The Global Jukebox: A public database of performing arts and culture

Preprint · May 2021

DOI: 10.31234/osf.io/4z97j

18 authors, including:

Anna Wood
Association for Cultural Equity at Hunter College
10 PUBLICATIONS **10** CITATIONS
SEE PROFILE

Carol Ember
Yale University
112 PUBLICATIONS **2,545** CITATIONS
SEE PROFILE

Hideo Daikoku
Keio University
5 PUBLICATIONS **2** CITATIONS
SEE PROFILE

John Mcbride
Ulsan National Institute of Science and Technology
9 PUBLICATIONS **7** CITATIONS
SEE PROFILE

The Global Jukebox: A Public Database of Performing Arts and Culture

Anna L.C. Wood[1*], Kathryn R. Kirby[2], Carol R. Ember[3], Stella Silbert[1], Hideo Daikoku[4], John McBride[5], Sam Passmore[4], Forrestine Paulay[6], Michael Flory[7], John Szinger[1], Gideon D'Arcangelo[8], Marco Guarino[1], Maisa Atayeva[1], Jesse Rifkin[1], Violet Baron[1], Miriam El Hajli, Martin Szinger[1], Patrick E. Savage[4*]

[1]Association for Cultural Equity (ACE), Hunter College, New York City, USA
[2]D-PLACE, University of Toronto, Canada
[3]Human Relations Area Files at Yale University, New Haven, CT, USA
[4]Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan
[5] Institute for Basic Science, South Korea
[6] ACE; Laban Institute for Movement Studies, New York, NY, USA
[7] NYS Institute for Basic Research in Developmental Disabilities
[8] Arup, New York, NY, USA
*Correspondence to: annalwood@gmail.com; psavage@sfc.keio.ac.jp

Please note: This is a non-peer-reviewed preprint. We welcome questions, comments, citation, and constructive criticism, bearing in mind that this is a non-peer-reviewed draft subject to revision. Supplementary materials will be added in future updates, at which time the data repository will be made public. Please direct correspondence to annalwood@gmail.com and psavage@sfc.keio.ac.jp.
Recommended citation: Wood ACL, Kirby KR, Ember CR, Silbert S, Daikoku H, McBride J, Passmore S, Paulay P, Flory M, Szinger J, D'Arcangelo G, Guarino M, Atayeva M, Rifkin J, Baron V, El Hajli M, Szinger M, Savage PE (2021) The Global Jukebox: A Public Database of Performing Arts and Culture. *PsyArXiv* preprint

Abstract

The lack of standardized cross-cultural databases has impeded scientific understanding of the role of the performing arts in other domains of human society. This paper introduces the Global Jukebox (theglobaljukebox.org) as a resource for comparative and cross-cultural study of the performing arts and culture. Its core is the Cantometrics dataset, encompassing standardized codings on 37 aspects of musical style for 5,779 traditional songs from 992 societies. The Cantometrics dataset has been cleaned and checked for reliability and accuracy. Also being released are seven additional datasets coding and describing instrumentation, conversation, popular music, vowel and consonant placement, breath management, social factors, and societies. For the first time, all digitized Global Jukebox data are being made available in open-access, machine-readable format, linked with streaming audiovisual files to the maximum extent allowed while respecting copyright and the wishes of culture-bearers. The data are cross-indexed with the Database of Peoples, Languages, and Cultures (D-PLACE) to allow researchers to test hypotheses about worldwide aesthetic patterns and traditions, including earlier findings by Alan Lomax and his research team regarding coevolutionary relationships between the performing arts, social structure and cultural history. The

Global Jukebox adds a large and detailed global database of the performing arts to enlarge our understanding of human cultural diversity.

Citation: Wood ACL, Kirby KR, Ember CR, Silbert S, Daikoku H, McBride J, Passmore S, Paulay P, Flory M, Szinger J, D'Arcangelo G, Guarino M, Atayeva M, Rifkin J, Baron V, El Hajli M, Szinger M, Savage PE (2021) The Global Jukebox: A Public Database of Performing Arts and Culture. *PsyArXiv* preprint.

Data and materials availability: All coded data and source code for data conversion/analysis are available at https://github.com/comp-music-lab/global-jukebox [to be made public soon]. All audiovisual files are available for streaming at http://theglobaljukebox.org, with some restrictions as explained in the text.

The datasets are archived with ZENODO, and the DOI provided by ZENODO should be used when citing particular releases of D-PLACE data (http://doi.org/ [to be updated]).

Funding: The Global Jukebox has been developed with support from the National Endowment for the Arts, the National Endowment for the Humanities, the Concordia Foundation, the Rock Foundation, and Odyssey Productions. PES, HD, and SP are supported by funding from the Yamaha corporation, a Grant-in-Aid from the Japan Society for the Promotion of Science (#19KK0064), and by grants from Keio University (Keio Global Research Institute and Keio Gijuku Academic Development Fund). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

1.Introduction

During the 20th century, anthropologists began organizing data on cross-cultural diversity in ways that could be systematically compared on a global scale. The *Ethnographic Atlas* [1] coded data on social structure, kinship, religion, and economy; the Human Relations Area Files (HRAF) [2] compiled and subject-indexed detailed ethnographic texts; Ethnologue [3] and Glottolog [4] catalogued linguistic diversity. These resources allow scientists to quantitatively test cross-cultural hypotheses using global data. The 21st century has seen a resurgence of interest in such global databases, now digitized, cross-compatible and many available online, stimulating new research and debate on the nature of cross-cultural diversity and the role of music in human evolution [5-12]. Alan Lomax and Conrad Arensberg's Expressive Style Research Project at Columbia University, which also began in the mid-20th century, was intended to complement resources like the *Ethnographic Atlas* for the domain of the performing arts by integrating cross-cultural classification schemes such as "Cantometrics" and "Choreometrics" ("canto" = song, "choreo" = dance, "metrics" = measure) through an interactive "Global Jukebox" [13-18]. However, while several subsequent studies have used methods and samples modeled after Cantometrics to analyze hundreds of traditional music recordings from around the world [19-23], the full Global Jukebox sample of data of over 5,000 coded performances was never made publicly available until now.

