Mol Genet Genomics (2012) 287:337–349
DOI 10.1007/s00438-012-0683-y

ORIGINAL PAPER

A comparative phylogenetic study of genetics and folk music

Horolma Pamjav · Zoltán Juhász · Andrea Zalán · Endre Németh · Bayarlkhagva Damdin

Received: 7 December 2011 / Accepted: 21 February 2012 / Published online: 4 March 2012

Abstract Computer-aided comparison of folk music from different nations is one of the newest research areas. We were intrigued to have identified some important similarities between phylogenetic studies and modern folk music. First of all, both of them use similar concepts and representation tools such as multidimensional scaling for modelling relationship between populations. This gave us the idea to investigate whether these connections are merely accidental or if they mirror population migrations from the past. We raised the question; does the complex structure of musical connections display a clear picture and can this system be interpreted by the genetic analysis? This study is the first to systematically investigate the incidental genetic background of the folk music context between different populations. Paternal (42 populations) and maternal lineages (56 populations) were compared based on Fst genetic distances of the Y chromosomal and mtDNA haplogroup frequencies. To test this hypothesis, the corresponding musical cultures were also compared using an automatic overlap analysis of parallel melody styles for 31 Eurasian nations. We found that close musical relations of populations indicate close genetic distances (<0.05) with a probability of 82%. It was observed that there is a significant correlation between population genetics and folk music; maternal lineages have a more important role in folk music traditions than paternal lineages. Furthermore, the combination of these disciplines establishing a new interdisciplinary research field of "music-genetics" can be an efficient tool to get a more comprehensive picture on the complex behaviour of populations in prehistoric time.

Keywords Y haplogroups · mtDNA haplogroups · Folk music · Music genetics · Human demographic history

Introduction

Computer-aided comparison of traditional music features from different cultures is one of the newest research areas. While a considerable amount of genetic studies has been carried out in connection with human evolution history and migrations, relations between genetic research and music cultures have never been investigated. Thus, we turned our attention to correlation patterns between genetic data and folk music melodies.

Human history is tightly related with a history of population migration. Patterns of genetic diversity provide information about population history because each major demographic event left an imprint on genomic diversity of populations. These demographic signatures are passed from generation to generation, thus the genomes of modern individuals reflect their demographic history. Studies for evolutionary history have benefited from analyses of

---

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-012-0683-y) contains supplementary material, which is available to authorized users.

H. Pamjav (✉) · A. Zalán · E. Németh
DNA Laboratory, Institute of Forensic Medicine,
Network of Forensic Science Institutes, Budapest,
PO Box 216, 1536, Hungary
e-mail: phorolma@hotmail.com

Z. Juhász
Department of Complex Systems, Research Institute for
Technical Physics and Materials Science, Budapest, Hungary

B. Damdin
Department of Molecular Biology, Faculty of Biology, National
University of Mongolia, Ulaanbaatar, Mongolia

mitochondrial DNA (mtDNA) and the non-recombining region of the Y chromosome. Because these regions of human genome are not shuffled by *recombination*, they are transmitted intact from one generation to the next, revealing the maternal and paternal lineages of a population. Populations share mtDNA or Y chromosome lineages as a result of common origins or *gene flow* (*admixture*). Use of mtDNA and the Y chromosome has been extensively studied and used in migration studies and in the analysis of population history and origins (Wells et al. 2001; Cinnioglu et al. 2004; Nasidze et al. 2005; Pakendorf et al. 2006). Y chromosome markers tend to show restricted regional distribution, or population specificity, making them ideal in marking unique migration events (Hammer et al. 1997). Similarly, the variation of mtDNA between different population groups can be used to estimate the time back to common ancestors. In the last few years, there has been significant progress in reconstructing the detailed genealogical branching order of the tree topologies for both mtDNA and the non-recombining region of the Y chromosome (Jobling and Tyler-Smith 2003; Underhill and Kivisild 2007; Karafet et al. 2008).

Affinities between populations may result from their common origin or from recent admixture due to geographic proximity. In particular, genetic distances between populations can generally be related to geographic distances, according to a model of isolation by distance (Cavalli-Sforza and Bodmer 1977). But this is not relevant in all cases.

According to earlier reports, correlation between genetic distance and language as a cultural marker has been debated (Nettle and Harriss 2003). We, however, set out to examine relationships between genetic distance and folk music cultures.

The comparative study of different folk music cultures goes back to the early 20th century. Bartók and Kodály identified certain musical styles of Hungarian folk music being common with Mari, Chuvash, as well as Anatolian melodies (Bartók 1949; Kodály 1971). These results initiated the expansion of a digitized international folksong database (the European Folksong Catalogue), and a software system for mathematical analysis of the data in the 1960s (Csébfalvy et al. 1965). Although this project, organized by the Hungarian Academy of Sciences was interrupted, development of computation tools led to the later resurrection of the idea (Toiviainen 1996, 2000; Leman 2000; Toiviainen and Eerola 2002; Kranenburg et al. 2009; Garbens et al. 2007; Huron 1996; Juhász (2009).

In recent years, we have developed and enhanced a software system which characterizes the overall similarity of different folk song databases using a scalar measure. The basic idea of this system is that musical information propagates and evolves by endless variation of melodies, since variation is an essential feature of most oral cultures. Cross-cultural analysis of different nations in Eurasia drew a well-interpretable network of “genetic” relations between musical traditions (Juhász 2000; Juhász and Sipos 2009).

This led us to the question whether these connections are merely accidental or can they possibly relate to the migration of populations in any way? Does the complex structure of musical connections display a clear picture and can we interpret this connection system on a genetic basis as well?

Materials and methods

Materials

In the present work, we describe the results of a computer-aided cross-cultural analysis of 31 representative Eurasian and North-American folksong collections, and a parallel study of genetic relations between the corresponding population groups. Our melody database contains digital notations of 31 cultures, each of them consisting of 1,000–2,500 melodies. The studied cultures are as follows: Chinese, Mongolian, Kyrgyz, Volga Region (Mari–Chuvash–Tatar–Votiac), Sicilian, Bulgarian, Azeri, Anatolian, Karachay, Hungarian, Slovakian, Moravian, Romanian, Finnish, Norwegian, German, Luxembourgish, French, Dutch, Irish–Scottish–English (one group), Spanish, Dakota, Komi, Khanty, Croatian–Serbian (Balkan group), Kurdish, Russian from the Pskov Area, Navajo and three regions of Poland: Great Poland, Warmia and Cassubia.

Simultaneously with the musical database investigation, we also calculated corresponding genetic distance matrices using widely used software tools (Schneider et al. 2000) from published data for those populations compared above, with the following differences: (1) no genetic data were found for the Luxembourgish, Karachay and Dakota Indians, (2) Y chromosomal and mtDNA genetic data were used for the Croatian; Tatar, Mari, Udmurt/Votiac and Chuvash (instead of Volga Region); Irish and Scottish populations separately, (3) genetic data of three Hungarian speaking groups (Hungarian, Szekler and Csango) were used (Szekler and Csango Y chromosomal data is unpublished). The Szeklers and Csangos are Hungarian speaking ethnic groups living in Romania, (4) one Polish population data were used for the Y and mtDNA genetic comparisons; whereas three Polish folksong databases arising from different areas of Poland were analysed.

Methods for genetic analysis

We calculated genetic distances (Fst) with AMOVA based on Y chromosomal and mtDNA haplogroup frequencies with Arlequin 2.0 software (Schneider et al. 2000). Multidimensional scaling (MDS) plots were constructed with ViSta 7.9.2.4 software.

