The author presents a new phylogeny for the Austronesian language family, and to argues that Tai-Kadai belongs to this family.

Main method: examination of numerals in Formosan languages (the indigenous languages of Taiwan)

Additional arguments:
- lexical items that fit the phylogeny
- Geographical distribution of descendant languages

1. The present consensual reconstruction of Proto-Austronesian numerals:
   1: *isa, *esa
   2: *duSa
   3: *telu
   4: *Sepat
   5: * lima
   6: *enem
   7: *pitu
   8: *walu
   9: *Siwa
   10: *puluq

In Formosan languages, where there is a reflex (modified version, “descendant”) of puluq as 10, there is also a reflex of Siwa as 9; where there is a reflex of Siwa as 9, there is also a reflex of walu as 8; if walu as 8 then enem as 6; if enem as 6 then lima as 5; if lima as 5 then as 7, but not in reverse order. (I.e., if a language has a reflex of walu as 8, it does not necessarily have a reflex of Siwa as 9.)
puluq 10→Siwa 9→walu 8→enem 6→lima 5→pitu 7

<table>
<thead>
<tr>
<th></th>
<th>pitu 7</th>
<th>lima 5</th>
<th>enem 6</th>
<th>walu 8</th>
<th>Siwa 9</th>
<th>puluq 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luilang</td>
<td>innai</td>
<td>(na)ulp</td>
<td>(na)tsulup</td>
<td>patulunai</td>
<td>satulunai</td>
<td>isit</td>
</tr>
<tr>
<td>Saiiasat</td>
<td>saivuseaha</td>
<td>rrasu</td>
<td>saivusa</td>
<td>makaspat</td>
<td>ra'ha</td>
<td>ranpon</td>
</tr>
<tr>
<td>Pazeh</td>
<td>xasebidusa</td>
<td>xasep</td>
<td>xasebusa</td>
<td>xasebaturu,</td>
<td>xasebituru</td>
<td>isit</td>
</tr>
<tr>
<td>Favorleng</td>
<td>naito</td>
<td>achab</td>
<td>nataap</td>
<td>maaspat</td>
<td>tannacho</td>
<td>zchiett</td>
</tr>
<tr>
<td>Taokas</td>
<td>yweto</td>
<td>hasap</td>
<td>tahap</td>
<td>mahalpat</td>
<td>tanasu</td>
<td>(ta)sisid</td>
</tr>
<tr>
<td>Sediq</td>
<td>pito</td>
<td>lima</td>
<td>mataro</td>
<td>maspat</td>
<td>manjali</td>
<td>maxal</td>
</tr>
<tr>
<td>Thao</td>
<td>pitu</td>
<td>rima</td>
<td>ka–turu,</td>
<td>kahspat,</td>
<td>tanacu</td>
<td>maqcin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>makal–turu–turu</td>
<td>maka(h)–shpa–shpat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siraya</td>
<td>pîttu</td>
<td>rima</td>
<td>nam</td>
<td>kuixpa</td>
<td>matuda</td>
<td>saat kîttian</td>
</tr>
<tr>
<td>Hoanya</td>
<td>pito</td>
<td>Lima (mi)un</td>
<td>(mi)alu</td>
<td>(a)sia</td>
<td>(miatasi)</td>
<td></td>
</tr>
<tr>
<td>Papora</td>
<td>pito</td>
<td>nema</td>
<td>(ne)nem</td>
<td>mahal</td>
<td>(me)siya</td>
<td>(me)tisi</td>
</tr>
<tr>
<td>Tsou</td>
<td>pîtu</td>
<td>eîmo</td>
<td>nómês</td>
<td>vôeu</td>
<td>sîo</td>
<td>màskê</td>
</tr>
<tr>
<td>Saaraoa</td>
<td>(k)upito</td>
<td>(k)ulima</td>
<td>(k)ânama</td>
<td>(k)ualo</td>
<td>(k)usia</td>
<td>(ku)ma:4ə</td>
</tr>
<tr>
<td>Kanabu</td>
<td>pîtu</td>
<td>rima</td>
<td>nam</td>
<td>(h)a:ru</td>
<td>sî:ya</td>
<td>ma:nê</td>
</tr>
<tr>
<td>Bunun</td>
<td>pîtu'</td>
<td>hima'</td>
<td>nuum</td>
<td>vau'</td>
<td>sîva'</td>
<td>mas'an</td>
</tr>
<tr>
<td>Rukai</td>
<td>pitu</td>
<td>Lima</td>
<td>eneme</td>
<td>vaLu</td>
<td>banata</td>
<td>manjeale</td>
</tr>
<tr>
<td>Paiwan</td>
<td>pitju</td>
<td>Lima</td>
<td>enem,</td>
<td>alu</td>
<td>sîva</td>
<td>puluq</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>unem</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puyuma</td>
<td>pîtu</td>
<td>Lima</td>
<td>nem</td>
<td>vaLu</td>
<td>iwa</td>
<td>puLu</td>
</tr>
<tr>
<td>Amis</td>
<td>pîtu</td>
<td>lima</td>
<td>'enem</td>
<td>falu</td>
<td>siwa</td>
<td>polo</td>
</tr>
<tr>
<td>Kavalan</td>
<td>pîtu</td>
<td>rima</td>
<td>'nem</td>
<td>waru</td>
<td>siwa</td>
<td>betin</td>
</tr>
<tr>
<td>Ketagalan</td>
<td>pîtu</td>
<td>tsjima</td>
<td>ânêm</td>
<td>wasu</td>
<td>siwa</td>
<td>labatan</td>
</tr>
<tr>
<td>PMP</td>
<td>*pitu</td>
<td>*lima</td>
<td>*enem</td>
<td>*walu</td>
<td>*siwa</td>
<td>*puluq</td>
</tr>
</tbody>
</table>

