Published in final edited form as:
Ann Hum Genet. 2010 November ; 74(6): 516–524. doi:10.1111/j.1469-1809.2010.00612.x.

Assessment of genetic ancestry and population substructure in Costa Rica by analysis of individuals with a familial history of mental disorder

Maia Segura-Wang[1], **Henriette Raventós**[1,2], **Michael Escamilla**[3], and **Ramiro Barrantes**[1,*]

[1]Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica

[2]Centro de Investigación en Biología Celular y Molecular, Universidad de Costa Rica, San José, Costa Rica

[3]Departments of Psychiatry and Cellular and Structural Biology, Psychiatric Genetics Research Center, University of Texas Health Science Center at San Antonio, San Antonio, Texas

Abstract

The population of Costa Rica has been considered valuable for locating susceptibility genes of complex disorders because of historical events and a gradual admixture process. We present an assessment of 426 unrelated individuals with a familial history of mental disorder and with ancestors born in the Central Valley, genotyped at 730 microsatellites to evaluate genetic diversity, ancestry and substructure at the general and regional population levels using quantitative methods. Low population substructure was found. Estimated mean ancestry proportions were 54%, 32% and 13% for European, Amerindian and African components respectively, with some regional variation. The Fst values obtained confirm the largest genetic similarity to Europeans. Subdivision of the Amerindians into individual populations revealed strong similarity to Chibchan groups. Analysis of the African ancestry showed high similarity to West and Central African populations. Gene ancestries from other African areas were also detected, probably resulting from ancestral admixture within Africa prior to colonial times. Our analyses show, in an ethnohistorical-genetic context, that gene flow and admixture are important components of Costa Rican population history. The results confirm the need to consider the particular regional genetic structure, the effects of genetic drift and the ancestry when designing and interpreting investigations of genetic traits in this population.

Keywords

Genetic structure; ancestry; admixture; STRs; mental disorders; Costa Rica

Introduction

The assessment of the genetic structure and evolutionary history of human populations is essential to properly design and interpret studies for the identification of genetic variants related to simple and complex inherited traits. Furthermore, it is also particularly important for the analysis of admixture mapping approaches, association and linkage studies, and the estimation of ancestry related risk differences in complex genetic diseases (Hoggart et al.,

*Corresponding author: Dr R Barrantes, Escuela de Biología, Universidad de Costa Rica, 2060 San José, Costa Rica. Tel: (506)25115622; ramiro.barrantes@ucr.ac.cr.

Supporting Information: Additional Supporting Information may be found in the online version of the article:

2003). These kinds of studies have been favored by the increasing availability of large numbers of informative markers, consequently promoting the development of new statistical and bioinformatics tools that allow the evaluation of the structure of human populations at a genome-wide level.

Undetected population stratification can generate spurious associations between phenotypic and genetic traits, producing false-positive results in association tests (Hoggart et al., 2003). Therefore, proper analysis of the structure and evolutionary history of a population are fundamental prior to explaining the relevance of specific genetic results. Previous linkage and association studies have often been based on *a priori* assumptions (not formally tested) about genetic homogeneity, as has occurred in the past for the population of the Central Valley of Costa Rica (León et al., 1992).

The history of the Costa Rican population, as well as other populations in Latin America (Sans, 2000), involves a complex process of admixture, mainly between groups of Amerindian, European and African ancestry, which constituted a multiethnic population by the 18th century (Meléndez, 1982; Lohse, 2005). However, most population genetic studies in the last 30 years in Costa Rica have focused primarily on Amerindian groups (Barrantes et al., 1990; Wang et al., 2007) and a few on African groups (Madrigal, 2006). Detailed investigations of the admixed population structure and ancestry are scant, and those available include a small number of individuals and markers, or include only specific geographic regions (Morera et al., 2003; Campos-Sánchez et al., 2006; Wang et al., 2008), limiting the understanding of the genetic constitution of this population, and therefore, its value for the detection of susceptibility genes.

Additionally, inference of the geographic origin of individuals is difficult to ascertain using data from broad ethnohistorical and anthropological investigations. Consequently, a thorough multidisciplinary approach, including genetic evidence, is necessary to determine the sources from which individuals originated and to estimate the proportions in which different parental populations contributed to the admixture process. Considering the above statements, this investigation intended to describe the nuclear genetic variability and the population substructure of the Costa Rican population descended from Central Valley ancestors, and to determine the possible ancestry of the populations that gave rise to the current admixed population. This is the first study to attempt a more precise reconstruction of the demographic history of the population of Costa Rica through different quantitative methods, together with support from ethnohistorical data, and the analysis of over 400 individuals genotyped at more than 700 highly polymorphic STR markers. The results obtained will assist in the analysis of future genetic studies in this population.