Y chromosomal haplogroups were combined into groups C, D, E, F, I, J2, K, L, N1c, PxR1a and R1a, so that published sources could be used for comparisons (Bosch et al. 2006; Cinnioglu et al. 2004; Di Gaetano et al. 2009; Hammer et al. 2006; Kayser et al. 2005; Maca-Meyer et al. 2003; Malhi et al. 2008; Nasidze et al. 2002, 2005; Pericić et al. 2005a, b; Petrejcíková et al. 2010; Pimenoff et al. 2008; Rosser et al. 2000; Semino et al. 2000; Tambets et al. 2004; Völgyi et al. 2009; Wells et al. 2001; Xue et al. 2006; Zastera et al. 2010).

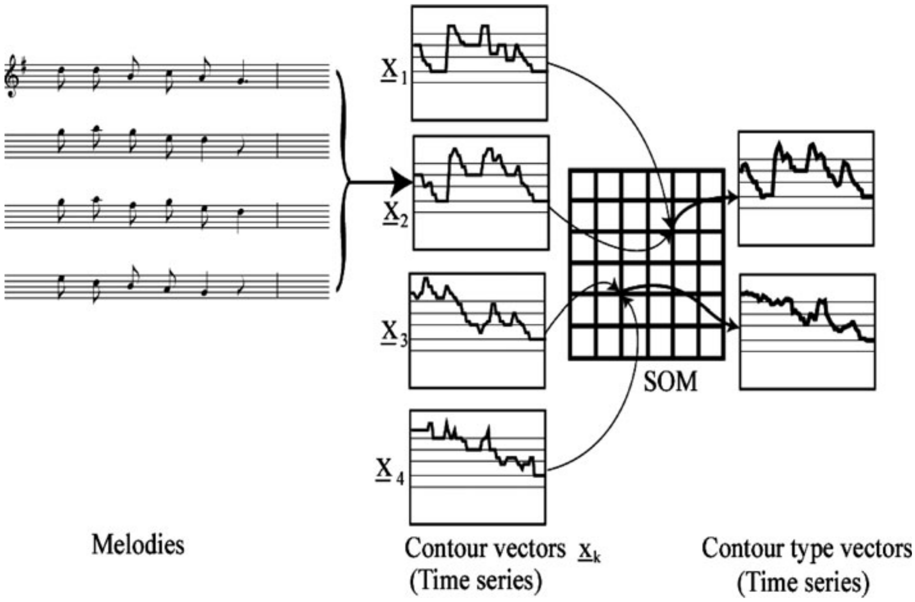
To compare maternal lineages, we used published mtDNA haplogroup frequencies (Bermisheva et al. 2004; Bosch et al. 2006; Brandstätter et al. 2007; Comas et al. 2004; Derbeneva et al. 2002; Helgason et al. 2001; Irwin et al. 2007; Maca-Meyer et al. 2003; Malyarchuk et al. 2002; Merriwether et al. 1996; Pakendorf et al. 2007; Passarino et al. 2002; Pericić et al. 2005a; Pimenoff et al. 2008; Quintana-Murci et al. 2004; Simoni et al. 2000; Starikovskaya et al. 2005; Yao et al. 2004; http://www.eupedia.com). Haplogroups were combined into groups A, B, C, D, E, F, G, HV, H, I, J, K, L, M (×Q, G, E, D, C, Z), N, P, Q, R, S, T, U × K, V, W, X, Y and Z.

Methods for musical analyses

We developed a system for automatic classification of melody contours based on a generally applied kind of artificial intelligence, the so-called self-organizing map (SOM) (Kohonen 1995). The operation of our system is demonstrated in Fig. 1. The first step is to construct melody contour vectors of *D* dimensions (*D* element pitch sequences) from the digital codes of music notations (Juhász 2006). Since *D* was the same for each melody, the contours could be compared to each other using a common Euclidean distance function defined in the same *D*-dimensional melody space, irrespective of their individual length. The contour vectors are sampled time functions of the pitch variation, the main rhythmic characteristics of the melodies are also mapped into the vector space determined by the contour vectors. At the same time, the tempo is completely normalised by this technique, but this is a useful and general tool of feature extraction in ethnomusicology. To avoid problems arising from different notation principles, the melodies of all collections were transposed to common final tone G. In most unison folk song cultures, the melodies have a well-defined tonic, and this tonic is identical to the final tone; therefore, this technique is musically relevant in the most studied cases. However, since this is not generally true for Western folksongs, the automatic transposition of these databases was corrected partly by experts, partly by a further algorithm.

After training, the SOM with the contour vectors of a given culture, the learned contour type vectors belonging to different grid locations of the map represented the most important contours characterizing the given culture (Juhász 2006). In other words, a well-trained SOM represents a musical language which is optimal for "understanding" melodies of its own culture. Using the terms of neural networks, the contour type vectors function as "receptors", "firing" or "being activated" when a melody contour is found to be similar enough to them. To compare two different "musical languages", one of them will be "excited" by the contour types of the other language. Common

Fig. 1 The generation of contour vectors and the learning of averaged contour type vectors using a SOM



musical features can be identified by contour types being activated by type vectors of the foreign language, and particular characteristics are also described by those having no relations in the other culture.

The degree of general relation between cultures A and B can be characterized by the overlap of their contour type sets. More exactly, we defined the degree of relation of the ith and jth cultures as $d_{i,j} = 1 - \frac{n_{i,j}}{N_i}$, where $n_{i,j}$ is the number of contour types of the ith culture activated by those of the jth one, and N_i are the total number of the contour types of the ith culture. Since the relation is not symmetric $\left(d_{i,j} \neq d_{j,i}\right)$, our definition will not satisfy the criteria of a distance function in the mathematical sense, but we can state that the closer two cultures are to each other, the smaller are the corresponding $d_{i,j}$ and $d_{j,i}$ values.

This rate, the relation of any musical culture to the others, can be characterized by a 31 dimensional vector containing the above defined degrees $d_{i,j}$. To get a clear picture of the very complicated system of relations described by these non-symmetric data, we applied a special version of the multidimensional scaling (MDS) method (Borg and Groenen 2005).

This technique is well known in genetics. Therefore, we are able to establish the relationships between musical cultures using the same method that is commonly used in the field of phylogenetic studies.

Results

The genetic relationships

Genetic distances (Fst) of Y chromosomal haplogroup frequencies between 42 populations compared (Online Resource 1) are displayed as an MDS plot (Fig. 2). Most populations compared tend to group into four clusters in the middle of the plot with two smaller clusters on the edges, whereas some population groups appear as outliers. Populations were determined to be members of a cluster on the basis that the genetic distances between them were less than 0.05, and also that as many populations as possible should be grouped. Three Hungarian speaking populations (Hungarian, Szekler and Csango) clustered together with two eastern European (Bulgarian and Serbian), two Tatar

