10-9-2020

Continental Origin for Q Haplogroup Patrilineages in Argentina and Paraguay

Laura S. Jurado Medina
CONICET-CIC-Universidad Nacional de La Plata

Paula B. Paz Sepúlveda
CONICET-CIC-Universidad Nacional de La Plata

Virginia Ramallo
Centro Nacional Patagónico, CONICET

Camila Sala
CONICET-CIC-Universidad Nacional de La Plata

Julieta Beltramo
CONICET-CIC-Universidad Nacional de La Plata

See next page for additional authors

Continental Origin for Q Haplogroup Patrilineages in Argentina and Paraguay

Laura S. Jurado Medina,[1] Paula B Paz Sepúlveda,[1] Virginia Ramallo,[3] Camila Sala,[1] Julieta Beltramo,[1,4] Marisol Schwab,[1] Josefina M B Motti,[5] María Rita Santos,[1] Mariela V. Cuello,[1] Susana Salceda,[6] José E Dipierri,[7] Emma L Alfaro Gómez,[7,8] Marina Muzzio,[1,2] Claudio M Bravi,[1,2] and Graciela Bailliet[1]*

[1]Instituto Multidisciplinario de Biología Celular, CONICET-CIC-Universidad Nacional de La Plata, La Plata, Argentina.

[2]Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, La Plata, Argentina.

[3]Centro Nacional Patagónico, CONICET, Puerto Madryn, Argentina.

[4]Laboratorio de Análisis Comparativo de ADN, Suprema Corte de Justicia de la Provincia de Buenos Aires, La Plata, Argentina.

[5]Laboratorio de Ecología Evolutiva Humana, NEIPHPA-FACSO, Universidad Nacional del Centro de la Provincia de Buenos Aires, Quequén, Argentina.

[6]División Antropología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, La Plata, Argentina.

[7]Instituto de Biología de la Altura, Facultad de Humanidades y Ciencias Sociales, Universidad Nacional de Jujuy, San Salvador de Jujuy, Argentina.

[8]Instituto de Ecorregiones Andinas, Universidad Nacional de Jujuy, San Salvador de Jujuy, Argentina.

*Correspondence to: Graciela Bailliet, Instituto Multidisciplinario de Biología Celular (IMBICE) CCT-CONICET-La Plata, CIC, Universidad Nacional de La Plata, Calle 526 Y

Camino General Belgrano. B1906APO La Plata, Buenos Aires, Argentina. E-mail: graciela.bailliet@gmail.com.

Short Title: Haplogroup-Q

KEY WORDS: Y CHROMOSOME, HAPLOGROUPS, HAPLOTYPES, SEQUENCING, SOUTH AMERICA.

Abstract

Haplogroup Q originated in Eurasia around 30,000 years ago. It is present in Y-chromosomes from Asia and Europe at rather low frequencies. Since America is undoubtedly one of the continents where this haplogroup is highly represented, it has been defined as one of the founding haplogroups. Its M3 clade has been early described as the most frequent, with Pan-American representation. However, it was also possible to find several other haplogroup Q clades at low frequencies. Numerous mutations have been described for haplogroup Q, allowing the analysis of its variability and the assignment of its geographic origin. We have analyzed 442 samples belonging to haplogroup Q of unrelated men from Argentina and Paraguay, but this work is specifically referred to 27 Q (xM3) lineages. We tested 3 SNPs by APLP, 3 for RFLP, 15 SNPs by Sanger sequencing, and 17 STRs. Our approach allowed us to identify 5 sub-haplogroups. Q-M3 and Q-CTS2730/Z780 are undoubtedly autochthonous lineages and represent the most frequent sub-haplogroups. With significant representation in self-defined aboriginal populations, their autochthonous status has been previously described. The aim of present work is to identify the continental origin of the remaining Q lineages. Thus, we analyzed the STR haplotypes for the samples of our series and compared them with haplotypes described by other authors for the rest of the world. Even when haplogroup Qs

have been extensively studied in America, some of them could have their origin in post Columbian human migration from Europe and Middle East.

Haplogroup Q lineages have a very wide geographic distribution. Q1 has been described in Siberia (Malyarchuk et al. 2011, Regueiro et al. 2013,), Lebanon, Syria, Jordan, Iran, and Kuwait (El-Sibaid et al. 2009), Pakistan and Afghanistan (Di Cristoforo et al. 2013). One of its sub-haplogroups, Q1a1-NWT01 described in Canadian Eskimo and Athapaskan speaking populations, Tłicho, Inuvialuit and Inupiat populations (Dulik et al. 2012). Q-M120 was described in Kungurtug from Tuva Republic, (Regueiro et al. 2013) and Mongolia (Battaglia et al. 2013), whereas Q-M25 was reported in Mongolia (Battaglia et al. 2013), in Kalamyks from East Europe (Malyarchuk et al. 2011), in Iran, and in Afghanistan (Di Cristoforo et al. 2013). Sub-haplogroup Q1b-M346) was described in Pakistan (Geppert et al. 2015), in South Siberia and East Europe (Malyarchuk et al. 2011), in Afghanistan, Pakistan, Iran, Kyrgyzstan, and Mongolia (Di Cristoforo et al. 2013). Sub-clades from Q1b, such as Q-L54, were present in Tuva Republic, Northeast Siberia and Mexico (Rigueiro et al. 2013), Canada (Dulik et al. 2012), America (Battaglia et al. 2013), and Kalamy (Geppert et al. 2015), while Q-L330 (Q1b1a3) was present in Mongolia (Battaglia et al. 2013). Haplogroup Q-M3 has been early described as a founder Y-chromosome lineage in America (Avena et al. 2009, Bisso-Machado et al. 2011, Bortolini et al. 2003, Geppert et al. 2011, Sala & Corach 2014, Seielstad et al. 2003, Vullo et al. 2015). Q-M3 is a very diverse haplogroup with many derived lineages: Q-M194, Q-SA01, Q-M557, Q-PV2, among many others (Battaglia et al. 2013, Dulik et al. 2012, Geppert et al. 2015, Jota et al. 2011, Jota et al. 2016, Regueiro et al. 2013). Recent publications describing Q-haplogroup by NGS have provided the complete scenario of Q differentiation in America (Grugni et al. 2019, Pinotti et al. 2019); we will further describe some specific details. The other Native American lineage derived from Q1b is Q-Z780 or Q-CTS2730 (present study), and its derivate lineage Q-L191 was also described in America (Geppert et al. 2015). Lineage Q-M378 (Q2a1) was only present in Panama, and in the North of South America (Battaglia et al. 2013) (Table S1).

Moreover, Q sub-clades have been identified in ancient human remains (Kivisild, 2017). The Q1a1b- Q-B143 sub-clade was described in an ancient human genome from Saqqaq culture in Greenland, 4,000 years ago (ya) (Grugni et al., 2019, Rasmussen et al., 2010) and its genomic variation showed biological affinity with Old World Arctic populations like Koryaks and Chukchis from Northeast Siberia (Raghavan et al., 2015, Rasmussen et al., 2010). The 10,300 year old On Your Knees Cave Man (OYKCM) from Alaska belonged to Q-M3 haplogroup (Kemp et al., 2007) and the Kennewick Man, dated between 8,340–9,200 ya, was also associated to the Q-M3 lineage (Rasmussen et al., 2015), whereas the Anzick Boy (12,707-12,556 ya), associated to Clovis Culture, showed the Q1b1a2-Q-Z780 sub-haplogroup (Grugni et al., 2019, Rasmussen et al., 2014). Combined ancient and modern DNA analysis allowed the inference of genetic changes during the last 11,000 years; showing that the Q1b1a2-Q-Z780 haplogroup, which is currently rare, was present in about a third of the ancient South Americans (Posth et.al., 2018).

In Argentina, 10 national censuses between 1869 and 2010 show high migration dynamics, where two main origins of migrants can be distinguished: overseas, mostly from Southern Europe, and Latin America, mainly from neighboring countries. Most of the European migration took place between 1880 and 1914, and after the First World War it began to decline dramatically (Lattes 1985). Migrants were mainly Italian and Spanish, estimated by 2001 in 351,135 immigrants out of the total population, 60% of which were older than 65 (INDEC, 2010). Arab migration reached its maximum between 1895 and 1947 (174,000 immigrants), most of them from present-day territories of Syria and Lebanon, and a minority from Palestine. Overseas migrants settled mainly in the Pampa region and littoral areas, and only 2,18% of them in the Northwest of Argentina. Arab migration was more evenly distributed, 28% of them established in Northwest Argentina (Liberali 2007). In some provinces such as La Rioja, the Syrian-Lebanese community exceeded the Italian and

Spanish (De Luca 2006). Considering the low demographic density of that region, the influence of the Arab community has been quite significant.