Materials and Methods

Samples and markers

Subjects were originally recruited for a population based association study of schizophrenia susceptibility genes in the Costa Rican Central Valley (National Institute of Health; Grant number: R01-MH61884; International Center for Genetic Engineering and Biotechnology Project; Grant number: CRP/COS98-01). All the probands were born in Costa Rica and were of putative mestizo descent with ancestors born in the Central Valley (i.e. at least two of the four grandparents were born in the Central Valley). Therefore, Costa Rican individuals self-identified as non-admixed were not included in this investigation. Trios of affected subject with the two parents or a sib were recruited in accordance with the principles of the Helsinki Declaration and with approval of the Institutional Review Boards of the University of Costa Rica and the University of Texas Health Science Center at San Antonio. All individuals signed informed consent previous to their participation. For this

analysis, a total of 426 unrelated individuals were genotyped for 730 polymorphic microsatellites drawn from Marshfield Screening Sets 16 and 54. The average missing data per locus was 5.4%.

Since the inclusion criterium for the present analysis was to be unrelated to other subjects in the study, the parents of affected subjects were selected and if these were not available, one affected or unaffected individual per family was randomly chosen. The status of being affected or unaffected was not selected for or against. In this way, of the 426 individuals, 131 had a diagnosis of schizophrenia, schizoaffective disorder, bipolar disorder, psychosis or major depression. The diagnostic process following DSMIV criteria is described elsewhere (Walss-Bass et al., 2005).

Peripheral blood was collected to create lymphoblastic cell lines. Genomic DNA was extracted from transformed lymphocytes by a standard procedure as described in Lahiri & Nurnberger (1991). Samples were collected in different regions of Costa Rica, as shown in Figure 1 and Table 1. Region delimitation was performed by geographic proximity based on ethnohistorical information (Morera et al., 2003), instead of political divisions.

Genetic variability

Allelic frequencies of all loci were estimated for the overall population and for each region separately using GDA, Genetic Data Analysis software (Lewis & Zaykin, 2001). The expected heterozygosity was computed for each locus and for each Costa Rican region using an unbiased estimator implemented in GDA. The mean heterozygosity across all loci was taken as the population and regional estimates. Each locus was tested for significant deviations from Hardy-Weinberg equilibrium using an exact test from GDA. Fst and pairwise LD between loci were additionally estimated using the same software. Since Hardy-Weinberg and pairwise LD tests showed that only 9 and 11 loci significantly deviated from equilibrium respectively, these loci were still used in further analyses, and the inclusion of these loci did not affect PCA, STRUCTURE nor ADMIXMAP results.

Population structure analysis

Population structure analysis was performed with Principal Components Analysis (PCA) and the Bayesian clustering method implemented in the STRUCTURE software package (Pritchard et al., 2000). PCA was performed with EIGENSTRAT, using each microsatellite allele as a marker and then scoring the presence or absence of the allele for each individual (Patterson et al., 2006).

To further identify possible clusters of individuals without prior assumptions of group membership, an unsupervised STRUCTURE analysis was used, varying the number of clusters k from 1 to 10. Ten replicates were run at each k. All runs assumed the F model of correlated allele frequencies across clusters, and were based on the admixture model. Runs were performed using 20000 steps burnin followed by an MCMC chain of 10000 iterations from which estimates were obtained.

Admixture analysis

Genotypic data of 662 microsatellites from 158 European individuals published by Rosenberg et al. (2005) and 2527 African individuals, comprising 121 populations, were obtained from the published database by Tishkoff et al. (2009) Similarly, data for the same microsatellites in 463 Native Americans, including the Chibchan populations Cabecar and Guaymi from Costa Rica and Panama, was obtained from Wang et al. (2007). The selection of the microsatellites was limited by the genotyped loci shared among these two datasets and the loci available in the Costa Rican mestizo sample. People from other ethnic groups, such

as Italian, Jewish, German and Chinese, known to have entered the country in more recent years, were not included in this study.

