

9-10-2013

Ancestry Informative Markers Clarify The Regional Admixture Variation In The Costa Rican Population

Rebeca Campos-Sánchez

Universidad de Costa Rica, San José, Costa Rica, rxc327@psu.edu

Henriette Raventós

Universidad de Costa Rica, San José, Costa Rica

Ramiro Barrantes

Universidad de Costa Rica, San José, Costa Rica, ramrio.barrantes@ucr.ac.cr

Recommended Citation

Campos-Sánchez, Rebeca; Raventós, Henriette; and Barrantes, Ramiro, "Ancestry Informative Markers Clarify The Regional Admixture Variation In The Costa Rican Population" (2013). *Human Biology Open Access Pre-Prints*. Paper 34.
http://digitalcommons.wayne.edu/humbiol_preprints/34

This Open Access Preprint is brought to you for free and open access by the WSU Press at DigitalCommons@WayneState. It has been accepted for inclusion in Human Biology Open Access Pre-Prints by an authorized administrator of DigitalCommons@WayneState. For more information, please contact digitalcommons@wayne.edu.

TITLE PAGE

TITLE OF THE PAPER:

ANCESTRY INFORMATIVE MARKERS CLARIFY THE REGIONAL
ADMIXTURE VARIATION IN THE COSTA RICAN POPULATION

AUTHOR'S NAMES:

REBECA CAMPOS-SÁNCHEZ^{a,1}, HENRIETTE RAVENTÓS^{a,b}, RAMIRO
BARRANTES^{b*}

^aUniversidad de Costa Rica, Centro de Investigación en Biología Celular y
Molecular, Universidad de Costa Rica, 2060 San José, Costa Rica.

^bUniversidad de Costa Rica, Escuela de Biología, Universidad de Costa Rica, 2060
San José, Costa Rica.

*Corresponding author post-publication: Tel: +506 25115622; Fax: +506 22247530.
E-mail: ramiro.barrantes@ucr.ac.cr

¹ Present address: The Pennsylvania State University, University Park, PA 16802.

Correspondence during review and publication: Rebeca Campos-Sánchez. Address:
306 Wartik Building, Penn State University. City: University Park, PA. Zip code: 16802.
Phone: 814 7530450. Email: rebcamsan@gmail.com, rxc327@psu.edu.

KEY WORDS:

historic perspective; population genetic structure; stratification; forensic databases;
genetic

studies; Costa Rica

ABSTRACT

The genetic structure of Costa Rica's population is complex, both by region and by individual, due to the admixture process that started during the 15th century and historical events thereafter. Previous studies have been done mostly on Amerindian populations and the Central Valley inhabitants using various microsatellites and mtDNA markers. Here, we study for the first time a random sample from all regions of the country with AIMS (Ancestry Informative Markers) to address the individual and regional admixture proportions. A sample of 160 male individuals was screened for 78 AIMS customized in a GoldenGate platform from Illumina. We observed that this small set of AIMS has the same power of hundreds of microsatellites and thousands of SNPs to evaluate admixture, with the benefit of reducing genotyping costs. This type of investigation is necessary to explore new genetic markers useful for forensic and genetic investigation. Our data showed a mean admixture proportion of 49.2% European, 37.8% Native American and 12.9% African, with a disproportionate admixture composition by region. In addition, when a fourth component, the Chinese, was included the proportions changed to 45.6% European, 33.5% Native American, 11.7% African, and 9.2% Chinese. The admixture trend is consistent among all regions (EUR>NAM>AFR) and individual admixture estimates vary broadly in each region. Though we did not find stratification in CRP, it is recommended to evaluate gene admixture in future genetic studies of Costa Rica, especially for the Caribbean region as it contains the largest proportion of African ancestry (30.9%).

Costa Rica's population (CRP) conformation is complex, starting by the admixture of natives (Barrantes *et al.* 1990) with Spaniards during the colonization period (15th and 16th centuries, (Meléndez 1982; Meléndez 1985), and then with African immigrants that entered the country as slaves on the 16th century and from the Caribbean countries during the 19th century (Bryce-Laporte 1962; Casey 1979; Chomsky 1995; Duncan 1972; Meléndez 1972; Stewart 1967). In addition, the Chinese immigration began in 1850 and has increased throughout the years (Bermúdez 2000; León 1987), nonetheless no scientific research has been done on its impact on the actual population until now. Furthermore, understanding admixture in this population is essential for disease susceptibility mapping studies, and Costa Rica's population has been extensively studied for psychiatric diseases (Contreras *et al.* 2010; Escamilla *et al.* 2009; Escamilla *et al.* 2007; Walss-Bass *et al.* 2006; Walss-Bass *et al.* 2009), longevity (Catri *et al.* 2011; Castri *et al.* 2009) and other disease studies (Leon *et al.* 1992).

Population admixture is best studied with genetic markers that show allele frequency differences between ancestral groups that originated in the population under analysis (Rosenberg *et al.* 2003). The most frequently used are the Ancestry Informative Markers (AIMs), single nucleotide polymorphisms (SNPs) that show large allele frequency differences (Galanter *et al.* 2012) and which have been studied on modern descendants of ancestral populations (Parra *et al.* 1998). AIMS can also be used to infer the geographic origin of an individual (Galanter *et al.* 2012). Previous studies have shown that these markers are useful in studies of Hispanics (Bonilla *et al.* 2004a), Mexicans (Martinez-Fierro *et al.* 2009; Martinez-Marignac *et al.* 2007; Tian *et al.* 2007), African Americans

(Parra *et al.* 1998), Native Americans (Klimentidis *et al.* 2009), and Puerto Ricans (Lai *et al.* 2009), among others. Another advantage of using SNPs over other markers (i.e. microsatellites or mtDNA sequences) is their low mutation rate (Budowle and van Daal 2008), which makes them ideal to study old population events as they reconstruct more accurate genotypes.

Previous studies in Costa Rica have addressed the admixture question, directly or indirectly, through population genetics (Morera *et al.* 2003) and forensic studies (Morales *et al.* 2001). All of these investigations possess a sampling bias towards a geographic region (mostly the Central Valley) or disease phenotype. Nonetheless, a diverse set of markers have been analysed such as blood groups and proteins (Morera *et al.* 2003), autosomal microsatellites (Segura-Wang *et al.* 2010; Wang *et al.* 2008), AIMs (Ruiz-Narvaez *et al.* 2010), and sex-specific markers in the mitochondrial and the Y-chromosome (Campos-Sanchez *et al.* 2006).

Here, we studied AIMs in a random sample (no disease associated) from the whole country to evaluate the genetic ancestry conformation of Costa Rica and its ethno-geographic regions. The estimated proportions of admixture from European, West Africans, Native Americans, and Chinese populations revealed different ancestral population proportions depending on the individuals' region of origin. In addition, we did not detect population stratification, and the individual admixture estimates vary broadly among the samples. The AIMs studied here could be used to address sample selection on future genetic studies, to understand historical records and for forensic applications (i.e. improving genetic population databases, identification of individual's origin). Moreover, we observed the power that a smaller set of AIMs has over hundreds of microsatellite

markers and thousands of SNPs to study admixture, which is beneficial as it reduces the costs of genotyping. We even suggest that AIMs should be included on forensic databases of Costa Rica and could plausibly be extended to Central America.

MATERIALS AND METHODS

Samples and ethno-geographic subdivision

Samples of 160 unrelated male individuals from the entire territory were randomly selected and classified into four regions using ethno-historic and geographic-political criteria (Morera *et al.* 2003). The four regions are: North (37 samples), Caribbean (21 samples), Central Valley (77 samples) and South (25 samples) (Figure 1). Sample sizes are proportional to the total population by region, that is the Central Valley is the largest settlement, North and South are intermediate and the Caribbean is the least inhabited. DNA was extracted with the phenol-chloroform method. This study was approved by the Ethics Committee of the University of Costa Rica.

