13 The GM genetic polymorphism in Taiwan aborigines

New data revealing remarkable differentiation patterns

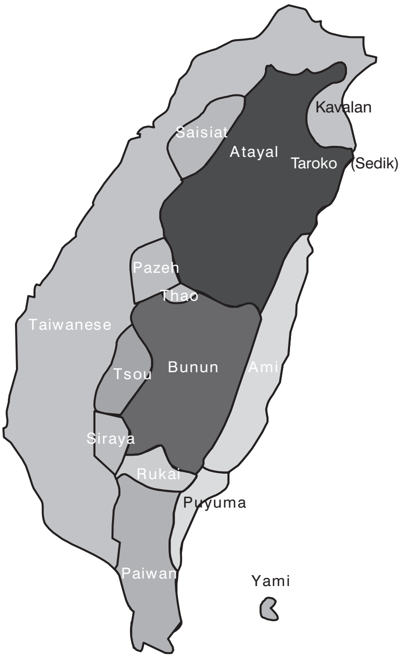
Alicia Sanchez-Mazas, Ludmilla Osipova,
Jean-Michel Dugoujon, Laurent Sagart and
Estella S. Poloni

Introduction

Taiwan (ex-Formosa), a large island of about 36,000 square kilometres located off the south-eastern coast of China, has aroused great interest among scientists from different disciplines whose study focuses on the history of settlement of East Asia and Oceania. The population of Taiwan was 22 million people in 2004, distributed among a large variety of populations. The Chinese, who for the most part migrated to the island in the 17th century ~~(1661)~~ after short-term occupations by the Portuguese (1544–82), Dutch (1624–62) and Spanish (1626–42), currently represent almost 98 per cent of its inhabitants. The main Chinese groups in Taiwan speak southeast Chinese dialects: the Minnan dialect (about 70 per cent) and the Hakka dialect (about 15 per cent). A further 12 per cent correspond to Han people who migrated from the mainland after World War II and are typically speakers of Mandarin, although most other Chinese dialects are also represented. Besides the Chinese, the island is also inhabited by aboriginal peoples who now represent about 2 per cent of the population. Twelve ‘tribes’, speaking distinctive Austronesian languages, some subdivided into dialects, are officially recognized (Map 13.I): the Ami, Atayal, Bunun, Kavalan, Paiwan, Puyuma, Rukai, Saisiat, Thao, Tsou and Sedik (including the Taroko), plus the Yami in Lan-Yu (‘Orchid’) Island. The Yami are not linguistically Formosan, but speak an Austronesian language within Batanic, a Northern Philippine group. Geographically, these populations are located on the east coast and in the mountains in the centre of the island. Twelve additional Formosan languages: Babuza, Basay, Hoanya, Ketagalan, Kulon, Luilang, Pazeh, Papora, Qauqaut, Siraya, Taokas and Trobiawan are extinct or on the verge of extinction (Pazeh). Most were spoken on the western and northern coasts of the island, the focal points of Chinese settlement.

Among
them, t

The Austronesian phylum is the second largest in the world after Niger-Congo in terms of the number of different languages spoken (1268 according to the Ethnologue).[1] Taiwan represents its northernmost geographic boundary. Linguists are generally agreed that all the non-Taiwanese Austronesian languages (spoken in



Map 13.1 Map of Taiwan showing the areas settled by the aboriginal populations (Austronesian) and the Taiwanese (Chinese).

the Philippines, Indonesia, Island and Coastal Melanesia, Micronesia, Polynesia, part of continental Southeast Asia, and Madagascar) belong to only one subgroup, ‘Malayo-Polynesian’ (Blust 1977), while opinions differ on whether Malayo-Polynesian is a primary branch of Austronesian (Blust 1999), or part of a branch also including some Formosan members such as Ami (Harvey 1979; Reid 1982; and many others). The number of primary Austronesian branches represented in Taiwan varies between two (Starosta 1995) and nine (Blust 1999). This makes Taiwan the most diverse area in the entire Austronesian area, and therefore, by Sapir’s principle (Sapir 1916), the most likely candidate for the Austronesian homeland. The Formosan languages are also archaic in various respects, which fits well with the idea of a Taiwanese homeland. Blust (1988) has given a highly plausible picture of the cultural and environmental changes involved in the Malayo-Polynesian migration, as reflected in the Austronesian lexicon. Finally, Gray and Jordan (2000), using lexical information and parsimony analysis, have produced a phylogeny for 77 languages of the Austronesian family that places Formosan languages at the top of the tree.

The thesis of a Formosan homeland is also supported by archaeological data. Since the 1970s, Peter Bellwood has argued for a close relationship between the geographical spread of Austronesian-speakers and the likely diffusion of farming societies identified by a sequence of Neolithic sites throughout the Austronesian area, with Taiwan representing the most ancient settlements (Bellwood 1978, 1997; Bellwood and Dizon, Chapter 1, this volume). In turn, the earliest Austronesian settlers are viewed as originating on the East Asia mainland: Chang Kwang-chih (1969) has identified ceramic sites on the mainland side of the Taiwan straits that are contemporary with, and culturally similar to, the earliest ceramic sites on Taiwan, pointing to the mainland as the origin of the Austronesian peopling of Taiwan (see also Blench, Chapter 4, this volume).

Those ideas gave rise to the so-called 'express-train to Polynesia' hypothesis (Diamond 1988) that has more recently been revisited by Peter Bellwood himself; indeed, this author considers that the train was not so express, since it took at least 1000 and perhaps 1500 years for Taiwanese people to migrate out of Taiwan since their first settlement in the island (Bellwood and Dizon, Chapter 1, this volume). His main argument is that, among other archaeological evidence, recent findings of carbonized grains at Nanguanli (south Taiwan) indicate rice and foxtail millet cultivation around 5,000 years ago in the island (Tsang 2005), and the likely earliest date for Taiwan Neolithic could then be around 3500 BCE. By contrast, although archaeological material found in the neighbouring Batan archipelago and the north of Philippines indicates a close cultural relationship with Taiwan, no Neolithic cultures appear to have moved from Taiwan until around 2500 BCE, or later (Bellwood and Dizon, Chapter 1, this volume); hence a temporal gap of at least 1000 years.

Several scenarios have been proposed for the initial differentiation of Austronesian languages on Taiwan. Starosta (1995) examined morphological innovations and presented a tree-like phylogeny taking the southwest coast of Taiwan to be the Austronesian homeland. In a study of the main sound changes in the Formosan languages, Blust (1999) could not detect a tree-like signal, and distinguished no less than 10 primary Austronesian branches, nine in Taiwan and a tenth – Malayo-Polynesian – outside of Taiwan. Another proposal, mostly based on lexical innovations in the numerals system, was proposed by Sagart (2005; and Chapter 5, this volume). It implies a geographic expansion of Formosan populations out of the northwest coast of Taiwan, where the first settlers arrived from China, southwards along the west coast to the south and southeast, from where migrations out of the island took place. This model has been integrated by Sagart (Chapter 5, this volume) into a larger scenario of Neolithic population movements in eastern China. According to his model, the Austronesian languages are a sister language to Sino-Tibetan ('STAN' hypothesis); Sagart also regards the Tai-Kadai languages of mainland Southeast Asia as a subgroup within Austronesian and a sister clade to Malayo-Polynesian.

Taiwan has been studied intensively from a genetic point of view. Most data come from field studies undertaken by Marie Lin's team to sample aboriginal populations, leading to the analysis of numerous genetic systems, among which blood groups and proteins, HLA, microsatellites and mtDNA (Lin and Broadberry 1998; Lin *et al*. 2000, 2005; Chu *et al*. 2001; Lee *et al*. 2002; Trejaut *et al*. 2005; Trejaut *et al*., Chapter 14, this volume). Some samples were also studied in other laboratories for mtDNA (Tajima *et al*. 2003), Y chromosome markers (Su *et al*.

2000; Capelli *et al*. 2001) and serum proteins (Matsumoto *et al*. 1972; Schanfield *et al*. 2002). All those studies reveal both a high level of genetic heterogeneity among Taiwanese aboriginal populations and some remarkable genetic features that differentiate them from mainland East Asian populations. The patterns of genetic variation among Taiwanese and between Taiwanese and other East Asian and Oceanic populations seem to be quite complex, and their relationship to the history of the peopling of this island still remains to be clarified.

