

## Introduction

Laurent Sagart, Roger Blench, Alicia Sanchez-Mazas

► **To cite this version:**

Laurent Sagart, Roger Blench, Alicia Sanchez-Mazas. Introduction. Laurent Sagart, Roger Blench et Alicia Sanchez-Mazas. *The Peopling of East Asia: Putting together archaeology, linguistics and genetics*, RoutledgeCurzon, pp.1-14, 2005. halshs-00104717

**HAL Id: halshs-00104717**

**<https://halshs.archives-ouvertes.fr/halshs-00104717>**

Submitted on 9 Oct 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## INTRODUCTION

5460 words

In the past ten years or so, important advances in our understanding of the formation of East Asian populations, historical cultures and language phyla have been made separately by geneticists, physical anthropologists, archaeologists and linguists. In particular, the genetics of East Asian populations have become the focus of intense scrutiny. The mapping of genetic markers, both classical and molecular, is progressing daily: geneticists are now proposing scenarios for the initial settlement of East Asia by modern humans, as well as for population movements in more recent times. Chinese archaeologists have shown conclusively that the origins of rice agriculture are to be sought in the mid-Yangzi region around 10,000 BP and that a millet-based agriculture developed in the Huang He Valley somewhat later. Linguists have been refining their reconstructions of the proto-languages of the main phyla of the region, and proposing evidence for genetic links to relate these phyla. The period of time they are considering is, by and large, the same period which saw the spread of domesticated plants. General hypotheses are being tested on East Asia: how congruent are languages and genes? and is the formation of language phyla linked with the beginnings of agriculture? Archaeologists, linguists and geneticists are attempting to unravel different aspects of the East Asian problem, sometimes proceeding independently, more often attempting to account for advances in other disciplines. It is important to

emphasise that there are conflicting hypotheses in each field and to clarify for other disciplines the significance of these hypotheses for their own interpretations.

## **FIVE BUILDING BLOCKS**

Before introducing the individual papers, we review current ideas on the classification of East Asian languages for the benefit of non-linguist readers.

Excluding Japanese, Korean, Ainu, and the Altaic languages (Mongolic, Turkic and Tungusic) spoken in the north and east of the region, there is near-universal agreement that the languages of East Asia fall into five phyla (Table 0.1), whose membership, by and large, is beyond dispute: Sino-Tibetan, Hmong-Mien, Tai-Kadai, Austroasiatic and Austronesian.

Table 0.1: Five East Asian phyla

<i>Phylum</i>	<i>Alternative name</i>	<i>Representative languages</i>	<i>Principal locations</i>	<i>Approximate date of ancestor</i>
Sino-Tibetan	Tibeto-Burman (van Driem)	Chinese, Tibetan, Burmese, Jingpo	China incl. Tibet, Burma, Nepal, Bhutan, NE India	6,000-7,000 BP
Hmong-Mien	Miao-Yao	Hmong, Mien, Ho Nte	South China, north Vietnam, Laos	2,500 BP
Tai-Kadai	Kra-Dai (Ostapirat), Daic	Thai, Lao, Kam, Li, Gelao	South China, Indochina, Burma	< 4,500 BP
Austroasiatic		Vietnamese, Khmer, Mon, Khasi, Munda	Indochina, Central Malaysia, NE India	7,000 BP
Austronesian		Atayal, Rukai, Paiwan, Tagalog, Malay, Malagasy, Hawaiian, Maori	Pacific islands except Australia & parts of New Guinea; Madagascar	5,500 BP

**Sino-Tibetan** is a large phylum of some 365 languages,<sup>1</sup> including Chinese and its ‘dialects’ (Sinitic), Tibetan, Burmese and Jingpo, and spoken over a vast unbroken area, mainly in China (including Tibet), Laos, Burma, India, Nepal and Bhutan. Its internal classification is disputed (see van Driem, this volume). Morphemes are mono- and iambisyllables (i.e., a major syllable preceded by a minor, unstressed syllable); many languages are tonal, but tones arose secondarily out of final laryngeal consonants; morphology is predominantly derivational and prefixal, with some suffixes and even infixes; word order is mostly Subject-Object-Verb but Chinese and Karen are Subject-Verb-Object. Chinese has also evolved in the direction of monosyllabicity, and loss of morphological alternations. The Sino-Tibetan proto-language is generally estimated to have been spoken some 6,000 or 7,000 years ago, but the location of the homeland is disputed, with arguments variously made for northern India, Sichuan, the Tibetan plateau, and the Yellow River valley in northern China.

