

Wb 23749

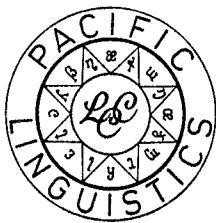
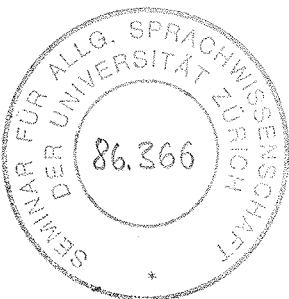
PACIFIC LINGUISTICS

Series B - No. 91

PROTO-SANGIRIC AND THE SANGIRIC LANGUAGES

by

J.N. Sneddon



Department of Linguistics  
Research School of Pacific Studies  
THE AUSTRALIAN NATIONAL UNIVERSITY

PACIFIC LINGUISTICS is issued through the Linguistic  
Circle of Canberra and consists of four series:

- SERIES A - Occasional Papers
- SERIES B - Monographs
- SERIES C - Books
- SERIES D - Special Publications

EDITOR: S.A. Wurm

ASSOCIATE EDITORS: D.C. Laycock, C.L. Voorhoeve, D.T. Tryon, T.E. Dutton

EDITORIAL ADVISERS:

B.W. Bender	John Lynch
University of Hawaii	University of Papua New Guinea
David Bradley	K.A. McElhanon
La Trobe University	University of Texas
A. Capell	H.P. McKaughan
University of Sydney	University of Hawaii
Michael G. Clyne	P. Mühlhäusler
Monash University	Linacre College, Oxford
S.H. Elbert	G.N. O'Grady
University of Hawaii	University of Victoria, B.C.
K.J. Franklin	A.K. Pawley
Summer Institute of Linguistics	University of Auckland
W.W. Glover	K.L. Pike University of Michigan;
Summer Institute of Linguistics	Summer Institute of Linguistics
G.W. Grace	E.C. Polomé
University of Hawaii	University of Texas
M.A.K. Halliday	Gillian Sankoff
University of Sydney	University of Pennsylvania
E. Haugen	W.A.L. Stokhof National Center for
Harvard University	Language Development, Jakarta;
A. Healey	University of Leiden
Summer Institute of Linguistics	E.M. Uhlenbeck
L.A. Hercus	University of Leiden
Australian National University	J.W.M. Verhaar
Nguyễn Đăng Liêm	Gonzaga University, Spokane
University of Hawaii	

All correspondence concerning PACIFIC LINGUISTICS, including  
orders and subscriptions, should be addressed to:

The Secretary  
PACIFIC LINGUISTICS  
Department of Linguistics  
Research School of Pacific Studies  
The Australian National University  
G.P.O. Box 4, Canberra, A.C.T. 2601  
Australia.

Copyright © The Author

First Published 1984

Typeset by Jeanette Coombes

Printed by A.N.U. Printing Service

Bound by Adriatic Bookbinders Pty. Ltd.

The editors are indebted to the Australian National University for assistance in  
the production of this series.

This publication was made possible by an initial grant from the Hunter Douglas  
Fund.

National Library of Australia Card Number and ISBN 0 85883 306 9

## PART TWO

### PROTO-SANGIRIC PHONOLOGY AND DIACHRONIC CHANGES

2.0. In Part Two PSan phonology is reconstructed and the interrelationships of the Sangiric languages are discussed.

First, the phonologies of the present-day languages are briefly stated, in section 2.1.

In section 2.2. PSan phonemes and their distribution within the word are given.

The evidence of the reconstructions and the development of PSan phonology into the present-day systems is stated in section 2.3. Phonological changes are also examined in that section for the light they shed on interrelationships among the languages.

In section 2.4. the relationships between the languages within the Sangiric group are further discussed.

