Converting DNA to Music: COMPOSALIGN

Todd Ingalls[a]   Georg Martius[b]   Marc Hellmuth[c]   Manja Marz[c*]
Sonja J. Prohaska[c]

[a] Arts, Media and Engineering, Arizona State University, Tempe, AZ 85289-8709, USA
[b] MPI Göttingen, Bunsenstrasse 10, 37073 Göttingen, Germany
[c] Bioinformatics, University of Leipzig, Härtelstr. 16-18, 04107 Leipzig, Germany
* Corresponding author

Abstract: Alignments are part of the most important data type in the field of comparative genomics. They can be abstracted to a character matrix derived from aligned sequences. A variety of biological questions forces the researcher to inspect these alignments. Our tool, called COMPOSALIGN, was developed to sonify large scale genomic data. The resulting musical composition is based on COMMON MUSIC and allows the mapping of genes to motifs and species to instruments. It enables the researcher to listen to the musical representation of the genome-wide alignment and contrasts a bioinformatician's sight-oriented work at the computer.

1   Introduction

Evolution and Selection shape the phenotype and genotype of an organism in an unique way. Homologous sequences are derived from a common ancestor by a sequence of selective changes and diverge over time. Multiple selective constraints on a genomic sequence constrain evolution and result in interesting structures, e.g. modularization. Evolutionarily shaped structures become discernible when sequences derived from a common ancestor are aligned. The result as well as the method is called "alignment". The data structure is a matrix, which is not only highly informative and story-telling for a biological expert but also patterned in a sometimes aesthetic way. Some patterns are visible when one of the numerous visualization tools is applied [RPC+00, KKZ+09, GJ05, LBB+07].

Nevertheless, the modular and structured nature of much music has struck many as providing opportunities to understand genomic data by translating it to sound [Ohn93, Ohn87, OO86]. However, only a few trials have been made to use music to convey the patterns to the interested party [HCL+99, HMR00, TM07, LWHC00]. All of them focus on single DNA or protein sequences. Early attempts transposed DNA sequences directly to music [OO86]. The assignment of two notes to each of the four characters (4 nucleotides) allowed for some flexibility to arrange notes to musical themes. Sonification of protein sequences offered a larger set of initial characters (20 amino acids) but was even more constrained and suffered from the creation of a monotonous string of notes without musical depth. Consideration of further properties [HM84, GS95, GS01, DC99] of characters or groups of characters and mathematical derivation based upon this additional information resulted in more exciting music but blurred the underlying information. A tool called

`gene2music` [TM07] can be used for automated conversion of protein-coding sequences to music. It maps the 20 amino acids on 13 chords, grouping chemically similar characters together while the chord duration is dependent on the frequency of the underlying codon. One system, PROMUSE [HCL$^+$99] deals with sonification of amino acid features as well as structural information and the similarity between related proteins along the sequences. This similarity between proteins and genomic sequences results from common ancestry and light variation and is of central importance to studies in evolution and genomics.

Presentation of highly complex, multidimensional data requires far more channels to transport information than can be handled in the visual channel alone. Visualization and animation are fairly well developed, however, research on the transport of information via sonification is only recently gaining some interest [HR05]. Surprisingly, the complexity of the information transported by the audio channel is usually low, even though musical compositions for entertainment or artistic purposes show highly complex structures. In a multi-media setting, Lodha et al. [LWHC00] showed that sonification can be efficient in disambiguating data in cases where visual presentation alone would be unclear. However, a direct comparison of the efficiency in auditory or visual information uptake is hard to perform. We can expect, however, that the perception of data via sonification and visualization is conceptually very different. Whether this can be beneficial for data presentation is an area we wish to continue to exam.

In this contribution we describe COMPOSALIGN, the first prototype for alignment sonification that translates genome-wide aligned data into a musical composition. Such an acoustic representation requires an unique mapping of alignment information onto musical features. While some mapping is easy to frame, we strive for a intuitive mapping that is easy to perceive and also lives up to the demand to be artistic, pleasant and interesting.