**Hmong-Mien** (also Miao-Yao) is a small and relatively coherent phylum of 32 languages, including the various Hmong ‘dialects’ and Ho Nte, Bunu, Mien etc. Hmong-Mien languages are spoken in scattered pockets, mainly in south China, but also in Laos, Thailand and Vietnam, by farming communities specializing in the exploitation of upland resources. Two branches: Hmongic (Hmong, Ho Nte, Bunu) and Mienic are usually recognized, but other phylogenies have been

proposed. Hmong-Mien has been very influenced by Chinese, to which it is now typologically very close. Only the most basic portion of the reconstructed Hmong-Mien vocabulary is *not* of Chinese origin. Some Chinese loanwords were already part of the Hmong-Mien proto-language; their phonological shape and cultural content suggest a date around 2,500 BP for proto-Hmong Mien. The homeland was most likely in the middle and lower Yangzi Valley. It has been suggested that proto-Hmong-Mien was the language of the state of Chu, a southern neighbour of China during the Zhou dynasty.

The 70 **Tai-Kadai** languages are spoken mainly in south China (including Hainan island), Thailand, Laos, Burma and Vietnam by communities of lowland rice farmers. Its most representative member and oldest literary language is Thai. Like Hmong-Mien, Tai-Kadai (and especially its Kam-Tai subgroup) has received much Chinese influence, and has come to resemble Chinese typologically, with monosyllables, tones and little overt morphology: it has also borrowed numerous Chinese loanwords. However, Benedict (1942) showed that a few languages spoken by small communities conserve more of the original vocabulary of the phylum. He referred to these conservative languages collectively as 'Kadai'. The internal subgrouping of Tai-Kadai is disputed. South China (Guangxi-Guizhou-Hainan) is the area of highest diversity and most Tai-Kadai languages outside of south China belong to southern and central Tai, two subgroups of Tai, itself a subgroup of Kam-Tai. The Tai-Kadai homeland was most likely in south China and the historically documented expansion of southern and central Tai occurred

towards the end of the first millennium CE. Evaluations of the age of Tai-Kadai vary considerably but a date < 4,000 BP appears plausible (Ostapirat, this volume). Another name for Tai-Kadai is **Kra-Dai**, used by Weera Ostapirat (2000; this volume) whose analysis of the internal subgrouping and age of the phylum differ from Benedict's.

**Austroasiatic** is a very diverse phylum of 168 languages whose original geographical unity has been lost, due to migration and to the intrusion of other languages in its midst. It is mainly spoken in Southeast Asia where the most representative languages are Khmer, Mon, and Vietnamese, but also in northern India (Khasi, Munda). Austroasiatic is often regarded as comprised of two branches, a western branch (Munda) and an eastern one (the remainder, including Khasi), but Diffloth (this volume) proposes a different phylogeny, with a central branch consisting of Khasi and Khmuic. Austroasiatic speakers tend to be rice farmers, but some communities in Central Malaysia, Nicobar and elsewhere maintain a foraging life-style. Austro-Asiatic languages have monosyllabic and iambisyllabic morphemes, with prefixal and infixal derivational morphology, SVO and head-modifier word order. Estimates of the age of the proto-language fall in the range 6,000 -7,000 BP, with a homeland presumably in the east, where diversity is highest (but see Diffloth, this volume).

**Austronesian** is a very large phylum of 1,262 languages covering the entire Pacific, excepting parts of New Guinea and surrounding islands, and Australia,

plus Madagascar and parts of south Vietnam. Some of the larger Austronesian languages are Malay, Javanese, Tagalog and Malagasy. Words typically have one, two or three syllables, with disyllables predominating. Syllables tend to be of a simple Consonant + Vowel type. Morphology is predominantly derivational, with prefixes, infixes and suffixes; in many languages, word order is verb-initial and head-modifier. There is growing agreement that the proto-language was spoken ca. 5,500 BP in Taiwan, by a population of millet and rice farmers who were skilled navigators adept at exploiting marine resources. By this view, Austronesian expansion occurred first in Taiwan, where diversity is highest. All the Austronesian languages outside of Taiwan have been shown by Robert Blust to share a few innovations exclusively, and are therefore considered to form a monophyletic taxon within Austronesian: Malayo-Polynesian. Whether Malayo-Polynesian is a primary branch of Austronesian, or merely a subgroup within one primary branch, is a matter of dispute.

### **AN EAST ASIAN COMPLEX OF PHYLOGENIES**

While the monophyletic status of the five phyla discussed in the preceding section is generally accepted, a number of proposals to integrate them into larger constructs, or macrophyla, have been put forward (Table 0.2). We will only be concerned here with theories currently defended by living linguists. For an overview of the early history of ideas on East Asian linguistic classification, the reader is referred to van Driem (2001).