The Europeans, Africans and Natives were taken as ancestral populations and were forced into separate clusters for a STRUCTURE supervised analysis of the Costa Rican mestizo sample, to estimate individual and regional ancestry proportions. The settings for each run were the same as in the unsupervised analysis. Different values of *k* where assessed. First, *k*=3 was used when considering Europeans, Native Americans and Africans as three independent groups. Then, the Amerindian ancestry component was separated into five linguistic stocks, and into 26 separate Amerindian populations (Wang et al., 2007); therefore *k*=7 and *k*=28 were used respectively. Finally, *k* was set at 7 when the African population was separated into five different regions (Tishkoff et al. 2009), while Europeans and Amerindians were forced to separate into independent clusters. Ten replicates at each *k* were performed and the mean estimate retained. The average membership coefficients across replicates for each *k* were used for plotting with DISTRUCT (Rosenberg, 2004). Individual ancestry proportions were also estimated using ADMIXMAP (Hoggart et al., 2004) for comparison.

To evaluate if the diagnosed mental disorders were dependent upon individual admixture, a logistic regression model was fitted. The presence or absence of the disorders was taken as the binary outcome variable, with a total of 130 affected individuals and 296 unaffected individuals. For this analysis the program ADMIXMAP (Hoggart et al., 2004) was used, with the hierarchical model of individual admixture. Runs used a burn-in of 500 iterations, followed by 3000 iterations, from which parameters were estimated. Prior allele frequencies from the three main parental populations were included in the analysis. Additionally, we used a χ^2 test of heterogeneity to assess the differences between admixture proportions estimated separately for the affected and unaffected groups with STRUCTURE supervised analyses.

Results

The mean heterozygosities obtained from the analysis of the 730 autosomal microsatellite markers are shown in Table 1. The mean heterozygosity for the entire sample (0.726) was not surprising given the nature of the markers. Only subtle differences in the heterozygosity values were observed between regions, and the estimates of observed and expected heterozygosities were not significantly different.

The results of the PCA indicate no evidence of population substructure, since no clustering of the samples was observed. Similar results, of no population substructure, were obtained if the Costa Rican population is analyzed without the parental populations (Figure S1). The two most informative components are shown (Figure 2). Instead, samples from all the different regions are roughly randomly distributed in the PCA plot. On the other hand, the samples of parental populations depicted in the plot are grouped together, forming three defined clusters. The Costa Rican population sample is located mainly between the European and Amerindian clusters, but also shows some expansion toward the African cluster, supporting the admixed history of the population. The pairwise Fst values between Costa Rican regions provide additional evidence of the lack of subpopulation clustering. Very low Fst values were observed among the regions, with the highest values between northern and southern populations (Table 2). In general, the Central region showed the lowest values when compared to all the other regions. When compared to the parental populations, the Costa Rican regions showed greatest genetic distances from the African population. On the other hand, the Fst values between Costa Rican populations and the Europeans are almost three times smaller than the other comparisons, indicating a lower

genetic differentiation from Europeans, which is also supported by the ancestry proportion analysis described below. All the pairwise comparisons between the Costa Rican regions and the parental populations showed larger differences than when comparing only subpopulations within Costa Rica.

Further evaluation of population substructure using the program STRUCTURE confirmed the mentioned results; no cluster containing the entire ancestry of any individual was found among the different runs performed, indicating the admixed composition of this population.

A more detailed estimation of this admixture was obtained by the supervised analysis with the STRUCTURE program, and separately by using ADMIXMAP. Both estimates were very similar and we obtained a positive correlation between ancestry proportions estimated with both programs. The results of the STRUCTURE program are presented in Table 3. No significant differences were observed between ancestry estimates in males and females (Wilcoxon signed rank test $p>0.05$). The proportion of European ancestry is the largest among the three main ancestry contributions considered in this study, but it is lower than the estimate obtained by Morera et al. (2003) and Wang et al. (2008), using the same set of markers, for a sample of individuals from the Central Valley of Costa Rica. Conversely, the African ancestry is smaller than the two others, but it is higher than previously published (Morera et al., 2003;Wang et al., 2008), and it is not only elevated in a limited number of outlier individuals. Individual admixture proportions were also estimated, showing similar values between each other and in different regions, which results in a low variance in individual ancestry proportions.

When the ancestry proportions were estimated for each geographic region separately, some differences between regions were observed (Table 3). Wilcoxon signed rank tests showed significant differences ($p<0.001$) in the European, Amerindian and African ancestry estimates. Similar to the general population of Costa Rica, all the regions have high mean European ancestry, ranging from 0.453 to 0.556. It is noteworthy to point out the relatively high levels of African ancestry estimated in this study, with the North Pacific region showing the highest value. The inclusion of a large number of genetic markers and individuals from different regions of Africa allowed a more detailed analysis of African ancestry.