2   Methods

2.1   Mapping

The main focus of our approach is to sonify the presence and absence of characters in the alignment such that their assignment to the corresponding sequence/species is clear. For simplicity, we assume that sequences are from different species, which allows us to refer to "different sequences" as "different species". However, the sources of the sequences are not essential for our theoretical framework but can be added in later steps. Therefore we have chosen the following mapping, formalized as follows:

A *musical motif* or *pattern* is an ordered set of notes and pauses played in one measure with a specific rhythm. Given a set $\mathbb{S}$ of species, a set $\mathbb{I}$ of instruments and a set $\mathbb{P}$ of (different) patterns, we assign to each species an instrument and a pattern played by the assigned instrument. Therefore, we define an injective function $f : \mathbb{S} \to \mathbb{A}$ with $\mathbb{A} = \{(x, y)$ with $x \in \mathbb{I}$ and $y \in \mathbb{P}\} = \mathbb{I} \times \mathbb{P}$, i.e. we assign to *every* species $S \in \mathbb{S}$ a value $f(S)$. Thus it holds $|\mathbb{S}| \leq |\mathbb{A}|$, since f is injective. Many mappings f fulfill the requirement that each species $S \in \mathbb{S}$ is determined and distinguishable from another species by its values

$f(S)$. The remaining degrees of freedom can be used to include auxiliary information such as the phylogenetic relationship of the species. Therefore, we assign instruments to species such that the relationships among the instruments reflect the relationship among species. However, this assignment is done by hand since the relatedness for instruments is a matter of perception. The usage of two independent features (x, y) with $x \in \mathbb{I}$ and $y \in \mathbb{P}$ to encode the species allows us to handle alignments with up to $|\mathbb{I} \times \mathbb{P}|$ species (here $10 \times 10 = 100$) and to represent two-dimensional phylogenetic information as returned by `SplitsTree` [HB06]. In addition to these 100 possibilities we provide 2 further motifs played by drums and cymbals, respectively. These rhythmical motifs are, in particular, useful to sonify outgroup species.

Given a sequence s we consider n units $u_1, \ldots, u_n$ which are, in particular, subsequences of s such that $\bigcup_{i=1}^{n} u_i \subseteq s$. Biologically, these units are referred to as characters in general, "genes" in this contribution. Moreover the units $u_1, \ldots, u_n$ are ordered, such that u_i occurs before u_j whenever $i < j$.

Each unit u_i can be absent, i.e. "0", or directed, i.e. "+" or "−" if present.

We are now able to define the following matrix A, also called *alignment*.

$$A_{i,j} = \begin{cases} + & \text{, if } u_i \text{ appears in species } S_j \text{ in } + \text{ orientation} \\ - & \text{, if } u_i \text{ appears in species } S_j \text{ in } - \text{ orientation} \\ 0 & \text{, else} \end{cases}$$

This means that all entries $A_{i,j} \neq 0$ for a fixed i are homologous. As explained we have assigned to every species a particular instrument playing a particular pattern. In general, an instrument and the corresponding pattern $f(S_j)$ assigned to species S_j plays during time interval i whenever unit u_i occurs in species S_j, i.e $A_{i,j} \neq 0$. Otherwise the instrument will rest. Whether $f(S_j)$ sounds or not is only dependent on $A_{i,j}$. However, three options can be set to highlight particular information:

Orientation. This option indicates whether a pattern is played forwards or backwards, depending on the orientation of the occurring unit. To be more precise let $f(S_j) = (I, P)$ and let unit u_i occur in species S_j. Then pattern P is played forwards or backwards, whenever $A_{i,j} = $ "+" or $A_{i,j} = $ "−", respectively. As a default $A_{i,j} \neq 0$ is set to $A_{i,j} = $ "+".

