

# Decrypting the Mitochondrial Gene Pool of Modern Panamanians

Ugo A. Perego<sup>1,9</sup>, Hovirag Lancioni<sup>2,9</sup>, Maribel Tribaldos<sup>3,9</sup>, Norman Angerhofer<sup>1,4</sup>, Jayne E. Ekins<sup>1</sup>, Anna Olivieri<sup>5</sup>, Scott R. Woodward<sup>1,4</sup>, Juan Miguel Pascale<sup>3</sup>, Richard Cooke<sup>6</sup>, Jorge Motta<sup>3</sup>, Alessandro Achilli<sup>2\*</sup>

**1** Sorenson Molecular Genealogy Foundation, Salt Lake City, Utah, United States of America, **2** Dipartimento di Biologia Cellulare e Ambientale, Università di Perugia, Perugia, Italy, **3** Instituto Conmemorativo Gorgas de Estudios de la Salud, Panama City, Panama, **4** Ancestry, Provo, Utah, United States of America, **5** Dipartimento di Biologia e Biotecnologie, Università di Pavia, Pavia, Italy, **6** Smithsonian Tropical Research Institute, Apdo, Ancón, Panama

## Abstract

The Isthmus of Panama—the narrow neck of land connecting the northern and southern American landmasses—was an obligatory corridor for the Paleo-Indians as they moved into South America. Archaeological evidence suggests an unbroken link between modern natives and their Paleo-Indian ancestors in some areas of Panama, even if the surviving indigenous groups account for only 12.3% of the total population. To evaluate if modern Panamanians have retained a larger fraction of the native pre-Columbian gene pool in their maternally-inherited mitochondrial genome, DNA samples and historical records were collected from more than 1500 volunteer participants living in the nine provinces and four indigenous territories of the Republic. Due to recent gene-flow, we detected ~14% African mitochondrial lineages, confirming the demographic impact of the Atlantic slave trade and subsequent African immigration into Panama from Caribbean islands, and a small European (~2%) component, indicating only a minor influence of colonialism on the maternal side. The majority (~83%) of Panamanian mtDNAs clustered into native pan-American lineages, mostly represented by haplogroup A2 (51%). These findings reveal an overwhelming native maternal legacy in today's Panama, which is in contrast with the overall concept of personal identity shared by many Panamanians. Moreover, the A2 sub-clades A2ad and A2af (with the previously named 6 bp Huetar deletion), when analyzed at the maximum level of resolution (26 entire mitochondrial genomes), confirm the major role of the Pacific coastal path in the peopling of North, Central and South America, and testify to the antiquity of native mitochondrial genomes in Panama.

**Citation:** Perego UA, Lancioni H, Tribaldos M, Angerhofer N, Ekins JE, et al. (2012) Decrypting the Mitochondrial Gene Pool of Modern Panamanians. *PLoS ONE* 7(6): e38337. doi:10.1371/journal.pone.0038337

**Editor:** David Caramelli, University of Florence, Italy

**Received:** March 22, 2012; **Accepted:** May 3, 2012; **Published:** June 4, 2012

**Copyright:** © 2012 Perego et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research received support from the Instituto Conmemorativo Gorgas de Estudios de la Salud (to JM), Gorgas Internal Grant for Genealogy Studies (to JMP), the Italian Ministry of the University: FIRB-Futuro in Ricerca 2008 (to AA and AO), Progetti Ricerca Interesse Nazionale 2009 (to AA), and the Sorenson Molecular Genealogy Foundation (to UAP, NA and SRW). The authors are grateful to all the donors for providing biological specimens, and to everyone at the Instituto Conmemorativo Gorgas and at the Sorenson Molecular Genealogy Foundation for their work on the preliminary data. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** NA and SRW are employed by a commercial company: Ancestry.com. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

\* E-mail: alessandro.achilli@unipg.it

**9** These authors contributed equally to this work.

## Introduction

Most genetic studies that focus on the population dynamics of the first human groups that moved from North to South America across the Central American isthmus were based on data collected exclusively from surviving indigenous Native American groups. However, after the dramatic encounters with occupying Europeans starting about 500 years ago, the cultural, demographic, ethnic, and genetic landscapes of the Western Hemisphere were changed irreversibly. Today's Native American populations are a non-random remnant of the multitude of culturally and socially diverse groups, which developed over the ~15 to ~20 millennia that have elapsed since the first human groups moved from north-east Asia into America across Beringia [1–19]. Reconstructing the history of any people using modern-day populations is often challenging since current populations likely do not represent the full extent of variation that existed in earlier populations that may

have changed vastly in composition in intervening years [20,21]. This is also true for the uniquely-positioned, narrow geographic region of Panama, a pivotal cross-road corridor that connects America's northern and southern landmasses [22].

The extant Panamanian ethnic groups comprise 12.3% (417,559 of 3,405,813) of the Panamanian population [23], exhibit remarkable cultural resilience, and speak languages, which are historically related at different time depths [24]. The majority *ethnia* are the Ngäbe (also known as Ngöbe), Kuna (also called Guna) and Emberá; smaller *ethnia* are the Wounaan, Bribri, and Naso (also called Teribe). The speakers of Ngäbere are by far the most numerous group (>260,000 individuals). However, it is unknown whether pre-Columbian indigenous inhabitants represented the same cultural population as the modern ones. Archaeological records, as well as vegetation history derived from lake sediment studies in the country of Panama, show that, after the initial arrival towards the end of the last glaciation, some

descendants of the earliest migrants remained on the isthmus, at least since Clovis times (13.2–12.8 Ky ago) [25], adapting their lifestyles to changing environmental and social conditions [24,26]. The demographic scenario of Panama, as in every country in America, changed dramatically with the arrival of European settlers and their African slaves. Conquest impacted local populations differentially. Survival was best in mountainous and Caribbean regions where the Spanish colonizers had little interest in settling or did not have the resources to do so. In some areas (e.g. the central and western Caribbean and much of the Darién) they were rebuffed by the Native people, e.g. the Ngäbe and Kuna who descend from populations, which have lived continuously in some areas of Panama for a very long time.

At Spanish contact, the landscape was occupied by hundreds of sedentary farming communities arranged into small chiefdoms. In central Panama, where archaeological and paleoecological records are most complete, it is likely that the inhabitants of polities encountered by the Spanish were largely derived from much earlier populations residing in the same area [24]. As for the historical-linguistic data, a small vocabulary of ~50 words of the Cueva language, which the Spanish say was spoken from central Panama to the Gulf of Urabá, is recorded in contact-period documents. It comprises some words that are cognate with modern Kuna and others with modern Waunaan [27]. This suggests that the Cuevan “language” may have been a *lingua franca* used for trade or that communities were multilingual [28,29]. It also implies a degree of linguistic continuity in this area across Spanish contact. In western Panama, less than 10 native words were recorded by 16<sup>th</sup> century chroniclers. Therefore it is impossible to determine whether these populations spoke recently extinct languages (Dorasque and Chánguena) or ancestral forms of the languages that are spoken today (Ngäbére, Buglére, and Naso). However, documentary evidence from after the 17<sup>th</sup> century makes it very likely that forms of these languages were spoken in pre-Columbian times.

