Revisiting the expansion of the Chamic language family

Marc Brunelle, Département de Linguistique, Université d'Ottawa

marc.brunelle@uottawa.ca

In this paper, I reconsider two historical scenarios that have become prevalent in the literature on Chamic languages. The first one is that Acehnese is an offshoot of Chamic that arrived in Sumatra directly from Champa (Blust 1992; Cowan 1991; Thurgood 1999, 2007). The second one is that Tsat, a Chamic language spoken by the Utsat people on the southern tip of Hainan, is a direct descendant of a Northern Chamic dialect closely related to Northern Raglai (Thurgood 1999, 2007). My goal here is not to reject earlier proposals in bulk, but rather to sort out the evidence and to establish ranges of historical scenarios compatible with the linguistic data.

I have two reasons for revisiting the two migration scenarios. The first one is methodological. Although multidisciplinary research is essential in understanding historical scenarios for which there is limited documentary evidence, specialists of one discipline are often inclined to choose fragmentary evidence from other fields to support their claims and tend to ignore inconsistent facts and negative evidence. In the cases presented here, weak historical evidence has become a keystone for dating linguistic changes that are in turn used for building further historical scenarios. Secondly, some of the linguistic evidence invoked for justifying the two Chamic migration scenarios is in my opinion weaker than it appears and needs to be revisited.

The structure of the paper is as follows. In §1, I summarize the two migration scenarios proposed by previous researchers. In §2 and §3, I review and evaluate the linguistic and historical evidence about the migration of Chamic speakers to Aceh and Hainan, respectively. Finally, in §4, I define a range of possible scenarios for each of the migrations in light of the conclusions of the previous sections.

Before I start, I need to briefly define some terminological conventions for non-linguists. Languages belong to *families*, and all languages in a family share a common *ancestor language*. These families can be further subdivided into *groups or subgroups* composed of *sister* languages that are the *daughters* of a descendant of the ancestor language. Hence, historical linguists normally conceptualize languages families as

structured into family trees. Languages are recognized as belonging to the same group (or branch in the tree) when they share *innovations*, i.e. have undergone identical phonological changes or borrowed identical words from other languages. This is based on the assumption that shared innovations are more likely to result from a single change in the mother language than from independent changes in sisters.

Historical linguists have a wide array of techniques to *reconstruct* the sound patterns and words of extinct languages by comparing the structures of their daughters. The term "Proto" is normally prefixed to the names of language groups to refer to the *reconstructed languages* ancestral to all their members. Hence, Proto-Chamic (PC) is the reconstructed ancestor of all Chamic languages. Another convention is to precede reconstructed forms with a star. Thus, *a means that the vowel 'a' has been reconstructed by comparing daughter languages, while [a] is a vowel that is actually attested in a living or documented language.

1. Introduction: Chamic expansion outside Indochina

The recognition that Chamic has relatives in Sumatra and Hainan is not recent. The close linguistic proximity between Acehnese and Chamic languages was established by Niemann (1891: 44), who briefly speculated that the Acehnese could have come from Champa. As for Tsat, it was recognized as a Chamic language by Paul Benedict in 1941 after being identified as Austronesian by Stübel and Meriggi (1937).

Despite these early findings, it is only in the 1990s that scholars started formulating historical accounts of the arrival of the Acehnese in Sumatra and the Utsat in Hainan. To my knowledge, the first such account is Cowan (1991), who proposed that the ancestors of the Acehnese and the Utsat left Champa as a consequence of *Nam Tiến*, the gradual conquest of the Cham mainland by the Vietnamese from the 10th to the 17th century. Shortly thereafter, Blust (1992) supported this idea (at least for Acehnese) and proposed that Acehnese speakers came from Champa before 1500 (and even possibly before 1200). A slightly different view has since been expressed by Sidwell (2006), who speculates that Acehnese speakers migrated from Champa around 400AD, possibly because of Chinese raids.

These proposals were further expanded by Thurgood (1999, 2007). Thurgood claims that the Acehnese and the Utsat are both descendants of Chams who fled Central Vietnam after the Vietnamese takeover of the Cham political centres of Indrapura (982) and Vijaya (1471). Two quotes can be used to summarize his views[1]:

> "Two migrations of Chamic speakers can be dated with confidence. The first migration involves what the literature on the Champa Kingdom calls the Northern Cham, Chamic speakers originally located around Hue […]. The linguistic record makes it clear that the modern Hainan Cham, found near Sanya City, and the Northern Roglai, now found in southern Vietnam, were once the same group." (Thurgood 2007, p. 5)

> "The second migration that can be dependably dated correlates with the fall of the southern capital in 1471 […]. The date 1471 fits remarkably well with the earliest date recorded for the Aceh dynasty, a date which is found on a Chinese bell (Reid 2006:10)." (Thurgood 2007, pp. 5-6)[2]

In the rest of this paper, I will argue that the ancestors of the Acehnese are unlikely to have ever transited through Champa and that the Utsat might have settled in Hainan later than 1471. My goal here is not to argue against specific authors or hypotheses, but to review the plausibility of the scenarios that have been previously formulated to account for the establishment of Tsat and Acehnese on Hainan and Sumatra. In order to do so, I will review the linguistic and historical evidence for both proposed migrations.

---

[1] In all fairness, Thurgood is less committed to specific dates in his 1999 book than in the 2007 paper from which these quotes are taken. He also seems to allow for the possibility that both the Tsat and the Acehnese could have migrated in distinct waves roughly correlating with the defeats of 982 and 1471.

