Latent Evolutionary Signatures: A General Framework for Analyzing Music and Cultural Evolution

Jonathan Warrell[1,*], Leonidas Salichos[1,*], Mark Gerstein[1]
[1] Program in Computational Biology and Bioinformatics, Yale University, * equal contribution

Abstract

Cultural processes of change bear many resemblances to biological evolution. Identifying underlying units of evolution, however, has remained elusive in non-biological domains, especially in music, where music evolution is often considered to be a loose metaphor. Here we introduce a general framework to jointly identify underlying units and their associated evolutionary processes using a latent modeling approach. Musical styles and principles of organization in dimensions such as harmony, melody and rhythm can be modeled as partly following an evolutionary process. Furthermore, we propose that such processes can be identified by extracting latent evolutionary signatures from musical corpora, analogous to the analysis of mutational signatures in evolutionary genomics, particularly in cancer. These latent signatures provide a generative code for each song, which allows us to identify broad trends and associations between songs and genres. To provide a test case, we analyze songs from the McGill Billboard dataset, in order to find popular chord transitions (k-mers), associate them with music genres and identify latent evolutionary signatures related to these transitions. First, we use a generalized singular value decomposition to identify associations between songs, motifs and genres, and then we use a deep generative model based on a Variational Autoencoder (VAE) framework to extract a latent code for each song. We tie these latent representations together across the dataset by incorporating an energy-based prior, which encourages songs close in evolutionary space to share similar codes. Using this framework, we are able to identify broad trends and genre-specific features of the dataset. Further, our evolutionary model outperforms non-evolutionary models in tasks such as period and genre prediction. To our knowledge, ours is the first computational approach to identify and quantify patterns of music evolution *de novo*.

1. Introduction

Molecular evolution involves the process of changing of a particular genomic sequence or code (DNA or RNA). When the DNA is replicated, each nucleotide (C/G/A/T) can sometimes be replaced by another in a process known as a mutation, and evolutionary models of increasing complexity can account for different mutation rates of transitioning from one nucleotide to another (Fig. 1). These mutations can be inherited and propagated to future generations through multiple levels of selection or without selection through genetic drift (1). Characterizing these nucleotide transitions, is an important task in different fields of molecular evolution, from phylogenetics (2), genotyping (3), microbial evolution (4) and cancer genomics (5, 6). Particularly, in cancer genomics, the underlying processes that cause these mutations (e.g. tobacco, UV, aging, etc) can be directly linked with specific mutational signatures, and specific cancer types (6–8).

In parallel to biological evolution, cultural evolution is also a theory of change, notably social change, where culture is defined as a socially learned behavior, and underlying structural constraints may influence transmission of behavior (9, 10). From the 19th century, different approaches have been used to study cultural evolution, including sociocultural evolutionism (11), social cycle theory (12) and historical materialism (13). Among other things, these theories disagree upon the interactions between environment and biological constraints, directionality of progress, means of change or progress, and levels of selection. Using a term coined by Richard Dawkins in 1976 (14), 'memetics' have been widely used to represent cultural information transfer, where a 'unit of culture' or 'meme' (idea, belief, behavior etc.) can propagate through imitation (or 'mimesis'). However, due to the lack of a stable 'code' for these memes - compared to DNA for genes - they have also been described as lacking in explanatory power (15, 16).

Music, as a cultural artifact, may be viewed as embedded in a process of cultural evolution (17). For instance, each song or work can influence and be influenced by other songs (for instance, via harmony, melody etc.), providing an analogue of replication and inheritance. Meanwhile, songs are continually subjected to changing cultural tastes that can act as a process of selection, while also respecting underlying functional constraints, such as harmonic syntax (18). Further, songs within a specific genre can have a level of homology, as a result of their shared lineage. While the above is suggestive as a viewpoint, the question of which types of influence should count as evolutionary (the underlying *memes*) is an open problem, as is the extent to which concepts such as mutation and the genotype-phenotype distinction can be applied to musical artifacts (19). Historically, different approaches have been applied to bridge the gap between music phylogenetics and evolution (10). Compared to language, where evolutionary relationships can be represented as phylogenetic trees or networks (20), music evolution is still often considered as a loose metaphor (21–24), where the evolutionary aspects are yet to be shown (9, 25).

Previous computational approaches to musical evolution have analyzed changes in audio features (26–28), interval use (27) and song selection (29). While previous methods have considered how characteristic variables extracted from such features change with time, such as PCA components (26) or measures of information-theoretic complexity (28), these are predefined, rather than extracted through fitting an integrated model. Further, while not explicitly evolutionary in focus, other approaches have considered changes in harmonic usage in time either through derived features from individual chords (30) or chord transitions using Hidden Markov Models (31). As above, these approaches primarily consider changes in surface features, as opposed to features learned via fitting an explicit evolutionary model.

To tackle the problem of identifying potential evolutionary processes in music, we adopt a machine learning perspective. We develop a latent evolutionary model which directly models the generation of observed musical data, such as chordal sequences from songs (26), using an underlying hidden binary code representation. We are influenced here by recent work in evolutionary genomics, which has shown that it is possible to extract signatures of mutational processes from cancer genomics data (7, 8), allowing latent factors to be inferred (6–8, 32, 33) (see Fig. 1). Specifically, we introduce a model that allows us to identify underlying evolutionary 'signatures' *de novo*, by optimizing the reconstruction error in a Variational Autoencoder framework (34), which can model arbitrarily complex generative processes through

