JOURNAL OF MICROBIOLOGY & BIOLOGY EDUCATION, March 2016, p. 129-132
DOI: http://dx.doi.org/10.1128/jmbe.v17i1.1028

More of an Art than a Science: Using Microbial DNA Sequences to Compose Music †

Peter E. Larsen
Biosciences Division, Argonne National Laboratory, Argonne, IL 60439

Bacteria are everywhere. Microbial ecology is emerging as a critical field for understanding the relationships between these ubiquitous bacterial communities, the environment, and human health. Next generation DNA sequencing technology provides us a powerful tool to indirectly observe the communities by sequencing and analyzing all of the bacterial DNA present in an environment. The results of the DNA sequencing experiments can generate gigabytes to terabytes of information, however, making it difficult for the citizen scientist to grasp and the educator to convey this data. Here, we present a method for interpreting massive amounts of microbial ecology data as musical performances, easily generated on any computer and using only commonly available or freely available software and the 'Microbial Bebop' algorithm. Using this approach, citizen scientists and biology educators can sonify complex data in a fun and interactive format, making it easier to communicate both the importance and the excitement of exploring the planet earth's largest ecosystem.

INTRODUCTION

It is our great privilege to live in a world teeming with life. We have discovered, however, that the great majority of living things on Earth are largely invisible to us. Microscopic bacteria are actually the dominant form of life on this planet (14). Communities of bacteria are found in nearly every earthly environment, from soil (7, 16) to sea (9), from the roots of trees (3, 12) to within our bodies (9), from deep underground (4, 13) to literally living among the clouds (2). These bacterial communities participate in every biogeochemical cycle on Earth and have a direct impact on our climate (1) and our health (10, 15). Recently, the use of ultra-high throughput DNA sequencing technology has opened new opportunities for investigating microbial ecology (5). As the majority of bacteria cannot be easily cultivated in the laboratory for direct experimental observations, DNA sequencing technology has made it possible to observe them indirectly, through analysis of their DNA or RNA extracted from an environmental sample. The effort to analyze this data is truly multidisciplinary, invoking molecular biology, ecology, bioinformatics, computational modeling, and statistics.

Finding a way to make these invisible but beautiful microscopic worlds accessible and exciting to nonspecialists can be a challenge. Modern microbial ecology is dependent on next-generation sequencing technologies and advanced computational analysis, and the scientific literature can be dense with requisite jargon. While the specialist might find beauty in the carefully pruned phylogenetic tree or insight in the principal component analysis, the nonspecialist may be left cold to the potential wonders of microbial ecosystems. Another difficultly in connecting modern microbial ecology to the citizen scientist is access to next generation molecular sequencing data and advanced computing capacity. Where the activities of microbiology and biology require increasingly powerful tools for data acquisition, what is the role of the citizen scientist? While it may be possible for citizen scientists to comb through images of distant Mars (http://beamartian.jpl.nasa.gov), track honeybee populations in their backyards (http://beespotter.org/), or help solve computationally intensive protein folding problems (http://fold.it/portal/), the opportunities for the citizen scientist to explore the microbial communities of their own world or within their own bodies are scarce. Of course, such citizen scientists can offer themselves up as the subjects of scientific inquiry (e.g., http://americangut.org/), but how then can they be made into more active participants? Data analysis is the fulcrum about which the citizen scientist might find the leverage to enter the field of microbial ecology, but the learning curve can be steep and computational requirements costly.

As a part of my role as a microbial ecologist, I endeavor to make the invisible visible. Sometimes, I also try to make the invisible audible. To the set of scientific disciplines available for microbial ecological analysis, art and music theory have now been added. In an effort to link microbial ecology data interaction with an interface that is fun and

Corresponding author. Mailing address: 9700 South Cass Avenue, Argonne, IL 60439. Phone: 630-525-3984. E-mail: plarsen@anl.gov.
†Supplemental materials available at http://jmbe.asm.org

©2016 Author(s). Published by the American Society for Microbiology. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial-NoDerivatives 4.0 International license (https://creativecommons.org/licenses/by-nc-nd/4.0/ and https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode), which grants the public the nonexclusive right to copy, distribute, or display the published work.

Downloaded from www.asmscience.org by
IP: 146.137.70.70
On: Sat, 02 Apr 2016 00:22:31

appealing to non-experts, we have previously presented a computational approach, Microbial Bebop, for the sonification of complex microbial ecology data, by which data are transformed into improvisational jazz-like compositions (8). In this manuscript, the algorithm behind Microbial Bebop is presented as a fun, easily accessed and manipulated tool, using commonly and freely available software, for educators and citizen scientists.