Table 0.2: Proposed macrophyla encompassing East Asian languages

Name of macrophylum	Main advocates	Proposed membership
Macro-Sino-Tibetan	Shafer (1966-1974), Li (1976), Xing (1999)	Chinese + Tibeto-Burman + Tai-Kadai (+ Hmong-Mien)
Austric	Schmidt (1906), Reid (1994), Blust (1998), Higham (1996)	Austronesian + Austroasiatic
Austro-Thai	Benedict (1942)	Austronesian + Tai-Kadai
Yangzian	Davies (1909), Haudricourt (1966), Peiros (1998), Starosta (this volume)	Austroasiatic + Hmong-Mien
Sino-Caucasian	Starostin (1991/1984)	Sino-Tibetan + north Caucasian + Ket
Sino-Tibetan-Austronesian	Sagart (2001)	Sino-Tibetan + Austronesian including Tai-Kadai
Greater Austric	Benedict (1942), Ruhlen (1991), Peiros (1998)	Austro-Thai + Austroasiatic
Macro-Austric	Schiller (1987)	Austronesian + Austroasiatic + Sino-Tibetan + Hmong-Mien + Tai-Kadai
Pan-Sino-Austronesian	Zhengzhang (1993, 1995), Pan (1995)	Austronesian + Austroasiatic + Sino-Tibetan + Hmong-Mien

		+ Tai-Kadai
East Asiatic (conjecture)	Starosta (this volume)	(Sino-Tibetan + Yangzian) + Austronesian
East Asiatic (conjecture)	Sagart (Chapter 15, this volume)	Sino-Tibetan-Austronesian + Yangzian

A view that the Sino-Tibetan and Tai-Kadai languages together form a large East Asian language macrophylum,<sup>2</sup> sometimes also including Hmong-Mien, was prevalent among students of East Asian languages under the name ‘Sino-Tibetan’ well into the second half of the 20<sup>th</sup> century. This theory (here ‘**Macro-Sino-Tibetan**’) was based on the observation that these languages share important traits, such as mono- or iambisyllabicity, tonality, and, for many of them, lack of overt morphology, as well as significant amounts of shared lexicon. Shafer (1966-1974) and Li Fang-kuei (1976) among others have been influential advocates of this theory, which is still popular in mainland China (Xing 1999). In a recent development Zhengzhang (1993, 1995) and Pan (1995) accept Sagart’s view of a genetic relationship between Chinese and Austronesian but (unlike Sagart) make the Austronesian languages part of Macro-Sino-Tibetan under the name ‘**Pan-Sino-Austronesian**’.

Complementary with Macro-Sino-Tibetan, the idea that the Austronesian and Austro-Asiatic phyla are the two primary branches of a larger ‘**Austriac**’ macrophylum is due to Schmidt (1906). Much of the lexical evidence presented

by Schmidt is no longer valid but the morphological evidence continues to be suggestive. Today Austric is defended by Reid (1994; this volume), Blust (1998) and Higham (1996:71) among others, but in a significant development, Reid (this volume) stresses that the Austronesian-Austroasiatic relationship need not be monophyletic, and that while he regards a genetic relationship of Austronesian and Austroasiatic as secure, Sino-Tibetan may be part of that relationship and stand closer to Austronesian than to Austroasiatic. Reid's position is close to that in Sagart (1994:303 and see below).

In the early 1940's Paul Benedict approached the classification of East Asian languages with the premise that the principal type of evidence for genetic relationships must come from basic vocabulary. He noticed lexical resemblances between Thai and Austronesian in lower numerals, personal pronouns and other basic vocabulary. At the same time, he argued that the strong typological resemblances between Thai and Chinese were not accompanied by significant amounts of shared basic vocabulary: he accordingly removed Thai from Sino-Tibetan, treating the relationship between Thai and Chinese as one of contact, with Chinese being on the receiving side. At first (Benedict 1942) he simply transferred Thai from Macro-Sino-Tibetan to the Austronesian side of Austric, which he then accepted, but in his later works, he eliminated Austroasiatic from the ensemble of Thai and Austronesian, these two now forming **Austro-Thai**. The result was a new overall configuration of East Asian linguistic classification, with three separate entities: a restricted **Sino-Tibetan** phylum in the north, consisting

of just Chinese and Tibeto-Burman, an isolated Austro-Asiatic phylum in the south-west, and an **Austro-Thai** phylum in the south-east, to which he eventually added Hmong-Mien (1975) and Japanese (1990).

Accepting Benedict's idea that the Sino-Tibetan languages are unrelated to any of the other languages of East Asia, Sergei Starostin (1991 [1984]), citing agreements in basic vocabulary with sound correspondences, sought to find their relatives in the languages of the north Caucasus and in Ket of the Yenisei Valley. This is the **Sino-Caucasian** hypothesis (see also Peiros 1998). Starostin envisions a proto-language spoken 10,000 years BP in a location west of East Asia, with Sino-Tibetan, and especially Chinese, being intrusive in East Asia.