[2] The bell cannot actually be linked to either Champa or the Aceh dynasty (Franke 1988, Salmon 2007). See §2.2.
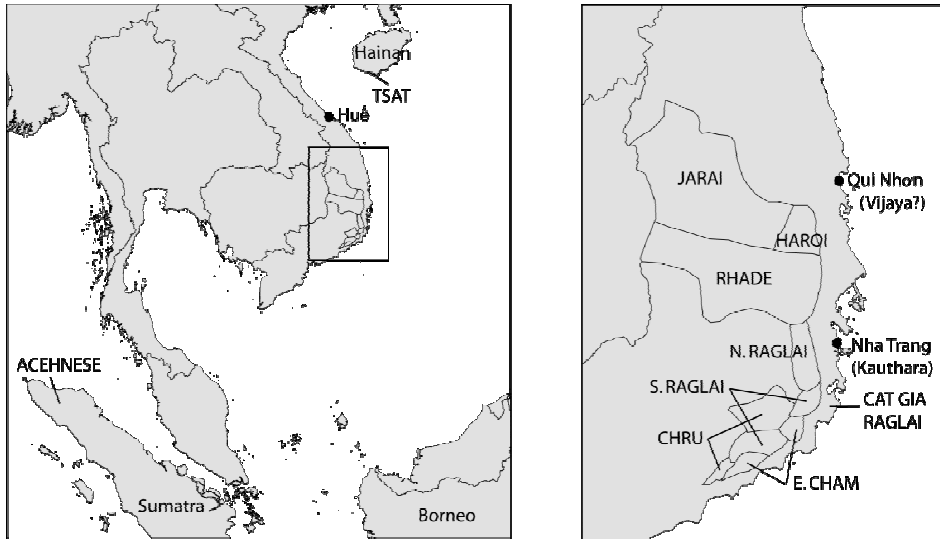
(1) Maps of languages and places mentioned in the paper (Partly adapted from Lee 1966)
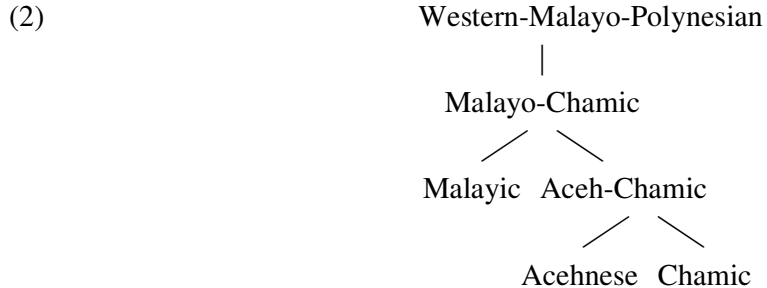

2. Acehnese

2.1. The linguistic evidence

The place of Acehnese and Chamic in the Malayo-Chamic family has long been recognized (Blagden 1929; Blust 1992; Collins 1975; Collins 1991; Cowan 1991; Durie 1985; Niemann 1891; Shorto 1975; Sidwell 2005; Thurgood 1999). Despite occasional dissenting views (Collins 1975; Collins 1991; Peiros 2008), most authors agree that Acehnese and Chamic languages are more closely related with each other than with other Malayo-Chamic languages.

Many names have been proposed for the language family that includes Acehnese and Chamic languages, but this does not reflect major disagreement about the actual sub-grouping of the languages[3]. The only contentious point is that Thurgood (1999) sees Acehnese as a Chamic language on par with other Chamic subgroups rather than as a subgroup that split earlier. I adopt the latter, more conservative view, in the genetic tree of Acehnese and Chamic given in (2). For the sake of simplicity, I will use the terms

---

[3] Achino-Cham (Blust 1992; Shorto 1975), Aceh-Chamic (Durie 1990; Sidwell 2006), Chamic (Thurgood 1999)

Aceh-Chamic to refer to the whole family and *Chamic* to refer to Mainland Chamic languages and Tsat.

(2)   Western-Malayo-Polynesian
                                    |
                              Malayo-Chamic
                             ╱   ╲
                      Malayic   Aceh-Chamic
                                        ╱   ╲
                              Acehnese   Chamic

Without being exhaustive, here are some non-controversial linguistic arguments for a close genetic relationship between Acehnese and Chamic (a systematic review of the linguistic evidence can be found in Sidwell 2005). These are all linguistic innovations that are shared by both subgroups but not by other Malayo-Chamic languages.

First of all, the Proto-Malayo-Polynesian vowel *ə merged with short *a in both Acehnese and Chamic languages. Many instances of short *a, by contrast, became long *aː in Aceh-Chamic and then diphthongized to [ɯə] in Acehnese (Shorto 1975; Sidwell 2006; Thurgood 1999). Second, Proto-Malayo-Polynesian high vowels *i and *u diphthongized to *ɛj and *ɔw in Aceh-Chamic final open syllables, but this diphthongization was blocked in vowels closed by a final *r. The final *r was then dropped, creating contrasts between [i ~ ɛj] and [u ~ ɔj] (Blust 1992; Sidwell 2006; Thurgood 1999). Third, the vowel of the initial syllable of Proto-Malayo-Polynesian disyllabic words was dropped before the consonants *l, *r and *h in both Acehnese and Chamic, leading to the development of the complex consonant clusters *Cr, *Cl, *Ch (Blust 1992; Sidwell 2006; Thurgood 1999). Finally, there is a set of Mon-Khmer loanwords common to both Acehnese and Chamic languages (Blust 1992; Shorto 1975; Thurgood 1999). Sidwell (2005) lists 28 such words.

The fact that Acehnese and Chamic languages share at least four innovations unattested in other Malayo-Chamic languages suggests that these features developed in a common ancestor language after its separation from Malayic. These innovations establish the close connection between Chamic and Acehnese, but they do not reveal

any special proximity between Acehnese and a specific Chamic subgroup. Once again, authors agree:

> "Does Acehnese share any features with particular Chamic sub-groups? (My impression is that it does not, and that therefore the separation of Chamic and Acehnese predates the breakup of the Chamic group)." (Durie 1990, p. 18)

> "Although future research may show that Acehnese is particularly close to one or another of the mainland Chamic languages, the evidence thus far suggests that the Acehnese left Champa before any strongly marked distinction developed among the Chamic languages. It appears that at the time of their departure, the Acehnese were the most northerly of the Chamic groups, covering an area now populated by, among others, the modern Katuic speakers." (Thurgood 1999, p. 42)

However, these two citations also highlight a fundamental difference in the interpretation of the similarities between Acehnese and Chamic. While Durie (1985) does not make claims about a direct connection between Acehnese and Champa, other authors are explicitly proposing that the Acehnese left central Vietnam, and assume that this migration occurred relatively late (Blust 1992; Cowan 1991; Thurgood 1999, 2007). Thurgood (2007), for instance, proposes two waves of migration in 986 and 1471. This is a problematic stance if one believes, as Thurgood reasonably does, that Chamic languages arrived on the coast of Vietnam "some time before 600 BC" (Thurgood 1999 16). It means that Chamic languages would have developed little dialectal differentiation even after being spoken on the Vietnamese coast for more than a thousand years.

Another contentious point raised in Thurgood's citation is the portrayal of Acehnese as the "most northerly of Chamic groups". The linguistic evidence for this claim is the presence of borrowings from Katuic, a Mon-Khmer subgroup, in its lexicon.

"Other evidence of a Chamic contact with Katuic include [sic] apparently Katuic borrowings into Chamic, particularly into Acehnese, and an apparent Austronesian morphological strata in Katu (Reid 1994), which one would presume were due to Chamic influences." (Thurgood 1999, p. 240)

Sidwell (2005, 2006) reviews the evidence and shows convincingly that there are no reasons to assume a special Katuic influence on Acehnese. Out of the 63 Acehnese-Chamic words that have Katuic cognates in Thurgood (1999)'s lexicon, only six appear closer to Katuic than Bahnaric (Sidwell 2007). Further, only three words seem to have been borrowed into Acehnese after the break-up of PC and none of them are closer to Katuic than other Mon-Khmer branches. The existence of Katuic loanwords in Acehnese is thus highly questionable.