From a genetic point of view, the uniparental markers, Y chromosome and mitochondrial DNA (mtDNA), have been widely employed in the past leading to the identification of a high degree of differentiation in specific regions and/or ethnic groups, and highlighting the existence of genetic structures even in geographically proximate populations [30–33]. An example of this differentiation is the “Huetar deletion,” a peculiar control-region 6-bp deletion between nucleotide pairs (np) 106 and 111, which was identified by Santos and Barrantes [34] in a sample of Huetar (Costa Rica) whose now extinct language belongs to the Chibchan stock (*sensu* Constenla Umaña [29]). This corresponds to the *MspI* site loss at np 104 within haplogroup A2 reported in several other Chibchan-speaking groups of Central America, including the Boruca, Bribri, Cabécar, Guaymí (Ngäbe and Bugle), Kuna, and Teribe [8,35–38]. It was concluded that this mutation might have originated several millennia ago, when it was likely that lower Central American populations spoke ancient variants of Chibchan languages [29], and formed small and mobile social units. Only later did they congregate into tribes or chiefdoms, which, in spite of frequent inter- and intra-group conflict, continually traded with each other. In the most productive areas, especially on the Pacific side, a considerable degree of sedentism, and moderate to strong social ranking were achieved [24].

In brief, the all-important 16<sup>th</sup> century is poorly understood. There is no evidence that the ancestral indigenous gene-pool of Panama (and lower Central America) was completely replaced. Some historical, linguistic and archaeological data point clearly towards continuity, at least in some areas of the Isthmus, deep in time and extending up to the Conquest era. If this is the case, the

populations of modern Panama should have retained at least a fraction of the native pre-Columbian gene-pool, possibly to a variable extent, given the differential degree of geographical and genetic isolation of the different Panamanian communities during the past five centuries. In order to address this issue, we present a summary of the mtDNA variation from a large Panamanian sample obtained from the mixed general population living in Panama’s urban areas as well as from autochthonous tribal groups. The analysis indicates that the study of mtDNA contributes to the understanding of evolutionary dynamics of the Native American population of this geographically-unique region.

## Results

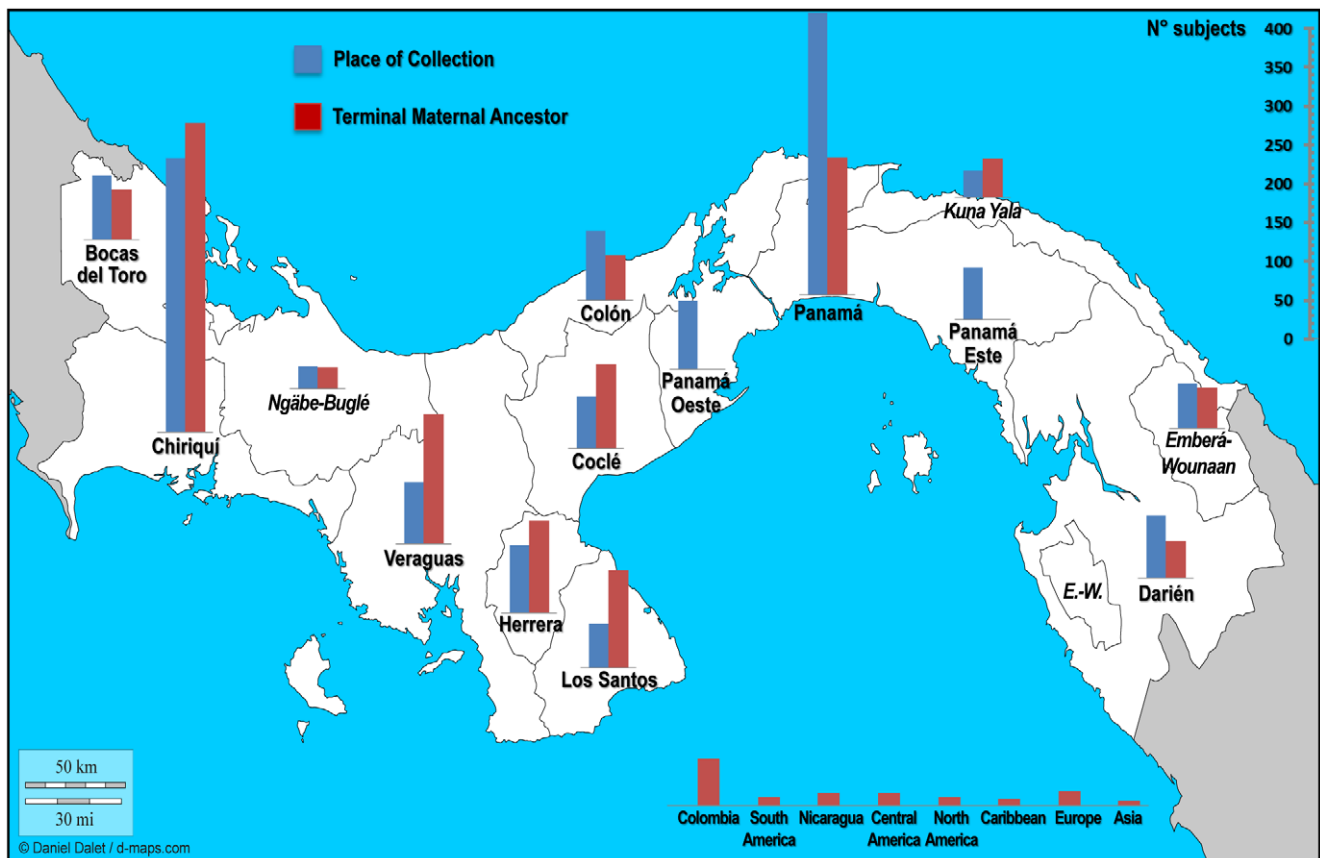
### Comparing Molecular and Genealogical Data

MtDNA profiles for 1565 samples – collected in different Panamanian provinces or Native American comarcas (Figure 1) – were determined by sequencing 1150 base pairs (bps) from nucleotide position (np) 16000 to np 580 (Tables S1 and S2), thus covering the entire control region and including the three hypervariable segments (HVS-I: nps 16024–16383, HVS-II: nps 057–372, and HVS-III: nps 438–576). Excluding gaps and ambiguous sites, 227 polymorphisms and 865 invariable sites were identified in the control-region sequences with a nucleotide diversity ( $\pi$ ) of 0.01144. The average number of nucleotide differences ( $k$ ) between two randomly chosen sequences is 13.118. A total of 375 different haplotypes were observed, with an observed high diversity index ( $Hd = 0.971$ ). These data confirm the efficacy of the sampling design, where related subjects were avoided. An accurate survey of mutational diagnostic motifs in the control region allowed the classification of mtDNAs into many haplogroups, sub-haplogroups, and paragroups, following the most updated mtDNA phylogeny and nomenclature [39] (Table S1).