The present study is a new contribution to the genetic history of Taiwan. Within the framework of a collaborative study, partly sponsored by the CNRS (OHLL programme), between the Mackay Memorial Hospital in Taipei (Marie Lin), the Siberian Branch of the Russian Academy of Sciences in Novosibirsk (L.O.), the CRLAO in Paris (L.S.), the Centre of Anthropology in Toulouse (J.-M.D.), and the AGP and LGB laboratories in Geneva (A.S.M. and E.S.P.), we investigated the GM polymorphism of human immunoglobulins in several aboriginal and non-aboriginal populations of Taiwan as well as in two non-Taiwanese populations from the Philippines and Thailand. Although many 'classical' genetic systems have already been studied in Taiwanese aborigines, the pattern of genetic variation within Taiwan has not yet been explored in detail for the GM polymorphism; moreover, these new data will help complement our extensive work on GM genetic variation in East Asia (Poloni *et al*. 2005).

The GM polymorphism

The GM polymorphism is characterized by a series of 'allotypes' detected by serological techniques on human antibodies (IgG immunoglobulins) circulating in the blood serum. Each individual is normally tested for the whole set of GM allotypes, which makes it possible to determine the GM phenotypic distribution in a given population sample and to estimate GM haplotype frequencies in the corresponding population. One particularity of the GM polymorphism is its high level of genetic heterogeneity among continental population groups at the world scale. By subdividing the world into 10 continental or subcontinental regions, Dugoujon *et al*. (2004) found a proportion of genetic diversity due to genetic differences among groups, or F_{CT}, of 39.15 per cent. This value is much higher than inter-population diversity components obtained among human continental groups for protein polymorphisms (Lewontin 1972), the human major histocompatibility complex HLA (Sanchez-Mazas 2007) or DNA markers (Barbujani *et al*. 1997; Excoffier and Hamilton 2003): it is of the order of 10 to 15 per cent for most of them. This difference may be due to the fact that the GM polymorphism is tested by serological typing, thereby providing only a broad description of its molecular variation. As a consequence, the frequencies of the most common haplotypes in each geographic area are ~~generally~~ probably overestimated, leading to an increased estimation of intergroup variation. However, GM haplotypes, like chromosome Y haplogroups (Underhill *et al*. 2000), are groups of phylogenetically related haplotypes which provide clear genetic information on worldwide and continental genetic differentiation, while more detailed molecular information is often more difficult to interpret. Using GM haplotypes, previous studies have shown that population genetic differentiations can be correlated to geography on a global scale (Sanchez-Mazas and Langaney 1988; Dugoujon *et al*. 2004), and to linguistic differentiation in some continental areas, like sub-Saharan Africa (Excoffier *et al*. 1987, 1991). GM is thus likely to reveal relevant information for human peopling history studies.

; Sanchez-Mazas and Poloni 2008

Populations tested and methods used

In this study, we contribute original GM data for eight aboriginal populations from Taiwan: the Siraya, Pazeh, Taroko (a group within Sedik), Atayal, Tsou, Bunun, Puyuma, and Yami. We also typed a Chinese population from Taiwan (the Minnan), an extra-Formosan population from Bataan Islands located between Taiwan and the Philippines, the Ivatan, whose language is closely related to that of the Yami, and a 'hill tribe' from Northern Thailand, the Akha (Sino-Tibetan). All samples were collected by Marie Lin and tested by a hemagglutination-inhibition technique for allotypes G$_1$M(1, 2, 3, 17), G$_2$M(23) and G$_3$M(5, 6, 13, 15, 16, 21) by L.O. in Novosibirsk. Data on the Ami were already available in the literature (Schanfield 1971; Schanfield *et al*. 2002). However, as no test for G$_2$M(23) was performed by these authors, we decided to collect an additional sample of 28 Ami individuals to check the presence of this allotype. Typings done by J.M.D. in Toulouse revealed that all individuals were G$_2$M(23) positive.

We estimated GM haplotype frequencies from complex phenotypes by a standard maximum-likelihood procedure, using the computer program GENEF. Consistent hypotheses on GM haplotype distributions were obtained for all populations except Ivatan (Table 13.1). For this population, however, satisfactory results were obtained when we estimated the frequencies of the GM broad haplotypes, i.e. not discriminated by the presence/absence of allotype G$_2$M(23) (see below). The hypothesis of Hardy-Weinberg equilibrium (HWE) was tested by chi-square. P-values were above 1 per cent (i.e. non significant) for all populations except Akha and Ivatan (Table 13.1). However, in these populations, the high chi-square value obtained for HWE test is due to two phenotypes with expected values lower than five individuals,[2] and the null hypothesis of HWE cannot be rejected on this basis. We can thus regard all populations tested as in HWE equilibrium.

We compared the new Taiwanese samples to data published in the literature: first to other Taiwanese, and second to a large database of East Asian populations tested for GM (*GeneVA* database maintained by ASM). As many populations were not typed for the allotype G$_2$M(23), we performed our analyses according to two different levels of genotypic resolution: a 'high resolution level', by including the information brought by G$_2$M(23), for a small set of populations (15), and a 'low resolution level', by ignoring the information contributed by this allotype, for a large set of populations (113). The Ivatan sample, for which G$_2$M(23) typing appeared to be problematic, was only used in the latter analyses. For the sake of clarity, we will talk of *GM broad haplotypes*, when we do not consider the information brought by G$_2$M(23) typings, and of *GM±23 sub-haplotypes*, when we consider that information.

Table 13.1 GM frequencies in Aboriginal populations from Taiwan

	Ahka	*Minnan*	*Pazeh*	*Siraya*	*Toroko*	*Atayal*	*Bunun*	*Tsou*	*Puyuma*	*Yami*	*Ami*		*Ivatan**
N	*41*	*50*	*74*	*52*	*56*	*60*	*49*	*50*	*52*	*69*	*28*		*48*
1,3;23;5*	0.406	0.5231	0.6492	0.7045	0.7267	0.75	0.7592	0.81	0.8553	0.971	1	1,3;0;5*	0.8959
1,3;-;5*	0.1672	0.0367	0.1683	0.1128	0.0323	0	0.0875	0	0.0774	0	0		
1,3;23;5	0	0.0202	0	0	0	0	0	0	0	0	0		~~0~~
A ~~1,17;21~~	0.1737	0.2668	0.1002	0.1019	0.1786	0.175	0.0714	0.18	0.026	0.029	0	1,17;;21	0.0625
1,17;23;21	0	0	0	0	0	0	0	0	0	0	0		
A ~~1,2,17;21~~	0.1046	0.1032	0.0417	0.052	0.0089	0.0083	0.0507	0	0.0221	0	0	1,2,17;;21	0.0104
A ~~1,17;5*~~	0	0.01	0	0	0	0	0	0.01	0.0096	0	0	1,17;;5*	0.0208
1,17;23;5*	0	0	0	0	0	0	0	0	0	0	0		
B ~~1,2,17;5*~~	0	0.01	0	0.0096	0	0	0.0105	0	0.0096	0	0	1,2,17;;5*	0.0104
B ~~1,17;10,11,13,15,16~~	0.0915	0.03	0.0203	0.0192	0	0	0	0	0	0	0		~~0~~
~~1,2,17;10,11,13,15,16~~	0	0	0	0	0	0.0083	0	0	0	0	0		~~0~~
1;23;21	0	0	0	0	0.0297	0.0584	0	0	0	0	0		~~0~~
~~1,2;15,16,21~~	0	0	0	0	0.0089	0	0	0	0	0	0		~~0~~
C ~~1,17~~	0.0448	0	0	0	0	0	0	0	0	0	0		~~0~~
C ~~1,17;16~~	0	0	0.0135	0	0	0	0	0	0	0	0		~~0~~
1,3;23;21	0.0122	0	0.0068	0	0.0149	0	0.0207	0	0	0	0		~~0~~
Number of haplotypes	7	8	7	6	7	5	6	3	6	2	1		5
HW-chi-2	42.82	6.86	9.29	22.08	25.02	4.11	5.83	2.75	14.91	0.06	n.a.		52.04
dl	14	19	17	12	14	8	10	3	10	1	n.a.		6
P-value	0.0001	0.9949	0.9307	0.0366	0.0344	0.8471	0.8293	0.4318	0.1354	0.8065	n.a.		0.0000
Gene diversity (h)	0.765	0.648	0.541	0.482	0.442	0.407	0.412	0.315	0.264	0.057	0.000		0.195

