AMERICAN JOURNAL OF HUMAN BIOLOGY 18:691-701 (2006)

Original Research Article

Y-STR Haplotype Diversity in Distinct Linguistic Groups From East Timor

LUÍS SOUTO,[1*] LEONOR GUSMÃO,[2] ANTÓNIO AMORIM,[2,3] FRANCISCO CÔRTE-REAL,[4] AND DUARTE N. VIEIRA[4]
[1]*Departamento de Biologia, Universidade de Aveiro, 3810-193 Aveiro, Portugal*
[2]*IPATIMUP, Instituto de Patologia e Imunologia Molecular da Universidade do Porto, 4200-465 Oporto, Portugal*
[3]*Faculdade de Ciências, Universidade do Porto, 4099-002 Oporto, Portugal*
[4]*Delegação de Coimbra, Instituto Nacional de Medicina Legal, 3000-213 Coimbra, Portugal, and Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal*

ABSTRACT East Timor is a country which harbors multiple ethnolinguistic groups generally assigned to an Austronesian or Papuan ancestry. The present study aimed to characterize Y-chromosome haplotype diversity in East Timor, and to test possible population structures based on linguistic and/or geographical information. Using a set of 12 Y-chromosome-specific STRs (DYS19, DYS389I and II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS437, DYS438, and DYS439), haplotypes were established in 342 individuals from 12 linguistic groups (Tetum, Kwaimina, Galoli, Wetarese, Dawan, Mambai, Kemak, Tokodede, Bunak, Makasai, Makalero, and Fataluku) belonging to the three major ethnolinguistic groups in East Timor: two from the Timorese-Austronesian branch (Fabronic and Ramelaic), and a third including languages related to a Trans-New Guinea phylum (Papuan). High values of haplotype diversity, average gene diversity, and mean number of pairwise differences per locus were found in all 12 linguistic groups, except for the Wetarese from the island of Ataúro. Analysis of genetic variance (AMOVA) and pairwise genetic distance analysis showed that the East Timor population is genetically structured, and if the Bunak and Wetarese are excluded, samples group well with respect to their language affinities, and furthermore, the most genetically homogeneous groups are those following the broad ethnolinguistic classifications. Bunak and Wetarese behave as outsider groups, and are genetically more closely related to populations classified in a different linguistic group. Am. J. Hum. Biol. 18:691–701, 2006.

The Democratic Republic of Timor Leste (the official name for East Timor), independent since May 20, 2002, is located in southeastern Asia, occupying the eastern half of Timor Island as well as the Oecussi enclave in the Indonesian west part of Timor, and the small islands of Ataúro and Jaco (Fig. 1).

The present-day population of Timor Leste amounts to slightly less than a million, but shows an increasing demographic tendency despite historically recent tragic events such as the Japanese invasion in World War II, the Indonesian invasion in 1975, and the postreferendum riots of 2001 (United Nations Population Fund (UNFPA), 2004).

The island of Timor has a long period of human occupation, with archaeological findings as old as 30,000–35,000 years ago (O'Connor et al., 2002). It is generally agreed that the island of Timor was subject to a previous Papuan migration (around 7000 AC), followed by a second important wave of Austronesian people.

In contrast to the west part of the island, East Timor is characterized by a diverse linguistic picture, where indigenous groups are associated with a particular language or dialect. However, the number of those linguistic groups is not consensual: from 31 as mentioned by Almeida (1955), to 16 (Gunn, 1999; Hull, 2004) or 17 as in the *Ethnologue* (Gordon, 2005). A scope of the different numbers and designations of language groups in several years in Timor can be followed in Esperança (2001). Some of the initial linguis-

Grant sponsor: Fundação para a Ciência e Tecnologia, FEDER; Grant number: POCI 2010.

*Correspondence to: Luís Souto, Departamento de Biologia, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal. E-mail: lsouto@bio.ua.pt

Received 10 March 2006; Revision received 17 May 2006; Accepted 31 May 2006

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/ajhb.20553

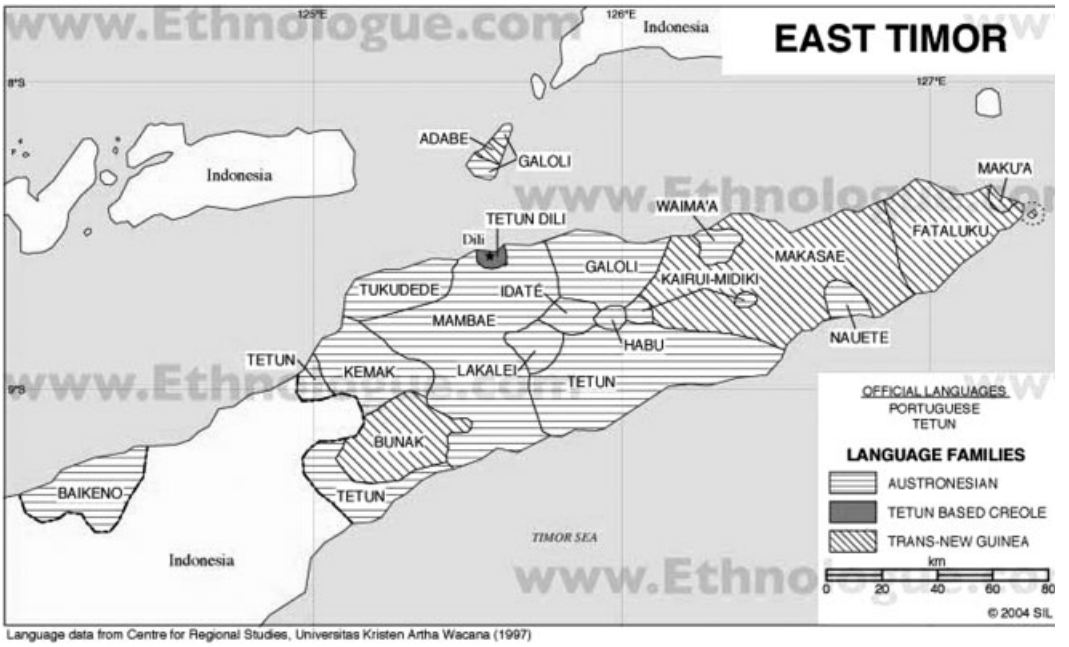

Fig. 1. Language families and groups {Ethnologue, Langauage data from Center for Regional Studies, Universitas Kristen Artha Wacana (1997)}. Different nomenclature/graphology used from the text: Tetun = Tetum (TET), Kairui-Midiki, Waima's (Waimaha), Nauete (Naueti) = Kwaimina (KW). Adabe = Wetarese (WET). Baikeno = Dawan (DAW). Mambae = Mambai (MAM0. Tukudede = Tokodede (TOK). Makasae = Makasai (also includes Makalere MKL). Lakalei, Idate and Maku'a were not sampled in this study.

tic groups have disappeared, and even now some exist as relics, with efforts underway to preserve them. This linguistic diversity is indeed a regional mark, as one can find about 25% of the world's total languages in this region (from Timor to the Solomon Islands) and some 1,000 languages only in New Guinea (Foley, 2000). The multiple Timorese languages are usually assigned to an Austronesian or non-Austronesian (also called Papuan) family, although the affiliation of some groups is controversial. It must be emphasized that the Papuan group is essentially defined by a negative statement toward Austronesian ancestry, than by a clear set of shared traces (Foley, 2000; Pawley and Ross, 1993). Papuan language groups are considered as a more archaic substrate, relating perhaps to the independent development of agriculture in Papua New Guinea and subsequent expansion to as far as Timor, followed by Austronesian migrations from southern China and Taiwan or alternatively from eastern Indonesia, in what has been considered the "fastest and widest expansion of prehistoric times" (Bellwood, 1991). The origins and path of this Austronesian expansion, however, are a matter of considerable debate (Rollet et al., 2002).

