Maternal Genetic Distance Between Sundanese and Javanese Populations in Indonesia

Wolly Candramila[1*], Sony Heru Sumarsono[2], Bambang Suryobroto[3], and Maelita Ramdani Moeis[2]

[1]Universitas Tanjungpura, Pontianak, Indonesia
[2]Institut Teknologi Bandung, Bandung, Indonesia
[3]Institut Pertanian Bogor, Bogor, Indonesia
wolly.candramila@fkip.untan.ac.id.

Abstract

The study of Austronesian languages classified the Sundanese language within the Malay-Sumbawa group along with Madurese, Aceh, Balinese, Sasak, Sumbawa, and Malay languages. However, The HUGO Pan-Asian SNP Consortium in 2010 reported that the genetic relationship between Sundanese and Javanese people was closer than the one between Sundanese and Malay people. The history of both kingdoms also confirmed the close relationship between Sundanese elites in the west and Javanese elites in the central and east parts of Java. Maternal lineage analysis was done on the mtDNA D-loop region (HVRI and II) from 118 unrelated Sundanese and 33 Javanese probands. We found 12 newly-formed haplotypes (or maternal lineages) as a divergence from one maternal lineage that was characterized by a normal mutation of 186C. These 12 haplotypes were commonly known as the current population of Sundanese and Javanese. Moreover, we also found 11 individual haplotypes that were predicted to be older than the common 12 haplotypes. These older haplotypes were characterized by normal mutations of C64T and C186G. Using a model of K2P, the maternal genetic distance between Sundanese and Javanese populations was 0.12555 or equal to 2190-4622 years of divergence based on the calibration of the mutation rate on both hypervariable regions.

1 Introduction

The Sundanese population is assumed to be the native people in the western part of Java, Indonesia. As a part of the cultural system of a society, the identity and attributes of ethnicity will be inherited and attached to the society member as it passed on from the parents to the next generation. However, the ethnical recognition was not just a part of the self-identification concept based on one's knowledge as a member of an ethnic group through heritability but also an emotional concept attached to it (Tajfel, 1981). Moreover, ethnical recognition was also a developmental process supported by one's acts and

choices throughout his life in a certain ethnic group (Phinney & Ong, 2007). In accordance to the increase of the acculturation process in the western part of Java as well as someone's self-recognition of being a member of a certain ethnic group, Sundanese people cannot be only considered as the native people in the western part of Java displayed by their Sundanese traditional way of life, but also by their self-recognition and emotional concept perceived by the people. However, the self-recognition and emotional concepts are a dynamic pattern considering people's migration and acculturation factors in the region. Based on the National Demography Census in 2010, the most updated number of the "so-called" Sundanese people was more than 36 million individuals in Indonesia.

Studies on the existence of Sundanese people had been done with many variables. Previous studies about Austronesian group languages categorized the Sundanese language into the Malay-Sumbawa group along with Madurese, Aceh, Balinese, Sasak, Sumbawa, and Malay languages (Bellwood, 2007; Bellwood, Fox, & Tryon, 2006; Blust, 2009). Meanwhile, The HUGO Pan-Asian SNP Consortium, et al. (2010) reported that the Sundanese people were genetically closer to the Javanese than to the Malay populations. The relatedness of Sundanese and Javanese populations was also depicted by the history of the mix-marriage between the Sundanese and Javanese elites during the old-kingdoms era, such as that found between a prince of Galuh Kingdom from Ciamis with a princess named Maharani Sima from Kalingga Kingdom in Central Java. The Sundanese population was also traced back into ±2000 years ago according to the historical evidence of Ciaruteun and Jambu inscriptions found in Bogor during the era of Tarumanagara Kingdom (358 A.D.) and the location of Salakanagara or Rajatapura that was mentioned by Ptolemaeus as Argyre (The City of Silver) and predicted to be founded in 150 A.D. in Teluk Lada of Pandeglang Region (Danasasmita, 2014).

The Sundanese population may be historically and genetically closer to the Javanese population than to other ethnicities in Indonesia, yet the current status of its genetic variation is not completely understood due to its large population and wide distribution in Indonesia. The study of genetic variation of the Sundanese population is becoming more interesting considering the existence of some traditional villages (called *kampung adat*) and their endogamous marital system. There is also limited information about the genetic makeup of these closed communities and their relationship with the other Sundanese people in the open communities. In this case, the study of the genetic variation of the Sundanese population should be considering both open and closed communities which were not described in the study by the HUGO Pan-Asian SNP Consortium, et al. (2010). The information on its genetic variation may be beneficial for the description of the distribution of certain genetic diseases in the population. In this research, we studied maternal lineage relationships and genetic distance between Sundanese and Javanese populations based on sequence variation on the D-loop region of mtDNA.