* Tests for GM(23) are not taken into account (see text); N: sample size; 5* stands for 5,10,11,13,14; n.a.: not available

A 1,17;-;21
A 1,2,17;-;21
A 1,17;-;5*

1,2,17;-;5*
B 1,17;-;10,11,13,15,16
B 1,2,17;-;10,11,13,15,16

1,2;-;15,16,21
1,17;-;-
C 1,17;-;16

Genetic structure of Taiwanese populations

GM frequencies

Previous global and continental surveys of the GM polymorphism (Steinberg and Cook 1981; Blanc *et al*. 1990; Dugoujon *et al*. 2004; Poloni *et al*. 2005) as well as statistics based on more than 450 population samples collected in our GM database indicate that four GM broad haplotypes are common in East Asia: GM*1,3;;5*, mostly observed in Southeast Asia and the Pacific, where it can reach frequencies of about 90 per cent; GM*1,17;;21 and GM*1,2,17;;21, commonly found in all continents, but more frequent in Northeast Asia, and even more so in Amerindians; and Gm*1,17;;10,11,13,15,16, which reaches its highest frequencies, of the order of 30 per cent, in Northeast Asia. Based on a subset of 96 population samples tested for G$_2$M(23), we also know that, among GM±23 sub-haplotypes, GM*1,3;23;5* is generally much more frequent than GM*1,3;-;5*, although both sub-haplotypes are commonly observed in East Asia and the Pacific; on the other hand, GM*1,17;23;21 and GM*1,2,17;23;21 are rare compared to GM*1,17;−;21 and GM*1,2,17;−;21, respectively, and GM*1,17;23;10,11,13,15,16 has never been observed.

The present study provides the first description of GM±23 sub-haplotypes in Taiwan and Lan-Yu. Their frequencies are shown in Plate 13.Ia[3] for 10 populations. GM*1,3;23;5* is the most frequent, as in the majority of East Asian and Pacific populations studied so far. The frequency of this haplotype is much higher in Taiwan aborigines than in the Minnan population (0.523). However, it also varies greatly among the former, from 0.649 in the Pazeh, to 0.855, in the Puyuma, and it is even higher in Yami from Lan-Yu (0.971). The new Ami sample tested by us was found to be monomorphic for GM*1,3;23;5*. However, as only 28 individuals were typed in this population, GM*1,3;23;5* frequency is probably closer to 0.95, as estimated for the broad GM haplotype GM*1,3;;5* in the two Ami samples tested by Schanfield (1971; Schenfield *et al*. 2002). GM*1,3;−;5* also exhibits heterogeneous frequencies among Taiwanese: it reaches 0.168 in the Pazeh, while it is not observed in Atayal, Tsou and Yami, and has a low frequency in Minnan (0.037). The third common sub-haplotype is GM*1,17;−;21, with the highest frequency in Tsou (0.180) and the lowest in Puyuma (0.026), while GM*1,2,17;−;21 is always observed at low frequencies (< 5 per cent) in Taiwan Aborigines. These two latter sub-haplotypes are much more frequent in the Minnan (0.267 and 0.103, respectively).

Taiwanese populations also exhibit many uncommon sub-haplotypes: GM*1,17;−;5* (≤ 1 per cent), in Puyuma and Tsou; GM*1,17;−;10,11,13,15,16, in Siraya and Pazeh (≤ 2 per cent); GM*1,3;23;21 (≤ 2 per cent), in Pazeh, Taroko, and Bunun, GM*1,2,17;−;5* (≤ 1 per cent), in Puyuma, Siraya, and Bunun; and GM*1,2,17;−;10,11,13,15,16 (≤ 1 per cent), in Atayal. Very unusual variants have also been considered to explain some rare phenotypes: GM*1;23;21, in Taroko and Atayal; GM*1,17;−;16 in Pazeh, and GM*1,2;−;15,16,21 in Taroko. The frequency of these variants are below 2 per cent, except for GM*1;23;21 (3 per cent in Taroko and 6 per cent in Atayal). An unusual variant is also observed in Minnan (GM*1,3;23;5).

Gene diversity

As indicated in Table 13.1, Taiwanese aborigines exhibit heterogeneous but generally low levels of gene diversity (here estimated by the expected heterozygosity *h*). The lowest *h* values are found in the Ami (0 when estimated in our Ami sample of 28 individuals, but 0.091–0.094 when estimated in the Ami samples tested by Schanfield and co-workers) and the Yami from Lan-Yu (0.057). Gene diversity is also rather low in other southern populations, like the Puyuma (0.264). We thus note a general decrease in heterozygosity from north to south when we plot this statistic against latitude (Plate 13.Ib), although the estimated correlation coefficients (r = 0.517, when we include the Yami, and r = 0.268, when we do not include them) are not significant. Actually, gene diversity is highest on the west coast (Pazeh, Siraya), and lowest on the south-east coast (Puyuma, Ami) and Lan-Yu (Yami), decreasing along a northwest to southeast direction. It corresponds to a gradual increase of the frequency of one haplotype, GM*1,3;23;5,10,11,13,14, to the detriment of the less frequent ones, meaning that all genetic profiles may be derived continuously from one another. However, we do not have samples from populations of the northeast coast such as the Kavalan. We are therefore not claiming that the pattern of decreasing diversity from north to south is found among populations of the Formosan east coast.

Genetic distance analyses

We performed a multidimensional-scaling analysis (or MDS) among the 10 Taiwanese populations tested in this study (including the Yami) on the basis of GM±23 sub-haplotypes frequencies (the highest level of resolution). To investigate the genetic relationships between Formosan aborigines and Chinese, we also included five mainland Han populations which had been tested for $G_2M(23)$ (Figure 13.1). Links were drawn between some populations to indicate the cases where genetic differentiation between them is not sustained statistically (at the 5 per cent level). We used two different tests, an exact test and a permutation test on pairwise F_{ST}s. The exact test is more powerful; i.e. in the present case, its power to detect genetic differentiation when such differentiation is true is higher; on the other hand, there is also a higher risk of erroneous detection of a differentiation. The permutation test is more conservative; i.e. in the present case, it will detect genetic differentiation only when the statistical result is clearly unambiguous; on the other hand, there is also a higher risk of missing true cases of genetic differentiation. Therefore, the solid lines drawn on the graph between some populations indicate the cases where populations are considered as undifferentiated genetically according to both tests, while the dotted lines have to be treated more cautiously as they represent cases where genetic differentiation can be concluded from the exact test, but not the permutation test.

