most optimal offspring with the least detuning of the melody, and which the least optimal*

248 *offspring with the most detuning of the melody?*. This gave us a ranking of the three offspring from
249 most to least optimal for each participant. We also offered a free text box prompted by the question
250 *Do you have any comments you wish to share?*. A copy of the survey can be found in the
251 supplementary materials. Our participants were principally recruited via mailing lists at the University
252 of Edinburgh. The survey was subject to ethics approval from the University of Edinburgh's SBS
253 Ethics Committee with reference dbarker-0004.

254

255 **Results**

256 Figure 1 illustrates the realised load across the 100 largest scaffolds in offspring produced by crossing
257 different combinations of pink pigeons. When an individual is crossed with itself, the realised load of
258 the offspring is relatively elevated, as shown in the raised peaks in Figure 1, panel i. For example, the
259 realised load of the offspring produced by the selfing cross PP1 x PP1 (i.e. 'Offspring A'
260 <https://soundcloud.com/sonified/pink-pigeons-offspring-a>) is significantly higher compared to that of
261 an outbred crosses PP2 x PP3 (i.e. 'Offspring B' [https://soundcloud.com/sonified/pink-pigeons-](https://soundcloud.com/sonified/pink-pigeons-offspring-b)
262 [offspring-b](https://soundcloud.com/sonified/pink-pigeons-offspring-b)) (T-test: $T=5.60$, $d.f.=1225$, $p<0.0001$). Due to the small effective population size of the
263 captive pink pigeon population, some individuals (e.g., PP2 and PP3) share a high genetic load.
264 Crosses between such individuals produce offspring with a significantly higher realised load (Figure
265 1, panel i; <https://soundcloud.com/sonified/pink-pigeons-offspring-b>) than crosses such as PP2 x PP6
266 (i.e. 'Offspring C' <https://soundcloud.com/sonified/pink-pigeons-offspring-c>) whose parents share
267 relatively few deleterious variants (Figure 1, panel iii) (T-test: $T=2.68$, $d.f.=1254$, $p<0.0075$).
268 Offspring of the latter cross (PP2 x PP6) are expected to have a higher fitness, and this is reflected by
269 the lower peaks in Figure 1, panel iii, and the relatively low level of detuning of the melody in the
270 corresponding soundtrack.

271

272 In the sonification experiment, and based on the answers of 98 respondents, the least optimal cross
273 resulting in offspring with the highest realised load was correctly identified in 85 out of 98 cases
274 (86.7%). The most optimal cross was correctly identified in 73 out of 98 cases (74.5%). Both the most
275 optimal and the least optimal were identified in 62 out of 98 responses (63.3%). These responses were

276 significantly better than random, in a binomial test taking the null hypothesis of a random trial with
277 $\pi \leq \frac{1}{6}$, $x = 62$, & $n = 98$: $p < 10^{-26}$. This provides evidence that our sonification method is
278 successful in communicating the fitness of potential pink pigeon offspring when compared to random
279 guessing.

280

281 The free text box prompted 18 responses (18.4% of participants). Six people stated that the task was
282 difficult with two of them citing auditory impairments as a cause, five expressed that they found joy
283 in the task, five stated an opinion that a pair were difficult to distinguish (four said A and C were very
284 similar, and one said A and B), and two people made jokes pitying how bad the inbreeding of B must
285 be. Of the six people (33.3% of free text responses) that cited the difficulty of the task, five of them
286 got the correct answers for both the most and least optimal. Out of this group, the only one to give
287 incorrect answers had identified themselves as having auditory impairment. Of the five that stated that
288 a pair were difficult to distinguish, four gave the correct answers for both most and least optimal.
289 Notably no-one said that B and C were difficult to distinguish, the pair with the largest difference.

290

291 **Discussion**

292 In this proof-of-concept study, we used sonification to explain to zoo visitors how genomic data can
293 be used to optimise the fitness of offspring by identifying parents that share little genetic load at the
294 same loci. Speak et al. (2024) quantified the fitness impacts of harmful variants (i.e. deleterious
295 mutations) across the genome using Combined Annotation-Dependent Depletion (CADD) scores,
296 studying the genome data of six pink pigeons. The study made virtual crosses between these
297 individuals and calculated the realised load of the potential offspring, assuming Mendelian
298 segregation of alleles (see Speak et al. 2024). The offspring of parents that share recessive deleterious
299 variants at the same genetic loci can inherit two copies of the same variant, which increases their
300 realised load and may result in inbreeding depression (Bertorelle et al. 2022).