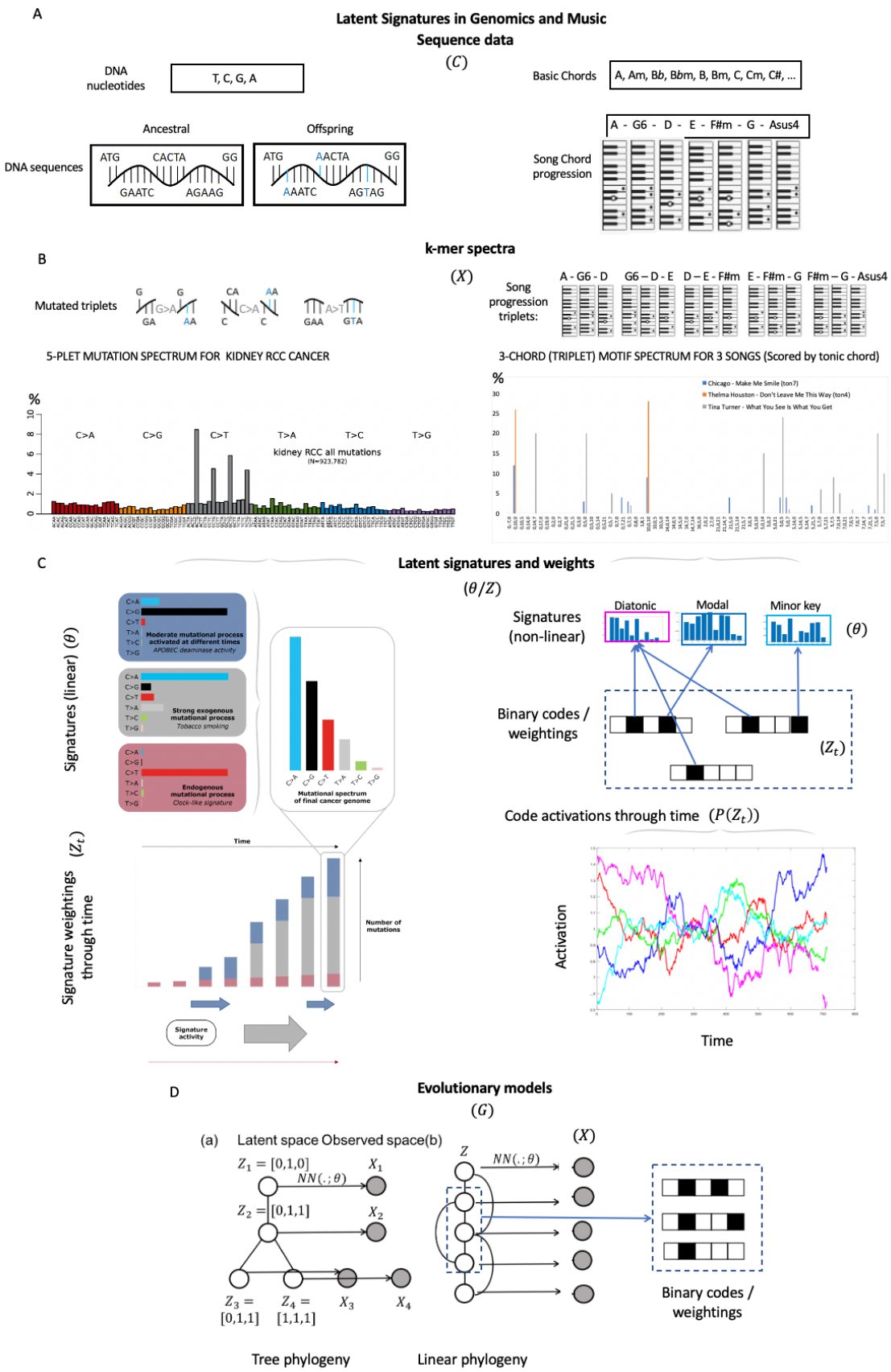

A
Latent Signatures in Genomics and Music
Sequence data
(C)
DNA nucleotides
T, C, G, A
Basic Chords
A, Am, Bb, Bbm, B, Bm, C, Cm, C#, ...
Ancestral
Offspring
DNA sequences
ATG CACTA GG
GAATC AGAAG
ATG AACTA GG
AAATC AGTAG
A - G6 - D - E - F#m - G - Asus4
Song Chord progression
B
k-mer spectra
(X)
Mutated triplets
Song progression triplets:
A - G6 - D
G6 – D - E
D – E - F#m
E - F#m - G
F#m – G - Asus4
5-PLET MUTATION SPECTRUM FOR KIDNEY RCC CANCER
3-CHORD (TRIPLET) MOTIF SPECTRUM FOR 3 SONGS (Scored by tonic chord)
C>A
C>G
C>T
T>A
T>C
T>G
kidney RCC all mutations
(N=923,782)
Chicago - Make Me Smile (ton7)
Thelma Houston - Don't Leave Me This Way (ton4)
Tina Turner - What You See Is What You Get
C
Latent signatures and weights
(θ/Z)
Signatures (linear) (θ)
Moderate mutational process activated at different times
APOBEC deaminase activity
Strong exogenous mutational process
Tobacco smoking
Endogenous mutational process
Clock-like signature
Mutational spectrum of final cancer genome
Signature weightings through time (Z_t)
Time
Number of mutations
Signature activity
Signatures (non-linear)
Diatonic
Modal
Minor key
(θ)
Binary codes / weightings
(Z_t)
Code activations through time (P(Z_t))
Activation
Time
D
Evolutionary models
(G)
(a) Latent space Observed space(b)
Z_1 = [0,1,0]
NN(.;θ)
X_1
Z_2 = [0,1,1]
X_2
Z_3 = [0,1,1]
Z_4 = [1,1,1]
X_3
X_4
Z
NN(.;θ)
(X)
Binary codes / weightings
Tree phylogeny
Linear phylogeny

Figure 1. Illustrates correspondence between latent signature models in evolutionary cancer genomics and musical evolution. (A) Biological evolution is a theory of change often signifying alterations in the genetic code of DNA nucleotide sequences. We draw parallels with music, by replacing the genetic code with a sequence of notes or chords. (B) While in biological evolution we derive a mutational spectrum by considering the point mutations from an ancestral to an offspring sequence in the form of k-*mers*, in music we construct a transitional motif spectrum that denotes the frequency of possible chord transitions within a song, song family, or category (e.g. composer, era, genre). As a biological example, we show a mutational spectrum observed in a type of Kidney cancer, as adapted from [6]. As a musical example, we show the motif spectrum of three popular songs for comparison, where chord transitions have been normalized by the tonic chord. (C) The mutational or motif spectrums can be, in turn, decomposed into latent weighted signatures in a linear or non-linear manner respectively. In the case of cancer, the mutational spectrum can be linearly reconstructed through its latent signatures. In the case of the music evolution model presented, a non-linear evolutionary neural network with binary codes is used to identify the corresponding latent signatures and their activations through time. (D) illustrates different forms of phylogenetic graph underlying the latent signatures model, with G defined by a tree (a) and temporal window function (b). Model variables are notated in parentheses, see text (Sec. 2.2).

using deep neural network decoders. Recent extensions of the VAE framework have considered adding extra structure to the latent space, such as graph structure (35) or cluster structure (36) to derive more interpretable representations for particular applications. In our case, we add an evolutionary structure to the latent space, by incorporating an energy-based prior, which encourages songs close in evolutionary space to share similar codes. The prior thus directly embeds a notion of *mutational distance* between codes, while the decoder allows a complex map to be learned between codes and observed *phenotypes*.

