Rev. Esp. Antrop. Fís. (2019)
Vol. 40: 27-37

World distribution of genetic diversity associated to musical ability

VILLANUEVA B. Y ALONSO S.

Dep. Genetics, Physical Anthropology and Animal Physiology- Fac Sciences and Technology.
University of the Basque Country UPV/EHU. Barrio Sarriena s/n 48940 Leioa Bizkaia.

Corresponding Author: bvillanueva.lan@gmail.com

Keywords:

SNP
Musical ability
PCA,
1000 Genomes Project

Recibido: 03-12-2019
Aceptado: 20-02-2020

Palabras claves:

SNP
Habilidad musical
ACP
Proyecto 1000 Genomas

RESUMEN

Los estudios de asociación del genoma completo (GWAS en inglés) se han convertido en una fuente de información cada vez más importante. Son muy útiles a la hora de correlacionar el genotipo humano con el fenotipo expresado. Hoy en día, incluso se ha conseguido información respecto a la variabilidad genética que está relacionada con las habilidades de percepción y producción musicales. Usando esa información, hemos llevado a cabo unos análisis de componentes principales con el fin de descubrir cómo está distribuida la variación genética relacionada con la habilidad musical en poblaciones mundiales. Además, también se calculó una puntuación para la habilidad musical esperada en cada uno de los individuos. Finalmente, descubrimos que los individuos se distribuyen en relación a su origen geográfico y en relación con su puntuación musical, siendo las poblaciones africanas las que mayor puntuación musical obtienen.

ABSTRACT

Genome-wide association studies (GWAS) have become a very important source of information. They are very useful at the time of correlating human genotype with the expressed phenotype. Today we even have information about genetic variability that is correlated to our music perception and music production abilities. Using that information, we performed a principal component analysis to discover how this variability (SNPs) is distributed geographically in world-wide human populations. In addition to that, we calculated a score for expected musical ability for each individual. Finally, we found that individuals show a distribution according to geography as well as a distribution based in their expected musical abilities, being African populations the ones with the highest expected musical ability.

ISSN: 2253-9921

Introduction

Genome-wide association studies (GWAS) are a way for scientists to identify inherited genetic variants associated with risk of disease or a particular trait. Using cases and controls and by looking at the frequencies of different genotypes present on these two groups; it is possible to associate a specific disease or trait to a determined allele. This method usually surveys the whole genome by means of a wide battery of Single Nucleotide Polymorphisms (SNPs). With more than 3.400 published papers and 62.000 unique SNP-related associations, GWAS are a very important source of information used by many researches (Ensembl blog, 2018).

These types of studies are not only useful from a biomedical or clinical perspective, but also for other phenotypic associations. This includes traits like lactase persistence, skin phenotype, adaptation to high altitudes, face morphology, or other traits like musical aptitude or sport performance (Han et al., 2008; Alkorta-Aranburu et al., 2012; Beleza et al., 2013; Pitsiladis et al., 2013; Ting Tan et al., 2014; Qiao et al., 2018).

With that idea in mind, several research groups have performed GWAS with the aim of linking SNPs or genes, to the individual's musical ability. As a result, more than 100 SNPs have been found related to this particular ability, which is in fact, a compound set of abilities ranging from perception to creative abilities. This can be measured for instance, by the Karma Music Test (KMT), which is a test that measures some potentially innate musical cognitive operations (Justus & Hustler, 2005; Karma, 2007). The test gives us information about the capability of the subject to perceive rhythms and sound patterns. Similarly, Carl Seashore´s subsets of pitch (SP) and time (ST) discrimination tests are another common type of tests used for determining musical perception (Seashore et al., 1960). In this case we can get information about the capability of the participants to discriminate in pitch and duration of the tones, a basic ability for good music perception. On many occasions, a combined test score (COMB) which is the sum of KMT, SP and ST (Ukkola-Vuoti et al., 2013; Oikkonen et al., 2015) is also calculated.

Among the abilities measured we can highlight Absolute pitch (AP), also known as "perfect pitch", which is the infrequent ability to identify or produce pitches perfectly without the help of an external reference (Profita & Bidder, 1988). This trait is used for measuring both music production and music perception abilities (Theusch et al., 2009). On the other hand, Pitch Production Accuracy (PPA) is the capability of an individual to correctly perform a pitch after hearing a specific pitch sample (Haumann et al., 2007).

These tests have been commonly applied by many researchers in European and Asian populations (Theusch et al., 2009; Hansoo et al., 2012; Oikkonen et al., 2015). Other music production abilities such as *arranging*, *composing* and more can also be tested in the individuals (Oikkonen et al., 2016). In this case only the presence or absence of the ability is checked and used for measuring music ability.

Interestingly, most of the publications do not only follow a typical GWAS approach, they also search on the segregation and transmission of these polymorphisms among individuals of different generations in the same family. This particular type of analysis is common in association studies regarding musical abilities and on many occasions this part of the paper has more weight than the GWAS itself.