Genotyping of AIMs

Each DNA was quantified and 500 ng were used for the Golden Gate Assay (Illumina Inc; San Diego, CA, USA). A total of 82 Ancestry Informative Markers (AIMs, Suppl Table 1) were customized for the assay and the allele assignment was done by BeadStudio 3.0 software. These markers present high allele frequency differences in European (EUR), Native American (NAM), African (AFR), and Chinese (CHB) descendants. Moreover, these AIMs have been used on admixture mapping studies and on individual admixture estimations of other Latin-American populations with ancestral

populations similar to that of Costa Rica (Bonilla *et al.* 2004b; Martinez-Marignac *et al.* 2007; Price *et al.* 2007; Tian *et al.* 2007).

Statistical analysis

Each marker was tested for Hardy Weinberg departures using the De Finetti software (<http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl>). Genetic distances were determined using GenAlEx (Peakall and Smouse 2006) and depicted on trees and MDS plots with MEGA version 3.1 (Kumar *et al.* 2004) and GenAlEx (Peakall and Smouse 2006), respectively.

ADMIXMAP 3.8 (Hoggart *et al.* 2004) for Windows was used to estimate individual and population admixture proportions. We also used this program to test for stratification using a test for residual allelic association between unlinked loci (Martinez-Marignac *et al.* 2007). We used the default parameters except for the following: “samples” of 5000 (iterations of the Markov chain), “burnin” of 200, “populations” of 3 or 4 depending on the model tested. Genotype data from West Africans, European Spaniards, Mesoamerican Amerindians, and Han Chinese from Beijing were used as parental populations. These parental allele frequencies were obtained from three sources: reference publications (Martinez-Marignac *et al.* 2007; Tian *et al.* 2007), Hapmap data extracted from the dbSNP at the NCBI webpage, and the 1000 Genomes project downloaded from SPSmart (Amigo *et al.* 2009; Amigo *et al.* 2008). The triangular plots were generated with the package klaR (Weihs *et al.* 2005) on R (Team 2011).

RESULTS

Markers selection and evaluation

The AIMs were selected from previous publications on Latin-American populations with ancestries similar to that of Costa Rica, because of their potential to study admixture (Bonilla *et al.* 2004b; Martinez-Marignac *et al.* 2007; Price *et al.* 2007; Tian *et al.* 2007). In our sample we obtained highly reliable genotyping profiles with the Golden Gate assay. Nonetheless, three out of 82 markers failed genotyping (rs1327805, rs1935946, rs983271) and rs983271 was monoallelic. For the rest of the markers, the analysis of Hardy Weinberg Equilibrium showed that four markers have significant deviances in the CRP after Bonferroni correction ($p < 0.0006$, data not shown). When we reanalyzed the data after deleting those markers, we obtained similar results; therefore we used all 78 markers in subsequent analysis (Suppl Table 1).

Admixture analysis with three ancestral populations

Our analysis of gene ancestry for Costa Rica with three ancestral populations revealed that 49.2% of the ancestry is European, 37.8% Native American and the remaining 12.9% African (Table 1). These results are fairly consistent with previous studies where the EUR component is predominant and the AFR component is the least abundant (Morera *et al.* 2003; Segura-Wang *et al.* 2010).

The proportions of admixture by region (Table 1) revealed that the European component is the largest in all regions. In the Central Valley it is 56.9%, followed by the South Region with 50.2%, the North with 44.1% and Caribbean with 40.1%. The second most important component is the Native American, where the North and South regions presented the largest proportions of 40.8% and 41.2%, respectively. The Central Valley showed an intermediate proportion of 36.4% and the smallest among the four regions studied was the Caribbean with 29%. In contrast, the Caribbean revealed the largest

African component (30.9%), as expected from their known ethno-historical distribution. Additionally, the second region with the largest proportion of African descent is the North Region (15.1%), also explained by the entrance of African slaves through the Pacific for the railroad construction (Bermúdez 2000; León 1987), and migration of slaves from the Central Valley to this region (Meléndez 1972; Meléndez and Duncan 1974). Lastly, the Central Valley and South Regions presented the smallest African component (6.7% and 8.5%, respectively).

Admixture analysis with four ancestral populations

When we evaluated a model with four ancestral populations, it showed 45.6% EUR, 33.5% NAM, 11.7% AFR, and 9.2% CHB components (Table 1). The trend per region and ancestry component was consistent with the three ancestral population model of admixture, with a slight reduction in the proportions for EUR, NAM and AFR. The Chinese component was highest on the North (6.6%) and South regions (6.1%), smallest on the Caribbean (4.2%) and intermediate in the Central Valley region (4.9%).

Individual admixture estimations

As expected, the variation in individual gene admixture is large, even for individuals belonging to the same region of the country. For the EUR component it ranged from 10.4% to 82.7%, the NAM from 7.3% to 72.4%, the AFR from 1.5% to 80.7%, and the Chinese from 1.3% to 17.4% (Figure 2). Although the variance among the proportions is large, there is a predominant overrepresentation of smaller African ancestry in most samples (82.5% individuals with <15%), as well as the Chinese (100% with <17%). In contrast, the European component (half of the sample with 50-60% ancestry) and Native American component (85% of the sample with 30-50% ancestry) showed intermediate

proportions as shown in Suppl Figure 1. As outliers, we found seven individuals in the Caribbean sample that have more than 65% African ancestry, and from the Central Valley Region we found three individuals that are more than 65% Native American and seven that are more than 70% European (data not shown). A triangular plot (Figure 2) illustrates the distribution of each sample according to its ancestry component and shows the clustering of most samples between NAM and EUR, independently if a three ancestral population (Figure 2) or four ancestral population estimation was used (Suppl Figure 2).

Stratification analysis

An important estimation based on our data is the plausible stratification of the population by region. Our results, based on a test for residual allelic association between 20 unlinked loci (Martinez-Marignac *et al.* 2007), showed no significant probability of subdivision in CRP. Though this result must be carefully evaluated, it could be a reflection of balanced migration of the population among the different regions, therefore reducing the effects of gene drift. This is consistent with previous studies that found no stratification in their samples from CRP (Ruiz-Narvaez *et al.* 2010; Segura-Wang *et al.* 2010; Wang *et al.* 2008).

Comparisons to worldwide populations

We estimated F_{st} distances among our four regions of study and other admixed and ancestral populations to evaluate their relationship. Clearly, the phylogenetic tree depiction (Suppl Figure 4) showed the Caribbean region as the most distant and closer to the Yoruban branch (YRI, African population). In addition, the most closely related regions were the Central Valley and the South. We observed that the F_{st} distances were

small among CRN, CRCV and CRS regions, and these three were more distant to CRC (data not shown), so we hypothesized that the separation of the regions could be due to the different proportions of African alleles in the sample. Therefore, we did a PCA analysis that revealed that the first and second components explained almost 48.64% of the variation. A clear separation was also observed between seven samples of the Caribbean and the rest of the samples (Suppl Fig. 4). Furthermore, a correlation of 94% ($p < 0.001$, Suppl Fig. 5) between the first component and the African ancestry estimated for all CRP samples confirmed our hypothesis.

The genetic distance analysis, depicted on the phylogenetic tree (Suppl Fig. 4), also revealed that the admixed populations included in the analysis were positioned closely to their most predominant ancestral population, confirming the efficiency of these 78 AIMS to identify admixture and ancestry.

We placed additional attention to the comparison of CRP to the other three admixed populations (Mexico, Colombia and Puerto Rico) from the 1000 Genome project (Amigo *et al.* 2008) and to their ancestral populations. Based on the PCA, the first component revealed a cluster of all the admixed individuals closer to the EUR and CHB ancestral populations, and a separated cluster represented by the AFR ancestral population and those individuals from the CRC with high African components (Suppl Figure 6). This plot is consistent with the estimations of individual admixture proportions for Costa Rica and with historical and genetic data for the admixed populations from Latin America, which are the result of a predominant Spanish and Native American blend. In addition, we generated a PCA plot for populations (data not shown) that revealed a cluster for the YRI and a separate cluster for the rest of the populations.