#### 2.1. MODERN PHONOLOGIES

2.1.0. The phonological system of the parent language is reconstructed from a comparison of the phonologies of its daughter languages. The phonologies of the five Sangiric languages are briefly discussed in this section. Some phonological phenomena shared by a number of languages are described separately in sections 2.1.6. and 2.1.7. and some points are left until section 2.3. for discussion.

##### 2.1.1. Sangir phonology

###### 2.1.1.1. Segmental phonemes

The phonology described here is that of the Manganitu dialect.<sup>21</sup> Manganitu has the following segmental phonemes:

p	t	k	?
b	d	g	
ɓ	r	ɠ	
m	n	ɳ	
	s		h
	l	!	
i		u	
e	ə	o	
	a		

When occurring after ə, nasals and l are lengthened: elo [éł:o] day, enuŋ [én:uŋ] six. All other consonants which can occur after ə are preglottalised except h: əbu? [é?bu?] to pull out, əsa [é?sa] one. Since these phenomena are predictable they are not indicated in the phonemic script.

ɓ and ɠ are bilabial and velar fricatives respectively and r is an alveolar flap or trill.

l is a voiced alveolar lateral while ! is a retroflexed lateral flap.<sup>22</sup>

All consonants except ɓ, ɠ, ! and ? occur initially; r occurs initially only in a handful of words, mainly in borrowings from other languages or other San dialects in which r corresponds to Manganitu h.

### 2.1.1.2. Stress

Stress is very largely predictable in San, falling on the penultimate syllable except when the final syllable is ə?, in which case it falls on the third-last syllable, e.g. lɪkudə? back (see section 2.1.6. for a discussion of final syllable ə?).

## 2.1.2. Sangil phonology

### 2.1.2.1. Segmental phonemes

Sangil (Sarangani dialect) has the following segmental phonemes:<sup>23</sup>

p	t	k	?
b	d	g	
ɓ		h	
m	n	ɳ	
	s		
	r		
	l	!	
w	y		
i		u	
e	ə	o	
	a		

The phonetic natures of l and ! are as described for San (see section 2.1.1.1.).

Nasals and l are lengthened after e, as in San, but other consonants are not preglottalised in this position as they are in San.

All consonants except h, w, y and ? occur word-initially although ! occurs initially only in a few words.

### 2.1.2.2. Stress

Stress placement is not predictable in Snl and stress contrasts occur, e.g. kápu to rub and kapú desirous. However, the number of words in which stress does not follow the regular pattern is very small. Regular stress falls on the penult, except if this contains e and the final syllable contains some other vowel, in which case stress falls on the final syllable, e.g. bátu stone, kégəŋ dried but təá? fly, kəŋkún fist.

The final syllable e? does not take stress and does not influence stress placement, e.g. líkude? back, səkóe? cough (see section 2.1.6.).

### 2.1.3. Talaud phonology

#### 2.1.3.1. Segmental phonemes

Tal (Salibabu dialect) has the following segmental phonemes:

p	t	k	?
b	d	g	
ɓ	r	h	
m	n	ŋ	
	s		
	ʐ		
	l	!	
i		u	
e	e	o	
a			

ʐ is a voiced retroflexed fricative.<sup>24</sup>

l and ! are as described for San (see section 2.1.1.1.).

Study of Tal has not been detailed enough to allow a reliable phonemic statement. Voiced stops b, d and g and their corresponding continuants, ɓ, r and h respectively, are in free variation word-initially, at least in some words, e.g. bobone ~ bobone mountain, donna ~ ronna leaf, gati ~ hati face. However, apparent contrast has been recorded word-initially, e.g. rurukka (is) sailing and durukka sail!. Usually only the continuants occur intervocally but the stops occur after some prefixes (see example below). For the present work stops and corresponding continuants are treated as phonemically separate.

In some dialects ʂ corresponds to Salibabu h.

Most consonants occur both singly and doubled. The choice between single and doubled consonants is, synchronically, very largely unpredictable. Some contrasts have been recorded, e.g. maballasa *will borrow* and mabballasa *is borrowing*, allaja *slave* and alanja *to swim*.