As a test case, we apply our model to analyze songs from the McGill Billboard dataset from 1958-1991 (30, 37). We represent each song as a distribution over chord transitions (k-*mers*), and use a generalized singular value decomposition to identify associations between songs, motifs and genres. We then apply our model to identify latent evolutionary signatures. We first interpret these signatures, by identifying them with features of changing harmonic syntax. We then evaluate the representations learned by their ability to perform period and genre prediction on hold-out data. Our evolutionary model significantly outperforms other non-evolutionary models on these tasks, suggesting the evolutionary structure is informative. To our knowledge, ours is the first approach to statistically test for evolutionary structure where the units of transmission are unknown and must be inferred. Our method thus provides evidence for music evolution via cultural progression beyond the scope of a 'loose metaphor'.

2. Latent evolutionary signatures model

2.1 Extracting Musical Features

We begin by describing our approach to extracting harmonic features from the McGill Billboard dataset songs (30, 37). In analogy to mutational spectra that are used to reconstruct specific signatures in cancer genomics (7, 8) (Fig. 1), we first determine the frequency of specific chord

motifs and chord transitions within each song, on which the latent evolutionary signatures will be based. The raw data consists of N songs, each represented by a sequence of chords $C_n = [(a_{n,1}, b_{n,1}) \dots (a_{n,l(n)}, b_{n,l(n)})]$, where $a_{n,i} \in \{0 \dots 11\}$ represents the root pitch-class of the i'th chord (letting Ab=0, A=1 etc.) of the n'th song, and $b_{n,i} \in \{0,1\}$ represents whether the chord is major or minor (0/1 resp.). For convenience, we do not encode added 7ths/9ths, inversions, or other chordal variants, and we prune the chordal sequences to remove chordal repetitions. Additionally, for each song we represent the tonality of the song by the pair $\left(\tau_n^{(1)}, \tau_n^{(2)}\right)$, where $\tau_n^{(1)} \in \{0 \dots 11\}$ represents the tonic, and $\tau_n^{(2)} \in \{0,1\}$ the mode (major/minor).

We generate a harmonic feature vector for each song, X_n, by applying a set of filters to each chord sequence. The first set of (basic) filters represent chordal transitions of length K. Here, the f'th filter is represented by a transition vector, $[t_{f,0}, \dots t_{f,K-2}]$, where $t_{f,i} \in \{0 \dots 11\}$, and a binary chord-type vector $[u_{f,0}, \dots u_{f,K-1}]$, where $u_{f,i} \in \{0,1\}$. The response of filter f to song n is calculated:

$$X_{n,f} = \sum_{i=1\dots l(n)-K+1} S_{n,f,i}^{(1)} S_{n,f,i}^{(2)}, \tag{1}$$

$$S_{\mathrm{n},f,i}^{(1)} = [\sum_{j=0\dots K-2} [\mathrm{mod}_{12}\left(a_{n,i+j+1} - a_{n,i+j}\right) \neq t_{f,j}] = 0],$$

$$S_{n,f,i}^{(2)} = \left[\sum_{j=0\dots K-1} \left[b_{n,i} \neq u_{f,j}\right] = 0\right].$$

Here, $S_{n,f,i}^{(1)}$ and $S_{n,f,i}^{(2)}$ are the 'scores' for filter f on song n at position i, representing agreement between the chord transitions and the major/minor chord types respectively. Further, [.] is the Iverson bracket, which is 1 for a true proposition, and 0 otherwise. Since there are 12^{K-1} transition vectors, and 2^K chord-type vectors, there are $12^{K-1} \cdot 2^K$ possible basic filters in total.

We also consider a second set of filters, which are normalized to the tonic/tonality of each song (τ-normalized). Here, the g'th filter is again represented by a transition vector, $[t_{g,0}, \dots t_{g,K-1}]$, where $t_{g,i} \in \{0 \dots 11\}$ and a binary chord-type vector $[u_{g,0}, \dots u_{g,K}]$, where $u_{g,i} \in \{0,1\}$, but here the transition vector represents the offset relative to the tonic. Additionally, an extra bit is added to u_g to represent the key of the song. The response of tonality normalized filter g to song n is calculated:

$$X_{n,g} = \sum_{i=1\dots l(n)-K+1} S_{n,g,i}^{(1)} S_{n,g,i}^{(2)}, \tag{2}$$

$$S_{n,g,i}^{(1)} = [\sum_{j=0\dots K-1} [\mathrm{mod}_{12}\left(a_{n,i+j} - \tau_n^{(1)}\right) \neq t_{g,j}] = 0],$$

$$S_{n,g,i}^{(2)} = [\sum_{j=0\dots K-1} \left[b_{n,i} \neq u_{g,j}\right] = 0]] \wedge \left[\tau_n^{(2)} = u_{g,j}\right].$$

For the tonality normalized filters, there are 12^K transition vectors, and 2^{K+1} chord-type vectors, leading to $12^K \cdot 2^{K+1}$ possible τ-normalized filters in total.

Both the basic filters and the τ-normalized filters described above can be represented using an indexing scheme for pairs of chords based on mod-12 arithmetic. We describe this indexing scheme in more detail in Sec. 3.1, and illustrate it in Fig. 2.

2.2 Model Formulation

We now describe our model for latent evolutionary signatures, and briefly outline our training algorithm for the model. The model requires as input a set of observed training data vectors $X_{n=1\ldots N}$, for which we use the harmonic features defined in the previous subsection (although other types of features may also be used). In addition, we require a weighted graph over the training examples G, which is a positive symmetric matrix of size $N \times N$, where $G(i,j)$ represents the closeness in evolutionary space of training samples i and j. For our current investigation, we predefine G by ordering the songs in the training set by date, and connecting each song to all other songs in a temporal window of size w on either side (Fig. 1D); hence $G(i,j) = [|i-j| \leq w]$.

The evolutionary model then fits a latent code to each song, $Z_n \in \{0,1\}^B$, where B is the number of bits in the latent code vectors, corresponding to the number of evolutionary signatures. Further, the model fits a neural network which provides a generative model of the observed feature vectors X_n from the latent codes. The likelihood of the model combines a reconstruction loss, with a prior over the codes, which penalizes large changes in the latent vectors between strongly related songs according to G:

$$P(X,Z) = P(X|Z)P(Z), \tag{3}$$

$$P(X|Z) = \prod_i P_{\text{gen}}(X_i|NN(Z_i;\theta)),$$

$$P(Z) \propto \exp\big(-E(Z)\big),$$

$$E(Z) = \sum_{i,j<i} G(i,j) \sum_b \gamma[Z_i(b) \neq Z_j(b)].$$

Here, $NN(Z_i;\theta)$ is a neural network parametrized by θ, $P_{\text{gen}}(.|.)$ is the generative model for the observed data (for instance, a Bernoulli model if the features of X_i are thresholded at 1, and the output of the neural network represents the probability that each feature is 1), and $E(Z)$ is an energy model defined over the latent vectors, which penalizes pairs of latent vectors by the product of their closeness in the underlying evolutionary graph and their Hamming distance weighted by γ. This form of $E(Z)$ is motivated by the considering a point-wise mutational process acting on code vectors between related songs. Further, we note that, while we have predefined G, alternatively a prior may be placed over G (for instance, enforcing that G has a tree or DAG structure, which respects temporal ordering, see Fig. 1), and it may be considered an additional parameter of the model likelihood.

