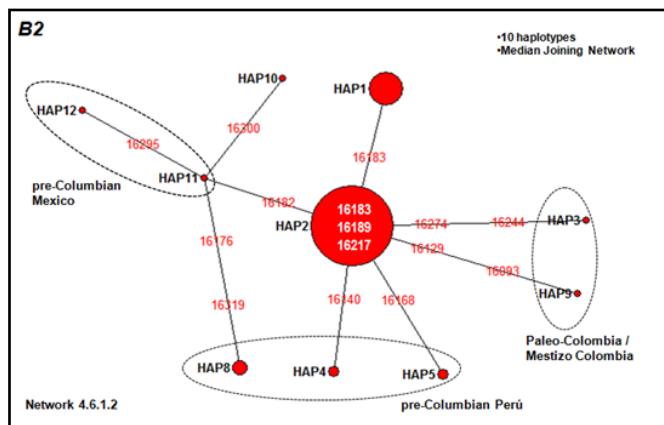


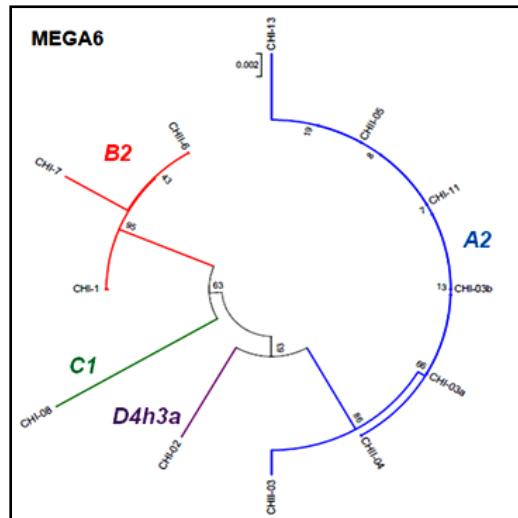
SUPPLEMENTAL MATERIALS

Supplemental Figure 1



Supplemental Figure 1. Phylogenetic Network of Haplotype B2 (138 ancient and modern mtDNAs HVR-I of the Americas: 10 haplotypes). The node size is related to its frequency. **HAP1** = ancient Perú, Ijka, Wayuú, and Mestizos-Colombian (Melton et al. 2007; Díaz-Matallana & Martínez-Cruzado 2010; Fehren-Schmitz et al. 2012); **HAP2**: Nodal Sequence of B2, **HAP2** = 2 Checuas (CHI-01, CHII-06) [TS], pre-Columbian Perú (Fehren-Schmitz et al. 2012), Infant Mummy from San José de Suaita, Santander-Colombia (~1350-350 yr BP) (AMEL XX) [Díaz-Matallana et al. 2010; Díaz-Matallana 2015; TS], Colombians Valle, CA, NA & SA (Salas et al. 2007); **HAP3** = 1 Checua (CHI-07) [TS]; **HAP4**, **HAP5** = pre-Columbian Perú; =pre-Columbian Perú; **HAP8** = ancient Perú (Fehren-Schmitz et al. 2012); **HAP9**, **HAP10** = Colombian [Salas et al. 2007]; **HAP11** & **12** = ancient Mexico (Mata-Míguez et al. 2012). TS: This Study. NA, CA & SA: North, Central and South America.

Supplemental Figure 2



Supplemental Figure 2. N-J Phylogenetic Tree of Paleo-Colombian Checua haplogroups. The evolutionary history was inferred using the Neighbor-Joining method (Saitou & Nei 1987). The optimal tree with the sum of branch length = 0.05688724 is shown. The percentage of replicate trees in which the associated taxa clustered together in the Bootstrap test (1000 replicates), are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Genetic distances between pairs of haplogroups were computed, ordered from highest to lowest, as follows: A2-C1 (0,027), D4h3a-C1 (0,026), C1-B2 (0,025), A2-B2 (0,023), D4h3a-B2 (0,022), and D4h3a-A2 (0,018). The analysis involved 12 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 356 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