Adriani (1911) does not refer to gemination except in reference to its occurrence stem-initially following prefix Cu-, e.g. Cu + dalanna *to walk* + duddalanna *is walking*. Steller (1913) points out that consonants are doubled following a where this is a reflex of a previous \*ə. He writes that this a is phonetically different from a elsewhere. However, no phonetic difference was noted in the speech of informants for this study except that a, like other vowels, is shorter in closed syllables than in open syllables.

Steller does not mention other instances of doubling. However, consonants are also often doubled preceding word-final a, e.g. papaidda *wing*, lanitta *sky*. This doubling occurs only before a where it is an historical addition (see section 2.1.6.) but not where it reflects PPh \*a, e.g. mata *eye* (PPh \*mata).

Doubling of the consonant before final a, as described above, does not always occur and is absent if the preceding consonant is doubled, e.g. allaja *slave*, annuma *six*, or is a nasal-stop cluster, e.g. sandaka *to lean*, sengota *to sail* or is Ň, e.g. bažata *west*, užasa *to wash*.<sup>25</sup>

Germinate stops are sometimes realised as preglottalised single consonants, in free variation with phonetically long consonants, e.g. pappuso [pa?puso] ~ [pap:uso] *heart*.

### 2.1.3.2. Stress

Although only a very brief study of this language was possible it appears to indicate that stress is non-phonemic. Stress usually falls on the penultimate syllable. When the final syllable is the additional vowel a stress often falls on the third-last syllable. However, there appear to be cases which do not conform to the above rules and further study of stress will be necessary. Tal stress is not indicated in this study.

### 2.1.4. Bantik phonology

#### 2.1.4.1. Segmental phonemes

Ban has the following segmental phonemes:

p	t	k	?
b	d	g	
m	n	ŋ	
	s		h
w	y		
i		u	
e	ə	o	
	a		

b, d and g are realised as voiced stops in all environments.

l is a retroflexed lateral flap [l̪] in all environments.<sup>26</sup>

#### 2.1.4.2. Stress

Word stress is unpredictable and therefore phonemic. Stress usually falls on the penultimate syllable: káyu *wood*, báley *house*, makapála? *thick*. In a large number of words it falls on the final syllable, this being synchronically unpredictable: mahémé *full*, mabahá? *heavy*, lumampán *to walk*. In other words stress falls on the third-last syllable: líkudu? *back*, dákele? *many*.

Stress contrast is shown by the following pairs: sáhan̪ *tree sp.* and sahán̪ *ant*, mamúku (stem búku) *to kneel* and mamukú (stem pukú) *to bend*, pahígi *well* and pahigí *knife*.

In the present work stress in Ban is indicated wherever it does not fall on the penultimate syllable of the word.

#### 2.1.5. Ratahan phonology

##### 2.1.5.1. Segmental phonemes

Rth has the following segmental phonemes:

p	t	k	?
b	d	g	
θ	r	h	
m	n	ŋ	
	s		(-)
	l		
w	y		
i		u	
e	ə	o	
	a		

Phoneme θ has the following allophones: voiceless bilabial fricative [ɸ] occurs word-finally and before a consonant: kukáθ [kukáɸ] *wing*, kukáθne [kukáɸne] *its wing*. A voiced bilabial or labiovelar continuant [θ ~ v] occurs elsewhere, being to some degree influenced by adjacent vowels. This set of allophones will henceforth be represented by [θ]: baley [θáley] *house*, libu [lɪθu] *to go around*.

Phoneme h has the following allophones: voiceless velar fricative [χ] occurs finally: balukáh [balukáχ] *chaff*. Voiceless glottal fricative [h] occurs elsewhere:ahasen [hasásen] *sand*.