We fit the model in Eq. 3 by optimizing the evidence lower-bound on the likelihood (ELBO) (34), while using a mean-field approximation as our variational distribution over the latent space (38). For our model, the ELBO has the form:

$$\mathcal{L} = \mathbb{E}_{Q(Z)}[\log P(X|Z)] - \mathrm{KL}(Q(Z)||P(Z)), \tag{4}$$

where $Q(Z)$ is the mean-field distribution across the latent space, and $\mathrm{KL}(.||.)$ is the KL divergence. For convenience, we assume that B is small enough that Q may be represented by an explicit discrete distribution across all code-vectors, with each song treated independently:

$$Q(Z) = \prod_i \sum_\beta q_{i,\beta}[Z_i = \beta], \tag{5}$$

where $\beta \in \{0,1\}^B$. The bound in Eq. 4 can be optimized by Variational Bayes Expectation-Maximization (VBEM) (39). For the E-step, this requires optimizing the local mean-field parameters $q_{i,\beta}$. Using standard mean-field results (38), this results in the local updates:

$$q_{i,\beta} \propto P_{\mathrm{gen}}(X_i|\beta) \exp\left(\sum_{j\neq n,\beta'} q_{j,\beta'} \exp\left(-E_{i,j}(\beta,\beta')\right)\right), \tag{6}$$

where $P_{\mathrm{gen}}(X_i|\beta) = P_{\mathrm{gen}}(X_i|NN(\beta;\theta))$, and

$$E_{i,j}(\beta,\beta') = G(i,j) \sum_b \gamma[\beta(b) \neq \beta'(b)]. \tag{7}$$

For the M-step, code-vectors may be sampled for each song from $Q(Z_i)$, and θ optimized by gradient descent to maximize $\log P_{\mathrm{gen}}(X_i|NN(Z_i;\theta))$.

2.3 Period and Genre Prediction

For period and genre prediction tasks, we split the data into a training and test partition, $X_{\mathrm{train}}, X_{\mathrm{test}}$, where X_{test} is formed by sampling every 5th song in chronological order. We first train the evolutionary signatures model on X_{train} using the approach in Sec. 2.2. We then infer the maximum a posterior latent representation for each test song independently, using the marginal distribution across the training latent codes as a prior. Hence for test song j, we find:

$$Z^*_{\mathrm{test},j} = \mathrm{argmax}_\beta \left(P_{\mathrm{gen}}\left(X_j \middle| NN(\beta;\theta)\right) P(\beta)\right), \tag{8}$$

where $P(\beta) = (1/N)\sum_i Q_i(\beta)$. Similarly, for each training song we find $Z^*_{\mathrm{train},i} = \mathrm{argmax}_\beta(Q_i(\beta))$. We then perform period prediction using a kernel-regression approach.

Hence, for the i'th training song in chronological order, we specify the label $y_{\text{train},i} = i/N$. For test song j, we then predict:

$$\hat{y}_{\text{test},j} = \left(\frac{1}{W}\right) \sum_{i} \exp\left(-\alpha H\left(Z^*_{\text{test},j}, Z^*_{\text{train},i}\right)\right) y_{\text{train},i}, \tag{9}$$

where $W = \sum_i \exp\left(-\alpha H\left(Z^*_{\text{test},j}, Z^*_{\text{train},i}\right)\right)$, $H(.,.)$ is the Hamming distance, and α is set by cross-validation as a hyper-parameter. To assess performance, we calculate the Pearson correlation coefficient between $\hat{y}_j$ and the ground-truth test labels, $y_{\text{test},j} = j/N_{\text{test}}$.
For genre prediction, each song may be assigned to multiple genres (rock, jazz, pop etc.). We treat assignment to each genre as a separate binary classification task. For a given genre, we assign labels $y_{\text{train},i} \in \{0,1\}$ for songs belonging to the genre versus not belonging respectively, and apply Eq. 9 to estimate a score for a test song with respect to the genre. Similarly, we assess performance by the Pearson correlation coefficient between the test scores and the ground-truth binary labels, and take the average Pearson correlation as a measure of genre prediction.

3. Results

3.1 McGill Billboard Dataset

We first describe the initial processing we apply to the McGill Billboard dataset (30, 37). We remove all duplicate songs, leaving us with 730 songs in total. We order these by date, and take every 5th song to form a testing set, thus splitting the data into a training and testing set of 584 and 146 songs respectively. We calculate basic and τ-normalized harmonic features for each song, as described in Sec. 2.1, for $K = 1, 2, 3, 4$. To create a compressed feature vector, we use only filters which are non-zero in at least 100 songs to form the data matrix, X.

Further, we use an indexing scheme for pairs of chords based on mod-12 arithmetic to encode the filters described in Sec. 2.1. First, to encode the τ-normalized filters, we identified the chord progression for each song in the database and assigned a chord value within a specific range, while correcting based on the song's tonic key (see Fig. 2a). More specifically, for chord Ab we assign an initial value of "0", for A:1, A#:2… and G:11. Then, we normalize to the tonic key, in such a way that, for chord G in tonic A, the normalized final value would be equal to 11-1=10. To distinguish between major and minor chords and tonic keys, we established a set of specific rules. For every minor key/chord, we add +24 to the initial value. Then we enforce specific key-chord associations to fall within ranges by further adding or subtracting ±12, if necessary. If both the tonic key and chord type are major, the normalized value should fall within [0,11]. If the tonic is a major key but the chord is minor, we add +24 to the key value, but potentially subtract 12 if necessary, in order for the new chord value to fall within [12,23]. Similarly, if both the tonic and chord keys are minor, the new value falls within [-1,-12], while if the song's tonic key is minor and the chord is in major, the new value should fall in [-13,-24]. Hence, if the song's tonic key is in G major and one chord from the song progression is A major, the final chord