PROCEDURE

The following is a protocol for educators and citizen scientist to transform potentially large datasets into musical compositions. The "Microbial Bebop" algorithm translates data into music using a computational method loosely inspired by improvisational jazz (stretching the definitions of "jazz," and perhaps even "music," to their absolute limit (11)). In a Microbial Bebop composition, the melody represents six to eight elements, out of potentially hundreds or thousands, from a biological dataset, the chord progression another element, and the pattern of note durations within a measure yet another element. The action of the Microbial Bebop algorithm is to synthesize those elements into a composition that obeys some of the dictates of music: melody, rhythm, and harmony. While a single musical measure represents one single observation, a complete composition is generated from all of the observations collected in an experiment. Owing to the vast amounts of data collected in an experiment, a single microbial ecology dataset can be amenable to a near infinite number of possible musical interpretations.

To follow this procedure, it is necessary to have the "Bebopilizer," a Microsoft Excel document available in Appendix 1, and ImproVisor (www.cs.hmc.edu/~keller/jazz/improvisor/), a free, open-source music notation program that can play back the compositions generated by the Bebopilizer. The Bebopilizer Microsoft Excel spreadsheet is comprised of multiple worksheets for users to enter their data and manipulate the parameters for generating compositions (Fig. 1).

The user's tabular data set is pasted into the Data worksheet (Fig. 1A). An example dataset has been provided in Appendix 1. The data must be formatted such that columns are individual observations and rows are the data types collected for each observation. While the input can be of nearly any number of columns, a number of observations between around eight and twenty-five may be best for a composition of a tolerable length. However many rows are in the data, the user will select between eight and ten for inclusion in any particular composition.

The worksheet Bebopilizer contains the setting options for generating the composition. The selections are:

- **Music Type:** This allows the user to select the meter and number of notes/data points per measure (Fig. 1B). "Jazz" is in 4/4 time with six notes per measure. "Blues" is in 12/8 time with eight notes per measure. "Waltz" is in 3/4 time with six notes per observation, played over two measures.
- There are two options for generating chords (Fig. 1C).
  - **Select data row:** This allows the user to select which data type, identified by its row number from worksheet Data, will be used to generate the composition's chords.
  - **Chord selection:** This allows the user to select the possible specific chords that will be used in the composition. Users can add additional chord selections of their own by adding them to the list of chords on worksheet Data_Chords.
- There are multiple options to select for melody generation (Fig. 1D).
  - **Range:** This allows the user to define the range of notes used to generate the melody, up to a value of 18 (two and a half octaves). The notes used are from the octatonic scale, but the user can change the intervals by modifying the list of notes on worksheet Data_Melody, column C.
  - **Row - Data 1 to 8:** This allows the user to select the data, identified by their row numbers from worksheet Data, which will be used to generate the melody. Notes 7 and 8 are only used when the Music Type is set to "Blues."
  - **Row - Notes:** The user can map a data row to the pattern of note durations and rests played during a single measure. The ambitious user can alter the possible patterns of note durations by making changes to worksheet Data_Melody, but care must be taken in modifications if the melody is to remain in sync with the chords in each measure.
- **Generate Melody:** Pressing this button (Fig. 1E) will generate a new Excel file that contains the necessary data to generate a melody using the program ImproVisor.

While the instructions here are enough to generate and play a new composition quickly, it is beyond the scope of this manuscript to provide complete instructions for use of ImproVisor. The reader will be well rewarded by becoming at least moderately familiar with the ImproVisor interface and user manual (www.cs.hmc.edu/~keller/jazz/improvisor/tutorials.html). Briefly, the number of measures and the time signature in ImproVisor need to be set to match the output for "Bebopilizer," i.e., 4/4, 12/8, or 3/4 time (Fig. 2A). Into the ImproVisor 'Textual Entry' bar (Fig. 2B), sequentially copy and paste the Chords and the Melody. The 'Rectify Melody to Chords' option in ImproVisor must then be selected (under Edit toolbar option, or else Shift-R). Not only will rectifying melody to chords produce compositions free of excessive discordant notes, but it will also uniquely modify the melody to the particular set of chords. Melodies generated with the

Downloaded from www.asmscience.org by
IP: 146.137.70.70
On: Sat, 02 Apr 2016 00:22:31

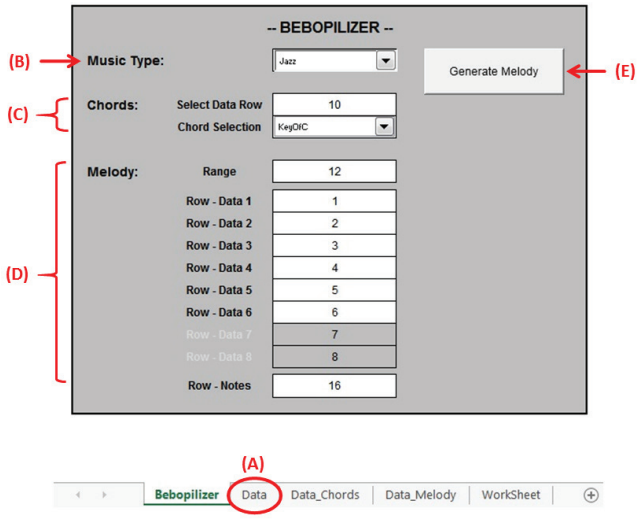

FIGURE 1. Bebopilizer User Form. Highlights A–E reference text from the Bebopilizer Procedure section.