Conservation. Conservation information is of central importance for a biological researcher. In some situations, units present in *all* species are the most interesting units which are analyzed in further detail. This option emphasizes units, present/conserved in all species. We have chosen to implement this as a change in harmony. Altering the harmony of a motif is done by a *diatonic transposition*. It shifts every pitch of a pattern by a fixed number of scale steps relative to the pattern's musical scale. To *every* pattern we apply a transposition that is selected with a probability depending on the patterns current scale whenever unit u_i is present in *all* species, Figure 2.1. The probability values, are in part based upon general principles of common practice tonal harmony [KP00] for making well-formed harmonic progressions. Thus a transposition maps a pattern P_j to pattern P'_j, which defines the new P_j. This process is well-known as *first-order Markov chain*. For all

Table 1: Transposition probabilities between Markov states: I maj6 – Tonic major sixth, ii m7 – Supertonic minor seventh, iii m7 – Mediant minor seventh, IV maj7 – Subdominant major seventh, V7 – Major Dominant seventh, vi m6 – Submediant minor sixth, vii o7 – Leading-tone diminished seventh.

from/to	I maj6	ii m7	iii m7	IV maj7	V7	vi min6	vii °
I maj6		0.2	0.2	0.2	0.1	0.2	0.1
ii m7				0.2		0.8	
iii m7				0.3		0.7	
IV maj7	0.3	0.4			0.2		0.1
V7	0.8					0.2	
vi min6		0.7		0.3			
vii °	0.8				0.2		

untuned idio- and membranophones P'_j equals P_j (i.e. the motifs cannot be transposed), in our case this holds for drums and cymbals. Notice that patterns P_j and P'_j are perceived as equal up to the change in scale.

Compression. Phylogenetic analyzes focus on differential information. In such a situation, conserved units are considered as uninformative. This option can be used to compress the detailed information in conserved units, while indicating the occurrence of a unit in all species. Under default options, the musical motif is played as it is. If we switch on the compression option and unit u_i is present in *all* species then for all species $S \in \mathbb{S}$ the chosen instruments are simultaneously playing the first note of each of the respective patterns $f(S)$ relative to their orientation, resulting in a so-called *tutti* chord.

2.2 Invertibility of the Mapping

Information representation, visualization as well as sonification, attempts to convey abstract information in intuitive ways. First, we require the information to be formally retrievable from the representation. In mathematical terms, the introduced mapping needs to be bijective, and thus provide an unique way to retrieve the information from the representation. Second, the information must be perceivable to the human ear. Therefore, we want to take advantage of the human sense of hearing.

If all options are set to "off", it is easy to see that we can determine the species S_i by their values $f(S_i)$ since $f : \mathbb{S} \to A' \subseteq \mathbb{A}$ with $A' = \{f(S) \text{ with } S \in \mathbb{S}\}$ is a bijective function.

Orientation – Induced Constraints. If we want to distinguish if a particular unit appears in forward or backward direction in species $S \in \mathbb{S}$ it must be possible to distinguish whether its motif is played forwards or backwards. Thus no symmetric patterns are allowed. Moreover, it is not allowed to have patterns $P, P' \in \mathbb{P}$ such that playing P backwards sounds just like P' in forward direction and vice versa.

Conservation – Induced Constraints. This option requires restrictions on instrument and pattern usage if we want to distinguish different species S by listening to their respective values $f(S)$. We will denote $f_1(S)$ and $f_2(S)$, as the instrument and the pattern of S, respectively. We can distinguish two cases. First, for all pairs of species S and S' holds that

the instruments are unequal ($f_1(S) \neq f_1(S')$). Then the choice of pattern is unrestricted, since each species is determined by its instrument. Second, if some species S and S' have the same instruments we have to distinguish them by their particular pattern. Thus it is not allowed that any composition of transpositions of $f_2(S)$ and $f_2(S')$, resp., leads to one and the same pattern even in scale. If the orientation option is switched on in addition, we have to make sure that no transposition leads to a symmetric pattern. By definition of the term transposition this case cannot occur if no pattern is originally symmetric.

Compression – Induced Constraints. Recall that this option is used to emphasize the occurrence of a unit in all species and to hide detailed information by means of compression. This could be realized in many ways. One of the simplest is the insertion of a single beep. Due to musical reasons, we decided to play the already mentioned tutti chord instead. We are aware that compression causes informational loss in most cases, e.g. orientation. However, we argue that the qualitative information "presence in all species given" is sufficient in most cases. Concerning the remaining cases, we suggest to omit the compression option.