301

302 Genomic data is likely to play an increasingly important role in maintaining the long-term viability of
303 zoo populations (Norman et al., 2019; van Oosterhout 2020). The efficacy of natural selection to
304 purge the genetic load might be compromised in zoos due to a variety of factors. First, the small
305 breeding population size reduces the effective population size, which increases genetic drift, thereby
306 reducing the efficacy of natural selection (Dussex et al. 2023). Second, the relatively benign
307 environmental conditions in zoos can interfere with hard selection by allowing individuals that might
308 have otherwise died in the natural environment to survive and reproduce in captivity (Armbruster and
309 Reed 2005). Third, in pedigreed zoo populations, conservation management aims to breed from
310 individuals that are genetically under-represented in the population to maximise founder
311 representation and founder genome equivalents (Lacy 1989). Although this helps to conserve genetic
312 diversity, it could also interfere with natural selection and the purging of the genetic load that may
313 occur in nature. Genomic analysis can detect harmful variants and reduce their impact by avoiding
314 crosses of individuals that share the same variants. The purpose of this study is to explain this concept
315 to laypersons, in particular, zoo visitors.

316

317 Our results show success compared to random guessing, and give evidence that the sonification
318 method successfully communicates the genetic load of the offspring. Interestingly, respondents who
319 gave feedback that the task was difficult proved to be more successful in completing the task than
320 those who gave no feedback. Perhaps this reflects that the task feels difficult, even while people show
321 quite a high success rate. This is an interesting future idea to explore, how a challenging public
322 engagement task is perceived by its audience impacts the effect it has on them.

323

324 Explaining the value of conservation genomics to the public in an understandable and playful manner
325 can make a meaningful contribution to conservation teaching. Future research directions could
326 involve the development of an *in-situ* installation at a zoo to investigate the impact of the idea in its
327 intended context. The future implementation of our sonification algorithm in a zoo context as a public
328 engagement game may have multiple impacts: raising awareness of genomics research and research
329 findings, meeting the conservation learning mission of zoos, mobilising changes in behaviour, and

330 also providing feedback to improve the public engagement research approach. Also the sound design
331 of the sonification could be developed to make a more emotionally compelling experience. As far as
332 we are aware, the approach to sonification by detuning the notes in a familiar melody to represent
333 ‘wrongness’ or ‘distance from an ideal outcome’ is a novel one, and many new applications of this
334 idea in different fields can be imagined.

335

336 Charles Darwin stated in *On the Origin of Species* “pigeons have been watched, and tended with the
337 utmost care, and loved by many people” (Darwin 1859). We hope our sonification work highlights
338 and supports the continuation of this tradition in the genomics era.

339

340 **Abbreviations**

341 CADD – Combined Annotation-Dependent Depletion

342 CSV – Comma Separated Values

343 DNA – Deoxyribonucleic Acid

344 MIDI – Musical Instrument Digital Interface

345 SNP - Single Nucleotide Polymorphism

346 UCE – Ultra-conserved Elements

347

348 **Data availability**

349 All supplementary materials including our algorithm, input midi file, example questionnaire, and the
350 data used to create our sonifications, are made available as a Jupyter Notebook available at

351 <https://github.com/sonified/pinkpigeons>.

352

353 **Authorship**

354 EJM, SAS, and CvO conceived the study and drafted the MS. EJM created the sonification algorithm,
355 sound files, questionnaire, and analysed results of the questionnaire. SAS produced the figure, and
356 CvO performed the statistical tests. All authors contributed intellectually to the study, and commented
357 and edited the manuscript.

358

359 Acknowledgements

360 Thanks to Daniel Barker and Thomas R. Meagher for feedback on the manuscript, and to Richard
361 Fitzpatrick and James Lamb for feedback on the sonification approach. EJM was funded by East of
362 Scotland Bioscience Doctoral Training Partnership (EASTBIO) funded by UKRI Biotechnology and
363 Biological Sciences Research Council (BBSRC) Grant Number BB/M010996/1. CvO was funded by
364 the Royal Society International Collaboration Awards (ICA\R1\201194) and the Earth and Life
365 Systems Alliance (ELSA), SAS was funded by a NERC ARIES PhD studentship (T209447) at the
366 UEA and a Research Training Support Grant (RTSG; 100162318RA1). HM was funded by an EMBO
367 fellowship (grant 1111–2018) and the European Union's Horizon 2020 research and innovation
368 program under Marie Skłodowska-Curie (grant 840519). For the purpose of open access, the author
369 has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript
370 version arising from this submission.