The voiced stops d and g do not contrast with their corresponding continuants, r and h respectively. The stops occur only after homorganic nasals, initially and medially, e.g. ndipa *rainbow*, indak *to breathe*, ngipu *soot*, munja? *to decrease*. Like h, r (trilled or flapped vibrant [r̪ ~ ř]) occurs initially, intervocalically and finally, e.g. rua *two*, arey *chin*, lar *foot*.

The situation with b and þ is basically the same, b occurring after m, e.g. mbulu *feather*, timbow *to float*, and þ occurring elsewhere, as exemplified above. However, in a few words b and þ occur in free variation intervocalically, e.g. loben ~ loben *big*, labah ~ labah *to throw*. In a few words only b occurs, e.g. masasabunu? *to fight*.

Although stops and corresponding continuants are in complementary distribution, apart from the few instances of contrast between b and þ mentioned above, it has been decided to represent them separately in Rth, at least for the purposes of this comparative study.<sup>27</sup> The main reason for distinguishing between the stops and continuants in this work is for the purpose of comparison with the other Sangiric languages. In San, Snl and Tal voiced stops and continuants, previously in complementary distribution, have become phonemically separate through subsequent sound changes and borrowing (see section 2.3.13.). To represent Rth voiced stops and corresponding continuants also as separate greatly facilitates discussion and description when comparing Rth with these languages.

Vowel-initial words are usually preceded by a non-phonemic glottal stop. Some words, however, never are; in these the initial vowel is either lengthened or has a breathy onset, which is very slight and varies with the following vowel. Preceding u velar constriction results in a faint velar fricative quality [χ]. Preceding i palatal constriction results in a faint palatal fricative quality [ç] to the onset of the vowel. These variants are hereafter all represented [h]. These are here recognised as variants of a phoneme represented by a dash over the initial vowel, (‐) in the chart of phonemes above. Thus: āmu? [a·mu? ~ hamu?] *root*, Ttik [i·tík ~ hitík] *swell*. This phoneme occurs medially only in reduplicated words: ūmitiTtík [hūmitihitík] *is swelling*.

One reason for recognising an initial consonant in such words is that they take infixes -um- and -in-, which otherwise occur only after stem-initial consonants (see the last example above). Also, diachronically, it reflects an earlier glottal fricative \*h, which still occurs in some other Sangiric languages. It still existed, initially and medially, when Niemann produced his wordlist (see section 2.3.9.).

Following a prefix this consonant is lost: ākir [á·kir ~ hákir] → muakir [muákir] *to tie up*.

It is felt preferable to use the diacritical device rather than a separate letter to represent this phoneme as it is more appropriate to its phonetic character, that of a breathy onset or slight lengthening of the vowel rather than a separate initial segment. However, the symbol <h> is still used to represent the glottal fricative phoneme h in the discussion of earlier stages of Rth.

### 2.1.5.2. Stress

Word stress is unpredictable and therefore phonemic. Stress usually falls on the penultimate syllable: káyu *wood*, báley *house*. In a large number of words stress falls on the final syllable: sukól *cough*, tiník *mosquito*. Stress contrast is shown by the following pairs: mundúpa *to put* and mundupá *to measure a span*, niú *winnowing pan* and niú *coconut*.

One exception to the occurrence of stress on the penultimate syllable is when this syllable contains a high vowel and is immediately followed by another vowel in a closed syllable. In such cases stress usually falls on the final syllable. This is further discussed, with examples, in section 2.3.23. When

stress occurs on a word-final high vowel the addition of a suffix may produce the conditions described above and, if so, stress shifts to the new final syllable, e.g. sumú *nasal mucus* + -an → sumuán *having a runny nose*.

Stress on the penultimate syllable is henceforth unmarked; where stress falls on any other syllable it is indicated in the script.

### 2.1.5.3. Long and short vowels

Phonetically long vowels occur in Rth: [ba:] *fire*, [tu:n] *order*. These contrast with short vowels, e.g. [tu:] *egg* and [tu] *embers*, [manú:k] *to draw water* and [mánuk] *fowl*.