**Conclusion:** contrary to the established view, these numerals (5-10) did not exist in Proto-Austronesian. They arose one after the other, in succession: first pitu 7, then lima 5 in a language that already had pitu 7, then enem 6 in a language that already had pitu 7 and lima 5 and so forth.

Numerals can be used to classify languages → these numerals are the basis of Sagart’s new phylogenetic tree.

The consensus that the numerals between 5-10 are proto-Austronesian relies on the fact that they are all present in Proto-Malayo-Polynesian, and the assumption that Malayo-Polynesian languages are an independent branch of Austronesian.

→If pitu 7, lima 5, enem 6, walu 8, Siwa 9 and puluq 10 all arose in Taiwan, then Malayo-Polynesian is not an independent branch, but part of a taxon that includes languages in Eastern Formosa.
The etymologies for pitu 7, walu 8 and Siwa 9

Those Formosan languages that do not have the familiar 5-10 numerals usually have analytic forms:

Additive:

Pazeh language:

6 = 5+1, xaseb-uza; 7 = 5+2, xaseb-i-dusa; 8 = xaseb-a-turu or xaseb-i-turu; 9 = xaseb-i-supat

(See Proto-Austronesian reconstructions above: 1: *isa, *esa; 2: *duSa; 3: *telu; 4: *Sepat)

Saitaoyak and Tarumyan are also additive.

Two varieties of Saisat have saivuseaha for 7: saivusa 6 + aha 1.

Multiplicative:

2x3 and 2x4 are common for 6 and 8. E.g.:

Sediq: materu 6 is based on *telu 3; maspat 8 based on *Sepat

Thao: katuru 6 based on *telu; kashpat 8 based on *Sepat

Saitaoyak: makaspat; Tarumyan: kaspat; Saisat, Favorlang, Taokas, Siraya also have 8 based on 4.

Subtractive forms for 9:

Sediq: maŋali, an imperative form of maŋal, “to take” (from ten);

Perhaps Saisat-Saitaoyak ra:ha, which contains aha, “one”;

etc.

These analytic forms for the numerals above 5 are prevalent on the west coast

→ probably the place of the earliest settlement, since it faces the continent

Two explanations for these analytic numerals:

1. After the settlement of Taiwan, the Proto-Austronesian numerals (enem 6, pitu 7, walu 8, Siwa 9, puluq 10) were replaced by analytic forms

2. Proto-Austronesian had stable forms for numerals 1-5, and no stable words for 6-9. The numerals between 5 and 10 were made up using additive, multiplicative and subtractive methods. The analytic forms in the west coast languages are their descendants. The familiar forms (enem 6, pitu 7, walu 8, Siwa 9) arose in succession in daughter languages, gradually replacing all the old analytic expressions.