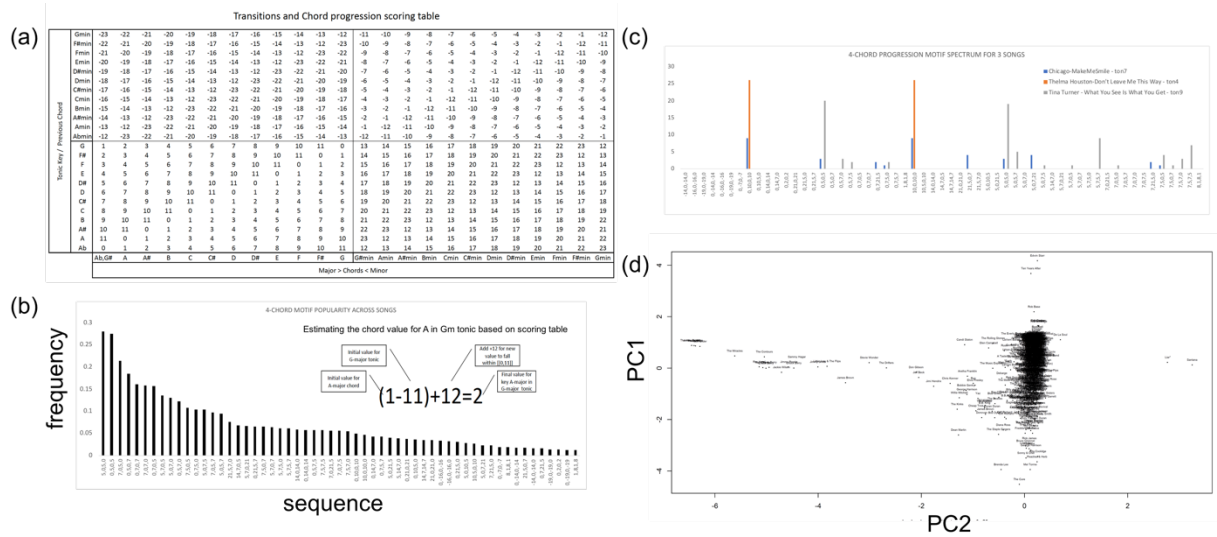


Figure 2. Quantifying chords/transitions to construct motif spectra and perform correspondence analysis with SVD. (a) illustrates indexing scheme, (b-c) plot motif frequencies and spectra, and (d) plots songs in SVD space. See Sec. 3.1

value assigned will be equal to 2 according to $(1 - 11) + 12 = 2$. In this way, we extracted each song's frequency of 3- and 4-mer vectors of consecutive chords based on the song's chord progression and key.

To encode the basic (non-normalized) filters, instead of correcting each chord to the tonic key, we obtained 2- and 3-mer vectors where each element represents the transition between two chords. By following the same set of conventions, we add +24 to the initial value if a chord is minor. If both chords *n* and *n*-1 are major, the transitional value is in [0,11], while if both chords *i* and *i*-1 are minor, the transitional value is in [-1,-12]. Similarly, if chord *i* is minor but *i*-1 is major, the transitional value is in [12,23], while if chord *n* is major but *n*-1 is minor, then the transitional value is in [-13,-24]. For example, the values for the progression Am ->C->D ->F are represented by the 3-mer vector {-21,2,3}.

This indexing scheme allows us to quantify prevalent 3,4-chord motifs (Fig. 2b), and to construct motif spectra for each song (Fig. 2c). Finally, we decompose every song's indexed residuals (as obtained from the chord transitions), using single value decomposition to identify shared similarities between songs and chord transitions. Fig. 2d, shows the correspondence analysis between principal components PC2 and PC8, while a heatmap of the indexed residuals for all songs is shown in the Supplement.

3.2 Evolutionary Signature Interpretation

We train models of evolutionary latent signatures (Sec. 2.2) using 3 types of harmonic features: (1) K=1 τ-normalized filters, corresponding to chord usage normalized by tonality, (2) K=4 basic filters, corresponding to transition sequences between groups of 4-chords, not normalized by tonality, and (3) K=4 τ-normalized filters, which are as in (2), but normalized by tonality. In each case, we fix the number of latent signatures to be 5 ($B = 5$), and use a single level neural

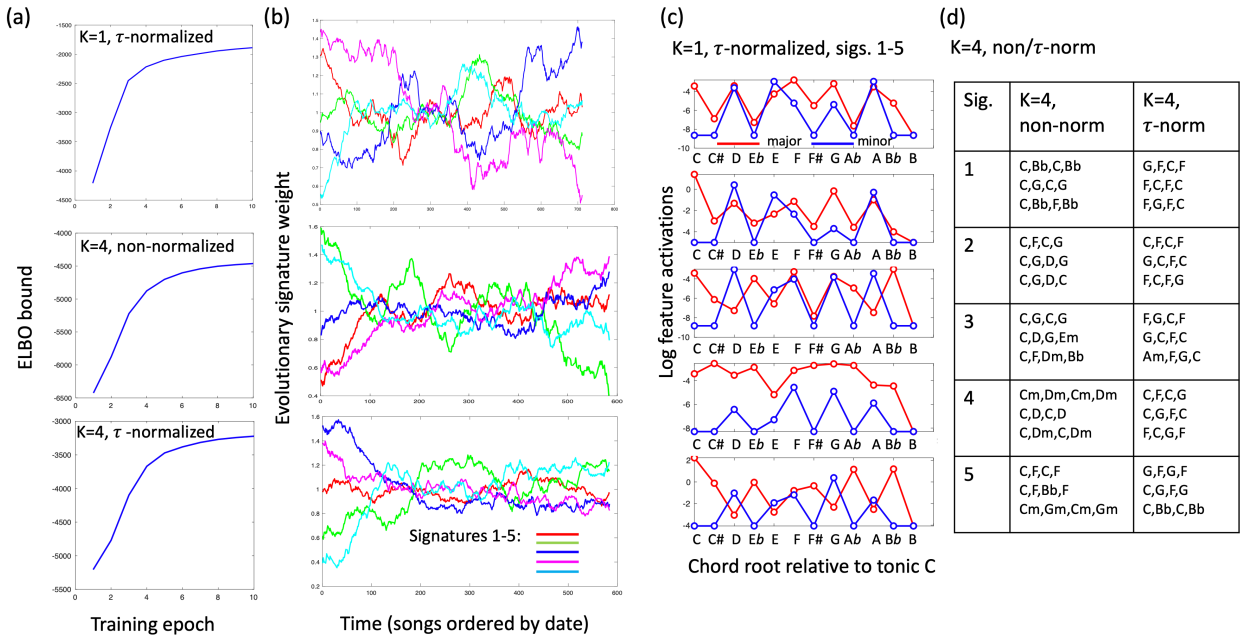
network to aid interpretability, whose output is a vector of Bernoulli probabilities, predicting whether each filter gives a non-zero response in a song or not (see Sec. 2.2). We train all models for 10 epochs of VBEM, taking 20 steps of gradient descent within each M-step to optimize the neural network parameters θ. We monitor the ELBO bound on the training log-likelihood and Pearson correlation (r) for period and genre prediction (discussed below) on the test set at each epoch, and set $\gamma = 1$ in Eq. 3 by cross-validation on r for period prediction (optimizing over $\gamma = \{0.1\ 1\ 10\}$).