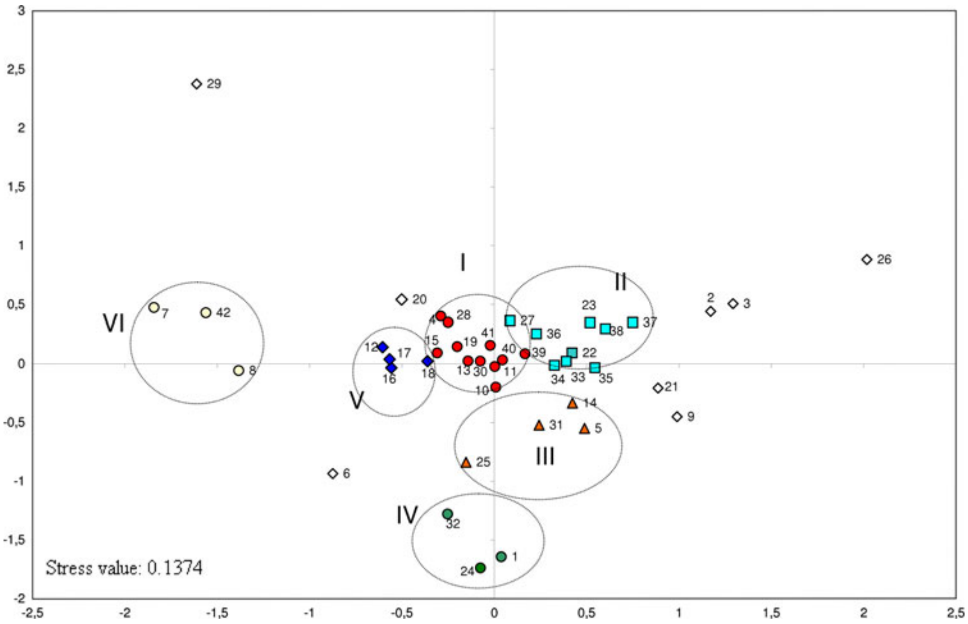


Fig. 2 Multidimensional scaling (MDS) plot constructed on Fst genetic distances of Y haplogroups of the populations compared *1* Mari (Semino et al. 2000) *2* Scottish (Rosser et al. 2000) *3* Irish (Rosser et al. 2000) *4* Bulgarian (Rosser et al. 2000) *5* Udmurt (Semino et al. 2000) *6* Chinese (Xue et al. 2006) *7* Ewenki pooled (Xue et al. 2006; Tambets et al. 2004) *8* Inner Mongolian (Xue et al. 2006) *9* Kyrgyz (Wells et al. 2001) *10* Kazan Tatar (Wells et al. 2001) *11* Crimean Tatar (Wells et al. 2001) *12* Azeri (Wells et al. 2001) *13* Kurdish (Wells et al. 2001) *14* Russian (Wells et al. 2001) *15* Sicilian (Di Gaetano et al. 2008) *16* Azeri (Nasidze et al. 2002) *17* Chechen (Nasidze et al. 2002) *18* Turkish (Cinnioglu et al. 2004) *19* Greek (Bosch et al. 2006) *20* Romanian (Bosch et al. 2006) *21* Polish (Kayser et al. 2005) *22* German (Kayser et al. 2005) *23* Spanish (Maca-Meyer et al. 2003) *24* Khanty (Pimenoff et al. 2008) *25* Mansi (Pimenoff et al. 2008) *26* Navajo (Malhi et al. 2008) *27* Croatian pooled (Peričić et al. 2005a, b) *28* Serbian (Peričić et al. 2005b) *29* Ainu (Hammer et al. 2006) *30* Kurdish (Nasidze et al. 2005) *31* Chuvash (Tambets et al. 2004) *32* Finnish (Tambets et al. 2004) *33* Norwegian (Tambets et al. 2004) *34* Slovakian (Petrejcíková et al. 2010) *35* Czech (Zastera et al. 2010) *36* French (Semino et al. 2000) *37* Dutch (Semino et al. 2000) *38* Italian (Semino et al. 2000) *39* Hungarian (Völgyi et al. 2009) *40* Szekler (unpublished data) *41* Csango (unpublished data) *42* Outer Mongolian (unpublished data)

(Crimean and Kazan), two Kurdish, a Sicilian and a Greek population group (Cluster I). Most western European (German, Dutch, Spanish, Norwegian and Italian) and three central European (Czech, Croatian and Slovakian) populations grouped into cluster II. Cluster III included three populations from the Volga Region (Udmurt, Chuvash and Mansi) and one Russian population. Three Finno-Ugric populations (Mari, Khanty and Finnish) clustered together in cluster IV. Two Azerian, Turkish and Chechen population groups clustered together in cluster V. Three Asian populations (Ewenki, Inner and Outer Mongolian) were included in Cluster VI.

To place these findings in context of female lineages, genetic distances of mtDNA haplogroups between the populations from published sources were calculated (Online Resource 2), and were again presented as an MDS plot (Fig. 3). Clusters were distinguished as a cluster in case the genetic distances between populations were less than 0.05, as well as in the case of the Y chromosomal MDS plot.

Three Hungarian speaking population groups (Hungarian, Szekler and Csango) grouped together in cluster I. Most European populations formed a fairly compact cluster in Fig. 3 (cluster II). Out of Finn-Ugric speaking people, two Finnish and Mari population groups included in cluster II and the rest of them belonged to other Europeans (14 population groups). Four other European populations (Sicilian, Norwegian, French and Spanish) formed cluster III. Cluster IV included three populations from the Volga Region (Chuvash, Udmurt and Tatar), two Komi groups (Ural), Azeri (Caucasus), Turkish (Anatolia) and two European populations (Italian and Bulgarian). Three Mansi (Ural), two Kyrgyz (Central Asia) and one Evenki (Siberia) population groups belonged to cluster V, whereas six Asian