2.3 Implementation

Our program COMPOSALIGN consists of a back end for the composition of the music using COMMON MUSIC [Tau] which runs in Gauche Scheme [Kaw]. COMMON MUSIC is a valuable toolbox for algorithmic composition and also for outputting MIDI data. It allows for a high level description of the compositional elements and convenient definition of the transformation process due to the expressive power of SCHEME. Additionally, there is a web front-end written in Haskell [tC] acting as a CGI program[1], which allows easy usage without the need to install additional software. The data flow is depicted in Figure 1.

The user can upload an input file. After the initial analysis of the file and automatic selection of settings the user has the opportunity to change various parameters. Among these are the selection of the reference sequence and the assignment of musical instrument and motifs to the individual sequences. The default settings are the ones discussed in this paper, however, depending on the biological question, a different assignment might be optimal.

The alignment data is transformed to music based on the settings. For this purpose, an appropriate SCHEME file is generated which is in turn processed by COMMON MUSIC to create a MIDI file. The SCHEME file contains the collection of motifs, the rules for the composition, and the mapping of the species to any of the twelve motifs and available instruments. The user can listen to or download the generated piece of music.

Input. We use a custom comma separated ASCII file type as input which is organized as follows. The input is a $n \times (3 \cdot m)$ matrix consisting of n rows for n units and m blocks of columns each of which holds the genomic start position, end position, and orientation of the unit for every m species. In each row all single columns are separated by a comma. If the unit is not present in a sequence, NA is used as the value for all 3 entries (start position, end position, and orientation). Comment lines start with a "#" symbol. The first block of

---

[1] http://www2.bioinf.uni-leipzig.de/ComposAlign/

Composition Rules and Motifs

Alignment

```
# D.melanogaster, D.yakuba, D.simulans
129; 301;  +; 7064; 202;  +; 108; 937;  +
101; 292;  +; 2202; 246;  +; 605; 585;  +
292; 143;  +;   NA;  NA, NA; 285; 753;  +
```

ComposAlign

Composed Piece of Music
as Midi File

Parameter,
Pattern and Instrument
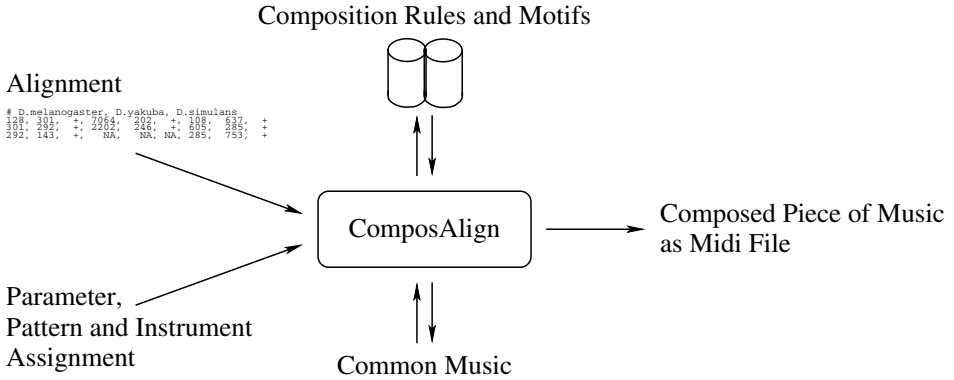Assignment

Common Music

Figure 1: Data flow diagram of COMPOSALIGN. An alignment (input data), parameter settings and the mapping of species to an instrument and pattern are given to COMPOSALIGN via the front-end `www2.bioinf.uni-leipzig.de/ComposAlign`. Using a list of prepared motifs and mapping rules a piece of music is composed.

columns is always treated as the reference species. In principle, it is possible to use any tabular data with absence/presence information for sonification with COMPOSALIGN. An example input file and the corresponding output files can be found in the supplemental material at `http://www2.bioinf.uni-leipzig.de/ComposAlign/`.

3 Application and Results

3.1 Application in Gene Annotation Alignments

For a real data application we have chosen the 12 fly species, each assigned to an unique instrument and pattern. One possible mapping is given by figure 2 and table 3. In all our applications the assignment of species to instruments and patterns fulfills the conditions of an unique mapping for all parameter settings except of the restriction that orientation information is lost in the case of compression, see Section 2.2.