2 Materials and Methods

2.1 DNA Sampling

DNA sampling was done from March 2011 until March 2013 in seven kampung adat and open communities in random settlement areas in West Java Province. The seven kampung adat included:

1. Kampung Cikondang in Lamajang Village of Pangalengan District in Bandung Regency,
2. Kampung Mahmud in Mekarrahayu Village of Margaasih District in Bandung Regency,
3. Kampung Dukuh in Ciroyom Village of Cikelet District in Garut Regency,
4. Kampung Naga in Salawu Village of Leles District in Garut Regency,
5. Kampung Sinarresmi in Sinarresmi Village of Cisolok District in Sukabumi Regency,
6. Kampung Urug in Sukajaya Village of Kiarapandak District in Bogor Regency, and
7. Kampung Kuta in Karangpaningal Village of Tambaksari District in Ciamis Regency.

DNA samples were collected from blood stored in a 1.5 ml Eppendorf tube with 2% EDTA as an anticoagulant agent. Each tube was coded for its sampling location and proband's field number. Blood samples were kept at room temperature until the next steps of the sample's analysis.

2.2   Probands Characteristics

Probands were pure Sundanese and Javanese ethnicity who was not related to each other and proved by tracing back their family lineages for three generations. All probands were healthy individuals and claimed that they were in a healthy condition. Pregnant women were also excluded. Informed consent was given to all sample candidates and signed only by research participants. Place origin of parents and grandparents were recorded in the questionnaire.

2.3   Genetic Variation Determination

Two hypervariable loci (HV1 and HV2) on the D-loop region of mtDNA were targeted at the position of 15990 – 409. The DNA extraction was done according to Sambrook, Fritsch, & Maniatis (1989) with some modifications. Target sequence amplification by polymerase chain reaction was performed using a pair of specific primer of L15990 (5'-TTAACTCCACCATTAGCAC C-3') as previously used by Sigurðardóttir, et al., (2000) and a new design of H409 (5'-CTGTTAAAAGTGCATACCG CC-3') for the Sundanese population. The amplification reaction was according to the protocol of *DreamTaq Green PCR Master Mix* (2X) (Thermo Fisher Scientific, Inc., 2013). The amplification condition was as followed: initial denaturation at 95°C for 3 minutes, denaturation at 95°C for 30 seconds, primer annealing at 54-58°C for 30 seconds, elongation at 72°C, and final elongation at 72°C for 10 minutes. The optimum temperature for primer annealing was varied and ranged between 54 to 58°C. DNA amplification was repeated for 30 – 35 cycles using T100™ Thermal Cycler (Bio-Rad Laboratories, Inc., USA). The amplification product was measured on 1% of agarose gel in the electrophoresis chamber with an electric current of 100 voltage for 30 minutes. A target product of 1000 bp was stained by soaking the gel in 0.5 mg/ml of EtBr liquid for 15 minutes followed with the de-staining process in dH_2O for 15 minutes to eliminate excessive EtBr in the gel. The DNA sequencing of the target product using the same primer pair bidirectionally was sent to Macrogen Inc. in South Korea.

2.4   Data Analysis

Genetic analysis on the D-loop sequence included data analysis of the sequence product in DNA Baserv4 (DNA Baser Sequence Assembler v4.x (2014), Heracle BioSoft SRL, www.DnaBaser.com), multiple alignments under MUSCLE 3.8 (Edgar, 2004) in MEGA 5.2 (Tamura, et al., 2011), determination of mutation types compared to homolog D-loop sequence of rCRS and other populations stored in NCBI, analysis of haplotypes motives and number, as well as building the phylogeny tree using Neighbor-Joining (NJ) method (Saitou & Nei, 1987) and genetic distance using Kimura 2 Parameters (K2P) model (Kimura, 1980).

3   Results and Discussion

The total number of samples included 89 Sundanese individuals from seven kampung adat and 30 individuals from open communities in 9 regencies in western Java, as well as 32 Javanese individuals from 21 cities and regencies in central and eastern Java. Ethical clearance was approved by The Ethics Committee of Health Research from the Faculty of Medicine at the University of Padjadjaran and Regional Hospital of dr. Hasan Sadikin in Bandung.