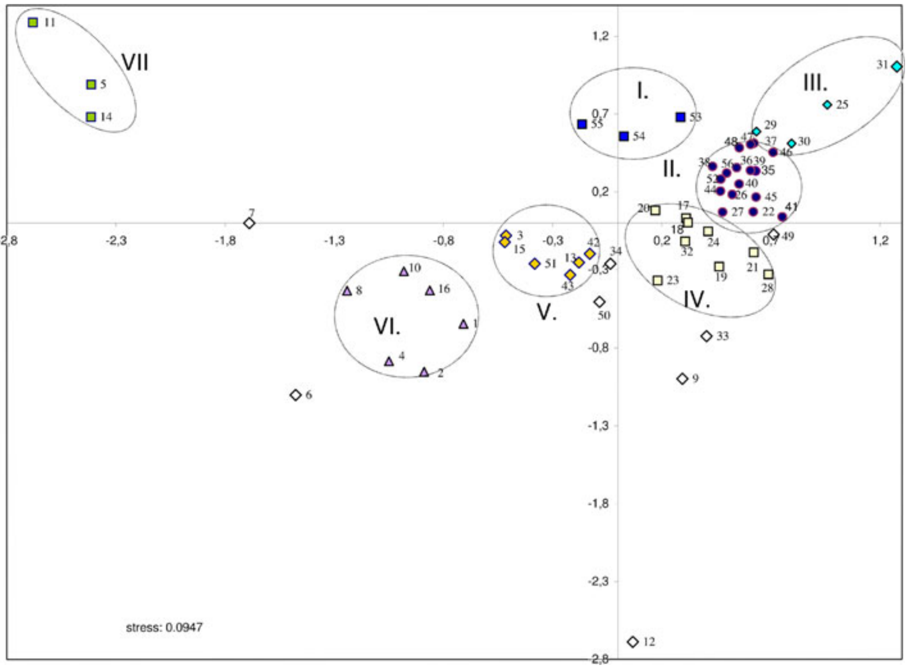


Fig. 3 Multidimensional scaling (MDS) plot constructed on Fst genetic distances of mtDNA haplogroups of the populations compared *1* Outer Mongolian (Yao et al. 2004) *2* Chinese (Yao et al. 2004) *3* Kyrgyz (Yao et al. 2004) *4* Inner Mongolian (Yao et al. 2004) *5* Evenks, Stony Tunguska (Pakendorf et al. 2007) *6* Evenks, Iengra (Pakendorf et al. 2007) *7* Evenks, Nyukhaza (Pakendorf et al. 2007) *8* Evenks, Yakut speaking (Pakendorf et al. 2007) *9* Crimean Tatar (Comas et al. 2004) *10* Kyrgyz (Comas et al. 2004) *11* Evenki (Merriwether et al. 1996) *12* Navajo (Merriwether et al. 1996) *13* Evenki (Starikovskaya et al. 2005) *14* Mansi (Starikovskaya et al. 2005) *15* Kyrgyz (Bermisheva et al. 2004) *16* Outer Mongolian (Bermisheva et al. 2004) *17* Azeri (Bermisheva et al. 2004) *18* Tatar (Bermisheva et al. 2004) *19* Chuvash (Bermisheva et al. 2004) *20* Komi, Permyak (Bermisheva et al. 2004) *21* Komi, Zyryan (Bermisheva et al. 2004) *22* Mari (Bermisheva et al. 2004) *23* Udmurt (Bermisheva et al. 2004) *24* Bulgarian (Simoni et al. 2000) *25* French (Simoni et al. 2000) *26* German (Simoni et al. 2000) *27* Finnish (Simoni et al. 2000) *28* Italian, Tuscany (Simoni et al. 2000) *29* Sicilian (Simoni et al. 2000) *30* Norwegian (Simoni et al. 2000) *31* Spanish, Central (Simoni et al. 2000) *32* Turkish, Anatolian (Quintana-Murci et al. 2004) *33* Kurdish, Iran (Quintana-Murci et al. 2004) *34* Kurdish, Turkmenistan (Quintana Murci et al. 2004) *35* Slovakian (Malyarchuk et al. 2002) *36* Czech (Malyarchuk et al. 2002) *37* Greek (Bosch et al. 2006) *38* Romanian (Bosch et al. 2006) *39* Polish (Malyarchuk et al. 2002) *41* Finnish (Derbeneva et al. 2002) *42* Mansi, Konda (Derbeneva et al. 2002) *43* Mansi, Lyamin (Derbeneva et al. 2002) *44* Norwegian (Passarino et al. 2002) *45* Russian (Helgason et al. 2001) *46* German (Helgason et al. 2001) *47* Irish (Helgason et al. 2001) *48* Scotish (Helgason et al. 2001) *49* Spanish, Cantabrian (Maca-Meyer et al. 2003) *50* Khanty (Pimenoff et al. 2008) *51* Mansi (Pimenoff et al. 2008) *52* Croatian (Pericic et al. 2005a) *53* Hungarian (Irwin et al. 2007) *54* Szekler (Brandstätter et al. 2007) *55* Csango (Brandstätter et al. 2007) *56* Dutch (http://www.eupedia.com)

population groups (3 Mongolian, Evenki, Kyrgyz and Chinese populations) were grouped into cluster VI (Fig. 3). Cluster VII included two Evenki (Siberia) and Mansi (Ural) population groups. Out of 56 population groups compared, the other ones were outliers as shown in Fig. 3.

The system of musical relations

The musical contacts revealed by the above described SOM-based analysis are represented in Fig. 4. The locations of the nodes representing the 31 cultures were determined according to the MDS principle. The lines indicate a very close musical contact between the connected nodes (both $d_{i,j}$ and $d_{j,i}$ are less than a threshold distance of 0.44), and all close musical contacts are indicated in Fig. 4. Although the aim of the algorithm is to minimize the average length of the edges, Fig. 4 shows that some of them remained rather long after learning. This fact shows that the algorithm has found an optimal, but not exact two-dimensional mapping of the complicated system of connections. However, the structure of the graph really reflects the following main properties: a very simplified description of the Chinese–Mongolian–Volga (IV) group can refer to the dominance of the melodies of high ambitus and pentatonic scale. The descending contours in this group have close contacts in Dakota, Anatolian Turkish, Karachay and Sicilian (V) folk music, although these latter variants are mainly diatonic. Another layer of Turkish music, containing diatonic melodies of low ambitus has close Romanian and Spanish contacts. On the other hand, a further significant layer in the Spanish corpus containing plagal melodies shows close French and Finnish relations. The Central European group (III: Hungarian–Slovakian–Great Polish corpora) has close contacts with almost all of the mentioned families, due to the particularly extended relations of Hungarian folk music. The close Chinese, Volga, Karachay, Sicilian, Turkish, Romanian, Spanish, Polish, Norwegian, Finnish, Irish–Scottish–English and Slovakian connections determine a central location for the Hungarian node, but this optimal position yields still rather long edges to the related cultures, due to the limitations of the two-dimensional approximation. Musical and historical analysis of the Hungarian melody layers being in contact with other cultures exceeds the aim of this paper, but some aspects have already been published in previous publications (Juhász and Sipos 2009).

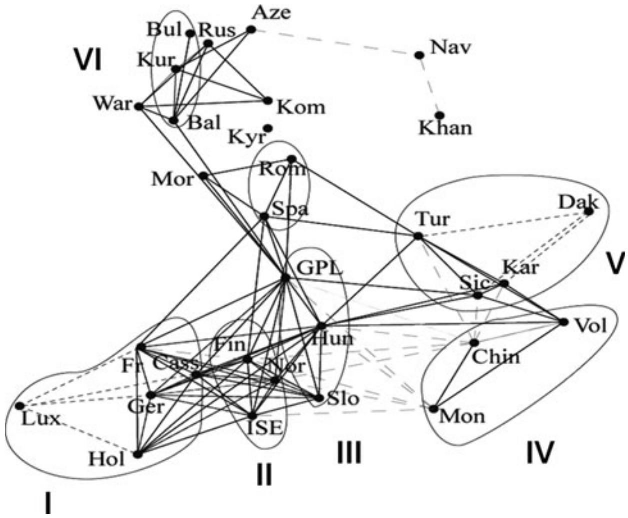


Fig. 4 Musical language groups determined automatically from the overlaps of 31 folk song cultures. *Aze* Azeri; *Bal* Balkan; *Bul* Bulgarian; *Cass* Cassubian (Poland); *Chin* Chinese; *Dak* Dakota; *Fin* Finnish; *Fr* French; *Ger* German; *GPL* Great (Central and Southern) Poland; *Hol* Holland; *Hun* Hungarian; *ISE* Irish–Scottish–English; *Kar* Karachay; Khan: Khanty; Kom: Komi; *Kur* Kurdish; *Kyr* Kyrgyz; *Lux* Luxembourgish; *Mor* Moravian; *Nav* Navajo; *Nor* Norwegian; *Rom* Romanian; *Rus* Russian; *Sic* Sicilian; *Slo* Slovakian; *Spa* Spanish; *Tur* Turkish; *Vol* Volga Region; *War* Warmia (Poland)

Up to this point, we aimed to the identify groups of cultures containing a significant amount of common melody types. The complementary approach is to search for large groups of melody types being simultaneously found in a common subset of the 31 cultures. To accomplish this task, we determined the totality of the important contour types appearing in our melody database by training a large SOM with all of the contour types previously learned by the 31 national/areal SOMs. Being in possession of this large “unified” contour type collection, the above question can be formulated as an algorithmic search as follows: we search for the largest subsets of the 31 melody corpora having the greatest amount of common elements in the unified contour type collection. Results are demonstrated in Fig. 5, where the common contour type vectors are assigned to points of the plane, and the resulting point system is arranged by an MDS algorithm. Thus, similar contour types are assigned to near locations of the plane, and the whole arrangement of the point system mirrors the overall similarity of relations of the common contour type collection. A search for four element subsets of cultures with the largest overlap found the Volga–Turkish–Karachay–Hungarian and the Chinese–Turkish–Karachay–Hungarian groups with respectively 46 and 45 totally common types. Black dots denoting these common types are located in a well-defined area of the melody map, consequently they may belong to a well-defined musical style with a continuous system of variants.

As a reference, we also show common melody types of four Western cultures in Fig. 5c. These types are located in a completely different area of the unified cloud, thus, they stand for completely different musical forms. The number of the totally common Western types (17) shows a much smaller musical core in the music of Western Europe than in the former groups. The search for groups of 7 cultures containing the most totally common melody types found the same subset completed by Sicilian, Finnish and Dakota, as well as Mongolian, Volga and Sicilian cultures. Figure 5d and e show that the totally common melody types form subsets of the larger black clouds shown in Fig. 5a, b. The genetic distance data of the above cultures are summarized in Table 1, where the close relations are highlighted in bold.