Sidwell (2006) also uncovers another interesting distribution: out of the 277 Acehnese-Chamic words of Mon-Khmer origin listed in Thurgood (1999), only 26 could have been borrowed simultaneously in Chamic and Acehnese (9.4%). Further, out of Thurgood's 179 Acehnese-Chamic words of unknown origin, only 18 are found in both subgroups (8.9%). Compare these proportions with the 203 (out of 285) Acehnese-Chamic words of Austronesian origin which are shared by both subgroups (71.2%). The much lower proportion of shared loanwords than shared native Austronesian vocabulary suggests that Acehnese separated from Chamic long before the bulk of loanwords entered the Chamic lexicon (Dyen 2001; Sidwell 2006).

A final issue pointed out by Sidwell (2005; 2006; 2007) is the large proportion of words of unknown origin in Thurgood's lexicon (179 out of 764 PC words). To this number, one should probably add a large number of Chamic words that are cognate with Mon-Khmer words found only in the branches of Bahnaric that were recently adjacent to Chamic-speaking areas and that are likely to have been borrowed into Bahnaric from Chamic. Sidwell (2006, 2007) interprets this high proportion of words of unknown origin to contact with an unknown language "typologically resembling MK at least in terms of word structure" (Sidwell, 2006; 198).

The picture that emerges from the linguistic evidence is once again complex, but suggests that Acehnese may have split from Chamic languages very early. Let me now review the relevant historical evidence.

2.2. The historical evidence

The claim that the Acehnese migrated to Sumatra from Champa was originally made by Niemann (1891), but Cowan is the first author who tried to systematically compare the linguistic and historical evidence in search of such a migration. In a 1991 paper, he presented three main pieces of evidence. He first used data from Malay chronicles:

"According to Shellabear's edition in Arabic characters (1313 A.H. = 1895/96 A.D., pp. 188-192; cf. also his edition in Latin characters of AD 1898, pp. 94-96), the king of Kuchi attacked Champa and occupied its capital Bal. The king of Champa died, and all the sons of the king fled with their following in all directions, no one knows where to, except for two sons: Shah Indra Berma (or Brama) and Shah Po Ling. They fled by boat with many followers ('orang banyak') and their wives and children, the first to Malacca, and Po Ling to Acheh. Indra Brama was well received by sultan Mansur Shah of Malacca, was converted to Islam and became the founder of the Cham colony there. Po Ling was the first of the kings of Acheh ('ialah raja asal raja Acheh')." (Cowan 1991, pp. 59-60)

As pointed out by Cowan, the occupation referred to in the chronicle seems to coincide with the Vietnamese capture of Vijaya-Bình Định in 1471(Kuchi = Giao Chỉ, or Northern Vietnam). Thurgood (2007) also sees a confirmation of this date in a Chinese bell which is "the earliest date recorded for the Aceh dynasty (p.6)". This is a bit of a leap, as this bell bears a date of 1469/1470, which predates the fall of Vijaya, and makes no mention of Champa or of the Aceh dynasty (Franke, et al. 1988; Salmon 2007). Furthermore, even if one takes the Sejarah Melayu literally, there is a big step between accepting that a Cham prince fled to Aceh (and even became its first king) and proposing that there was a large enough migration to explain the linguistic replacement of the existing population (a problem also noted in Thurgood 2007).

Perhaps for this reason, Cowan (1991) turns to other evidence, like the existence of a small town named Jeumpa on the Northern coast of Aceh, a name that he attributes to the presence of an early Cham settlement:

> "The cognacy of the Achehnese *Jeumpa* and the name of the ancient Asian Mainland kingdom must be ascribed to relations between the two, just as in modern times colonists in overseas countries, particularly for instance in the Americas, often took with them geographical names from the mother country to apply them to the new settlements. This leads automatically to the conclusion that there, in Jeumpa, is to be located an area where Cham colonists settled in North Sumatra." (Cowan 1991, p. 66)

If the resemblance between the names Jeumpa and Champa is not a mere coincidence, the name of the settlement could have been chosen because of a symbolic link to Champa (or even the Indian town of Champa), or even because of the presence of a small Cham outpost in Jeumpa. None of these explanations entails the existence of a large Cham community in Aceh, but Cowan does not consider these possibilities. According to him, there was a sizeable Cham population in Northern Aceh and the most probable cause of their migration was:

> "…the wars with Cambodia of 1177-1203 and especially the final phase of them, when the Khmers, in retaliation for a Cham surprise attack (Maspero *o.c.*:164) [1928], invaded Champa, took the capital, and made king Jaya Indravarman IV prisoner. Champa was divided into two kingdoms, Vijaya in the North and Panrang in the South *(ibid.*:165*)*. The latter was attacked again by the Cambodians in 1203 and its king fled abroad. 'Il arrive au port de Co'-la en août 1203, suivi de toute sa famille et de nombre de ses fidèles sur une flotte de plus de deux cents jonques et y demandait asile' (*ibid.*:167), which he did not get, and the king, Maspero concludes, 'reprit la mer et 1'histoire ne nous dit pas ce qu'il devint' *(ibid.).*" (Cowan 1991, p. 67)

The problem with this speculative account is that the port of Cơ-la cited by Maspéro is located in Nghệ An in north-central Vietnam (Maspero 1928, p. 167ff). It probably corresponds to modern Cửa Lò, but in any case, it is definitely not in the vicinity of Aceh.

A scenario in which a large number of Cham speakers would have taken over the entire Northern tip of Sumatra and be the direct ancestors of the 3.5 millions modern speakers of Acehnese is at best improbable (as Blagden pointed out as early as 1929). However, one must also consider the possibility that Cham speakers could have settled in one location in Northern Sumatra and benefited from enough economic and military power to gradually conquer their neighbours and spread their language. This scenario is consistent with the account given by Marco Polo, who described eight Sumatran kingdoms, each with his own language (Durie 1985, p. 2), and also with the expansion of the Acehnese sultanate from the 12th to the 17th century. That said, we must keep in mind that Marco Polo's descriptions are not always trustworthy (he describes the inhabitants of Lambri, one of the eight Sumatran kingdoms, as having tails), and that even if there was language diversity in northern Sumatra, this does not tell us much about a migration from Champa. After all, it is possible that Acehnese replaced closely related languages.

Brief mentions of other possible scenarios for explaining the proximity between Chamic and Acehnese include an Acehnese origin of the Chams (Dyen 2001) and a Funanese origin of the Acehnese (Durie 1985; Sidwell 2005). However, neither of these scenarios is backed up by historical and linguistic evidence.