Until now, only a few studies have been conducted to characterize the genetic diversity of East Timor. A previous study on 15 autosomal STRs, for a general population sample of 186 individuals living in Timor, revealed a high level of diversity and low differentiation with other populations from the region, namely Indonesia, Malaysia, and Macao (Souto et al., 2005). Nevertheless, when analyzing Y-chromosome-specific STRs (markers that are more sensitive to genetic drift effects) a statistically nonsignificant difference was observed in the comparison of East Timor and a coastal sample from Papua New Guinea (Souto et al., 2006).

Using a set of 12 Y-chromosome-specific STRs (DYS19, DYS389I and II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS437, DYS438, and DYS439), we aimed to characterize the Y-chromosome haplotype diversity inside 12 different East Timor linguistic groups (Tetum, Kwaimina, Galoli, Wetarese, Dawan, Mambai, Kemak, Tokodede, Bunak, Makasai, Makalero, and Fataluku). These groups belong to

TABLE 1. Y-STR haplotype statistics (intrapopulation analysis) all 12 loci considered. N, sample size; NH, number of different haplotypes; HD, haplotype diversity; AGD, average gene diversity; MNPD, mean number of pairwise differences

Metagroup	Group	Code	N	NH	HD	AGD	MNPD
Austronesian	Tétun	TET	33	31	0.996 ± 0.009	0.644 ± 0.345	7.081 ± 3.410
Fabronic N = 90	Kwaimina	KW	18	15	0.980 ± 0.024	0.634 ± 0.350	6.974 ± 3.438
	Galoli	GAL	15	12	0.971 ± 0.033	0.630 ± 0.352	6.933 ± 3.455
	Wetarese	WET	14	10	0.923 ± 0.060	0.347 ± 0.208	3.813 ± 2.041
	Dawan	DAW	10	10	1.00 ± 0.045	0.685 ± 0.395	7.533 ± 3.844
Austronesian	Mambai	MAM	75	58	0.991 ± 0.004	0.635 ± 0.334	6.984 ± 3.319
Ramelaic N = 130	Kemak	KEM	28	28	1.000 ± 0.010	0.637 ± 0.344	7.011 ± 3.395
	Tokodede	TOK	27	21	0.974 ± 0.020	0.647 ± 0.349	7.117 ± 3.446
Trans-New Guinean	Bunak	BUN	24	19	0.982 ± 0.016	0.505 ± 0.280	5.558 ± 2.767
(Papuan) N = 122	Makasai	MAK	80	46	0.975 ± 0.007	0.602 ± 0.318	6.617 ± 3.158
	Makalero	MKL	5	4	0.900 ± 0.161	0.573 ± 0.382	6.300 ± 3.599
	Fataluku	FAT	13	11	0.962 ± 0.050	0.604 ± 0.343	6.641 ± 3.353
Total	East Timor		342	219	0.995 ± 0.000	0.639 ± 0.335	7.027 ± 3.308

the three major linguistic clusters in East Timor: two from the Timorese-Austronesian branch (Fabronic and Ramelaic), and a third including languages related to a Trans-New Guinea phylum (Papuan).

We also intended to test if linguistic and/or geographic factors led to any relevant structuring of male lineages in East Timor.

MATERIALS AND METHODS

Samples

Samples from 342 unrelated blood male donors, representing several districts of East Timor, were collected on FTA® cards, upon informed consent. DNA extraction followed the Chelex protocol of Walsh et al. (1991), using 1.2-mm punches.

Upon inquiries to each volunteer, and according to anthropological and linguistic information, samples were classified, following the criteria of Hull (2004), in 12 linguistic groups: Tetum (TET, N = 33); Kwaimina (which includes Kairui, Waimaha, Midiki, and Naueti, KW, N = 18); Galoli (GAL, N = 15); Wetarese (which includes dialects from the island of Ataúro, WET, N = 14); Dawan (Baikenu dialect, representing those from the Oecussi enclave, DAW, N = 10); Mambai (MAM, N = 75); Kemak (KEM, N = 28); Tokodede (TOK, N = 27); Bunak (BUN, N = 24); Makasai (MAK, N = 80); Makalero (MKL, N = 5); and Fataluku (FAT, N = 13). The geographic distribution of these groups is shown in Figure 1.

Throughout this text, we follow the nomenclature of Hull (2004).

The ethnolinguistic subgroups are gathered, following the same classification, in: *Fabronic* and *Ramelaic*, which together form the Timoric (Timorese-Austronesian) branch, and a third group, *Papuan*, which includes languages that are related to a Trans-New Guinea phylum and are supposed to represent an aboriginal (pre-Austronesian) signature (detailed in Table 1).

Y-STR typing

The 12 Y-chromosome-specific STRs (DYS19, DYS389I and II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS437, DYS438, and DYS439) were amplified with the PowerPlex® Y System kit (Promega Corp.), following the manufacturer's instructions, and analyzed in ABI 310 automatic sequencers (Applied Biosystems) with Genescan 2.1 Analysis software. Allele designation was based on comparison with the allelic ladders provided in the PowerPlex® Y System kit (Promega Corp.).

Statistical analyses

Haplotype frequency and diversity, gene diversity, and mean number of pairwise difference values were calculated for all population samples with ARLEQUIN version 3.0 software (Excoffier et al., 2005). Genetic distances and analysis of molecular variance (AMOVA) were computed with ARLEQUIN version 3.0 software (Excoffier et al., 2005), assuming the stepwise mutational model (sum of squared size differences, R_{st}; Slatkin, 1995). Therefore, DYS385 was not considered, and the number of repeats in DYS389I was subtracted from DYS389II. The Neighbor program, implemented in the PHYLIP 3.65 (Felsenstein, 1989) software package (http://evolution.genetics.

washington.edu/phylip.html), was used to draw unrooted neighbor-joining trees, based on genetic distance values among populations. The tree was visualized with Treeview software (Page, 1996).

RESULTS

Genetic diversity analysis

Table 2 is a list of the different Y-STR haplotypes found in each East Timor ethnolinguistic group studied here. This list includes the profile of 138 individuals studied in a previous work (Souto et al., 2006), plus 204 new haplotypes. In the global sample of 342 individuals, 218 different haplotypes were found. The total number of samples studied in each ethnolinguistic group, and the number of different haplotypes, are given in Table 1. The average diversity per locus and the mean number of pairwise difference values found inside each group were similar, except for the Wetarese and the Bunak groups, presenting the lowest values. For the Wetarese group, haplotype diversity is also below the average value of 0.971 found in the 12 groups, which is more likely due to genetic drift in the colonization of Ataúro Island and subsequent low gene flow from the main territory. On the other hand, Bunak presents higher haplotype diversity than average, which seems to indicate that the low average locus diversity and mean number of pairwise differences are most probably a consequence of some isolation. Makalero is the group in which the lowest haplotype diversity was found. Nevertheless, since it does not present low average locus diversity and mean number of pairwise differences, it is possible that this is a consequence of the low number of Makalero samples available in this study. Therefore, it is expected that, with an increase in sample size, this group will present diversity levels close to the average.