Fig. 3 shows the output of the training for these three models. The ELBO monotonically increases, and begins to exhibit incremental improvements at 10 epochs for all models (Fig. 3a). Fig. 3b shows the signature activations for each song arranged chronologically, corresponding to $Q(Z_{i,j} = 1)$, the probability signature j is 1 in the i'th song under the variational posterior (normalized by the mean per-signature activation). All three models exhibit evolutionary signatures with net positive and negative trends over time, as well as signatures with peaks at specific time periods. Figs. 3c and d provide further viewpoints on the models to help interpret the signatures learned. In Fig. 3c, the log of the activation over the K=1 harmonic features is plotted when each signature is turned on in turn. Since the K=1 τ-normalized model uses single chords, there are only 48 possible features corresponding to the 24 major/minor chords at each possible offset from the tonic, under major and minor tonalities. The major tonality activations are visualized (the minor tonality activations showed little variation between signatures, possibly due to a lack of training data in minor keys). A number of prominent features can be observed, such as the primarily diatonic distributions in signature 1, the up-weighting of major chords in signature 4, and the emphasis on *b*VI and *b*VII chords in signatures 3 and 5, with a de-emphasis on the dominant (7) in the latter. Fig. 3d summarizes chordal sequences that receive the largest activations in the K=4 models for each signature, where the non-normalized sequences are notated to begin on C or Cm, and the τ –normalized sequences are notated relative to a tonic on C. The table shows the top 3 sequences from each signature, after removing sequences which are rotationally equivalent. The full ranking for each sequence is provided in the Supplement. Prominent features include signatures emphasizing only primary chords (I, IV, V), signatures involving mixed major and minor chords, signatures involving whole-tone alternations and/or *b*7 emphasis (C,Bb,C,Bb).

3.3 Comparing Models on Prediction Tasks

We use the kernel-regression approach of Sec. 2.3 to compare a variety of architectures for the evolutionary signatures model, as well as to compare the value of the signatures learned against other latent representations and approaches to performing these tasks. In each case, we plot the Pearson correlation of the predicted periods and genres on the hold-out test data as described in Sec. 2.3. For convenience, we compare all models using the K=1 τ-normalized harmonic features as inputs, while varying the size of the latent space, and number of layers in the neural network (NN) generative model (θ).

Fig. 4a shows that performance on both prediction tasks increases during training, although model performance is not necessarily optimized for both tasks at the same training epoch (training shown for $B = 5$, 1 layer NN). Fig. 4b further shows that on both tasks, the



Sig.	K=4, non-norm	K=4, τ-norm
1	C,Bb,C,Bb C,G,C,G C,Bb,F,Bb	G,F,C,F F,C,F,C F,G,F,C
2	C,F,C,G C,G,D,G C,G,D,C	C,F,C,F G,C,F,C F,C,F,G
3	C,G,C,G C,D,G,Em C,F,Dm,Bb	F,G,C,F G,C,F,C Am,F,G,C
4	Cm,Dm,Cm,Dm C,D,C,D C,Dm,C,Dm	C,F,C,G C,G,F,C F,C,G,F
5	C,F,C,F C,F,Bb,F Cm,Gm,Cm,Gm	G,F,G,F C,G,F,G C,Bb,C,Bb

Figure 3. Evolutionary signature interpretation. (a) shows the ELBO bound over 10 epochs of training for three models of evolutionary signatures. (b) plots the normalized evolutionary signature weights over time for the signatures learned. (c-d) provides a summary of the features prioritized by the three models (chords and sequences). See Sec. 3.2 for details.

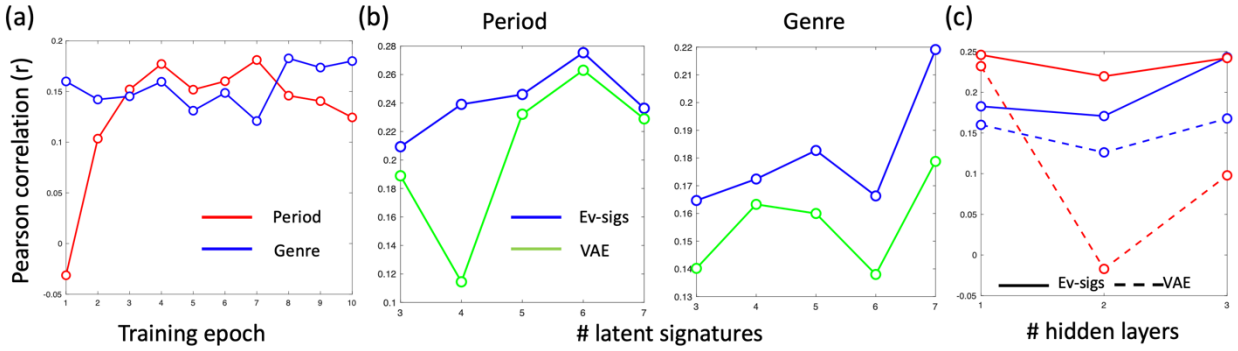


Figure 4. Comparing period and genre prediction performance across models. (a) shows genre and period prediction during training across epochs. (b) compares performance on models with different numbers of latent signatures (evolutionary and VAE models), while (c) compares models including neural nets with different depths. See Sec. 3.3.

evolutionary signatures model outperforms predictions using the latent representations learned by a VAE with the same-sized latent space. Further, the plots suggest that ~6 latent signatures are optimal for generalization in the period prediction task, while more signatures are beneficial for genre classification. We test a maximum of $B = 7$ latent signatures, to allow us to perform exact optimization of $Q(Z_i)$ over all configurations in the latent space for a given song, although using a $Q(Z_i)$ which factors over both songs and signatures would allow us to test larger models.

Fig. 4c then tests the performance of evolutionary signatures and VAE models when the number of layers is varied in the generative NN (fixing $B = 5$ for both); as shown, the evolutionary signatures learned are consistently more informative than the VAE latent representations ($p = 3e - 5$, sign-test across all model comparisons). Further, we tested the performance of the kernel-regression predictor when applied directly to the raw features to the models in Fig. 4. This gave $r = 0.220, 0.217$ for period and genre performance respectively, which are substantially lower than $r = 0.275, 0.243$ for the best performing evolutionary signatures models in Fig. 4 ($p = 2e - 6$ and $p = 1e - 5$ for paired t-tests on the period and genre tasks respectively, using per-instance squared-error and cross-entropy respectively, the latter corresponding to an increase in test classification accuracy from 0.609 to 0.627).