In short, historical evidence in favour of the arrival of large contingents of Cham refugees to Aceh because of the fall of Cham principalities to the Vietnamese (or another invader) is to this date non-existent. There are numerous reports of diplomatic relations and marriages-of-state between Cham rulers and other Southeast Asian polities, and the presence of Cham merchants throughout Southeast Asia is well-attested (Reid 2000), but it is unlikely that a major migration in the past millennium would have occurred without leaving more archaeological or historical traces.

3. The Utsat of Hainan

3.1. The linguistic evidence

Tsat was first recognized as Chamic by Benedict (1941), but the first attempt at using linguistic evidence for dating the arrival of the Utsat in Hainan was made by Cowan (1991). Cowan speculates that there were two waves of migration from Champa to Hainan: in 986 and in 1471.

Cowan uses two linguistic arguments to support an early migration in 986. The first one is that while Cham (i.e. Eastern Cham) has raised *a to [ɨ] after nasals, Tsat still preserves [a] in this environment. If Tsat is indeed related to Eastern Cham, this suggests that the split between the two languages occurred before a-raising happened, which, for unexplained reasons, Cowan assumes must have taken place in the distant past. Cowan's second argument is that Cham has borrowed Malay numerals while Tsat has Kadai numerals. However, as we now know that Cham has regularly inherited Malayo-Chamic numerals, this argument has to be abandoned (Blust 2010).

Cowan's evidence for a second wave of migration to Hainan in 1471 is based on two linguistic innovations shared by Tsat and Cham. First, while Highlands Chamic preserves word-final stops (i.e. consonants involving a full closure of the oral cavity), Cham has mostly debuccalised them (i.e. *-p has become [-wʔ], *-c has become [-jʔ] and *-k has become [-ʔ]). Tsat has gone even further and has dropped final stops altogether, replacing them with tones (Maddieson and Pang 1993; Thurgood 1993). Second, Tsat and Cham have both lost their voicing contrast in syllable-initial obstruents (i.e. the difference between b~p, d~t and g~k), while Highlands Chamic languages still preserve it. The fact that Cham and Tsat share features unattested in Highlands Chamic is taken by Cowan as evidence for a late migration of the Utsat from the coast after the break-up of Proto-Chamic into divergent sister languages.
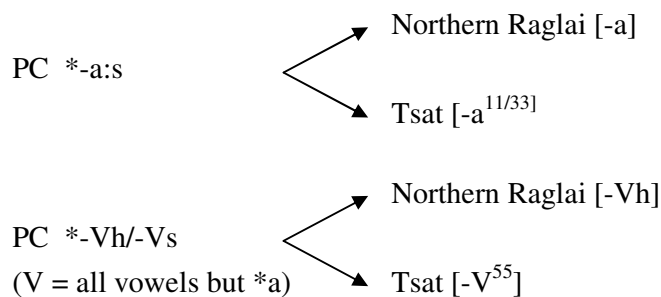
Unfortunately, these facts alone do not support the view that there were two waves of migration to Hainan. Cowan's arguments are actually more consistent with a single migration at a time when coda stops were already debuccalised and onset voicing lost, but when post-nasal a-raising had not yet taken place in Eastern Cham (if Tsat did in fact branch out from Eastern Cham). Since post-nasal a-raising is marked in the *akhăr thrah* script used in Cham manuscripts from the 17th century onwards, this means that

the migration to Hainan would have taken place at some point after Cham separated from Highlands Chamic (a split that is undated), but before the 17th century.

However, additional data must be considered. Thurgood (1999) puts forward strong linguistic evidence that Tsat is more closely related to Northern Raglai (which he treats as a Highlands Chamic language) than to Cham. Tsat and Northern Raglai share two innovations unattested in other Chamic languages: they have both lost the PC final consonant *-s in *-a:s rhymes and they both show evidence of preplosion of nasal codas (a *coda* is a syllable-final consonant). These innovations need to be reviewed in detail.

In Northern Raglai, the Proto-Chamic codas *-h and *-s have merged into [-h]. The only exception to this merger is the Proto-Chamic rhyme *-a:s, in which *-s was dropped, yielding [-a]. A similar distribution is found in Tsat: in that language, Proto-Chamic *-h and *-s have merged into *-h, which was then replaced with a high level tone (marked [55]). Moreover, the exact same exception is found in Tsat as in Northern Raglai: PC *-a:s has lost its final *-s. This can be inferred from the fact that words that had the rhyme *-a:s in Proto-Chamic all have low or mid level tones (marked [11] and [33]) in Tsat, two tones that are normally found in words that had open syllables (i.e. syllables without codas) in Proto-Chamic. This is illustrated in (4) below:

(4)   Northern Raglai and Tsat reflexes of Proto-Chamic (PC) *-s and *-h (based on Thurgood 1999)

PC *-a:s   Northern Raglai [-a]
                               Tsat [-a$^{11/33}$]

PC *-Vh/-Vs   Northern Raglai [-Vh]
(V = all vowels but *a)   Tsat [-V^{55}]

Note, however, that the changes summarized in (4) are not without exceptions. Four irregular words shared by Northern Raglai and Tsat are listed in Thurgood (1999) and reproduced in (5). The lack of correspondence between the Northern Raglai and Tsat finals in these forms could be exceptional, but the fact that the Indic loans 'vehicle' and

'horse' are treated differently in the two languages could also indicate that they were borrowed independently. As it is unlikely that the Utsat borrowed Indic words, directly or via Indianised languages, after their migration to Hainan (especially if it happened as late as the end of the 10th century), this could be an argument against a late separation of Northern Raglai and Tsat.

(5)   Proto-Chamic *ratus 'hundred', N. Raglai [ratuh], **Tsat [tu$^{33]}$**

Proto-Chamic *ɓuh NEG, N. Raglai [ɓuh], **Tsat [pu$^{33]}$**

Post- Proto-Chamic *radɛh 'vehicle', N. Raglai [radeh], **Tsat [the$^{11]}$**

Post- Proto-Chamic *asɛh 'horse', N. Raglai [aseh], **Tsat [se$^{33]}$**

The second innovation shared by Northern Raglai and Tsat is an apparent preplosion process (*-m, *-n, *-ŋ > *-p͡m, *-t͡n, *-k͡ŋ) that would have led to the glottalisation of nasal codas in Tsat (*-p͡m and *-t͡n > [-nʔ]; *-k͡ŋ > [-ŋʔ]) and to their denasalisation in Northern Raglai (*-p͡m, *-t͡n, *-k͡ŋ > [-p], [-t], [-k]) (Thurgood 1999). Interestingly, an identical exception to this process is found in both languages: the presence of a nasal onset (an *onset* is a syllable-initial consonant) results in the maintenance of Proto-Chamic nasal codas[4]. Obviously, the existence of similar innovations subject to the exact same exception in the two languages is strong evidence for grouping them together. However, the details of the preplosion process are less straightforward than they first appear. As Thurgood (1999) explains, the effects of preplosion in Tsat seem limited to nasal codas following short [a] and [ɔ]. There are too few examples of words containing Proto-Chamic nasal codas in Thurgood's Tsat lexicon to establish clear correspondences for all vowels, but there is enough to show that the PC vowel *u is treated differently in Tsat and Northern Raglai. Proto-Chamic rhymes of the *-uN type surface with a denasalized coda in Northern Raglai, but with a nasalized coda in Tsat, as shown in (6).