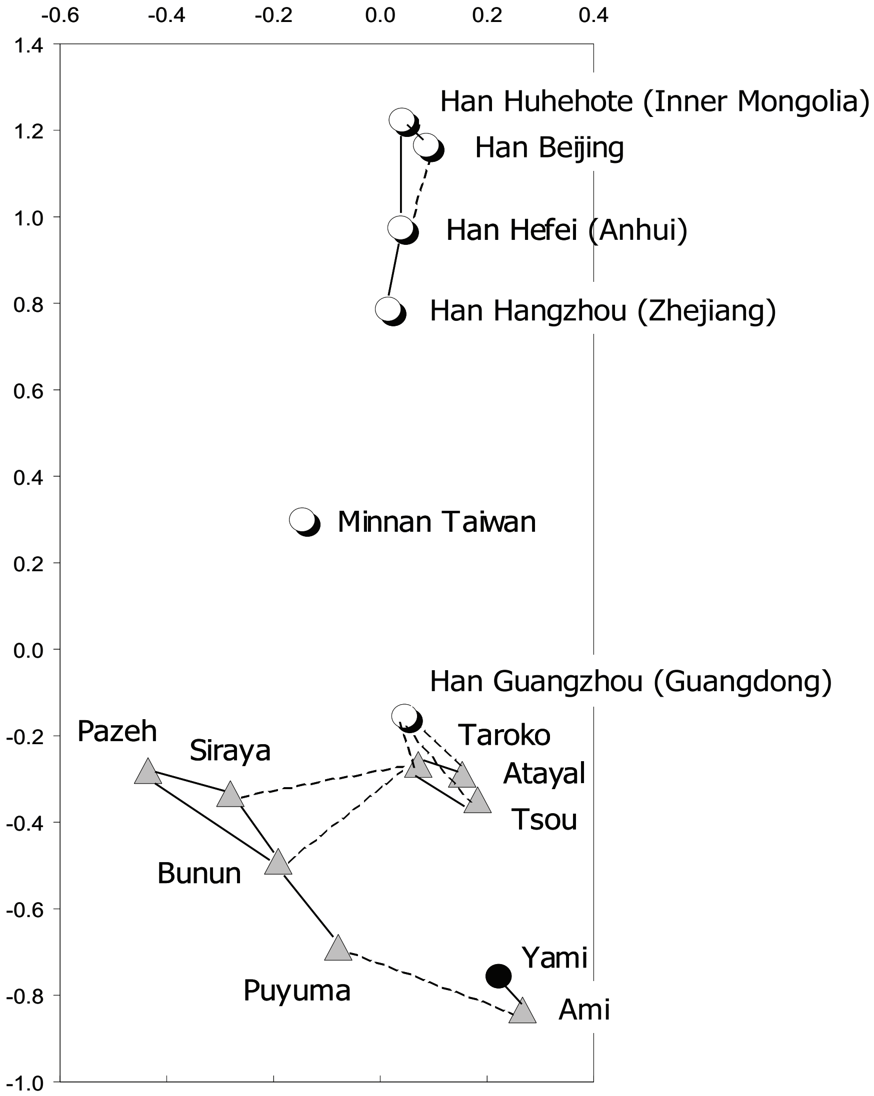


Figure 13.1 Multidimensional scaling analysis among 15 populations from Taiwan and China based on the frequencies of GM ± 23 sub-haplotypes. Prevosti's *et al*. (1975) distances have been used. Solid lines link populations that are undifferentiated at the 5% level according to an exact test of population differentiation. Dotted lines link populations that are undifferentiated at the 5% level according to a permutation test on pairwise Fst's, but not to the exact test. Final stress = 0.041 (excellent). White circles: Sino-Tibetans; triangles: Austronesians from Taiwan; black circle: Extra-Formosan Austronesians.

The MDS (characterized by a very good stress value of 0.041) segregates four Northern and Central Han populations at the top, the Minnan from Taiwan in a central position, and the Southern Han from Guangzhou, the eight Austronesian populations from Taiwan and the extra-Formosan Yami at the bottom. The four Northern and Central Han populations are genetically linked to each other according to a pattern recalling their relative geographic position. The Minnan are genetically intermediate between the Northern and the Southern Han but are very distant from both groups and significantly differentiated from all populations. Within the last group, we observe three clusters of genetically related populations: the Atayal, Taroko and Tsou; the Pazeh, Siraya, Bunun and Puyuma; and the Ami and Yami. In addition, we find that the Southern Han from Guangzhou are genetically very close to the Taroko, Atayal and Tsou, from whom they are not differentiated significantly according to the permutation statistical test (dotted lines). Genetic relationships are also suggested between the Taroko and both the Siraya and Bunun, and between the Puyuma and the Ami according both to the permutation test and to their relative position in the MDS.

The results of this analysis indicate a genetic continuity from the populations of the west coast of Taiwan (Pazeh, Siraya) to those of the southeast coast (Puyuma), the South-Central Bunun being in between. This pattern further extends to the Ami, located on the east coast, and to the extra-Formosan Yami on Lan-Yu. It corresponds to a marked decrease in gene diversity (Plate 13.Ib). On the other hand, the northern Atayal and Taroko and the south-central Tsou form a homogeneous cluster, relatively separated from the former, again with a decreasing level of genetic diversity from the Atayal and Taroko to the Tsou (Plate 13.Ib). A reduction in gene diversity is generally explained by genetic drift. These observations, then, suggest that genetic differentiations occurred, on the one hand, from the west coast to the southeast coast of Taiwan via a south-central region (Bunun location), and, on the other hand, from the north of Taiwan southwards (to Tsou location) in the mountainous central regions of the island (but see another interpretation below). Moreover, it is worth noting that the northern Taiwanese group is much closer genetically to the Han from Guangdong Province in southern China than to the Minnan from Taiwan. Actually, this is true for all Taiwanese tribes, suggesting a genetic relationship between aboriginal populations from Taiwan and southeastern Chinese.

Comparison to other East Asian populations

We pursued our study by comparing the Taiwanese populations to other populations from East and Southeast Asia. Those data were gathered from the literature and were already used in our previous work (Poloni *et al*. 2005). We excluded all populations located in the northern part of East Asia to keep only those speaking languages belonging to the Sino-Tibetan, Tai-Kadai, Austroasiatic, Hmong-Mien and Austronesian linguistic families. A total of 113 populations representing 15,904 individuals were compared. Unfortunately, these comparisons were only done by using the frequencies of broad GM haplotypes (thus with a lower level of resolution than for the analysis presented in Figure 13.1) because $G_2M(23)$ was most often not tested in those populations.

We performed a first MDS (Plate 13.II) including the whole set 113 populations (a good stress of 0.085 is obtained). The graph is basically similar to the MDS among 102 populations presented in our former work (Poloni *et al*. 2005: figure 15.3), except that it also incorporates the 11 populations tested in the present study (all new samples but the Ami, due to its low sample size),[4] and that it uses a different genetic distance. Indeed, we computed the distances of Prevosti (*et al*. 1975) rather than Reynolds (*et al*. 1983), because the former generally discriminates the populations better than the latter, while both of them produce a similar differentiation pattern (personal results). Using Prevosti's distance is thus more useful when numerous populations are genetically very close, like here.

As observed by Poloni *et al*. (2005), many populations tend to group together according to their linguistic affiliation (Northern Tibeto-Burman; Northern Chinese; Wu and Southeastern Mandarin; and other Tibeto-Burman), although without any discontinuity between them; on the other hand, substantial overlapping is observed for the other groups.

We thus performed a second MDS by keeping only those 63 populations speaking Southern Chinese dialects, Austroasiatic, Tai-Kadai, Hmong-Mien and Austronesian languages (Figure 13.2). Despite a general overlapping of most groups (a less good but still acceptable stress of 0.135 is here obtained), some results are worthy of mention: first, all Southern Chinese are displayed at the top of the plot, all Tai-Kadai at the bottom-left, and all Austronesians (including both Taiwanese and Extra-Formosan) at the bottom right with a few exceptions corresponding to very isolated populations (one of the Aetas and the Mamanwas) who have probably undergone rapid genetic drift. Those three groups are thus relatively well discriminated from each other, despite their very different levels of genetic heterogeneity. This is harder to say for the Hmong-Mien, only represented by three populations, and the Austroasiatic-speakers, which are very dispersed due to their high level of inter-population diversity (Poloni *et al*. 2005). Secondly, among the Chinese populations, those located in the Guangdong province (except in one case) are closer genetically to the Southeast Asian groups (Tai-Kadai, Austronesian, Hmong-Mien and Austroasiatic) than those from other regions. On the other hand, the Chinese from Taiwan (here represented by several populations) are genetically closer to populations from regions further inland in China, such as Fujian, Jiangxi and Hunan, and, to a lesser extent, Sichuan. Finally, Taiwan aborigines share distinct relationships with non-Taiwanese populations: in particular, this confirms the general pattern in Figure 13.1 where the Ami (here represented by two samples) and Puyuma are the most distant from the southern Chinese. The Saisiat, who were not tested for $G_2M(23)$ and thus not represented in Figure 13.1, cluster with the Siraya and Pazeh. Also, the Ami and the Puyuma are very close to the Extra-Formosan Austronesians from Lan-Yu (Yami), the Philippines (Ivatan) and Borneo (Kadazan).