371

372 Ethics

373 This research was approved by the Ethics Committee at the School of Biological Sciences, University
374 of Edinburgh (Reference number dbarker-0004). All methods were carried out in accordance with
375 relevant guidelines and regulations. Regarding the genomic data, publicly available DNA sequencing
376 data were analysed, and no animals were used in this research. The data used is available at

377 <https://github.com/sonified/pinkpigeons>

378

379 Conflict of interest statement

380 We declare no conflict of interests.

381

382 References

383 Armbruster, P. and Reed, D.H., 2005. "Inbreeding depression in benign and stressful environments."
384 *Heredity*, 95(3):235-242. <https://doi.org/10.1038/sj.hdy.6800721>

385

- 386 Ballora, M. 2014. "Sonification, science and popular music: In search of the 'wow'." *Organised*
387 *sound* 19(1):30-40. <https://doi.org/10.1017/S1355771813000381>
388
- 389 Ballou, J., and Lacy, R. C., 1995. "Identifying genetically important individuals for management of
390 genetic variation in pedigreed populations." In *Population management for survival and recovery*,
391 edited by J. D Ballou, 76– 111. Columbia University Press.
392
- 393 Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H.E. and van
394 Oosterhout, C., 2022. "Genetic load: genomic estimates and applications in non-model
395 animals." *Nature Reviews Genetics*, 23(8):492-503. <https://doi.org/10.1038/s41576-022-00448-x>
396
- 397 Carr, N. and Cohen, S., 2011. "The public face of zoos: Images of entertainment, education and
398 conservation." *Anthrozoös*, 24(2):175-189. <https://doi.org/10.2752/175303711X12998632257620>
399
- 400 Cook, C.N. and Sgrò, C.M., 2018. "Poor understanding of evolutionary theory is a barrier to effective
401 conservation management." *Conservation Letters*, 12(2):e12619. <https://doi.org/10.1111/conl.12619>
402
- 403 Crow, J.F., 1970. "Genetic loads and the cost of natural selection". In: Kojima, Ki. (eds)
404 *Mathematical Topics in Population Genetics*. Biomathematics, vol 1. Springer, Berlin, Heidelberg.
405 https://doi.org/10.1007/978-3-642-46244-3_5
- 406 Darwin, Charles, 1859. "Variation under domestication." In *On the Origin of Species*, 7-43. John
407 Murray, London.
408
- 409 DeWoody, J.A., Harder, A.M., Mathur, S. and Willoughby, J.R., 2021. "The long-standing
410 significance of genetic diversity in conservation." *Molecular Ecology*, 30(17):4147-4154.
411 <https://doi.org/10.1111/mec.16051>
412

- 413 Dussex, N., Morales, H.E., Grossen, C., Dalén, L. and van Oosterhout, C., 2023. “Purging and
414 accumulation of genetic load in conservation.” *Trends in Ecology & Evolution*, 38(10):961-969.
415 <https://doi.org/10.1016/j.tree.2023.05.008>
416
- 417 Grond, F. and Berger, J., 2011. “Parameter mapping sonification.” In *The sonification handbook*,
418 edited by Thomas Hermann, Andy Hunt, and John G. Neuhoff, 363-398. Logos, Berlin.
419
- 420 HM Government, 2020. “Genome UK – the future of healthcare.”
421
- 422 Hoban, S.M., Hauffe, H.C., Pérez-Espona, S., Arntzen, J.W., Bertorelle, G., Bryja, J., Frith, K.,
423 Gaggiotti, O.E., Galbusera, P., Godoy, J.A. and Hoelzel, A.R., 2013. “Bringing genetic diversity to
424 the forefront of conservation policy and management.” *Conservation Genetics Resources*, 5:593-598.
425 <https://doi.org/10.1007/s12686-013-9859-y>
426
- 427 Hohenlohe, P.A., Funk, W.C., and Rajora, O.P., 2021. “Population genomics for wildlife conservation
428 and management.” *Molecular Ecology* 30(1): 62-82. <https://doi.org/10.1111/mec.15720>
429
- 430 Jackson, H.A., Percival-Alwyn, L., Ryan, C., Albeshr, M.F., Venturi, L., Morales, H.E., Mathers,
431 T.C., Cocker, J., Speak, S.A., Accinelli, G.G., Barker, T., ... van Oosterhout, C., 2022. “Genomic
432 erosion in a demographically recovered bird species during conservation rescue.” *Conservation*
433 *Biology*, 36(4): p.e13918. <https://doi.org/10.1111/cobi.13918>
434
- 435 Kluyver, T. et al., 2016. “Jupyter Notebooks – a publishing format for reproducible computational
436 workflows.” In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited
437 by F. Loizides & B. Schmidt, 87–90. IOS Press BV
438
- 439 Lacy R.C., 1989. “Analysis of founder representation in pedigrees: founder equivalents and founder
440 genome equivalents.” *Zoo Biology*. 8(2):111-23. <https://doi.org/10.1002/zoo.1430080203>

441

442 López-Cortegano, E., Moreno, E. and García-Dorado, A., 2021. “Genetic purging in captive
443 endangered ungulates with extremely low effective population sizes.” *Heredity*, 127(5):433-442.