---

[4] Thurgood (1999) proposes a uniform preplosion of all nasal codas, followed by a renasalisation in syllables with nasal onsets.

(6) PC ʔiduŋ 'nose'   N. Raglai idũk   Tsat tʰuŋ¹¹

PC jarum 'needle'   N. Raglai jurup   Tsat sun¹¹

PC pʰun 'tree'   N. Raglai pʰut   Tsat pʰun³³

PC rabung 'bamboo shoot'   N. Raglai rubuk   Tsat pʰuŋ¹¹

PC trun 'descend'   N. Raglai trut   Tsat tsun³³

Besides *u, PC long vowels also trigger different changes in nasal codas. In Tsat, nasal codas fail to undergo glottalisation after long vowels, but they denasalize regularly in Northern Raglai. In any case, the most salient difference between the two languages is the result of preplosion itself. While Northern Raglai exhibits complete denasalisation, Tsat shows final glottalisation (or even sometimes vowel nasalization accompanied by the loss of the final nasal). It is not unreasonable to consider that these two outputs stem from a single original process, but the path described by Thurgood crucially relies on a non-standard technique of "reconstitution" (rather than on a regular reconstruction) based on four "sandhi" forms whose use and conditioning are not described or explained. Lastly, the presence of spontaneous glottalisation in the Tsat reflexes of PC *-ay and *-aw, realized as [-aːiʔ] and [-aːwʔ] leaves open the possibility that final glottalisation has sources other than preplosion. Therefore, while there is no doubt that the evolution of *-aːs and of nasal codas are important clues in finding who the closest cousins of the Utsat are, more evidence is needed before we can claim for certain that they are the Northern Raglai. In fact, other Raglai dialects (Southern and Cát Gia Raglai), though poorly described, also exhibit coda denasalisation (Lee 1998; Nguyễn 2003).

Other pieces of linguistic evidence also need to be reconsidered besides nasals and final -h, such as monosyllabization, a feature that Tsat shares with Eastern Cham. In addition, onset devoicing and the loss of codas are two Tsat innovations not only found in Cham, but also in other Coastal Chamic languages like Haroi, Southern Raglai and Cát Gia Raglai (Brunelle 2009; Cowan 1991; Lee 1998; Nguyễn 2003). Thurgood dismisses this last point by treating it as the result of a convergence in linguistic sub-areas:

"The devoicing in Tsat and the Coastal Chamic are instructive for two reasons: first, the two occurred independently of each other and, second, the reflexes correlate with different linguistic sub-areas. In the southern Vietnamese highlands, Rade, Jarai, Chru, and N. Roglai have preserved the original PC [Proto-Chamic] voicing; in the Hainan linguistic sub-area Tsat has devoiced, and in the linguistic sub-area along the coastline, the Haroi and Cham have undergone devoicing." (Thurgood 1999, p.83)

However, rejecting the relevance of devoicing off-hand (by circularly decreeing that it occurred independently in Tsat and Coastal Chamic) is no more logical than retaining it and dismissing nasal preplosion as irrelevant. Depending on the innovations one chooses to focus on, one obtains very different conclusions about the subgrouping of Tsat. A successful classification will have to take all innovations into account.

A last type of evidence militates against classifying Tsat and Northern Raglai in the same subgroup. A survey of Thurgood (1999)'s lexicon reveals that a much larger proportion of the Chamic words borrowed from Mon-Khmer are preserved in Northern Raglai than in Tsat (68.6% vs. 18.4%). Interestingly, the difference in rates of retention of the native Austronesian vocabulary is not as large (81.8% vs. 52.3%). The proportions are given in (7).

(7) Proportion of PC lexicon found in Northern Raglai and Tsat, by origin

PC words of Austronesian origin	Number of words in Thurgood's lexicon	**285**	
	Number of cognates in Northern Raglai	233	(81.8%)
	Number of cognates in Tsat	149	(52.3%)
	Number of cognates in both	137	(48.1%)
PC words of Mon-Khmer origin	Number of words in Thurgood's lexicon	**277**	
	Number of cognates in Northern Raglai	190	(68.6%)
	Number of cognates in Tsat	51	(18.4%)
	Number of cognates in both	39	(14.1%)
PC words of uncertain origin	Number of words in Thurgood's lexicon	**179**	
	Number of cognates in Northern Raglai	126	(70.4%)
	Number of cognates in Tsat	30	(16.8%)
	Number of cognates in both	28	(15.6%)

A first possible interpretation of these data is that Chamic words were replaced with Chinese and Tai-Kadai loanwords (or even new words) after arrival in Hainan. However, this interpretation forces us to propose that while Tsat replaced about 30% more native Austronesian words than Northern Raglai, it replaced 50% more Mon-Khmer loanwords than the latter. This is extremely unlikely, because speakers of a language are not usually aware of the origin of the words they use and thus cannot consciously decide to keep or replace words based on their provenance. This leaves a second possibility: a large proportion of the Mon-Khmer vocabulary reconstructed to PC by Thurgood was actually borrowed into Northern Raglai after the separation between the two languages. This could either be because the Tsat left to Hainan very early (before other Chamic languages underwent most of their contacts with Mon-Khmer) or more likely, because the forms reconstructed to PC by Thurgood (1999) have actually been borrowed independently from very similar Mon-Khmer languages after the break-up of PC (and in the case at stake, after the break-up of the common ancestor of Northern Raglai and Tsat). Independent borrowing from distinct but related Mon-

Khmer languages is probable if Mon-Khmer-speaking populations were found in (or close to) coastal areas of Champa rather than being confined to the highlands (Hardy and Nguyễn 2012; Vickery 2005).