Sagart argues that the second explanation is true.
Pazah 5, xasep has cognates in other west coast languages:

Favorlang achab, Saisat a:seb → reflexes of *RaCep

Taokas: hasap → appears to reflect *qaCep

The numerals added to *RaCep in Pazeh 6,7,8,9 are the Proto-Austronesian numerals (1: *isa, *esa; 2: *duSa; 3: *telu; 4: *Sepat) with regular developments.

The full additive forms found in Pazeh are not observed elsewhere, but shorter forms in some west coast languages are indicative of their former existence:

- The final three syllables of Pazeh xasebaturu, the additive form for '8', are paralleled in Luilang '8' patulu-nai (where -nai is detachable, compare '7' in -na and '9' satulu-nai, with sa- = '1' in Luilang).
- One dialect of Siraya (Tsuchida, Yamada and Moriguchi 1991, point M2) has sipat nine: 8 given the phonetic proximity with PA *Sepat 'four', this is almost certainly based on an additive 5+4 form.

Based on the above, Sagart proposes a new etymology for *pitu 7, *walu 8 and *Siwa 9:

\[
\begin{array}{cccccccc}
\text{x} & \text{a} & \text{s} & \text{e} & \text{b} & \text{i} & \text{d} & \text{u} & \text{s} \\
\text{p} & \text{i} & \text{t} & \text{u} & \\
\end{array}
\]

\[
\begin{array}{cccccccc}
\text{x} & \text{a} & \text{s} & \text{e} & \text{b} & \text{a} & \text{t} & \text{u} & \text{r} & \text{u} & \\
\text{w} & \text{a} & \text{l} & \text{u} & \\
\end{array}
\]

\[
\begin{array}{cccccccc}
\text{x} & \text{a} & \text{s} & \text{e} & \text{b} & \text{i} & \text{u} & \text{s} & \text{p} & \text{a} & \text{t} & \\
\text{S} & \text{i} & \text{w} & \text{a} & \\
\end{array}
\]

The sound changes can be explained by the rules of Pazeh and Austronesian phonology.

→ "consensus Proto-Austronesian" forms for 7, 8, and 9 arose as reduced forms of earlier additive expressions.

Based on the accepted sound correspondences between Pazeh and Proto-Austronesian, this process can be reconstructed as follows:
Although they are arbitrary, the changes supposed here

- are natural changes: assimilations, cluster simplifications, schwa deletions, lenitions, stress-conditioned prunings, not outrageous changes like p > r, or l > m, or i > q;
- affect at least two forms (except for changes 1 and 6): two changes (#3, 4) affect the entire paradigm (three forms). By definition, \textit{ad hoc} changes would affect only one form. The relatively marked lenition -pa\textbf{–} > -wa\textbf{–} affects two forms;
- do not change the vowels, except for schwas: this is a general tendency of later An phonetic evolution;
- do not affect the points of articulation of the consonants.

Why \textit{wali 8} and \textit{Siwa 9} appear only in Formosan languages that have \textit{pitu 7}?

\textit{pitu 7}, \textit{wali 8} and \textit{Siwa 9} all appear at stage 5, but for \textit{wali 8} and \textit{Siwa 9}, there were multiplicative and subtractive competitors, while for \textit{pitu 7} there were none

\textit{all the daughters of the stage-5 language have a reflex of \textit{pitu 7}, but only some of them show \textit{wali 8} and \textit{Siwa 9}.}

Those which do not have \textit{wali 8} and \textit{Siwa 9} (Atayal, Sediq, Thao, Favorlang, taokas and Siraya) all have a multiplicative form for 8 and either a subtractive form or a form of unknown origin for 9:
These innovations (pitu 7, walu 8, Siwa 9) are ideal criteria for subgrouping Austronesian languages:

- Their etymology is known and the direction of the innovation is certain (in contrast, the etymologies of enem 6 and puluq 10 are not known)

- The risk that they might have spread by contact is minimized by the fact that they belong to the (relatively) basic vocabulary

- The risk that each of these innovations was made several times independently is almost nonexistent, given the complexity of the six-stage process and the fact that the resulting cognates obey the habitual Austronesian sound correspondences

→ the pitu 7, walu 8 and Siwa 9 innovations form the backbone of Sagart’s new Austronesian phylogeny:

PAn (Proto-Austronesian): the language spoken by the first neolithic settlers of Taiwan, from the moment they set foot on the island to the moment when their language broke up into two or more dialects some generations later

Pituish: the hypothetical daughter language of PAn in which pitu 7 was first invented. (Might have had also wa(t)lu 8 and Siwa 9, but these were later replaced.)

Walu-Siwaish: the language in which Siwa and walu eliminated all their competitors and became the only words for 8 and 9.

Proto-Malayo-Polynesian shares pitu 7, walu 8, and Siwa9, therefore it is not an independent branch of Proto-Austronesian, but part of Walu-Siwaic, a branch of Pituic.

Hoanya and Papra appear to have reflexes of walu 8 and Siwa 9, but these are probably loans from southern Tsoic.
Enriching the phylogeny

Additional lexical and morphological characters that are compatible with the phylogeny:

Well-known cognates sets often assigned to the Proto-Austronesian level, but which Sagart argues are post-PAn:

The remaining three numerals: lima 5, enem 6, puluq 10

Lima, ‘5’:
- *lima is also the PAn word for ‘hand’. The older meaning is presumably ‘hand’ → lima as ‘5’ is innovation
- *RaCep, and not lima, is used in additive expressions based on 5 → lima as ‘5’ is innovation.

RaCep 5 is distributed only in the upper region of Sagart’s phylogeny, specifically in Pazeh, Saisat and Taokas.

Lima 5 is universal in Pituic as ‘5’, except in Taokas and Favorlang.

→ Pituish might have been a common ancestor to Taokas and Favorlang on the one hand and to “Limaish” on the other

or

→ lima already had the meaning of ‘5’ in Pituish, but ReCap 5 was not eliminated yet (two competing forms)

Enem, ‘6’:

The etymology of this word is unknown, but there are good reasons to suppose that it too is a post-PAn innovation: some of the highest languages in Sagart’s phylogeny still have additive forms for ‘6’: Pazeh xaseb-uza, Sediq ma-teru, Thao ka-turu etc. Some have words for 6 of unknown origin. The other Formosan languages, including all on the east coast and Proto-Malayo-Polynesian, have reflexes of enem 6 → it’s an innovation

Puluq ‘10’:

It occurs in Rukai (some varieties), Paiwan, Puyuma, Amis and Proto-Malayo-Polynesian, all of which are Walu-Siwaic languages. Other languages show reflexes of #sa-iCit, which was probably the Proto-Austronesian form; #masehaN, which was probably a Pituish invention that competed with sa-iCit and puluq; and in Saisat ranpon, which is an isolated form and probably a local innovation.
*CawiN, ‘year’:

According to Blunt, the Proto-Austronesian word for ‘year’ was *kawaS

→ indeed, reflexes of kawaS only occur in the higher regions of Sagart’s phylogeny: Pazeh, Saisiat, Atayalic, Thao.

Another form for ‘year’, *CawiN has reflexes in Rukai, Kanakanabu, Saaora, Paiwan, Bunun and Siraya.

→ *CawiN displaced *kawaS as ‘year’ in a language ancestral to Siraya and Walu-Siwaic, i.e., Enemish.

*-mu ‘2nd person singular genitive’, i.e., “your”, singular:

Originally: *-Su ‘your, singular’; *-mu “your, plural”

A politeness shift probably occurred, and *-mu became ‘your, singular’ in Ketagalan and Proto-Malayo-Polynesian. → innovation of the “Muish” language.

However, Kavalan, which forms the North-East Formosan with Ketagalan, shows a reflex of –Su, not –mu, as 2nd person singular genitive. It is possible that Kavalan eliminated the polite pronoun in favor of the non-polite form.