In Figure 13.3, we also represented the average genetic distance between each aboriginal population from Taiwan and the populations belonging to

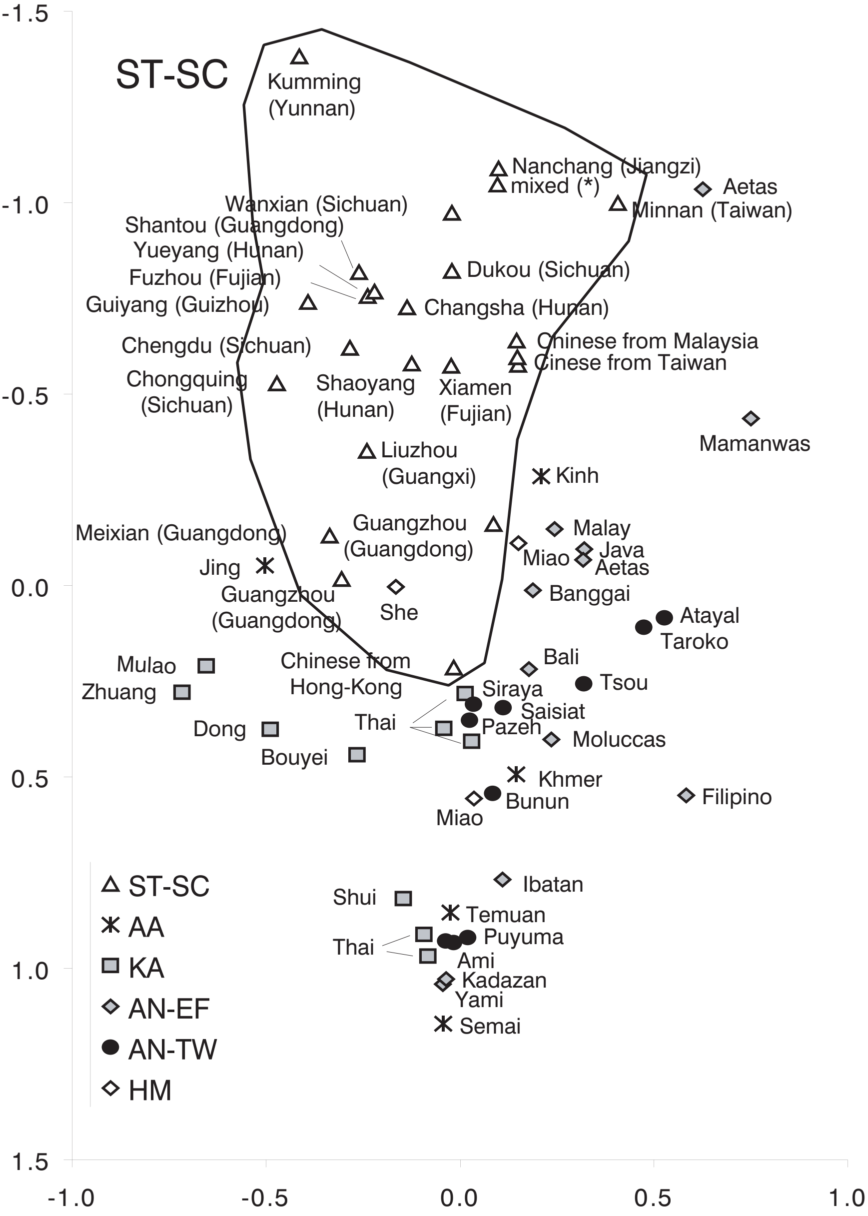


Figure 13.2 Multidimensional scaling analysis among 63 populations from Southeast Asia excluding all Sino-Tibetans but the Southern Chinese (grouped). Prevosti's et al. (1975) genetic distances have been computed from the frequencies of GM broad haplotypes. Final stress = 0.135 (fair). The 'mixed' sample is composed of individuals from Fujian, Zhejiang, Jiangxi, and Hunan. ST-SC: Sino-Tibetan, Southern Chinese (Min, Xiang, Gan, Hakka, Yue); AA: Austroasiatic; KA: Tai-Kadai; AN-EF: Extra-Formosan Austronesian; AN-TW: Autronesian from Taiwan; HM: Hmong-Mien.

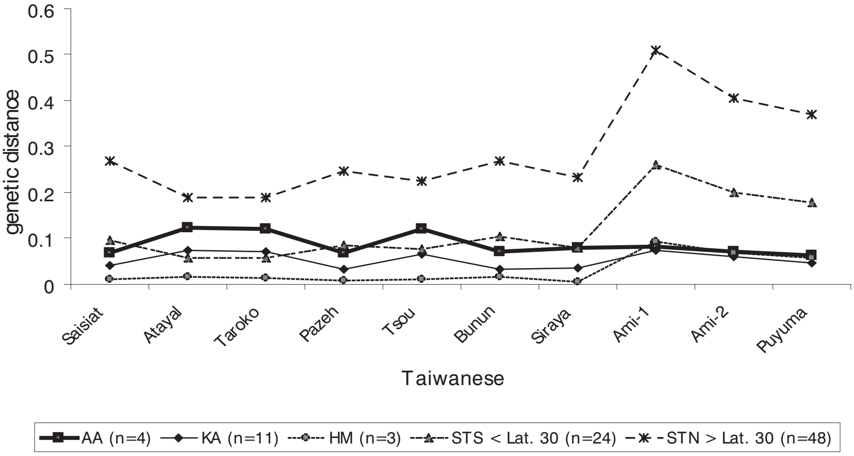


Figure 13.3 Average Reynold's *et al*. (1983) genetic distances between each aboriginal population from Taiwan tested in this study (except Ami-1, Ami-2 and Saisiat typed by Schanfield *et al*. 1971, 2002), and the populations belonging to different linguistically defined groups of Southeast Asia (with their number given in parentheses). The Taiwanese populations are ordered according to their latitude. AA: Austroasiatic; KA: Tai-Kadai; HM: Hmong-Mien; STS: Southern Sino-Tibetan, located below latitude 30ºN; NTS: Northern Sino-Tibetan, located above latitude 30ºN.

different linguistically defined groups of Southeast Asia (Austroasiatic, Tai-Kadai, Hmong-Mien, Southern Sino-Tibetan (populations located below latitude 30°N) and Northern Sino-Tibetan (populations located above latitude 30°N). This graph confirms that Taiwanese aboriginals are closer to southern than to northern Chinese and that Ami and Puyuma are the most distant from both of them. In addition, we note that, on average, Ami and Puyuma are at equal distance from all other groups, i.e. Austroasiatic, Tai-Kadai and Hmong-Mien, while all other Taiwanese are slightly closer to Hmong-Mien than to Austroasiatic, the Tai-Kadai (and, often, the southern Chinese) lying in between. However, such differences may not be significant.

Discussion

The present study provides a new and appreciable examination of the genetic diversity of Taiwanese populations thanks to the complete analysis of the GM polymorphism, only partially tested before on some populations of that island. The results reveal an important heterogeneity among its aboriginal populations: there are marked differences in their internal gene diversity, in their genetic distances from each other, and in their genetic relationships to other East Asian populations.

First, in agreement with studies of other genetic markers (Lin and Broadberry 1998; Schanfield *et al*. 2002; Sewerin *et al*. 2002; Sanchez-Mazas *et al*. 2005), the Ami are very unusual for GM, as they are almost monomorphic for the predominant haplotype in East Asians, GM*1,3;23;5*. This implies that the Ami, while being today the largest aboriginal group with about 140,000 people (Chu 1997), underwent an extreme bottleneck, losing their genetic diversity from an original gene pool where GM*1,3;23;5* was already very frequent. Such a loss of diversity is compatible with observations in a previous HLA study (Sanchez-Mazas *et al*. 2005) where we emphasized that only seven HLA-DRB1 alleles were detected in that population compared to an average of 20.6 in 48 other East Asian populations.

The uniqueness of the Ami has been discussed by other authors in relation to either natural selection or affinities to other, more distant, population groups. The idea that GM*1,3;;5* would be advantageous in an environment endemic to malaria has been invoked to explain the high frequency of this haplotype in lowland areas like the east coast of Taiwan (Schanfield *et al*. 2002). In relation to another region of the world, i.e. Sardinia, a similar explanation was formerly given for allotype $G_2M(23)$, which was then considered to be selectively advantageous (Piazza *et al*. 1976). However, no other evidence has ever corroborated these hypotheses. On the contrary, many counter-examples can be given where such a correlation is not verified. For example, other plains tribes of Taiwan, like the Pazeh and Siraya, have lower frequencies of GM*1,3;23;5*; we also know that $G_2M(23)$ is very rare in sub-Saharan Africa (Blanc *et al*. 1990), where malaria is endemic. Therefore, we do not have enough evidence to accept the hypothesis of this type of selection acting on the GM system.