In the end, the linguistic evidence is far from conclusive. It does suggest that Tsat and Northern Raglai are not-so-distant relatives, but it also points to the possibility of a relationship with other Chamic languages. One scenario that could explain why Tsat seems to share features with many different Chamic languages is the existence of a patchwork of closely related Chamic dialects in late southern Champa. These dialects would have shared the same incipient changes (optional coda denasalisation, partial debuccalisation, trend towards the loss of *-h after *a:, tendency to devoicing in onset stops), but at different relative frequencies. The ancestors of the Utsat would have brought along a subset of that variation to Hainan. Over centuries, they would have retained some variants and abandoned others, while their mainland cousins were in the meantime selecting their own variants, leading to a situation in which innovations are shared with too many different sister languages to allow the reconstitution of a traditional classification tree. Obviously, in order to obtain a more complete scenario, other factors such as dialectal contact and convergence will also need to be considered (a view partly articulated in Lee 1998), as well as the possibility that the forebears of the Utsat did not originate from a single location but from several areas of the Cham homeland, in which case Tsat could be a *koinè*, or a compromise variety between divergent dialects.

3.2. The historical evidence

The historical evidence for a Chamic migration to Hainan is based on Utsat tradition and on Chinese historical sources. The two oral traditions recorded by Stübel (1937) and cited in Benedict (1941) are that the Utsat came from Guangdong under the Song dynasty (960-1279) or that they came from Vietnam in the early 16th century. Chinese sources partly support the two stories:

"Zheng (1986:37) notes that the *History of the Song Dynasty* (960-1279) makes it clear that some of the northern Cham went to Hainan. Specifically, in 986, Pu-Luo-E and a hundred of his clan arrived, having not just fled Zhancheng (Champa) but

having been harassed by the people of Jiaozhi, the name given by the Northern Sung dynasty to northern parts of Vietnam." (Thurgood 1999, p. 22)

"With reference to Hainan, again citing from Zheng (1986:37), in 1486 the *True Records of the Emperor Xian Zong of the Ming Dynasty* (1368-1644) record over 1000 new refugees in Hainan from Zhancheng (Champa). With reference to Guangzhou (Canton City), in 988 AD the Guangzhou records report 310 refugees from Zhancheng." (Thurgood 1999, p. 22)

What is less clear is where exactly these Chamic speakers came from and why they left Champa (or whatever polity the term "Zhancheng" was used to refer to). Thurgood (2007) explicitly tries to establish a link between the two waves of migrants and the fall of Indrapura (somewhere near present-day Huế) and Vijaya, respectively.

"However, one part of the group is now on Hainan Island and the other part is now inland in southern Vietnam; the modern distribution only makes sense if they both descend from the Northern Cham, that is, the northern part of the Champa kingdom, with the merchants fleeing to Hainan and the non-merchant class moving south with the fall of the northern capital. This scenario also explains why the Northern Roglai are reputed to have had in their position some of the royal regalia from the kingdom of Champa---when the northern capital fell, it was the royal capital. In addition, the Chinese dynastic records date the arrival of the Hainan Cham. The northern capital of Champa, Indrapura, fell in 982, splitting the Northern Cham into two, with those engaged in trading migrating to the southern tip of Hainan Island while the remaining Northern Cham, most of those engaged in farming, crafts, and the like, migrated south. The Chinese records show these refugees arriving in Hainan and paying tribute to the Chinese just four years after the fall of Indrapura." (Thurgood 2007, p.5)

"The second migration that can be dependably dated correlates with the fall of the southern capital [Vijaya] in 1471." (Thurgood 2007, p. 5)

Although the match between the dates found in Chinese records and two notable Cham military defeats is striking, the passages above over-interpret the evidence. First of all, Champa was not a centralized kingdom, but a feudal alliance of more or less united polities (Gay 1994; Southworth 2001; Vickery 2005). Each small polity had its own rulers and its own regalia. In present-day south-central Vietnam, a set of royal clothes and jewellery is still associated with each Cham royal temple and under the custody of a specific Chamic village (Cham, Raglai or Chru). Secondly, Thurgood does not explain how he concludes that merchants fled to Hainan while non-merchants went south after the fall of Indrapura. The idea that farmers and artisans would have fled 500 kilometres away to Khánh Hòa to become the Northern Raglai is rather implausible. Such a long journey would not have been necessary given that the fall of Indrapura only entailed the loss of Quảng Bình: refugees would have been able to find asylum much closer (Southworth 2001; Vickery 2005). Besides, if the Northern Raglai were culturally and linguistically related to the people who were already living in Khánh Hòa, one wonders why they did not assimilate but instead maintained a separate linguistic identity for 500 years. Finally, today's Northern Raglai live in the mountains of Khánh Hòa province, not on its coast, which means that upon arrival in the south, their ancestors would have had to establish themselves in a very unfamiliar environment. This is all very implausible.

Another problem with Thurgood's interpretation of the historical data is the assumption that the Northern Raglai are totally distinct from neighbouring Southern and Cát Gia Raglai. While Raglai groups are not as homogeneous as their common name suggests, Raglai dialects all seem to share some features such as vowel nasalization and the denasalisation of coda stops (Lee 1998; Nguyễn 2003). In the end, if there is really a close linguistic proximity between Tsat and Northern Raglai, perhaps it is simpler to claim that the ancestors of the Utsat came from further south than assumed by Thurgood, which would also explain why Tsat shares features with present-day Coastal Chamic languages. There is in fact some evidence that a late migration to Hainan came from the southern part of the Cham territory. The *True Records of the Emperor Xian Zong of the Ming Dynasty* actually mention that Gu-Lai, who led a thousand Chams to Hainan in 1486, was the brother of Zhai-ya-ma-wu-an, the ruler of a small principality "in the south of the original Champa" (Wade 2012, p. 16).

4. Revisiting the Chamic past

4.1. *The Aceh-Chamic homeland and the settlement in Mainland Southeast Asia*

What is the range of possible scenarios accounting for the linguistic relation between Chamic and Acehnese? In order to answer that question, we must review the history of Austronesian settlement in Mainland Southeast Asia from the time when the two languages were still undifferentiated. In a recent paper, Robert Blust (2010) proposes that Malayo-Chamic can be grouped with the languages of Northern Borneo (and probably Moken, Rejang and Sundanese) into a Greater Northern Borneo family that was spoken in Borneo. At some point, Malayo-Chamic speakers would have branched out and migrated to Mainland Southeast Asia (Blust 1992). Blust (1992) speculates that they would have left the Kapuas river basin around 300-200 BC and that they would have formed a dialectal chain ranging from the Malay Peninsula to Central Vietnam. The latter idea was first proposed by Blagden (1929).

Leaving speculation aside, what is well established is that Aceh-Chamic speakers once formed a single linguistic group in Borneo (possibly not yet differentiated from Malayo-Chamic) and that they sailed to the Mainland in the last few centuries BC. Based on the linguistic evidence reviewed above, we also know that before splitting, they underwent several common phonological innovations and borrowed a sizeable amount of common vocabulary from either an atypical Mon-Khmer language or a language typologically similar to Mon-Khmer (Sidwell 2006). When and where could these innovations have happened?