Pairwise genetic distance analysis

Pairwise genetic distances between all samples, assuming the stepwise mutation model (sum of squared size differences, R_{st}), are given in Table 3. Inside the Austronesian-Fabronic group, the highest genetic distances were found in comparisons involving the Wetarese, yielding significant *P*-values, at a 5% level, apart from the comparison of Wetarese vs. Galoli. Low, nonsignificant R_{st} values were found for the remaining Austronesian-Fabronic comparisons.

Inside the Austronesian-Ramelaic sample, low nonsignificant distances were found between all population pairs.

Within Trans-New Guinean (Papuan) samples, no significant distances were found between Makasai, Makalero, and Fataluku, but Bunak yielded highly significant differences in the comparisons with all samples, including those from other major language groups.

In the comparison of populations among different groups, apart from the high genetic distance between Bunak and any of the others, on average, the highest genetic distances are observed between Austronesian-Ramelaic and Trans-New Guinean (Papuan) samples, and the lowest between Austronesian-Fabronic and Austronesian-Ramelaic populations. Wetarese is an exception since, although included in the Austronesian-Fabronic group, it presents higher distances in comparison with samples from the Austronesian-Ramelaic than with those from the Trans-New Guinean (Papuan).

In summary, these results show that the East Timor population is genetically structured, and if Bunak and Wetarese are excluded, samples group well with respect to their language affinities. However, on average, Bunak presents lower genetic distances with Austronesian-Ramelaic populations than with Papuan ones, and Wetarese with Papuan instead of Austronesian-Fabronic populations. Moreover, although Kwaimina shows the lowest genetic distance values in comparison with other Austronesian-Fabronic populations, higher average distances were found in comparison with Austronesian-Ramelaic than with their neighbor Papuan groups.

Analysis of molecular variance

The AMOVA shows that, if samples are joined in a single group, although 92.67% of the variation is observed within populations, a significant proportion can still be attributed to differences among populations (G I, in Table 4). To explain the variation due to differences among populations, an AMOVA was performed considering different arrangements of the 12 ethnolinguistic groups. The first attempts used the traditional Austronesian vs. Papuan classification (G II, Table 4), followed by separation into three groups: Fabronic (Austronesian), Ramelaic (Austronesian), and Papuan (G III; Table 4). Nevertheless, the results did not support such grouping

TABLE 2. Haplotype distribution through different ethnolinguistic groups in East Timor

Haplotype code: DYS19/DYS389I/DYS389II/ DYS390/DYS391/DYS392/DYS393/ DYS385/DYS437/DYS438/DYS439	TET	KW	GAL	WET	DAW	MAM	KEM	TOK	BUN	MAK	MKL	FAT
104TL: 14/14/31/21/10/11/13/12–14/14/10/11	1	–	–	–	–	–	–	–	–	–	–	–
119TL: 15/13/31/25/11/12/13/14–20/14/10/13	1	–	–	–	–	–	–	–	–	–	–	–
125TL: 14/13/27/23/10/13/13/12–20/15/9/11	1	–	–	–	–	–	–	–	–	–	–	–
203TL: 15/12/28/24/10/13/13/12–16/15/10/12	1	–	2	–	–	–	–	–	–	–	–	–
207TL: 14/14/29/23/10/14/14/13–20/14/10/12	1	–	–	–	–	–	–	–	–	–	–	–
239TL: 16/15/31/22/10/11/14/14–19/14/11/12	1	–	–	–	–	–	–	–	–	–	–	–
267TL: 15/12/28/23/10/14/13/13–14/14/10/11	1	–	–	–	–	–	–	–	–	–	–	–
272TL: 15/13/29/22/10/11/14/14–19/14/11/11	1	–	–	–	–	2	2	–	1	–	–	–
286TL: 16/12/28/21/10/12/13/14–19/14/10/11	1	–	–	–	–	–	–	–	–	–	–	–