Nor does our knowledge of the GM system support the idea that Ami are related to more distant population groups. Although some HLA alleles and haplotypes shared by Ami and Australian Aborigines or Papuans could be taken as a possible evidence of a remote relationship between Taiwanese and an ancient peopling of Sundaland (Lin *et al*. 2005~~; Sanchez-Mazas *et al*. in preparation~~), Australians, Melanesians as well as Amerindians are characterized by completely different GM genetic profiles, where GM*1,12;21 or GM*1,17;5* or GM*1,2,17;21 are the most frequent haplotypes, and GM*1,3;5* is either rare or not observed (Steinberg and Cook 1981; Sanchez-Mazas and Pellegrini 1990). Therefore, the genetic profile observed in the Ami is undeniably an extreme differentiation of a Southeast Asian genetic pool. Moreover, the results shown in Plates 13.I and 13.II and Figure 13.1 reveal a correlation between genetic and geographic proximity: the Ami are genetically related to the Puyuma, a geographically close population located in the south of Taiwan, the Yami, just situated off the southeast coast of Taiwan, the Ivatan, living a little bit further south in the Philippines, and the Kadazan from Borneo. Interestingly, the Puyuma, while being in genetic continuity with the Ami, exhibit many rare haplotypes in addition to their very high GM*1,3;23;5 frequency. The Ami thus probably underwent a extreme bottleneck either when they reached the southeast of Taiwan or later, a hypothesis already proposed in a previous work based on an analysis of the HLA-DRB1 polymorphism (Sanchez-Mazas *et al*. 2005).

As opposed to the Ami, two groups of Taiwanese populations are characterized by a high level of gene diversity, but distinct genetic profiles. On one hand, the Pazeh, Siraya and Bunun (plus Saisiat, when only broad GM haplotypes are considered) exhibit several haplotypes with middling frequencies, i.e. GM*1,3;–;5*, GM*1,2,17;21, and GM*1,17;10,11,13,15,16. We observe a genetic continuity between these populations and the Puyuma and Ami, with a loss of diversity in the latter. On the other hand, the Atayal, Taroko and Tsou show GM*1,2,17;21 at a relatively high frequency, in addition to GM*1,3;23;5*. Atayal and Taroko populations also share a rare haplotype, GM*1;23;21, while the Tsou, with only two haplotypes, exhibit a much lower genetic diversity (Plate 13.Ib).

The two groups of populations described above could represent the descendents of an initial split of Austronesian populations into two primary branches: Atayal-Tsouic versus the rest of Austronesian. However the chances that the similarity we observe between speakers of Atayalic languages (Atayal and Taroko) and the Tsou are due to independent parallel developments are significant, because the Tsou basically have only two haplotypes, which are also present, albeit at lower frequencies, among the Siraya and other populations. The location of the Tsou in the midst of Taiwan's central mountain range is conducive to genetic drift, which could also account for the simplification of a Siraya-type profile into a Tsou-type profile (see below).

It is often argued (Lee *et al*. 2002) that the 'plains tribes' of Taiwan, i.e. those living in the west coast (Pazeh, Siraya), have been 'sinicized', that is, underwent high levels of gene flow from neighbouring Chinese people, this explaining their greater proximity to the latter ('gene flow hypothesis'). However, our study does not show any peculiar genetic relationship between the 'plains tribes' and the local Chinese, who began settling the island more than three centuries ago. First, the Minnan are only distantly related to the Taiwanese aborigines, whether from the plains or the mountains (Figures 13.1 and 13.2). Second, although Siraya and Pazeh (i.e. 'plains tribes') lie closer to some populations from South China (Hong Kong) according to our low-resolution analysis (Figure 13.2), this is not true when we consider GM typings at high resolution (which includes G_2M23 typings), where Taroko and Atayal (i.e. 'mountain tribes') are closer than the 'plains tribes' to the southern Chinese from Guangdong Province (Figure 13.1). Moreover, the level of genetic diversity is not different in the 'plains' and 'mountain tribes' mentioned above (Plate 13.Ib), while one would expect the 'plains tribes' to be more diversified under an admixture hypothesis. Different levels of gene flow from the Chinese do not explain the observed GM diversity.

Having excluded natural selection and Chinese gene flow, the only explanation left is that the GM genetic patterns observed in Taiwan are mainly the signal of its ancient peopling history. Two questions may then be asked: first, is it possible to define a geographic origin for the Taiwanese Austronesians? And, second, what have we learnt about Taiwanese differentiation?

Based on our analyses at a continental scale, we have shown that Taiwanese are on average almost equally distant from the Hmong-Mien, Tai-Kadai, south Chinese and even Austroasiatic population groups (except the Ami and Puyuma who are genetically far from the Chinese) and that most Southeast Asian groups overlap (Figures 13.3 and 13.2, respectively). Therefore, due to the low resolution of our data, we cannot answer the first question. However, the MDS presented in Figure 13.2 indicates that all Southeast Asian populations, including Austronesians from Taiwan and most Extra-Formosans, appear to be closer to Chinese populations from Hong Kong, located near the Pearl River delta in southerly Guangdong Province, while the Minnan (and other ethnically undefined Chinese from Taiwan) share a similar genetic profile with Chinese populations from Fujian (across from Taiwan on the mainland) and provinces further inland like Jiangxi, Hunan and even Sichuan. Thus, if we accept, like most scholars today, the theory of a Chinese origin of Austronesians, the general proximity of Formosans to Hong Kong Chinese could support an origin of Formosan populations in the Pearl River delta region, as argued by Tsang (2005) on the basis of similarity in pottery types; but it is also compatible with an origin on the coast of Fujian (Chang 1969) assuming that the present-day population of Fujian reflects Chinese expansion in the 1st millennium BCE (incorporating a high level of genetic diversity through gene flow with populations from other regions and/or linguistic groups), while the present-day population of Guangdong more faithfully reflects the genetic make-up of the early southeast China coast at the time of the Austronesian movement to Taiwan. A passage from Fujian to northwestern Taiwan is advocated by Sagart (Chapter 5, this volume) on the grounds that (a) the sea crossing is shortest there; (b) Taiwan is visible from the Fujian side; (c) the Austronesian languages of Northwestern Taiwan are the most diverse and show none of the post-Proto-Austronesian innovations in the numeral system. In our study, that the populations of northwestern Taiwan are the most diverse in the island fits better with a northwestern passage than with a southwestern, which Tsang's hypothesis of a Pearl River delta origin implies.

As mentioned above, the first peopling of Taiwan was eventually followed by an early differentiation into two main groups: a northern group from whom Taroko and Atayal (and maybe Tsou) differentiated, and a western group including the other populations (Pazeh, Siraya, Bunun, Puyuma and Ami). An attractive explanation to account for the variability observed in the western group is that populations differentiated from northwestern to south and southeastern regions, losing their genetic diversity as they moved gradually along the west, south and east coasts, and/or partly through central regions (where the Bunun currently live). Such a scenario is in agreement with the linguistic theory proposed by Sagart (2005; and Chapter 5, this volume), although in Sagart's thesis Tsou is a southern language whose sisters are Bunun, Rukai and Kavalan, without any particular proximity to Atayalic. However, we propose that the Tsou underwent rapid genetic drift due to their isolation in the central mountains of Taiwan; their genetic closeness to Atayal and Taroko would then be an accidental convergence due to an increase of the GM*1,17;–;21 haplotype frequency (the second most frequent after GM+1,3;23;5,10,11,13,14) and the loss of all rarer haplotypes through genetic drift from any southern genetic pool. Indeed, other 'classical' and HLA polymorphisms do not reveal such a relationship (Lin *et al*. 2005, and personal results not shown) and results based on DNA data indicate an intense bottleneck for the Tsou (Chen *et al*. 2001) and/or the Atayal (Chen *et al*. 2001; Shepard *et al*. 2005).