Discussion

Genetic relationships based on MDS plots

To evaluate genetic relationships between populations investigated, Fst genetic distances were calculated and the general structure of the distance matrix was depicted by MDS plots. The genetic clustering of the populations can be interpreted as the degree of similarity of population structures that often reflect common ancestry or gene flow. Both Y chromosomal and mtDNA MDS plots were constructed from nearly the same set of populations depending on genetic data found. MDS plots of Y chromosome and mtDNA showed six and seven distinct clusters suggesting genetic affiliations between populations based on geographical and historical background. Most central and south-eastern European population groups clustered together indicating close genetic relationships (Fig. 2, cluster I). Genetic affinity of two Kurdish (Turkmenistan, Turkey–Georgia) and two Tatar population groups included in Cluster I together with other Europeans could not be explained by geographic proximity, but we found convincing musical correlations between many members of this group [for instance: Hungarian–Tatar (Volga), Kurd–Serbian–Croatian (Balkan)–Azeri—see Fig. 4 and below]. The close genetic relationships that were detected among populations included in clusters II, III, IV, V and VI can be interpreted by admixture due to geographical proximity. Ainu, Chinese, Romanian, Mansi and Navajo Indian population groups were outliers as shown in Fig. 2. Scottish–Irish (2 and 3) and Polish–Kyrgyz (9 and 21) population groups showed close affinities to each other (Fig. 2).

The MDS plot of mtDNA represented seven distinct clusters; generally more populations are grouped into a cluster based on threshold value comparing with Y chromosomal MDS plot. Close genetic affinity between the three Hungarian speaking population groups is evident (Cluster I). Almost all European population groups compared (17 populations) based on the Fst genetic distances of mtDNA haplogroups belong to a more compact cluster than that of Y chromosomal comparison (Fig. 3, Cluster II). This observation could also be explained by admixture. Population affiliations between French, Spanish, Sicilian and Norwegian peoples are probably due to geographic and historical backgrounds (cluster III). Population groups included in cluster IV reflect an interesting genetic relationship: Genetic affinities of three groups of the Volga Region (Chuvash, Udmurt, and Tatar) and two Komi groups in Ural are probably associated with geographic proximity. The same can be said for the Azeri, Turkish and Bulgarian populations, but the Italian population is exceptional in this cluster. Interestingly, the Mansi, Kyrgyz

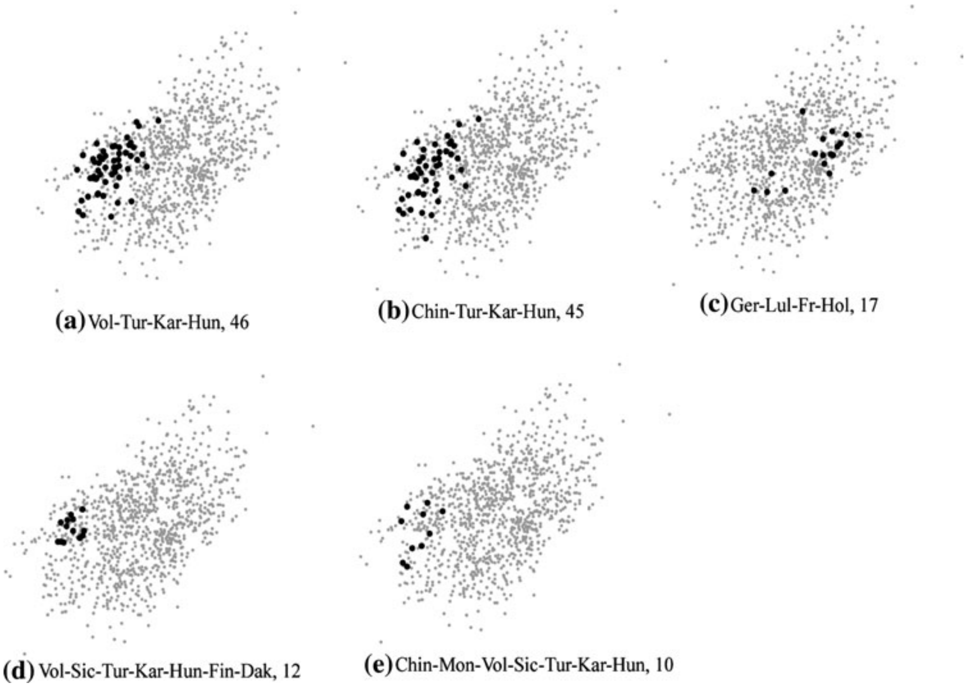


Fig. 5 The largest overlaps of musical cultures on the great common melody map. *Chin* Chinese; Fr: French; *Ger* German; *Hol* Holland; *Hun*: Hungarian; *Kar* Karachay; *Lul* Luxembourgish; *Spa* Spanish; *Tur* Turkish; *Vol* Volga Region

Table 1 Genetic distances (Fst) of male and females lineages compared

	Fst values (females)											
	Chinese	Mongolian	Mari	Udmurt	Tatar	Chuvash	Sicilian	Turkish	Hungarian	Szekler	Csango	Finnish
Fst values (males)												
Chinese	–	**0.012**	0.195	0.137	0.150	0.169	0.214	0.139	0.189	0.158	0.153	0.169
Mongolian	0.322	–	0.147	0.098	0.104	0.117	0.175	0.090	0.159	0.127	0.127	0.124
Mari	0.346	0.458	–	0.053	**0.007**	**0.012**	0.024	**0.017**	0.054	0.064	0.095	**0.003**
Udmurt	0.272	0.394	0.165	–	**0.025**	**0.036**	0.109	**0.025**	0.086	0.081	0.110	**0.047**
Tatar	0.187	0.303	0.223	0.059	–	**0.003**	**0.040**	**−0.002**	0.051	0.052	0.068	**0.004**
Chuvash	0.197	0.372	0.133	**−0.015**	**−0.001**	–	0.079	**0.005**	0.099	0.094	0.099	**0.013**
Sicilian	0.191	0.246	0.309	0.189	0.056	0.136	–	**0.054**	**0.023**	**0.040**	0.065	**0.022**
Turkish	0.164	0.247	0.299	0.198	**0.049**	0.126	**0.004**	–	0.060	**0.049**	0.060	**0.012**
Hungarian	0.222	0.314	0.291	0.105	**0.011**	0.071	0.074	0.080	–	**0.007**	**0.043**	**0.045**
Szekler	0.201	0.279	0.260	0.091	**0.001**	0.053	**0.022**	**0.031**	**0.011**	–	**0.028**	**0.049**
Csango	0.272	0.318	0.325	0.144	**0.012**	0.096	**0.053**	0.055	**0.005**	**0.010**	–	0.064
Finnish	0.324	0.407	**0.048**	0.174	0.147	0.123	0.255	0.247	0.226	0.200	0.235	–

Genetic distances between Chinese, Mongolian, Volga Region (Mari, Udmurt, Tatar, Chuvash), Turkish, Hungarian speaking people (Hungarian, Szekler and Csango), Sicilian, Turkish and Finnish populations are summarized. Bold characters indicate that Fst genetic distances are less than 0.05

Fst genetic distances of Y chromosomal and mtDNA haplogroup frequencies were chosen from the Tables in Online Resource 1 (ESM_1) and 2 (ESM_2)

and Evenki groups mirrored a close genetic affinity to each other, despite greater geographical distances (Fig. 3, Cluster V). Close genetic distances between the populations included in clusters VI and VII are acceptable due to historical backgrounds.

The largest genetic distances were detected at Ainu, Chinese and Navajo Indian populations in the MDS of Y chromosomal comparison as well as at two Evenki and Navajo in mtDNA comparison. These outlier populations are labelled as empty diamonds in Figs. 2 and 3.

Comparisons of Y chromosomal and mtDNA studies in the investigated populations have indicated greater levels of mtDNA lineage sharing among populations, suggesting that females may have accomplished more mobility during human evolution history than males. This observation is consistent with results previously reported (Bamshad et al. 2001).

Musical and genetic correlations

The results of the comparison of folk music corpora of 31 populations are summarized in Online Resource 3. The tones of the edges in Fig. 4 refer to correspondence of musical and genetic similarities. Thick black edges indicate a simultaneously close musical and genetic relation between the corresponding populations; while long dash lines connect populations with a close musical contact without a close genetic relation. Short dash lines indicate the cases when we could not find genetic data belonging to the related population. No data were available for Dakota, Karachay and Luxembourgish populations in either the mitochondrial or male lineages. In order to draw as detailed a picture of musical contacts as possible, we did not eliminate these cultures from the musical analysis, although we confined ourselves to the cultures characterised by simultaneously existing musical and genetic data when analysing the correlation between music and genetics. Therefore, the incomplete data do not influence our accuracy.