Markers that have a greater frequency in individuals with a trait of interest across or within pedigrees are identified by means of linkage analysis, then parametric linkage analysis is done to determine this relationship between the marker locus and the trait locus (Schnell & Sun, 2012). Within this context we can define the posterior probability of linkage (PPL) and the posterior probability of linkage disequilibrium (PPLD). PPL and PPLD are interpreted as the probability of a trait gene being linked or associated to the marker examined (Vieland et al., 2011; Oikkonen et al., 2015). The ratio of the above-mentioned probabilities is the LOD score, a high value of which represents a significant evidence of linkage between the two.

Thus, the main objective of this project is to analyse how the genetic variability (SNPs) associated

to musical ability is distributed geographically in world-wide human populations. If this distribution were different from the global distribution of human genetic diversity, we could argue that other factors beyond demography are exerting an influence on this trait.

Materials and Methods

Google Scholar (https://scholar.google.es/) and NCBI (https://www.ncbi.nlm.nih.gov/) were used as search databases, using keywords like GWAS, musical ability, SNP, music perception and music production. We selected 6 papers Theusch et al., (2009), Hansoo et al., (2012); Ukkola-Vuoti et al., (2013); Ting Tan et al., (2014); Oikkonen et al., (2015) and Oikkonen et al. (2016) out of which 90 SNPs related to musical ability were retrieved for analysis, (the SNPs were sequenced out of European and Asian populations most of the times). Some of them were related to music perception (KMT, SP, ST or COMB), while others are related to music production abilities (PPA, arranging, composing etc.). They have not been separated among categories, so they are considered just as "SNPs related to music ability", (or "musical SNPs" for short).

Next, genomic data of individuals from the 1kGP (The 1000 Genomes Project Consortium, 2015) was downloaded. Out of the 1092 individuals available we initially selected 264 random individuals from 13 world populations (approximately 20 individuals per population), which were grouped into 10 regions: African-American (ASW population), North Europe (CEU and GBR populations), South Europe (IBS and TSI populations), Finland (FIN population), East Asia (CHB and JPT populations), East Africa (LWK population), West Africa (YRI population), and Central America (MXL, PUR and CLM populations). The 1kGP abbreviations and their meanings are available here: https://www.internationalgenome.org/category/population/. Note that the Finish population was not included in any other European population because it usually appears as an outlying population (Huckins et al., 2014). Data was downloaded from the Phase 1 repository as VCF files, one per chromosome (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/integrated_call_sets/).

In a second step, another sample composed of just European individuals (N = 291) was also considered for the analysis (the European individuals from the previous world sample were kept). In this case population acronyms were kept as in the 1kGP: FIN, CEU, GBR, IBS and TSI.

Out of the whole genomes of the individuals downloaded, the genotypic information on the 90 musical SNPs was filtered using VCFtools software (Danecek et al., 2011). As controls, 4 additional sets of random SNPs were chosen and filtered out of the VCF file which contains the whole genome information of the original 264 individuals. These random SNPs were chosen so they had the same characteristics in terms of location (same chromosome region for example) as the target musical SNPs by means of SNPsnap (Pers et al., 2015).

Then we carried out a Principal Component Analysis (PCA) with the VCF file containing the genotypes of the 90 musical SNPs for 264 individuals from all over the world. This was carried out in Galaxy (http://usegalaxy.org; Afgan et al., 2018) a free online portal to perform multiple bioinformatic analysis. The resulting VCF file from VCFtools processing was uploaded to the Galaxy server where, a Principal Component Analysis was done. The same procedure was applied to the 4 random matched, control VCF files (world population) and the VCF file containing the genotypes of the (same) 90 musical SNPs for the European population (N = 291).

Additionally, a music score based on the number of alleles that favour music ability was calculated for each individual. Unfortunately, GWAS on music ability are often focused on the heritability of these genotypes (by calculating LOD, PPL or PPLD scores). Out of the SNPs discovered, few of them have detailed data about the specific alleles that have a positive and a negative effect on the musical-ability phenotype. Most of them just conclude that a region or SNP is related to the trait without specifying the effect of the alleles. Out of the detailed 19 SNPs, there are some that are in

linkage disequilibrium (LD) with each other, so they cannot be used for the musical score, as we would be counting the same information more than once. At the end, just 4 independent SNPs were chosen for this task (Table 1).

Table 1: The 4 SNPs used for the calculation of the musical score of each individual.

Chromosome Nº	Base pair position	SNP	Effect allele
4	96367953-96367954	rs9307160	C
4	110958163-110958164	rs6845765	C
4	113355410-113355411	rs17628408	G
4	115640580-115640581	rs12510781	G

Excel was used to calculate the musical score of each individual. Per each effect allele of the individuals a value of 1 was added to their scores (i.e. an homozygous genotype for the allele associated to the trait would add 2 points, whereas a heterozygous genotype would only add 1 point, and 0 if homozygous for the non-associated allele). With 4 different SNPs each with 2 possible alleles, the total scores could range from 0 to 8. Afterwards, as the highest and lowest marks were infrequent this 0-8 scale was transformed into a categorical scale (A to D, in which A is the highest and D is the lowest category, see Table 2).