Supplemental Material Pairwise distances

Supplemental Table 1. Estimates of Pairwise distances between sequences (Checua individuals)

	1	2	3	4	5	6	7	8	9	10	11	12
1. CHI-02												
2. CHI-03a	0.017											
3. CHI-03b	0.017	0.000										
4. CHI-08	0.026	0.026	0.026									
5. CHI-11	0.017	0.000	0.000	0.026								
6. CHI-13	0.023	0.006	0.006	0.032	0.006							
7. CHII-03	0.020	0.003	0.003	0.029	0.003	0.008						
8. CHII-04	0.017	0.000	0.000	0.026	0.000	0.006	0.003					
9. CHII-05	0.017	0.000	0.000	0.026	0.000	0.006	0.003	0.000				
10. CHI-01	0.020	0.020	0.023	0.020	0.026	0.017	0.020	0.020				
11. CHI-07	0.026	0.026	0.029	0.026	0.032	0.023	0.026	0.026	0.006			
12. CHII-06	0.020	0.020	0.023	0.020	0.026	0.017	0.020	0.020	0.000	0.006		

The number of base substitutions per site from averaging over all sequence pairs is shown. Analyses were conducted using the Maximum Composite Likelihood model. The analysis involved 12 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 356 positions in the final dataset (MEGA6) (Tamura et al. 2013).

Supplemental Material ANOVAs

- In statistics, **ANOVA (ANalysis Of VAriance between groups)** is a collection of statistical models used in order to analyze the differences among group means and their associated procedures, developed by statistician and geneticist / evolutionary biologist R.A. Fisher and is sometimes known as "Anova Fisher" or "analysis of variance Fisher" due to the use of the distribution F Fisher for contrasting hypothesis. **One-way analysis of variance (one-way ANOVA)** is used to compare means of three or more samples (F distribution). This technique can be used only for numerical data.

Supplemental Table 2. Number of HVR-I Haplotypes in Amerindians from Central, South America, and the Caribbean

Population / Community / Ethnic Group	Size	No. Haplotypes	Region	Ref.
Ngöbe	46	7	Central America	(Melton et al. 2013)
Huetar	29	7	Central America	(Melton et al. 2013)
Kuna	63	7	Central America	(Batista et al. 1995)
Ciboney – pre-Columbian	15	10	Caribbean Cuba	(Lalueza-Fox et al. 2003)
Taíno – pre-Columbian	19	11	Caribbean Dominican Rep.	(Lalueza-Fox et al. 2001)
Arsario	28	4	Caribbean Colombia	(Melton et al. 2007)
Ijka	31	3	Caribbean Colombia	(Melton et al. 2007)
Kogi	21	3	Caribbean Colombia	(Melton et al. 2007)
Guahibo	59	12	Amazon Venezuela	(Vona et al. 2005)
Gaviao	27	7	Amazon Brazil	(Lewis et al. 2005)
Zoró	29	8	Amazon Brazil	(Melton et al. 2007)
Xavante	25	4	Amazon Brazil	(Melton et al. 2007)
Guane – pre-Columbian, Santander	17	9	Andes Colombia	(Casas-Vargas et al. 2011)
Checua – Paleo-Colombian, Cundinamarca	12	7	Andes Colombia	(This Study)
Cayapa	30	8	Andes Ecuador	(Rickards et al. 1999)
Palpa - pre-Columbian (Middle Horizon)	6	6	Palpa Perú	(Fehren-Schmitz et al. 2010)
Mapuche	39	13	Andes Argentina	(Melton et al. 2007)

ANOVA for Number of HVR-I haplotypes:

Group A: pre-Columbian communities [Ciboney, Taíno, Guane, Paleo-Colombian Checua (This Study), Palpa]

Group B: Native Americans Central America (Ngöbe, Huetar, Kuna).

Group C: Native Americans Caribbean Colombia (Arsario, Ijka, Kogi).

Group D: Native Americans Amazon (Guahibo, Gaviao, Zoró, Xavante).

Group E: Native Americans Andes South America (Cayapa, Mapuche).