In some other cases, multiple genetic data correspond to the same musical collection. This can be illustrated by the Mari, Chuvash, Tatar and Udmurt genetic distance data, being mapped to the same musical database named "Volga", containing a unified folksong collection of these nations.

Utilizing this "Volga" musical culture as an example, the genetic distances from the Turkish population are 0.3, 0.2, 0.049 and 0.13, respectively, for Mari, Udmurt, Tatar and Chuvash Y chromosomal data. Notwithstanding that three of these data are much higher than 0.05, the Turkish–Tatar distance fulfils our requirement; therefore, we can say that close contact of the Turkish musical tradition in the Volga-region can be traced back to the close genetic relation between the Turkish and Tatar people. Moreover, the mitochondrial genetic distance between Turkish and Tatar populations becomes merely 0.002, therefore, this latter data can be accepted as the characteristic distance of the given populations (see Table 1). As a consequence of this consideration, the simultaneously close musical and genetic relations are indicated by the thick black edges connecting the Turkish and Volga nodes in Fig. 4.

We can say in general that a genetic relation between certain characteristic parts of two populations indicates an earlier physical and biological contact; therefore, the musical relation can be traced back to a probable cultural interaction or cooperation of the ancestors of the given groups. Therefore, we state the concept of close genetic relation for our case as follows: a close genetic relation belonging to a pair of musical cultures is established on condition that at least one of the genetic distance data determined for any subsets of the corresponding populations is less than the critical threshold value.

Since the genetic distance is in the range of 0–1 by definition, our choice of the threshold of 0.05 is a rather rigorous requirement for establishing a close genetic contact. This requirement is satisfied by 21% of the Y chromosomal and 37% of the mitochondrial data.

Counting up the number of cases when a close relation was detected between

1. genetic (G)
2. musical (M) and
3. both musical and genetic (X) data, the conditional probabilities of the events that
   (a) a close genetic relation implies a close musical contact, as well as
   (b) a close musical relation implies a close genetic contact becomes

$$p(m|g) = X/G = 0.28 \text{ and } p(g|m) = X/M = 0.82 \text{ respectively}$$

According to the low probability of the first case, close genetic relations of populations are definitely not appropriate indicators of similar musical cultures. However, the high value of the second conditional probability verifies that a significant similarity of folk music cultures detects significant genetic connections in 82% of the studied cases. In other words, genetic relations can be predicted from the musical relations with a hit probability of 82%.

The dependence of the correlation of the musical and genetic relations on the genetic threshold value is represented in Fig. 6. In this figure, curve 2 shows the variation of the $p(g|m) = X/M$ conditional probability as a function of the genetic threshold value. In other words, curve 2 shows the variation of the frequency of close genetic contacts within the group of closely related musical cultures. In the case of total independence of genetic and musical relations, this curve would be essentially identical to that of showing the ratio of the close genetic contacts within the whole set of the studied 31 cultures. However, curve 1, representing this latter frequency, is significantly and consequently lower than curve 2. This observation verifies that the correlation being found between the musical and genetic relations for the genetic threshold of 0.05 is standing independently of the choice of the threshold value. Curve 3 shows the $p(m|g) = X/G$ probability as a function of the genetic distance threshold (the frequency of the musical relations within the genetically related cultures). In contrast to the above cases, this curve shows a decreasing tendency, therefore, our statement that no musical contacts can be predicted from genetic relations also proved to be independent of the genetic threshold.

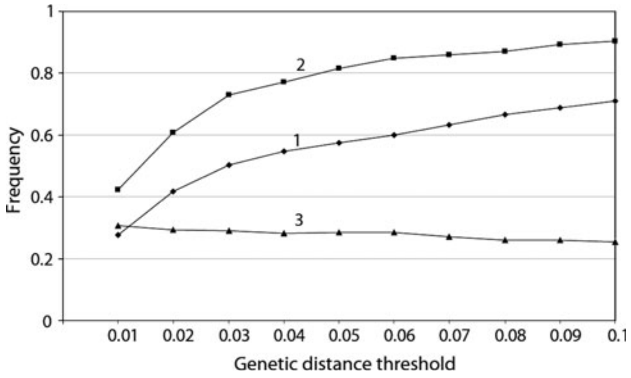


Fig. 6 The dependence of some characteristic frequencies on the genetic threshold value. *1* The frequency of the close genetic contacts within the whole set of the 31 cultures. *2* The frequency of the close genetic contacts within the group of the musically related cultures ($p(g|m) = X/M$). *3* The frequency of the close musical contacts within the group of genetically related cultures ($p(m|g) = X/G$)

Comparing Figs. 2 and 3 with Fig. 4, one can see that merely partial correspondences can be found between the genetic and musical clusters. For instance, the close Bulgarian–Azeri, Serbian–Kurdish, Hungarian–Tatar, Irish–Dutch, Chuvash–Udmurt and Finnish–Mari musical relations really connect cultures belonging to the same genetic cluster in Fig. 2, but the Turkish–Hungarian, Mongolian–Chuvash, Finnish–Scottish, etc. musical relations do not fit into this system at all. This fact is a natural consequence of the asymmetric conditions indicated by the above discussed asymmetric conditional probability values. At the same time, the high predictability of the genetic contacts from musical relations is expressed in the dominance of the thick black edges in Fig. 4.

As a musical illustration of the weight of this latter result, we show some examples of similar Hungarian, Norwegian and Appalachian (Appalachian people are mainly of Scottish and Irish origin) songs in Note 1 (Online Resource 4). The musical analysis showed mutually intensive contacts among these cultures, but the high geographical distances made these results questionable before the comparison to the genetic data. However, the genetic distances of mtDNA data show a consequent correlation with the musical results: 0.0017, 0.024 and 0.024 for Norwegian–Appalachian (Scottish), Norwegian–Hungarian and Hungarian–Appalachian relations, respectively In the light of these data, the melodies in Note 1 (Online Resource 4) can no longer be considered to be accidentally similar "compositions" of three completely independently evolving cultures. It is worth mentioning here that the musical analysis itself excludes this assumption, because the connecting lines in Fig. 4 refer to an extremely high number of overlapping melody types in the three cultures. Moreover, our examples show that these overlaps cover a significant amount of common types simultaneously appearing in all of the three cultures. The first example (Online Resource 4a) contains variants of a melody type of domed contour with an ascending fifth transposition between the first and second sections, while the melodies of example Online Resource 4b represent a descending contour type. These significantly different forms illustrate the existence of a rather varied group of melody forms being simultaneously present in the three cultures.

Accomplishing the above probability analysis for mitochondrial as well as male genetic distances separately, we found $p(m|g) = 0.27$ and $p(g|m) = 0.77$ for mitochondrial and $p(m|g) = 0.36$ and $p(g|m) = 0.35$ for Y chromosomal data. These results show that musical relations indicate dominantly female genetic contacts. Can this observation be attributed to brides transferring mitochondrial genetic information to far populations of their fiancés, or to mothers transmitting their musical mother tongue to their children even in foreign surroundings? The answer is open. All in all, women seem to be much more important actors in the cross-cultural transfer of musical and genetic information than men, indicating more mobility during our evolutionary history as described in the section of genetic relationship of the populations.

The above-mentioned examples of simultaneously existing Appalachian–Norwegian–Hungarian variants of common contour types pose some further questions: Are there major or minor groups of common musical styles which exist simultaneously in many different cultures, and can these data be correlated with parallel genetic relations? A positive answer to these questions would reveal common crystallization points of musical cultures and verify that these connections can be traced back to historic or prehistoric reasons.