Table 2: The original possible numerical scores for expected musical ability and the resumed categorical scale from A to D. Each categorical scale covers a quartile of the total number of scores. "*" there were no individuals with that score.

Numerical music score	Categorical music score
0	*
1	*
2	D
3	D
4	C
5	B
6	A
7	A
8	*

All subsequent data were processed with R v3.5.3 (R Core Team, 2019) and RStudio v1.1.463 (RStudio Team, 2016). First, the optimal number of clusters (in which the individuals are classified according to the PCA coordinates) was calculated using mclust v5.4.3 (Scrucca et al., 2016) an R package which can calculate model-based optimal number of clusters. Afterwards, ggplot2 (Wickham, 2016) and Kmisc (Kevin, 2013) were used for plotting the generated data from the PCA and drawing the areas which define the clusters. When looking for significant differences in the data an exact Fisher´s test was performed using the VassarStats web service (VassarStats, 2019) and rcompanion v2.2.1. R package (Mangiafico, 2019). A p-value of 0.05 is considered as the nominal threshold for statistical significance but the Bonferroni correction was applied when multiple tests were performed (Bonferroni, 1936).

Results

The optimal number of clusters for the 4 PCAs using the 4 sets of control SNPs in the world population was always 3 (Figure 1). The clustering itself was very similar in all 4 sets (Figure 2). As the disposition of the second set was considered the best for later explanations, only that set is referred to when talking about control PCA with random SNPs (Figure 3). The 3 different clusters show the typical geographical clustering of the individuals. Cluster 1 is composed of African individuals, cluster 2 is mostly of Asian individuals and cluster 3 is a mixture of European and the Central American populations (Figure 3).

In the PCA with the SNPs related to musical ability (musical SNPs), the optimal number of clusters was 4, one cluster more than in the case of the control SNPs. In Figure 4 we can see that cluster 1 is kept almost the same in both PCAs. Cluster 2 loses some individuals, but it is still mainly formed by Asian individuals. However, individuals that form cluster 3 (in the PCA of the control SNPs) suffer a severe rearrangement. So much, that we can consider that they form two different clusters in the PCA with musical SNPs (named as clusters 3 and 4).

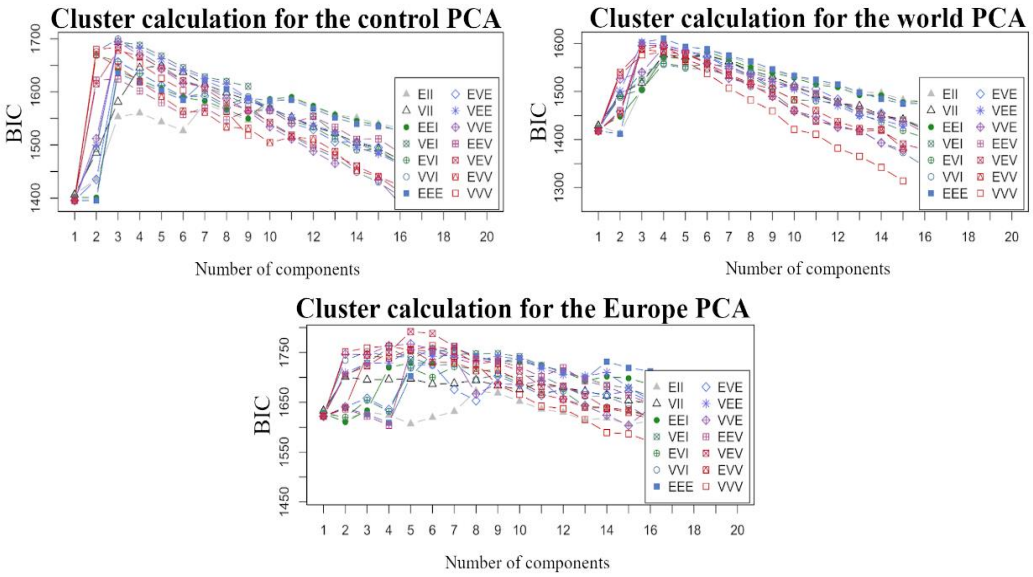

Figure 1: Model-based optimal number of clusters calculations performed using mclust R package. The X axis represents the number of clusters and in the Y axis Bayesian Information Criterion (BIC) is showed for different clustering methods (EII to VVV).

Some Asian individuals now appear in this two clusters but still, it is mostly European and Central American individuals the ones who constitute these two new clusters (Figure 4).

The distributions of musical scores in the world analysis with musical SNPs point out that cluster 1 shows the highest mean musical score (4.85). On the other hand, cluster 2 has the lowest score (3.35). We can also see that cluster 3 has a slightly higher musical score than cluster 4 (see Table 3).