287 TL: 14/14/29/23/10/14/14/13–19/14/10/11	1	–	–	–	–	4	–	–	1	–	–	–
294TL: 15/12/28/20/10/12/13/14–20/14/10/13	1	–	–	–	–	–	–	–	–	–	–	–
45TL: 16/13/31/21/10/11/15/16–17/14/9/12	1	–	–	–	–	–	–	–	–	–	–	–
470TL: 15/12/28/20/10/13/13/14–19/14/10/11	1	–	–	–	–	–	–	–	–	–	–	–
47TL: 14/12/29/24/10/14/13/12–12/14/9/13	2	–	–	–	–	–	–	–	–	4	–	–
493TL: 14/12/28/26/11/13/14/14–15/16/11/12	1	–	–	–	–	–	–	–	–	–	–	–
70TL: 16/12/29/23/10/14/13/11–14/14/10/11	1	–	–	1	–	2	1	–	–	–	–	–
78TL: 15/14/32/23/10/11/13/13–13/14/10/10	1	–	–	–	–	–	–	–	–	–	–	–
98TL: 15/13/29/22/10/11/15/14–19/14/11/12	1	–	–	–	–	–	–	–	–	–	–	–
210TL: 14/13/29/24/11/13/13/11–14/15/12/12	1	–	–	–	–	–	–	–	–	–	–	–
30TL: 15/12/28/21/10/12/13/14–19/14/10/11	2	1	–	–	–	–	–	–	–	1	–	–
329TL: 17/12/29/23/10/14/13/13–14/14/10/11	1	–	–	–	–	–	–	–	–	–	–	–
336TL: 14/13/29/21/10/12/13/11–14/14/10/13	1	–	–	–	–	–	–	–	–	–	–	–
337TL: 15/14/30/25/10/10/13/14–19/14/10/13	1	–	–	–	–	–	–	–	–	–	–	–
350TL: 14/14/30/22/10/13/13/13–19/15/9/12	1	–	–	–	–	–	–	–	–	–	–	–
352TL: 14/13/30/24/10/14/13/12–14/14/9/13	1	–	–	–	–	1	–	–	–	–	–	–
419TL: 14/13/29/23/10/13/13/12–20/15/9/11	1	–	–	–	–	–	–	–	–	–	–	–
433TL: 14/13/31/25/10/14/12/12–14/14/9/11	1	–	–	–	–	–	–	–	–	–	–	–
422TL: 16/13/29/23/10/14/12/13–15/16/10/12	1	–	–	–	–	–	–	–	–	–	–	–
433TL: 14/13/31/25/10/14/12/12–14/14/9/11	1	0	–	–	–	–	–	–	–	–	–	–
455TL: 16/13/28/23/11/13/13/13–16/16/10/13	1	0	–	–	–	–	–	–	–	–	–	–
464TL: 14/14/30/23/10/13/13/12–19/15/9/12	1	2	–	–	–	–	–	–	–	–	–	–
127TL: 15/14/32/22/11/11/13/13–13/14/10/10	–	1	–	–	–	–	–	–	–	–	–	–
134TL: 15/14/32/22/10/11/13/13–13/14/10/10	–	1	–	–	–	1	–	–	–	2	–	–
13TL: 15/13/31/22/10/11/13/13–13/14/10/10	–	1	–	–	–	–	–	–	–	–	–	–
176TL: 13/13/30/24/10/11/14/14–18/14/10/12	–	1	–	–	–	–	–	–	–	–	–	–
189TL: 14/29/23/10/14/14/13–20/14/10/12	–	1	–	–	–	2	–	–	2	–	–	–
20TL: 15/12/28/22/10/12/13/14–20/14/10/11	–	2	–	–	–	–	–	–	–	1	1	–
227TL: 14/13/31/25/10/13/12/12–14/14/9/11	–	2	–	–	–	–	–	–	–	3	–	–
270TL: 15/12/28/21/9/12/13/14–20/14/10/11	–	1	–	–	–	–	–	–	–	–	–	–
278TL: 16/13/29/23/11/13/13/14–15/16/11/12	–	1	–	–	–	–	–	–	–	–	–	–
285TL: 15/12/28/21/10/12/13/14–20/14/10/11	–	1	–	–	–	–	–	–	–	5	–	–
50TL: 16/12/30/21/10/12/13/14–17/14/10/13	–	1	1	–	–	–	–	–	–	2	–	–
95TL: 13/13/31/23/10/11/14/14–18/14/10/12	–	1	–	–	–	–	–	–	–	–	–	–
96TL: 15/12/28/22/10/12/14/14–21/14/10/14	–	1	–	–	–	–	–	–	–	–	–	–
290TL: 16/13/31/21/11/11/13/14–17/14/11/11	–	–	1	–	–	–	–	–	–	–	–	–
303TL: 15/12/28/24/10/13/13/12–12/15/10/14	–	–	1	–	–	–	1	–	–	–	–	–
305TL: 15/13/30/21/10/11/14/14–14/14/11/11	–	–	1	–	–	–	–	–	–	–	–	–
31TL: 14/12/28/24/10/14/14/12–19/14/10/11	–	–	2	–	–	4	–	–	–	–	–	–
471TL: 14/13/29/23/10/13/13/12–20/14/9/11	–	–	2	–	–	1	–	–	–	–	–	–
519TL: 14/12/29/24/10/14/13/12–13/14/9/13	–	–	1	–	–	–	–	–	–	–	–	–
171TL: 13/13/29/22/11/11/13/15–17/14/10/12	–	–	1	–	–	–	–	–	–	–	–	–
172TL: 14/14/29/23/10/14/14/13–18/14/10/11	–	–	1	–	–	1	–	–	–	–	–	–
338TL: 15/13/27/21/10/11/14/14–18/14/11/12	–	–	1	–	–	–	–	–	–	–	–	–
340TL: 13/14/33/22/10/11/13/15–16/14/10/12	–	–	1	–	–	–	–	–	–	–	–	–
427TL: 15/12/28/20/10/12/13/14–19/14/10/11	–	–	1	4	–	–	–	1	–	–	–	–
241TL: 15/12/28/21/10/12/13/14–14/14/10/11	–	–	–	1	–	1	–	–	–	–	–	–
296TL: 15/12/28/20/10/12/13/14–19/14/10/12	–	–	–	1	–	–	–	–	–	–	–	–
302TL: 15/12/28/24/10/15/13/12–16/14/10/14	–	–	–	1	–	–	–	–	–	–	–	–
306TL: 16/13/31/23/10/14/13/12–13/14/9/13	–	–	–	1	–	–	–	–	–	–	–	–
308TL: 15/12/28/22/10/12/13/14–19/14/10/11	–	–	–	2	1	1	–	–	–	–	–	–
310TL: 15/12/28/24/10/13/13/12–16/14/10/13	–	–	–	1	–	–	–	–	–	–	–	–
313TL: 15/12/28/22/10/12/13/14–14/14/10/11	–	–	–	1	–	–	–	–	–	–	–	–
314TL: 14/13/30/21/10/12/13/11–14/14/10/13	–	–	–	1	–	–	–	–	–	–	–	–