Starting in 1990, Sagart cited sound correspondences and agreements in vocabulary, basic and non-basic, as well as in morphology, to argue for a genetic relationship between Chinese and Austronesian —the **Sino-Austronesian** theory. In its first version (1993), Chinese was closer to Austronesian than to Tibeto-Burman, but more recently (2001), Sino-Austronesian has two branches, Sino-Tibetan and Austronesian. To reflect this change, Sagart now calls the resulting macrophylum '**Sino-Tibetan-Austronesian**'. The proto-language is identified with the speech of the first rice and millet farmers in the Huang He Valley around 8,000 BP. Sagart also claims that Tai-Kadai is a branch of the Austronesian phylum (this volume, Chapter 10), rather than a separate phylum. **Sino-Tibetan-**

**Austronesian** thus unites Sino-Tibetan, Austronesian and Tai-Kadai into one macrophylum.

Complementary with Sino-Tibetan-Austronesian, a theory claiming that Austroasiatic and Hmong-Mien are two branches of larger macrophylum has its origins in Davies (1909); it was later defended by Haudricourt (1966), Pejros and Shnirelman (1998:155 ff.), who cite Yakhontov as another precursor, and Starosta (this volume). It relies on shared elements of basic vocabulary. As there is no accepted term for this construct we will use Starosta's '**Yangzian**' (so named because Starosta places the homeland in the Yangzi valley).

Benedict's fleeting consideration of a macrophylum consolidating Austric and Austro-Thai, soon abandoned by him, was taken up by Ruhlen (1991) and Peiros (1998). The name they use is 'Austric', but clearly this is different from Schmidt's Austric (limited to Austronesian and Austro-Asiatic). We will use the term '**greater Austric**' to refer to this construct. Pejros and Shnirelman (1998) date its disintegration to the ninth to eighth millennium BCE.

Then come global proposals which aim at unifying all of the five language phyla of East Asia: both Schiller's **Macro-Austric** (Schiller 1987) and Zhengzhang's **Pan-Sino-Austronesian** (Zhengzhang 1993) consolidate Sino-Tibetan, Austro-Tai, Hmong-Mien and Austroasiatic into a macrophylum without an explicit subgrouping. Sagart (1994: 303), acknowledging the validity of some of Reid's

morphological arguments, argues speculatively for a higher-level unity between his Sino-Austronesian (then including Tibeto-Burman) and Austroasiatic, a view close to that expressed by Reid (this volume). A further version of this conjecture, in which Hmong-Mien is added as a third primary branch, is subjected to genetic testing in Chapter 15 of this volume under the name **East Asiatic**. Starosta's **East Asiatic** (this volume) is a conjecture consolidating Sino-Tibetan and Yangzian, and Sino-Tibetan-Yangzian further with Austronesian. Starosta's and Sagart's versions of East Asiatic differ in their internal subgrouping, despite having the same name.

## **THE PAPERS**

This volume consists of three sections: archaeology (Chapters 1-4), linguistics (Chapters 5-10), genetics and physical anthropology (Chapters 12-17) which all address the general issues of the peopling of east Asia and the formation of its populations, material cultures and language phyla.

### **Section I Archaeology**

Chapter 1 by Peter Bellwood considers the general hypothesis that many language phyla dispersed as a consequence of the adoption of agriculture in the light of recent archaeological evidence from East Asia. New dates for rice in Taiwan provide additional support for agriculture as the engine of expansion for Austronesian while the dates for the Yangzi valley allows us to explore the interface between different phyla. The time difference between the earliest dates

for rice and for foxtail millet in northern China lead Bellwood to formulate a scenario in which only one transition to agriculture occurred in East Asia when rice was domesticated in the Yangzi Valley: under this scenario, foxtail millet is a secondary domesticate, brought into cultivation in the Huang He basin as the earliest domesticated rice economy expanded beyond its natural limits. This scenario is alternative to that presented by Tracey Lu in Chapter 3.

Roger Blench's paper (Chapter 2) discusses the ethnolinguistic geography of the East Asian region and in particular the imbalance between the single dominant group in each country and a scatter of numerically small minorities, a pattern not found in other continents. It attributes this to the spread of paddy rice agriculture and looks at linguistic reconstructions of rice terminology to support this. Wet and dry rice turn out to have very different modes of dispersal and it is clear that dry rice had only a limited impact on linguistic diversification.