*manuk, ‘bird’:

Reflected in Proto-Malayo-Polynesian and Ketagalan, but in no other language of Taiwan. Reflexes of *qayam ‘bird’ are widespread, including Kavalan. But *manuk did not replace *qayam, which is reflected in Malayo-Polynesian languages in some cases still as ‘bird’, but more often as ‘domesticated animal’

→*manuk arose in Muish, and took the meaning of ‘wild bird’, while *qayam meant ‘domesticated bird’, and later extended its meaning to ‘domesticated animal’ in general, leaving *manuk free to shift to ‘domesticated fowl, chicken’, or not. Kavalan, a Muish language, abandoned *manuk, keeping *qayam ad the only word for ‘bird’.
Fitting morphological innovations in the model

Starosta (2001) list eight morphological characters of Formosan languages.

Six of these are found in Saisiat and/or Pazeh (high in the phylogeny) and elsewhere in Taiwan → Proto-Austronesian features

The remaining two are post-PAn innovations → potentially informative for early Austronesian subgrouping.

1. a-prefixation marking future in verbs (Thao, Tsouic, Rukai, Amis):

their absence in Saisiat and Pazeh could be the result of loss or of incomplete description. If neither possibility applies, the process can be an innovation of Pituish. In any case we need to suppose several independent events of loss of this feature at later times.

2. paŋ-prefixation deriving instrumental nouns out of verbs (Amis, Malayo-Polynesian)

Most likely a post-PAn innovation. Amis, an east coast language, is the only Formosan language to show this feature. Probably an innovation of Walu-Siwaish.
Fitting phonological innovations into the model

Phonological mergers are convenient features in subgrouping, because one can be sure that they are inventions. At the same time, they can spread by contact across language boundaries, which means they are not very reliable. Sagart argues that 5 important mergers in Formosan languages are better explained as areal events than as phylogeny-defining events, and the most useful information they provide is geographical: what languages were in contact at the time when a change took place. This information can be used to probe the location of the Formosan precursor of Proto-Malayo-Polynesian: that language must have been spoken near the precursors of Bunun and Kavalan-Ketagalan, since these are the languages with which it shares the most sound changes → probably an east coast location.

The position of Tai-Kadai

The modern Tai-Kadai languages are spoken in parts of South China, Vietnam, Laos, Thailand, Burma and Assam. The Tai-Kadai languages outside of south-east China and adjacent areas of Vietnam are all within the Tai subgroup, and are very homogeneous. This is due to the expansion of Tai speakers in late first and early second millennia AD. The area of highest diversity is in the northeast part of the Tai-Kadai domain: in Hainan Island, in northern Vietnam, and in the Chinese provinces of Guangxi and Guizhou. This is presumably where the TK homeland was located, not earlier than 2000 BC.

Benedict proposed in 1942 that Austronesian and Tai-Kadai families are coordinate within an ‘Austro-Tai’ macrophyllum. He based his argument on the fact that they share a remarkable set of very basic vocabulary items (personal pronouns, numerals, body part terms, basic verbs). His proposal didn’t meet full approval at the time.

Sagart and Ostapirat argue that sets in Benedict’s basic vocabulary comparisons can be isolated which exhibit strong phonetic regularities:

<table>
<thead>
<tr>
<th></th>
<th>PAN</th>
<th>PMP</th>
<th>Tai</th>
<th>Lakkia</th>
</tr>
</thead>
<tbody>
<tr>
<td>die</td>
<td>maCa</td>
<td>matay</td>
<td>ta:i1</td>
<td>plei1</td>
</tr>
<tr>
<td>eye</td>
<td>maCa</td>
<td>mata</td>
<td>ta1</td>
<td>pla1</td>
</tr>
<tr>
<td>bird</td>
<td>manuk</td>
<td>nok8</td>
<td>mlok7</td>
<td></td>
</tr>
</tbody>
</table>

Sagart argues that the Tai-Kadai forms can be accounted for on the ground of Proto-Austronesian and Proto-Malayo-Polynesian forms, e.g.:

maCa → mCa → pCa → pta → pla, ta

Further evidence: Buyang, a Tai-Kadai language that preserved several Austronesian disyllabic words, whereas most Tai-Kadai languages reflect Austronesians disyllables as monosyllables.
In these Buyang words, the first syllable is reduced: the vowel is always /a/, the syllable is toneless, and the inventory of initial consonants is limited to a few. Yet Buyang shows that Austronesian words in Proto-TK were still disyllabic.