DISCUSSION

As is known from historical records, Costa Rica was built mainly from the admixture of three ancestral populations starting in the 15th century (Acuña 2009; Barrantes *et al.* 1990; Obando 2004; Russell Lohse 2005). In addition, the Chinese component was integrated for the first time into the population during the mid-19th century, primarily as labour force for the Pacific railroad construction and spreading afterwards to the Caribbean and other regions since then (Bermúdez 2000; León 1987). The interaction and movements of the people following these events resulted in the complex regional ancestry conformation confirmed by our results.

European and Native American ancestry predominates in CRP

As expected, the European component is the highest and is evenly distributed throughout all regions. A similar distribution was revealed for the Native American ancestry component. Our results confirm the process of admixture among the Spanish and the original residents of Costa Rica, a process that started during the Colonial times and continues to this day. It is known from ethno-historical records that the population of Native Americans was dramatically decimated during the Colonial period as in many other Latin American countries (Barrantes 1993; Crosby 1986). Nonetheless, the offspring from Spanish men and Native American women carried the genetic diversity from the Amerindian population (Crosby 1986; Meléndez 1982) that we now detect.

African ancestry component varies among regions

We observed that the African component is approximately 11% for most of the country. But the singularities revealed by region can be understood from an historic

perspective. Most of the African descendants established in the Caribbean region and were isolated for centuries because of racism (Madrigal *et al.* 2001). This situation explains the scarce migration to the rest of the country (Madrigal *et al.* 2001; Russell Lohse 2005) and the low proportion of African alleles, especially in the Central Valley. The other region with a significant representation of African descendants is the North (the Guanacaste province). In this case, the first African slaves entered through the Pacific coast and established there, among other places (Klein and Vinson III 2007). This is supported by our result of over 25% of African descent alleles in some individuals from the North Region, a reflection of historical admixture and migration into this region.

Chinese ancestry component is widely spread in the CRP

Including the Chinese ancestral population in the analysis showed the importance of this component in the formation of the Costa Rican population. Although it is small (up to 9%) compared to the other three ancestral populations, it should be considered an important component for forensic applications as it is widely spread, especially in the North and South Regions (Bermúdez 2000; León 1987).

Individual admixture estimations

From our analysis we observed large differences between individual admixtures in each region sampled, something never reported before in such detail. The random sample evaluated here allowed us to clarify the composition of regions, and individuals within the regions. Therefore, this new perspective of individual admixture diversity should be considered when studying susceptibility genes, as there may be individuals in a sample with huge divergent genetic backgrounds confounding or diffusing important signatures of association. This also reflects the complexity of CRP from a historical perspective, as

the belief was on EUR and NAM admixture predominantly, but now we show that the AFR component is significant in some individuals not only from the Caribbean. Moreover the CHB ancestry is reflected on every region and considerably on few individuals in our study.

Regional analysis reveals important differences

It is evident, by the ADMIXMAP analysis, that common patterns of admixture (but not uniform) took place for the Central Valley, the North and South regions. Also the genetic distances depicted by phylogenetic trees and PCA plots showed the closeness of these regions, even when compared with other admixed populations. The phylogenetic tree shows a stronger interaction between the Central Valley and the South Region. Moreover, historical records document the intense migration among them (Perez Briglioni 2010) confirming our estimations. It is also remarkable that the history of isolation of the African descendants from the Caribbean Region (Madrigal *et al.* 2001) is also proved by our genetic analysis which shows that 33% of the Caribbean sample has more than 62% African genes. This is also evident by the low African component in most individuals from other regions. The power of this study at the regional level resides in the random sampling of the whole country. Therefore, we could refine ancestry estimations of the underexplored regions (i.e. North, South and Caribbean) and revealed the differences among them (Table 1).

Absence of stratification

The lack of stratification observed in CRP might seem contradictory to the diverse regional and individual proportions of admixture reported here. This could be the result of the small group of markers used for the estimation (20 unlinked AIMs out of 78).

Nevertheless, the similar proportions of EUR and NAM ancestry throughout Costa Rica could also be responsible for the absence of stratification. The impact of this result on genetic studies is important as it implies that most regions of Costa Rica have similar Spanish-Native American proportions. Nonetheless, this conclusion should not exclude the need for a stratification analysis on all genetic drug or disease susceptibility mapping efforts done on this population (Morera and Barrantes 2004), as the African and Chinese ancestries are highly present in random individuals throughout the country. Additionally, our results support that the “genetic isolate hypothesis” of the Central Valley is wrong. It was believed that few founder European individuals and their descendants populated the region; therefore a significant homogeneity was expected to exist (Freimer *et al.* 1996; Morera *et al.* 2003; Segura-Wang *et al.* 2010). Nevertheless, we observed a large proportion of Native American component in the sample (Morera and Barrantes 2004) proving the diversity of the region.

Comparison to other studies in CRP

Previous population genetic studies of the CRP have addressed the admixture question (Madrigal *et al.* 2001; Morera *et al.* 2003; Ruiz-Narvaez *et al.* 2010; Segura-Wang *et al.* 2010; Wang *et al.* 2008; Wang *et al.* 2010). Our results are consistent with the major trends, where the EUR component is predominant throughout the country, followed by the NAM component and lastly AFR (EUR>NAM>AFR). Nevertheless, the regional analysis differs partially because of: different sample sizes, sampling bias towards a disease phenotype, markers used, regional subdivision of the sample, different ancestral population datasets, and even the program used for the analysis.

The first study published on the Caribbean population (375 samples, 4 autosomal markers) subdivided the sample in two groups by self-ethnic identification (Madrigal *et al.* 2001). One group identified as Afro-Caribbean possessed high AFR components (75.95%) and equally shared EUR and NAM (10.47% and 13.57%, respectively). The other group identified as Hispanic-Caribbean revealed a larger EUR ancestry (58.66%), intermediate NAM (33.8%) and smaller AFR (7.51%). Our random sample resembles more the Hispanic-Caribbean sample from Madrigal and collaborators (2001), but with an increased AFR component mostly due to few samples with individual AFR admixture >65%.

The North region was first studied by Wang *et al.* (2010). A sample of 1301 women was selected for an HPV-related (Human Papilloma Virus) study and genotyped with 27904 SNPs, which could result in a bias of admixture estimates. However, our results are consistent with theirs and a 1% difference is observed for the EUR component (43%) and NAM (38%) in the North region. In addition, they observed a 4% residual Asian ancestry, which was higher in our study (7%). This is the only time that Asian ancestry was studied.

Two additional studies (Morera *et al.* 2003; Segura-Wang *et al.* 2010) analyzed different samples from the entire CRP and, similar to them, we subdivided the population into regions. Morera *et al.* (2003) used a random sample and analyzed 11 classic genetic markers on 2196 individuals. They estimated a 61% EUR ancestry, 30% NAM and 9% AFR, which implies an overestimation of EUR and underestimation of NAM and AFR compared to our results. Moreover, their analysis by region reflects that the Caribbean

sample has a larger EUR and smaller AFR component, comparable to the Hispanic-Caribbean from Madrigal *et al.* (2001).

Segura-Wang and collaborators (2010) studied a large sample of individuals (426) with 730 microsatellites. Their sample selection (families with mental disorders) could have resulted in a biased admixture estimate and indeed up to 5.6% differences were observed in the mean admixture estimations (54.1% EUR, 32.2% NAM and 13.7% AFR ancestry) compared to our random sample. The regional analysis also shows large contrasts with our analysis; where the major differences are reflected on the Caribbean (also known as Atlantic) sample overestimating the EUR (51.9%) and NAM (35.5%) component and underestimating the AFR (12.6%).

From the preceding summary, it is evident that the Caribbean region shows the most variability on admixture estimates due to its historical complexity reflected at the gene level. Even the simple subdivision of the samples by regions did result in significant differences that could be important for sample selection in disease studies, specially the subdivision of the North Region into Pacific (or Chorotega) and North.