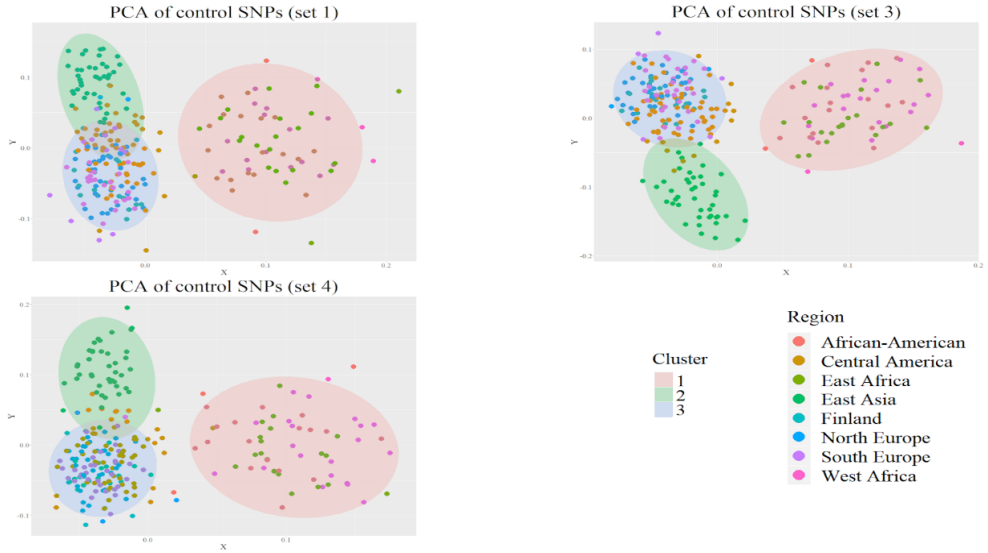

Figure 2: Principal Component Analysis (PCA) made by using the first, third and fourth set of control SNPs (X and Y eigenvectors explain 15.6%, 17.5% and 16,7% of variance respectively). Individuals are represented with a dot and the respective region of each one is indicated by a colour. Clusters are drawn as big coloured circles and the corresponding number of cluster is shown in the legend.

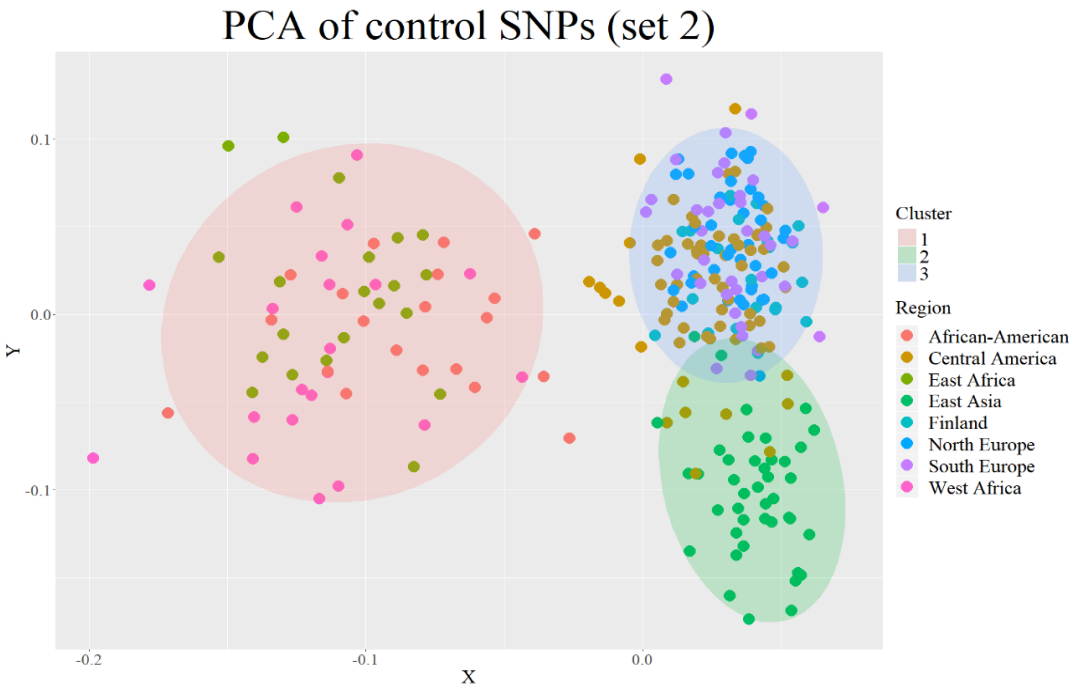

Figure 3: Principal Component Analysis (PCA) made by using the second set of control SNPs (X and Y eigenvectors explain 18.1% of variance). Individuals are represented with a dot and the respective region of each one is indicated by a colour. Clusters are drawn as big coloured circles and the corresponding number of cluster is shown in the legend.

Figure 5 shows a more in-depth view of how categorical (A to D) musical score of individuals is formed in each cluster. Data is shown as proportions up to 1. In cluster 1 we can see many individuals with the best possible score (A). In cluster 2 this A score is absent, and we can find many type D individuals. Clusters 3 and 4 have mainly individuals with B and C scores.