4. Discussion

We have argued that the evolutionary processes underlying developments in musical styles, syntax and genres may be identified by explicitly incorporating evolutionary constraints into a generative model of a musical corpus. We are motivated by recent techniques in evolutionary genomics, which have identified 'mutational signatures' that underlie evolutionary processes in the context of cancer genomics. For this purpose, we propose a model of latent evolutionary signatures, which learns a latent 'code' for each song in a corpus, and jointly optimizes the codes to reconstruct the observed musical features and respect an underlying evolutionary structure. We note, however, two differences between our approach and mutational signature models in genomics: First, mutational signatures are not explicitly optimized to incorporate evolutionary constraints ((7, 8) use PCA components); Second, mutational processes generate variation in genotypes, rather than acting directly as 'codes' themselves (although they may be viewed as latent factors in a two-level evolutionary process, as discussed in (40)). Finally, we trained our model on a variety of harmonic features extracted from the McGill Billboard dataset, and showed that incorporating latent evolutionary structure led to more informative features (than matched VAE models) for the tasks of period and genre prediction. We conclude by discussing the relevance of our model in both musical and evolutionary contexts.

Musicological Interpretation. Several features of the evolutionary signatures we extract are consistent with previous analyses of the McGill Billboard dataset. The decrease over time of our signature 4 from Fig. 3c is consistent with observations in (37) that minor chord usage increases with time over the dataset, and the increase over time of our signatures 3 and 5 from Fig. 5c, and signatures 4 and 5 from Fig. 5d for the non-normalized and τ-normalized models respectively, corresponds to the observation in (31) that *b*VII is becomes increasingly important, taking on the role of a substitute dominant in later periods. The decrease of our signature 1 from Fig. 3c with time, corresponding to traditional diatonic harmony, is also consistent with the increased use of modal harmonies and transitions noted in (31). In our model, these trends are tied to evolutionary signatures which link features that act together (for instance, *b*VII is tied to distinct signatures (3 and 5) in Fig. 3c, the former retaining a dominant emphasis and containing a moderate weight on *b*VI, while the latter reverses this weighting).

Deep Memetics. As noted, the influence of musical (and cultural) entities on one another may be highly varied. For an evolutionary process to take shape, though, there must be some regularity in the transmission of 'cultural units' or 'memes'. The problem of identifying these

units *a priori* may be viewed as an intractable task (9, 22, 23, 25). Our model demonstrates that, instead, such features can be identified statistically through the process of fitting an evolutionary model to observed data. In this way, we do not need to explicitly define the memes underlying the process, but may discover them *de novo* through model fitting. In general, we do not expect these units to be directly observable as surface features of the phenotypes in question. We thus focus on discovering underlying *deep memes*, which live in a latent space (either continuous or discrete, where the latter may also be viewed as a code), and whose relationship to the observed phenotypes may in general be highly complex, modeled by an arbitrary neural network. In this way, the meme-phenotype relationship is directly analogous to the gene-phenotype relationship, both involving a complex generative process. As we described in Sec. 2, a latent evolutionary process may thus be formulated as a deep latent model, whose latent variables respect evolutionary relationships defined by an underlying ancestral graph (which may itself be discovered) between the modeled entities through similarity. The degree to which the deep memes so discovered by fitting models of this kind are explanatory, may be assessed through statistically testing the model fit against matched models which do not include evolutionary structure. Such techniques may be further explored in other cultural and biological contexts, for instance, language evolution, and cancer mutational processes.

Characterizing musical change as an evolutionary process. As discussed, to define a musical evolutionary process, we require a relationship corresponding to the ancestral or parental relationship in biological evolution. In our model, we propose that the appropriate relationship is one of *influence* (or *mimesis*, see (14, 41)), which may be defined in information theoretic terms (40). The nature of the influence may be highly varied, for instance some parent songs have merely a stylistic influence on their offspring, while others influence particular motifs and phrases. Further, such a web of influences may generate a neutral or adaptive evolutionary process. Formally, an *adaptive* evolutionary process requires there to be a dependency of the number of offspring on a heritable phenotype of an individual (either a single phenotype, or a combination of phenotypes). In contrast, a *neutral* evolutionary process may still exhibit changes in the phenotypes of individuals, but these are due to statistical sampling effects (drift) rather than systematic dependencies between phenotype and number of offspring. In the context of musical evolution, such a definition translates to the net influence of a particular song or work on others. The type of dependencies with respect to phenotype that may be relevant here include factors such as cultural tastes, memorability and emotional salience or valence. Environmental factors may also play a role (for instance, popularity of the artist), but these will not lead to adaptive processes, since they are not heritable. Further, we note that while factors such as the chart ranking or radio air-time may be treated as proxies for fitness of a song, they are themselves not equivalent to net influence (for instance, a relatively unpopular artist may have high fitness by being a 'songwriter's songwriter' or a 'composer's composer'). Finally, evolutionary processes may be single or multileveled, and frameworks for *group* and *multilevel selection* may naturally be applied to evolution of different kinds of musical entity (40, 42). Musical genres, for instance, may on one level be regarded as similar to 'musical species', where influence relations tend to be stronger within genres than between them. At another level though, genres themselves may be viewed as more or less fit, having the potential to generate more or less further genres (or styles) historically (type II multilevel selection, see (42). Similarly, an artist's output as a whole may be regarded as a 'group' with its own fitness, which may both enhance the fitness of the component songs (type I multilevel selection), or the number

of subsequent artists influenced by the artist's oeuvre as a whole (type II). Our latent evolutionary framework may be naturally extended to further distinguish between different types of evolution in music and other cultural spheres, for example by introducing a fitness parameter, or fitting a two-level process to identify genre-level effects (see (40)).

Further extensions and applications. We finally note some further extensions to our model. First, although we focus on harmonic features in our analysis, the model may be applied to any domain (for instance, melody, rhythm), as well as cross-domain analysis. In addition, our approach may be extended so that the underlying graph G is learned jointly with the other model parameters. Optimizing G provides a means of detecting influences between songs, and learning fine-grained evolutionary signatures that reflect specific influences (as opposed to using the coarse approximation of influence due to temporal proximity). Such a model also provides a context for exploring features which are novel to particular songs and artists, in the context of evolutionary innovation (through mutation and neutral processes, see (43)). Finally, as we show, the latent representations learned by our model may be used for many tasks of interest. A particularly interesting application is recommender systems (29), where an individual's taste provides a selective environment in which a predictive evolutionary system may learn and interact.