Upon evaluating the genealogical data (with an average value of  $3.13 \pm 0.87$  ancestral generations) we found that 149 of the collected samples have a last known forebear on the maternal line from abroad, which means a non-Panamanian terminal maternal ancestor (TMA, Figure 1): 48.3% of them from South America (mostly from Colombia, 40.9%), followed by Central America (21.5%, half from Nicaragua), Europe (12.8%), North America (7.4%), Caribbean (6.0%), and Asia (4.0%). None of the TMAs were reported as being from Africa, probably due to the lack of genealogical records available during the slave trade era. It is worth noting that samples are all consistently assigned to haplogroups typical of the TMA area of origin (Table S1). Panamanian pedigrees were also analyzed in detail, and a search was conducted using the Sorenson Molecular Genealogy Foundation mtDNA database [40] in order to identify samples that might come from the same family (distant relatives) on the maternal side. A total of 123 cross-related samples (119 with a TMA from Panama, the other four from other countries) were identified: 42 pairs and 13 triplets (Table S1). Only one member for each of these extended families was retained. After excluding two samples with origins external to Panama (Costa Rica and Colombia), our final analyses were performed utilizing 1350 autochthonous and unrelated samples.

### Haplogroup Distribution

Approximately one fourth of the subjects with TMA from Panama were from the provinces of Chiriquí (28%), followed by Panamá (13%), and then Veraguas (12%) (Figure 1). These selected samples were assigned to more than a hundred different haplogroups and sub-haplogroups (Tables 1 and S1), grouped on



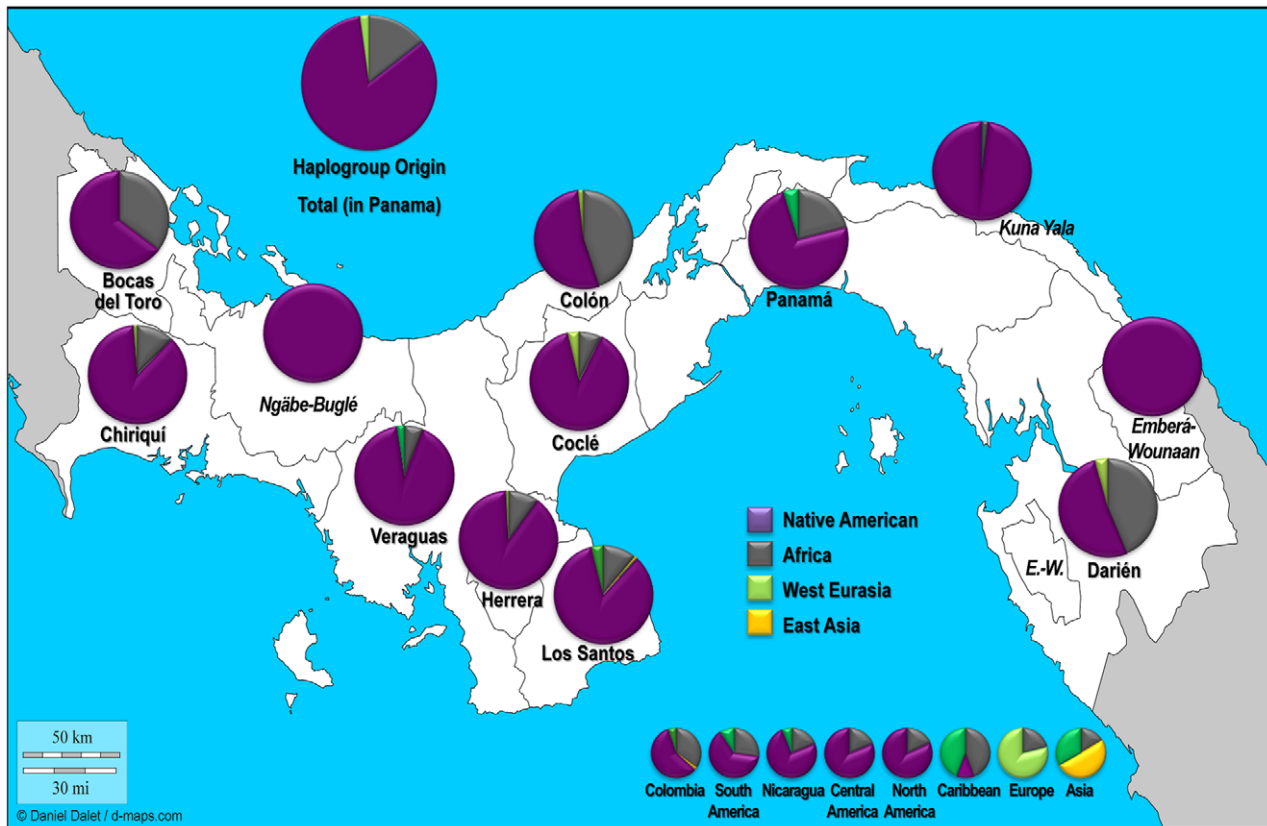
**Figure 1. Distributions in Panama of the 1565 samples analyzed in this paper.** Bars show both place of collection and terminal maternal ancestor (TMA) origin. This means the origin of the last known ancestor on the maternal side of the recorded pedigree.  
doi:10.1371/journal.pone.0038337.g001

the basis of their geographic/ethnic prevalence (Figure 2). A total of 83.5% of the analyzed lineages are of Native American origin [1,6,8,41]. Not surprisingly, a significant percentage of sub-Saharan African [42–44] (14.4%) mtDNAs were also detected (although none of the genealogical records gathered indicated such ancestry), while a few Western Eurasian [45,46] (2.1%) and East Asian [47,48] clades (only one G1a1 in Los Santos) were identified. It is worth noting that the highest frequency of European haplogroups is in the Panamá province (5.1%), geographically coinciding with a spike in the African lineage signals (20.8%). However, sub-Saharan lineages are most common across the Caribbean (Bocas del Toro, 35.0%; Colón, 45.6%) and in the easternmost province of Panama (Darién, 42.2%). All the four common “pan-American” haplogroups (A2, B2, C1, and D1) are well represented (83.5% overall), but none of the rare Native American haplogroups (D4h3a, X2a, and C4c) were observed. More than half of the Panamanians belong to haplogroup A2 (51.1%), the most common native lineage observed in Central America [49]. When comparing the haplogroup origin distribution between provinces and comarcas the difference is highly significant ( $\chi^2$   $p$ -value < 0.0001), since, as expected, virtually all the samples with a TMA from the three comarcas (Ngäbe-Buglé, Emberá-Wounaan, and Kuna Yala) show Native American mtDNAs (99.2%). Intriguingly, the frequency of Native American haplogroups is quite different among the three comarcas ( $\chi^2$   $p$ -value < 0.0001), with a prevalence of A2 in Kuna Yala (77.1%) and C1/D1 in Emberá-Wounaan (40.9%/6.8%), while the Ngäbe and Buglé together are about half A2 and half B2. Subsequently, when

taking into account the distribution of indigenous populations in Panama, these data become even more interesting: e.g. high percentages of C1 and D1 (6.7% and 2.2%, respectively) stood out in the Darién province, where the Emberá and Wounaan indigenous people are most prevalent [23]. Actually, the distribution of Native American haplogroups is significantly different ( $\chi^2$   $p$ -value = 0.0004) between the eastern (southern) area of Panama (Panamá Gulf provinces plus Colón and Kuna Yala) and the western (northern) area. In fact, haplogroup C1 and D1 frequencies are noticeably higher in the eastern than the western area.