Latin-American immigration came from neighboring countries: Bolivia, Brazil, Chile, Paraguay, and Uruguay. As shown by national censuses, their flow modified the configuration of Argentina. Thus, out of 1,5 million immigrants by 2001, 67% had come from Latin-America, most of them from neighboring countries, 28.2% from Europe, and 4.8% from Asia and Middle East (Pacecca 2008). The 1947 census reported a total of 2,4 million immigrants, out of which 8.6% had come from neighboring countries and 91% from countries not bordering Argentina (INDEC 2010).

Historical evidence shows that the massive overseas migration concentrated in the Central East (Pampas region) and Northeast. This pattern was clearly determined by genetic analyses in native and "criollos" populations (Demarchi & Mitchell 2004, Dipierri et al. 1998) and admixture analyses (Avena et al. 2012, Fejerman et al. 2005, Martínez-Marignac et al. 2004, Muzzio et al. 2018, Sala et al. 1998, Wang et al. 2008).

The aim of this work is to determine the continental origin of Q (xM3) lineages present in Argentina through a worldwide Q haplotype comparison. Thus, we considered aboriginal individuals from Gran Chaco Region (including two populations from Paraguay), and the rest of the samples from urban populations, mainly provincial capitals. These populations are admixed, and most individuals could not estimate their ancestry farther than three generations back, either aboriginal or European. In this way we can determine which lineages are Native American and which ones come from Europe or the Middle East, and contribute to a better understanding of the various American population processes, from the first settlers to recent migrations.

Materials and Methods

Biological Samples

We analyzed 442 biological samples of voluntary male donors from 22 locations in Argentina and Paraguay, all of them previously assigned to haplogroup Q M242 (Table 1). Self-ascribed Native American individuals were present in Gran Chaco Ecoregion: in Argentina, Wichi and Toba from Laguna Yema and Ingeniero Juárez, Formosa, and Wichi, Toba and Chorote from Santa Victoria Este, Salta; and in Paraguay, Lengua and Ayoreo. Gran Chaco population has been widely studied from anthropological, archaeological and ethnohistorical perspectives (Braunstein 2005, Calandra & Salceda 2007, Demarchi et al. 2001, 2009, Susnik 1972).

This region shows a complex linguistic pattern, with about 18 to 20 different languages (Braunstein & Miller 1999, Fabre 2007, Unicef 2009). Some of these are rarely spoken and could be soon replaced by hegemonic languages or language admixture. In several cases, contact and spreading of languages have been described. This is coherent with the traditional social organization of Gran Chaco populations, characterized by units progressively inclusive and familiar alliances that advance in a territory.

The remaining samples represent urban populations of cities from different provinces, from hospitals and blood-banks, through the volunteer participation of donors.

All of our sampling and recruiting protocols, as well as this project, have been approved by Argentine Ethics Committees. Donors provided informed consent and completed a questionnaire about their ancestry and birth place. Samples were coded and submitted in anonymity for DNA testing.

Genotyping

Twenty one SNPs were analyzed through three different techniques (Table S2): RFLP (Bailliet et al. 2009) (Table S3), APLP (Jurado Medina et al. 2014) , and Sanger sequencing

of 400 to 800 bp fragments with known flanking and previously described SNPs (Battaglia et al. 2013, Y Chromosome Consortium 2002, Karafet et al. 2008, Shen et al. 2004, Underhill et al. 2001) (https://www.familytreedna.com/groups/y-dna-q) (Table S4). Due to sequencing difficulties, we replaced Z780 by CTS2730A since both determine the same phylogenetic status.

We analyzed 17 microsatellites (AmpFLSTR® Yfiler® Life Technologies) using Genescan v. 3.7 (Applied Biosystems). Electrophoresis was performed in an ABBI 3130 analyzer. Our Laboratory has been granted the YHRD Quality Test Certificate for Yfiler analysis.

We built median-joining networks (Bandelt, Forster & Röhl 1999) using the NETWORK software (fluxus-engineering.com) with weights assigned following Muzzio et al. (2010). We used Rho dating methods (Forster et al. 1996, Saillard et al. 2000), considering the rate based on population comparison 6.9 104/ STR/generation (Zhivotovsky et al. 2004).

Due to the different number of STRs analyzed in each work, networks showed different STR number: Figure 2: Q1(xQ1a.Q1b) sub-haplogroup median-joining network , we considered 10 STRs (DYS19, DYS389ab, DYS390, DYS392, DYS393, DYS437, DYS438, DYS439); Figure 3: Q1a2-M25 sub-haplogroup median-joining network, 16 STRs (DYS19, DYS385ab, DYS389ab, YS390, DYS391, DYS392, DTS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYSYGATA H4) ; Figure 4: Q1b-M346 sub-haplogroup median-joining network, 10 STRs (DYS19, DYS389ab, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439) ; Figure 5: Q1b1a2-CTS2730 and haplotypes Q-L54 from bibliography median-joining network, 14 STRs (DYS19, DYS390, DYS391, DYS392, DYS393, GATAH4, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458).

The geographic distribution of different sub-haplogroups was analyzed by building isofrequency graphs with Surfer12 software, applying the Kriging algorithm each time that the number of samples was higher than 20 (Golden Software 2002).

Results

Out of 442 samples from Q-M242 haplogroup, 3/442 were Q-P36.2*, 1/442 Q-M25, 2/442 Q-M346*, 21/442 Q-CTS2730*, and 415/442 Q-M3. Their phylogenetic status is summarized in Figure 1 and their geographic distribution is shown in Table S1.

Three samples showing a derived state from M242 and P36.2, and ancestral state for F1096, M346, L565, and P89.1 were assigned to Q1* (Q-P36.2*). Since we considered that P36.2 was the marker allowing the identification of Q1* (Jota et al. 2011), we did not analyze M275 which determines Q2 (ISOGG 2019-20, http://www.isogg.org/ tree/ISOGG_YDNA_SNP_Index.html).

Q-P36.2 came from 3 different Argentine provinces: Mendoza (haplotype 1), Santiago del Estero (haplotype 2), and La Rioja (haplotype 3) (Tables S1 and S5). Figure 2 shows a median network of 25 Q1* haplotypes based on 10 STRs from Middle East, Tuva Republic, Canada, and Siberia (haplotypes 4 to 28, Table S5)(Dulik et al. 2012, El-Sibai et al. 2009, Malyarchuk et al. 2011, Regueiro et al. 2013, Zalloua et al. 2008) (Table S5)

We found one sample from Catamarca, Argentina, with Q1a2 (Q-M25), showing derived state for M242, P36.2, F1096, and M25 (haplotype 29) (Tables S1 and S5). Figure 3 shows the median-joining network of this sub-haplogroup including 27 haplotypes based on 16 STRs from Afghanistan, Iran, and China (haplotypes 30 to 56, Table S5) (Di Cristofaro et al. 2013) (Table S5).

In two samples from San Juan, Argentina, we found Q1b*- (Q-M346*), determined by derived state for M346 but ancestral state for L940, L213, and M323 (haplotypes 57 and

58, Tables S1 and S5). Figure 4 shows the median-joining network including 10 STR haplotypes from 163 lineages from Peru and Bolivia, South America; California, USA; Kyrgyz Republic; Mongolia; Afghanistan; Iran; Europe; Asia; South Siberia; China; and Africa (haplotypes 59 to 219, Table S5) (Di Cristofaro et al. 2013, Dulik et al. 2012, Lacau et al. 2012, Liu et al. 2014, Malyarchuk et al. 2011, Msaidie et al. 2011, Sandoval et al. 2013) (Table S5). We included 24 M346 lineages from Sandoval et al. (2013); they are relevant because of their geographic location, close to Argentina. We considered that haplotypes were Q-M346 (xL54). There is no evidence of L54 analyses for these samples in the original publication.

Twenty-one Q1b1a2-(Q-CTS2730) showed derived state for M346, L213, L53, L54, and CTS2730 (Z780), ancestral state for CTS11969, L191, L456, L718 and L565, and wide geographic distribution (Table S1). Figure 5 shows a median-joining network based on 14 STRs including 19 haplotypes from Argentina and Paraguay (haplotypes 165 to 183, Table S5), and three from Canada (haplotypes 239 to 242) (Dulik et al. 2012) (Table S5), and Q-L54 (x Q-M3), from a wide geographic distribution: Tuva Republic, Northeast Siberia, Mexico, Bolivia, Colombia, Peru , Nicaragua, Panama , and Paraguay (Battaglia et al. 2013, Dulik et al. 2012, Regueiro et al. 2013) (Table S5).