The NJ phylogeny tree of the samples showed a very close relationship between Sundanese and Javanese populations. According to the mutation pattern on HV1 and HV2, we found 12 common haplotypes or maternal lineages of relatively new divergence (A-L) and 11 individual haplotypes (M1-M11) of older lineages (Figure 1). The eleven presumed older haplotypes were found only in one or two individuals of Sundanese or Javanese, therefore the typical mutation for the haplotype could not be determined unless for C64T and C186G which were found in the whole haplotypes. Meanwhile, the newer 12 haplotypes found both in Sundanese and Javanese populations were shown to be the results of the divergence from one maternal lineage. This maternal lineage gave birth to most varieties of the current Sundanese and Javanese populations (A-L) which was characterized by the normal mutation of 186C. Among these twelve haplotypes, Haplotype A had diverged more into 4 sub-haplotypes (A1-A4). The sample distribution for each haplotype is shown in Table 1.
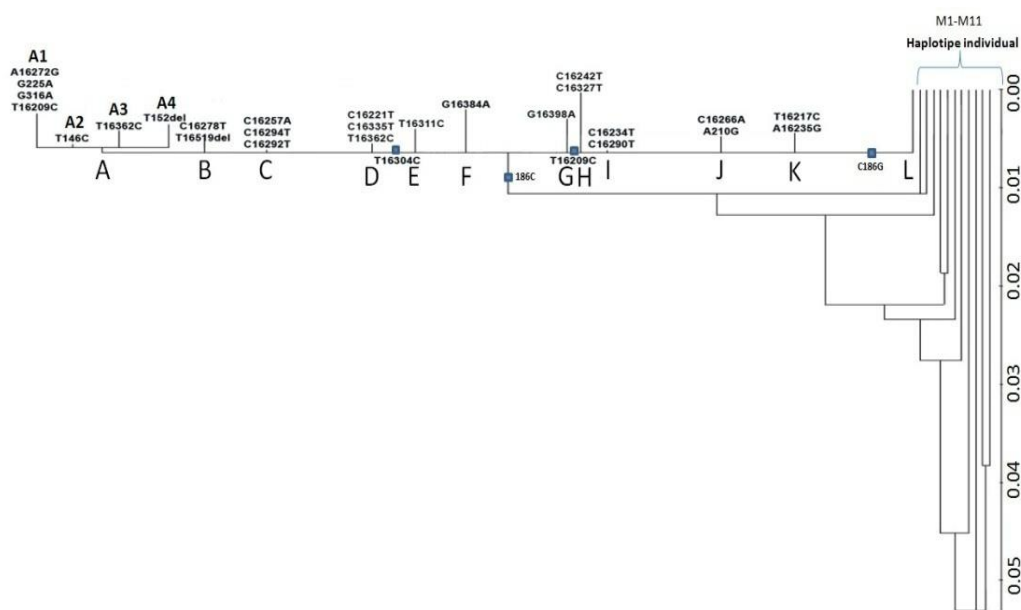
Figure 1: Phylogeny tree of Sundanese (n=118) and Javanese (n=32) populations built with Neighbor Joining method. Twelve new haplotypes (A-L) were assumed to be the core genetic varieties of the current maternal lineages of Sundanese and Javanese populations descended from one older haplotype with typical mutation of 186C. Eleven individual haplotypes (M1-M11) among Sundanese and Javanese populations had normal mutations of C64T and C186G which predicted to be the remnant of the older maternal lineages that had proceeded a bottleneck effect throughout the divergence of the founding population.

Haplotype A showed the highest frequency in Kampung Urug and Kampung Dukuh; Haplotype B and E in Kampung Cikondang; Haplotype C in Kampung Sinarresmi and Sundanese open communities; Haplotype E in Kampung Kuta; and Haplotype J in Kampung Mahmud, Kampung Naga, Kampung Kuta, and Javanese populations. The similar haplotype with the highest frequency indicated that both clusters showed a closer relationship than to other clusters with different haplotypes. Interestingly, every cluster showed more than one haplotype which meant that there was more than one maternal lineage constructed each community. Among the seven *kampung adat*, Kampung Mahmud and Kampung Naga showed similarities in lowest haplotype variation and haplotype with the highest frequency. Meanwhile, Haplotype J was found in almost all sampling communities both in Sundanese

and Javanese populations except that in Kampung Dukuh. It assumed that Haplotype J constructed most of both current Sundanese and Javanese populations.