### Haplogroup A2

A total of 198 different A2 haplotypes (comprising 689 mtDNAs) were identified in Panama with the most common (53 mtDNAs) having the following mutational motif: 16111, 16223, 16290, 16319, 16360, 16362, 89, 106–111d, 146, 153, 198, 235, 263, 309.1C, 309.2C, 315.1C, and 522–523d. This and 79 other haplotypes were found to share the motif 64@, 73@, 106–111d, and 16360 (a total of 326 mtDNAs). The deletion of six base pairs (nps 106–111) was reported previously in the Huetars of Costa Rica and several other Chibchan-speaking groups of Central America. The additional control-region mutations detected in these haplotypes allow for a better classification of this particular A2 subclade, here preliminarily named A2af. A phylogenetic analysis of these mitochondrial genomes was performed through a network structure (Figure 3), resulting in four different sub-branches with a large prevalence of a particular one marked by the



**Figure 2. Spatial frequency (%) distributions of mtDNA haplogroups among the analyzed samples.**  
doi:10.1371/journal.pone.0038337.g002

89 transition. We employed the recent control-region mutation rate published by Soares et al. [50] to date the entire A2af clade at  $23.24 \pm 8.96$  ka.

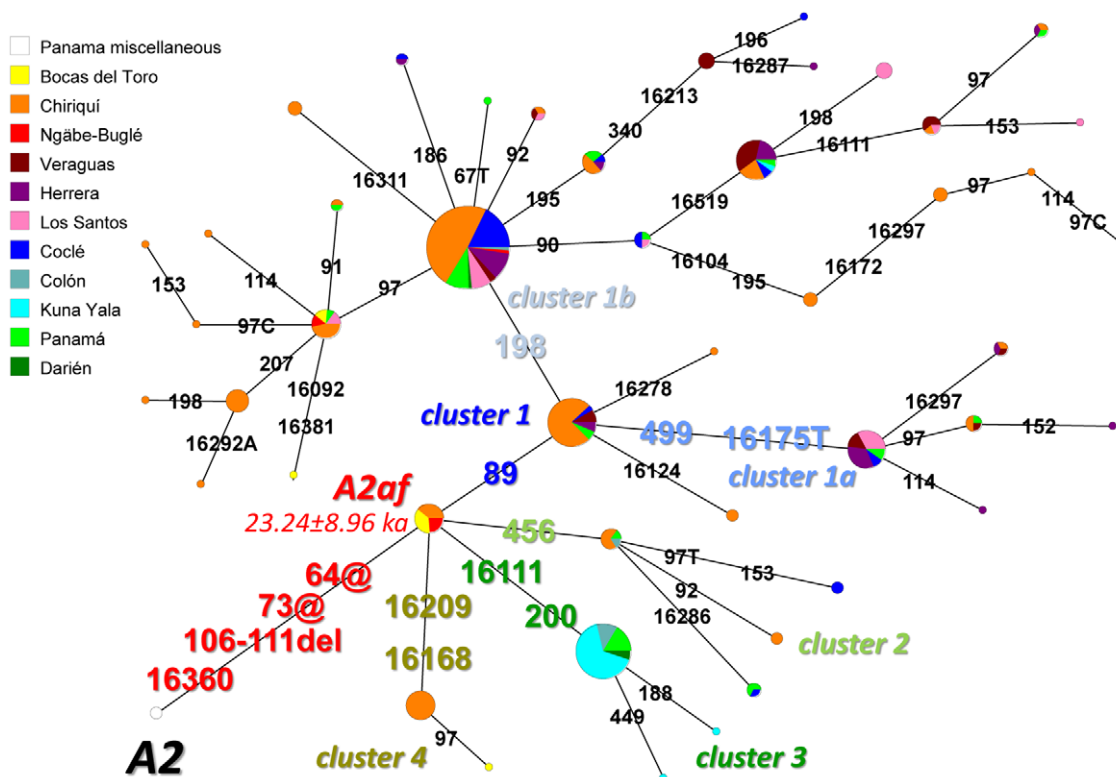
Entire Mitochondrial Genomes

To increase the resolution of our analyses, we also evaluated entire mitochondrial genomes. A total of 18 novel A2af mitochondrial genomes were completely sequenced (Table 2): 16

**Table 1. Frequency distribution (%) of major haplogroups in Panama.**

Province/Comarca	n°	A2	B2	C1	D1	Native	L0	L1	L2	L3	Africa	HV	JT	N1'2	U	West Eurasia	G	East Asia
Bocas del Toro	60	46.67	18.33	-	-	<b>65.00</b>	-	3.33	18.33	13.33	<b>35.00</b>	-	-	-	-	<b>0.00</b>	-	<b>0.00</b>
Chiriquí	378	60.05	24.07	1.06	1.32	<b>86.51</b>	1.59	4.76	4.23	1.59	<b>12.17</b>	0.79	0.53	-	-	<b>1.32</b>	-	<b>0.00</b>
Coclé	109	45.87	35.78	6.42	0.92	<b>88.99</b>	-	0.92	2.75	3.67	<b>7.34</b>	1.83	0.92	0.92	-	<b>3.67</b>	-	<b>0.00</b>
Colón	57	33.33	19.30	1.75	-	<b>54.39</b>	-	5.26	26.32	12.28	<b>43.86</b>	-	-	-	1.75	<b>1.75</b>	-	<b>0.00</b>
Darién	45	28.89	17.78	6.67	2.22	<b>55.56</b>	-	11.11	20.00	8.89	<b>40.00</b>	2.22	-	-	2.22	<b>4.44</b>	-	<b>0.00</b>
Herrera	116	61.21	24.14	3.45	-	<b>88.79</b>	-	-	4.31	6.03	<b>10.34</b>	-	-	-	0.86	<b>0.86</b>	-	<b>0.00</b>
Los Santos	122	56.56	24.59	2.46	0.82	<b>84.43</b>	0.82	-	2.46	7.38	<b>10.66</b>	3.28	-	-	0.82	<b>4.10</b>	0.82	<b>0.82</b>
Panamá	178	43.82	25.84	2.81	1.69	<b>74.16</b>	-	5.06	6.18	9.55	<b>20.79</b>	4.49	-	-	0.56	<b>5.06</b>	-	<b>0.00</b>
Veraguas	166	46.39	39.16	6.02	-	<b>91.57</b>	0.60	0.60	2.41	2.41	<b>6.02</b>	2.41	-	-	-	<b>2.41</b>	-	<b>0.00</b>
<b>Total Provinces</b>	<b>1231</b>	<b>51.34</b>	<b>26.73</b>	<b>3.01</b>	<b>0.89</b>	<b>81.97</b>	<b>0.65</b>	<b>3.17</b>	<b>6.26</b>	<b>5.36</b>	<b>15.43</b>	<b>1.79</b>	<b>0.24</b>	<b>0.08</b>	<b>0.41</b>	<b>2.52</b>	<b>0.08</b>	<b>0.08</b>
Emberá-Wounaan	44	15.91	36.36	40.91	6.82	<b>100.00</b>	-	-	-	-	<b>0.00</b>	-	-	-	-	<b>0.00</b>	-	<b>0.00</b>
Kuna Yala	48	77.08	12.50	8.33	-	<b>97.92</b>	-	-	2.08	-	<b>2.08</b>	-	-	-	-	<b>0.00</b>	-	<b>0.00</b>
Ngäbe-Buglé	27	48.15	48.15	3.70	-	<b>100.00</b>	-	-	-	-	<b>0.00</b>	-	-	-	-	<b>0.00</b>	-	<b>0.00</b>
<b>Total Comarcas</b>	<b>119</b>	<b>47.90</b>	<b>29.41</b>	<b>19.33</b>	<b>2.52</b>	<b>99.16</b>	-	-	<b>0.84</b>	-	<b>0.84</b>	-	-	-	-	<b>0.00</b>	-	<b>0.00</b>
<b>Grand Total</b>	<b>1350</b>	<b>51.04</b>	<b>26.96</b>	<b>4.44</b>	<b>1.04</b>	<b>83.48</b>	<b>0.59</b>	<b>2.89</b>	<b>5.78</b>	<b>4.89</b>	<b>0.84</b>	<b>1.63</b>	<b>0.22</b>	<b>0.07</b>	<b>0.37</b>	<b>2.30</b>	<b>0.07</b>	<b>0.07</b>