The Global Jukebox is an interactive online resource for exploring music and other performing arts cross-culturally. On it, the immediacy of field recordings representing the full range of the world's music can be experienced with reference to ethnological, ethnographic, environmental, linguistic, and geographic context, with song lyrics and testimonials by first-hand observers, musicians, and culture members. The Jukebox thus bridges the sciences and the humanities. It encompasses thousands of examples of singing, dancing, speaking, instrumentation, and other performing arts from over 1,000 societies, transformed into an online form that can be used for research, education, and cultural activism. Each example is classified and coded by aesthetic and organizational features that can be compared cross-culturally, making it possible to explore relationships between music, dance, speech, social life and the environment.

This article announces the long-anticipated publication of the Global Jukebox and the raw coded data in downloadable form of Cantometrics and six additional studies and two supporting datasets (see Table 1), on https://theglobaljukebox.org/ and on https://d-place.org/. By releasing these data to the public on two accessible web platforms, we hope to enrich scientific data on the expressive arts, to support cultural diversity, and to facilitate the practice of cultural equity in homes, classrooms, and research organizations.

2. Theoretical Background

Scientific cross-cultural comparison of music had been conducted since the late 19th century when the invention of the phonograph made relatively objective cross-cultural comparison of sound possible for the first time, birthing the field of comparative musicology, which later became known as "ethnomusicology" [24-29]. But until Alan Lomax's research on expressive culture, such comparisons were generally limited to relatively small samples of recordings and tended to emphasize aspects of pitch and rhythmic structure that are privileged in Western staff notation. In intellectual partnership with the anthropologist Conrad Arensberg, Lomax worked with several multidisciplinary teams from the 1960s through the mid-1990s to collect and analyze thousands of examples of recorded music, dance, and conversation from all world regions using a radically different approach that, in addition to traditional musical features, emphasized aspects of performance style and social interaction through sonic indicators of emotional states, social cohesion and differentiation, synchrony, and group organization [14-18] (see S3 for details).

The principal research goals were to learn (1) whether equivalent performance patterns appear in different classes of performance, and (2) whether such codes, or markers of performance style, relate to the fundamentals of human society. Lomax theorized that performance style is related to social structure and evolutionary history. These hypotheses found support through statistical analysis of thousands of examples from hundreds of societies (though cf. [14, 18] for discussion of methodological criticisms). Figure 1 shows an adapted version of a diagram that Lomax used to illustrate his discovery of large clusters, or mega-families, of performance style, markedly consistent with patterns of ancient human settlement, subsistence, and migration [Table S2].

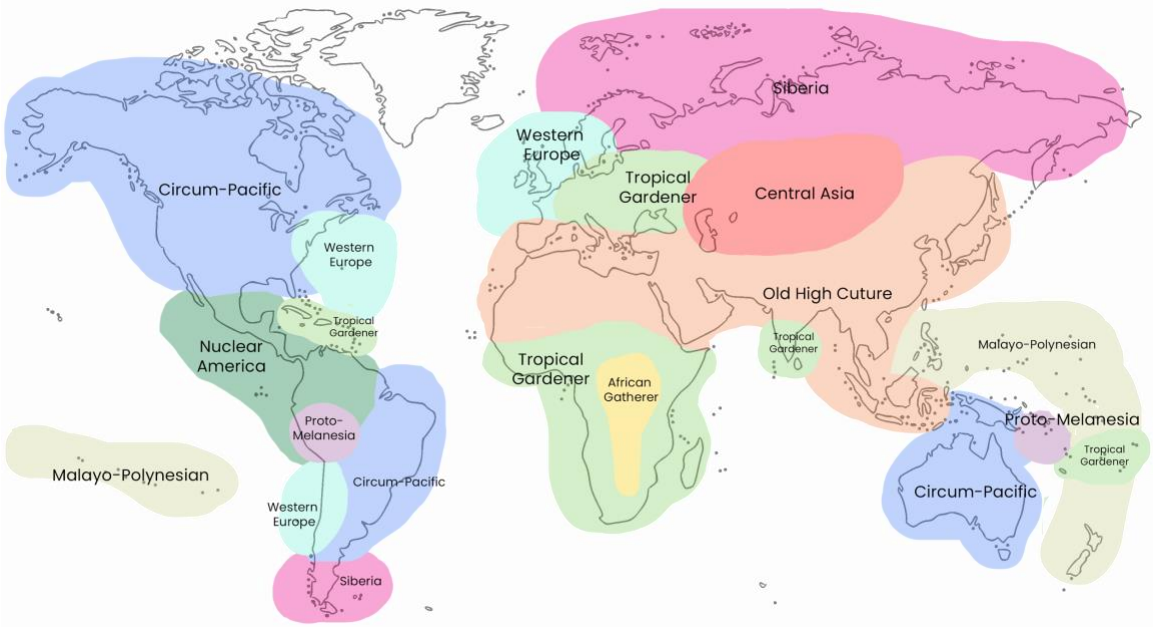

Figure 1. Global map of 10 song-style regions identified by factor analyses of Cantometrics data (adapted from [15]).

3. The Data

3.1 Background

In the 1980s, the multiple datasets were incorporated into a single multimedia, relational database that Lomax called the Global Jukebox, which would synthesize and complete the multi-decade project. Lomax's poor health, technological limitations, and the intellectual climate of the eighties and nineties cut these plans short. In 2010, ALCW, an anthropologist, and GD'A (Arup), a world leader in architectural digital design, with a team of ethnomusicologists, designers and programmers, stepped in to realize the project.

3.2..Datasets of the Global Jukebox

Foundational to the Jukebox are eleven cross-cultural studies on distinct but related aspects of the performing arts (Table 1; see full descriptions in Supplementary Material S3.1 and Table S4 for "In progress" datasets not included in the current release). This article focuses on describing the largest and most comprehensive dataset, **Cantometrics** , which uses 37 features to code musical style for 5,779 audio recordings of traditional songs from 992 societies. Where possible, Cantometrics was sampled from the more than 1,200 societies for which cultural data had already been coded in George P. Murdock's Ethnographic Atlas to facilitate comparison of song style and social structure [14, 17, 18] .