In Chapter 3, Tracey Lu presents a discussion of the archaeological dates for millets and rice in East Asia, with emphasis on the Chinese mainland. She argues that there are two distinct foci for the transition to agriculture: one in the mid-Yellow River region, based on millet, with early antecedents in the final Palaeolithic of Xiachuan culture of Shanxi; and another in the mid-Yangzi for rice, with antecedents in Jiangxi and Hunan. The question of millet cultivation in Taiwan is given special consideration. The paper includes a map of cereal-yielding sites with dates.

Taiwan archaeologist Tsang Cheng-hwa reports in Chapter 4 on the recently excavated Ta-Pen-Keng site in southwest Taiwan which has yielded the earliest dates for cultivated rice (3,000-2,500 BCE) on the island so far, and the first findings of cultivated grains of millet ever, also dated ca. 3,000-2,500 BCE. These remarkable findings indicate that the earliest Austronesian communities engaged in rice and millet agriculture, as pointed out by Bellwood in Chapter 1. Based on similarities in the material culture, Tsang argues that the most probable homeland of the Austronesians is in the Pearl River delta in Guangdong Province in China. The paper is accompanied by clear photographs of rice and millet grains, as well as of artefacts found at Nan-kuan-li.

## **Section II Linguistics**

Austroasiatic is one of the least-known language phyla in the world and many of its languages remain inaccessible and unmapped. Using new reconstructions based on unpublished fieldwork, Diffloth argued in his oral presentation for an early-period dispersal of shifting cultivators using hillsides along the watersheds of SE Asia and NE India river valleys. The present short contribution in Chapter 5 sets out the Austroasiatic reconstructed forms for terms related to rice-cultivation and faunal terms as a contribution towards eventually locating the homeland of Austroasiatic speakers as well as his latest ‘tree’ of the internal structure of Austroasiatic.



George van Driem has published a series of papers challenging the conventional internal classification of Sino-Tibetan and suggests that the whole phylum must be rethought, arguing in particular for an incorporation of Sinitic and Bodic in the same subgroup. He presents ‘an informed but agnostic picture of Tibeto-Burman subgroups’ in Chapter 6 and uses both recent archaeological and genetic data to make an argument for the homeland of Sino-Tibetan in Sichuan.

Weera Ostapirat has been at the forefront of gathering new data on the Tai-Kadai languages in China and has recently published a new reconstruction of ‘Proto-Kra [=Kadai]’. Using this material, in Chapter 7 he makes a convincing case for a genetic link between Tai-Kadai and Austronesian, using sound correspondences from lexical cognates. He shows that Tai-Kadai preserves early distinctions in the Austronesian languages, typical of the West and central Formosan languages, such as the distinction between PAN \*t and \*C, and between PAN \*n and \*N. He concludes that if, as Sagart argues in Chapter 10, the Tai-Kadai languages are a subgroup within Austronesian, rather than being a related phylum, then they are more likely to be outside the clade which includes the languages of the Formosan East coast and Malayo-Polynesian.

Chapter 8 by Lawrence Reid, currently the most prominent advocate of the Austric theory, critically examines the supporting lexical evidence presented by L.V. Hayes, concluding that limited parts of it are admissible. He also reviews the morphosyntactic evidence presented to date and answers some criticisms of earlier

publications. Reid reiterates the validity of the Austronesian-Austroasiatic genetic connection but, in an important development, concludes, in view of the evidence presented by Sagart linking Sino-Tibetan and Austronesian, that the relationship between Austroasiatic and Austronesian may turn out to be more remote than earlier considered, and the Austric phylum as traditionally defined not monophyletic, but could include Sino-Tibetan as well.

Sagart first proposed a genetic link between Sinitic and Austronesian in 1990, based essentially on shared lexicon, sound correspondences and shared morphology. In Chapter 9 he presents an improved argument for Sino-Tibetan-Austronesian, a theory which claims the Sino-Tibetan and Austronesian families are related. The proposed proto-language (PSTAN) would originate in the millet culture of northern China in the mid-Huang He valley between 8,500 and 7,500 BP, and the Taiwan millet culture would thus be a retained feature from this epoch.

Chapter 10, also by Sagart, presents a new theory of the origin of Tai-Kadai. Instead of being co-ordinate with Austronesian as Benedict argues, it is viewed as an offshoot of Proto-Austronesian, belonging to the clade which includes several of the languages of the Formosan East coast and Malayo-Polynesian. Evidence comes from lexical and morphological features in the vocabulary Thai (broadly speaking) shares with Austronesian: in particular Tai-Kadai shares with Malayo-Polynesian some characteristic innovations in the second-person pronouns.