doi:10.1371/journal.pone.0038337.t001



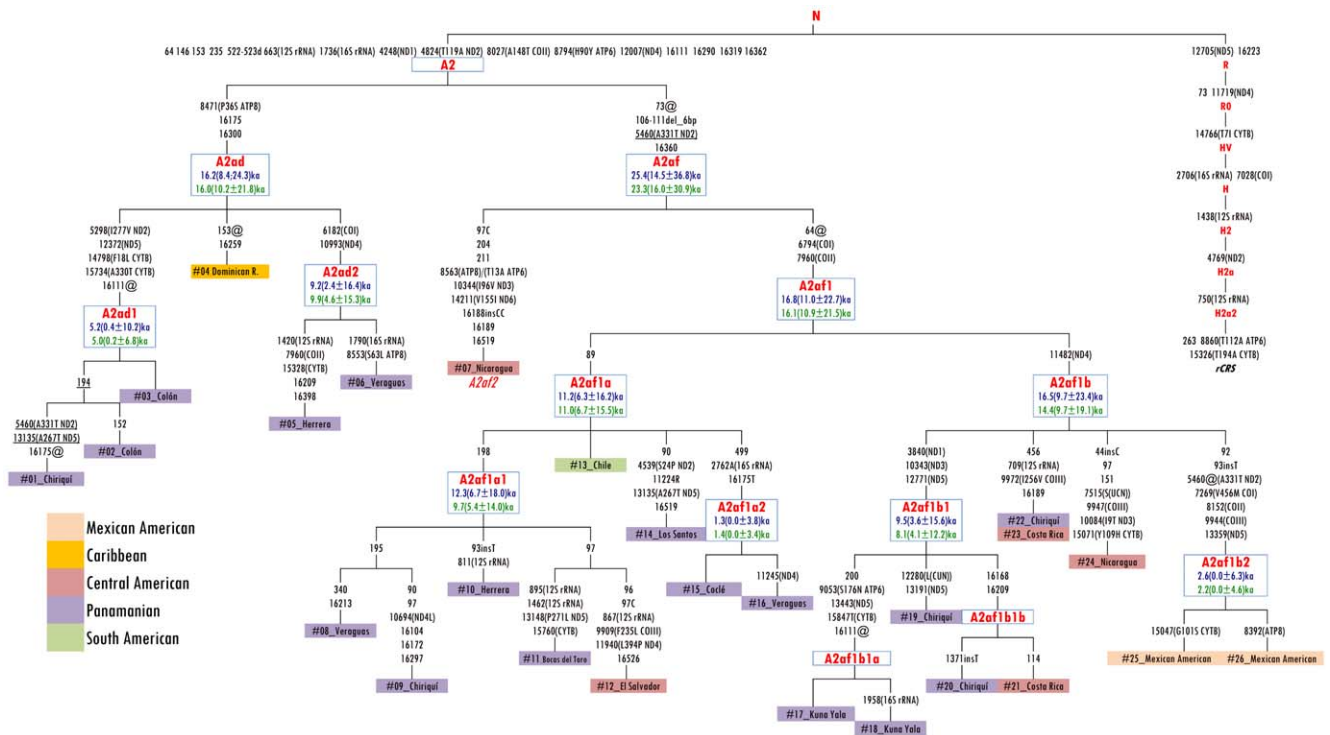
**Figure 3. Network of A2af control-region haplotypes from Panama subdivided according to their geographic origin.** The mutations on the connecting branches refer to the (revised) Cambridge reference sequence (rCRS) [52]. Markers of different clusters are in colors. Mutations are transitions unless the base change is explicitly indicated. Insertions, deletions and heteroplasmic mutations were excluded, with the notable exception of the 106–111 6 bp deletion. The size of each circle is proportional to the haplotype frequency and geographical origins are indicated by different colors. Coalescence ages of A2af and A2af1 are also reported using the control-region mutation rate reported by Soares *et al.* [50]. doi:10.1371/journal.pone.0038337.g003

were collected in Panama (listed in Table S1); two others, collected in El Salvador and Chile, were already available in the SMGF dataset [40]. Overall, when looking at the TMA of these 18 samples, 12 were from Panama, two from Costa Rica, two from Nicaragua, one from El Salvador, and one from Chile. The evolutionary history of their mtDNAs was inferred by a parsimony approach and compared to two other Mexican American mtDNAs (MA145 and MA148) reported by Kumar *et al.* [51]. The latter two samples were misclassified as A2s, but they actually carry a clear A2af mutational motif. Through the phylogeny of Figure 4, rooted with the (revised) Cambridge Reference Sequence rCRS [52], we found that the Nicaraguan sequence #07 did not cluster either with A2af, or any of the known Old World A2 branches [39]. Its presence allows us to better define the A2af basal motif (73@, 106–111d, 5460, 16360), while the reversion at np 64, previously thought to be ancestral to the A2af (Figure 3), characterizes a major sub-branch A2af1 together with two coding-region transitions at nps 6794 and 7960. On the other hand, the Nicaraguan sequence #07 most likely indicates an additional and very rare Native American A2af sub-clade, here named A2af2. Concerning the major cluster of the phylogeny, we were able to date the terminal maternal ancestor of haplogroup A2af1 at ~17 ka ago. The complete sequence analysis confirms the main sub-branching (A2af1a) marked by the control-region mutation 89, already detected in the control-region network analysis, but reveals also an additional sister branch defined only by a coding region transition at np 11482, here named A2af1b. Surprisingly, a member of A2af1a is from Chile and was one of the two A2af

mtDNAs from South America that were found in the SMGF control-region database, the other one was from Colombia. Considering the overlapping patterns of Native lineages (obtained in some recent papers [5,49]) when comparing haplogroup frequency distributions from general-mixed populations to that of Native American tribes or communities, we proceeded to analyze the incidence of A2af among the 79,928 records (as of March 2<sup>nd</sup>, 2012) of the SMGF database (Table S3A). Figure 5 shows that this peculiar haplogroup is detected at low frequencies along the Pacific coast and the western side of the Andes, but with great incidences in lower Central America, having its highest peaks in Costa Rica (12.16%) and Panama (24.15%). Previously, A2's HVS-I haplotypes carrying the 6-bp deletion were observed almost exclusively in Central America (Table S3B), especially (frequency >10%) among the Huetars (56%) [36] and the Bribri (74%) [38] of Costa Rica, and in Panama among the Ngäbe (9–12%) [32] [41] and the Kuna (53%) [41].