Figure 5. Q-CTS2730 and haplotypes Q-L54 from bibliography median-joining network.

Figures S1 and S2 show the isofrequency graphs for Q1a2a1b-CTS2730 and Q1a2a1a1-M3 respectively, which were the only haplogroups with N higher than 20.

Discussion

Our approach allowed us to determine that most samples from Q (xM3) were in fact Q1a2a1b (Q-CTS2730). In our series of samples, it was the second most frequent sub-haplogroup after Q-M3. The origin of the rest of Q sub-haplogroups seems to be controversial; however,

lineage STR networks all over the world provide evidence of relatedness with European and Middle Eastern lineages.

We included Q1a lineages from Northeast Siberia (Regueiro et al. 2013), Tlingit from Canada (Dulik et al. 2012, Schurr et al. 2012), Siria, Jordan and Iran (El-Sibai et al. 2009), Koryaks from Kamchatka (Malyarchuk et al. 2011), and Lebanon (Zalloua et al. 2008). The median-joining network built with 10 Y-STRs showed that the Argentine lineages from Villa Tulumaya (haplotype 1), Santiago del Estero (haplotype 2), and Chepes (haplotype 3) clustered together with Syria, Jordan, Iran, and Lebanon haplotypes. The Villa Tulumaya haplotype was identical to two Jordan haplotypes (Figure 2). The network showed two groups of samples, one corresponding to Middle East and our samples, and another one to samples from Russia. Between these two groups we found the Tlingit sample from Canada, closer to the Middle East than the Russian group. The TMRCA calculated for the complete network through Rho was 10,612.14 ya (SD 1,998.21). There are few estimations of TMRCA for Q1a and dates referring to sub-haplogroup Q-M120 available are older than our estimation (Huang et al. 2018); since ours was calculated through the analysis of 10 STRs in 28 samples, the low number of haplotypes might have caused underestimation.

The very restricted geographic distribution and the clustering with Middle East haplotypes in the median-joining network evidence the allochthonous origin of these lineages. Their geographic distribution in Argentina (from Mendoza, La Rioja, and Santiago del Estero provinces) is consistent with a recent genetic flow from Middle East (Figure 2) (1914 Argentine population census, http://www.indec.gov.ar/bicentenario /pdf/1914.pdf). In the case of Chepes, many donors declared Arab ancestors from Turkey, Syria, Lebanon, among others.

Q1a1b-M25 has been previously described at high frequencies in Turkmen (31%) (Di Cristofaro et al. 2013), but is less common in Mongolia (Battaglia et al. 2013), Kalmyk from

Eastern Europe (Malyarchuk et al. 2011), Iran (Sengupta et al. 2006) (haplotypes not available), and Anatolia (Cinnioğlu et al. 2004). It was not found in other population analyses (Dulik et al. 2012, Regueiro et al. 2013). Only one individual from Catamarca, Northwest Argentina, carried this haplogroup (haplotype 29). The median-joining network based on 16 STR haplotypes showed most of the haplotypes organized into one cluster and the Catamarca haplotype very distant from it (Figure 3). Rho value for the distance between haplotype 31 and haplotype 29 was 6,762.22 ya (SD1017,19), similar to the complete network value of 7,160 ya (SD1081,05). Interestingly, these values are higher than those calculated by Zhong et al. (2011). Our results for this group are inconclusive due to the absence of other haplotypes from different geographical areas to compare with. Nonetheless, the strongest hypothesis suggests a foreign origin for this sample because Catamarca received 9% of overseas migration between 1921 and 1924; it is possible that this lineage derived from Middle East, from where 174,000 migrants arrived (INDEC 2010, Lattes 1985).

Q1a2-M346 was detected in Chumash from North America (Dulik et al. 2012), in Bolivia and Peru from South America (Sandoval et al. 2013), China (Liu et al. 2014), Northwest Asia, Afghanistan, Iran, Pakistan, Mongolia, South and West Siberia, Europe, Africa, and America (Di Cristofaro et al. 2013, Lacau et al. 2012, Malyarchuk et al. 2011, Msaidie et al. 2011, Sandoval et al. 2013). The median-joining network (Figure 4) showed a large cluster highly represented by Siberian haplotypes; Middle East, America, and Europe haplotypes radiated from this Siberian cluster. Two individuals from San Juan, Argentina, presented this haplogroup, one of them (haplotype 57) connected with Siberian haplotypes through a series of median vectors and very close to other Native American lineages from Peru and Bolivia included in the network (Sandoval et al. 2013). Haplotype 58 was at the end of a branch of Middle East and Asia lineages that initiated in the Siberian cluster. Our NGS results for this haplotype (data not shown) reveal derived state for Z5902, Z36057,

Z36058, Z36059, but ancestral state for Y6794 and L717. This lineage origin is closer to South Asia (Grugni et al. 2019). The complete network showed a TMRCA of 17,005 ya (SD 1,790). TMRCA from the center of the Siberian cluster (haplotype 63) to haplotype 58 was 550.77 ya (SD 91.79). From the same center to haplotype 57, TMRCA value was 716 ya (SD 206.69). Date estimates from different authors and methods show a great diversity in this sub-haplogroup (10,000 ya to 2,000 ya).

Though both San Juan (Argentina) haplotypes are Q-M346, they belong to different migration origins. Both donors declared in the questionnaire that the origin of their paternal lineage was local and that they did not remember their family history further generations ago. Lineage 58 probably belongs to a post Columbian migration, since an internal migration could have led this individual from Tucuman, Santiago del Estero, Jujuy, or La Rioja where Middle East migration was more significant (INDEC, 2010), to San Juan, in Central-West Argentina. Lineage 57 should be further analyzed in order to determine its origin. It was close to Siberian and other Native American haplotypes from Peru and Bolivia (Sandoval et al. 2013). Thus, we propose here a Native American status for this Q-M346 lineage. As detailed in Material and Methods, some of Q-M346 from Sandoval et al. (2013) should be Q-L54. Analyses in our network showed short distance of some of them to our sample 57, what should indicate that they probably are Q-M346 (xL54). If this were the case, Q-M346 could be a new founding lineage. This should be finally solved by further analyses.

Q-CTS2730 (Q1a2a1b) represented the second most frequent native lineage in Argentina and was the only Q (xM3) lineage present in Ayoreo and Lengua populations from Paraguay, though it was not present in any of Wichi, Toba and Chorote individuals from Formosa and Salta, the other group of self- identified Native Americans analyzed in this series of samples. We included in the network 3 haplotypes from Q-L54 sub-haplogroup (xM3, L191, L330, L334, L400, L401, L456) from North American Gwich'in samples (Dulik

et al. 2012), and 120 Q-L54 haplotypes (Battaglia et al. 2013, Regueiro et al. 2013). Figure 5 is a big network with a medial axis that connected all the haplotypes in different clusters. Our series of lineages (220-238, black) and those from Dulik et al. (2012) (239-242, yellow) are closely related to this axis, and connected to it by medial vectors. Haplotypes from our series of samples are distributed in the whole network. Though, some haplotypes from Battaglia et al. (2013) and Regueiro et al. (2013) are grouped into a cluster, what should represent more than one sub-haplogroup. Previous reports provided evidence of the antiquity of lineages Q-Z780 and described their high variability. In 2016, three sub-lineages were described: Q-SA03 and QSa02 in Colombian individuals, and Q-SA29 from southeast Brazil (Jota et al. 2016). In relation with NGS data, the coalescence time 17.0 [15.0–19.3] kya for Q-CTS1780/CTS2730/Z78 results quite striking (Pinotti et al. 2019), since this value is very close to our coalescence time obtained by Rho estimation: 16.543.94 ya (SD 2,090.8). Q-CTS2730 haplogroup seems to be an ancient lineage that entered the Americas as early as Q-M3 did.

These isofrequency graphs from Q- Q1a2a1b-CTS2730 (Figure S1) showed uniform distribution all over the area analyzed, with the maximum frequency for Paraguay (Ayoreo and Lengua populations) and Calingasta (San Juan, Argentina). Calingasta is a very small city in a valley surrounded by mountains, in which Motti et al. (2013) found high proportion of Native American ancestry and very low frequency of European lineages.