All the Costa Rican regions show similar mean Amerindian ancestry proportions, although there are small variations between regions. The Central region has the lowest mean Amerindian ancestry, while the South Pacific region shows the highest estimate. The influence of the sample characteristics, specifically the inclusion of individuals with a diagnosis of schizophrenia or a related mental disorder, over the ancestry estimates was not significant according to ADMIXMAP logistic regression results was found ($p=0.135$), and no significant differences were found in admixture proportions estimated for affected and unaffected individuals separately ($\chi^2=0.002$, $p>0.05$).

The more detailed analysis of ancestry components showed that different regions within Costa Rica have similar mean ancestry estimates with respect to the Amerindian and African ancestries. The partitioning of the Amerindian ancestry into five linguistic groups showed, as expected, that the corresponding genetic component in the admixed population of Costa Rica is most similar to the Amerindian populations that currently occupy the same geographic region (Supplemental Figure S2), specifically, to the Chibchan groups. This result is part of a general tendency observed for some urban centers in Latin America, using the same set of genetic markers (Wang et al., 2008). However, similarities to other non-Chibchan populations currently distributed throughout Latin America were also observed (Supplemental Figure S2).

Regarding the partitioning of the African genetic ancestry, the results obtained are among the first genome-wide analyses concerning the geographic origin of African immigrants in a Latin American population. For the general admixed population, Table 4 shows that the African component is most similar to Western and Central African regions, confirming ethnohistorical sources (Olien, 1980;Lohse, 2005;Midlo, 2005). Differences in African ancestry between Costa Rican regions were not significant. Smaller genetic ancestry from other regions within Africa was also observed, and could provide evidence about the probable origins of African immigrants. In comparison, the ancestry proportions of African Americans from the United States (Tishkoff et al., 2009) also show large West and Central African ancestry proportions.

Discussion

The present study aimed to assess the genetic structure and population ancestry of the Costa Rican population using a large set of highly polymorphic microsatellite markers. The lack of population substructure found was expected given that no true separation between regions, either by geographic or cultural limits, currently exists. Instead, in the last decades it has been common for individuals to move among the different regions, favoring the gene flow between them. Nevertheless, Fst values show a general tendency, in accordance with results from classical genetic markers (Morera et al., 2003), in which the northern and southern geographic regions are the most genetically different. Additionally, Fst values between the Costa Rican regions and parental populations, and the position of the admixed population in the PCA, give a broad picture of the ancestrally admixed population of Costa Rica.

The mean proportions of Amerindian, European and African ancestries for the entire sample of the Costa Rican population, and within the regions, are in general agreement with previous historical and genetic data (Meléndez, 1982; Morera et al., 2003; Lohse, 2005; Wang et al., 2008). In fact, based on historical records, this population has been considered a multiracial society (Fonseca et al., 2003), that resulted from the gradual settlement and admixture of Spanish and African individuals since the 16th century, as occurred in other populations of Latin America (Lohse, 2005; Midlo, 2005). However, some differences can be outlined. Morera et al. (2003) reported similar estimates of the accumulated Amerindian and European proportions, 61% and 30% respectively, whereas the African ancestry reported was 9%. Nevertheless, they used a different method and classical markers to estimate the admixture.

On the other hand, it is important to notice the remarkable difference in the estimate of African ancestry for the Costa Rican population, 14% in the present study, as compared to a proportion of less than 5% obtained in a similar study using STRUCTURE but with a much smaller sample comprising only individuals from the Central Valley of Costa Rica (Wang et al., 2008). These discordant results might be attributed to differences in the nature of the selected samples for analysis, such as the inclusion of patients with schizophrenia or other mental disorders in the present study. However if any effect was present, it must have been small and consequently undetected by the ADMIXMAP analysis of the relationship between individual admixture and the mental disorders considered.

The genetic similarity to other Amerindian populations, which was observed to a lesser extent, suggests that the different Amerindian groups used as parental populations are genetically related. Genetic similarity with Amerindians from more distant regions might be attributed to the migration processes that took place before and during the settlement of these groups.