Finally, our approach does not allow us to propose any time depth for the peopling of Taiwan. Based on mtDNA lineages, Trejaut *et al*. (2005; and Chapter 14, this volume) support the idea that 'the prevalence of the gene pool of aboriginal Taiwanese initiated from late Pleistocene settlers'. Genetic evidence for ancient settlement is also sustained by Hill *et al*. (2007) from mtDNA analyses, although the two research groups do not agree on the role of Taiwan in the spread of Austronesians towards the Pacific. According to Trejaut *et al*. (Chapter 14, this volume), the origin of the Austronesian migration can be traced back to Taiwan (or to a wider region encompassing Taiwan, the Philippines and Borneo), while Hill *et al*. (2007) claim that a mid-Holocene migration from that region is demographically minor. It is worth recalling, here, besides the caution to consider when using molecular dating (see Blench, Ross and Sanchez-Mazas, Introduction, this volume), that mtDNA is maternally transmitted; therefore, those studies reconstruct the phylogeny of female lineages only, and the demographic history of men may have been quite different than that of women.

To our view, there is no reason to exclude *a priori* a contribution of Palaeolithic populations to the present Taiwanese genetic diversity, as archaeological excavations indicate that Taiwan was settled by modern humans at least 15,000 years ago and probably as early as 30,000 years ago based on the remains of the Changpin culture (Sung 1978). A possible introgression of ancient mtDNA lineages to the Taiwanese genetic pool, with an eventual origin in Island Southeast Asia, as suggested by some mtDNA shallow lineages (Trejaut *et al*., Chapter 14, this volume), does not contradict the hypothesis that most of the GM variability currently observed in Taiwan was shaped during the Neolithic. A similar situation occurs in Europe, where demic diffusion of Neolithic farmers explains a high percentage of the variability observed for classical and DNA genetic markers in Europe, where early settlements by modern humans in the Palaeolithic are well-documented (Ammerman and Cavalli-Sforza 1984; Cavalli-Sforza *et al*. 1994; Chikhi *et al*. 1998, 2002), despite some debate on that subject (Barbujani and Chikhi 2006). Likewise, our results fit the model of an Austronesian expansion from Taiwan to the Philippines and Indonesia in relatively recent times, as the progressive loss of genetic diversity towards the south would be more difficult to reconcile with the main thrust of peopling coming from the south, as advocated by some geneticists (Oppenheimer and Richards 2002; Hill *et al*. 2007).

Acknowledgements

This work was supported by grant # 3100A0-112651 of the Swiss National Foundation (FNS) to A.S.M. and by the French CNRS OHLL (Origine de l'Homme, du Langage et des Langues) action to E.S.P. and L.S. We express our warmest thanks to Professor An-Vu-Trieu for providing Khmer samples from Viet Nam.

Notes

1 http://www.ethnologue.com.

2 The maximum-likelihood procedure did not converge for this population, meaning that no satisfactory GM haplotype frequency distribution was obtained.

3 To simplify, we represented the Minnan at the top of the island, but in reality, the Minnan occupy a very large area (see Map 13.I).

4 In the present analysis, the same samples as in Poloni *et al*. (2005) have been considered, except that we have added the 11 populations tested for GM in the present study, plus one Khmer sample from Viet Nam provided by An-Vu-Trieu and tested for GM by J.M.D, plus two Taiwanese samples (Ami and Saisiat) tested by Schanfiels *et al*. (2002) and that we have excluded three Taiwanese samples tested by Matsumoto *et al*. (1972) due to very low resolution typing.

References

Ammerman, A.J. and Cavalli-Sforza, L.L. (1984) *The Neolithic Transition and the Genetics of Populations in Europe*, Princeton, NJ: Princeton University Press.

Barbujani, G. and Chikhi, L. (2006) 'DNAs from the European Neolithic', *Heredity*, 97: 84–5.

Barbujani, G., Magagni, A., Minch, E. and Cavalli-Sforza, L.L. (1997) 'An appointment of human DNA diversity', *Proceedings of the National Academy of Science USA*, 94: 4516–19.

Bellwood, P. (1978) *Man's Conquest of the Pacific: The Prehistory of Southeast Asia and Oceania*, Oxford: Oxford University Press.

Bellwood, P. (1997) *Prehistory of the Indo-Malaysian Archipelago*, Honolulu, HI: University of Hawaii Press.

Bellwood, P. (2005) 'Examining the farming/language dispersal hypothesis in the East Asian context', in L. Sagart, R. Blench and A. Sanchez-Mazas (eds) *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, pp. 17–30, London and New York, NY: RoutledgeCurzon.

Blanc, M., Sanchez-Mazas, A., Hubert van Blyenburgh, N., Sevin, A., Pison, G. and Langaney, A. (1990) 'Inter-ethnic genetic differentiation: Gm polymorphism in Eastern Senegal', *American Journal of Human Genetics*, 46: 383–92.

Blust, R. (1977) 'The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report', *University of Hawai'i Working Papers in Linguistics*, 9(2): 1–15.

Blust, R. (1988) 'The Austronesian homeland: a linguistic perspective', *Asian Perspectives*, 26(1): 45–67.

Blust, R. (1999) 'Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics', in E. Zeitoun and P.J.-K. Li (eds) *Selected Papers from the Eighth International Conference on Austronesian Linguistics* (Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, 1), Taipei: Academia Sinica.

Capelli, C., Wilson, J.F., Richards, M., Stumpf, M.P., Gratrix, F., Oppenheimer, S., Underhill, P., Pascali, V.L., Ko, T.M. and Goldstein, D.B. (2001) 'A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania', *American Journal of Human Genetics*, 68: 432–43.

Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1994) *The History and Geography of Human Genes*, Princeton, NJ: Princeton University Press.

Chang, K.-c. (1969) *Fengpitou, Tapenkeng and the Prehistory of Taiwan*, New Haven, CT: Yale University Press.

Chen, Y.-F., Chen, S.-J., Chen, C.-H., Chu, C.L., Hsu, M., Yeh, J.-I. and Wang, L.Y. (2001) 'A study of microsatellite loci on Y chromosome of Taiwan aborigines', International Symposium on Austronesian Cultures: Issues Relating to Taiwan.

Chikhi, L., Destro-Bisol, G., Bertorelle, G., Pascali, V. and Barbujani, G. (1998) 'Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool', *Proceedings of the National Academy of Science USA*, 95: 9053–8.

Chikhi, L., Nichols, R.A., Barbujani, G. and Beaumont, M.A. (2002) 'Y genetic data support the Neolithic demic diffusion model', *Proceedings of the National Academy of Science USA*, 99: 11008–13.

Chu, C.C., Lin, M.,Nakajima, F., Lee, H. L., Chang, S. L., Juji, T. and Tokunaga, K. (2001) 'Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines', *Tissue Antigens*, 58: 9–18.

Chu, H.L. (1997) *An Introduction to the Indigenous Culture of Taiwan*, Taipei: Charity Printing Industrial Co.

Diamond, J.M. (1988) 'Express train to Polynesia', *Nature*, 336: 307–8.

Dugoujon, J.-M., Hazout, S., Loirat, F., Mourrieras, B., Crouau-Roy, B. and Sanchez-Mazas, A. (2004) 'The GM haplotype diversity of 82 populations over the world suggests a centrifugal model of human migrations', *American Journal of Physical Anthropology*, 125: 175–92.

Excoffier, L. and Hamilton, G. (2003) 'Comment on "Genetic structure of human populations"', *Science*, 300: 1877.

Excoffier, L., Harding, R., Sokal, R.R., Pellegrini, B. and Sanchez-Mazas, A. (1991) 'Spatial differentiation of RH and GM haplotype frequencies in Sub-Saharan Africa and its relation to linguistic affinities', *Human Biology*, 63: 273–97.

Excoffier, L., Pellegrini, P., Sanchez-Mazas, A., Simon, C. and Langaney, A. (1987) 'Genetics and history of Sub-Saharan Africa', *Yearbook of Physical Anthropology*, 30: 151–94.

Gray, R.D. and Jordan, F.M. (2000) 'Language trees support the express-train sequence of Austronesian expansion', *Nature*, 405: 1052–5.

Harvey, M. (1979) *Subgroups in Austronesian*, Canberra: Australian National University.