The best Austronesian – Tai-Kadai comparisons are personal pronouns and numerals:

<table>
<thead>
<tr>
<th>Buyang</th>
<th>PAN</th>
<th>PMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ku₅⁴</td>
<td>-ku</td>
</tr>
<tr>
<td>thou</td>
<td>ma₃¹²</td>
<td>-Su</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Buyang</th>
<th>PAN</th>
<th>PMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>ca₅⁴</td>
<td>duSa</td>
</tr>
<tr>
<td>3</td>
<td>tu₅⁴</td>
<td>telu</td>
</tr>
<tr>
<td>4</td>
<td>pa₅⁴</td>
<td>Sepat</td>
</tr>
<tr>
<td>5</td>
<td>ma₃¹²</td>
<td>RaCep</td>
</tr>
<tr>
<td>6</td>
<td>nam₅⁴</td>
<td>-----</td>
</tr>
<tr>
<td>7</td>
<td>tu₃¹²</td>
<td>-----</td>
</tr>
<tr>
<td>8</td>
<td>ma₀ du₃¹²</td>
<td>-----</td>
</tr>
<tr>
<td>9</td>
<td>va¹¹</td>
<td>-----</td>
</tr>
<tr>
<td>1₀</td>
<td>put₅⁴</td>
<td>sa-iCit</td>
</tr>
</tbody>
</table>
Under Sagart’s theory of subgrouping, sharing the *-mu ‘2sg-genitive’ and *manuk ‘bird’ innovations with Proto-Malayo-Polynesian and with Ketagalan makes the Formosan Ancestor of Tai-Kadai (FATK) a mic language— it should have the post-Proto-Austronesian innovations discussed above, such as the short forms of the numerals 7, 8 and 9

→ these predictions are verified:

<table>
<thead>
<tr>
<th></th>
<th>PPM</th>
<th>P–Kra (Ostapirat 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>five</td>
<td>lima</td>
<td>r–ma A</td>
</tr>
<tr>
<td>six</td>
<td>enem</td>
<td>x–nam A</td>
</tr>
<tr>
<td>seven</td>
<td>pitu</td>
<td>t–ru A / c–tu A²⁶</td>
</tr>
<tr>
<td>eight</td>
<td>walu</td>
<td>m–ru A</td>
</tr>
<tr>
<td>nine</td>
<td>Siwa</td>
<td>s–ya wa B</td>
</tr>
<tr>
<td>ten</td>
<td>puluq</td>
<td>pwlot D</td>
</tr>
<tr>
<td>2sg</td>
<td>-mu</td>
<td>ma A/B</td>
</tr>
<tr>
<td>bird</td>
<td>manuk</td>
<td>ηok D</td>
</tr>
</tbody>
</table>

The original Tai-Kadai numerals are only preserved in the Hlai and Kra branches, in the Kam-Tai branch they have been replaced with Chinese numerals.
The Austronesian phylogeny in space

Sagart’s phylogeny is verified by geography: the locations for Proto-Austronesian, Pituish, Enemish, Walu-Siwaish and Muish can be determined based on the locations of their direct descendants. The most likely location of PAn is the region of Luilang, Saisiat and Pazeh, in the north-west of Taiwan. Pituish must have been spoken in the western plains somewhat to the south of Proto-Austronesian, being ancestral to Atayalic, Favorlang, Taokas, Thao, Papora and Hoanya. Enemish must have been spoken more to the south on the West coast, towards the area occupied by Siraya in recent times. Walu-Siwaish may have been spoken near the southern tip of the island, or on the south-east coast. Muish was probably further north along the east coast, as suggested by the location of its modern descendants Kavalan and Ketagalan and of Bunun.