Our study of the African ancestry of the population of Costa Rica is the first to show the relative similarities of the admixed population of Costa Rica to groups of different African geographic origin. Historical sources have shown the highly complex population substructure in Africa, and the process of African migrations within Africa, and to Europe and the Americas (Olien, 1980; Curtin et al., 1995; Midlo, 2005; Lohse, 2005; Klein & Vinson, 2007; Tishkoff et al., 2009; Bryc et al., 2010). The first Africans arrived in Costa Rica in the early 1500s and were most likely descendants of other Africans born in the Iberian Peninsula (Curtin et al., 1995; Lohse, 2005); others had already lived in other Spanish colonies, increasing the possibilities of having a mixed genetic ancestry (Olien, 1980; Cáceres, 2000; Lohse, 2005). Afterwards, Africans were brought to the Americas directly from West and West Central Africa (Curtin et al., 1995).

As a result of this particular movement of Africans to Costa Rica different types of matings took place and generated the unique patterns of genome admixture observed. The diverse historical processes of African settlement in specific regions of America might be largely responsible for the differences in genetic composition when comparing the Costa Rican population to other populations of Latin America (Bedoya et al., 2006; Silva-Zolezzi et al., 2009) and to African Americans from the United States (Tishkoff et al., 2009).

The largest genetic similarity to Central and West Africans which we found is in agreement with previous data. For example, United States African Americans also have large levels of ancestry from West African populations. The main difference between the African ancestry estimates in Costa Ricans and United States African Americans is that the later show a lower contribution from Eastern and Southern African populations. These differences can be attributed to historical events, and the exact places of origin in Africa are difficult to identify. With a few exceptions, there are hardly any references about slave trade from East Africa (Klein & Vinson, 2007). For these reasons, a feasible explanation about the presence of ancestry proportions from the Eastern and Southern regions of Africa in the population of Costa Rica is the existence of gene flow between African groups prior to the slave trade (Curtin et al., 1995). Further explanations for the presence of non Sub-Saharan African genetic influence could be the presence of African ancestry genes in the Spanish settlers that arrived in America (Curtin et al., 1995; Lohse, 2005); the admixture of African slaves with the Creoles and the mestizo population; and the immigration of people from West Indies, mostly from Jamaica, at the end of the 19th century (Midlo, 2005). The impact of Africans in America should not be extrapolated from one place to another, due to differences in the patterns of introduction of Africans, their gender proportions and mating patterns, the size of Amerindian populations already present in a particular region, and the cultural and social policies of specific European institutions (Midlo, 2005).

The described scenario, and the differences with other populations, can and should be taken into consideration when analyzing the impact of the genetic structure on studies of complex diseases, as in our sample, and for linkage and association investigations, since even subtle population stratification in large samples can increase considerably the false-positive rate (Pritchard & Donnelly, 2001). A balance between different demographic and historical events, which could have similar consequences, must be incorporated to establish correct hypothesis tests and to explain results generated from this kind of study. The present investigation also suggests that efforts to select a less stratified sample, through selection of subjects with ancestors born in the Central Valley (Escamilla et al., 1996), may indeed be supported, even if the current subjects in such studies are born in different regions of Costa Rica and likely have a different recent ancestral history.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported in part by the University of Costa Rica (MSW, RB, HR), Consejo Nacional de Investigaciones Científicas y Tecnológicas CONICIT (MSW), Florida Ice & Farm (RB, HR), the National Institutes of Health grant RO1-MH61884 (ME), the International Center for Genetic Engineering and Biotechnology CRP/COS98-01 (HR).

References

Barrantes R, Smouse P, Mohrenweiser H, Gershowitz H, Azofeifa J, Arias T, Neel J. Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panamá, and a consensus taxonomy based on genetic and linguistic affinity. Am J Hum Genet. 1990; 46:63–84. [PubMed: 2294755]

Bedoya G, Montoya P, García J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, Ruiz-Linares A. Admixture dynamics in Hispanics: A shift in the nuclear genetic ancestry of a South American population isolate. Proc Natl Acad Sci. 2006; 103:7234–7239. [PubMed: 16648268]

Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, Froment A, Bodo JM, Wambebe C, Tishkoff SA, Bustamante CD. Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc Natl Acad Sci USA. 2010; 107:786–791. [PubMed: 20080753]

Cáceres, R. Negros, mulatos, esclavos y libertos en la Costa Rica del siglo XVII. México: Instituto Panamericano de Geografía e Historia; 2000.