Ht	Sub-haplotype	Sample's percentage in each community (%)								
		SUB	SSS	SMM	SCP	SDG	SNT	SKC	SOC	J
A	A1	48					20		17	
	A2				7				3	3
	A3	4			7	44		9	3	9
	A4									6
B					14					15
C		9	38			11		9	28	9
D		13			7					
E				7	14	11				6
F		4					20	27		3
G			25							
H						22				
I				29						
J		4	13	64	7		60	27	7	18
K		13	13						10	12
L									3	

Table 1: Sample distribution of each haplotype in each community
Legends:
SOC = Sundanese-open community; SCP = Sundanese-Cikondang Pangalengan; SDG = Sundanese-Dukuh Garut; SKC = Sundanese-Kuta Ciamis; SMM = Sundanese-Mahmud Margaasih; SNT = Sundanese- Naga Tasimalaya; SSS = Sundanese-Sinarresmi Sukabumi; SUB = Sundanese-Urug Bogor.

Generally, Sundanese population showed more varieties of haplotypes than Javanese. Among the twelve haplotypes, five of them were not found in Javanese population. These five haplotypes were only found in Sundanese population, including Haplotype D, G, H, I, and L, yet these haplotypes were also not found in all clusters. Haplotype D was only found in Kampung Urug (13%) and Kampung Cikondang (7%); Haplotype G in Kampung Sinarresmi (25%); Haplotype H in Kampung Dukuh (22%), Haplotype I in Kampung Mahmud (29%); and Haplotype L in Sundanese open communities. The existence of these five haplotypes was probably connected to the differentiation of ethnic recognition by the Sundanese and Javanese people.

Differed from the eleven individual haplotypes (M1-M11) that performed a mutation of C186G, the twelve common haplotypes (A-L) found in Sundanese and Javanese populations showed a typical mutation of 186C. Meanwhile, Haplotype L still carried a mutation of C64T as found in M1-M11. Interestingly, the C64T mutation was found in African sequence origin (*Accession number* D38112) but not in all Asian sequences compared (Japan origin *Accession number* AF346989 and AF346990, China AY255146, Papua New Guinea AY289083, and AY289092, and India AY13976). Unlike Haplotype L, the C64T mutation found in A2 was more likely a convergent evolution instead of a direct descendant from the previous ancestor. Moreover, the mutations of C64T and C186G were typical for maternal lineages in Java Island populations in the past differed from other Asian populations.

Haplotype A, B, C, and D shared a common ancestor with T16304C mutation, meanwhile, Haplotype H, I, J, and K carried a mutation of T16209C. Haplotype E, F, and G performed only one additional mutation from the ancestor with 186C mutation. Haplotype E had a typical mutation of T16311C, Haplotype F G16384A, and Haplotype G G16398A. Haplotype A diverged more into four sub-haplotypes of A1, A2, A3, and A4. Haplotype A1 showed the highest number of mutations,

including A16272G, G225A, G316A, and T16209C. Haplotype A2 had a mutation of T146C, Haplotype A3 had T1632C, and Haplotype A4 had a T deletion at position 152. Haplotype A1 was common in Kampung Urug, meanwhile, Haplotype A3 was common in Kampung Dukuh.

According to the mutation types among all samples of Sundanese and Javanese and using the model of K2P (Kimura 1980), the genetic distance between Sundanese and Javanese population was 0.12555 or equals to the divergence time of 2190 until 4622 years based on the calibration of mutation rate on both hypervariable regions (1.64273×10^{-7} substitution per nucleotide per year for HV1 and 2.29640×10^{-7} substitution per nucleotide per year for HV2) on the D-loop by (Soares, et al., 2009). Meanwhile, the genetic distance between kampung adat was varied started from the lowest of 0.01345 or equals to 32 until 387 years of divergence time between Kampung Dukuh and Kampung Sinarresmi to the highest of 0.04144 or equals to 637 until 830 years of divergence time between Kampung Cikondang and Kampung Kuta (Table 2). Low divergence time indicated that both members of the kampung adat were still conducting mixed marriage or the establishment of the kampung adat was still early.