(*Continued*)

TABLE 2. Haplotype distribution through different ethnolinguistic groups in East Timor (Continued)

Haplotype code: DYS19/DYS389I/DYS389II/ DYS390/DYS391/DYS392/DYS393/ DYS385/DYS437/DYS438/DYS439	TET	KW	GAL	WET	DAW	MAM	KEM	TOK	BUN	MAK	MKL	FAT
120TL: 14/12/29/24/10/14/13/12–14/14/9/13	–	–	–	–	1	–	–	–	–	–	–	–
193TL: 14/13/31/24/11/14/13/12–12/14/9/12	–	–	–	–	1	–	–	1	–	–	–	–
194TL: 15/12/29/24/11/13/13/12–15/16/11/11	–	–	–	–	1	–	–	–	–	–	–	–
257TL: 14/13/29/23/10/13/13/12–22/15/9/11	–	–	–	–	1	–	–	–	–	–	–	–
448TL: 16/14/30/24/11/13/14/13–15/16/11/13	–	–	–	–	1	–	–	–	–	–	–	–
450TL: 13/14/32/23/10/11/13/15–16/14/10/13	–	–	–	–	1	–	–	–	–	–	–	–
86TL: 15/14/30/23/10/13/13/14–15/14/10/11	–	–	–	–	1	–	–	–	–	–	–	–
87TL: 15/12/29/21/10/12/14/14–19/14/10/12	–	–	–	–	1	–	–	–	–	–	–	–
170TL: 17/13/30/21/10/12/13/14–19/14/10/11	–	–	–	–	1	–	–	–	–	–	–	–
143TL: 15/13/29/25/11/13/13/12–12/16/11/12	–	–	–	–	–	1	–	–	–	–	–	–
156TL: 16/13/30/22/10/11/13/14–17/14/11/11	–	–	–	–	–	1	–	–	–	–	–	–
16TL: 16/13/29/23/11/14/12/13–15/16/10/12	–	–	–	–	–	1	–	–	–	5	–	–
187TL: 15/12/28/24/10/14/14/12–19/14/10/11	–	–	–	–	–	1	–	–	–	–	–	–
190TL: 15/12/28/24/10/13/13/12–12/15/10/13	–	–	–	–	–	3	–	–	–	–	–	–
197TL: 15/12/29/24/10/13/13/12–12/15/10/13	–	–	–	–	–	2	–	–	–	–	–	–
1TL: 15/14/28/23/10/13/13/12–19/15/9/11	–	–	–	–	–	1	–	–	–	–	–	–
206TL: 15/13/29/21/10/11/13/16–16/14/11/11	–	–	–	–	–	1	–	–	–	–	–	–
209TL: 15/12/28/24/10/13/13/12–13/15/10/13	–	–	–	–	–	1	–	–	–	–	–	–
223TL: 15/13/29/25/11/13/13/12–15/16/11/13	–	–	–	–	–	1	–	–	–	–	–	–
22TL: 17/12/28/25/10/13/12/12–19/14/10/11	–	–	–	–	–	1	1	–	–	–	–	–
231TL: 15/13/29/23/10/13/13/12–14/15/12/11	–	–	–	–	–	1	–	–	–	–	–	–
235TL: 14/12/27/25/11/14/14/14–15/16/11/13	–	–	–	–	–	1	–	–	–	–	–	–
256TL: 14/13/31/24/10/11/14/14–16/14/10/13	–	–	–	–	–	1	–	–	–	–	–	–
262TL: 15/12/28/21/10/12/13/16–19/14/10/12	–	–	–	–	–	1	–	–	–	–	–	–
266TL: 14/14/28/23/10/13/13/12–18/15/9/11	–	–	–	–	–	1	–	–	–	–	–	–
27TL: 16/13/29/23/11/13/13/13–15/15/10/13	–	–	–	–	–	2	1	4	–	–	–	–
281TL: 15/13/30/21/11/11/13/14–18/14/11/11	–	–	–	–	–	1	–	–	–	–	–	–
330TL: 15/13/31/21/11/11/13/14–17/14/11/11	–	–	–	–	–	1	–	–	–	–	–	–
370TL: 15/13/29/25/11/13/13/12–15/16/11/12	–	–	–	–	–	2	1	1	–	–	–	–
396TL: 14/12/28/24/10/14/14/12–18/14/10/12	–	–	–	–	–	2	–	–	–	–	–	–
441TL: 16/12/28/23/10/14/12/11–14/14/10/11	–	–	–	–	–	1	–	–	–	–	–	–
444TL: 15/12/28/21/10/12/13/14–20/14/10/12	–	–	–	–	–	1	–	–	–	–	–	–
465TL: 14/12/28/24/10/14/14/13–19/14/10/11	–	–	–	–	–	1	–	–	–	–	–	–
492TL: 14/12/28/24/10/14/14/12–20/14/10/11	–	–	–	–	–	1	–	–	–	–	–	–
51TL: 15/13/29/22/10/11/14/14–18/14/11/11	–	–	–	–	–	1	1	–	–	–	–	–
59TL: 15/13/29/25/10/13/13/12–15/16/11/12	–	–	–	–	–	1	–	–	–	–	–	–
60TL: 17/13/29/23/11/14/12/13–15/16/10/12	–	–	–	–	–	1	–	–	–	–	–	–
65TL: 15/14/29/23/10/14/14/13–21/14/11/12	–	–	–	–	–	1	–	–	–	–	–	–
88TL: 15/12/29/24/10/13/13/11–14/15/12/12	–	–	–	–	–	1	–	–	–	–	–	–
173TL: 15/13/29/21/10/11/14/13–18/14/12/13	–	–	–	–	–	1	–	–	–	–	–	–
201TL: 14/15/31/23/10/12/13/12–19/15/9/11	–	–	–	–	–	1	–	–	–	–	–	–
341TL: 16/13/31/22/10/11/13/12–13/14/10/10	–	–	–	–	–	1	–	–	–	1	–	–
343TL: 14/13/30/24/9/15/13/12–12/14/9/11	–	–	–	–	–	1	–	–	–	–	–	–
344TL: 14/14/30/23/11/13/13/12–19/14/9/11	–	–	–	–	–	1	–	–	–	–	–	–
356TL: 16/12/28/23/10/14/14/11–14/14/10/11	–	–	–	–	–	2	–	–	–	–	–	–
361TL: 15/12/28/23/10/12/13/14–19/14/10/11	–	–	–	–	–	1	–	–	–	–	–	–
372TL: 15/13/30/21/11/11/14/14–19/14/11/11	–	–	–	–	–	1	–	–	–	–	–	–
373TL: 15/13/31/21/10/11/14/15–19/14/11/12	–	–	–	–	–	1	–	–	–	–	–	–
374TL: 15/13/29/24/11/13/13/12–15/16/11/12	–	–	–	–	–	1	–	–	–	–	–	–
377TL: 16/12/28/23/10/14/13/11–14/14/10/12	–	–	–	–	–	1	–	–	–	–	–	–
383TL: 14/13/29/22/10/13/14/13–21/15/9/12	–	–	–	–	–	1	–	–	–	–	–	–
403TL: 15/13/29/24/12/13/13/12–15/16/11/12	–	–	–	–	–	1	–	–	–	–	–	–
428TL: 14/14/30/23/11/13/13/12–19/15/9/11	–	–	–	–	–	1	–	–	–	–	–	–
463TL: 14/14/29/23/10/14/14/13–19/14/10/12	–	–	–	–	–	1	1	–	1	–	–	–
62TL: 15/13/30/22/10/11/14/14–18/14/11/11	–	–	–	–	–	1	–	2	–	–	–	–
106TL: 15/12/28/23/10/14/13/14–14/14/10/12	–	–	–	–	–	–	1	–	–	–	–	–
144TL: 15/12/28/24/10/13/13/12–16/15/10/14	–	–	–	–	–	–	1	–	–	–	–	–
188TL: 14/14/29/23/10/14/13/13–20/14/10/12	–	–	–	–	–	–	1	–	–	–	–	–
192TL: 16/12/29/21/10/12/13/14–19/14/10/11	–	–	–	–	–	–	1	–	–	–	–	–
211TL: 15/13/29/24/10/13/13/12–12/15/10/14	–	–	–	–	–	–	1	–	–	–	–	–
213TL: 14/13/30/24/10/11/12/12–15/16/10/11	–	–	–	–	–	–	1	–	–	–	–	–
225TL: 15/13/31/24/10/13/13/13–15/16/11/12	–	–	–	–	–	–	1	–	–	–	–	–
226TL: 15/12/28/21/10/12/13/13–19/14/10/11	–	–	–	–	–	–	1	–	–	–	–	–

(*Continued*)

TABLE 2. Haplotype distribution through different ethnolinguistic groups in East Timor (Continued)