444 <https://doi.org/10.1038/s41437-021-00473-2>

445

446 Martin, E.J., Meagher, T.R. and Barker, D., 2021. “Using sound to understand protein sequence data:
447 new sonification algorithms for protein sequences and multiple sequence alignments.” *BMC*

448 *Bioinformatics* 22:456. <https://doi.org/10.1186/s12859-021-04362-7>

449

450 Middleton, A., Adams, A., Aidid, H., Atutornu, J., Boraschi, D., Borra, J., Bircan, T., Burch, C.,
451 Costa, A., Dickinson, A., Enticknap, A., Galloway, C., Gale, F., Garlick, E., Haydon, E., Henriques,
452 S., Mitchell, M., Milne, R., Monaghan, J., Morley, K. I., ... Parkinson, S., 2023. “Public engagement
453 with genomics.” *Wellcome open research*, 8:310. <https://doi.org/10.12688/wellcomeopenres.19473.2>

454

455 Norman A.J., Putnam A.S., and Ivy J.A., 2019. “Use of molecular data in zoo and aquarium collection
456 management: benefits, challenges, and best practices.” *Zoo Biology*, 38(1):106-118.

457 <https://doi.org/10.1002/zoo.21451>

458

459 R Core Team, 2021. *R: A language and environment for statistical computing*. R Foundation for
460 Statistical Computing, Vienna, Austria.

461

462 Rittenbruch, M. et al., 2022. “Collaborative sense-making in genomic research: The role of
463 visualisation.” *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4477-4489.

464 <https://doi.org/10.1109/TVCG.2021.3090746>

465

466 Plaisier, H., Meagher, T.R. and Barker, D., 2021. “DNA sonification for public engagement in
467 bioinformatics.” *BMC Res Notes*, 14:273. <https://doi.org/10.1186/s13104-021-05685-7>

468

- 469 Sandbrook, C., Adams, W.M. and Monteferri, B., 2015. "Digital games and biodiversity
470 conservation." *Conservation Letters*, 8(2):118-124. <https://doi.org/10.1111/conl.12113>
471
- 472 Speak, S., Birley, T., Bortoluzzi, C., Clark, M., Percival-Alwyn, L., Morales, H. and van Oosterhout,
473 C., 2024. "Genomics-informed captive breeding can reduce inbreeding depression and the genetic
474 load in zoo populations." *Molecular Ecology Resources*, 2024;00:e13967.
475 <https://doi.org/10.1111/1755-0998.13967>
476
- 477 Temple, M.D., 2021. "Real-time audio and visual display of the Coronavirus genome." *BMC*
478 *Bioinformatics* 21:431. <https://doi.org/10.1186/s12859-020-03760-7>
479
- 480 van Oosterhout, C., 2020. "Mutation load is the spectre of species conservation." *Nature Ecology &*
481 *Evolution*, 4(8):1004-1006. <https://doi.org/10.1038/s41559-020-1204-8>
482
- 483 van Oosterhout, C., 2021. "Conservation genetics: 50 Years and counting." *Conservation Letters*,
484 14(2):e12789. <https://doi.org/10.1111/conl.12789>
485
- 486 van Rossum, G. & Drake, F.L., 2009. *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA.
487
- 488 Wickham, H., 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York.
489
- 490 Wickham H., Averick M., Bryan J., Chang W., McGowan L.D., François R., Grolemund G., Hayes
491 A., Henry L., Hester J., Kuhn M., Pedersen T.L., Miller E., Bache S.M., Müller K., Ooms J.,
492 Robinson D., Seidel D.P., Spinu V., Takahashi K., Vaughan D., Wilke C., Woo K., Yutani H., 2019.
493 "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43):1686.
494 <https://doi.org/10.21105/joss.01686>
495

496 **Figure Legends**

497 **Table 1. Survey results and sum of CADD scores across 100 largest scaffolds.** A total of 98
498 responses are listed here. The CADD scores listed are the sum of CADD scores for the largest 100
499 scaffolds for each potential offspring, and they provide an estimate of the realised load expected in the
500 offspring.

501

502 **Figure 1. The realised load across the 100 largest scaffolds in offspring produced by crossing**
503 **different combinations of pink pigeons.** Panel i shows a selfing-cross PP1 x PP1 that results in a
504 very high realised load ('Offspring B' with corresponding sonification:
505 <https://soundcloud.com/sonified/pink-pigeons-offspring-b>). Panel ii is cross PP2 x PP3, which also
506 produces offspring with a high realised load ('Offspring A' with corresponding sonification:
507 <https://soundcloud.com/sonified/pink-pigeons-offspring-a>). Panel iii is cross PP2 x PP6, which
508 resulted in offspring with a low realised load ('Offspring C' with corresponding sonification:
509 <https://soundcloud.com/sonified/pink-pigeons-offspring-c>). The scaffolds were ordered alphabetically
510 by their name, which does not correspond to size or location on the genome.