Under the name 'Proto-East-Asian' the late Stanley Starosta presents a conjecture in Chapter 11 unifying all five-language phyla of East Asia, accompanied by an explicit scenario linking linguistics with archaeology. Starosta's conjecture involves an ancestral language spoken some 8,000-8,500 years BP on the north China plain by an expanding population of millet farmers identified with the Cishan-Peiligang culture. The first to break off was a group identified as the pre-Austronesians, who were located on the eastern seaboard of China (Dawenkou and Hemudu cultures): one subgroup reached Taiwan, acquiring rice agriculture along the way. In Taiwan these people became the proto-Austronesians and started diversifying into the various Austronesian branches, including Tai-Kadai (Starosta accepts Sagart's view, presented in Chapter 10, that the Tai-Kadai phylum is a subgroup of Austronesian, rather than a distinct phylum). Meanwhile, those who stayed at home in the north China plain expanded south toward the Yangzi region, forming a southern, or Yangzian branch, later to diversify into Hmong-Mien and Austroasiatic, while the others still in the northern China plain evolved into the Tibeto-Burman phylum (Starosta accepts Driem's understanding of this phylum, with the associated terminology). Some linguistic characteristics of each proposed node in the tree are outlined.

### **Section III Genetics and physical anthropology**

In Chapter 12, physical anthropologist Michael Pietrusewsky analyses the available craniometric data of modern and near-modern indigenous inhabitants of

East Asia and Oceania using multivariate analyses on a total of 2,805 male crania. The study suggests a major subdivision into an East Asian/Pacific group and an Australo-Melanesian group, supporting the hypothesis of two separate colonization events involving morphologically distinct populations. An early differentiation of Southeast and East/Northeast Asian populations also emerges from the data. On the other hand, the results challenge views based on archaeology and historical linguistics by proposing a homeland for Pacific peoples in island Southeast Asia rather than China/Taiwan.

The other papers in this section focus on genetics. In Chapter 13, immunogeneticist Marie Lin and co-workers present a large synopsis of classical and HLA polymorphisms in aboriginal peoples of Taiwan. Very peculiar genetic traits and a high inter-tribal diversity are observed in this island, suggesting long isolation of small populations. Although Taiwanese peoples are genetically related to insular Southeast Asians, the authors also suggest a possible link between the Ami of the east coast of Taiwan and Australo-Melanesians. Overall, they argue, present Taiwanese differentiations indicate a complex peopling history possibly starting before 12,000 years BP when the island was still connected to the continent.

The significance of DNA markers in the reconstruction of East Asian prehistory is addressed by geneticist Chu Jiayou (Chapter 14) whose paper describes the remarkable diversity of Chinese populations (especially in Yunnan Province) and summarizes two recently published works on microsatellite and Y chromosome

polymorphisms in China. His main conclusion supports a unique origin of all modern humans rather than a multiregional model of *Homo sapiens* origins.

Chapters 15 and 16 are two contributions by geneticists Estella Poloni and Alicia Sanchez-Mazas in collaboration with linguists Guillaume Jacques and Laurent Sagart. They compare the genetic structure of East Asian populations to the linguistic structure observed in this continent by analysing large sets of genetic data for two blood polymorphisms (RH and GM) and the HLA-DRB1 locus of the major histocompatibility complex. Using an analysis of variance framework, both studies indicate a significant correspondence between linguistic and genetic differentiation in East Asia, although the genetic landscape of human populations is closely related to geography showing a pattern of continuous differentiation along a north-to-south axis. In Chapter 15, Poloni and her collaborators also compare the RH and GM variation against three competing linguistic phylogenies, i.e. Sagart's hypothesis of a main East-Asian macrophylum, a combination of the greater Austric and Sino-Caucasian hypotheses, and a null hypothesis, assuming no genetic relationships between the main East Asian phyla. The authors conclude that the data do not yet permit us to discriminate between the three hypotheses.

In Chapter 16, Sanchez-Mazas et al. also discuss the observed HLA-DRB1 genetic diversity in each East Asian linguistic phylum in relation to several models of human differentiation based on the variation of two genetic diversity indexes, the diversity *among* and *within* populations, respectively. A main

difference is observed between continental East Asians and the insular populations represented by Austronesians who probably experienced rapid genetic differentiations. Based on the frequencies of peculiar HLA-DRB1 alleles, a close historical relationship is also tentatively proposed between extra-Formosans and populations from the east coast of Taiwan, in particular the Amis. This view is alternative to that presented by Marie Lin et al. in Chapter 13.