Q-M3 is the most frequent lineage in Native American populations. In Argentina, it shows a North-South decreasing frequency gradient for Northern aboriginal populations (Figure S2). The high number of samples in our collection allowed us to detect even those lineages with less than 1% frequency. This finding required individual STR analyses in order to identify the geographic origin of these samples, being 17 STRs enough for our goal. We

were able to identify lineages with an allochthonous origin; however, the fine-scale origin of some samples remains unassigned.

Conclusion

We determined Q sub-haplogroups of allochthonous origin and presented a network including similar lineages from the rest of the world. In most of the cases, lineages from Argentina clustered together with those from Middle East. We identified 2 sub-haplogroups autochthonous for America: Q-M3, and Q-CTS2730, one lineage would be also autochthonous, but needs further analyzes (Q-M346*). While the presence of other Q lineages in Argentina seems to be due to recent extra-continental migrations. The detailed analysis of Q lineages results crucial before assigning their continental origin, thus avoiding an over-estimation of Native American lineages and allowing us to see the complex migration history of this continent.

Data Availability Statement

After this work is accepted for publication, data will be available at the Repository of the National University of La Plata, Argentina: http://sedici.unlp.edu.ar/

Acknowledgments

We thank all DNA voluntary donors for making this work possible. We specially thank Dario A. Demarchi for critical review of the manuscript. Laura S Jurado Medina was a doctoral fellow of Consejo Nacional de Investigaciones Científicas y Técnicas de la Republica Argentina (CONICET). This investigation was funded by ANPCyT through grant PICT 2013 425.

Received 8 January 2020; accepted for publication 20 August 2020.

Literature Cited

Avena, S. A., M. L. Parolin, C. B. Dejean et al. 2009. Mezcla génica y linajes uniparentales en Comodoro Rivadavia (provincia de Chubut, Argentina). *Rev. Argent. Antropol. Biol.* 11:25–41.

Avena, S., M. Via, E. Ziv et al. 2012. Heterogeneity in genetic admixture across different regions of Argentina. *PLoS One* 7:e34695.

Bailliet, G., V. Ramallo, M. Muzzio et al. 2009. Brief communication: Restricted geographic distribution for Y-Q* paragroup in South America. *Am. J. Phys. Anthropol.* 140:578–582.

Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37–48.

Battaglia, V., V. Grugni, U. A. Perego et al. 2013. The first peopling of South America: New evidence from Y-chromosome haplogroup Q. *PLoS One* 8:e71390.

Bisso-Machado, R., M. S. Jota, V. Ramallo et al. 2011. Distribution of Y-chromosome Q lineages in Native Americans. *Am. J. Hum. Biol.* 23:563–566.

Bortolini, M. C., F. M. Salzano, M. G. Thomas et al. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am. J. Hum. Genet.* 73:524–539.

Braunstein, J. 2005. Los pueblos indígenas del Gran Chaco. *Mundo de Antes* 4:127–137.

Braunstein, J., and E. Miller. 1999. Ethnohistorical introduction. In *Peoples of the Gran Chaco*, E. S. Miller, ed. Westport, CT: Bergin & Garvey, 1–22.

Calandra, H., and S. Salceda. 2004. El territorio y sus ocupantes: ¿Qué, quienes, cómo y cuándo? *Folia Hist. Nordeste* 15:107–128.

Cinnioğlu, C., R. King, T. Kivisild et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum. Genet.* 114:127–148.

De Luca, J. 2006. *La Inmigración Sirio-Libanesa en la Argentina.* Seminario de Inmigración/Emigración. Universidad Nacional de Buenos Aires.

Demarchi, D. A. 2009. Microsatélites, distancias genéticas y estructura de poblaciones nativas sudamericanas. *Rev. Argent. Antropol. Biol.* 11:73–88.

Demarchi, D. A., and R. J. Mitchell. 2004. Genetic structure and gene flow in Gran Chaco populations of Argentina: Evidence from Y-chromosome markers. *Hum. Biol.* 76:413–429.

Demarchi, D. A., G. M. Panzetta-Dutari, C. C. Motran et al. 2001. Mitochondrial DNA haplogroups in Amerindian populations from the Gran Chaco. *Am. J. Phys. Anthropol.* 115:199–203.

Di Cristofaro, J., E. Pennarun, S. Mazières et al. 2013. Afghan Hindu Kush: Where Eurasian sub-continent gene flows converge. *PLoS One* 8:e76748.

Dipierri, J. E., E. Alfaro, V. L. Martinez-Marignac et al. 1998. Paternal directional mating in two Amerindian sub-populations located at different altitudes in northwestern Argentina. *Hum. Biol.* 70:1,001–1,010.

Dulik, M. C., A. C. Owings, J. B. Gaieski et al. 2012. Y-chromosome analysis reveals genetic divergence and new founding Native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc. Natl. Acad. Sci. U. S. A.* 109:8,471–8,476.

El-Sibai, M., D. E. Platt, M. Haber et al. 2009. Geographical structure of the Y-chromosomal genetic landscape of the Levant: A coastal-inland contrast. *Ann. Hum. Genet.* 73:568–581.

Fabre, A. 2007. Morfosintaxis de los clasificadores posesivos en las lenguas del Gran Chaco (Argentina, Bolivia y Paraguay). *UniverSOS* 4:67–85.

Fejerman, L., F. R. Carnese, A. S. Goicoechea et al. 2005. African ancestry of the population of Buenos Aires. *Am. J. Phys. Anthropol.* 128:164–170.

Forster, P., R. Harding, A. Torroni et al. 1996. Origin and evolution of Native American mtDNA variation: A reappraisal. *Am. J. Hum. Genet.* 59:935–945.

Geppert, M., Q. Ayub, Y. Xue et al. 2015. Identification of new SNPs in Native South American populations by resequencing the Y chromosome. *Forensic Sci. Int. Genet. Suppl. Ser.* 15:111–114.

Geppert, M., M. Baeta, C. Núñez et al. 2011. Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of Native populations in South America. *Forensic Sci. Int. Genet. Suppl. Ser.* 5:100–104.

Grugni, V., A. Raveane, L. Ongaro et al. 2019. Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. *BMC Biol.* 17:1–14.

Huang, Y. Z., H. Pamjav, P. Flegontov et al. 2018. Dispersals of the Siberian Y-chromosome haplogroup Q in Eurasia. *Mol. Genet. Genomics* 293:107–117.

Instituto Nacional de Estadística y Censos (INDEC). 2010. Censo del bicentenario resultados definitivos. https://www.indec.gob.ar/indec/web/Nivel4-Tema-2-41-135.

Jota, M. S., D. R. Lacerda, J. R. Sandoval et al. 2011. A new subhaplogroup of Native American Y-chromosomes from the Andes. *Am. J. Phys. Anthropol.* 146:553–559.

Jota, M. S., D. R. Lacerda, J. R. Sandoval et al. 2016. New Native South American Y chromosome lineages. *J. Hum. Genet.* 61:1–11.

Jurado Medina, L. S., M. Muzzio, M. Schwab et al. 2014. Human Y-chromosome SNP characterization by multiplex amplified product-length polymorphism analysis. *Electrophoresis* 35:2,524–2,527.

Karafet, T. M., F. L. Mendez, M. B. Meilerman et al. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18:830–838.

Kemp, B. M., R. S. Malhi, J. McDonough et al. 2007. Genetic analysis of early Holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *Am. J. Phys. Anthropol.* 132:605–621.

Kivisild, T. 2017. The study of human Y chromosome variation through ancient DNA. *Hum. Genet.* 136:529–546.

Lacau, H., T. Gayden, M. Regueiro et al. 2012. Afghanistan from a Y-chromosome perspective. *Eur. J. Hum. Genet.* 20:1,063–1,070.

Lattes, A. E. 1985. *Migraciones hacia America Latina y el Caribe desde Principios del Siglo XX*. Cuaderno del CENEP No 35. Buenos Aires, Argentina: Buenos Aires Centro de Estudios de Población.

Liberali, A. 2007. Cultura árabe en la provincia de Salta – Argentina. *Espacio y Desarrollo* 19:179–188.

Liu, Y., L. Liao, M. Gu et al. 2014. Population genetics for 17 Y-STR loci in a Chinese Han population sample from Mudanjiang city, Northeast China. *Forensic Sci. Int. Genet. Suppl. Ser.* 13:E16–E17.

Malyarchuk, B., M. Derenko, G. Denisova et al. 2011. Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *J. Hum. Genet.* 56:583–588.

Martínez-Marignac, V. L., B. Bertoni, E. J. Parra et al. 2004. Characterization of admixture in an urban sample from Buenos Aires, Argentina: Using uniparentally and biparentally inherited genetic markers. *Hum. Biol.* 76:543–557.