The answer to these exciting questions is given in Fig. 5, where we have shown that the largest subsets of totally common melody types exist in the quartets of the Volga–Turkish–Karachay–Hungarian and the Chinese–Turkish–Karachay–Hungarian groups with 46 and 45 totally common types, respectively. These types are situated in a continuous part of the common music map, therefore, we can state that they form a cohesive musical style with many melody types of certain common characteristics (for instance, most of them have a descending contour with a range of an octave.). Continuing the systematic search among larger subsets of the 31 cultures, we found that the ensembles having the largest overlaps are the right extensions of the above quartets: The Volga–Sicilian–Turkish–Karachay–Hungarian–Finnish–Dakota and Chinese–Mongol–Volga–Sicilian–Turkish–Karachay–Hungarian cultures have the largest sets of common melody types among the seven element groups (12 and 10, respectively). Apart from the lack of Dakota and Karachay data, the close musical relations of the former group totally correlate with the known genetic relations (See Table 1). It is regrettable that the Karachay data is not available, however, most of the genetic distances calculated for the neighbouring Chechen people show close genetic connections to Hungarian, Turkish, Sicilian and Tatar (Volga) populations. According to these data and some recently published results, a close genetic relationship of the Karachays to the mentioned populations can be expected (Sen 2010).

Turning our attention to the second septet, we found large genetic distances of the Chinese people from the members of the very compact Volga–Sicilian–Turkish–Hungarian group. Thus, groups b and e in Fig. 5 are surely not completely coherent in a genetic sense; hopefully, this anomaly may also be explained by some future analyses.

The common area of the above-discussed musical cultures in the melody map is totally separate from that of the German–{Luxembourgian–Lorrain}–French–Holland quartet, where the genetic distances—being less than 0.03 for any pair—also indicate very intensive genetic contacts (see Fig. 5c). The fact that we found a totally disjunctive area for the common "Western" and "Eastern" melody forms can be interpreted by the existence of two essentially different musical language groups. Musical examples illustrating the typical common melody forms in Western and Eastern language groups are found in Notes 2–4 (Online Resource 5–7). The relatively low number of totally common types in the Western cultures (17) shows that their pair-wise large overlaps contain mainly different types; therefore, they form a much less extended core-style than the Eastern group.

What melody forms construct the totally common musical styles simultaneously existing in so many cultures? An extended analysis of this core-style exceeds the limitations of this study; therefore, we illustrate it merely by some examples in Notes 3 and 4 (Online Resource 6–7).

These melodies, being variants of two common contour types of high ambitus, gradually descending from the octave, indicate a well-defined common style residing in all cultures classified above as members of one of the two "Eastern" groups (Fig. 5).

One can hardly imagine a different explanation for such common musical style than the existence of a common "parent language" as an initial crystallization point of more or less independent musical evolutions. Of course, this statement cannot be considered as a real conclusion of the present work. It is an assumption that needs further historical, archaeological, etc. investigation, but we can state that it is supported by the genetic correlation discussed here. Earlier studies of the contacts of Hungarian folk music have drawn the conclusion that descending, sometimes fifth transposing melodies of high ambit constitute the most important part of the contact melodies between Hungarian as well as Mari, Chuvash (Kodály 1971), Anatolian (Bartók 1949; Sipos 2000, 2001, 2006) and Mongolian (Sipos 1997) folk music. Our results confirm that the correspondences found by these independent studies were not coincidental, since they can be deduced from the existence of the common core of the "Eastern" cultures. According to the above results, most of the populations conserving this musical style in their current culture have common genetic roots; consequently, the development of this musical culture in very early times can be attributed to their common ancestors.

In conclusion, using the current methods of genetics, data mining, artificial intelligence research and musicology, we have drawn the system of musical connections of 31 cultures in Eurasia and verified that these contacts can be traced back to genetic basis in 82% of the studied cultures, while close genetic relations of populations are definitely not appropriate indicators of similar musical cultures (28%). Such a result can be easily accepted in the case of people living recently in the same area, like the Western countries, Slovakians and Hungarians, Finns and Norwegians, etc. However, a significant part of our results indicate simultaneous musical and genetic similarity of peoples who live very far from each other over a very long period. For instance, the simultaneous musical and genetic contacts of Hungarians and Norwegians, Tatars and Sicilians, Sicilians and Turks, etc., emphasize that musical contacts may arise from very early times, thus, oral musical traditions may conserve prehistoric musical forms and styles. The collaboration of genetics and computational musicology was a crucial requirement for uncovering this "hidden" connection. We identified a set of melody forms constructing the most extended common (ancient) musical style appearing in many cultures simultaneously. Furthermore, we showed that most of the cultures having this common musical heritage are also closely related in a genetic sense establishing a new interdisciplinary research field as "*music-genetics*". This result led to the conclusion that musical styles arise from a common musical parent language of historic and prehistoric origin.

Acknowledgments This work was supported by the Hungarian National Research Found (grant no. K81954), and the Network of Forensic Science Institutes (NFSI), Ministry of Public Administration and Justice. We would like to say special thanks to Dr. Eva Susa (General Director of the Network of Forensic Science Institutes) for her support. We thank both reviewers for their constructive comments and suggestions and Ati Rosselet for the English editing.

References

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB (2001) genetic evidence on the origins of Indian caste populations. Genome Res 11(6):994–1004

Bartók B (1949) On collecting folk songs in Turkey. Tempo, New Ser, No. 13, Bartok Number (Autumn, 1949), pp 15–19

Bermisheva MA, Kutuev IA, Korshunova TY, Dubova NA, Villems R, Khusnutdinova EK (2004) Phylogeographic analysis of mitochondrial dna in the nogays: a strong mixture of maternal lineages from Eastern and Western Eurasia. Mol Biol 38:516–523

Borg I, Groenen P (2005) Modern multidimensional scaling: theory and applications, 2nd edn. Springer, New York

Bosch E, Calafell F, González-Neira A, Flaiz C, Mateu E, Scheil HG, Huckenbeck W, Efremovska L, Mikerezi I, Xirotiris N, Grasa C, Schmidt H, Comas D (2006) Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann Hum Genet 70:459–487

Brandstätter A, Egyed B, Zimmermann B, Duftner N, Padar Z, Parson W (2007) Migration rates and genetic structure of two Hungarian ethnic groups in Transylvania, Romania. Ann Hum Genet 71:791–803

Cavalli-Sforza LL, Bodmer WF (1977) The genetics of human populations. W.H. Freeman and Co., San Francisco

Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA (2004) Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114:127–148

Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J (2004) Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. Eur J Hum Genet 12:495–504

Csébfalvy K, Havass M, Járdányi P, Vargyas L (1965) Systematization of tunes by computers. Studia Musicologica 7:253–257

Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI (2002) Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. Am J Hum Genet 70:1009–1014

Di Gaetano C, Cerutti N, Crobu F, Robino C, Inturri S, Gino S, Guarrera S, Underhill PA, King RJ, Romano V, Cali F, Gasparini M, Matullo G, Salerno A, Torre C, Piazza A (2009) Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. Eur J Hum Genet 17:91–99

Garbens J, Kranenburg P, Volk A, Wiering F, Veltcamp R, Grijp L (2007) Using Pitch Stability Among a Group of Aligned Query Melodies to Retrieve Unidentified Variant Melodies. In: Proceedings of the 8th ISMIR conference Vienna 2007

Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL (1997) The geographic distribution of human Y chromosome variation. Genetics 145(3):787–805

Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S (2006) Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. J Hum Genet 51:47–58

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, Sykes B (2001) mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet 68:723–737

Huron D (1996) The melodic arch in Western folksongs. Comput Musicol 10:3–23

Irwin J, Egyed B, Saunier J, Szamosi G, O'Callaghan J, Padar Z, Parsons TJ (2007) Hungarian mtDNA population databases from Budapest and the Baranya county Roma. Int J Legal Med 121:377–383