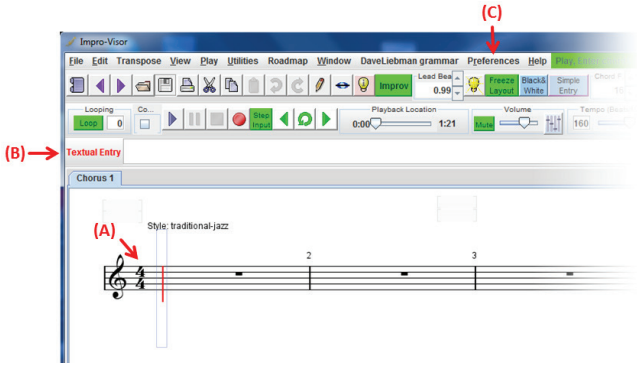

FIGURE 2. ImproVisor Screenshot. Highlights A–C reference text from the ImproVisor Methods section.

same data will be different when played with chords generated from an alternate data selection.

Once this has been accomplished, the full range of ImproVisor options for selection of tempo, musical instruments, and rhythm types can be modified by the user (Preferences (Fig. 2C) is a good place to start with this). ImproVisor provides the opportunity to output the composition as sheet music PDFs or as MIDI music files. A variety of free web-based tools is available (e.g., HammieNet at www.hamienet.com/midi2mp3) for converting MIDI files into mp3 files.

A series of Microbial Bebop examples is provided (Table 1) to give the user a sampling of the kinds of data that can be used and the sorts of music that can be made. The data are as diverse as a decade of marine microbiology observations to daily measurements of a human's symbiotic bacteria. The performers range from a laptop computer to the children of a primary school choir. For the reader and potential novice explorer of microbial ecosystems, there are two excellent sources of available data for analysis that may be easily used: MG-RAST (http://metagenomics.anl.gov/) and IMG/M (https://img.jgi.doe.gov/cgi-bin/edu/main.cgi).

CONCLUSION

It is unlikely that the sonification of data will ever overtake the humble bar graph as the preferred method for easily communicating complex data. Likewise, I think it improbable that listening to a specific data sonification alone will ever lead to greater insights into that data (although I would be very happy to be proven wrong in this case). There are however, overlaps between the skills needed for manipulating large datasets for use in data sonification and data analysis that can be identified and put into practice by the science educator. For example, there are several ways the initial data can be considered. Log transformation, relative ratios, or rate of change are all ways that data can be precomputed before being entered into the Bebopilizer. While it is possible to select data at random for inclusion into the composition, data selection provides the opportunity to consider what variables are expected to interact within the dataset, or to identify data types with patterns that correlate, positively or negatively, with other data types. The activities associated with this sort of analysis of the data may lead to additional insights or to the mastery of the skills required to manipulate large datasets. For the purpose of education, outreach, and citizen science, it is the act of generating the composition, selecting the components, and thinking about how those components might interact, not the composition itself, which provides the opportunity to interact meaningfully with the data. Most importantly however, this sort of analysis brings the opportunity to have fun while discussing and analyzing data, by generating musical compositions that can be played and shared. And of course, if nothing is gained other than the appreciation of the beauty that underlies these important and pervasive bacterial communities, that will be more than enough.

SUPPLEMENTAL MATERIALS

Appendix 1: Microsoft Excel spreadsheet Bebopilizer

ACKNOWLEDGMENTS

This work was supported by the United States Department of Energy under Contract DE-AC02-06CH11357. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The author declares that there are no conflicts of interest.

REFERENCES

1. **Bardgett, R. D., C. Freeman, and N. J. Ostle.** 2008. Microbial contributions to climate change through carbon cycle feedbacks. ISME J. **2**(8):805–814.

Downloaded from www.asmscience.org by
IP: 146.137.70.70
On: Sat, 02 Apr 2016 00:22:31

TABLE 1.
Examples of Microbial Bebop Compositions.