Future directions

Here we show that 78 AIMs are enough to obtain appropriate ancestry resolution in CRP, which can be used on other disease related samples. An important difficulty we found is the lack of genotyping data on NAM populations relevant to CRP conformation. Recently a new set of 446 AIMs was developed to study admixture in Latin American individuals, proven useful in populations with similar history to Costa Rica, as Colombia (Galanter *et al.* 2012). Although this set of AIMs represents a more comprehensive genome-wide selection and the corresponding ancestral populations are finely genotyped,

it must be evaluated whether smaller subsets can be used for admixture estimations to reduce genotyping costs. Furthermore, extending this analysis to other Central American countries could result in a more comprehensive understanding of the evolution and interrelationships of these populations, with significant applications on forensic databases and disease susceptibility research.

Although no broad admixture proportion differences among regions were found in our study, except for the Caribbean, the high individual differences should be considered when performing forensic identification and determining origin of a sample. These markers are possibly useful to identify African descendants in the Caribbean region, but further studies are necessary for descendants of Native Americans from the Chibchan groups. The genetic population conformation and individual differences described for the first time in this study, reveals the need to construct a genetic database with random samples localized by geographical regions as an ideal reference for forensic investigation. Another potential use would be for the identification of the population's origin for forensic evidence, in which mitochondrial DNA or Y-chromosome markers cannot discriminate, if the DNA is highly degraded or the sample material is extremely small (i.e. 11-M Madrid bomb attack investigation (Phillips *et al.* 2009)). AIMs have also been shown to be very promising for admixture mapping when disease predisposition is linked to ancestry (Tian *et al.* 2007). This could also be useful for genetic mapping of complex disorders in Costa Rica with an adequately selected panel of markers.

ACKNOWLEDGEMENTS

We would like to thank: Florida Ice and Farm for the financial support, the University of Costa Rica (Grant 111-A6-047) and the University of Texas Health Science Centre at San Antonio (Robin Leach's lab) for the use of their facilities to analyze the samples.

LITERATURE CITED

- Acuña M. 2009. Mestizajes en la provincia de Costa Rica 1690-1821. San José: Universidad de Costa Rica.
- Amigo J, Phillips C, Salas A *et al.* 2009. Viability of in-house datamarting approaches for population genetics analysis of SNP genotypes. *BMC Bioinformatics* 10 Suppl 3:S5.
- Amigo J, Salas A, Phillips C *et al.* 2008. SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access. *BMC Bioinformatics* 9:428.
- Barrantes R. 1993. Diversidad genética y mezcla racial en los amerindios de Costa Rica y Panamá. *Rev Biol Trop* 41:379-384.
- Barrantes R, Smouse PE, Mohnenweiser HW *et al.* 1990. Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am J Hum Genet* 46(1):63-84.
- Bermúdez Q. 2000. El contexto internacional de la inmigración china en Costa Rica (1850-1980) [Maestría Profesional de Historia]. San José, Costa Rica: Universidad de Costa Rica.
- Bonilla C, Parra EJ, Pfaff CL *et al.* 2004a. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann Hum Genet* 68:139-153.
- Bonilla C, Shriver MD, Parra EJ *et al.* 2004b. Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York city. *Hum Genet* 115(1):57-68.
- Bryce-Laporte R. 1962. Social relations and cultural persistence (change) among Jamaicans in a rural area of Costa Rica [Doctoral]: University of Puerto Rico.
- Budowle B, and van Daal A. 2008. Forensically relevant SNP classes. *Biotechniques* 44(5):603-610.
- Campos-Sanchez R, Barrantes R, Silva S *et al.* 2006. 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* 78(5):551-563.
- Casey J. 1979. Limón: 1880-1940. San José, Costa Rica: Editorial Costa Rica.
- Castri L, Madrigal L, Melendez-Obando M *et al.* 2011. Mitochondrial polymorphisms associated with differential longevity do not impact lifetime-reproductive success. *Am J Hum Biol* 23(2):225-227.
- Castri L, Melendez-Obando M, Villegas-Palma R *et al.* 2009. Mitochondrial polymorphisms are associated both with increased and decreased longevity. *Hum Hered* 67(3):147-153.
- Chomsky A. 1995. Afro-Jamaican traditions and labor force organizing on United Fruit Company plantations in Costa Rica, 1910. *J Soc Hist* 28:837-855.

- Contreras J, Hernandez S, Quezada P *et al.* 2010. Association of serotonin transporter promoter gene polymorphism (5-HTTLPR) with depression in Costa Rican schizophrenic patients. *J Neurogenet* 24(2):83-89.
- Crosby AW. 1986. *Ecological imperialism: the biological expansion of Europe, 900-1900*: Cambridge University Press.
- Duncan Q. 1972. *El negro Antillano: inmigración y presencia*. San José, Costa Rica: Editorial Costa Rica.
- Escamilla M, Hare E, Dassori AM *et al.* 2009. A schizophrenia gene locus on chromosome 17q21 in a new set of families of Mexican and central american ancestry: evidence from the NIMH Genetics of schizophrenia in latino populations study. *Am J Psychiatry* 166(4):442-449.
- Escamilla MA, Ontiveros A, Nicolini H *et al.* 2007. A genome-wide scan for schizophrenia and psychosis susceptibility loci in families of Mexican and Central American ancestry. *Am J Med Genet B Neuropsychiatr Genet* 144B(2):193-199.
- Freimer NB, Reus VI, Escamilla M *et al.* 1996. An approach to investigating linkage for bipolar disorder using large Costa Rican pedigrees. *Am J Med Genet* 67(3):254-263.
- Galanter JM, Fernandez-Lopez JC, Gignoux CR *et al.* 2012. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet* 8(3):e1002554.
- Hoggart CJ, Shriver MD, Kittles RA *et al.* 2004. Design and analysis of admixture mapping studies. *Am J Hum Genet* 74(5):965-978.
- Klein H, and Vinson III B. 2007. *African slavery in Latin America and the Caribbean*. New York: Oxford University Press.
- Klimentidis YC, Miller GF, and Shriver MD. 2009. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *Am J Phys Anthropol* 138(4):375-383.
- Kumar S, Tamura K, and Nei M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5(2):150-163.
- Lai CQ, Tucker KL, Choudhry S *et al.* 2009. Population admixture associated with disease prevalence in the Boston Puerto Rican health study. *Hum Genet* 125(2):199-209.
- León MG. 1987. *Chinese migration on the Atlantic of Costa Rica. The economic adaptation of an Asian minority in a pluralistic society [Doctoral]*. New Orleans, Louisiana: Tulane University.
- Leon PE, Raventos H, Lynch E *et al.* 1992. The gene for an inherited form of deafness maps to chromosome 5q31. *Proc Natl Acad Sci U S A* 89(11):5181-5184.
- Madrigal L, Ware B, Miller R *et al.* 2001. Ethnicity, gene flow, and population subdivision in Limon, Costa Rica. *Am J Phys Anthropol* 114(2):99-108.
- Martinez-Fierro ML, Beuten J, Leach RJ *et al.* 2009. Ancestry informative markers and admixture proportions in northeastern Mexico. *J Hum Genet* 54(9):504-509.
- Martinez-Marignac VL, Valladares A, Cameron E *et al.* 2007. Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. *Hum Genet* 120(6):807-819.