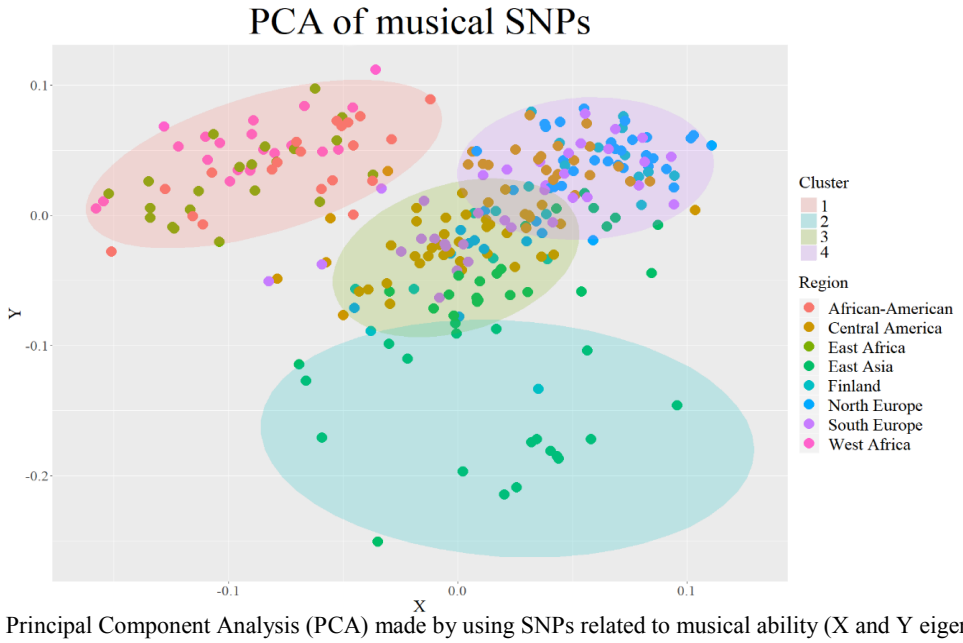

Figure 4: Principal Component Analysis (PCA) made by using SNPs related to musical ability (X and Y eigenvectors explain 15.9% of variance). Individuals are represented with a dot and the respective region of each one is indicated by a colour. Clusters are drawn as big coloured circles and the corresponding number of cluster is shown in the legend.

Table 3: Mean musical score per cluster of the world and European populations. Number of individuals is also shown as "N".

World population			
Cluster	**Mean individual musical score**	**Standard Deviation**	**N**
1	4.859	0.990	64
2	3.353	0.996	17
3	4.106	0.900	85
4	4.031	0.913	98
European population			
Cluster	**Mean individual musical score**	**Standard Deviation**	**N**
1	3.989	0.905	95
2	4.425	1.130	40
3	3.930	0.910	43
4	4.048	0.930	84
5	3.759	0.951	29

In cluster 1, Bonferroni-corrected p-values of the differences in the A score proportions were all significant ($p = 0.0078$, $p = 0.0005$ and $p = 0.0003$ respectively for cluster 1 vs 2, 1 vs. 3 and 1 vs. 4). We can see the same when looking at individuals with D score ($p = 0.0001$, $p = 0.0034$ and $p = 0.0048$ respectively for cluster 1 vs 2, 1 vs 3 and 1 vs 4). On the other hand, no significant differences were found between the rest of the clusters (2, 3 and 4). Clusters 3 and 4 suppose the main difference between the control and world analysis with musical SNPs (Figures 3 and 4), but we found no differences in geographical origin of individuals, nor with musical score.

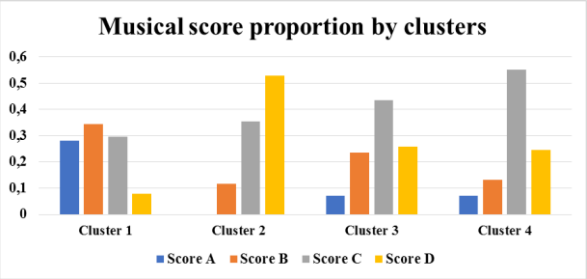

Figure 5: Proportional frequency of each musical score by each cluster of the world population.

Prompted by the split observed in the European population, we decided to perform a PCA with an expanded set of European individuals only. This deeper analysis of the European population showed that the optimal number of clusters was 5 (see Figure 6). Geography could not explain that distribution as Fisher exact tests between clusters resulted on non-significant p-values (data not shown). In terms of mean musical score, cluster 2 has the highest mean musical score (4.425), followed by clusters 4, 1 and 3 respectively (Table 3). The last one is cluster 5 (3.75).

The differences in musical score proportion between different clusters were tested like before, by a Fisher exact test. In this case significant p-values showed up when comparing A score proportion of cluster 2 against clusters 1 and 3 ($p = 0.012$ and $p = 0.010$, respectively). However, differences with clusters 4 and 5 ($p = 0.0218$ and $p = 0.0441$, respectively) were non-significant (after Bonferroni correction), even if cluster 5 has only one A type individual (Figure 7). Also, a low p-value was obtained when comparing D score in cluster 2 against cluster 5 ($p = 0.0485$, no Bonferroni correction needed in this case, as it was a single comparison). All the proportions are shown in Figure 7.