Microbial Bebop Composition	Link	Data Analysis
Blues for Elle	https://www.youtube.com/watch?v=EwZJj2k61ns	This composition highlights seasonal patterns in marine physical parameters at the L4 Station in the Western English Channel. The chords are generated from seasonal changes in photosynthetically active radiation. The melody of each measure is comprised of eight notes, each mapped to a physical environmental parameter, in the following order: temperature, soluble reactive phosphate, nitrate, nitrite, saline, silicate, and chlorophyll A concentrations.
Fifty Degrees North, Four Degrees West	https://www.youtube.com/watch?v=IRb3-J2ABYA	All of the data in this composition derive from 12 observed time points collected at monthly intervals at the L4 Station during 2007. The composition is composed of seven choruses. Each chorus has the same chord progression of 12 measures each in which chords are derived from monthly measures of temperature and chlorophyll A concentrations. The first and last chorus melodies are environmental parameter data as in *Blues for Elle*. The melody in each of the second through sixth choruses is generated from the relative abundances of one of the five most common microbial taxa: Rickettsiales, Rhodobacteriales, Flavobacteriales, Cyanobacteria, and Pseudomondales. A different 'instrument' is used to represent each microbial taxon.
Mycorrhizal Waltz	https://www.youtube.com/watch?v=V3GQ1HfINLk&feature=youtu.be	This composition is drawn from the interaction between a soil fungus and gene expression in tree roots (data unpublished). The melody is taken from the relative gene expression of signaling molecules from the soil fungus over time and across multiple conditions. The chords are taken from the differential gene expression of anti-fungal pathogen genes by the plant root.
A Microbiome Musical	https://www.youtube.com/watch?v=gJTBGtZIHFg	Final-year students at Stellenbosch University, South Africa, used the Microbial Bebop algorithm to generate a series of compositions derived from human gut microbiome data and had their compositions performed by students in the Eikestad Primary School Choir.

Sample Microbial Bebop Compositions "Blues for Elle" and "Fifty Degrees North, Four Degrees West" were initially presented in the Microbial Bebop manuscript (15). "Mycorrhizal Waltz" is first presented here and is derived from unpublished aspen root transcriptomic data. "A Microbiome Musical" data and methods are detailed here: http://stellenboschmicrobiomemusical.blogspot.com.

2. **Bowers, R. M., et al.** 2009. Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei. Appl. Environ. Microbiol. **75**(15):5121–5130.
3. **Dennis, P. G., A. J. Miller, and P. R. Hirsch.** 2010. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? FEMS Microbiol. Ecol. **72**(3):313–327.
4. **Edwards, R. A., et al.** 2006. Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics **7**:57.
5. **Gilbert, J. A., B. Laverock, B. Temperton, S. Thomas, M. Muhling, and M. Hughes.** 2011. Metagenomics methods. Mol. Biol. **733**:173–183.
6. **Gosalbes, M. J., J. J. Abellan, A. Durban, A. E. Perez-Cobas, A. Latorre, and A. Moya.** 2012. Metagenomics of human microbiome: beyond 16s rDNA. Clin. Microbiol. Infect. **18**(Suppl 4):47–49.
7. **Hettiarachchi, G., L. Jassogne, D. Chittleborough, and A. McNeill.** 2009. Distribution and speciation of nutrient elements around micropores. Soil Sci. Soc. Am. J. **73**(4):1319–1326.
8. **Larsen, P., and J. Gilbert.** 2013. Microbial bebop: creating music from complex dynamics in microbial ecology. PLoS One **8**(3):e58119.
9. **O'dor, R. K., K. Fennel, and E. Vanden Berghe.** 2009. A one ocean model of biodiversity. Deep-Sea Res. Pt. II. **56**(19–20):1816–1823.
10. **Ramakrishna, B. S.** 2013. Role of the gut microbiota in human nutrition and metabolism. J. Gastroenterol. Hepatol. **28**(Suppl 4):9–17.
11. **Sax and Violence.** [Online.] https://www.youtube.com/watch?v=CgfZVNv6w2E.
12. **Semenov, A. M., A. H. C. van Bruggen, and V. V. Zelenev.** 1999. Moving waves of bacterial populations and total organic carbon along roots of wheat. Microb. Ecol. **37**(2):116–128.
13. **Teske, A., and K. B. Sorensen.** 2008. Uncultured archaea in deep marine subsurface sediments: have we caught them all? ISME J. **2**(1):3–18.
14. **Whitman, W. B., D. C. Coleman, and W. J. Wiebe.** 1998. Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. U. S. A. **95**(12):6578–6583.
15. **Yoon, S. S., E. K. Kim, and W. J. Lee.** 2015. Functional genomic and metagenomic approaches to understanding gut microbiota-animal mutualism. Curr. Opin. Microbiol. **24**:38-46.
16. **Young, I. M., and J. W. Crawford.** 2004. Interactions and self-organization in the soil-microbe complex. Science **304**(5677):1634–1637.

Downloaded from www.asmscience.org by
IP: 146.137.70.70
On: Sat, 02 Apr 2016 00:22:31