- Meléndez C. 1972. Aspectos sobre la inmigración Jamaicana. San José, Costa Rica: Editorial Costa Rica.
- Meléndez C. 1982. Conquistadores y pobladores: orígenes histórico-sociales de los costarricenses. San José, Costa Rica: EUNED.
- Meléndez C. 1985. Bosquejo para una historia social costarricense antes de la independencia. San José, Costa Rica: Editorial Costa Rica.
- Meléndez C, and Duncan Q. 1974. El negro en Costa Rica: antología. San José, Costa Rica: Editorial Costa Rica.
- Morales AI, Morera B, Jimenez-Arce G *et al.* 2001. Allele frequencies of markers LDLR, GYPA, D7S8, HBG, GC, HLA-DQA1 and D1S80 in the general and minority populations of Costa Rica. *Forensic SciInt* 124(1):1-4.
- Morera B, and Barrantes R. 2004. Is the Central Valley of Costa Rica a genetic isolate? *Revista De Biología Tropical* 52(3):629-644.
- Morera B, Barrantes R, and Marin-Rojas R. 2003. Gene admixture in the Costa Rican population. *Ann Hum Genet* 67(Pt 1):71-80.
- Obando MOM. 2004. The importance of genealogy applied to genetic research in Costa Rica. *Revista De Biología Tropical* 52(3):423-450.
- Parra EJ, Marcini A, Akey J *et al.* 1998. Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63(6):1839-1851.
- Peakall R, and Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6(1):288-295.
- Perez Briglioni H. 2010. La población de Costa Rica, 1750-2000: Editorial UCR.
- Phillips C, Prieto L, Fondevila M *et al.* 2009. Ancestry Analysis in the 11-M Madrid Bomb Attack Investigation. *PLoS One* 4(8).
- Price AL, Patterson N, Yu F *et al.* 2007. A genomewide admixture map for Latino populations. *Am J Hum Genet* 80(6):1024-1036.
- Rosenberg NA, Li LM, Ward R *et al.* 2003. Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* 73(6):1402-1422.
- Ruiz-Narvaez EA, Bare L, Arellano A *et al.* 2010. West African and Amerindian ancestry and risk of myocardial infarction and metabolic syndrome in the Central Valley population of Costa Rica. *Hum Genet* 127(6):629-638.
- Russell Lohse K. 2005. Africans and their descendants in Colonial Costa Rica, 1600-1750. Austin, TX: University of Texas at Austin. 653 p.
- Segura-Wang M, Raventos H, Escamilla M *et al.* 2010. Assessment of Genetic Ancestry and Population Substructure in Costa Rica by Analysis of Individuals with a Familial History of Mental Disorder. *Ann Hum Genet* 74:516-524.
- Stewart W. 1967. Keith y Costa Rica. San José, Costa Rica: Editorial Costa Rica.
- Team RDC. 2011. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Tian C, Hinds DA, Shigeta R *et al.* 2007. A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. *Am J Hum Genet* 80(6):1014-1023.
- Walss-Bass C, Montero AP, Armas R *et al.* 2006. Linkage disequilibrium analyses in the Costa Rican population suggests discrete gene loci for schizophrenia at 8p23.1 and 8q13.3. *Psychiatr Genet* 16(4):159-168.

- Walss-Bass C, Soto-Bernardini MC, Johnson-Pais T *et al.* 2009. Methionine sulfoxide reductase: a novel schizophrenia candidate gene. *Am J Med Genet B Neuropsychiatr Genet* 150B(2):219-225.
- Wang S, Ray N, Rojas W *et al.* 2008. Geographic patterns of genome admixture in Latin American mestizos. *Plos Genetics* 4(3).
- Wang Z, Hildesheim A, Wang SS *et al.* 2010. Genetic admixture and population substructure in Guanacaste Costa Rica. *PLoS One* 5(10):e13336.
- Weihs C, Ligges U, Luebke K *et al.* 2005. klaR Analyzing German Business Cycles. In: Baier D, Decker R, and Schmidt-Thieme L, editors. *Data Analysis and Decision Support*. Berlin: Springer-Verlag. p 335-343.

Table 1. Mean admixture proportions on four regions and total for Costa Rica, estimated by 78 AIMs, using two separate models with three and four ancestral populations. (EUR: European, NAM: Native American, AFR: African, CHB: Chinese ancestry)

Region (number of samples)	EUR	NAM	AFR	CHB
Model with 3 ancestral pop:				
North Region (37)	0.441	0.408	0.151	
Caribbean Region (21)	0.401	0.290	0.309	
Central Valley (77)	0.569	0.364	0.067	
South Region (25)	0.502	0.412	0.085	
Total Costa Rica	0.492	0.378	0.129	
Model with 4 ancestral pop:				
North Region (37)	0.422	0.371	0.141	0.066
Caribbean Region (21)	0.389	0.265	0.305	0.042
Central Valley (77)	0.553	0.335	0.063	0.049
South Region (25)	0.485	0.377	0.077	0.061
Total Costa Rica	0.456	0.335	0.117	0.092



Figure 1. Map of Costa Rica and the four regions of study (North, Central Valley, Caribbean, and South), including the number of samples in parenthesis