Supplemental Table 3. ANOVA: Results

The results of an ANOVA statistical test performed at 17:49 on 21-JAN-2015

NOTE: this is too few data: 2 in group E This group will be ignored.

Source of Variation	Sum of Squares	d.f.	Mean	F
Between	55.38	3	18.46	4.012
Error	50.62	11	4.602	
Total	106.0	14		
The probability of this result, assuming the null hypothesis, is 0.037.				
Group A: Number of items= 5 6.00 7.00 9.00 10.0 11.0				
Mean = 8.6000 95% confidence interval for Mean: 6.489 thru 10.71 Standard Deviation = 2.07 High = 11.00 Low = 6.000 Median = 9.000 Average Absolute Deviation from Median = 1.60				
Group B: Number of items= 3 7.00 7.00 7.00				
Mean = 7.0000 95% confidence interval for Mean: 4.274 thru 9.726 Standard Deviation = 0.00 High = 7.000 Low = 7.000 Median = 7.000 Average Absolute Deviation from Median = 0.00				
Group C: Number of items= 3 3.00 3.00 4.00				
Mean = 3.3333 95% confidence interval for Mean: 0.6075 thru 6.059 Standard Deviation = 0.577 High = 4.000 Low = 3.000 Median = 3.000 Average Absolute Deviation from Median = 0.333				
Group D: Number of items= 4 4.00 7.00 8.00 12.0				
Mean = 7.7500 95% confidence interval for Mean: 5.389 thru 10.11 Standard Deviation = 3.30 High = 12.00 Low = 4.000 Median = 7.500 Average Absolute Deviation from Median = 2.25				

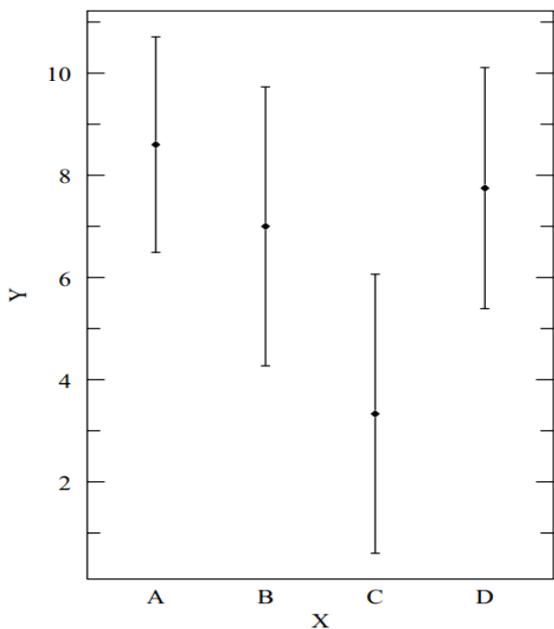
(See Suppl. Figures 3, 4) [Kirkman 1996]

Hypothesis ANOVA

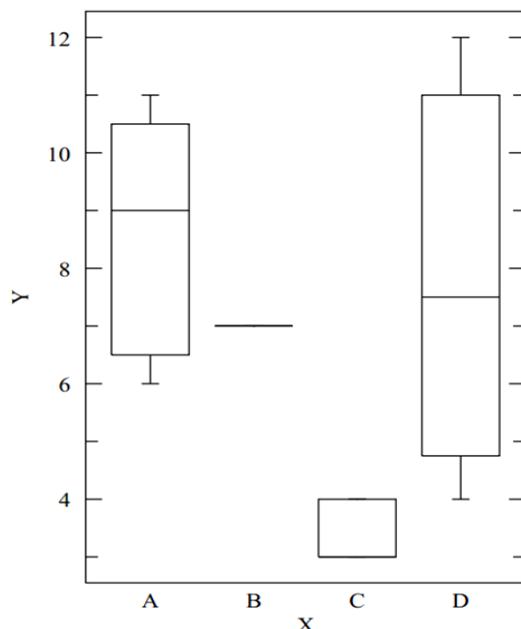
$$H_0: \mu(A) = \mu(B) = \mu(C) = \mu(D)$$

$$H_a: \mu(A) \neq \mu(B) \neq \mu(C) \neq \mu(D)$$

$p < 0.05$. The null hypothesis is rejected. The results are significant at 5%.