Msaidie, S., A. Ducourneau, G. Boetsch et al. 2011. Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean. *Eur. J. Hum. Genet*. 19:89–94.

Muzzio, M., J. C. Muzzio, C. M. Bravi et al. 2010. Technical note: A method for assignment of the weight of characters. *Am. J. Phys. Anthropol.* 143:488–492.

Muzzio, M., J. M. B. Motti, P. B. Paz Sepulveda et al. 2018. Population structure in Argentina. *PLoS One* 13:e0196325.

Pacecca, M. I., and C. Courtis. 2008. *Inmigración Contemporánea en Argentina: Dinámicas y Políticas*. Santiago, Chile: Naciones Unidas.

Pinotti, T., A. Bergström, M. Geppert et al. 2019. Y chromosome sequences reveal a short Beringian Standstill, rapid expansion, and early population structure of Native American founders. *Curr. Biol.* 29:149–157.

Posth, C., N. Nakatsuka, I. Lazaridis et al. 2018. Reconstructing the deep population history of Central and South America. *Cell* 175:1,185–1,197.

Raghavan, M., M. Steinrücken, K. Harris et al. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349:aab3884.

Rasmussen, M., S. L. Anzick, M. R. Waters et al. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506:225–229.

Rasmussen, M., Y. Li, S. Lindgreen et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757–762.

Rasmussen, M., M. Sikora, A. Albrechtsen et al. 2015. The ancestry and affiliations of Kennewick Man. *Nature* 523:455–458.

Regueiro, M., J. Alvarez, D. Rowold et al. 2013. On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists. *Am. J. Phys. Anthropol.* 150:333–348.

Saillard, J., P. Forster, N. Lynnerup et al. 2000. mtDNA variation among Greenland Eskimos: The edge of the Beringian expansion. *Am. J. Hum. Genet.* 67:718–726.

Sala, A., and D. Corach. 2014. Analysis of admixture and genetic structure of two Native American groups of Southern Argentinean Patagonia. *Mol. Biol. Rep.* 41:1,533–1,543.

Sala, A., G. Penacino, and D. Corach. 1998. Comparison of allele frequencies of eight STR loci from Argentinian Amerindian and European populations. *Hum. Biol.* 70:937–947.

Sandoval, J. R., D. R. Lacerda, M. S. A. Jota et al. 2013. The genetic history of Indigenous populations of the Peruvian and Bolivian Altiplano: The legacy of the Uros. *PLoS One* 8:e73006.

Schurr, T. G., M. C. Dulik, A. C. Owings et al. 2012. Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska. *Am. J. Phys. Anthropol.* 148:422–435.

Seielstad, M., N. Yuldasheva, N. Singh et al. 2003. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am. J. Hum. Genet.* 73:700–705.

Sengupta, S., L. A. Zhivotovsky, R. King et al. 2006. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both Indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am. J. Hum. Genet.* 78:202–221.

Shen, P., T. Lavi, T. Kivisild et al. 2004. Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum. Mutat.* 24:248–260.

Sušnik, B. 1972. *Dimensiones Migratorias y Pautas Culturales de los Pueblos del Gran Chaco y su Periferia.* Resistencia, Argentina: Instituto de Historia, Facultad de Humanidades, Universidad del Nordeste.

Underhill, P. A., G. Passarino, A. A. Lin et al. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65:43–62.

Vullo, C., V. Gomes, C. Romanini et al. 2015. Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations. *Int. J. Legal Med.* 129:673–680.

Wang, S., N. Ray, W. Rojas et al. 2008. Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet.* 4:e1000037.

Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.*12:339–348.

Zalloua, P. A., Y. Xue, J. Khalife et al. 2008. Y-chromosomal diversity in Lebanon is structured by recent historical events. *Am. J. Hum. Genet.* 82:873–882.

Zhivotovsky, L. A., P. A. Underhill, C. Cinnioğlu et al. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* 74:50–61.

Zhong, H., H. Shi, X. B. Qi et al. 2011. Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol. Biol. Evol.* 28:717–727.

Table 1. Geographic Location and Number of Samples Analyzed

Province	*Locations*	*Latitud*	*Longitud*	*N*	*N Q*
Jujuy	**La Quiaca**	-22,24	-65,72	73	28
	Maimara	-23,88	-65,09	105	54
	Humahuaca	-23,12	-65,21	42	31
	San Salvador de Jujuy	-24,28	-64,78	37	18
Salta	**Salta**	-24,83	-64,59	81	35
Formosa	**Formosa**	-26,11	-58,1	141	141
Santiago del Estero	**Santiago del Estero**	-27,55	-64,46	154	9
Tucuman	**Tucumán**	-26,66	-65,09	158	2
Catamarca	**Santa Maria**	-26,94	-66,41	64	13
	Belén	-28,03	-67,29	56	14
	San Jose	-28,3	-65,87	7	3
	Catamarca	-28,77	-65,4	96	15
La Rioja	**La Rioja**	-30,27	-66,29	86	10
	Chepes	-31,29	-66,73	31	3
San Juan	**Jachal**	-30,00	-68,49	11	4
	Calingasta	-31,63	-69,75	22	4
	San Juan	-31,77	-68,8	77	6
Mendoza	**Villa Tulumaya**	-32,79	-68,49	36	7
	Malargue	-35,17	-68,87	41	4
	Mendoza	-33,47	-68,49	78	8
Paraguay	**Yalve Sanga**	-22,35	-59,53	9	9
	Fieladelfia	-22,20	-60,01	24	24
Total number of individuals				1429	442

N: number of samples from men

N-Q: number of Q-haplogroups

Supplementary Table S1. Sub-haplogroup Frequency in Populations Analyzed, and Haplogroup Q Frequencies Worldwide

[See supplemental Excel file.]

Supplementary Table S2. SNP Considered to Discriminate Sub-haplogroups Q

SNP	2013(GRCh38)/hg38	References	Methods	References
M242	13527976	Karafet et al 2008	APLP	Jurado Medina 2014
P36.2	12384646	YCC,2002	Sequencing	Present work
M346	3019115	Karafet et al 2008	APLP	Jurado Medina 2014
P89.1	13359859	INT. SOC. GENETIC GENEALOGY	Sequencing	Present work
F1096	8219532	INT. SOC. GENETIC GENEALOGY	Sequencing	Present work
NWT01	3020042	Battaglia et al 2013	Sequencing	Present work
M25	19704778	Underhill et al 2001	RFLP	Present work
L565	16984504	INT. SOC. GENETIC GENEALOGY	Sequencing	Present work
M120	19745508	Underhill et al 2001	RFLP	Present work
M323	19705832	Shen et al 2004	Sequencing	Present work
L53	19480410	FAMILY TREE	Sequencing	Present work
L54	21130896	FAMILY TREE	RFLP	Present work
L213	8306992	INT. SOC. GENETIC GENEALOGY	Sequencing	Present work
L456	8367023	FAMILY TREE	Sequencing	Present work
L718	17054795	FAMILY TREE	Sequencing	Present work

L940	6978239	FAMILY TREE	Sequencing	Present work
L191	3019338	FAMILY TREE	Sequencing	Present work
L334	12361348	FAMILY TREE	Sequencing	Present work
CTS2730	12328516	INT. SOC. GENETIC GENEALOGY	Sequencing	Present work
CTS11969	21229412	INT. SOC. GENETIC GENEALOGY	Sequencing	Present work
M3	16984483	Karafet et al 2008	APLP	Jurado Medina 2014

Supplementary Table S3. SNP Analyzed and Oligonucleotide Sequence Information When SNP Was Detected by PCR-RFLP

SNP	Oligonucleotides	Fragment	Enzime	Ancestral state	RFLP
L54	F: TTGTTTGTTGACCCCGTTTT	485	MnII	A	A 218-146-59-62/G 364-59-62
	R: ATTGCTGCTTTTGCTCCACT				
M25	F: CACACAAAACAAGAACCGTGA	172	EcoR I	G	G 82-88 / C 18-64-88
	R: GCACTGACACAAGTTATCTCCC				
M120	F: TGGACAGATTACAGTAAACCTTCAAC	123	BspHI	T	T 100-23 / C 123
	R: GTATAATTTCCCTTAAAAACATCATG				

Supplementary Table S4. SNP Analyzed and Oligonucleotide Sequence Information When SNP Was Detected by Sanger Sequencing