Haplotype code: DYS19/DYS389I/DYS389II/ DYS390/DYS391/DYS392/DYS393/ DYS385/DYS437/DYS438/DYS439	TET	KW	GAL	WET	DAW	MAM	KEM	TOK	BUN	MAK	MKL	FAT
258TL: 14/14/30/26/11/11/13/12–16/14/10/13	–	–	–	–	–	–	1	–	–	–	–	–
288TL: 15/12/28/24/10/13/13/12–16/15/10/13	–	–	–	–	–	–	1	–	–	–	–	–
384TL: 14/14/29/23/10/12/14/20–20/14/10/12	–	–	–	–	–	–	1	–	–	–	–	–
467TL: 15/14/31/22/9/11/13/14–17/14/11/11	–	–	–	–	–	–	1	–	–	–	–	–
472TL: 16/13/30/21/10/11/13/14–18/14/11/10	–	–	–	–	–	–	1	–	–	–	–	–
501TL: 14/15/30/23/10/14/14/13–19/14/10/11	–	–	–	–	–	–	1	–	–	–	–	–
509TL: 15/13/29/21/10/12/13/14–18/14/10/11	–	–	–	–	–	–	1	–	–	–	–	–
360TL: 15/13/29/21/10/11/14/14–17/14/11/11	–	–	–	–	–	–	1	–	–	–	–	–
376TL: 16/13/28/25/11/13/14/15–20/14/10/13	–	–	–	–	–	–	1	–	–	–	–	–
382TL: 14/14/29/23/10/12/14/14–20/14/10/12	–	–	–	–	–	–	1	–	–	–	–	–
451TL: 14/13/29/21/10/11/13/13–15/14/10/12	–	–	–	–	–	–	1	–	–	–	–	–
457TL: 15/12/28/24/10/13/13/12–17/15/10/13	–	–	–	–	–	–	1	–	–	–	–	–
140TL: 15/13/29/21/11/10/13/14–14/14/10/11	–	–	–	–	–	–	–	1	–	–	–	–
162TL: 13/13/30/24/10/11/13/15–18/14/10/12	–	–	–	–	–	–	–	1	–	–	–	–
195TL: 15/12/28/24/10/13/13/12–16/14/10/12	–	–	–	–	–	–	–	1	–	–	–	–
230TL: 16/13/29/23/11/13/12/13–15/15/10/13	–	–	–	–	–	–	–	1	–	–	–	–
232TL: 15/13/31/25/11/14/13/12–13/14/9/12	–	–	–	–	–	–	–	1	–	–	–	–
250TL: 15/12/29/21/9/12/13/12–14/14/10/12	–	–	–	–	–	–	–	1	–	–	–	–
486TL: 15/14/29/23/11/11/13/12–17/14/10/12	–	–	–	–	–	–	–	1	–	–	–	–
5TL: 15/13/31/25/11/14/13/12–12/14/9/12	–	–	–	–	–	–	–	1	–	–	–	–
615TL: 14/13/29/23/10/13/13/12–19/16/9/11	–	–	–	–	–	–	–	2	–	–	–	–
618TL: 15/12/28/20/10/14/13/14–18/14/10/11	–	–	–	–	–	–	–	2	–	–	–	–
69TL: 14/14/30/25/11/11/13/12–16/14/10/13	–	–	–	–	–	–	–	1	1	–	–	–
92TL: 16/13/30/22/9/11/14/13–19/14/11/12	–	–	–	–	–	–	–	1	–	–	–	–
97TL: 14/14/29/23/10/14/14/13–20/14/10/13	–	–	–	–	–	–	–	1	–	–	–	–
381TL: 16/12/28/23/10/14/13/11–14/14/10/11	–	–	–	–	–	–	–	1	–	1	–	–
174TL: 14/13/28/23/10/14/14/13–19/14/10/11	–	–	–	–	–	–	–	1	–	2	–	–
368TL: 16/12/31/25/10/13/12/12–19/14/10/11	–	–	–	–	–	–	–	1	–	1	–	–
118TL: 14/13/28/23/10/14/14/13–21/14/10/12	–	–	–	–	–	–	–	–	2	1	–	–
169TL: 15/13/29/21/10/12/13/14–18/14/10/11	–	–	–	–	–	–	–	–	1	1	–	–
242TL: 16/13/30/23/10/15/14/12–14/12/10/12	–	–	–	–	–	–	–	–	1	1	–	–
263TL: 15/13/29/21/10/11/14/14–18/14/11/12	–	–	–	–	–	–	–	–	1	1	–	–
279TL: 14/14/29/23/10/14/13/14–19/14/10/12	–	–	–	–	–	–	–	–	2	1	–	–
447TL: 14/14/29/23/10/14/14/13–21/14/10/12	–	–	–	–	–	–	–	–	2	1	–	–
460TL: 15/13/29/23/10/11/14/14–18/14/11/11	–	–	–	–	–	–	–	–	2	1	–	–
221TL: 14/14/29/23/10/14/14/13–19/14/10/13	–	–	–	–	–	–	–	–	1	1	–	–
348TL: 14/13/29/22/10/13/14/21–21/15/10/15	–	–	–	–	–	–	–	–	1	1	–	–
354TL: 14/14/28/23/10/13/13/12–19/15/9/11	–	–	–	–	–	–	–	–	1	1	–	–
359TL: 14/14/29/23/10/14/14/13–20/14/10/11	–	–	–	–	–	–	–	–	1	1	–	–
388TL: 14/15/30/23/10/14/14/13–19/14/9/13	–	–	–	–	–	–	–	–	1	–	–	–
429TL: 16/13/30/23/10/12/14/12–14/12/10/13	–	–	–	–	–	–	–	–	1	–	–	–
442TL: 13/13/31/24/10/11/13/15–15/15/10/12	–	–	–	–	–	–	–	–	1	–	–	–
102TL: 15/14/31/20/10/11/13/13–16/14/11/13	–	–	–	–	–	–	–	–	–	–	1	–
108TL: 16/12/28/20/10/12/13/16–19/14/10/12	–	–	–	–	–	–	–	–	–	–	1	–
113TL: 15/12/28/22/10/12/13/13–21/14/10/11	–	–	–	–	–	–	–	–	–	–	2	–
117TL: 17/13/29/23/10/14/12/13–16/16/10/12	–	–	–	–	–	–	–	–	–	–	5	–
11TL: 15/12/28/23/10/12/13/14–20/14/10/11	–	–	–	–	–	–	–	–	–	–	1	–
123TL: 16/12/30/21/10/12/13/14–17/14/10/12	–	–	–	–	–	–	–	–	–	–	5	–
129TL: 15/12/28/21/10/12/13/13–20/14/10/12	–	–	–	–	–	–	–	–	–	–	1	–
133TL: 15/12/28/22/10/12/13/14–22/14/10/12	–	–	–	–	–	–	–	–	–	–	1	–
146TL: 15/14/32/22/10/11/13/13–14/14/10/10	–	–	–	–	–	–	–	–	–	–	1	–
165TL: 13/13/30/22/10/11/13/17–17/14/10/12	–	–	–	–	–	–	–	–	–	1	–	–
168TL: 13/13/30/22/11/11/13/15–17/14/10/13	–	–	–	–	–	–	–	–	–	7	–	–
180TL: 15/14/31/21/10/11/13/13–16/14/11/13	–	–	–	–	–	–	–	–	–	1	–	–
181TL: 15/12/29/21/10/13/13/15–16/14/10/11	–	–	–	–	–	–	–	–	–	1	–	–
191TL: 14/12/29/24/10/14/13/11–12/14/9/12	–	–	–	–	–	–	–	–	–	1	–	–
19TL: 16/13/30/23/11/14/12/13–15/16/10/12	–	–	–	–	–	–	–	–	–	1	–	–
216TL: 16/12/28/20/10/12/13/16–19/14/10/11	–	–	–	–	–	–	–	–	–	1	–	–
217TL: 16/13/29/23/11/14/13/13–15/16/10/12	–	–	–	–	–	–	–	–	–	1	–	–
236TL: 16/12/28/21/10/12/13/14–15/16/10/12	–	–	–	–	–	–	–	–	–	1	–	–
245TL: 15/13/31/24/10/13/14/16–19/14/10/12	–	–	–	–	–	–	–	–	–	1	–	–
255TL: 15/12/28/22/10/12/13/14–21/14/10/12	–	–	–	–	–	–	–	–	–	1	–	–
265TL: 15/12/28/21/10/12/13/13–20/14/10/11	–	–	–	–	–	–	–	–	–	1	–	–

(*Continued*)

TABLE 2. Haplotype distribution through different ethnolinguistic groups in East Timor (Continued)