A first option would be in Borneo, before leaving for the Mainland. The main problem with this view is that the evidence that Mon-Khmer or related languages were ever spoken in Borneo is limited (but see Adelaar 1995; Blench 2010). A second option would be that Aceh-Chamic speakers first settled somewhere in Mainland Southeast Asia, were in contact with Mon-Khmer or other Austroasiatic speakers (Mon-Khmer is a branch of Austroasiatic), and then split into distinct groups, one eventually going west to Aceh while another went east to central Vietnam. Note that both of these options leave open the possibility that Aceh-Chamic speakers once occupied a much wider territory than what is usually accepted, which would concur with Blagden's (1929) idea of a dialectal chain. The last option, the only one that has really been explored so far, is

that Aceh-Chamic speakers went straight from Borneo to Vietnam and that the ancestors of the Acehnese later left Vietnam to undertake a long migration all the way to northern Sumatra. Even if this view was correct, such a migration could have taken place very early on and would not have to be correlated with any of the attested events of the history of Champa. Regardless of the scenario chosen, the absence of obvious similarities between Acehnese and a specific Cham subgroup suggests that Aceh-Chamic must have split very early on. If Chamic speakers were already in Vietnam in the 3rd century BC, as is usually accepted, it is almost impossible that no dialectal differentiation would have emerged by 986 AD.

In the end, I believe that we should keep our options open until we have more evidence. We are unlikely to obtain fresh historical and linguistic data on the distant past of Acehnese, but new archaeological or genetic evidence (Peng, et al. 2010) could help us narrow down the range of possibilities in the near future.

4.2. *The migration to Hainan*

Where did the Utsat migrate from, and when? There is no doubt that they came from central Vietnam. More specifically, shared linguistic innovations suggest that they are related to speakers of languages that are currently spoken in the southern part of original Chamic-speaking territories. While some features are only shared with Northern Raglai (PC *a:s to *a), others are shared with all Raglai dialects (behaviour of nasal codas) and some are much closer to Eastern Cham and Southern Raglai (loss of onset voicing, monosyllabization, debuccalisation).

While Thurgood (1999)'s claim that Tsat is closely related to Northern Raglai is supported by enough linguistic evidence to be taken seriously, the contention that Northern Raglai and the Utsat originally came from Indrapura is dubious. In the absence of any evidence that Northern Raglai was ever spoken close to the northern edge of Chamic-speaking territories, it is more economical to postulate that the Utsat left from present-day south-central provinces, where both Raglai and Cham[5] are or were recently spoken. This is not incompatible with the view that the Utsat have taken refuge in

---

[5] The autonym *Utsat*, and the native name of the language, *Tsat*, which are both derived from 'Cham', could themselves be indications that the ancestors of the Utsat viewed themselves as Cham rather than as members of a Chamic group with a different name. However, as ethnonyms shift and change easily, this is at best peripheral evidence.

Hainan after the conquest of the area by the Vietnamese from 1471 (fall of Vijaya-Bình Định) to 1651 (fall of Kauthara-Nha Trang).

This does not rule out the possibility that some Chamic speakers settled in Hainan much earlier, which seems consistent with the waves of refugees recorded in Chinese chronicles. In fact, Hainan could have been a safe haven or a commercial outpost for Chamic speakers for several centuries, with a more or less regular influx of new speakers. However, this does not mean that these early refugees were the direct ancestors of the Utsat.

A complicating issue is the fact that Tsat has far fewer PC words borrowed from Mon-Khmer than the Chamic languages with which it shares innovations. As mentioned in §3, this could either be due to the fact that many of these forms are not reconstructible to PC (contrary to Thurgood (1999)'s analysis), or that the Utsat left for Hainan much earlier than previously assumed (before PC). Given all the innovations that Tsat shares with Cham and Raglai dialects, the first option seems the most plausible.

5. Conclusion

In this paper, I have reexamined the linguistic and historical evidence surrounding two Cham migrations outside Mainland Southeast Asia. This reexamination leads me to propose that Acehnese probably split from Chamic much earlier than usually believed, and might never have transited through Central Vietnam. The Utsat, on the other hand, probably left later than usually proposed (15th-17th century), from an area corresponding to modern-day Khánh Hòa or Ninh Thuận in south-central Vietnam.

The goal of this paper was to reopen the discussion by expanding the range of possible historical and linguistic scenarios that need to be considered in our exploration of Cham migrations outside Mainland Southeast Asia. I believe this discussion should be multidisciplinary as it is almost impossible for a single researcher to command all the various types of methodologies and data sources at hand. The linguistic and historical evidence addressed here should be combined with archaeological, anthropological and genetic evidence in order to draw more definitive conclusions.

6. Acknowledgements

I would like to thank Paul Sidwell, Pittayawat Pittayaporn and the audience of the conference *New Research in Cham Studies*, held in Paris in June 2012, for their feedback and comments on earlier versions of this paper. Special thanks to Claudine Salmon for providing me with information about the Chinese bell mentioned in the paper. I also want to thank Phạm Thị Thanh Hiền for preparing the two maps, Paul Hirschbühler for his help in clarifying Dutch passages from Niemann (1891) and Daryl Chow for proofreading.

7. References

Adelaar, K. Alexander. "Borneo as a Cross-Roads for Comparative Austronesian Linguistics." In *The Austronesians: Historical and Comparative Perspectives*, edited by Peter Bellwood, James J. Fox and Darrel Tryon. 81-102. Canberra: Australian National University, 1995.

Benedict, Paul K. "A Cham Colony on the Island of Hainan." *Harvard Journal of Asiatic Studies* 6, no. 2 (1941): 129-34.

Blagden, Charles O. "Achinese and Mon-Khmer." In *Feestbundel, Uitgegeven Door Het Koninklijk Bataviaasch Genootschap Van Kunsten En Wetenschappen Bij Gelegenheid Van Zijn 150 Jarig Bestaan 1778-1928. Vol. 1*. 35-38. Weltevreden: G. Kolff, 1929.

Blench, Roger. "Was There an Austroasiatic Presence in Island Southeast Asia Prior to the Austronesian Expansion?". *Bulletin of the IndoPacific Prehistory* 30 (2010): 133-44.

Blust, Robert. "The Austronesian Settlement of Mainland Southeast Asia." *Papers from the the Second Annual Meeting of the Southeast Asian Linguistics Society* (1992): 25-83.

———. "The Greater North Borneo Hypothesis." *Oceanic Linguistics* 49, no. 1 (2010): 44-118.

Brunelle, Marc. "Diglossia and Monosyllabization in Eastern Cham: A Sociolinguistic Study." In *Variation in Indigenous Minority Languages*, edited by James Stanford and Dennis Preston. 47-75. Amsterdam: John Benjamins, 2009.