Geneticist Peter Underhill presents a complete overview of Y chromosome diversity in East Asia and Oceania in Chapter 17 by synthesising the data of 3,702 samples from 73 populations analysed by different authors to produce a broad phylogeny. East Asian lineages are derived from a unique ancestor that developed into three main branches. The author relates these lineages to different migration events, notably a first migration from Africa into southern Asia via a coastal route, and an early settlement of Asia by successful colonizers displaced to the geographic margins by pressure from more recent migrations. On the other hand, the two complementary graphs of Y chromosome frequencies in Asia/Oceania presented by Underhill reveal intricate genetic relationships which suggest a highly complex history of the peopling of these continents.

## **BROAD THEMES**

Deep similarities between the language phyla of East Asia have led scholars to believe that they reflect genetic connections and proposals for macrophyla have a long history. However, these proposals are themselves highly diverse and

certainly some similarities must be explained by early contacts, for example, the ancient strata of Sinitic lexemes in Hmong-Mien. It is also true that a history of intense bilingualism has caused some phyla to undergo dramatic morphological restructuring thereby concealing similarities; witness the encapsulation of Tai-Kadai within Sinitic. A proposal that has had particular longevity is the Austric proposal, uniting Austronesian and Austroasiatic. Originally put forward by Schmidt, it has had a significant revival in the 1990s in the publications of Reid and La Vaughn Hayes. Blust is now a supporter and archaeologists such as Higham have adopted it to explain patterns of East Asian prehistory. Others, such as Diffloth and Sagart, oppose it and a consensus may be emerging that the relationship is not as neat as a single clade, but rather that Austroasiatic and Austronesian fit together in a larger macrophylum that includes all the phyla under discussion in different configurations. Similarly, Austro-Thai, first put forward by Benedict, is now gathering support from Ostapirat and Sagart, although they differ in their interpretations of the structure of this relationship. The key to disentangling such high-level relationships is more complete reconstruction of proto-languages, a particularly urgent task in the case of Sino-Tibetan.

Peter Bellwood has been an active promoter of the notion that language expansions have been driven by agriculture, a hypothesis that has itself expanded out of the Austronesian region to cover much of the world (for a recent restatement see Diamond & Bellwood 2003). This has been a major stimulus to

the field and has gathered much support in various areas. Some language phyla do demonstrate such a wealth of reconstructions in the field of agriculture that it is economical to suppose that its introduction was the engine of their expansion. This is true, for example, in Austronesian and Tai-Kadai. However, in other phyla, such as Sino-Tibetan and Austroasiatic, reconstructions are fewer and appear to reflect principally cereal cultivation. It is also important to emphasise that reconstructions of single crop names can simply reflect the presence of wild forms; for agriculture to be given this starring role more breadth is required. What is stimulating is that archaeology and linguistics can come together to throw up hypotheses and test each other's models; and the pace at which new archaeobotanical material is appearing will surely change the picture of agriculture in East Asia rapidly in the coming decade.

Macrophyla proposals have a venerable history in the field, but the comparison of genetic variation and linguistic classification, pioneered by the teams of Luca Cavalli-Sforza, Robert Sokal and André Langaney among others is less than two decades old (Excoffier et al. 1987, 1991; Cavalli-Sforza et al. 1988, 1992; Sokal et al. 1988, 1992). The potential of both classical and DNA polymorphisms for assessing the historical relatedness or level of admixture between human populations appears to be enormous, but it is clear from the analyses both here and in related texts that their interpretation should be kept within reasonable limits. Genetic studies allow us to offer major narratives of the peopling of East Asia, but not to decide between specific transphylic hypotheses. This is partly a matter of



sampling: because the indigenous populations of Taiwan have been so intensively studied, observations such as the special status of the Amis (Lin and colleagues, Chapter 13) can be made. But this is also a matter of evolution; genes and languages, even when deriving from a common origin, do not evolve at the same rate, and the levels of gene flow across linguistic boundaries may also vary greatly around the world. While keeping such limitations in mind, we believe that our understanding of human peopling history can be considerably improved by putting together the three disciplines, archaeology, linguistics and genetics.

#### REFERENCES

- Benedict, P.K. (1942) 'Thai, Kadai and Indonesian: a new alignment in Southeastern Asia' *American Anthropologist*, *n.s.*, 44, 576-601.
- (1975) *Austro-Thai: language and culture*, New Haven: HRAF Press.
- (1990) *Japanese/Austro-Tai*, Ann Arbor: Karoma.
- Blust, R. (1998) Beyond the Austronesian homeland: the Austric hypothesis and its implications for archaeology, in Ward H. Goodenough (ed.) *Prehistoric settlement of the Pacific*. Philadelphia: American Philosophical Society.
- Cavalli-Sforza, L. L., Piazza, A., Menozzi, P. and Mountain, J. (1988) 'Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data', *Proceedings of the National Academy of Science USA*, 85:6002-6.