Supplemental Figure 3. Plot the Means Group of HVR-I Haplotypes with 95% Confidence Intervals. X: Groups; Y: 95% Confidence interval for haplotypes Means.
(Kirkman 1996)



Supplemental Figure 4. Box Plot for HVR-I haplotypes data. X: Groups, Y: Haplotypes data.

Supplemental Table 4. One-Way ANOVA from Summary Data (Also known as *Between-Subjects* ANOVA)

	Number of Subjects	Mean	Standard Deviation		
Group A:	68	8.60	2.07		
Group B:	138	7.00	0.00		
Group C:	80	3.33	0.57		
Group D:	140	7.75	3.30		
RESULTS:					
	SS	d.f.	MS	F	p
Between:	1,315.294	3	438.465	101.36	0.000
Within:	1,826.422	422	4.328		
Total:	3,141.859	425			

STATCALC v3 Beta (Soper 2015)

Hypothesis One-Way ANOVA

$H_0: \mu(A) = \mu(B) = \mu(C) = \mu(D)$

$H_a:$ At least two of the means are different from each other.

Critical value for the F-Distribution:

➤ 0.05 significance level: d.f. (3, 422)

Critical F-Value = 2.62604577

$F > F_{critical}$: the null hypothesis is rejected.

➤ 0.01 significance level: d.f. (3, 422)

Critical F-value= 3.82833023

$F > F_{critical}$: the null hypothesis is rejected.

$p < 0.05, p < 0.01$. The results are significant at 5% and 1% significance level.

STATCALC v3 Beta (Soper 2015)

Supplemental Material Principal Component Analysis (PCA): MaxStat Pro v3.60 (© 2015).

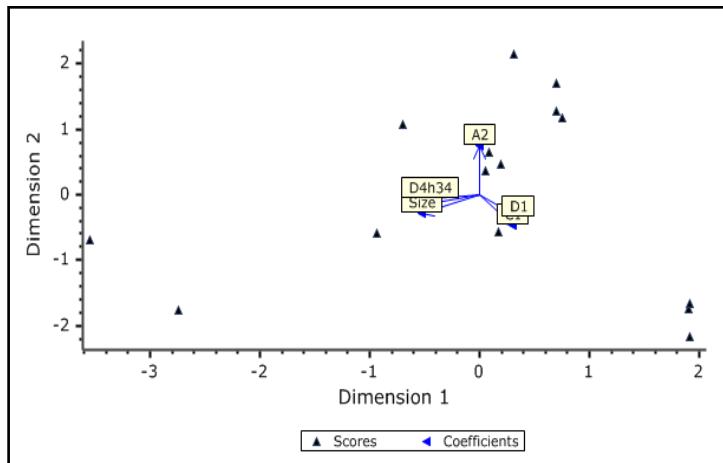
- ❖ The goal of PCA: To decompose a data table with correlated measurements into a new set of uncorrelated variables. Given a high correlation between two or more variables it can be concluded that these variables are quite redundant thus share the same driving principle in defining the outcome of interest. The new uncorrelated variables are called, depending upon the context, principal components, factors, eigenvectors, singular vectors, or loadings. Different statistical software may compute eigenvectors with different signs depending on the used algorithm, and inconsistent signs do not mean incorrect eigenvectors.
- ❖ Each unit is also assigned a set of scores which correspond to its projection on the components. The results of the analysis are often presented with graphs plotting the projections of the units onto the components, and the loadings of the variables. The graphs are called biplots.
- ❖ Is a way of identifying patterns in data, expressing the data in such a way as to highlight their similarities and differences.