We also introduce several datasets representing more detailed analyses of performance style using many of the audio recordings coded by Cantometrics, as well as additional recordings: **Minutage** uses 36 additional variables to code breathing and phrasing patterns for 687 songs from 113 societies, while **Phonotactics** uses 51 variables to code vowel and consonant use for 338 songs from 45 societies. In addition, we include the **Instruments** dataset, which uses 14 different

variables to code structural and functional aspects of 1,780 instruments from 152 societies, **Ensembles** with 776 cases from 153 societies, the **Urban Strain** (a popular song study**)** with 378 songs and dances from North America, and the **Parlametrics** dataset, which uses 52 variables to code aspects of spoken conversation for 188 audio recordings from 158 societies (Table 1). Other datasets in the process of being digitized and prepared for release in a future publication are Choreometrics, Personnel and Orchestra, Song Texts, Vocal Qualities, Popular Songs (an extension of the Urban Strain set not yet coded), and several derivative studies on song style and social structure (see S3 for details).

Table 1. Global Jukebox Datasets Included in this Public Release

Dataset	Description	Variables	Performances/ Cases	Societies
1.Cantometrics	Aspects of singing style, song structure, social organization of singers and relationships between vocal and instrumental parts	37	5,779 songs	992
2.Minutage	Phrasing and breath patterns in song	36	687 songs	113
3.Phonotactics	Vowel and consonant frequency patterns in singing	51	338 songs	45
4.Ensemble Study	Bibliographic information on ensembles worldwide	12	776 ensembles	153
5.Instrument Study	Bibliographic information and classification of instruments across ensembles worldwide	14	1,780 instruments	152
6.Parlametrics	Conversational style	52	188 conversations	158
7.Urban Strain	North American popular music and dance styles	18	378 popular songs	178*
Supporting Datasets				
Social Factors	Cross-cultural sample coded for geography, population size, subsistence, political structure, gender roles, kinship & family structure, property, social stratification, sexuality, games, theology (adapted and expanded from early versions of the Ethnographic Atlas)	38	1,310 societies	1,310
Societies (Cultures)	Ethnographic descriptors for sampled societies: lat/long, geographic, ethnological,	50	1,246 societies	1,246

	linguistic, climate, & terrain classes; historical subsistence, sources.			

*To be categorized separately as the classification of urban musical cultures requires a different scheme from the one used in cross-cultural research on traditional societies.

Finally, we are publishing data on social structure adapted by Lomax, Arensberg, Barbara Ayres and their colleagues from Ethnographic Atlas data and used in Lomax et al.'s previous analyses of relationships between music and culture ("**Social Factors**" and "**Societies**"). These data appear in updated versions of the Ethnographic Atlas now available online as part of the Database of Places, Language, Culture and Environment (D-PLACE; d-place.org)[8]. Their inclusion here will allow replication or extension of Lomax et al.'s original analyses.

In order to further extend the types of analyses possible with each dataset, careful efforts have been made to verify and make explicit: (1) the year(s) in which the song recordings for each society were made; (2) the geographic coordinates of the recorded society; and (3) the language (or dialect, where available) spoken by the society (as indicated by a glottocode [4]). The data is stored in a series of CSV tables, within a cross-linguistic data format (CLDF) framework [30], which are released via Zenodo. The audio recordings of songs and their accompanying Cantometrics codings are available in interactive format at http://theglobaljukebox.org (see Section 3.4 for caveats regarding restrictions on audio recording availability).

3.3. Selection of Audio Examples

The Global Jukebox is based on analysis of audiovisual recordings (song, dance, conversation, etc.). Lomax sampled primarily folk and indigenous songs for Cantometrics, although small samples of traditional art songs were also included (e.g., Hindustani/Carnatic traditions of South Asia; Central European Baroque and other classical singing; various examples of Chinese opera, Javanese Gamelan, modern jazz, etc.). In his own research practice, Lomax sought out the oldest and most typical songs and performance styles because he believed they would shed light on how singing voices the touchstones of emotion and personality development in a culture [18]. He also held extensive consultations with the singers and culture holders he recorded, which in many instances amounted to mini- or full autobiographical accounts (e.g., "Mr. Jelly Roll", a full-length curated autobiography [31] and Interviews with Bessie Jones at www.archive.culturalequity.org [32]; see [33] for a book-length review). Representativeness within a tradition, performance quality, and aesthetic criteria were important in selecting songs for Cantometrics, Lomax made every effort to include women's songs and female performances, and even those of children, although these weren't always available at the time. Final selections were based on subjective but informed decisions by Lomax and experts who recorded the songs. Lomax also drew upon his own extensive fieldwork, from oral histories, recording liner notes, and from the accounts of ethnographers and collectors.

An assumption of Cantometrics, based on Lomax's field experience and early listening sessions with Victor Grauer (co-inventor of Cantometrics with Lomax), was that the same features of song style would appear throughout all or most of the genres of singing in a society. Extensive piloting

during the development of the Cantometric song sample led Lomax and Grauer to conclude that approximately ten songs were sufficient to capture the key elements of a song style in most folk and indigenous societies. This assumption has been critiqued and it could be systematically retested, but it is debatable whether a better option was available given the limitations of available recordings and resources to devote to manually listening to and coding these recordings (cf. discussion in [14]).

3.4. Availability of Audiovisual Recordings

Alan Lomax had research and publishing agreements with collectors, filmmakers, repositories, and publishers, and the Association for Cultural Equity (ACE) has obtained or requested streaming rights for the Global Jukebox. But distributing such recordings presents challenges in terms of copyright and respecting the wishes of the societies who have made the recordings. It is therefore not possible to make most of the recordings of songs freely available for download. However, we have negotiated agreements to allow 5,134 of the songs to be available for listening via streaming on the Global Jukebox website. We are in the process of requesting permission from 60 North American and 15 Australian indigenous societies to stream 457 songs (we have been unable to locate audio for the remaining 188 coded songs) and will make these available as requests are granted (not all will necessarily agree).