The distribution of the European individuals of world clusters 3 and 4 into the second European analysis is shown in table 4. The segregation of people is very different in each cluster: >75% of individuals from world cluster 4 appear in European clusters 1 and 3 but we cannot find there any individual from world cluster 3. Instead 75% of them appear in European clusters 2 and 4. These 2 distributions are significantly different ($p = 7.794e\text{-}13$).

Discussion

After analysing the geographical distribution of a set of SNPs of genes related to the inner ear and other neurocognitive functions, which are key for a good musical ability (Oikkonen et al., 2015), the obtained results show that a different clustering of individuals arises when classifying the same worldwide populations with random SNPs, or SNPs related to musical ability (Figure 1). Thus, cluster 1 shows the highest musical score (Table 3), which is probably due to its highest proportion of type A individuals and at the same time, the lowest proportion of type D individuals

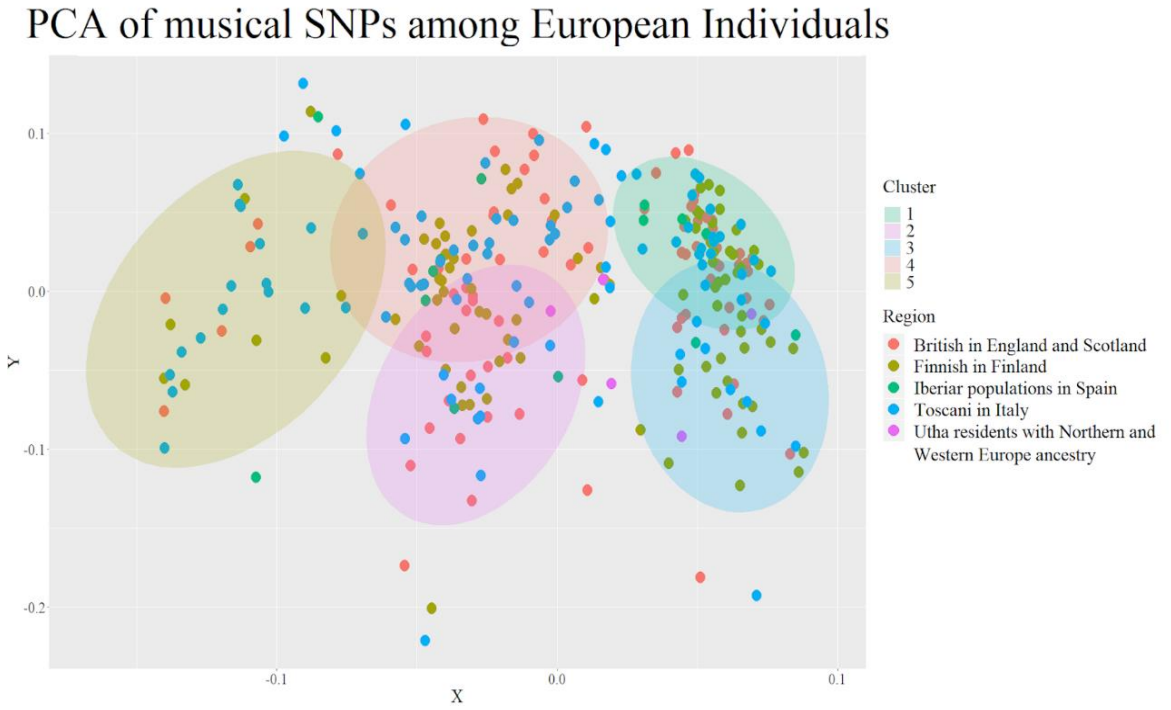

Figure 6: Principal Component Analysis (PCA) made by using SNPs related to musical ability with the European population (X and Y eigenvectors explain 13.8% of variance). Individuals are represented with a dot and the respective country of each one is indicated by a colour. Clusters are drawn as big coloured circles and the corresponding number of cluster is shown in the legend.

(Figure 5). In addition to that, cluster 1 is composed of individuals from Africa of with African ancestry (Figure 4). These proportions are quite different from those found in every other cluster of the world analysis with musical SNPs. Even with the Bonferroni correction, the differences of type A individuals and type D individuals were significant against any other cluster of the world analysis (musical SNPs). In addition to this, cluster 1 does not change substantially from the control cluster 1 (Figure 3). Considering these evidences, it seems that cluster 1 is determined by both the geographical origin of the individuals and by the musical score, to some extent. This suggests that there is some relationship between an African ancestry and a high musical score. Note that musical score aims to represent the expected musical ability, which could be a complex interaction between genetic and environmental effects.

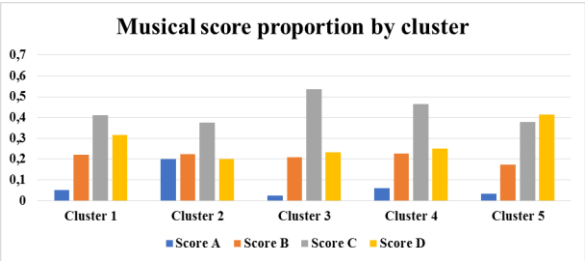

Figure 7: Proportional frequency of each musical score by each cluster of the European population.