Supplemental Table 5. Estimated frequencies of HVR-I Amerindian haplogroups in Ancient and Modern Groups (Raw Data)

	Population / Group	Size	A2	B2	C1	D1	D4h3a	Other	Ref.
ANCIENT PEOPLE	Xaltocan full (Mexico)	25	0.480	0.240	0.040	0.240	0.000	0.000	(Mata-Miguez et al. 2012)
	Taino (Dominican Rep.)	19	0.000	0.000	0.790	0.210	0.000	0.000	(Lalueza-Fox et al. 2001)
	Ciboney (Cuba)	15	0.070	0.000	0.600	0.330	0.000	0.000	(Lalueza-Fox et al. 2003)
	Paleo-Colombian Chocua (Cundinamarca)	12	0.583	0.250	0.083	0.000	0.083	0.000	(This Study)
	Guane (Santander, Colombia)	17	0.350	0.410	0.000	0.240	0.000	0.000	(Casas-Vargas et al. 2011)
	Paracas Palpa (Perú)	28	0.070	0.000	0.140	0.790	0.000	0.000	(Fehren-Schmitz et al. 2010)
	O-Humahuaca, Jujuy (Argentina)	32	0.469	0.250	0.125	0.156	0.000	0.000	(Mendisco et al. 2014)
MODERN PEOPLE	Maya (Mesoamerica)	27	0.520	0.220	0.150	0.070	0.000	0.040	(Melton et al. 2007)
	O-Huetar (Costa Rica)	27	0.700	0.040	0.000	0.260	0.000	0.000	(Melton et al. 2013)
	Emberá (Panamá)	44	0.230	0.520	0.250	0.000	0.000	0.000	(Kolman & Bermingham 1997)
	Arsario (Caribbean Colombia)	28	0.710	0.000	0.290	0.000	0.000	0.000	(Melton et al. 2007)
	Ijka (Caribbean Colombia)	31	0.900	0.030	0.060	0.000	0.000	0.000	(Melton et al. 2007)
	Kogi (Caribbean Colombia)	21	0.810	0.000	0.190	0.000	0.000	0.000	(Melton et al. 2007)
	Wayuú (Caribbean Colombia)	30	0.270	0.270	0.430	0.000	0.000	0.000	(Melton et al. 2007)
	Mestizo Bogotá, DC (Colombia)	25	0.360	0.440	0.080	0.080	0.000	0.040	(Díaz-Matallana & Martínez-Cruzado 2010)
	Cayapa (Ecuador)	120	0.290	0.400	0.090	0.000	0.220	0.000	(Rickards et al. 1999)
	Mestizo Jujuy Province (Argentina)	177	0.139	0.567	0.178	0.089	0.011	0.000	(Cardoso et al. 2013)

Supplemental Table 6. Results PCA.

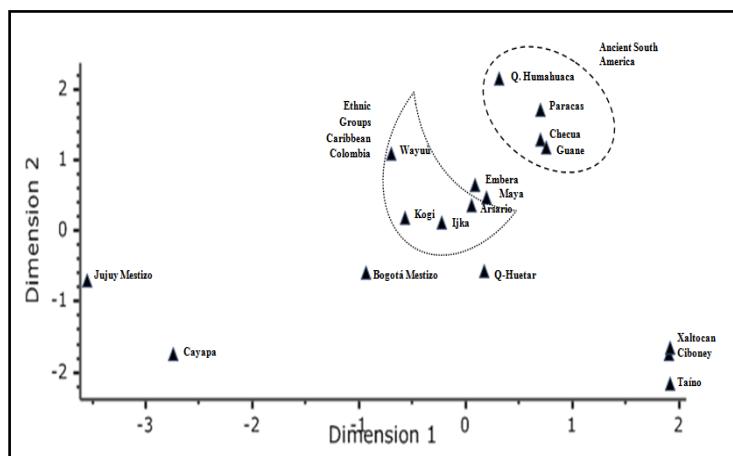
Principal Component Analysis				
Eigenvalues				Cumulative
Var A (Size)				36.053
Var B (A2)				64.456
Var C (B2)				80.405
Var D (C1)				93.063
Var E (D1)				99.987
Var F (D4h34)				100.000
Eigenvectors				
C (B2)		D (C1)	E (D1)	F (D4h34)
-0.037		0.031	0.803	-0.002
0.053		0.072	0.261	0.598
-0.129		-0.535	-0.401	0.455
0.681		0.016	0.074	0.480
-0.711		0.231	0.072	0.438
0.101		0.809	-0.342	0.114
Factor loadings				
C (B2)		D (C1)	E (D1)	F (D4h34)
-0.037		0.027	0.517	0.000
0.052		0.063	0.168	0.017
-0.126		-0.466	-0.258	0.013
0.666		0.014	0.048	0.014
-0.695		0.202	0.046	0.012
0.099		0.705	-0.220	0.003
Z Scores				
C (B2)		D (C1)	E (D1)	F (D4h34)
-0.881		-0.232	-0.171	0.019
1.652		0.297	-0.021	0.010
0.647		0.437	-0.050	0.005
0.279		0.691	-0.916	0.012
-1.135		-0.730	-0.800	0.029
-2.440		0.942	0.201	-0.014
-0.331		-0.347	-0.071	0.023
0.085		-0.355	-0.078	
-0.908		0.382	0.472	
0.383		-1.294	-0.630	
0.952		0.214	0.587	
0.243		0.170	0.688	
0.663		0.228	0.518	
1.131		-0.613	-0.292	
-0.341		-0.981	-0.749	
0.303		2.281	-0.310	
-0.303		-1.090	1.623	