→ the first Neolithic settlers arrived on the northwest tip from the continent, and their descendants migrated southward, circling the island counter-clockwise. Geographic conditions prevented them from moving clockwise towards the north-east or across the island. This also explains why the descendants of each subgroup still inhabit their ancestral areas.

Figure 2. The An settlement of Taiwan with the MP and TK migrations
[a]: The pre-Austronesians, from NE China, expand southward along the SE China seaboard in the 5th and early 4th millennia BCE, they cultivate rice, foxtail millet, exploit marine resources, practice tooth evulsion. [b]: the Nanri and Pingtan islands, from which the top of Mt Xueshan (3884 m, at center of 200-km radius visibility circle) can be seen, are reached. From there one group crosses to Taiwan c. 3500 BCE, while [c] the rest continues expanding in a SW direction towards the Pearl River Delta. [1]: location of earliest An (PAn-speaking) settlements on Taiwan. [2]: location of Pituish, [3] location of Enemish, [4] location of Walu-Siwaish, [5] location of Muish, [6] Tai-Kadai migration, [7] Malayo-Polynesian migration, c. 2000 BCE.
Finally, under the present interpretation, FAMP and FATK, the two Muic languages whose speakers left Taiwan to settle other regions, were probably located in the north–east or north of the island, where the last available agricultural lands had been. The MP and TK migrations out of Taiwan thus appear motivated by the need to find new agricultural lands. It is probably no coincidence that the site of Yuan-Shan near Taipei, in the region where Ketagalan was spoken until the early 20th century, has significant connections to the earliest neolithic of the Philippines (Bellwood 1997: 215).

The time scale of early Austronesian settlement of Taiwan

The primary evidence for the time scale of the Austronesian settlement of Taiwan comes from archaeology. Bellwood's estimates for the date of the initial Austronesian settlement of Taiwan, inferred from the earliest Ta–Pen–K'eng radiocarbon dates (plus a few hundred years for good measure), and from the earliest neolithic sites in the Cagayan Valley in the northern Philippines, are c. 5500 BP for the former and ca. 4000 BP for the latter (Bellwood 2004). During that period, the daughter languages of PAn were presumably confined to Taiwan. While these are provisional dates, they provide approximate external limits between which the full settlement of Taiwan must have taken place: in the present framework, the initial settlement of the Philippines, presumably by PMP speakers, could not have occurred until Muish had already broken up into its three components. Likewise, the initial Tai–Kadai settlement on the mainland, by FATK speakers, could not have occurred before the breakup of Muish. It is relevant here to recall that Ostapirat estimates the date of PTK to be no older than 4000 BP, simultaneous with the Cagayan dates.
Archeology and language: some conjectures on pre-Austronesian times

I have argued elsewhere (Saqart, in press, b) that the pre-Austronesians spoke a language related to Sino-Tibetan, and that they reached Taiwan from a location in NE China where millet and rice were cultivated, and where ritual evulsion of the upper lateral incisors in boys and girls was practiced. The eastern China seaboard region north of the Yangzi estuary, from north Jiangsu to north Shandong, is the one area in East Asia where the distribution of these three traits overlaps in the period before the arrival of the Austronesians in Taiwan: thus both rice and millet were cultivated in Xihe in north Shandong (Wright 2004) c. 8000 BP and in Longqiu zhuang in the lower Huai basin c. 7000–5000 BP. Tooth evulsion is attested from 6500 BP on in Shandong and north Jiangsu (Han and Nakahashi 1996). We may surmise that before they reached Taiwan, the pre-Austronesians were expanding southward along the coastal plains of central-eastern China in Jiangsu, Zhejiang and north Fujian. We can expect that archaeological sites with rice, Setaria, tooth evulsion, and a technology intermediate between the Dawenkou culture of north-east China and Ta-Pen-K'eng of Taiwan will eventually appear there.

At Nan-kuan-li, near Tainan in south-west Taiwan, a neolithic culture having rice, millet and practicing ritual tooth ablation 5000-4500 years ago was discovered. This culture has close affinities with the Neolithic cultures of Hong Kong and the Pearl River Delta.

⇒ Probably most cultures are descended from a common precursor on the Fujian coast.