The second most common A2 subgroup in Panama is defined by the control-region mutational motif 16175–16300, representing about 3% of modern Panamanians. This new clade, named A2ad, was found in the SMGF database (and in the literature, Table S3B) in mostly the same countries as A2af, but at lower frequencies (Figure 5). This similarity is both spatial and temporal. In fact, A2ad was dated at approximately 16 ka ago, by looking at the entire sequence variation of five randomly selected Panamanian mtDNAs and one published Dominican sample [1] belonging to this clade (Figure 4).





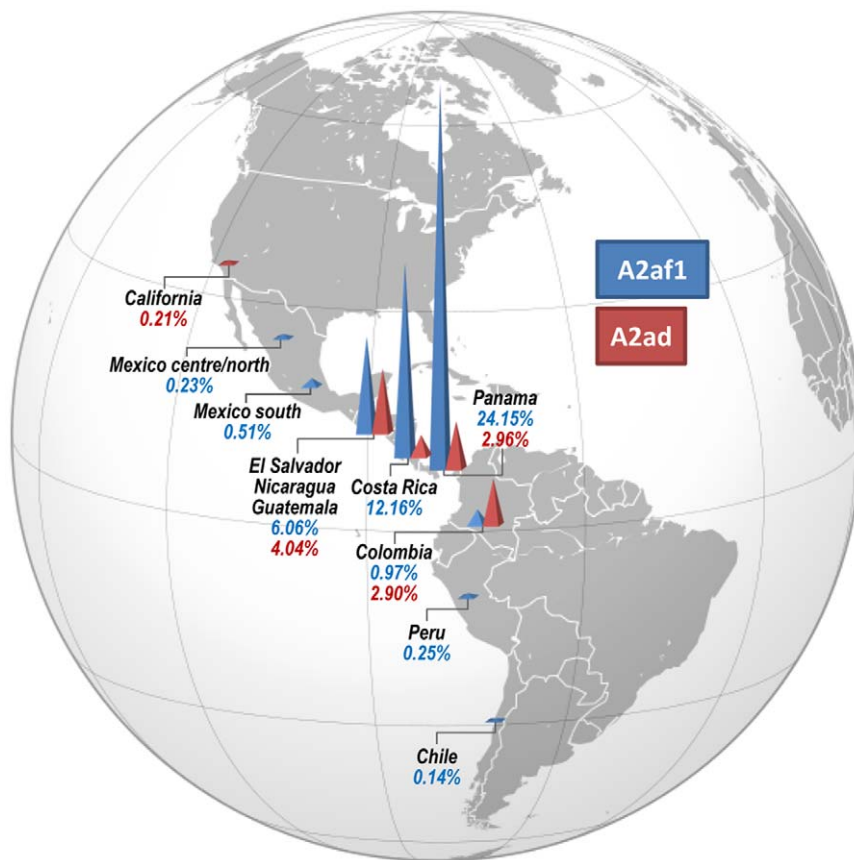
**Figure 4. Phylogeny of complete mtDNA sequences belonging to haplogroups A2ad and A2af.** The tree was rooted by using the reference sequence rCRS that is indicated for reading off sequence motifs. All sequences are new except for #04, #25 and #26 (Table 2). Mutations are shown on the branches; they are transitions unless a base is explicitly indicated. Suffixes indicate: reversions (@), transversions (to A, G, C, or T), indels (ins, del), gene locus, synonymous or non-synonymous changes. Recurrent mutations within the A2 branch are underlined. Any length variations in the C-stretch between nucleotides 303–315 and 16184–16193 were disregarded. Additional information regarding each mtDNA is available on Table 2. Coalescence times were calculated by converting into years [50] the averaged distance ( $\rho$ , in blue) and the maximum likelihood (ML, in green) estimate, calculated by considering all the substitutions on the entire mitochondrial genome. doi:10.1371/journal.pone.0038337.g004

**Discussion**

According to the last official census [23], less than 420,000 Panamanians (12.3%) recognize themselves as Native American, while about 313,000 declare to be Afro-descendants (9.2%). Most of the Native American population live in the three comarcas (95%) where people of African descent are less than 1%; thus when looking at the provinces’ general population, indigenous percentage decreases to 7.1%, while Afro-descendants remain almost the same (9.7%). In this study, we provided a comprehensive overview of the mitochondrial gene pool of Panamanians, with ~80% of mtDNAs belonging to the native lineages A2, B2, C1, and D1, with A2 being the predominant clade. A similar extreme pattern was actually observed in another Central American country, El Salvador [53], where the African lineages are virtually absent (only one mtDNA of sub-Saharan origin). In Panama, the sub-Saharan component consists of 57 haplogroups and sub-haplogroups accounting for 14.4% of modern Panamanians, mostly identified in the Pacific provinces of Panamá and Darién and the Caribbean provinces of Colón and Bocas del Toro, where the Atlantic slave trade and more recent migrations from the Caribbean islands and northern Colombia clearly had a more relevant demographic impact [54,55]. The oldest African population stems from African slaves since all Spanish settlements would have had some. There would have been concentrations around early towns such as Panama la Vieja, Natá, Los Santos, Nombre de Dios and Portobelo and also at mines like the Concepción mine in Caribbean Veraguas where several thousand slaves were employed in the late 16<sup>th</sup> century [56]. African lineages

found in Darién (Figure 2) are likely to derive directly from slaves who escaped Spanish rule and were known as “*cimarrones*,” although the more recent immigration of people of African descent from northernmost Colombia should be taken into account. This population and others settled between Colón and Kuna Yala, and along the western Caribbean, speak Spanish, and are culturally different from the descendants of more recent immigrants from French- and English-speaking islands in the Caribbean who came to work on the Panama Canal. Another group of African people are the “turtlers” (*tortugueros*) who settled in the western Caribbean mostly from English-speaking islands such as Providence and the Cayman Islands. The cemetery at Drago on Isla Colón has grave stones of people born in Cayman in 1790 AD [57].

In conclusion, our molecular data reveal an overwhelming native maternal legacy in the modern Panamanian population. It seems that the Spanish conquistadores and additional more recent European demographic influences did not contribute significantly to today’s genetic composition of Panama, at least with regard to the maternal side. These data are in contrast with the overall concept of personal origin shared by many Panamanians. Moreover, through the micro-phylogeographic approach proposed with the current Panamanian dataset, we were able to confirm a distinct sub-structure of native lineage distributions. Haplogroups C1 and D1 harbour much higher frequencies in the eastern (southern) area of Panama, particularly in Darién, where the Emberá and Wounaan indigenous people now prevail. A2 is prevalent in Kuna Yala, where 62.5% of people carry the



**Figure 5. Spatial distribution of A2af and A2ad mtDNAs identified in general mixed populations.** Exact values are listed in Table S3A. doi:10.1371/journal.pone.0038337.g005

previously called Huetar deletion (see introduction), thus belonging to the newly defined A2af haplogroup. Complete sequence analyses of this and the other most common A2 sub-lineage in Panama (A2ad) places their founder ages at more than 10 ka ago, highlighting an ancient expansion and settlement through this area. These two lineages are not found in North America and along the Atlantic coast, but can be observed at low frequencies as far south as Peru and Chile (Valparaiso). Considering the most recently accepted age estimate for haplogroup A2 in the American continent as a whole at 15–19 ka ago [5,7,51] and as a proxy for the time of expansion of Paleo-Indians into the Americas, it can be suggested that the initial settlement of Panama occurred fairly rapidly after the initial colonization of the American continent. These data fully support the hypothesis that the Pacific coast was the major entry point and diffusion route for the earliest human settlers. Moreover, the antiquity and high frequency of subclade A2af provides evidence of the existing mitochondrial DNA legacy between modern Panamanians and America's first inhabitants.