We attempted to sonify data of this kind in a flexible way. These motifs were designed so that they could be placed in various registers. They were also created with varied contours and rhythms to aid in them being individually perceivable in a musical texture.

We used the gene annotations and gene correspondences of chromosome 3R from *D. melanogaster* and the other 11 sequenced Drosophilid genomes as input [Con07]. The input is a matrix $345 \times (3 \cdot 12)$, i.e. 345 genes (units) and 12 species. The genes are given by their genomic sequence interval and their orientation. We sorted the genes by the start position in the reference species (here *D. melanogaster*). Furthermore, we used a relative orientation information, with the orientation of *D. melanogaster* genes set to "+", and the orientation for other genes given by "+" or "-" when the orientation is 'the same' or 'reverse' compared to *D. melanogaster*, respectively.

Figure 2: Panel A shows the 12 motifs in forward orientation. Panel B shows the assignment of instruments to the transposed motifs from panel A. The transpositions are based on appropriate instrument ranges. E.g., motif 1 is transposed up two octaves to sound in a more typical flute range. When motif 2 is set to clarinet it is transposed up an octave in order for it to be perceptible when other instruments sound. The motifs 11 and 12 are for untuned instruments only and will be assigned to snare drums and cymbal, respectively, in all our applications.

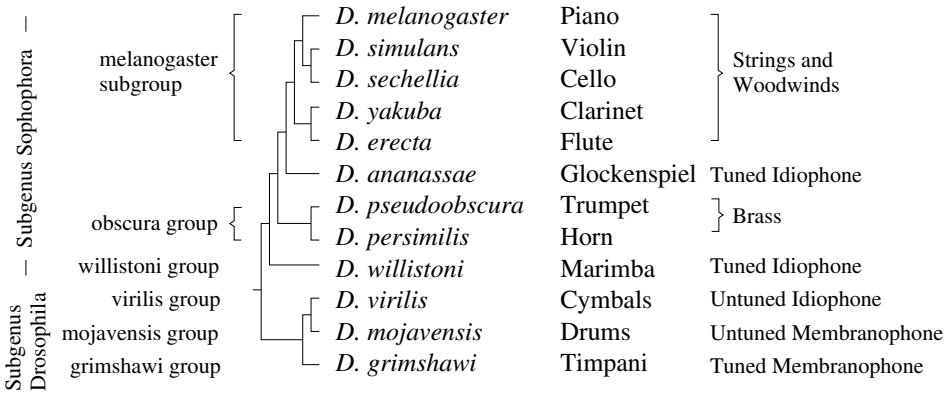|		
D. melanogaster	Piano	
D. simulans	Violin	Strings and
D. sechellia	Cello	Woodwinds
D. yakuba	Clarinet	
D. erecta	Flute	
D. ananassae	Glockenspiel	Tuned Idiophone
D. pseudoobscura	Trumpet	Brass
D. persimilis	Horn	
D. willistoni	Marimba	Tuned Idiophone
D. virilis	Cymbals	Untuned Idiophone
D. mojavensis	Drums	Untuned Membranophone
D. grimshawi	Timpani	Tuned Membranophone

Figure 3: Mapping of fly species to instruments. The tree on the left-hand side represents the topology of the phylogenetic tree [Con07]. Branch lengths are arbitrary.

Moreover, we wanted to have the instrumentation reflect the relative closeness of each species. This closeness is part of a biologist's expert knowledge and reflected in the tree in Figure 3. Of the 12 Drosophila species, five are very closely related – *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*. One of them, *D. melanogaster*, is the model organism and reference species, which we placed in a continuous motif played by the piano, since this provided the basis for the rest of the music. Furthermore, we looked to place the other four in strings and woodwinds so as to provide some similarity but also enough timbral and register difference so they could be distinguished (Figure 3).

As currently implemented each measure takes 2 seconds resulting in a piece of music, 11.5 minutes long, for all 345 genes.

3.2 Evaluation

In Section 2.2 we have formally shown that the selection of an unique instrument and pattern for each species will allow an unique mapping under certain restrictions. However, it remains to be evaluated how the sonification is perceived by the user. The following analysis of COMPOSALIGN is based on impressions of 50 un-trained, non-musician test persons. The example described in 3.1 is only one of several tested cases with various setting.