SNP	Sequence	Fragment
P36.2	F: AGCATCTCCACACAGCACAC	629
	R: CATCCATCTACCTACATACCTGTCA	
P89.1	F: GGAGACATGTTTTTATTTCGCTTT	829
	R: TCCTCTTTCAATAATTGTTCCTTT	
F1096	F: ACAGCCAAGGCATTAGAAGG	638
	R:TGTTTTCAACTGGACCTTGC	
NWT01	F: TTTTTGCAATGGGAAAGGAC	688
	R: AAACCCCGCTTTCAGGTAAT	
L565	F: TTACGAAGCCTGGTTCAGTG	679
	R: TCAGTCTCCTCCCAGCAAGT	
M323	F: AGGAGAAAGCAGACCGTGAA	608
	R: AGACCCAAGGGGAGAAAAGA	
L53	F: TGTTTGCCCAAAATTAACCA	388
	R: TTAAATGACCCCCTGAGTGC	
L213	F: CACCCGGATCAAAAGGAGTA	627
	R: GCAGTGGCAACAAAATTCAA	
L456	F: CACCCGGATCAAAAGGAGTA	627
	R: GCAGTGGCAACAAAATTCAA	
L718	F: GGGGTCTCACACCTGGATAA	643
	R: TTTTGGGGTGGATATTTGGA	
L940	F: CCACCACCCAAATAGGCTAC	649
	R: ATTTTGCATGAGGACGAAGC	
L191	F: GGCCTGAAAATGTGGAAAGA	473
	R: CCTGGTGTTCCTCCACTCAT	

L334	F: GATCCTGCCCATTGTTGAGT	557
	R: TCACAATCTCCCCTCTGGTC	
CTS2730	F: GCTCCACGGCTTATTCTCAA	725
	R: GCTCCAGATTGAGTTTTTGCT	
CTS11969	F: GAGGCAAGGTCATGTCACAG	617
	R: TTGGCACAAACGTCATCATT	

Supplementary Table S5. STR Haplotypes of Present Work Samples and Haplotypes Extracted from References: Numeric Codes of Haplotypes Are the Same Codes as Those in Median Joining Network

[See supplemental Excel file.]

Figure Captions

Figure 1. Phylogenetic status of the samples analyzed.

Figure 2. Q1x(Q1a.Q1b) sub-haplogroup median-joining network.

Figure 3. Q1a2 sub-haplogroup median-joining network.

Figure 4. Q1b- Q-M346 sub-haplogroup median-joining network.

Figure 5. Q-CTS2730 and haplotypes Q-L54(xM3)from bibliography median-joining network.

Supplementary Figure S1. Geographic distribution of Q1a2a1b-CTS2730 sub-haplogroup drawn by means of the Surfer12 software with Kriging algorithm.

Supplementary Figure S2. Geographic distribution of Q1a2a1a1-Q-M3 haplogroup drawn by means of the Surfer12 software with Kriging algorithm.

Figure 1.

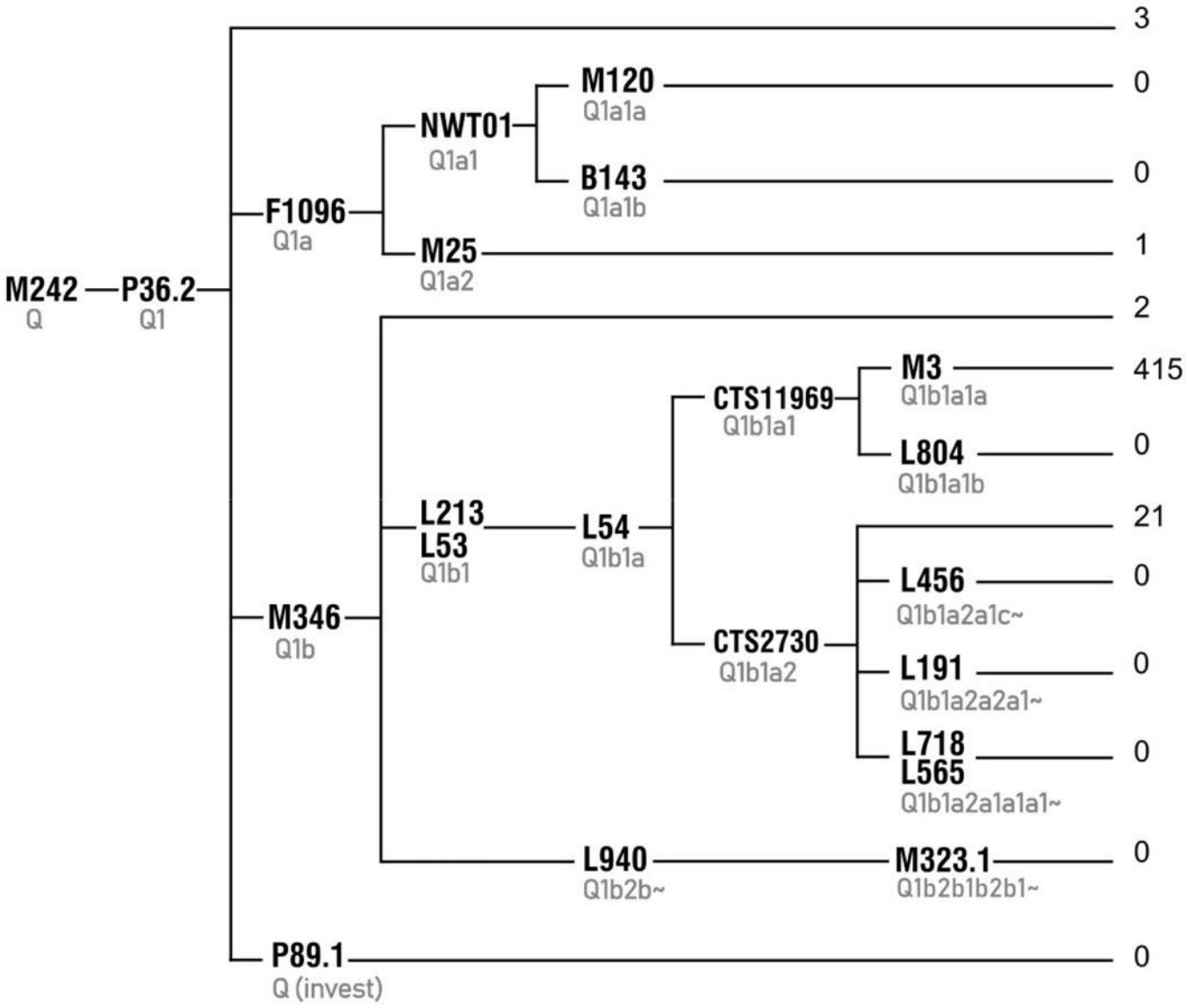


Figure 2.