3.6. Choice of Coded Performance Variables

Variables for Choreometrics, Cantometrics, Phonotactics, Parlametrics, and Minutage were developed in field observation, intensive listening/viewing, and experimentation. Lomax and his collaborators looked for performance qualities with significant worldwide ranges and with a role in shaping performance traditions. Variables requiring fine distinctions to be made were discarded. Such work required intensive, repeated listening, viewing, and cross-comparison. Lomax and Grauer [34] transcended the Western musical canon to probe cohesiveness and differentiation among performers, synchrony, social and musical organization, vocal timbre, rhythmic relationships within and between orchestra and voice, and emphasis on text and ornament. Together with melodic, rhythmic and structural features, these factors outline a broad matrix of aesthetic and social codes, or conventions that are fairly consistent at the cultural level. Following a similar methodology, specialist teams developed parameters for coding dance, speech and additional aspects of music cross-culturally, producing multiple datasets. The inclusion of speech in this project, alongside forms more typically recognized as art, like song and dance, reflects Lomax's intentions to study the cross-cultural aesthetic patterns of expressive culture in its various forms rather than on the basis of conventional definitions of art. Full details for most codings systems are being republished or published for the first time [34, 35]. The Global Jukebox website contains a fully digitized training course ("Songs of Earth: Aesthetic and Social Codes in Music"), with recordings and coding guides to allow researchers to interpret existing Cantometrics codings and to become trained Cantometric coders capable of adding new codings. New introductions, coding guides, and results for the other datasets are available in [35].

3.7. Performance Data Sources

Performance data came from recorded examples of song and speech, and filmed examples of dance and movement. As a documentarian himself, Alan Lomax had great faith in the recorded (and

filmed) medium and what it could communicate. Lomax already had a substantial library of world music on records and field tapes recorded himself or sent to him by colleagues, but he and Grauer spent a year acquiring more recordings to fill in the gaps of their first sample of over 2,000 recordings. Only later, in stages, was it possible to obtain material from Eastern Europe, the former Soviet Union, rural China, India, and the Pacific.

3.8. Performance Metadata

Metadata for each coded song performance includes a unique coding identification number; source society; audio reference numbers and information; song and performer information; setting and context if known; collector, publishing and archival data and year recorded; comments; source tags; and additional descriptive information. Figure 2 is a screenshot of metadata and codings for one song. Metadata for Cantometrics, Phonotactics, Minutage, Parlametrics, and Instruments and Ensembles are summarized in Table S5 in Supplementary Material.

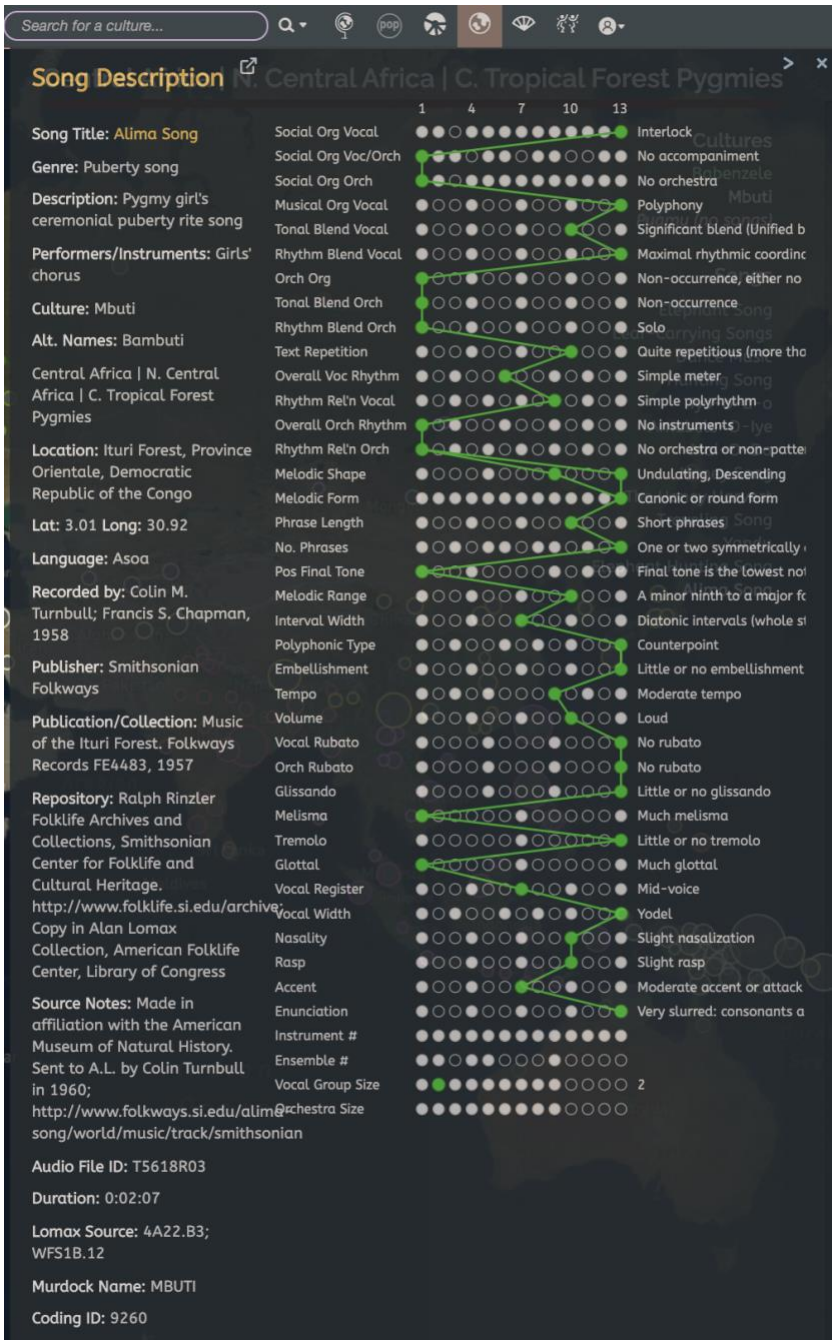

Figure 2. Screenshot of metadata and Cantometric codings of the Mbuti "Alima Song". For detailed explanation of the definitions of each of the 37 Cantometric variables shown here, see [34, 35] and the online training examples in the "Songs of Earth" section at http://theglobaljukebox.org.