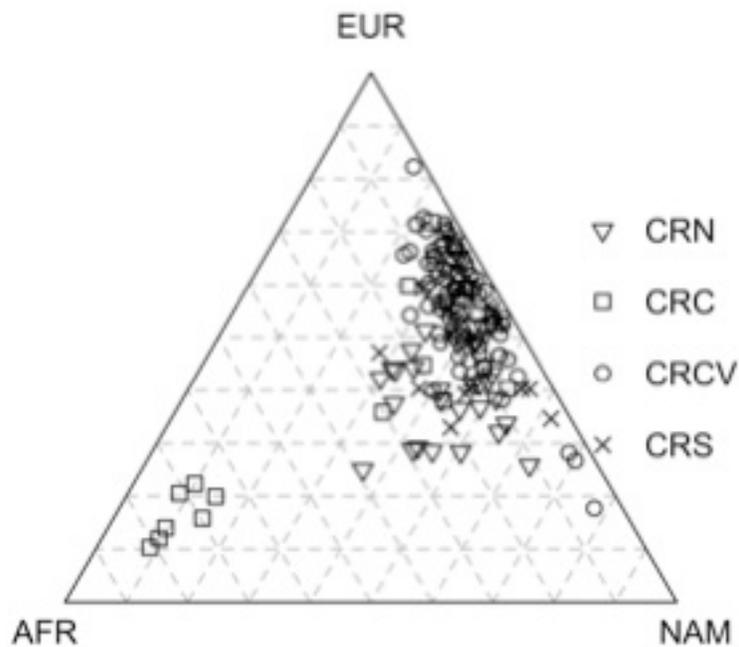
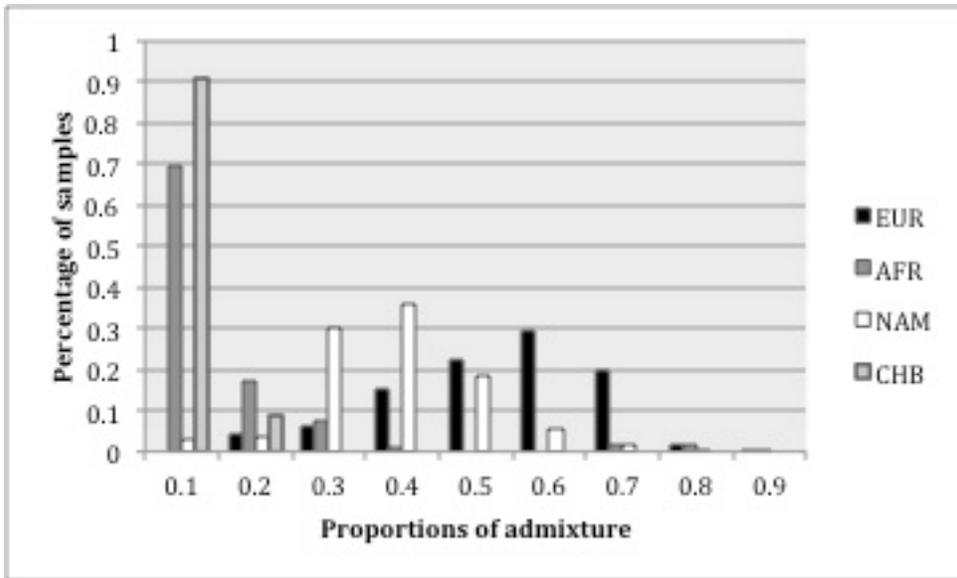
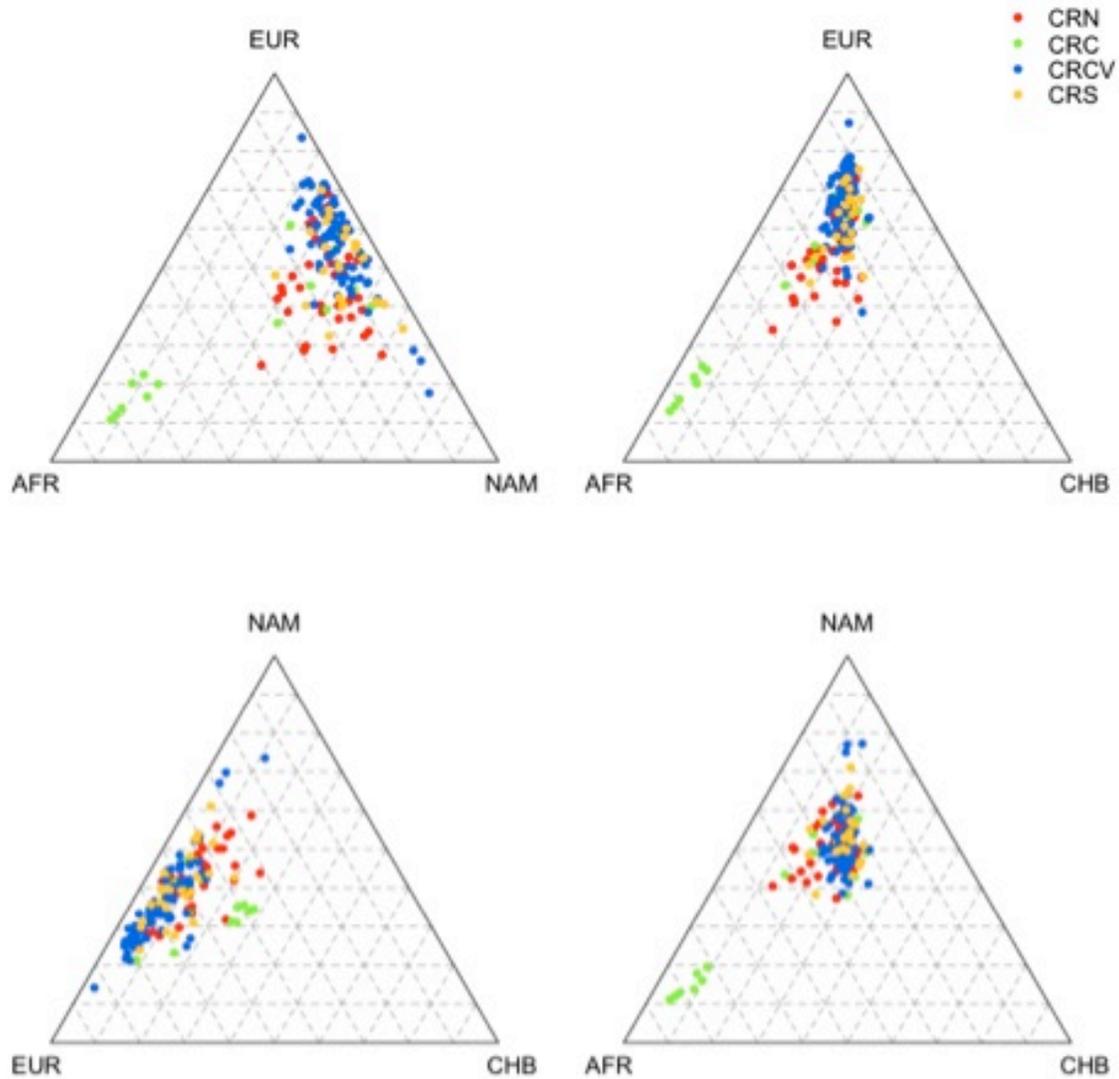


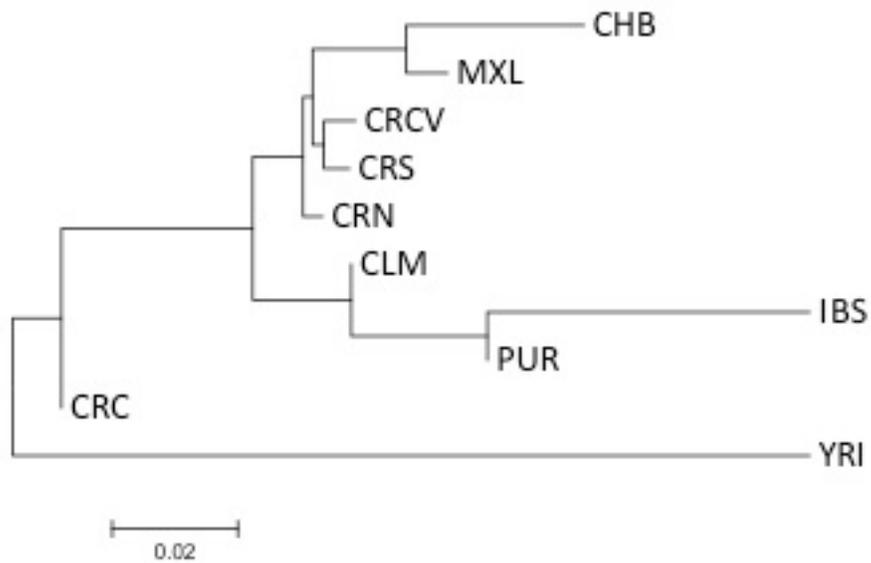
Figure 2. Triangular plot showing the individual admixture proportions estimated with 78 AIMS in 160 random samples from Costa Rica. (EUR: European, AFR: African, NAM: Native American, CRN: Costa Rica North, CRC: Costa Rica Caribbean, CRCV: Costa Rica Central Valley, CRS: Costa Rica South)



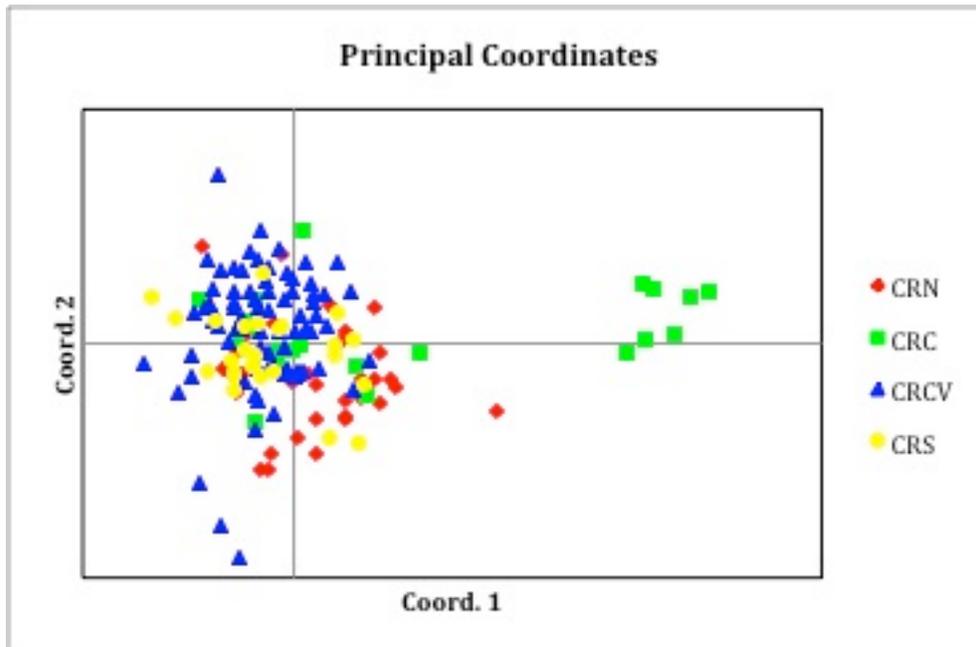
Suppl Figure 1. Proportions of admixture on the sample from Costa Rica based on 78
AIMs. (EUR: European, AFR: African, NAM: Native American, CHB: Chinese ancestry)