- Cavalli-Sforza, L. L., Minch, E. and Mountain, J. L. (1992) 'Coevolution of genes and languages revisited', *Proceedings of the National Academy of Science USA*, 89:5620-4.
- Davies, H. R. (1909) *Yunnan: the link between India and the Yangzi*, Cambridge: University Press.
- Diamond, J. & Bellwood, P. (2003) 'Farmers and their languages: the first expansions', *Science*, 300:597-603.
- van Driem, G. (2001) *Languages of the Himalayas. An ethnolinguistic handbook of the Greater Himalayan Region*, 2 vols, Leiden: Brill.
- 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, 3:273-97.
- 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.
- Haudricourt, A.-G. (1966) 'The limits and connections of Austroasiatic in the Northeast', in Norman H. Zide (ed.) *Studies in Comparative Austroasiatic Linguistics*, 44-56, The Hague/Paris: Mouton.
- Higham, C. (1996) *The Bronze Age of Southeast Asia*, Cambridge: Cambridge University Press.
- Li, F.K. (1976) 'Sino-Tai', *Computational Analyses of Asian and African Languages*, 3: 39-48.

- Ostapirat, W. (2000) 'Proto-Kra', *Linguistics of the Tibeto-Burman Area* 23.1.
- Pan, W.Y. (1995) 'Dui Hua-Ao yuxi jiashuo de ruogan zhichi cailiao', in W. Wang (ed.) *The Ancestry of the Chinese language*, Journal of Chinese Linguistics monograph series n°8, 113-144.
- Peiros, I. (1998) *Comparative Linguistics in Southeast Asia*. Canberra: Pacific Linguistics.
- Peiros, I. and V. Shnirelman (1998) 'Rice in Southeast Asia: a regional interdisciplinary approach', in R. Blench and M. Spriggs (eds) *Archaeology and Language, archaeological data and Linguistic hypotheses* Vol. II: 379-389, London: Routledge.
- Reid, L. (1994) 'Morphological evidence for Austric', *Oceanic Linguistics* 33, 2: 323-344.
- Ruhlen, M. (1991) *A guide to the world's languages Vol 1: Classification*, Stanford University Press.
- Sagart, L. (1993) 'Chinese and Austronesian: evidence for a genetic relationship' *Journal of Chinese Linguistics* 21, 1 : 1-62.
- (1994) 'Old Chinese and Proto-Austronesian evidence for Sino-Austronesian', *Oceanic Linguistics* 33 (2): 271-308.
- (2001) Connections across the south Pacific: a personal synthesis, paper for *Pacific Neighborhood Consortium conference*, Hongkong, January 2001.
- Schiller, E. (1987) 'Which way did they grow? (Morphology and the Austro-Tai/(Macro)-Austric debate)', *Proceedings of the 13th Annual Meeting of the Berkeley Linguistic Society* 235-246.

- Schmidt, W. (1906) *Die Mon-Khmer Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens*, Braunschweig: Friedrich Vieweg und Sohn.
- Shafer, R. (1966-1974) *Introduction to Sino-Tibetan*, Wiesbaden: Otto Harrassowitz.
- Sokal, R. R., Oden, N. L. and Thomson, B. A. (1988) 'Genetic changes across language boundaries in Europe', *American Journal of Physical Anthropology*, 76:337-61.
- (1992) 'Origins of the Indo-Europeans: genetic evidence', *Proceedings of the National Academy of Science USA*, 89:7669-73.
- Starostin, S. (1991 [1984]) 'On the Hypothesis of a genetic connection between the Sino-Tibetan languages and the Yeniseian and North-Caucasian languages', translation and introduction by William H. Baxter III, in V. Shevoroshkin (ed.) *Dene-Sino-Caucasian*, Bochum: Brockmeyer.
- Xing, G.W. (1999) *Han-Tai yu bijiao shouce*. Beijing: Shangwu.
- Zhengzhang, S.F. (1993) 'The root of Austro-Tai languages is in Sino-Austronesian', Paper presented at the Conference on Asia-Mainland/Austronesian connections (CAMAC), Honolulu, May 10-13, 1993.
- (1995) 'Hanyu yu qinshu yu tongyuan genci ji fuzhui chengfen bijiaoshang de zedui wenti', in W. Wang (ed.) *The Ancestry of the Chinese language*, Journal of Chinese Linguistics monograph series n° 8, 269-282.

---

<sup>1</sup> Numbers of languages per phylum cited here are from the *Ethnologue*

[http://www.ethnologue.com/family\\_index.asp](http://www.ethnologue.com/family_index.asp) (accessed July 2003).

<sup>2</sup> The position of the then little-known Hmong-Mien languages was a question mark, but recent versions of the theory, especially in China, make Hmong-Mien a part of Macro-Sino-Tibetan.