	Javanese	SOC	SCP	SDG	SKC	SMM	SNT	SSS
Javanese								
SOC	0.01988							
SCP	0.02974	0.02879						
SDG	0.01788	0.01602	0.02714					
SKC	0.03376	0.03306	0.04144	0.03151				
SMM	0.01973	0.01793	0.02901	0.01608	0.03291			
SNT	0.02026	0.01949	0.02960	0.01798	0.03188	0.01973		
SSS	0.01756	0.01581	0.02785	0.01345	0.03059	0.01599	0.01768	
SUB	0.01970	0.01732	0.02859	0.01575	0.03309	0.01783	0.01850	0.01660

Table 2: Genetic distance estimates between Sundanese and Javanese based on the model K2P (Kimura, 1980)

Legends:

SOC = Sundanese-open communities; SCP = Sundanese-Cikondang Pangalengan; SDG = Sundanese-Dukuh Garut; SKC = Sundanese-Kuta Ciamis; SMM = Sundanese-Mahmud Margaasih; SNT = Sundanese-Naga Tasimalaya; SSS = Sundanese-Sinarresmi Sukabumi; SUB = Sundanese-Urug Bogor.

The estimate of divergence time between Sundanese and Javanese populations is following the historical age of the existence of Salakanagara, a city that assumed to have existed since 150 C. It also means that the western Java population have existed for at least 2000 years ago. The relatively early divergence time resulted in many genetic similarities in both Sundanese and Javanese populations. The frequent event of a mixed marriage between populations in Indonesia which was not limited by geographical condition as stated by Handoko, et al. (2001) indicated the potential of more genetic similarities with other western Indonesia populations. The low genetic distance between Sundanese and Javanese populations also confirmed the higher genetic relationship between Sundanese and Javanese populations than between Sundanese and other populations in Indonesia as stated by The HUGO Pan-Asian SNP Consortium, et al. (2010).

4 Conclusion

Sundanese and Javanese populations showed high similarities in genetic variation especially those observed on HV1 and HV2 regions of mtDNA D-loop. Sundanese. Twelve common haplotypes were assumed to be the current maternal lineages of Sundanese, yet only seven of them were found in Javanese. Eleven individual haplotypes owned by one or two individuals were randomly found in both

Sundanese and Javanese people which was predicted to be the remnant of the previous maternal lineages that proceeded bottleneck effect throughout the differentiation of ethnicities between the western and the central and eastern Java. The establishment of different ethnic recognition was believed to be happening between 2190 until 4622 years ago based on the value of the genetic distance between Sundanese and Javanese populations.

Acknowledgment

References

Bellwood P. (2007). *Prehistory of the Indo-Malaysian Archipelago.* Canberra: Australian National University E Press.

Bellwood P, Fox JJ, and Tryon D. (2006). *The Austronesian: Historical and Comparative Perspectives.* Canberra: Australian National University E Press.

Blust R. (2009). *The Austronesian Languages (Pacific Linguistics, 602).* Canberra: Research School of Pacific and Asian Studies, Australian National University.

Danasasmita S. (2014). *Menemukan Kerajaan Sunda. In Bahasa.* Bandung: PT Kiblat Buku Utama.

Edgar R. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics, 5*, 1-19.

Handoko H, Lum J, Gustiani, Rismalia, Kartapradja H, Sofro A, and Marzuki S. (2001). Length Variation in the COII-tRNALys Intergenic Region of Mitochondrial DNA in Indonesian Populations. *Human Biology, 73*, 205-223.

Kimura M. (1980). A Simple Method for Estimating Evolutionary Rates of Base Substitutions Through Comparative Studies of Nucleotide Sequence. *Journal of Molecular Evolution, 16*, 111-120.

Phinney J and Ong A. (2007). Conceptualization and Measurement of Ethnic Identity: Current Status and Future Directions. *Journal of Counseling Psychology, 54*, 271-281.

Saitou N and Nei M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution, 4*, 406-425.

Sambrook J, Fritsch E, and Maniatis T. (1989). *Molecular Cloning: A Laboratory Manual 2nd Edition.* New York: Cold Spring Harbor Laboratory Press.

Sigurðardóttir S, Helgason A, Gulcher J, Stefansson K, and Donnelly P. (2000). The Mutation Rate in the Human mtDNA Control Region. *American Journal of Human Genetics, 66*, 1599-1609.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, . . . and Richards M. (2009). Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *The American Journal of Human Genetics, 84*, 740-759.

Tajfel H. (1981). *Human Groups and Social Categories: Studies in Social Psychology.* New York: Cambridge University Press.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. (2011). Mega5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution, 28*, 2731-2739.

The HUGO Pan-Asian SNP Consortium, et al. (2010). Mapping Human Genetic Diversity in Asia. *Science, 326*, 1541-1545.