Haplotype code: DYS19/DYS389I/DYS389II/ DYS390/DYS391/DYS392/DYS393/ DYS385/DYS437/DYS438/DYS439	TET	KW	GAL	WET	DAW	MAM	KEM	TOK	BUN	MAK	MKL	FAT
268TL: 13/13/31/23/10/11/12/15–16/14/10/12	–	–	–	–	–	–	–	–	–	2	–	–
275TL: 16/12/29/21/10/12/13/14–17/14/10/12	–	–	–	–	–	–	–	–	–	1	–	–
29TL: 13/13/31/22/10/11/13/15–18/14/10/12	–	–	–	–	–	–	–	–	–	1	–	–
300TL: 15/14/32/22/10/11/13/13–13/14/10/11	–	–	–	–	–	–	–	–	–	1	–	–
491TL: 16/12/28/21/10/12/13/16–19/14/10/11	–	–	–	–	–	–	–	–	–	3	–	–
499TL: 15/13/30/21/10/12/13/14–20/14/10/11	–	–	–	–	–	–	–	–	–	1	–	–
500TL: 13/13/31/22/11/11/14/14–18/14/10/12	–	–	–	–	–	–	–	–	–	1	–	–
504TL: 15/14/32/21/10/11/13/13–13/14/10/10	–	–	–	–	–	–	–	–	–	1	–	–
56TL: 14/13/29/23/10/13/13/12–18/15/9/12	–	–	–	–	–	–	–	–	–	1	–	–
58TL: 13/14/31/22/11/11/13/14–16/14/10/12	–	–	–	–	–	–	–	–	–	1	–	–
80TL: 14/12/28/24/10/13/14/12–13/14/8/12	–	–	–	–	–	–	–	–	–	1	–	–
321TL: 14/12/29/24/10/14/13/12–12/14/9/12	–	–	–	–	–	–	–	–	–	2	–	–
378TL: 15/14/31/22/10/11/13/13–13/14/10/10	–	–	–	–	–	–	–	–	–	1	–	–
379TL: 15/12/28/25/10/13/13/10–16/14/10/15	–	–	–	–	–	–	–	–	–	1	–	–
409TL: 16/12/30/21/10/12/12/14–17/14/10/12	–	–	–	–	–	–	–	–	–	1	–	–
482TL: 14/12/28/22/10/14/11/14–14/15/10/12	–	–	–	–	–	–	–	–	–	1	–	–
253TL: 16/12/28/20/10/12/13/15–18/14/10/11	–	–	–	–	–	–	–	–	–	–	2	–
49TL: 14/13/31/24/11/15/14/12–12/14/9/11	–	–	–	–	–	–	–	–	–	–	1	–
79TL: 14/13/30/24/10/15/14/12–12/14/9/12	–	–	–	–	–	–	–	–	–	–	1	–
151TL: 14/14/30/22/10/13/14/13–21/15/9/13	–	–	–	–	–	–	–	–	–	–	–	1
183TL: 16/14/31/22/11/13/12/13–13/14/10/13	–	–	–	–	–	–	–	–	–	–	–	1
186TL: 16/14/31/22/10/13/12/13–13/14/10/12	–	–	–	–	–	–	–	–	–	–	–	3
219TL: 14/12/28/25/11/13/14/14–15/16/11/11	–	–	–	–	–	–	–	–	–	–	–	1
237TL: 13/14/30/22/11/11/13/14–16/14/10/12	–	–	–	–	–	–	–	–	–	–	–	1
248TL: 16/12/28/20/10/12/13/17–18/14/10/11	–	–	–	–	–	–	–	–	–	–	–	1
291TL: 16/14/31/22/9/13/12/13–13/14/10/12	–	–	–	–	–	–	–	–	–	–	–	1
476TL: 14/13/29/22/10/13/13/14–21/16/9/13	–	–	–	–	–	–	–	–	–	–	–	1
478TL: 14/13/30/22/10/13/13/13–22/15/9/12	–	–	–	–	–	–	–	–	–	–	–	1
518TL: 16/12/28/20/10/12/13/16–20/14/10/11	–	–	–	–	–	–	–	–	–	–	–	1

TABLE 3. Sum of squared-differences R_{st} between samples from 12 different ethnolinguistic groups in East Timor

	Austronesian-Fabronic					Austronesian-Ramelaic			PAPUAN		
	TET	KWA	GAL	WET	DAW	MAM	KEM	TOK	BUN	MAK	MKL
KWA	0.012										
GAL	−0.029	−0.015									
WET	0.097*	0.103*	0.083								
DAW	−0.041	−0.022	−0.036	0.141*							
MAM	0.012	0.108**	0.019	0.198**	0.005						
KEM	−0.010	0.056*	0.002	0.160*	−0.009	0.015					
TOK	−0.022	0.023	−0.010	0.114*	−0.057	0.013	−0.011				
BUN	0.080**	0.205**	0.089*	0.310**	0.148**	0.102**	0.096**	0.128**			
MAK	0.050*	0.006	0.041	0.041	0.007	0.142**	0.092**	0.040*	0.251**		
MKL	−0.002	0.012	−0.022	−0.032	−0.020	0.077	0.096	0.012	0.197*	0.016	
FAT	0.028	0.035	0.056	0.095*	0.018	0.150**	0.078*	0.035	0.198**	0.013	0.039

*Significant *P*-values at 0.05 level.
**Significant *P*-values at 0.005 level (applying Bonferroni correction for multiple tests).

strategies, since the variation among populations within groups was much higher than that observed among groups. However, when considering geographic criteria over ethnolinguistic affiliation, i.e., assigning an ethnolinguistic label to a western or eastern provenience (G IV, Table 4; see also map, Fig. 1), the percentage of variation among groups increased, while the variation within groups decreased, although we still have significant among-populations, within-group variation ($P = 0.000$). A similar result was found when, following genetic distance analysis, Bunak was included in the Ramelaic group and Wetarese

TABLE 4. AMOVA results from R_{st} pairwise distances with different tentative groupings

Grouping	Source of variation (% of variance)		
	Among groups	Among populations within groups	Within populations
G I: 1 group		7.33	92.67
G II: 2 groups: Austronesian/Papuan	0.54	7.02	92.44
G III: 3 groups: Fabronic/Ramelaic/Papua	0.90	6.62	92.48
G IV: 2 groups: West/East	8.44	2.43	89.13
G V: 3 groups: Fabronic/Ramelaic (plus BUN)/ Papua (plus WET)	6.28	2.54	91.18
G VI: 1 group (excluding WET and BUN)		4.89	95.11
G VII: 2 groups Austronesian/Papuan (excluding WET and BUN)	6.82	1.15	92.03
G VIII: 3 groups: Fabronic/Ramelaic/Papua (excluding WET and BUN)	5.89	0.22	93.89
G IX: 2 groups: West/East (excluding WET and BUN)	7.66	0.38	91.96

in Papuan, instead of their original placements (G V, Table 4).

Given that both diversity and genetic distance analysis show that Bunak and Wetarese are outsider populations, the AMOVA was repeated, using the same grouping strategy but excluding data from these two populations (G VI–IX, Table 4). The results show that no significant among-populations, within-group variation is observed when grouping the samples using the Austronesian, Ramelaic, and Papuan (0.22%; $P = 0.391$) or Austronesian vs. Papuan classifications (1.15%; $P = 0.092$). However, when samples are divided according to geography (Table 4, G IX), the percentage of variation explained by differences among populations within groups, although low (0.38%), is still significant ($P = 0.008$).

In summary, AMOVA supports the prior genetic distance results, demonstrating that, genetically, both Bunak and Wetarese are more closely related to populations from a different linguistic group, and that the most genetically homogeneous groups are those following the three wide language groups classification, i.e., Fabronic/Ramelaic/Papuan, provided that Bunak and Wetarese are set apart.

DISCUSSION

According to anthropological and linguistic information, one would expect to find a heterogeneous picture of the genetic background of East Timor. Indeed, our genetic results are in agreement with the high levels of ethnolinguistic diversity of the population of East Timor.

Although the linguistic substrate of East Timor has been discussed, no genetic data were previously presented concerning the genetic structure of this population and possible genetic and linguistic relationships. Several linguistic classifications of the East Timor population were attempted, all suggesting a Papuan vs. Austronesian influence; we followed a three-language grouping according to Hull (2004). Provided that two ethnolinguistic groups are set apart, we found that, despite the high diversity within each major group, the genetic data support the traditional Papuan vs. Austronesian dichotomy and even the three-language combination tested. These data are more significant if we realize that linguistic differences do not necessarily mean cultural differences (McWilliams, personal communication, 2004). Curiously, 2 of the 12 groups studied (Wetarese and Bunak), differ from the previous picture, present a very distinct Y-lineage profile, and therefore can be considered outgroups. The Wetarese, including the people from Ataúro Island opposite Díli, display signs of genetic drift and low gene flow from main Timor. Also, the Wetarese show greater affinity toward the Papuan than toward *Austronesian-Fabronic*. As expected, Galoli is the *Austronesian-Fabronic* group presenting the lowest genetic distance with the Wetarese, in agreement with known linguistic data also placing Galoli in Ataúro (see map, Fig. 1) and suggesting some gene flow between them.