4. Sampling Societies and Links to Other Datasets

4.1. Sampling Societies. The Jukebox is designed to be an experience of the expressive arts as *cultural* phenomena. We use the term "societies" as an alternative to the term "cultures" or other terms for defining cultural groups. The datasets (songs, instruments, conversation, etc.) are related through their source societies. The Global Jukebox follows the *Ethnographic Atlas* in adopting

ethno-linguistically defined societies as a key unit of analysis. The Jukebox differs from the *Ethnographic Atlas* in that in most cases a given society is not represented by a single set of coded data, but instead contains multiple examples of different performances. For example, the Mbuti society is represented by 12 different song samples, each of which has its own set of Cantometric codings (cf. Figure 2). Performances (songs, dances, etc.) thus represent a finer unit of analysis beneath the larger unit of societies, and make it possible account for local/regional historical, sociological and aesthetic influences, as well as for musical diversity within societies. [36, 37]. Societies can also be related to one another at higher levels of organization (e.g., people, Koppen climate/terrain [38], language family, geographic region, etc.; see Fig. S5). For visualization on the maps and linguistic trees of D-PLACE, songs and other performance data are condensed into a single "modal profile" for a given society (cf. [8] for examples of how D-PLACE can be used to explore patterns in culture and their relationship to population history (via language) and geography.

A total of 1,246 indigenous and folk societies across thirteen world regions are represented in the Global Jukebox, plus 472 popular song cultures. This accounts for societies that were sampled in the primary studies listed in Table 1, with the exception of Instruments and Ensembles. Of the total number of Global Jukebox societies, 1,195 are linked with coded data (including Choreometrics data coding dance style, which will be published in the future), and of that number, 992 are included in the Cantometrics set. Of the 992 societies sampled in the Cantometrics study, XXX are from the Ethnographic Atlas, while YYY have been added to fill sampling gaps in in some areas not represented in the *Ethnographic Atlas* (parts of Europe, the Caribbean, China, India, and Euro- and Afro-America). Most of these new cultures were then coded for selected Ethnographic Atlas features, released here as part of the 'Social Factors' dataset.

The cases included in the"Ensembles" and "Instruments" datasets depart from the other datasets in that they only partially conform to our definition of society. Because these studies use bibliographic sources rather than specific audio recordings, the societal designations made by the original investigators were often necessarily much broader than those in the other datasets (they 'lump' many societies that are 'split' in other datasets). Future research with scholars who specialize in musical instruments will be necessary to match the Instruments/Ensembles data with our more detailed societal data.

4.2. Links to Other Cross-Cultural Datasets

One goal of this release of Lomax's datasets is to enable researchers to map and reanalyze the kinds of relationships between the arts and society originally explored by Lomax (cf. Figure 4 for an example comparing distributions of song style and social structure). In addition to the xxx Lomax societies represented in the Ethnographic Atlas, integration with D-PLACE will allow each society to be matched to corresponding societies in the Binford Hunter-Gatherer, Western North American Indian, Standard Cross Cultural Sample, and eHRAF World Cultures datasets (and other cultural datasets as they are integrated to D-PLACE) [39]. Over the past several years, Ethnographic Atlas data sources have been revised and brought up to date (see D-PLACE [d-place.org] [8]), making it easier for researchers to return to primary sources when a coded cultural variable of interest is not available. eHRAF World Cultures [ehraf world cultures.yale.edu] [2]) greatly facilitates this

process, by providing subject-indexed, fully digitized and searchable primary ethnographic documents. Currently, xxx Jukebox societies are covered by eHRAF, and HRAF plans to include documents that match the time and place of the Cantometrics data over the next two years as part of eHRAF World Cultures.

5. Data curation, Cleaning, and Validation of the Cantometric dataset

Each of the 7 Global Jukebox datasets and the Social Factors supporting dataset listed in Table 1 have their own coding schemes and criteria. Fully cleaning each dataset takes a great deal of effort and resources, and we currently cannot fully clean all datasets for publication. We decided to focus our cleaning efforts on the largest and most influential dataset of 5,779 Cantometric codings of songs (documented below). We decided to simultaneously publish the other 6 partially cleaned datasets listed in Table 1 in order to make the materials available as widely as possibly, but wish to emphasize that only the Cantometric dataset has been fully cleaned and validated.

Supplementary Material S6 provides a detailed description of the data curation, digitization, cleaning, and validation process. In such a large database of subjective codings, we cannot guarantee complete absence of errors or disagreement. However, our analyses suggest that both coding reliability (mean $\kappa = 0.54$) and validity (~0.4% error rate) are at acceptable levels (see S5 and Table S6 for details).

6. Analysis

Lomax and colleagues published numerous analyses of Cantometrics and other Global Jukebox data. These results were reviewed by Wood [17, 18] and Savage [14], and are summarized in S2. Unfortunately, the original analyses cannot be exactly reproduced, as it is not possible to reconstruct the exact datasets and procedures that were used for specific previous analyses before data/code-sharing services were widely available. For reanalyses of Cantometric data and analyses of similar data, see [19, 21, 22, 36, 37, 40-45].

The range in the original analyses and reanalyses show that there are many ways to analyze this rich dataset. For example, a given society may contain a stylistically diverse set of songs with substantially differing codings. Such diversity can be condensed into a single "modal profile", average similarity metric, etc. in order to analyze relationships at the unit of society, or clustered and analyzed separately at the unit of individual songs [21-23, 40-46]. An example modal profile is mapped in Fig. 3 for the Cantometric variable "Embellishment". Cantometrics data can also be refined and analyzed to capture musical diversity within societies, although it is essential to account for the nature of such diversity and to ascertain if the same patterns appear in all or most of the societies under consideration [36].