Number of Organisms/Instruments. Depending on the education in the arts of the test persons, up to 12 instruments could be recognized. Most people felt confident to distinguish six instruments. If distinction of more (instrument) tracks is desired the majority of people need to be trained to more clearly differentiate the instruments or patterns. We might also want to consider to utilize other types of instrumental or synthesized sounds which would be more easily identified by untrained users.

In the case of 2 or 3 species, the composition was described as "musically pleasing" and

users found it easy to hear which genes were present in which species. However, the ability to resolve the presence/absence pattern decreased rapidly with the number of different instruments and/or motifs playing per measure. Nevertheless, presence/absence of genes that involve groups of species, was still found easy to hear.

Most people who concentrated on a specific instrument and tried to observe the presence/absence at a specific time point, found the correct solution independently of the number of instruments played concurrently.

Conserved Sites – Changes in Harmony. The introduction of changes in harmony based on the local context improved the artistic value of the output and the listeners attention span. All participants had the impression of a much more interesting piece of music, if the conservation option was used. Apart from this aesthetic effect, it also helped emphasized conservation and to draw the listeners attention to conserved regions.

Conserved Sites – Compressed Units. While this sets the presence of m (while m is the total number of species in the alignment) and less than m species clearly apart from each other, it also causes a time compression and allows the user to focus on the data where the absence/presence patters are more informative from a biological perspective. For emphasis of conservation, users preferred the compression option over the conservation option.

All test persons were enthusiastic after including changes in harmony and compressed chords about the musical variability. The outcome was described much more "happier", "interesting, irregular", "less crowded", "rhythmically interesting" and "dramatic". The interrogation also provides an intriguing result in that certain choices that were made largely for aesthetic reasons also appear to make the sonification more legible to users.

Orientation of a Gene – Forward and Backward Motifs. The asymmetry of the individual motifs, some of which are clearly ascending, is an essential attribute to sonify a character's direction information. The character of the motifs allows the user still to identify the mirrored motifs as belonging to the same motif. The results sound pleasant, however most test persons found it difficult to follow which motifs were reversed when several instruments played at the same time. It is unclear if the ear needs some training only or if it might be necessary to explore other strategies which may help in communicating this information.

Mapping – Assignment of instruments and patterns. Using different settings we expected to find combinations that might sound unpleasant. Given an uncommon combination of instruments (e.g. drums, marimba and trumpet) most people found the outcome to be surprisingly rich in character and interesting. When various outputs for the same data file were heard with different instruments and patterns in place, the participants felt that this emphasized the underlying structure in the data.

4 Conclusion and Future Work

To date, COMPOSALIGN is the first prototype of an alignment sonification tool. Existing sonification methods for single biological sequences map each individual characters (e.g.

nucleotides or amino acids) on single notes or chords. We decided to map one character to a measure. This had mainly two effects. First, it added the necessary degrees of freedoms to encode more information and still allowed us to take compositional aspects into account and make it sound pleasant. Second, it stretched the information onto a larger time interval, allowed organized presentation of the information with a measure and therefore insured that the information was easy to perceive. COMPOSALIGN draws its power from the motif design and mapping rules that are modular and flexible. Also, biological sequence alignments are particularly suited for sonification since individual elements of information become blurred in a composition when researcher's become more interested in the overall picture (e.g. groups of species with a conspicuous absence/presence pattern in the sequences). It might turn out that music is a suitable medium to convey information on different levels of resolution at the same time. This leads us immediately to the question: Can sonification compete with or outperform the currently dominating visualization? If not, is sonification able to transport a certain kind of information better than visualization? The omnipresence of visualization might suggest a better performance in all respects. However, to perform a fair test, a competitive sonification tool first needs to be developed. Our prototype is just a small step in this direction.

Based on the experience gained during our project, we intend to construct a mapping for alignments that allows us to add different kinds of additional/contextual information (e.g. lengths of characters, distance between characters, higher order annotation, phastcons score). An interactive interface shall allow the user to edit the parameters on runtime and display the scores and alignment in flying windows. This shall allow the interested user to play (with) his/her alignment.