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612

Juhász Z (2000) A model of variation in the music of a Hungarian ethnic group. J New Music Res 29:159–172

Juhász Z (2006) A systematic comparison of different European folk music traditions using self-organizing maps. J New Music Res 35:95–112

Juhász Z (2009) Motive identification in 22 folksong corpora using dynamic time warping and self organising maps. In: Hirata K, Tzanetakis G (eds) Proceedings of the 10th Internapional Society for Music Information retrieval Conference. International Society for Music Information Retrieval, Kobe, pp 171–176

Juhász Z, Sipos J (2009) A comparative analysis of Eurasian folksong corpora, using self organising maps. J Interdiscip Music Stud 4:1–16

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838

Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, Henke J, Henke L, Hohoff C, Illing A, Jonkisz A, Kuzniar P, Lebioda A, Lessig R, Lewicki S, Maciejewska A, Monies DM, Pawłowski R, Poetsch M, Schmid D, Schmidt U, Schneider PM, Stradmann-Bellinghausen B, Szibor R, Wegener R, Wozniak M, Zoledziewska M, Roewer L, Dobosz T, Ploski R (2005) Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. Hum Genet 117:428–443

Kodály Z (1971) Folk music of Hungary. Corvina, Budapest

Kohonen T (1995) Self-organising maps. Springer, Berlin

Kranenburg P, Volk A, Wiering F, Veltkamp RC (2009) Musical models for folk-song melody alignment. In: Hirata K, Tzanetakis G, Yosh K (eds) 10th International Society for Music Information Retrieval Conference (ISMIR 2009), pp 507–512

Leman M (2000) An auditory model of the role of short-term memory in probe-tone ratings. Music Perception 17:481–509

Maca-Meyer N, Sánchez-Velasco P, Flores C, Larruga JM, González AM, Oterino A, Leyva-Cobián F (2003) Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). Ann Hum Genet 67:329–339

Malhi RS, Gonzalez-Oliver A, Schroeder KB, Kemp BM, Greenberg JA, Dobrowski SZ, Smith DG, Resendez A, Karafet T, Hammer M, Zegura S, Brovko T (2008) Distribution of Y chromosomes among native North Americans: a study of Athapaskan population history. Am J Phys Anthropol 137:412–424

Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Woźniak M, Miścicka-Sliwka D (2002) Mitochondrial DNA variability in Poles and Russians. Ann Hum Genet 66:261–283

Merriwether DA, Hall WW, Vahlne A, Ferrell RE (1996) mtDNA variation indicates Mongolia may have been the source for the founding population for the New World. Am J Hum Genet 59:204–212

Nasidze I, Sarkisian T, Kerimov A, Stoneking M (2002) Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. Hum Genet 112:255–261

Nasidze I, Quinque D, Ozturk M, Bendukidze N, Stoneking M (2005) MtDNA and Y-chromosome variation in Kurdish groups. Ann Hum Genet 69:401–412

Nettle D, Harriss L (2003) Genetic and linguistic affinities between human populations in Eurasia and West Africa. Hum Biol 75(3):331–344

Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP, Stoneking M (2006) Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. Hum Genet 120:334–353

Pakendorf B, Novgorodov IN, Osakovskij VL, Stoneking M (2007) Mating patterns amongst Siberian reindeer herders: inferences from mtDNA and Y-chromosomal analyses. Am J Phys Anthropol 133:1013–1027

Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Børresen-Dale AL, Underhill PA (2002) Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. Eur J Hum Genet 10:521–529

Peričić M, Barać Lauc L, Martinović Klarić I, Janićijević B, Rudan P (2005a) Review of Croatian genetic heritage as revealed by mitochondrial DNA and Y chromosomal lineages. Croat Med J 46:502–513

Peričić M, Lauc LB, Klarić IM, Rootsi S, Janićijevic B, Rudan I, Terzić R, Colak I, Kvesić A, Popović D, Sijacki A, Behluli I, Dordevic D, Efremovska L, Bajec DD, Stefanović BD, Villems R, Rudan P (2005b) High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. Mol Biol Evol 22:1964–1975

Petrejčíková E, Soták M, Bernasovska J, Bernasovský I, Sovicová A, Bôžiková A, Boroňová I, Gabriková D, Švíčková P, Mačeková S, Čherhová V (2010) The genetic structure of the Slovak population revealed by Y-chromosome polymorphisms. Anthoropol Sci 118:23–30

Pimenoff VN, Comas D, Palo JU, Vershubsky G, Kozlov A, Sajantila A (2008) Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. Eur J Hum Genet 16:1254–1264

Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K (2004) Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. Am J Hum Genet 74:827–845

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526–1543

Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis. Ver 2.000, Genetics and Biometry Lab, Department of Anthropology, University of Geneva

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. Science 290:1155–1159

Sen A (2010) The genetic history of the Karachays: insights from mtDNA and Y-chromosome evidence. https://www.sas.upenn.edu/anthropology/undergraduate/theses/Sen

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000) Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet 66:262–278

Sipos J (1997) Similar musical structure in Turkish, Mongolian, Tungus and Hungarian folk music. In: Berta Á (ed) Historical and linguistic interaction between Inner-Asia and Europe. Szeged, pp 305–317

Sipos J (2000) In the wake of Bartók in Anatolia. European Folklore Institute, Budapest (Bibliotheca Traditionis Europeae 2)

Sipos J (2001) Egy most felfedezett belső-mongóliai kvintváltó stílus és magyar vonatkozásai. Ethnographia 112, pp 1–80

Sipos J (2006) Comparative analysis of Hungarian and Turkic folk music. Edition of TIKA (Türk İşbirliği ve Kalkınma İdaresi Başkanlığı) and the Hungarian Embassy, Ankara

Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, Wallace DC (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. Ann Hum Genet 69:67–89

Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli EL, Tolk HV, Reidla M, Metspalu E, Pliss L, Balanovsky O, Pshenichnov A, Balanovska E, Gubina M, Zhadanov S, Osipova L, Damba L, Voevoda M, Kutuev I, Bermisheva M, Khusnutdinova E, Gusar V, Grechanina E, Parik J, Pennarun E, Richard C, Chaventre A, Moisan JP, Barác L, Pericić M, Rudan P, Terzić R, Mikerezi I, Krumina A, Baumanis V, Koziel S, Rickards O, De Stefano GF, Anagnou N, Pappa KI, Michalodimitrakis E, Ferák V, Füredi S, Komel R, Beckman L, Villems R (2004) The western and eastern roots of the Saami–the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. Am J Hum Genet 74:661–682

Toiviainen P (1996) Optimizing auditory images and distance metrics for self-organizing timbre maps. J New Music Res 25:1–30

Toiviainen P (2000) Symbolic AI versus connectionism in music research. In: Mirinda E (ed) Readings in music and artificial intelligence. Harwood Academic Publishers, Amsterdam, pp 47–68

Toiviainen P, Eerola T (2002) A computational model of melodic similarity based on multiple representations and self-organizing maps. In: Stevens C, Burham D, McPherson G, Schubert E, Rewick J (eds) Proceedings of the 7th International Conference on Music Perception and Cognition, Causal Productions, Sidney, pp 236–239

Underhill PA, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu Rev Genet 41:539–564

Völgyi A, Zalán A, Szvetnik E, Pamjav H (2009) Hungarian population data for 11 Y-STR and 49 Y-SNP markers. Forensic Sci Int Genet 3:e27–e28

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244–10249

Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurles ME, Yang H, Tyler-Smith C (2006) Male demography in East Asia: a north-south contrast in human population expansion times. Genetics 172:2431–2439

Yao YG, Kong QP, Wang CY, Zhu CL, Zhang YP (2004) Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. Mol Biol Evol 21:2265–2280

Zastera J, Roewer L, Willuweit S, Sekerka P, Benesova L, Minarik M (2010) Assembly of a large Y-STR haplotype database for the Czech population and investigation of its substructure. Forensic Sci Int Genet 4:e75–e78