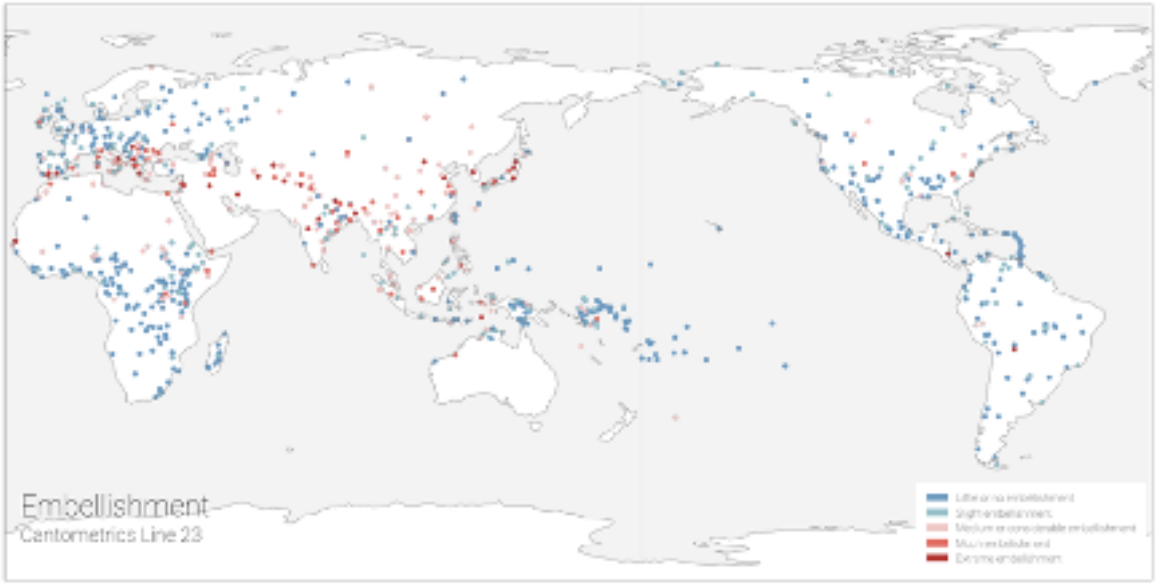

Figure 3. Example map of the global distribution of one Cantometric variable, "Embellishment", after condensing the 5,779 coded songs into 992 "modal profiles" (i.e., visualizing only the most common Embellishment coding for each of the 992 societies. Embellishment is an ornamenting device in which rapid, ephemeral notes ornament the main melodic line, but are distinct from it. The distribution of much embellishment clearly outlines the entire region of Old Eurasian culture ("Old High Culture" in Figure 1), which includes the Circum-Mediterranean, North Africa, the Arabian Peninsula, the Middle East, Western Asia, South Asia, Southeast Asia, and East Asia, where layered systems of stratification with castes, slaves, and landless laborers, and large state governments have also prevailed (cf. the map of the "Social Stratification" variable at D-PLACE: https://d-place.org/parameters/SCCS158#2/14.3/151.2).

Figure 3 shows a prominent "Silk Road" area of mostly red colour representing highly embellished singing stretching from the Mediterranean to Indonesia that is differentiated from the surrounding predominantly blue colours in Eastern and Central Europe, sub-Saharan Africa, Oceania, and the Americas where singing is mostly unembellished. Lomax's original analyses found correlations between embellished singing and stratified societies (cf. the map of the "Social Stratification" at D-PLACE: https://d-place.org/parameters/SCCS158#2/14.3/151.2 to compare correspondences and dissociations in their distributions). Lomax also found many other potential correlations between singing style and social structure, though critics have pointed out methodological caveats (e.g., autocorrelation, correction for multiple comparison [14, 48]). Now that the full Global Jukebox data are public and linked with D-PLACE, such correlations can be reanalyzed using modern phylogeographic methods [7, 8, 49, 50].

7. Ethics, Rights and Consent

The purpose of the Jukebox is to provide a platform for the world's music and for the data to speak for itself, without judging a diverse range of performative approaches by Western standards. The criteria for analyzing and coding the primary data of songs, dances, etc. are based on listening to global samples and identifying their myriad approaches to performance [49]. Alan Lomax's research and its integration into the Global Jukebox platform sprang from an approach to cross-

cultural performance analysis that was based on his fieldwork with singers and musicians. Guiding Lomax's efforts was his belief that every society's music must be appreciated on its own terms, according to its own aesthetic standards, given equitable airplay, and play a significant part in education [48].

The Association for Cultural Equity is committed to obtaining permission to stream the media examples that have been studied and analyzed for the Global Jukebox. As did Lomax, ACE seeks out the estates of artists recorded by Lomax, and their descendents and estates receive fees and royalties from licensing and sales. Repatriation of Lomax's recordings to their communities of origin, in partnership with those communities, is ongoing and has reached ~50 communities, descendents of artists, and national libraries. North American and Australian indigenous audio samples will not be streamed on the Jukebox unless each community has agreed.

ACE supports emerging leaders from endangered cultures as they learn, with their communities, to document, describe, steward and reappraise their expressive traditions; communities of origin must own and control the results. To improve ethical practices, ACE convenes with cultural advocates from diverse communities [51, 52]. Lomax's recordings and research are freely available online at www.archive.culturalequity.org and www.theglobaljukebox.org, and culture-bearers and other experts are invited to correct, interpret and add their documentation to the songs and metadata.

8. Conclusion

The full publication of Global Jukebox data represents the culmination of sixty years of research by one of the world's most influential scholars of music [33, 53]. We are making these data publicly available in order to encourage their use, improvement, and expansion through diverse intercultural and interdisciplinary collaborations. We also hope to encourage further scientific research into music, dance, speech and other arts as primary rather than ancillary factors in human history and evolution, as well as to deepen our understanding of cross-cultural diversity at a time when it is more important than ever before.

How to cite the Global Jukebox

Research that uses data from the Global Jukebox should cite both the original source(s) of the data and this paper (e.g., research using data from the Cantometrics dataset: "Lomax (1968); Wood et al. 2021)." The reference list should include the date that data were accessed and URL for the Global Jukebox (http://theglobaljukebox.org), in addition to the full reference for Lomax (1968). Additionally, each dataset (table 1) is versioned and stored on Zenodo. Users can cite the specific dataset and version used by visiting [zenodo website].

Acknowledgements

Alan Lomax, with Michael del Rio, imagined and realized the essence of the Global Jukebox. It was based on Lomax's years of fieldwork and experimentation, and thirty years of research on performance style with Conrad M. Arensberg and specialist collaborators. George P. Murdock's foundational work in cross-cultural anthropology, which led to the Ethnographic Atlas, and Norman Berkowitz, a cutting edge programmer and statistician, each played a leading role in operationalizing theories and secondary hypotheses. Richard Smith brought the website out of the

dustbin of obsolete programming languages and obsolete hard drives. Gideon D'Arcangelo helped to bring the Jukebox to life as it is now being developed; Jeff Feddersen (Original Design); Ray Cha (Wireframe); Alona Weiss (Additional Design); Kiki Smith-Archiapatti (Design/Content); Martin Szinger (Developer); Steve Rosenthal and the Magic Shop (audio digital transfers, restoration); Forrestine Paulay (Choreometrics); Karen Kohn Bradley, Frederick Curry, Meriam Lobel, Sinclair O'Gaga, Onye Ozuzu, Miriam Philips, Susan Wiesner (Dance Analysis and Social Justice); Patricia Campbell (Curriculum); Todd Harvey, Jorge Arevalo Mateus, Bruno Nettl, Anthony Seeger, Michael Tenzer, Philip Yampolsky (Ethnomusicology advisors), Karen Claman, Kathleen Rivera (Researchers); Miriam Elhajli (Latin American folk sample); Don Fleming (Popular Song); Victor Grauer (Cantometrics Advisor); Sergio Bonanzinga, Judith Cohen, Lamont Pearly III, Mark Slobin (Journeys); Herb Sturz.