5. References

1. M. Lynch, *et al.*, Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* (2016) https:/doi.org/10.1038/nrg.2016.104.
2. L. Salichos, A. Rokas, Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* (2013) https:/doi.org/10.1038/nature12130.
3. A. C. Syvänen, Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* (2001) https:/doi.org/10.1038/35103535.
4. P. Simmonds, P. Aiewsakun, A. Katzourakis, Prisoners of war — host adaptation and its constraints on virus evolution. *Nat. Rev. Microbiol.* (2019) https:/doi.org/10.1038/s41579-018-0120-2.
5. L. Salichos, W. Meyerson, J. Warrell, M. Gerstein, Estimating growth patterns and driver effects in tumor evolution from individual samples. *Nat. Commun.* (2020) https:/doi.org/10.1038/s41467-020-14407-9.
6. S. Kumar, *et al.*, Passenger Mutations in More Than 2,500 Cancer Genomes: Overall Molecular Functional Impact and Consequences. *Cell* (2020) https:/doi.org/10.1016/j.cell.2020.01.032.
7. L. B. Alexandrov, *et al.*, The repertoire of mutational signatures in human cancer. *Nature* (2020) https:/doi.org/10.1038/s41586-020-1943-3.
8. L. B. Alexandrov, *et al.*, Signatures of mutational processes in human cancer. *Nature* (2013) https:/doi.org/10.1038/nature12477.
9. A. Mesoudi, How Cultural Evolutionary Theory Can Inform Social Psychology and Vice Versa. *Psychol. Rev.* (2009) https:/doi.org/10.1037/a0017062.
10. P. E. Savage, Cultural evolution of music. *Palgrave Commun.* (2019) https:/doi.org/10.1057/s41599-019-0221-1.
11. J. Klüver, The socio-cultural evolution of our species. The history and possible future of human societies and civilizations. *EMBO Rep.* (2008)

https:/doi.org/10.1038/embor.2008.35.
12. J. Russell, R. Cohn, *Social Cycle Theory* (Book on Demand, 2012).
13. K. Marx, F. Engels, "The German ideology" in *Social Theory Re-Wired: New Connections to Classical and Contemporary Perspectives: Second Edition*, (2016) https:/doi.org/10.4324/9781315775357.
14. R. Dawkins, *The selfish gene* (Oxford University Press, 1976).
15. , The Skeptic encyclopedia of pseudoscience. *Choice Rev. Online* (2003) https:/doi.org/10.5860/choice.40-4360.
16. M. Midgley, *The solitary self: Darwin and the selfish gene* (2010) https:/doi.org/10.1017/UPO9781844654833.
17. P. Savage, S. Brown, Toward a new comparative musicology. *Anal. Approaches to World Music* (2013) https:/doi.org/10.31234/osf.io/q3egp.
18. P. E. Savage, S. Brown, E. Sakai, T. E. Currie, Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U. S. A.* (2015) https:/doi.org/10.1073/pnas.1414495112.
19. M. Rahaim, What else do we say when we say "music evolves?" *World Music* (2006).
20. S. C. Levinson, R. D. Gray, Tools from evolutionary biology shed new light on the diversification of languages. *Trends Cogn. Sci.* (2012) https:/doi.org/10.1016/j.tics.2012.01.007.
21. J. Fracchia, R. C. Lewontin, The price of metaphor. *Hist. Theory* (2005) https:/doi.org/10.1111/j.1468-2303.2005.00305.x.
22. J. Fracchia, R. C. Lewontin, Does culture evolve? *Hist. Theory* (1999) https:/doi.org/10.1111/0018-2656.00104.
23. W. G. Runciman, Forum: Does Culture Evolve? *Hist. Theory* (2005).
24. S. Pinker, "The False Allure of Group Selection" in *The Handbook of Evolutionary Psychology*, (2015) https:/doi.org/10.1002/9781119125563.evpsych236.
25. R. Spencer, The missing unit of inheritance: Cultural Evolution's elephant in the room. *J. Evol. Psychol.* (2012) https:/doi.org/10.1556/jep.10.2012.3.4.
26. M. Mauch, R. M. MacCallum, M. Levy, A. M. Leroi, The evolution of popular music: USA 1960–2010. *R. Soc. Open Sci.* (2015) https:/doi.org/10.1098/rsos.150081.
27. E. Nakamura, K. Kaneko, Statistical Evolutionary Laws in Music Styles. *Sci. Rep.* (2019) https:/doi.org/10.1038/s41598-019-52380-6.
28. T. Parmer, Y. Y. Ahn, Evolution of the informational complexity of contemporary western music in *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019*, (2019).
29. M. Quadrana, M. Reznakova, T. Ye, E. Schmidt, H. Vahabi, *Modeling Musical Taste Evolution with Recurrent Neural Networks* (2018).
30. J. A. Burgoyne, J. Wild, I. Fujinaga, An expert ground-truth set for audio chord recognition and music analysis in *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011*, (2011).
31. C. W. White, I. Quinn, Chord Context and Harmonic Function in Tonal Music. *Music Theory Spectr.* (2018) https:/doi.org/10.1093/mts/mty021.
32. L. B. Alexandrov, M. R. Stratton, Mutational signatures: The patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* (2014) https:/doi.org/10.1016/j.gde.2013.11.014.
33. M. Gerstung, *et al.*, The evolutionary history of 2,658 cancers. *Nature* (2020)

https:/doi.org/10.1038/s41586-019-1907-7.
34. D. P. Kingma, M. Welling, Auto-encoding variational bayes in *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, (2014).
35. V. Fortuin, M. Hüser, F. Locatello, H. Strathmann, G. Rätsch, Som-Vae: Interpretable discrete representation learning on time series in *7th International Conference on Learning Representations, ICLR 2019*, (2019).
36. M. Amodio, *et al.*, Abstract 5306: SAUCIE: Sparse autoencoder for unsupervised clustering, imputation, and embedding in (2018) https:/doi.org/10.1158/1538-7445.am2018-5306.
37. J. A. Burgoyne, J. Wild, I. Fujinaga, Compositional data analysis of harmonic structures in popular music in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, (2013) https:/doi.org/10.1007/978-3-642-39357-0_4.
38. D. Koller, N. Friedman, *Probabilistic Graphical Models: Principles* (1974) https:/doi.org/10.2307/478142.
39. D. M. Blei, A. Kucukelbir, J. D. McAuliffe, Variational Inference: A Review for Statisticians. *J. Am. Stat. Assoc.* (2017) https:/doi.org/10.1080/01621459.2017.1285773.
40. J. Warrell, M. Gerstein, Cyclic and Multilevel Causation in Evolutionary Processes. *bioRxiv*, 830422 (2019).
41. E. Jablonka, M. J. Lamb, Z. Anna, *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life* (2014) https:/doi.org/10.1172/jci27017.
42. S. Okasha, *Evolution and the Levels of Selection* (2007) https:/doi.org/10.1093/acprof:oso/9780199267972.001.0001.
43. G. P. Wagner, *Homology, genes, and evolutionary innovation* (2014) https:/doi.org/10.5860/choice.52-0829.