Campos-Sánchez R, Barrantes R, Silva S, Escamilla M, Ontiveros A, Nicolini H, Mendoza R, Munoz R, Raventos H. Genetic structure analysis of three Hispanic populations from Costa Rica, Mexico, and the southwestern United States using Y-chromosome STR markers and mtDNA sequences. Hum Biol. 2006; 78:551–63. [PubMed: 17506286]

Curtin, P.; Feierman, S.; Thompson, L.; Vansina, J. African History from earliest times to independence. London, UK: Longman; 1995.

Escamilla MA, Spesny M, Reus VI, Gallegos A, Meza L, Molina J, Sandkuijl LA, Fournier E, Leon PE, Smith LB, Freimer NB. Use of linkage disequilibrium approaches to map genes for bipolar disorder in the Costa Rican population. Am J Med Genet. 1996; 67:244–253. [PubMed: 8725743]

Fonseca, E.; Alvarenga, P.; Solórzano, JC. Costa Rica en el siglo XVIII. San José, Costa Rica: Editorial de la Universidad de Costa Rica; 2003.

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue P. Control of confounding of genetic associations in stratified populations. Am J Hum Genet. 2003; 72:1492–1504. [PubMed: 12817591]

Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. Design and analysis of admixture mapping studies. Am J Hum Genet. 2004; 74:965–978. [PubMed: 15088268]

Klein, H.; Vinson, B, III. African slavery in Latin America and the Caribbean. New York: Oxford University Press; 2007.

Lahiri D, Nurnberger J. A rapid non-enzimatic method for the preparation of HMW DNA from blood for RFLP studies. Nucl Acids Res. 1991; 19:5444. [PubMed: 1681511]

León P, Raventos H, Lynch E, Morrow J, King MC. The gene for an inherited form of deafness maps to chromosome 5q31. Proc Natl Acad Sci USA. 1992; 89:5181–5184. [PubMed: 1350680]

Lewis, PO.; Zaykin, D. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). 2001. Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html

Lohse, K. PhD Dissertation. University of Texas at Austin; USA: 2005. Africans and their descendants in colonial Costa Rica, 1600-1750.

Madrigal, L. Human biology of Afro-Caribbean populations. NY, USA: Cambridge University Press; 2006.

Meléndez, C. Conquistadores y pobladores. Orígenes histórico sociales de los costarricenses. San José, Costa Rica: EUNED; 1982.

Midlo, G. Slavery and African Ethnicities in the Americas: Restoring the Links. USA: The University of North Carolina Press; 2005.

Morera B, Barrantes R, Marin-Rojas R. Gene admixture in the Costa Rican population. Ann Hum Genet. 2003; 67:71–80. [PubMed: 12556237]

Olien M. Black and part-Black populations in colonial Costa Rica: Ethnohistorical resources and problems. Ethnohistory. 1980; 27:13–29.

Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]

Pritchard JK, Donnelly PJ. Case-control studies of association in structured or admixed populations. Theor Pop Biol. 2001; 60:227–237. [PubMed: 11855957]

Pritchard JK, Stephens M, Donnelly PJ. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–959. [PubMed: 10835412]

Rosenberg NA. Distruct: a program for the graphical display of population structure. Molec Ecol Notes. 2004; 4:137–138.

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW. Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genetics. 2005; 1:660–671.

Sans M. Admixture studies in Latin America: from the 20th to the 21st century. Hum Biol. 2000; 72:155–177. [PubMed: 10721616]

Silva-Zolezzi I, Hidalgo-Miranda A, Estrada-Gil J, Fernandez-Lopez JC, Uribe-Figueroa L, Contreras A, Balam-Ortiz E, Bosque-Plata L, Velazquez-Fernandez D, Lara C, Goya R, Hernandez-Lemus E, Davila C, Barrientos E, March S, Jimenez-Sanchez G. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. Proc Natl Acad Sci USA. 2009; 106:8611–8616. [PubMed: 19433783]

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. The genetic structure and history of Africans and African Americans. Science. 2009; 324:1035–1044. [PubMed: 19407144]

Walss-Bass C, Escamilla MA, Raventos H, Montero AP, Armas R, Dassori A, Contreras S, Liu W, Medina R, Balderas TG, Levinson D, Pereira R, Pereira M, Atmella I, Nesmith L, Leach R, Almasy L. Evidence of genetic overlap of schizophrenia and bipolar disorder: linkage disequilibrium analysis of chromosome 18 in the Costa Rican population. Am J Med Genet B Neuropsychiatr Genet. 2005; 139B:54–60. [PubMed: 16152570]