The Global Jukebox is a project of the Association for Cultural Equity (Culturalequity.org), a 501c(3) non-profit charitable organization and custodian of the Alan Lomax Archive. Founded by Alan Lomax in 1983, ACE's mission is to stimulate cultural equity through fostering research, dissemination, and sustainability of the world's traditional expressive practices. It endeavors to reconnect people and communities with their creative heritage through open access and mutual engagement. Lomax's original recordings and papers were deposited with the American Folklife Center of The Library of Congress; ACE retains digital copies which it uses in repatriation, publication, and collaborative initiatives with source communities.

Hunter College of CUNY
ACE is grateful to Hunter College of CUNY for its generous support and for giving us space for the Lomax archive and to work.

Author Contributions
Text: ALCW, PES, KRK, CRE, SS, JS, MF; Design & technical direction: GD; Research direction: ALCW; Reconstruction of code and database structure: RS; Website design: ALCW, JF; Website development: JS, MS; Restoration, digitization, audio and video: SR and MS; Charts and Figures: SS, ALCW, HD, SP; Reliability/validity analyses: PES, ALCW, JM; Data cleaning/curation: SS, JM, HD, SP, ALCW, PES; Matching of Global Jukebox and D-PLACE societies: KRK.

References

1. Murdock, G. P. (1967). Ethnographic atlas: A summary. *Ethnology*, *6*(2), 109–236.
2. HRAF. eHRAF World Cultures. https://ehrafworldcultures.yale.edu/ehrafe/
3. Wyclifee Bible Translators. (1951). *Ethnologue: Languages of the world*. SIL International.
4. Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2020). *Glottolog 4.2.1*. Max Planck Institute for the Science of Human History. https://doi.org/10.5281/zenodo.3754591

5. Currie, T. E., & Mace, R. (2009). Political complexity predicts the spread of ethnolinguistic groups. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(18), 7339–7344. https://doi.org/10.1073/pnas.0804698106

6. Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science, 332*, 346–349. https://doi.org/10.1126/science.1199295

7. Botero, C. A., Gardner, B., Kirby, K. R., Bulbulia, J., Gavin, M. C., & Gray, R. D. (2014). The ecology of religious beliefs. *Proceedings of the National Academy of Sciences of the United States of America, 111*(47), 16784–16789. https://doi.org/10.1073/pnas.1408701111

8. Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., Leehr, D., Low, B. S., McCarter, J., Divale, W., & Gavin, M. C. (2016). D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. *Plos One*, *11*(7), e0158391. https://doi.org/10.1371/journal.pone.0158391

9. Turchin, P., Currie, T. E., Whitehouse, H., François, P., Feeney, K., Mullins, D., Hoyer, D., Collins, C., Grohmann, S., Savage, P. E., Mendel-Gleason, G., Turner, E., Dupeyron, A., Cioni, E., Reddish, J., Levine, J., Jordan, G., Brandl, E., Williams, A., … Spencer, C. (2018). Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *Proceedings of the National Academy of Sciences of the United States of America, 115*(2), E144–E151. https://doi.org/10.1073/pnas.1708800115

10. Turchin, P., Francois, P., Hoyer, D., Nugent, S., Savage, P. E., Brandl, E., Cioni, E., Levine, J., Reddish, J., Altaweel, M., Anderson, E., Bol, P., Carballo, D., Covey, R. A., Feinman, G., Korotayev, A., Kradin, N., Larson, J., Peregrine, P., … Whitehouse, H. (2019). Explaining the rise of moralizing religions: A test of competing hypotheses using the Seshat Databank. *SocArXiv* preprint. https://doi.org/10.31235/osf.io/2v59j

11. Blasi, D. E., Moran, S., Moisik, S. R., Widmer, P., Dediu, D., & Bickel, B. (2019). Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science*, *363*(6432). https://doi.org/10.1126/science.aav3218

12. Slingerland, E., Atkinson, Q. D., Ember, C. R., Sheehan, O., Muthukrishna, M., Bulbulia, J., & Gray, R. D. (2020). Coding culture: Challenges and recommendations for comparative cultural databases. *Evolutionary Human Sciences*, *2*, e29. https://doi.org/10.1017/ehs.2020.30

13. Cohen, R. D. (Ed.) (2003). *Alan Lomax: Selected writings, 1934-1997*. Routledge.

14. Savage, P. E. (2018). Alan Lomax's Cantometrics Project: A comprehensive review. Music & Science, 1, 1–19. https://doi.org/10.1177/2059204318786084

15. Lomax, A. (1980). Factors of musical style. In S. Diamond (Ed.), *Theory & practice: Essays presented to Gene Weltfish* (pp. 29–58). Mouton.

16. Lomax, A. (1968). *Folk song style and culture*. Washington, DC: American Association for the Advancement of Science.

17. Wood, A. L. C. (2018a). "Like a cry from the heart": An insider's view of the genesis of Alan Lomax's ideas and the legacy of his research, Part I. Ethnomusicology, 62(2), 230-264.

18. Wood, A. L. C. (2018b). "Like a cry from the heart": An insider's view on the genesis of Alan Lomax's ideas and the legacy of his research: Part II. Ethnomusicology, 62(3), 403-438.

19. Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

20. Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*, eaax0868. https://doi.org/10.1126/science.aax0868

21. Matsumae, H., Ranacher, P., Savage, P. E., Blasi, D. E., Currie, T. E., Koganebuchi, K., Nishida, N., Sato, T., Tanabe, H., Tajima, A., Brown, S., Stoneking, M., Shimizu, K. K., Oota, H., & Bickel, B. (In press). Exploring correlations in genetic and cultural variation across language families in Northeast Asia. *Science Advances.* Preprint: https://doi.org/10.1101/513929

22. Savage, P. E., Matsumae, H., Oota, H., Stoneking, M., Currie, T. E., Tajima, A., Gillan, M., & Brown, S. (2015). How "circumpolar" is Ainu music? Musical and genetic perspectives on the history of the Japanese archipelago. *Ethnomusicology Forum*, *24*(3), 443–467. https://doi.org/10.1080/17411912.2015.1084236

23. Brown, S., Savage, P. E., Ko, A. M.-S., Stoneking, M., Ko, Y.-C., Loo, J.-H., & Trejaut, J. A. (2014). Correlations in the population structure of music, genes and language. Proceedings of the Royal Society B: Biological Sciences, 281(1774), 20132072. https://doi.org/10.1098/rspb.2013.2072

24. Stumpf, C. (1911). *The origins of music [translated 2012]* (D. Trippett (Trans.)). Oxford University Press.

25. Hornbostel, E. M. von. (1975). *Hornbostel opera omnia* (K. P. Wachmann, D. Christensen, & H.-P. Reinecke (Eds.)). Marinus Nijhoff.

26. Sachs, C. (1962). *The wellsprings of music* (J. Kunst (Ed.)). M. Nijhoff.

27. Nettl, B., & Bohlman, P. V. (Eds.). (1991). *Comparative musicology and anthropology of music: Essays on the history of ethnomusicology*. University of Chicago Press.

28. Schneider, A. (2006). Comparative and Systematic Musicology in Relation to Ethnomusicology: A Historical and Methodological Survey. *Ethnomusicology*, *50*(2). http://www.jstor.org/stable/10.2307/20174451

29. Savage, P. E., & Brown, S. (2013). Toward a new comparative musicology. *Analytical Approaches to World Music*, *2*(2), 148–197. https://doi.org/10.31234/osf.io/q3egp

30. Forkel, R., List, J.-M., Greenhill, S. J., Rzymski, C., Bank, S., Cysouw, M., … Gray, R. D. (2018). Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. Scientific Data, 5(1), 1–10. https://doi.org/10.1038/sdata.2018.205

31. Lomax, A. (1950). *Mister Jelly Roll: the fortunes of Jelly Roll Morton, New Orleans Creole and "inventor of jazz."* New York: Duell, Sloan and Pearce.
32. Jones, B. (1961, October 3-31). *Interviews by Alan Lomax.* Bessie Jones 1961-1962, Association for Cultural Equity, Lomax Digital Archive. https://archive.culturalequity.org/node/765
33. Szwed, J. (2010). Alan Lomax: The man who recorded the world. Viking.

34. Lomax, A., & Grauer, V. (1968). The Cantometric coding book. In A. Lomax (Ed.), *Folk song style and culture* (pp. 34–74). American Association for the Advancement of Science.

35. Wood, A. L. C. (2021). *Songs of Earth: Aesthetic and Social Codes in Music*. ACE/University Press of Mississippi, Oxford, MS. (In Press.)

36. Rzeszutek, T., Savage, P. E., & Brown, S. (2012). The structure of cross-cultural musical diversity. Proceedings of the Royal Society B: Biological Sciences, 279(1733), 1606–1612. https://doi.org/10.1098/rspb.2011.1750

37. Daikoku, H., Wood, A. L. C., & Savage, P. E. (2021). Musical diversity in India: A preliminary computational study using Cantometrics. *Keio SFC Journal*, *20*(2), 34–61.

38. Peel, M.C., Finlayson, B.L. & McMahon, T.A. (2007). Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci., 11, 1633-1644.

39. Ember C.R., Page H, Martin MM, O'Leary T.A (1992). A computerized concordance of cross-cultural samples. New Haven: Human Relations Area Files.

40. Busby, G. (2006). *Finding the blues: An investigation into the origins and evolution of African-American music* [M.Sc. thesis: University of London]. http://users.ox.ac.uk/~some2456/docs/Busby_GBJ_finding_the_blues_2006.pdf

41. Leroi, A. M., & Swire, J. (2006). The recovery of the past. *The World of Music*, *48*(3), 43–54.Callaway, E. (2007). Music is in our genes. *Nature News*, *December*. https://doi.org/10.1038/news.2007.359

42. Du Toit, D. (2011). *Cultural variation of music [thesis, Department of Psychology]* University of Auckland.

43. Grauer, V. A. (2006). Echoes of our forgotten ancestors. *The World of Music*, *48*(2), 5–58.

44. Flory, M., et al? (2021?). Cluster analysis of geographic patterns in Cantometrics. *[PsyArXiv Preprint?]*. https://doi.org/???

45. Passmore, S., et al? (2021?). Global coevolution of music and culture [preliminary title?]. *[PsyArXiv Preprint?]*. https://doi.org/????

46. Savage, P. E., & Brown, S. (2014). Mapping music: Cluster analysis of song-type frequencies within and between cultures. Ethnomusicology, 58(1), 133–155. https://doi.org/10.5406/ethnomusicology.58.1.0133

47.

48. Erickson, E. E. (1976). Tradition and evolution in song style: A reanalysis of Cantometric data. *Cross-Cultural Research*, *11*(4), 277–308. https://doi.org/10.1177/106939717601100403

49. Levinson, S. C., & Gray, R. D. (2012). Tools from evolutionary biology shed new light on the diversification of languages. *Trends in Cognitive Sciences*, *16*(3), 167–173.

50. Gray, R. D., & Watts, J. (2017). Cultural macroevolution matters. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), 7846–7852. https://doi.org/10.1073/pnas.1620746114

51. Titon, J. T. (1992). Music, the public interest, and the practice of ethnomusicology. Ethnomusicology, 36(3), 315–322.

52. Pettan, S., & Titon, J. T. (Eds.). (2015). The Oxford Handbook of Applied Ethnomusicology. Oxford: Oxford University Press.

53. Pareles, J. (2002). Alan Lomax, who raised the voice of folk music in the U.S., dies at 87. The New York Times, A1,A11. http://www.nytimes.com/2002/07/20/arts/alan-lomax-who-raised-voice-of-folk-music-in-us-dies-at-87.html

Supplementary Material [to be added soon]