## Materials and Methods

### Ethics Statement

All experimental procedures and individual written informed consent, obtained from all donors, were reviewed and approved by the Comité Nacional de Bioética de la Investigación of Panama and by the Western Institutional Review Board, Olympia, Washington (USA).

### Specimen Collection

A total of 1565 saliva samples were collected from healthy unrelated individuals in different areas of Panama in collaboration with the Gorgas Memorial Institute for Health Studies. The field sampling was undertaken with the kind help of local assistants.

Mouthwash rinsing was the primary method of biological specimen collection using a method (GenetiRinse™) that comprises the use of 10cc of commercially available mint-flavored Scope™ mouthwash in a 15cc volume leak-free Nalgene™ plastic container. Participants swished the 10cc of mouthwash for 45 seconds and then spat the mouthwash back into its original container. Participants were asked to abstain from eating or drinking for at least 30 minutes prior to the mouthwash rinse. Total DNA was extracted from mouthwash using standard commercial kits (Qiamp DNA Blood Maxi Kit, Qiagen) and stored at  $-20^{\circ}\text{C}$ .

### Analysis of mtDNA Control-Regions

The first step of the mtDNA molecular analyses consisted of DNA PCR amplification of the control region using the primers listed in Table S2. After PCR, fragments were purified using the ExoSAP-IT® enzymatic system (Exonuclease I and Shrimp Alkaline Phosphatase, GE Healthcare) and Cycle Sequencing was performed by application of ABI PRISM™ BigDye Terminator chemistry. A protocol including three forward and three reverse sequencing primers was used (Table S2). Primer redundancy is employed particularly for mtDNAs harbouring the transition T16189C, which creates a poly-C tail that causes premature termination of the sequencing reaction. The additional

**Table 2.** Sources and haplogroup affiliation for the A2ad and A2af complete mtDNA sequences.

ID (Fig. 4)	Place of Collection	TMA Origin	Clade	GenBank	Original ID
#01	Chiriquí	Panama (Chiriquí)	A2ad1	JQ965966	S-947838
#02	Colón	Panama (Colón)	A2ad1	JQ965967	S-947655
#03	Colón	Panama (Colón)	A2ad1	JQ965968	S-947625
#04	Dominican R	Dominican R	A2ad	EF079873	#7_A2(Tor22)
#05	Herrera	Panama (Herrera)	A2ad2	JQ965969	S-996787
#06	Panamá	Panama (Veraguas)	A2ad2	JQ965970	S-915431
#07	Panamá	Nicaragua	A2af	JQ965971	S-915367
#08	Panamá	Panama (Veraguas)	A2af1a1	JQ965972	S-992683
#09	Chiriquí	Panama (Chiriquí)	A2af1a1	JQ965973	S-948115
#10	Darién	Panama (Herrera)	A2af1a1	JQ965974	S-943682
#11	Bocas del Toro	Panama (Bocas del Toro)	A2af1a1	JQ965975	S-948541
#12	El Salvador	El Salvador	A2af1a1	JQ965976	S-635621
#13	Chile	Chile	A2af1a	JQ965977	S-686714
#14	Los Santos	Panama (Los Santos)	A2af1a	JQ965978	S-996672
#15	Chiriquí	Panama (Coclé)	A2af1a2	JQ965979	S-947916
#16	Veraguas	Panama (Veraguas)	A2af1a2	JQ965980	S-992645
#17	Panamá	Panama (Kuna Yala)	A2af1b1a	JQ965981	S-937137
#18	Panamá	Panama (Kuna Yala)	A2af1b1a	JQ965982	S-915331
#19	Bocas del Toro	Panama (Chiriquí)	A2af1b1	JQ965983	S-948757
#20	Chiriquí	Panama (Chiriquí)	A2af1b1b	JQ965984	S-947228
#21	Bocas del Toro	Costa Rica	A2af1b1b	JQ965985	S-948517
#22	Chiriquí	Panama (Chiriquí)	A2af1b	JQ965986	S-948533
#23	Chiriquí	Costa Rica	A2af1b	JQ965987	S-948446
#24	Chiriquí	Nicaragua	A2af1b	JQ965988	S-948713
#25	US (Texas)	Mexican American	A2af1b2	HQ012106	MA148
#26	US (Texas)	Mexican American	A2af1b2	HQ012104	MA145

doi:10.1371/journal.pone.0038337.t002

reverse primers solve the problem by completing the sequence information with multiple reads.

Electropherograms were aligned, assembled, and compared using the software Sequencher<sup>TM</sup> 5.1 (Gene Codes). Finally the mutational differences relative to the Cambridge reference sequence (rCRS) [52] were accurately analyzed in order to identify mutational motifs for haplogroup classification following the most updated human mitochondrial phylogeny [39].

The median-joining network of control-region haplotypes observed in 326 A2af mtDNAs was constructed by using the Network 4.6 software program (<http://www.fluxus-engineering.com>). The time estimates option was employed to measure the age of an ancestral node in mutational units. This mutational age was then converted into years by the recent control-region mutation rate published by Soares et al. [50].

### Analysis of Entire Mitochondrial Genomes

Sequencing of entire mtDNA genomes belonging to haplogroups A2ad and A2af were performed as previously described [5,58]. In order to obtain coalescence times, we directly calculated the average distance ( $\rho$ ) of the haplotypes of a clade to the respective root haplotype, accompanied by a heuristic estimate of the standard error ( $\sigma$ ) calculated from an estimate of the genealogy. PAML 4.5 [59] was used to calculate maximum likelihood (ML) estimates, assuming the HKY85 mutation model (with indels ignored, as usual) with gamma-distributed rates

(approximated by a discrete distribution with 32 categories) and three partitions: HVS-I (positions 16051–16400), HVS-II (positions 68–263), and the remainder. These calculations were performed on entire mtDNA haplotypes (excluding the mutations 16182C, 16183C, 16194C and 16519). Mutational distances were converted into years using the substitution rate for the entire molecule of about one mutation every 3,624 years [50].