Wang S, Lewis C, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano F, Petzl-Erler ML, Tsuneto L, Llop E, Rothhammer F, Excoffier L, Feldman M, Rosenberg N, Ruiz-Linares A. Genetic variation and population structure in Native Americans. PloS Genet. 2007; 3:e185. [PubMed: 18039031]

Wang S, Ray N, Rojas W, Parra M, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado A, Camrena B, Nicolini N, Klitz W, Barrantes R, Molina J, Freimer N, Bortolini M, Salzano F, Petzl-Erler M, Tsuneto L, Dipierri J, Alfaro E, Bailliet G, Bianchi N, Llop E, Rothhammer F, Excoffier L, Ruiz-Linares A. Geographic patterns of genome admixture in Latin American mestizos. PLoS Genet. 2008; 4:e1000037. [PubMed: 18369456]

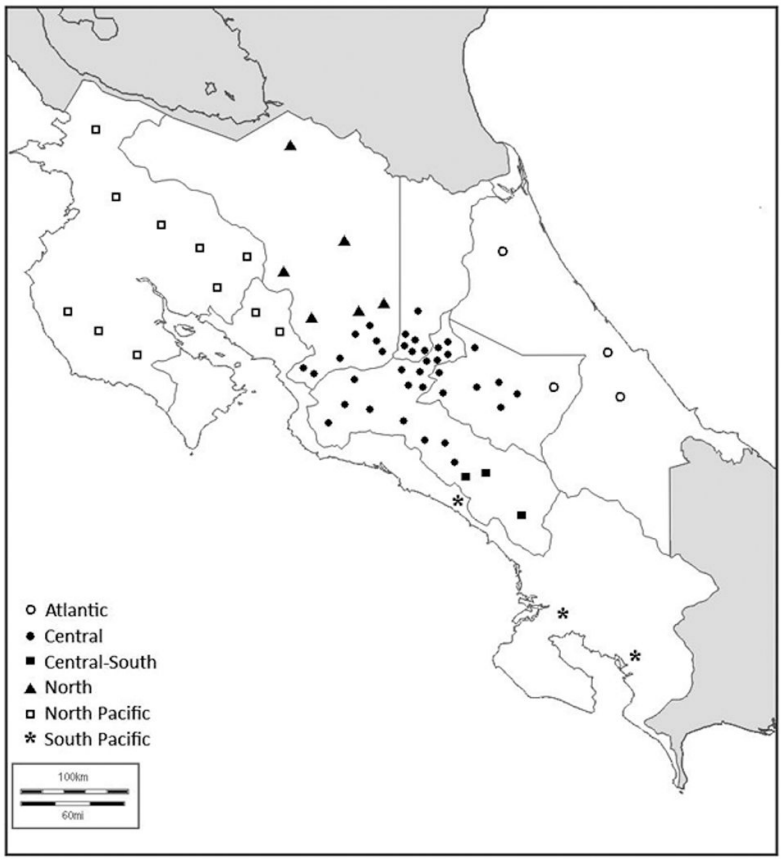

Figure 1.
Geographic location of individual places of birth assigned to specific geographic regions. Regions were defined following ethnohistorical records (Morera et al., 2003).

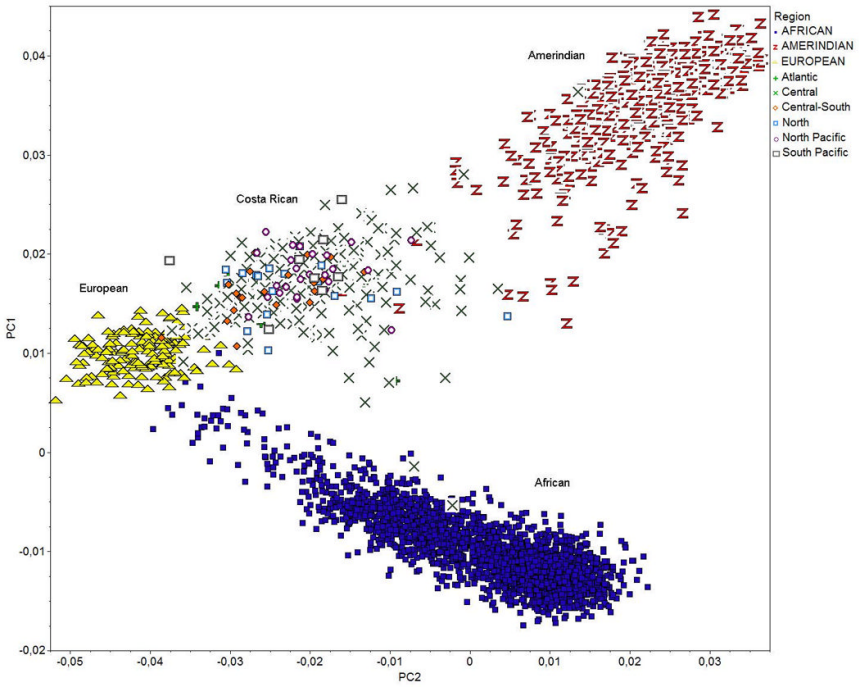

Figure 2.