The Bunak are linguistically segregated, as they are not closely related to any other language (Gordon, 2005), do not understand and are not understood by their (Austronesian) neighbors, and are quite differentiated from other Papuan-like languages due to geographical isolation (Hull, 2004). Indeed, the Bunak are located in a mountain region and are sur-

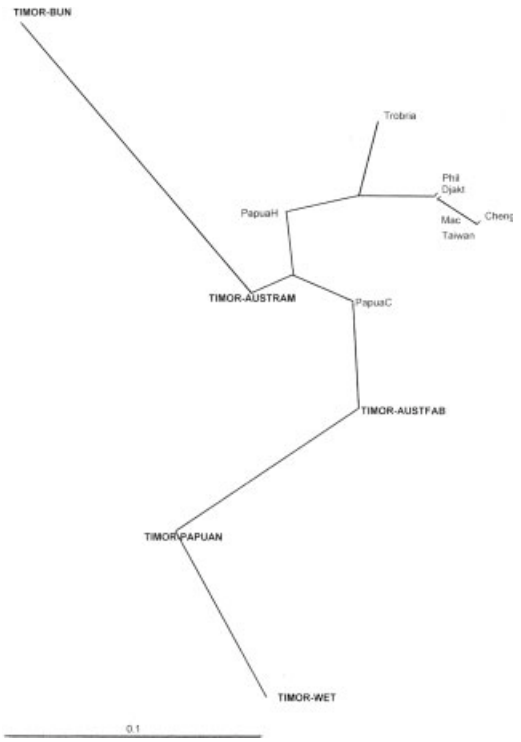

Fig. 2. Genetic distances between East Timor samples (Grouped, in bold) and other populations from region: PapuaC, Papuan-Coastal; PapuaH, Papuan Highlands; Trob, Trobriand Islands; Phil, Luzon (Philipines); Djakt, Djakarta (Indonesia); Mac, Macao; Cheng, Chengdu (China); Taiw, Taiwan. East Timor samples: TIMOR-WET, Wetarese; TIMOR-BUN, Bunak; TIMOR-AUSTFAB, Austronesian-Fabronic group; TIMOR-AUSTRAM, Austrnesian-Ramelaic group.

rounded by populations belonging to a different linguistic branch (Fig. 1), and are believed to have spread from the mountains to the seaside in contrast to all the other languages of Timor (Esperança, 2001). Thus, the Bunak represent a highly differentiated group according to our data, in which geography, rather than language, seems to play a major role. It should be mentioned that even today the communications between different districts of East Timor are not easy (e.g., from Díli to Baucau, in the district of Lautem, a 120-km distance would take a 3-hr journey, and in some districts, it would take more than 6 hr to reach the district capital from one village; GERTIL, 2002). This, along with a history of the pulverization of local kingdoms (in opposition to what happened in the west part of Timor) into dozens of "sucos," i.e., local hierarchic population groups that persist nowadays, can explain the high degree of intrapopulation differentiation.

In a previous comparison of an East Timor general sample with others from the region, significant differences were found with all populations considered, except for one New Guinea Papuan (Souto et al., 2006). Using data from the same populations (Djakarta, Indonesia, Trobriand Islands, and Highland and Coastal Papua New Guinea, Kayser et al., 2001; Luzon, Philippines, Tan et al., 2004; Macao, China, Gusmão et al., 2000; Chengdu, China, Hidding and Schmitt, 2000; and Taiwan, Wu and Pu, 2001), the present East Timor population ethnolinguistic groups are genetically closer to Papuan than to Southeast Asians, even those classified as "Austronesian" (Fig. 2).

To raise a final remark on the presence of "exotic" haplotypes, 15 samples in total were classifiable as being of European (Portuguese) and Chinese (Macau) male ancestry. Although those findings certainly should not be neglected in an overall study of East Timor genetic diversity, as they represent the important influence of colonial times, we did not consider these samples as beyond the scope of the present research of East Timor ethnolinguistic groups.

ACKNOWLEDGMENTS

The authors thank Fundação Oriente, the East Timor Ministry of Health, Universidade Nacional de Timor Leste, the Health Services of the former Portuguese Mission in East Timor, Prof. Amadeu Soares (Departamento de Biologia, Universidade de Aveiro), and particularly, the East Timorese volunteers. IPATIMUP is partially supported by Fundação para a Ciência e a Tecnologia through Programa Operacional Ciência, Tecnologia e Inovação.

LITERATURE CITED

Almeida A. 1955. Contribuição para o estudo da antropologia serológica dos nativos de Timor Português, de Macau e de S. Tomé e Príncipe. Estudos Ultramarinos 5:1–4.

Bellwood P. 1991. The Austronesian dispersal and the origin of languages. Sci Am 265:70–75.

Esperança JP. 2001. Estudos de linguística Timorense. Aveiro: Sul Associação de Cooperação para o Desenvolvimento Aveiro.

Excoffier L, Laval G, Schneider S. 2005. Arlequin version 3.0: an integrated software package for population genetics data analysis. Evol Bioinfo Online 1:47–50.

Felsenstein J. 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics 5:164–166.

Foley W. 2000. The languages of New Guinea. Annu Rev Anthropol 29:357–404.

GERTIL, Grupo de Estudos de Reconstrução de Timor Leste, Faculdade de Arquitectura da Universidade Técnica de Lisboa. 2002. Atlas de Timor Leste. Lisbon: Lidel.

Gordon RG Jr, editor. 2005. Ethnologue: languages of the world,15th ed. Dallas: SIL International. Online version: http://www.ethnologue.com/.

Gunn G. 1999. A sociedade Timorense. In: Timor Loro Sae: 500 anos. Lisbon: Livros do Oriente. p 33–54.

Gusmão L, Amorim A, Martins N, Prata MJ, Gonzalez-Neira A, Lareu MV, Carracedo A. 2000. Y chromosome STR haplotypes in Macau (SE China). Prog Forensic Genet 8:324–326.

Hidding M, Schmitt C. 2000. Haplotype frequencies and population data of nine Y-chromosomal STR polymorphisms in a German and a Chinese population. Forensic Sci Int 113:47–53.

Hull G. 2004. The languages of East Timor, some basic facts. http://www.shlrc.mq.edu.au/~leccles/langs.html

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Knijff P. 2001. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 68:990–1018.

O'Connor S, Mattew S, Veth P. 2002. Excavation at Lene Hara Cave establishes occupation in East Timor at least 30000–35000 years ago. Antiquity 76:45–50.

Page RDM. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358.

Pawley A, Ross M. 1993. Austronesian historical linguistics and culture history. Annu Rev Anthropol 22:425–459.

Rolett BV, Tianlong J, Gongwu L. 2002. Early seafaring in the Taiwan Strait and the search for Autronesian origins. J East Asian Archaeol 4:307–319.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.

Souto L, Alves C, Gusmão L, Ferreira E, Amorim A, Corte-Real F, Vieira DN. 2005. Population data on 15 autosomal STRs in a sample from East Timor. Forensic Sci Int 155:77–80.

Souto L, Gusmão L, Ferreira E, Amorim A, Corte-Real F, Vieira DN. 2006. Y-chromosome STR haplotypes in East Timor: forensic evaluation and population data. Forensic Sci Int 156:2–3.

Tan M, Delfin F, Calacal G, Roewer L, De Ungria MCA. 2004. Y-chromosome STR allele frequency distribution among male Filipinos in the National Capital region for forensic applications. Philipp J Sci 132:95–101.

UNFPA. 2004. Census Timor-Leste 2004. http://dne.mopf.gov.tp

Walsh PS, Metzer DA, Higuchi R. 1991. Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506–513.

Wu FC, Pu CE. 2001. Multiplex DNA typing of short tandem repeat loci on Y chromosome of Chinese population in Taiwan. Forensic Sci Int 120:213–222.