### Supporting Information

**Table S1** Control-region haplotypes (relative to rCRS) and haplogroup/sub-haplogroup classification (based on PhyloTree, Built 14) of the 1565 mtDNAs collected in Panama and deposited in the SMGF database (<http://www.smgf.org>). (XLS)

**Table S2** Oligonucleotides used for amplifying and sequencing the entire control region. (PDF)

**Table S3** Distribution of haplogroups A2ad and A2af (**a**) in the SMGF database (general mixed populations) and (**b**) in the literature (native samples and forensic/population cohorts). (PDF)



## Author Contributions

Conceived and designed the experiments: UAP MT SRW JMP JM AA. Performed the experiments: UAP HL AO. Analyzed the data: UAP MT

## References

- Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, et al. (2008) The phylogeny of the four pan-American mtDNA haplogroups: implications for evolutionary and disease studies. *PLoS ONE* 3: e1764.
- Bodner M, Perego UA, Huber G, Fendt L, Röck AW, et al. (2012) Rapid coastal spread of First Americans: novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res* Epub Feb 14.
- Hooshiar Kashani B, Perego UA, Olivieri A, Angerhofer N, Gandini F, et al. (2012) Mitochondrial haplogroup C4c: a rare lineage entering America through the ice-free corridor? *Am J Phys Anthropol* 147: 35–39.
- Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, et al. (2012) Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am J Hum Genet* 90: 229–246.
- Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, et al. (2009) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19: 1–8.
- O'Rourke DH, Raff JA (2010) The human genetic history of the Americas: the final frontier. *Curr Biol* 20: R202–207.
- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogó MR, et al. (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82: 583–592.
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, et al. (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53: 563–590.
- Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59: 935–945.
- Kaufman T, Golla V (2000) Language groupings in the New World: their reliability and usability in cross-disciplinary studies. In: Renfrew C, ed. *America past, America present: genes and language in the Americas and beyond*. Cambridge: McDonald Institute for Archaeological Research. pp 47–57.
- Schurr TG, Sherry ST (2004) Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol* 16: 420–439.
- Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, et al. (2007) Genetic variation and population structure in Native Americans. *PLoS Genet* 3: e185.
- Wallace DC, Torroni A (1992) American Indian prehistory as written in the mitochondrial DNA: a review. *Hum Biol* 64: 403–416.
- Waters MR, Stafford TWJ (2007) Redefining the age of Clovis: implications for the peopling of the Americas. *Science* 315: 1122–1126.
- Gilbert MTP, Jenkins DL, Götherström A, Naveran N, Sanchez JJ, et al. (2008) DNA from pre-Clovis human coprolites in Oregon, North America. *Science* 320: 786–789.
- Goebel T, Waters MR, Dikova M (2003) The archaeology of Ushki Lake, Kamchatka, and the Pleistocene peopling of the Americas. *Science* 301: 501–505.
- Goebel T, Waters MR, O'Rourke DH (2008) The late Pleistocene dispersal of modern humans in the Americas. *Science* 319: 1497–1502.
- Dillehay TD, Ramirez C, Pino M, Collins MB, Rossen J, et al. (2008) Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* 320: 784–786.
- O'Rourke DH (2009) Human migrations: the two roads taken. *Curr Biol* 19: R203–205.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463: 757–762.
- Crawford MH (1998) *The origins of Native Americans: evidence from anthropological genetics*. Cambridge; New York: Cambridge University Press. XV: 308.
- Harding RC (2006) *The history of Panama*. Westport, Conn.: Greenwood Press. xviii: 153.
- Censos Nacionales Available: <http://estadisticas.contraloria.gob.pa/Resultados2010/>.
- Cooke R (2005) Prehistory of native Americans on the Central American land bridge: Colonization, dispersal, and divergence. *J Archaeol Res* 13: 129–187.
- Dillehay TD (2009) Probing deeper into first American studies. *Proc Natl Acad Sci U S A* 106: 971–978.
- Piperno DR, Bush M, Flenley J, Gosling W (2011) Prehistoric human occupation and impacts on Neotropical forest landscapes during the Late Pleistocene and Early/Middle Holocene. *Tropical Rainforest Responses to Climatic Change*: Springer Berlin Heidelberg. pp 185–212.
- Romoli K (1987) *Los de la lengua de cueva: los grupos indígenas del istmo oriental en la época de la conquista española*: Instituto Colombiano de Antropología.
- Barrantes R, Smouse PE, Mohrenweiser HW, Gershowitz H, Azofeifa J, 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: 63–84.
- Constelma Umaña A (1991) *Las lenguas del Área Intermedia: introducción a su estudio areal*; Editorial de la Universidad de Costa Rica SJ, editor. 216 p.
- Ruiz-Narváez EA, Santos FR, Carvalho-Silva DR, Azofeifa J, Barrantes R, et al. (2005) Genetic variation of the Y chromosome in Chibcha-speaking Amerindians of Costa Rica and Panama. *Hum Biol* 77: 71–91.
- Batista O, Kolman CJ, Bermingham E (1995) Mitochondrial DNA diversity in the Kuna Amerinds of Panamá. *Hum Mol Genet* 4: 921–929.
- Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, et al. (1995) Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá. *Genetics* 140: 275–283.
- Ascunce MS, González-Oliver A, Mulligan CJ (2008) Y-chromosome variability in four Native American populations from Panama. *Hum Biol* 80: 287–302.
- Santos M, Barrantes R (1994) D-loop mtDNA deletion as a unique marker of Chibchan Amerindians. *Am J Hum Genet* 55: 413–414.
- Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994) Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. *Proc Natl Acad Sci U S A* 91: 1158–1162.
- Santos M, Ward RH, Barrantes R (1994) mtDNA variation in the Chibcha Amerindian Huetaar from Costa Rica. *Hum Biol* 66: 963–977.
- Rickards O, Martínez-Labarga C, Lum JK, De Stefano GF, Cann RL (1999) mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. *Am J Hum Genet* 65: 519–530.
- Santos M, Barrantes R (1994) Direct screening of a mitochondrial DNA deletion valuable for Amerindian evolutionary research. *Hum Genet* 93: 435–436.
- van Oven M, Kayser M Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30: E386–394. Available: <http://www.phylotree.org>.
- SMGF The Sorenson Molecular Genealogy Foundation Mitochondrial Database. Available: <http://www.smgf.org>.
- Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, et al. (2007) Beringian standstill and spread of Native American founders. *PLoS ONE* 2: e829.
- Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22: 339–345.
- Behar DM, Villemers R, Soodyall H, Blue-Smith J, Pereira L, et al. (2008) The dawn of human matrilineal diversity. *Am J Hum Genet* 82: 1130–1140.
- Campbell MC, Tishkoff SA (2010) The evolution of human genetic and phenotypic variation in Africa. *Curr Biol* 20: R166–173.
- Soares P, Achilli A, Semino O, Davies W, Macaulay V, et al. (2010) The archaeogenetics of Europe. *Curr Biol* 20: R174–183.
- Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, et al. (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144: 1835–1850.
- Stoneking M, Delfin F (2010) The human genetic history of East Asia: weaving a complex tapestry. *Curr Biol* 20: R188–193.
- Kong Q-P, Bandelt H-J, Sun C, Yao Y-G, Salas A, et al. (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet* 15: 2076–2086.
- Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, et al. (2010) The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20: 1174–1179.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, et al. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84: 740–759.
- Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, et al. (2011) Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol Biol* 11: 293.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147.
- Salas A, Lovo-Gómez J, Alvarez-Iglesias V, Cerezo M, Lareu MV, et al. (2009) Mitochondrial echoes of first settlement and genetic continuity in El Salvador. *PLoS ONE* 4: e6882.
- Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, et al. (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* 74: 454–465.
- Salas A, Richards M, Lareu MV, Sobrino B, Silva S, et al. (2005) Shipwrecks and founder effects: divergent demographic histories reflected in Caribbean mtDNA. *Am J Phys Anthropol* 128: 855–860.
- Castillero Calvo A (1995) *Conquista, evangelización y resistencia: triunfo o fracaso de la política indigenista?* Panama City: Editorial Mariano Arosemena.
- Dalton R (2011) Panama's Big Ambition. *Nature* 469: 462–463.

58. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75: 910–918.
59. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13: 555–556.