Principal components analysis based on individual genotypes of the Costa Rican sample and the three parental populations. The two most informative components are shown. The Costa Rican population is divided into six regions corresponding to those shown in Figure 1.

Table 1

Sample size and mean observed and expected heterozygosities for each region of Costa Rica included in this investigation. Standard deviations are shown in parenthesis.

Region	N	Ho	He
Atlantic	20	0.724 (0.129)	0.711 (0.084)
Central	322	0.727 (0.079)	0.732 (0.075)
Central-South	21	0.729 (0.122)	0.712 (0.083)
North	18	0.725 (0.125)	0.712 (0.086)
North Pacific	25	0.729 (0.119)	0.724 (0.078)
South Pacific	8	0.725 (0.172)	0.686 (0.099)
Total	414	0.726 (0.127)	0.713 (0.086)

Table 2

F_{ST} (x100) values between regions within Costa Rica, and between these and Amerindian, European and African parental populations. Genotypic information includes 426 individuals of Costa Rica from the present investigation, and of parental populations obtained from Rosenberg et al. (2002), Wang et al. (2007) and Tishkoff et al. (2008).

	AFRICAN	EUROPEAN	AMERINDIAN	Atlantic	Central	Central-South	North	North Pacific
EUROPEAN	7.71							
AMERINDIAN	13.95	11.96						
Atlantic	7.69	2.26	7.94					
Central	7.88	2.05	7.11	0.13				
Central-South	7.70	2.17	7.98	0.24	0.05			
North	7.10	1.82	7.57	0.12	0.10	0.01		
North Pacific	6.54	2.63	7.05	0.01	0.09	0.19	0.34	
South Pacific	6.12	2.42	7.21	0.08	0.03	0.20	0.43	0.28

Table 3

Mean ancestry proportions and standard deviations, estimated using STRUCTURE program, for each region of the Costa Rican population and for the whole population.

Region	EUROPEAN	AMERINDIAN	AFRICAN
Atlantic	0.519 (0.070)	0.355 (0.076)	0.126 (0.037)
Central	0.552 (0.087)	0.318 (0.083)	0.131 (0.036)
Central-South	0.556 (0.069)	0.319 (0.071)	0.125 (0.028)
North	0.536 (0.099)	0.329 (0.076)	0.135 (0.046)
North Pacific	0.463 (0.156)	0.340 (0.093)	0.197 (0.078)
South Pacific	0.453 (0.144)	0.374 (0.085)	0.172 (0.101)
Total	0.541 (0.097)	0.322 (0.084)	0.137 (0.055)

Table 4

STRUCTURE results of relative mean ancestry proportions and standard deviations for each region of the Costa Rican population in relation to the African (AF) groups divided into five geographic areas. The European and Amerindian ancestry components are not shown. Data from African populations were obtained from a published database (Tishkoff et al., 2009).

Region	CENTRAL AF	EASTERN AF	SAHARAN AF	SOUTHERN AF	WESTERN AF
Atlantic	0.204 (0.016)	0.180 (0.008)	0.103 (0.008)	0.131 (0.006)	0.247 (0.016)
Central	0.224 (0.016)	0.172 (0.010)	0.110 (0.012)	0.118 (0.006)	0.272 (0.016)
Central-South	0.294 (0.013)	0.162 (0.007)	0.076 (0.006)	0.172 (0.005)	0.174 (0.012)
North	0.227 (0.025)	0.164 (0.009)	0.098 (0.007)	0.136 (0.005)	0.234 (0.019)
North Pacific	0.261 (0.020)	0.191 (0.024)	0.099 (0.012)	0.150 (0.004)	0.461 (0.047)
South Pacific	0.233 (0.019)	0.174 (0.016)	0.085 (0.008)	0.220 (0.012)	0.288 (0.062)
Total	0.223 (0.017)	0.175 (0.013)	0.108 (0.011)	0.121 (0.006)	0.299 (0.034)