The effect alleles of the 4 SNPs used to build the score where defined in previous studies on Mongolian populations (Hansoo et al., 2012). Despite that, the highest musical scores where not found in Asian populations so for the exploratory purposes of this work (and taking obviously into consideration the limitations inherent to our work), we can tentatively consider the SNPs as valid for the scoring task. Cluster 2 is the opposite of cluster 1 in terms of musical score (world analysis with musical SNPs). It has the highest

Table 4: Distribution of only the European individuals present in clusters 3 and 4 of the world analysis into the newly formed clusters of the only European analysis. The distribution is shown in proportions out of the total of European individuals of the world cluster 3 and cluster 4. The European destiny clusters are ordered from left to right according to the mean score of each cluster (left highest, right lowest). Number of individuals is also shown as "N".

	Proportion of cluster of origin from first (world) analysis into cluster of destination in the second (European) analysis					
Cluster of origin from the first (world) analysis	**Cluster 2**	**Cluster 4**	**Cluster 1**	**Cluster 3**	**Cluster 5**	**N**
From Cluster 3	0.2500	0.5000	0	0	0.2500	85
From Cluster 4	0.0784	0.1176	0.5490	0.2549	0	98

proportion of type D individuals (52%), no type A individuals at all (Figure 5) and the lowest mean musical score (3.35, Table 3). Despite that, the low musical score may have had a small influence in this PCA as differences in proportion of individuals with A and D scores are significantly different only against the ones seen in cluster 1, not against the ones in clusters 3 and 4. It is important to mention that cluster 2 also has a very marked geographical origin: all individuals except one are from East Asia. Like in the case of cluster 1, the distribution does not change substantially between the control analysis and world analysis. Further work with an increased number of SNPs from which we could build a musical score would be convenient to reveal possible associations.

Looking closer at the distribution of European individuals from the world analysis (which were majority in clusters 3 and 4) in the second analysis with the extended European individuals, revealed that individuals from world cluster 3 were distributed very differently from individuals of world cluster 4. This spatial distribution is statistically significant ($p = 7.794e-13$) and is based on SNPs related to musical ability (Table 4). Despite that, it cannot be well explained by the musical score because, in this case, the musical score distribution does not correlate with the PCA distribution. Possibly due to the small number of SNPs used for the calculation of the "rough musical score" and/or a lack of precision in the estimation of the individual weight of each SNP in the calculation of the musical score. Including more individuals from the 1kGP could also improve the results and reduce the sampling effect.

In conclusion, performing a PCA of world populations with SNPs related to musical ability, reflects both geographical distribution and musical score (as calculated in this work). African populations are not only clearly separated from the rest populations geographically, but also by their mean musical score, suggesting a relationship between African population and good expected musical ability. Taking our results as preliminary and again considering the limitations of this work, we could speculate that environmental forces other than demography, have also shaped the genetic distribution of this reduced set of SNPs associated to musical abilities.

Acknowledgments

SA is funded by the Basque Government (Project IT1138-16 to consildated groups). The authors wish to thank the two anonymous reviewers for significantly improving the quality of the manuscript.

References

Afgan E., Baker D., Batut B., van den Beek M., Bouvier D., Čech M., Blankenberg D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46(1): 537-544.

Alkorta-Aranburu G., Beall CM., Witonsky D.B., Gebremedhin A., Pritchard J.K., Di Rienzo A. (2012). The Genetic Architecture of Adaptations to High Altitude in Ethiopia. PloS Genet 8(12): e1003110. doi: https://doi.org/10.1371/journal.pgen.1003110.

Beleza S., Johnson N.A., Candille S.I., Absher D.M., Coram M.A., Lopes J., Campos J., Araujo I.I., Anderson T.M.,

Vilhjámsson B.J., Nordborg M., Correia e Silva A., Shriver M.D., Rocha J., Barsh G.S., Tang H. (2013). Genetic Architecture of Skin and Eye Color in an African-European Admixed Population. PloS Genet 9(3): e1003372. doi: https://doi.org/10.1371/journal.pgen.1003372.

Bonferroni C.E. (1936). Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8(3): 62.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R., 1000 Genomes Project Analysis Group (2011). The Variant Call Format and VCFtools. Bioinformatics 27(15): 2156-2158.

Ensembl blog (2018). Ten years of the GWAS Catalog-Past, present and future, URL: http://www.ensembl.info/2018/06/28/ten-years-of-the-gwas-catalog-past-present-and-future/, 7/06/2019.

Han J., Kraft P., Nan H., Guo Q., Chen C., Qureshi A., Hankinson S.E., Hu F.B., Duffy D.L., Zhao Z.Z., Martin N.G., Montgomery G.W., Hayward N.K., Thomas G., Hoover R.N., Chanock S., Hunter D.J. (2008). A Genome-Wide Association Study Identifies Novel Alleles Associated with Hair Color and Skin Pigmentation. PloS Genet 4(5): e1000074 doi: https://doi.org/10.1371/journal.pgen.1000074.