Collins, Ira V. "The Austro-Asiatic Substratum in Acehnese." Berkeley, 1975.

Collins, James T. "Chamic, Malay and Acehnese: The Malay World and Malayic Languages." In *Le Campa Et Le Monde Malais*. 109-22. Paris: Publications du Centre d'histoire et Civilisations de la Péninsule Indochinoise, 1991.

Cowan, Hendrik K.J. "Achehnese Dialects in Connection with Chamic Migrations." In *Vical 2 Western Austronesian and Contact Languages: Papers from the Fifth International Conference on Austronesian Linguistics*, edited by Ray Harlow. 53-83. Auckland: Linguistic Society of New Zealand, 1991.

Durie, Mark. *A Grammar of Acehnese on the Basis of a Dialect of North Aceh.* Verhandelingen Van Het Koninklijk Instituut Voor Taal-, Land- En Volkenkunde, 112. Dordrecht; Cinnaminson, NJ: Foris Publications, 1985.

———. "Proto-Chamic and Achenese Mid Vowels: Towards Proto-Aceh-Chamic." University of Melbourne, ms., 1990.

Dyen, Isidore. "From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change. With an Appendix of Chamic Reconstructions and Loanwords." Review of Thurgood, Graham. *Anthropological Linguistics* 43, no. 3 (fall 2001): 390-94.

Franke, Wolfgang , Claudine Salmon, and Anthony Siu. *Chinese Epigraphic Materials in Indonesia, Vol I Sumatra*. Singapore: The South Seas, 1988.

Gay, Bernard. "New Perspectives on the Ethnic Composition of Champa." In *Proceedings of the Seminar on Champa*, edited by Đình Tế Huỳnh. 43-52. Rancho Cordova: Southeast Asia Community Ressource Center, 1994.

Hardy, Andrew, and Tiến Đông Nguyễn. "Multi-Disciplinary Evidence for a Hypothesis on the Champa Territory and Population of Quảng Ngãi." In *New Research in Cham Studies: an International Conference*. EFEO, Paris, 2012.

Lee, Ernest W. "The Contribution of Cat Gia Roglai to Chamic." In *Papers in Southeast Asian Linguistics No.15: Further Chamic Studies*, edited by David D. Thomas. 31-54. Canberra: Pacific Linguistics - Series A, 1998.

Maddieson, Ian, and Keng-Fong Pang. "Tone in Utsat." In *Tonality in Austronesian Languages*, edited by Jerold Edmondson and Kenneth Gregerson. 75-89. Honolulu: U of Hawaii Press, 1993.

Maspero, Georges. *Le Royaume De Champa* [The Kingdom of Champa]. Paris and Brussels: G. Van Oest, 1928.

Nguyễn, Văn Huệ. "Nasalized Vowels and the Changing of Final Consonants in the Raglai Language." In *Những Vấn Đề Văn Hóa Và Ngôn Ngữ Raglai*. Hồ Chí Minh City: Nhà xuất bản Đại học Quốc gia, 2003.

Niemann, George K. "Bijdrage Tot De Kennis Der Verhouding Van Het Tjam Tot De Talen Van Indonesië. [Contribution to the knowledge of the relationship of Cham to the languages of Indonesia]" *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 40 (1891): 27-44.

Peiros, Ilia. "Malayic, Chamic and Aceh: Some Lexicostatistical Remarks." In *Lsaul 06: Language and Text in the Austronesian World*, edited by Yury A. Lander and Alexander K. Ogloblin. 73-88. Munchen: Lincom Europa, 2008.

Peng, Min-Sheng, Quang Huy Ho, Dang Khoa Pham, Trieu An Vu, Hua-Wei Wang, Yong-Gang Yao, Qing-Peng Kong, and Ya-Ping Zhang. "Tracing the Austronesian Footprint in Mainland Southeast Asia: A Perspective from Mitochondrial DNA." *Molecular Biology and Evoluation* 27, no. 10 (2010): 2417-30.

Reid, Anthony. "Chams in the Southeast Asian Maritime System." In *Charting the Shape of Early Modern Southeast Asia*. 39-55. Singapore: Institute of Southeast Asian Studies, 2000.

Salmon, Claudine. "Transnational Networks as Reflected in Epigraphy: The Case of Chinese Buddhist Bells in Southeast Asia." In *Chinese Overseas: Migration, Research and Documentation*, edited by Chee-Beng Tan, Colin Storey and Julia Zimmerman. 23-84. Hong Kong: The Chinese University of Hong Kong, 2007.

Shorto, H.L. "Achinese and Mainland Austronesian." *Bulletin of the School of Oriental and African Studies* 38, no. 1 (1975): 81-102.

Sidwell, Paul. "Acehnese and the Aceh-Chamic Language Family." In *Chamic and Beyond: Studies in Mainland Austronesian Languages* edited by Anthony Grant and Paul Sidwell. 211-46. Canberra: Research School of Pacific and Asian Studies, Australian National University 2005.

———. "Dating the Separation of Acehnese and Chamic by Etymological Analysis of the Aceh-Chamic Lexicon ". *Mon-Khmer Studies* 36 (2006): 187-206.

———. "The Mon-Khmer Substrate in Chamic: Chamic, Bahnaric and Katuic Contact " Paper presented at the SEALS XII: Papers from the 12th annual meeting of the Southeast Asian Linguistics Society 2002, De Kalb, 2007.

Southworth, William A. "The Origins of Campa in Vietnam: A Preliminary Overview." SOAS, 2001.

Stübel, Hans, and Piero Meriggi. *Die Li-Stimme Der Insel Hainan, Ein Beitrag Zur Volkskunde Südchinas* [The Li family of the island of Hainan, a contribution to the folklore of Southern China]. Berlin: Walter de Gruyter, 1937.

Thurgood, Graham. *From Ancient Cham to Modern Dialects : Two Thousand Years of Language Contact and Change*. Oceanic Linguistics Special Publication ; No. 28. Honolulu: University of Hawai'i Press, 1999.

———. "The Historical Place of Acehnese: The Known and the Unknown." In *First International Conference of Aceh and Indian Ocean Studies*, 12, 2007.

———. "Phan Rang Cham and Utsat: Tonogenetic Themes and Variants." In *Tonality in Austronesian Languages*, edited by Jerold Edmondson and Kenneth Gregerson. 91-106. Honolulu: U of Hawaii Press, 1993.

Vickery, Michael. "Champa Revised." *Asia Research Institute Working Papers*, no. 37 (2005).

Wade, Geoff. "Champa in the Ming Reign Annals (Ming Shi-Lu) 14th-17th Centuries." In *New Research in Cham Studies: an International Conference*. Paris, 2012.