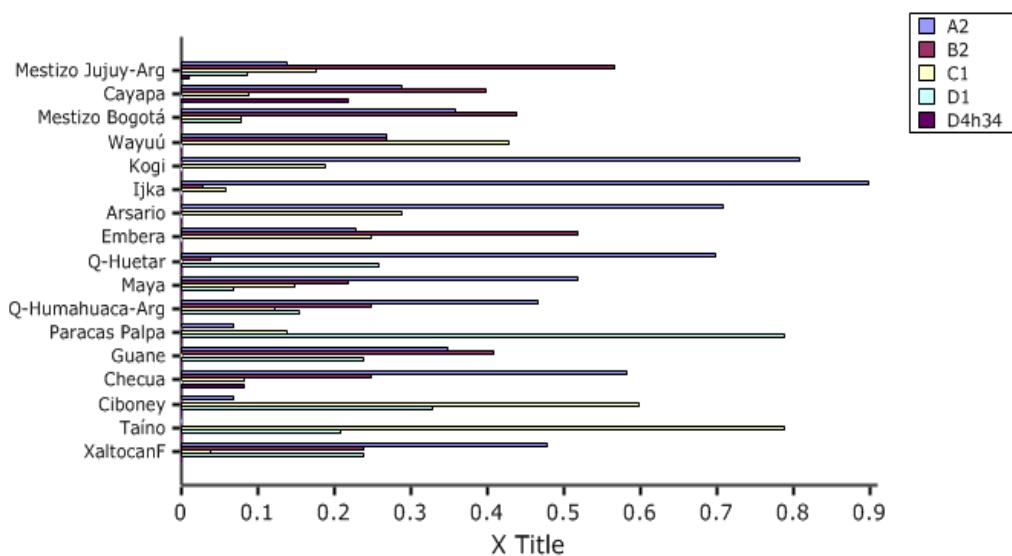
Supplemental Figure 5. PCA biplot: Scores and coefficients.

Supplemental Table 7. Scores PCA

No.	Population / Group	▲ Scores
1	Mestizo Jujuy	-0.694
2	Cayapa	-1.75
3	Mestizo Bogotá	-0.595
4	Wayuú	1.081
5	Kogi	0.181
6	Ijka	0.115
7	Arsario	0.358
8	Emberá	0.652
9	Q-Huetar	-0.575
10	Maya	0.477
11	Q. Humahuaca, Jujuy	2.147
12	Paracas Palpa	1.709
13	Guane	1.175
14	Paleo-Colombian Checua	1.286
15	Ciboney	-1.749
16	Taíno	-2.165
17	Xaltocan (full)	-1.653



Supplemental Figure 6. Final PCA biplot.



Supplemental Figure 7. Comparison of HVR-I frequencies of Amerindian haplogroups in Ancient and Modern Groups.