Hansoo P., Seungbok L., Hyun-Jin K., Young Seok J., Jong-Yeon S., Dongwan H., von Grotthuss M., Lee D-S, Park C., Hayeon Kim J., Boran K., Yoo Y.J., Sung-Il C., Joohon S., Lee C., Jong-Il K., Jeong-Sun S. (2012). Comprehensive genomic analyses associate UGT8 variants with musical ability in a Mongolian population. J Med Genet 49 (12): 747-752.

Haumann S., Muhler R., Ziese M., Specht H. (2007). Discrimination of musical pitch with cochlear implants. HNO 55: 613-619.

Huckins L., Boraska V., Franklin C., Floyd J., Southam L., GCAN; WTCCC3, Sullivan P.F., Bulik C.M., Collier D.A., Tyler-Smith C., Zeggini E., Tachmazidou I. GCAN; WTCCC3. (2014). Using ancestry-informative markers to identify fine structure across 15 populations of European origin. Eur J Hum Genet 22: 1190-1200.

Justus T., Hustler JJ. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. Music Percept 23: 1-27.

Karma K. (2007). Musical Aptitude Definition and Measure Validation: Ecological Validity can Endanger the Construct Validity of Musical Aptitude Tests. Psychomusicology 19(2): 79-90.

Kevin U. (2013). Kmisc: Kevin Miscellaneous. R package version 0.5.0. URL: https://CRAN.R-project.org/package=Kmisc.

Lowry R., VassarStats: Website for Statistical Computation, URL: http://vassarstats.net/tab2x2.html, 16/06/2019.

Mangiafico S. (2019). rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.2.1. URL: https://CRAN.R-project.org/package=rcompanion.

Oikkonen J., Huang Y., Onkamo P., Ukkola-Vuoti L., Raijas P., Karma K., Vieland V.J., Järvelä I. (2015). A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. Mol Psychiatry 20: 275-282.

Oikkonen J., Kuusi T., Peltonen P., Raijas P., Ukkola-Vuoti L., Karma K., Onkamo P., Järvelä I. (2016). Creative Activities in Music-A Genome-Wide Linkage Analysis. PloS ONE 11(2): e0148679. doi: https://doi.org/10.1371/journal.pone.0148679.

Pers T.H., Timshel P., Hirschhorn J.N. (2015). SNPsnap: A Web-based tool for identification and annotation of matched SNPs. Bioinformatics 31(3): 418-420.

Pitsiladis Y., Wang G., Wolfarth B., Scott R., Fuku N., Mikami E., He Z., Fiuza-Luces C., Eynon N., Lucia A. (2013). Genomics of elite sporting performance: What little we know and necessary advances. Br J Sports Med 47(9): 550-555.

Qiao L., Yang Y., Fu P., Hu S., Zhou H., Peng S., Tan J., Lu Y., Lou H., Lu D., Wu S., Guo J., Jin L., Guan Y., Wang S., Xu S., Tang K.(2018). Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction. J Genet Genomics 45(8): 419-432.

R Core Team. (2019). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. URL: https://www.R-project.org/.

RStudio Team. (2016). RStudio: Integrated Development for R. RStudio. URL: http://www.rstudio.com.

Schnell A.H., Sun X. (2012). Model-based linkage analysis of a quantitative trait. In R.C. Elston, J. M. Satagopan, S. Sun (Eds.). Statistical Human Genetics: Methods and Protocols: 263–284. New York: Humana Press, New York.

Scrucca L., Fop M., Murphy T.B., Raftery A.E. (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. The R Journal 8(1): 205-233.

Seashore C.E., Lewis D., Saetveit J.G. (1960) Seashore measures of musical talents. Manual. New York: The Psychological Corp.

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. Nature 526: 68-74. doi: https://doi.org/10.1038/nature15393.

Theusch E., Basu A., Gitschier J. (2009). Genome-wide Study of Families with Absolute Pitch Reveals Linkage to 8q24.21 and Locus Heterogeneity. Am J Hum Genet 85: 112-119.

Ting Tan Y., McPherson G.E., Peretz I., Berkovic S.F., Wilson S.J. (2014). The genetic basis of music ability. Front Psychol 5: 658. doi: https://doi.org/10.3389/fpsyg.2014.00658.

Ukkola-Vuoti L., Kanduri C., Oikkonen J., Buck G., Blancher C., Raijas P., Karma K., Lähdesmäki H., Järvelä I. (2013). Genome-Wide Copy Number Variation Analysis in Extended Families and Unrelated Individuals Characterized for Musical Aptitude and Creativity in Music. PloS ONE 8(2): e56356. doi: 10.1371/journal.pone.0056356.

Vieland V.J., Huang Y., Seok S.C., Burian J., Catalyurek U., O'Connell J., Segre A., Valentine-Cooper W. (2011). Kelvin: a software package for rigorous measurement of statistical evidence in human genetics. Hum Hered 72: 276-288.

Wickham H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.