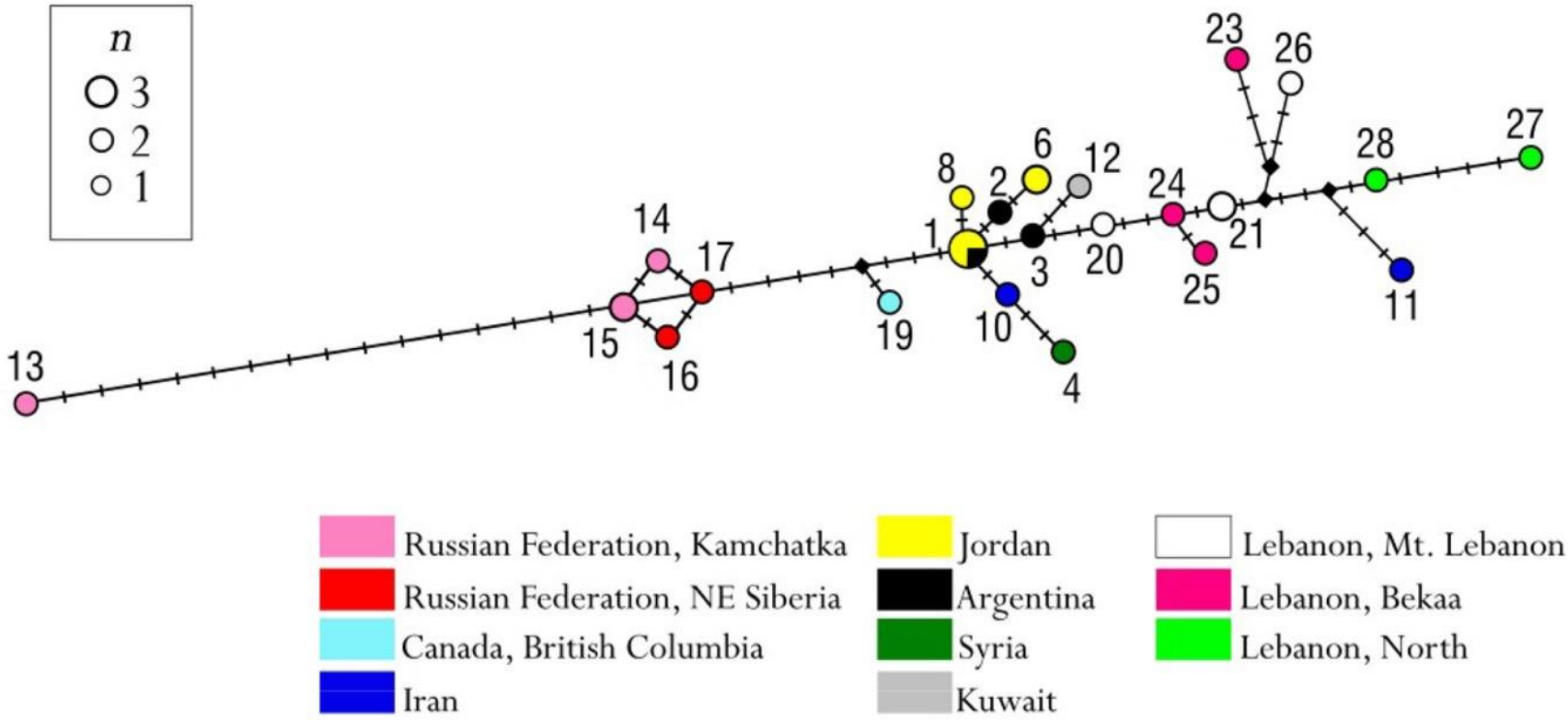


Figure 3.

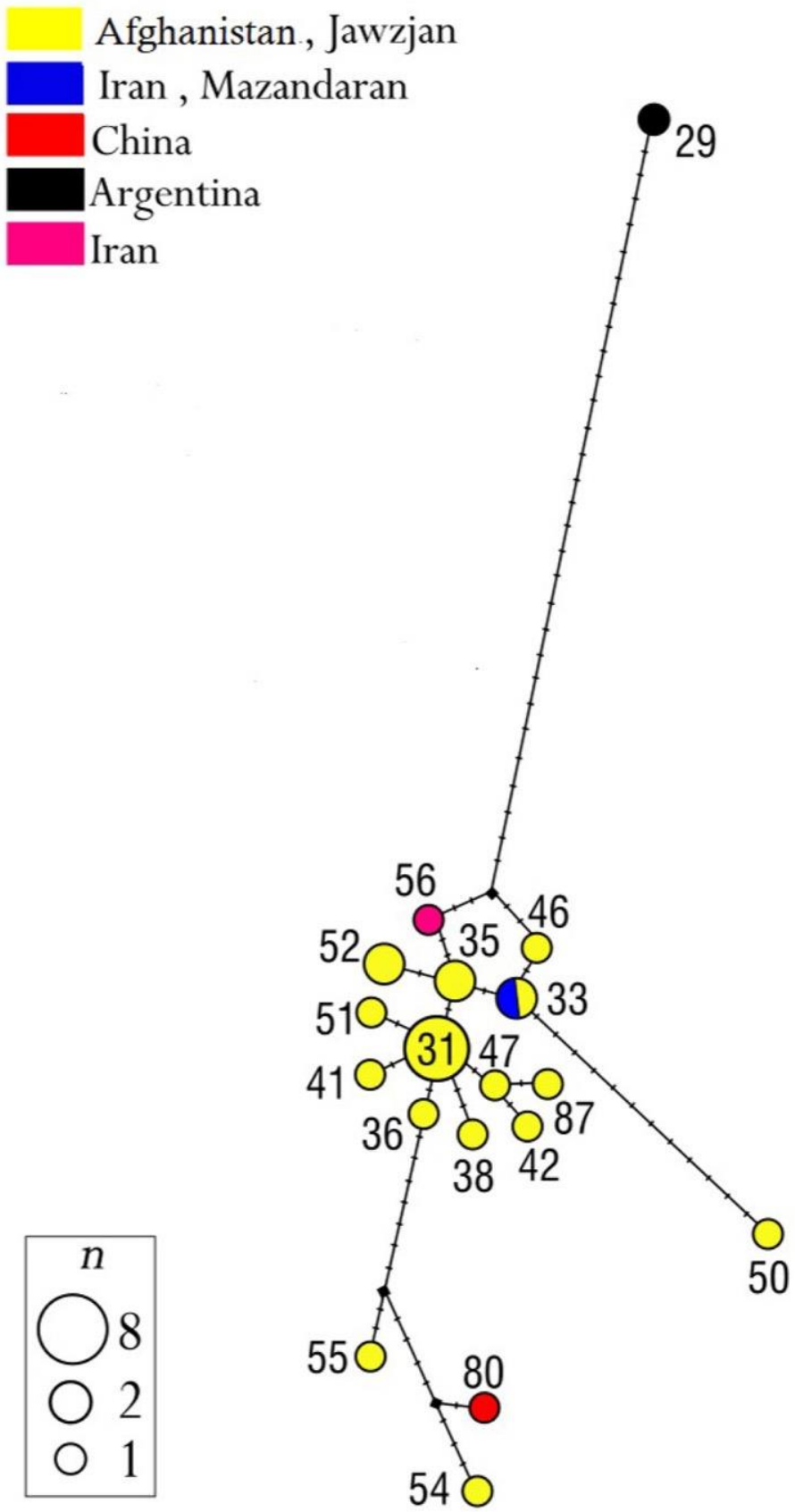


Figure 4.