"Play is the highest form of research." (quote by Albert Einstein)

Acknowledgments.

References

[Con07]   Drosophila 12 Genomes Consortium.   Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, 450(7167):203–218, Nov 2007.

[DC99]   John Dunn and Mary Ann Clark. Life Music: The Sonification of Proteins. *Leonardo*, 32(1):25–32, 1999.

[GJ05]   S Griffiths-Jones.   RALEE–RNA ALignment editor in Emacs.   *Bioinformatics*, 21(2):257–259, Jan 2005.

[GS95]   P Gena and C Strom. Musical synthesis of DNA sequences. In *in XI Colloquio di Informatica Musicale*, pages 203–204, Bologna, I, 1995.

[GS01]   P Gena and C Strom. A physiological approach to DNA music. In *in Proceedings of CADE 2001*, pages 81–86, Glasgow, UK, 2001. Glasgow School of Art Press.

[HB06]   D H Huson and D Bryant. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*, 23(2):254–267, Feb 2006.

[HCL+99]   M D Hansen, E Charp, S Lodha, D Meads, and A Pang. PROMUSE: a system for multi-media data presentation of protein structural alignments. *Pac Symp Biocomput*, pages 368–379, 1999.

[HM84]   K Hayashi and N Munakata. Basically musical. *Nature*, 310(5973):96–96, Jul 1984.

[HMR00]   T Hermann, P Meinicke, and H Ritter. Principal Curve Sonification. In *in Proceedings of the Int. Conf. on Auditory Display*, pages 81–86, 2000.

[HR05]   Thomas Hermann and Helge Ritter. Crystallization sonification of high-dimensional datasets. *ACM Trans. Applied Perception*, 2(4):550–558, 10 2005.

[Kaw]   Shiro Kawai. Gauche Scheme - `http://practical-scheme.net/gauche/index.html`.

[KKZ+09]   R M Kuhn, D Karolchik, A S Zweig, T Wang, K E Smith, K R Rosenbloom, B Rhead, B J Raney, A Pohl, M Pheasant, L Meyer, F Hsu, A S Hinrichs, R A Harte, B Giardine, P Fujita, M Diekhans, T Dreszer, H Clawson, G P Barber, D Haussler, and W J Kent. The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res*, 37(Database issue):755–761, Jan 2009.

[KP00]   Stefan M. Kostka and Dorothy Payne. *Tonal Harmony, with an introduction to twentieth-century music*. McGraw-Hill, Boston, 4th edition, 2000.

[LBB+07]   M A Larkin, G Blackshields, N P Brown, R Chenna, P A McGettigan, H McWilliam, F Valentin, I M Wallace, A Wilm, R Lopez, J D Thompson, T J Gibson, and D G Higgins. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948, Nov 2007.

[LWHC00]   Suresh K Lodha, Doug Whitmore, Marc Hansen, and Eric Charp. Analysis and user evaluation of a musical-visual system: Does music make any difference. In *in Proceedings of the Int. Conf. on Auditory Displays*, pages 167–172, 2000.

[Ohn87]   S Ohno. Repetition as the essence of life on this earth: music and genes. *Haematol Blood Transfus*, 31:511–518, 1987.

[Ohn93]   S Ohno. A song in praise of peptide palindromes. *Leukemia*, 7 Suppl 2:157–159, Aug 1993.

[OO86]   S Ohno and M Ohno. The all pervasive principle of repetitious recurrence governs not only coding sequence construction but also human endeavor in musical composition. *Immunogenetics*, 24(2):71–78, 1986.

[RPC+00]   K Rutherford, J Parkhill, J Crook, T Horsnell, P Rice, M A Rajandream, and B Barrell. Artemis: sequence visualization and annotation. *Bioinformatics*, 16(10):944–945, Oct 2000.

[Tau]   Heinrich Taube. Common Music Website - `http://commonmusic.sourceforge.net/doc/cm.html`.

[tC]   HASKELL Community. Common Music Website - `http://www.haskell.org/`.

[TM07]   R Takahashi and J H Miller. Conversion of amino-acid sequence in proteins to classical music: search for auditory patterns. *Genome Biol*, 8(5):405–405, 2007.