Hill, C., Soares, P., Mormina, M., Macaulay, V., Clarke, D., Blumbach, P.B., Vizuete-Forster, M., Forster, P., Bulbeck, D., Oppenheimer, S. and Richards, M. (2007) 'A mitochondrial stratigraphy for island southeast Asia', *American Journal of Human Genetics*, 80: 29–43. Epub 16 Nov. 2006.

Lee, J.C.-I., Lin, M., Tsai, L.-C., Hsu, C.-M., Hsieh, H.-M., Huang, N.-E., Shih, R.T.-P., Wun, J.-H., Chang, J.-G., Ko, Y.-C., Tzeng, C.-H. and Linacre, A. (2002) 'Population study of polymorphic microsatellite DNA in Taiwan', *Forensic Science Journal*, 1: 31–7.

Lewontin, R.C. (1972) 'The apportionment of human diversity', in T. Dobzhansky, M.K. Hecht and W.C. Steere (eds) *Evolutionary Biology 6*, New York: Appleton-Century-Crofts.

Lin, M. and Broadberry, R.E. (1998) 'Immunohematology in Taiwan', *Transfusion Medicine Reviews*, 12: 56–72.

Lin, M., Chu, C.C., Broadberry, R.E., Yu, L.C., Loo, J.H. and Trejaut, J.A. (2005) 'Genetic diversity of Taiwan's indigenous peoples: possible relationship with insular Southeast Asia', in L. Sagart, R. Blench and A. Sanchez-Mazas (eds) *The Peopling of East Asia:*

Putting Together Archaeology, Linguistics and Genetics, pp. 230–47, London and New York: RoutledgeCurzon.

Lin, M., Chu, C.C., Lee, H.L., Chang, S.L., Ohashi, J., Tokunaga, K., Akaza, T. and Juji, T. (2000) 'Heterogeneity of Taiwan's indigenous population: possible relation to prehistoric Mongoloid dispersals', *Tissue Antigens*, 55: 1–9.

Matsumoto, H., Miyazaki, T., Fong, J.M. and Mabuchi, Y. (1972) 'Gm and Inv allotypes of the Takasago tribes in Taiwan', *Japanese Journal of Human Genetics*, 17: 27–37.

Oppenheimer, S. and Richards, M. (2002) 'Polynesians: devolved Taiwanese rice farmers or Wallacean maritime traders with fishing, foraging, and horticultural skills?', in P. Bellwood and C. Renfrew (eds) *Examining the Farming/Language Dispersal Hypothesis*, pp. 287–97, Cambridge: McDonald Institute for Archaeological Research.

Piazza, A., Loghem, E. van, Lange, G. de, Curtoni, E.S., Ulizzi, L. and Terrenato, L. (1976) 'Immunoglobulin allotypes in Sardinia', *American Journal of Human Genetics*, 28: 77–86.

Poloni, E.S., Sanchez-Mazas, A., Jacques, G. and Sagart, L. (2005) 'Comparing linguistic and genetic relationships among East Asian populations: a study of the RH and GM polymorphisms', in L. Sagart, R. Blench and A. Sanchez-Mazas (eds) *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, pp. 252–72, London and New York: RoutledgeCurzon.

Prevosti, A., Ocaña, J. and Alonso, G. (1975) 'Distances between populations of Drosophila subobscura based on chromosome arrangement frequencies', *Theoretical and Applied Genetics*, 45: 231–41.

Reid, L. (1982) 'The demise of Proto-Philippines', in A. Halim, L. Carrington and S.A. Wurm (eds) *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2, *Tracking the Travellers*, pp. 201–16, Canberra: Pacific Linguistics.

Reynolds, J., Weir, B.S. and Cockerham, C.C. (1983) 'Estimation for the coancestry coefficient: basis for a short-term genetic distance', *Genetics*, 105: 767–79.

Sagart, L. (2005) 'Sino-Tibetan-Austronesian: an updated and improved argument', in L. Sagart, R. Blench and A. Sanchez-Mazas (eds) *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, pp. 161–76, London and New York: RoutledgeCurzon.

Sanchez-Mazas, A. (2007) 'An apportionment of human HLA diversity', *Tissue Antigens*, 69, suppl.1: 198–202.

Sanchez-Mazas, A. and Langaney, A. (1988) 'Common genetic pools between human populations', *American Journal of Human Genetics*, 78: 161–6.

Sanchez-Mazas, A. and Pellegrini, B. (1990) 'Polymorphismes Rhésus, Gm et HLA et histoire de l'Homme moderne', *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, NS, 2: 57–76.

Sanchez-Mazas, A., Poloni, E.S., Jacques, G. and Sagart, L. (2005) 'HLA genetic diversity and linguistic variation in East Asia', in L. Sagart, R. Blench and A. Sanchez-Mazas (eds) *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, pp. 273–96, London and New York: RoutledgeCurzon.

Sanchez-Mazas, A., Poloni, E.S. and Sagart, L. (in preparation) 'The peopling of East Asia from the point of view of HLA markers'.

Sapir, E. (1916) 'Time perspective in aboriginal American culture: a study in method', in D.G. Mandelbaum (eds) *Selected Writings of Edward Sapir*, Berkeley and Los Angeles, CA: University of California Press.

Schanfield, M.S. (1971) 'Population Studies on the Gm and Inv antigens in Asia and Oceania', PhD thesis, University of Michigan, Ann Arbor.

Schanfield, M.S., Ohkura, K., Lin, M., Shyu, R. and Gershowitz, H. (2002) 'Immunoglobulin allotypes among Taiwan aborigines: evidence of malarial selection could affect studies of population affinity', *Human Biology*, 74: 363–79.

Sewerin, B., Cuza, F.J., Szmulewicz, M.N., Rowold, D.J., Bertrand-Garcia, R.L. and Herrera, R.J. (2002) 'On the genetic uniqueness of the Ami aborigines of Formosa', *American Journal of Physical Anthropology*, 119: 240–8.

Shepard, E.M., Chow, R.A., Suafo'a, E., Addison, D., Perez-Miranda, A.M., Garcia-Bertrand, R.L. and Herrera, R.J. (2005) 'Autosomal STR variation in five Austronesian populations', *Human Biology*, 77: 825–51.

Starosta, S. (1995) 'A grammatical subgrouping of Formosan languages', in P.J.-k. Li, D.-a. Ho, Y.-k. Huang, C.-h. Tsang and C.-y. Tseng (eds) *Austronesian Studies Relating to Taiwan* (Symposium Series of the Institute of History and Philology), Taipei: Academia Sinica.

Steinberg, A.G. and Cook, C.E. (1981) *The Distribution of the Human Immunoglobulin Allotypes*, Oxford: Oxford University Press.

Su, B., Jin, L., Underhill, P., Martinson, J., Saha, N., McGarvey, S.T., Shriver, M.D., Chu, J., Oefner, P., Chakraborty, R. and Deka, R. (2000) 'Polynesian origins: insights from the Y chromosome', *Proceedings of the National Academy of Science USA*, 97: 8225–8.

Sung, W.H. (1978) 'Prehistoric Taiwan', Proceedings of Taiwan Historical Research Society, Taipei.

Tajima, A., Sun, C.S., Pan, I.H., Ishida, T., Saitou, N. and Horai, S. (2003) 'Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese', *Human Genetics*, 113: 24–33.

Trejaut, J.A., Kivisild, T., Loo, J.H., Lee, C.L., He, C.L., Hsu, C.J., Lee, Z.Y. and Lin, M. (2005) 'Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations', *Public Library of Science (PLoS) Biology*, 3: e247. Epub 5 July 2005.

Tsang, C.-h. (2005) 'Recent discoveries at the Tapenkeng culture sites in Taiwan: implications for the problem of Austronesian origins', in L. Sagart, R. Blench and A. Sanchez-Mazas (eds) *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, pp. 63–73, London and New York: RoutledgeCurzon.

Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J.R., Mehdi, S.Q., Seielstad, M.T., Wells, R.S., Piazza, A., Davis, R.W., Feldman, M.W., Cavalli-Sforza, L.L. and Oefner, P.J. (2000) 'Y chromosome sequence variation and the history of human populations', *Nature Genetics*, 26: 358–61.