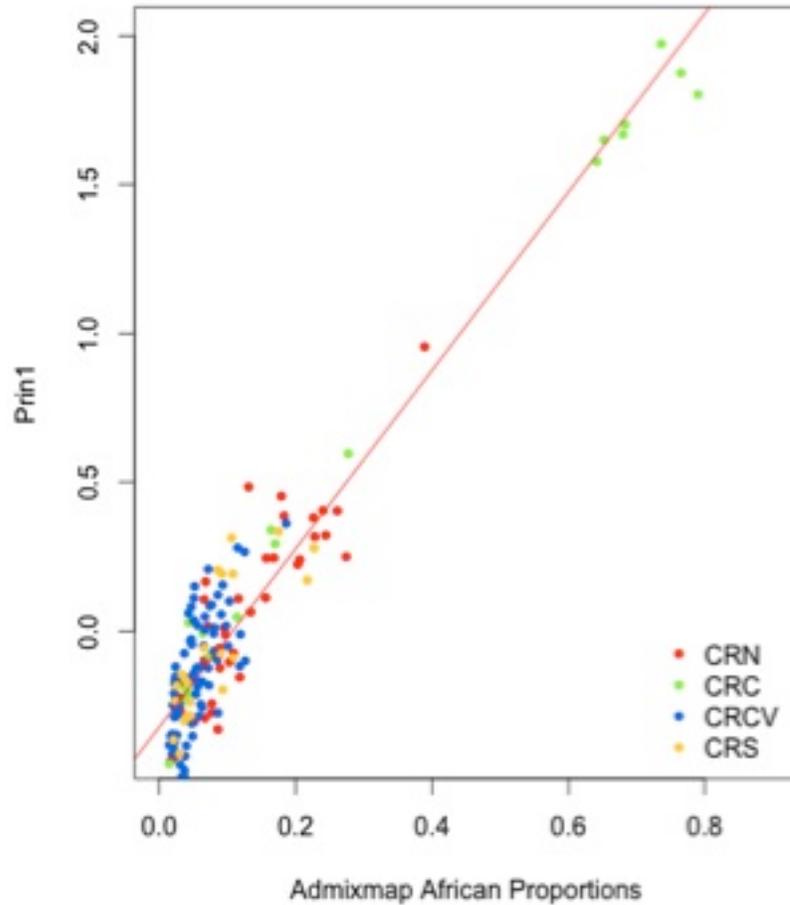
Suppl Figure 2. Individual admixture as three way comparisons estimated using four ancestral populations for Costa Rica studied with 78 AIMS. (CRN: Costa Rica North, CRC: Costa Rica Caribbean, CRCV: Costa Rica Central Valley, CRS: Costa Rica South region)



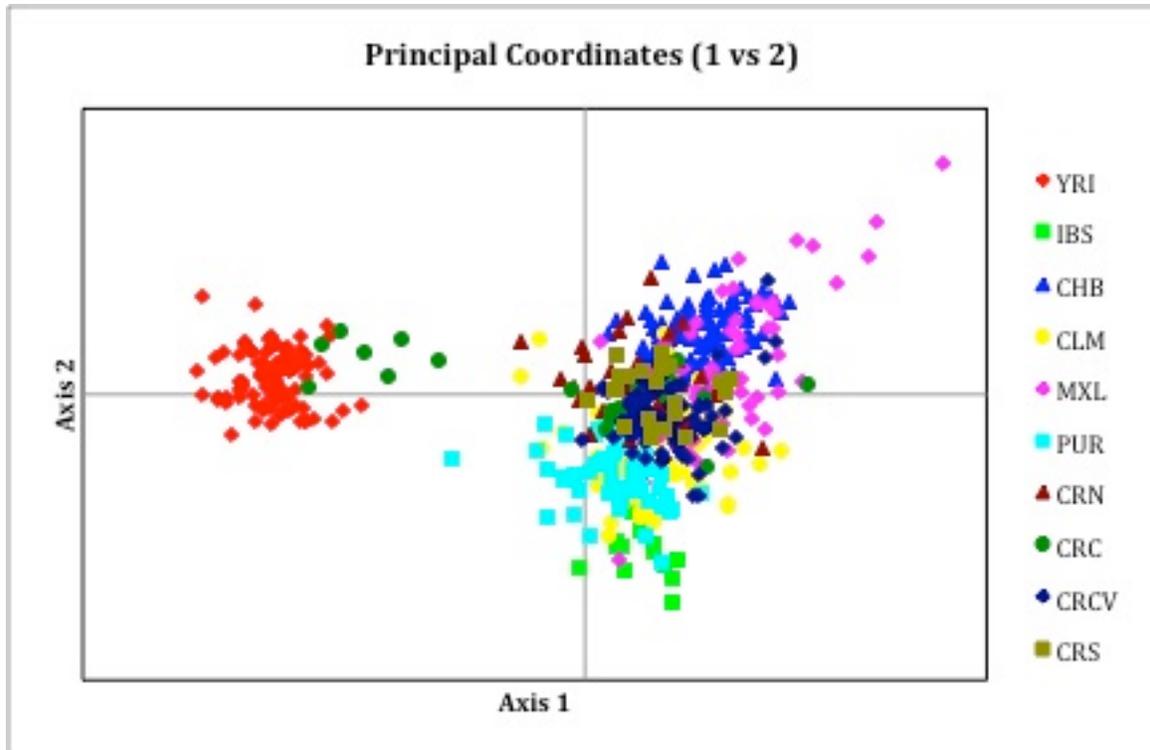
Suppl Figure 3. Phylogenetic tree based on F_{st} distances and the Neighbour joining algorithm on four regions of Costa Rica in comparison with other admixed and ancestral populations, analyzed with 78 AIMs. (YRI Yoruba, IBS Iberia, CHB China, CLM Colombia, MXL Mexico, PUR Puerto Rico, CRN Costa Rica North Region, CRC Costa Rica Caribbean Region, CRCV Costa Rica Central Valley, CRS Costa Rica South Region)



Suppl Figure 4. PCA of 160 samples studied with 78 AIMS color-coded by region of sampling. (CRN: Costa Rica North, CRC: Costa Rica Caribbean, CRCV: Costa Rica Central Valley, CRS: Costa Rica South region)



Suppl Figure 5. Correlation between the first component value and the African ancestry proportion obtained by ADMIXMAP (Pearson correlation coefficient = 0.94; $p < 0.001$). (CRN: Costa Rica North, CRC: Costa Rica Caribbean, CRCV: Costa Rica Central Valley, CRS: Costa Rica South region)



Suppl Figure 6. PCA of individual genetic distances (F_{st}) between 10 populations genotyped for 78 AIMs. (YRI Yoruba, IBS Iberia, CHB China, CLM Colombia, MXL Mexico, PUR Puerto Rico, CRN Costa Rica North Region, CRC Costa Rica Caribbean Region, CRCV Costa Rica Central Valley, CRS Costa Rica South Region)