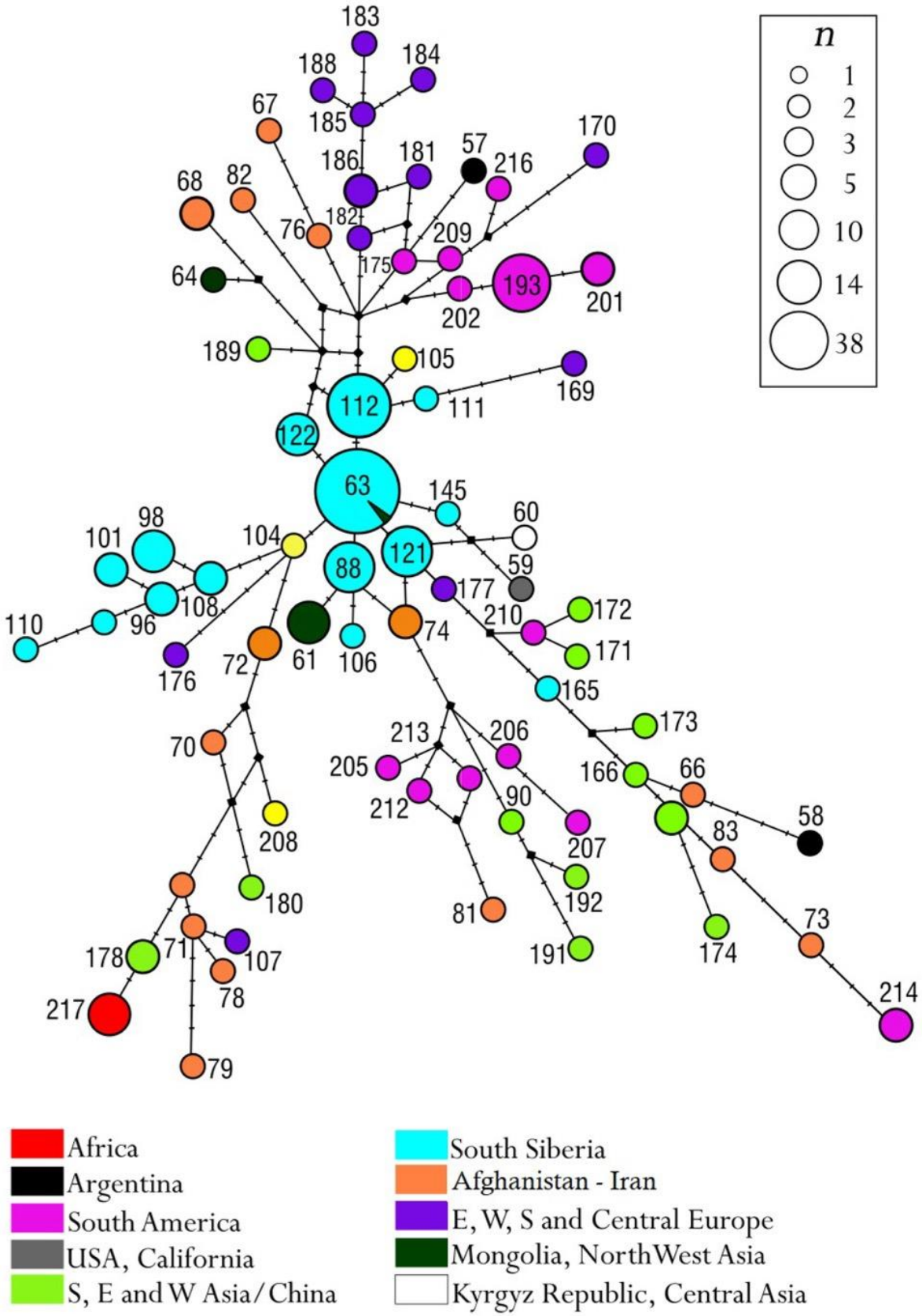
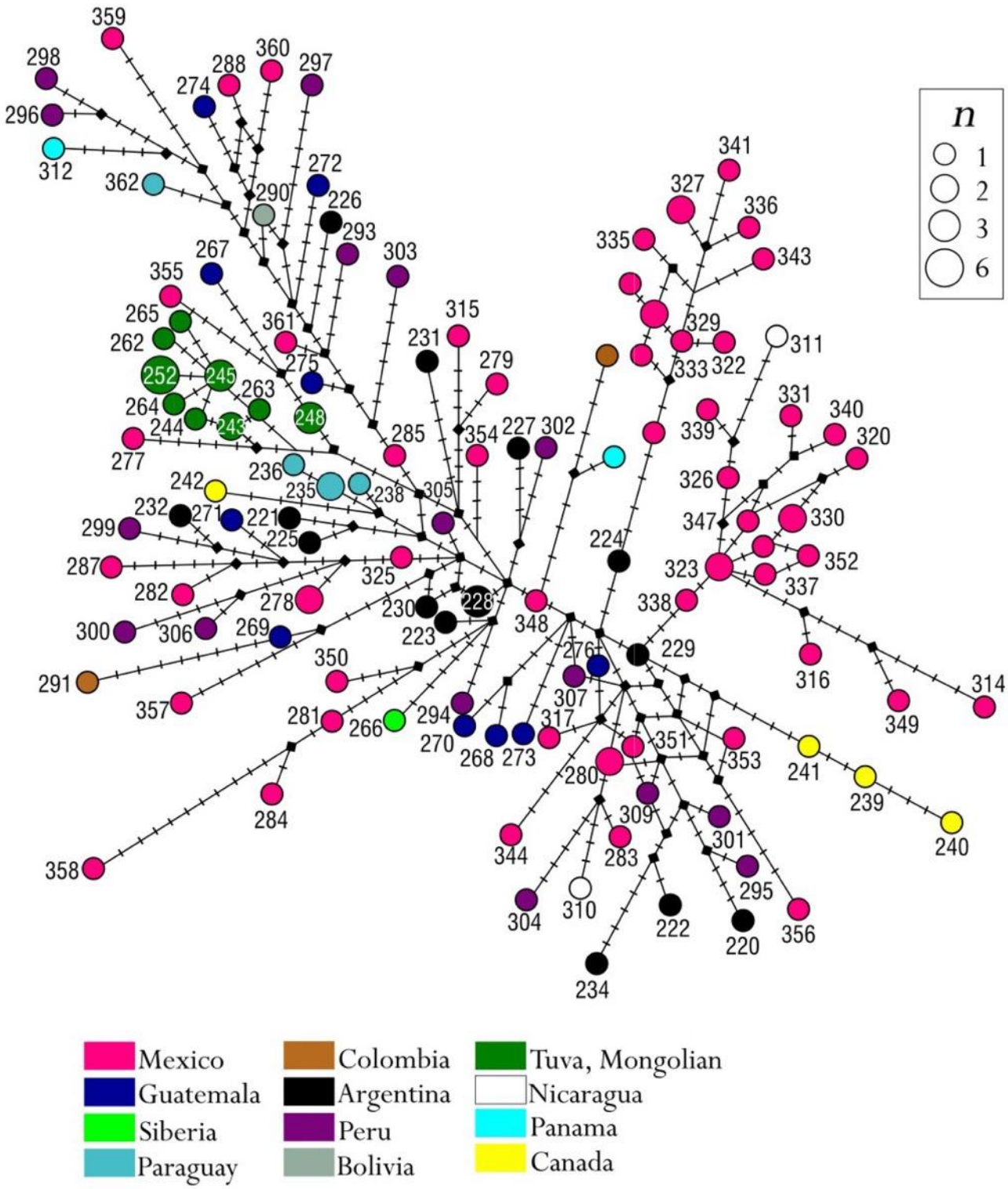
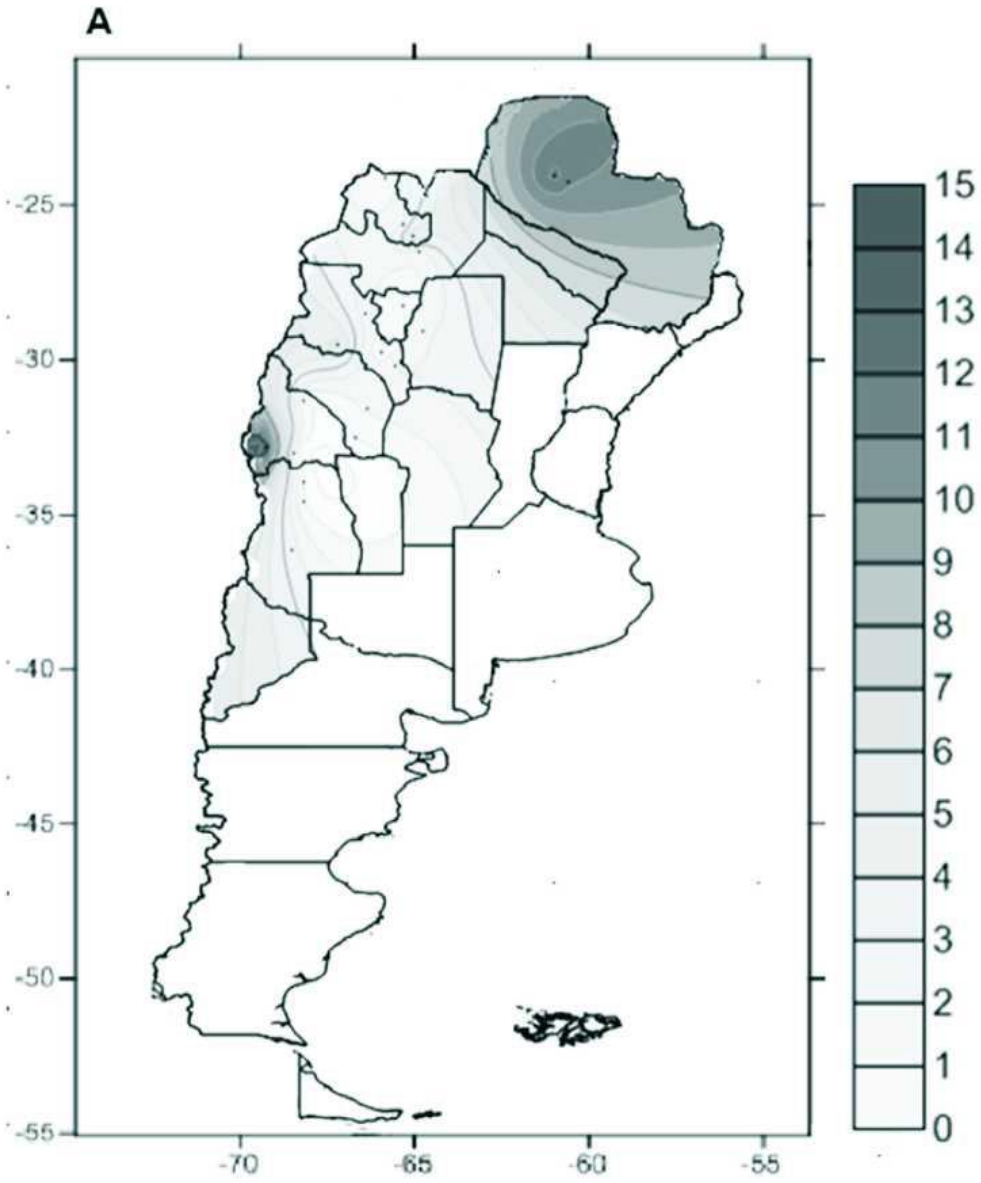


Figure 5.



Supplementary Figure S1.



Supplementary Figure S2.