Suppl Table 1. Ancestral population allele frequencies for 78 AIMs studied in a Costa Rican sample

rs number	Chromosome	European	Native American	African	Chinese ⁺	Reference
2752	1	0.563	0.278	0.153	0.593	Martinez-Marignac <i>et al.</i> 2007
6003	1	0.104	0.019	0.698	0.031	Martinez-Marignac <i>et al.</i> 2007
723822	1	0.083	0.864	0.219	0.474	Martinez-Marignac <i>et al.</i> 2007
725667	1	0.12	0.001	0.708	0.000	Martinez-Marignac <i>et al.</i> 2007
963170	1	0.146	0.922	0.001	0.495	Martinez-Marignac <i>et al.</i> 2007
1008984	1	0.881	0.276	0.34	0.577	Martinez-Marignac <i>et al.</i> 2007
1506069	1	0.028	0.368	0.927	0.273	Martinez-Marignac <i>et al.</i> 2007
2065160	1	0.079	0.838	0.504	0.773	Martinez-Marignac <i>et al.</i> 2007
2225251	1	0.348	0.808	0.955	0.536	Martinez-Marignac <i>et al.</i> 2007
2479409	1	0.66	0.03	0.775 *	0.304	Tian <i>et al.</i> 2007
2814778	1	0.993	0.999	0.002	1.000	Martinez-Marignac <i>et al.</i> 2007
4908736	1	0.86	0.19	0.683 *	0.330	Tian <i>et al.</i> 2007
3287	2	0.786	0.868	0.285	0.918	Martinez-Marignac <i>et al.</i> 2007
1435090	2	0.24	0.847	0.203	0.588	Martinez-Marignac <i>et al.</i> 2007
1861498	2	0.792	0.991	0.116	0.928	Martinez-Marignac <i>et al.</i> 2007
6730157	2	0.63	0	0.000 *	0.000	Tian <i>et al.</i> 2007
7595509	2	0.143	NA	0.011	0.191	Amigo <i>et al.</i> 2008
768324	3	0.043	0.76	0.205	0.273	Martinez-Marignac <i>et al.</i> 2007
1344870	3	0.967	0.06	0.941	0.716	Martinez-Marignac <i>et al.</i> 2007
1465648	3	0.783	0.9	0.109	0.763	Martinez-Marignac <i>et al.</i> 2007
2317212	3	0.922	0.146	0.295	0.531	Martinez-Marignac <i>et al.</i> 2007
2613964	3	0.321	NA	0.369	0.309	Amigo <i>et al.</i> 2008
719776	4	0.88	0.852	0.052	0.655	Martinez-Marignac <i>et al.</i> 2007
951784	4	0.171	0.702	0.074	0.247	Martinez-Marignac <i>et al.</i> 2007
1112828	4	0.829	0.113	0.94	0.387	Martinez-Marignac <i>et al.</i> 2007
1403454	4	0.139	0.885	0.026	0.531	Martinez-Marignac <i>et al.</i> 2007
2702414	4	0.09	0.79	0.083 *	0.356	Tian <i>et al.</i> 2007
3309	5	0.3	0.711	0.4	0.763	Martinez-Marignac <i>et al.</i> 2007
3340	5	0.797	0.216	0.92	0.701	Martinez-Marignac <i>et al.</i> 2007
26247	5	0.8	0.16	0.367 *	0.340	Tian <i>et al.</i> 2007
1461227	5	0.111	0.825	0.409	0.881	Martinez-Marignac <i>et al.</i> 2007
1881826	6	0.885	0.2	0.794	0.876	Martinez-Marignac <i>et al.</i> 2007
1935946	6	0.536	NA	0.159	0.072	Amigo <i>et al.</i> 2008
2001144	6	0.92	0.2	0.306 ⁺	0.567	Tian <i>et al.</i> 2007
1320892	7	0.74	0.105	0.904	0.294	Martinez-Marignac <i>et al.</i> 2007
1469179	7	0.357	NA	0.324	0.954	Amigo <i>et al.</i> 2008
2341823	7	0.833	0.575	0.126	0.722	Martinez-Marignac <i>et al.</i> 2007
2396676	7	0.779	0.575	0.129	0.670	Martinez-Marignac <i>et al.</i> 2007

285	8	0.494	0.439	0.965	0.289	Martinez-Marignac <i>et al.</i> 2007
1373302	8	0.287	0.921	0.351	0.562	Martinez-Marignac <i>et al.</i> 2007
1808089	8	0.417	0.966	0.397	0.392	Martinez-Marignac <i>et al.</i> 2007
4130405	8	0.85	0.29	1.000 *	0.546	Tian <i>et al.</i> 2007
2695	9	0.167	0.964	0.271	0.443	Martinez-Marignac <i>et al.</i> 2007
1327805	9	0.071	NA	0.670	0.598	Amigo <i>et al.</i> 2008
1928415	9	0.817	0.999	0.25	0.995	Martinez-Marignac <i>et al.</i> 2007
1980888	9	0.933	0.052	0.801	0.515	Martinez-Marignac <i>et al.</i> 2007
2149589	9	0.43	0.93	0.358 *	0.593	Tian <i>et al.</i> 2007
2888998	9	0.87	0.52	0.717 *	0.294	Tian <i>et al.</i> 2007
563654	10	0.07	0.6	0.400 *	0.191	Tian <i>et al.</i> 2007
1594335	10	0.7	0.76	0.206	0.778	Martinez-Marignac <i>et al.</i> 2007
1891760	10	0.379	0.94	0.203	0.572	Martinez-Marignac <i>et al.</i> 2007
2207782	10	0.347	0.948	0.905	0.758	Martinez-Marignac <i>et al.</i> 2007
1042602	11	0.485	0.027	0.004	0.000	Martinez-Marignac <i>et al.</i> 2007
1079598	11	0.929	NA	0.909	0.546	Amigo <i>et al.</i> 2008
1487214	11	0.064	0.161	0.842	0.216	Martinez-Marignac <i>et al.</i> 2007
1800498	11	0.63	0.077	0.135	0.041	Martinez-Marignac <i>et al.</i> 2007
5443	12	0.681	0.741	0.199	0.546	Martinez-Marignac <i>et al.</i> 2007
726391	12	0.778	0.5	0.056	0.474	Martinez-Marignac <i>et al.</i> 2007
4767387	12	0.03	0.72	0.400 *	0.392	Tian <i>et al.</i> 2007
717091	13	0.191	0.319	0.779	0.304	Martinez-Marignac <i>et al.</i> 2007
2078588	13	0.925	0.875	0.023	0.722	Martinez-Marignac <i>et al.</i> 2007
4885346	13	0.071	NA	0.222	0.438	Amigo <i>et al.</i> 2008
2862	15	0.142	0.704	0.382	0.490	Martinez-Marignac <i>et al.</i> 2007
4646	15	0.287	0.739	0.321	0.289	Martinez-Marignac <i>et al.</i> 2007
724729	15	0.895	0.999	0.115	0.866	Martinez-Marignac <i>et al.</i> 2007
1800404	15	0.636	0.491	0.133	0.407	Martinez-Marignac <i>et al.</i> 2007
292932	16	0.001	0.727	0.001	0.263	Martinez-Marignac <i>et al.</i> 2007
764679	16	0.056	0.625	0.187	0.490	Martinez-Marignac <i>et al.</i> 2007
2816	17	0.517	0.061	0.001	0.067	Martinez-Marignac <i>et al.</i> 2007
1014263	17	0.9	0.22	1.000 ⁺	0.603	Tian <i>et al.</i> 2007
1074075	17	0.266	0.008	0.868	0.330	Martinez-Marignac <i>et al.</i> 2007
1369290	18	0.071	0.001	0.9	0.000	Martinez-Marignac <i>et al.</i> 2007
1464612	18	0.94	0.29	0.966 ⁺	0.644	Tian <i>et al.</i> 2007
718092	20	0.153	0.824	0.76	0.629	Martinez-Marignac <i>et al.</i> 2007
1877751	20	0.67	0	0.900 *	0.144	Tian <i>et al.</i> 2007
718387	21	0.906	0.893	0.174	0.665	Martinez-Marignac <i>et al.</i> 2007
2829556	21	0.09	0.73	0.125 *	0.608	Tian <i>et al.</i> 2007
461915	22	0.143	NA	0.335	0.634	Amigo <i>et al.</i> 2008

* Obtained from the Hapmap Yoruba sample at the dbSNP website (NCBI)

⁺ Obtained from 1000 Genome Project (Amigo *et al.* 2008)

NA not available