Phonetically long vowels are interpreted as sequences of two identical short vowels. Where phonetically stress falls on a long vowel it is treated phonemically as falling on the first of the two identical vowels,<sup>28</sup> e.g. [manú:k] manúuk *to draw water*, [rá:mi] ráami *deserted rice-field*. In such cases stress is not further indicated in the phonemic script. However, it is indicated where it falls elsewhere, e.g. [ka:bí] kaabí *last night*.

A long vowel reduces to a short vowel if another vowel occurs adjacent to it. Thus ka- + kii *to grate* + -an → kakián *grater*.

### 2.1.6. The paragoge

In San, Snl and Ban an extra syllable is added after all final consonants except n and ?. This syllable is θ? in San and Snl and V? in Ban, where V assimilates to the preceding vowel.

As mentioned in the descriptions of the individual languages above, this syllable does not take or influence stress, the preceding syllable acting as the word-final syllable for the purpose of stress placement. Thus San úhasə?, Snl úrase?, Ban úhasa? *to wash*. Here stress is regular in San and Snl, falling on the 'second-last' syllable; the support syllable is distinctive and stress placement is therefore predictable. But for Ban the support syllable is not distinctive as the sequence V? is often the final syllable of the root morpheme, e.g. kinaka? *older sibling*. Therefore for Ban the support syllable must be counted when stress placement is calculated and this, in addition to other factors (see section 2.3.1.), results in unpredictable stress placement. Thus úhasa?, with stress on the antepenultimate syllable, and kinaka?, with stress on the penultimate.

For ease of description and because of its precise definition the term paragoge will henceforth be used to refer to this support syllable.<sup>29</sup>

In Tal the paragoge is vowel a, usually accompanied by doubling of the preceding consonant, as described in section 2.1.3.1. In Tal the paragoge occurs after all final consonants, including nasals, e.g. užasa *to wash*, itumma *black*.

Rth is the only language in which there is no paragoge. Thus uras *to wash*, itum *black*.

In all languages the paragoge is lost if there is a suffix beginning with a vowel. Thus, San, Ban uhaseŋ, Snl uraseŋ, Tal užasanna *will be washed* (with the paragoge reappearing after the suffix-final n in Tal). One exception is with San roots of one syllable, which do not lose the paragoge before a vowel-initial

suffix, e.g. *halə?* *work* + -e → *halə?e* *his work*. Loss of the paragoge is synchronically unpredictable in Snl when suffix -e *his, its* occurs, e.g. *bagə?* *strength* + -e → *bagə?e* *its strength* but *bade?* *payment* + -e → *bade* *its payment*.

In a systematic phonemic statement the paragoge would be non-phonemic in San as its occurrence is predictable, e.g. for uhas the only possible pronunciation would be [úhasə?].

In Snl the diachronic loss of \*! subsequent to the development of the paragoge (see section 2.3.10.) has resulted in its not being entirely predictable, as can be shown with the word *səkəe?* *cough* (from earlier \*səkələ?). From the representation *səko* the surface form *səkəe?* could not be predicted as the paragoge does not normally immediately follow a vowel.<sup>30</sup> Further, the irregular retention of the paragoge with suffix -e, as described above, requires it to be given phonemic status.

Because of the discussion it requires (see section 2.3.3.) and to conform with all previous publications on these languages, in this study the paragoge is represented wherever it occurs.

### 2.1.7. Final nasals

In San, Snl and Ban the only word-final nasal is ɳ. This frequently reflects a previous \*m or \*n (see section 2.3.5.).

When a suffix is added in Ban, ɳ is replaced by the original nasal, e.g. *inuŋ* *to drink* + -an → *inuman* *will be drunk*, *suaŋ* *to plant* + -en → *suanen* *will be planted*.

In San ɳ generally remains when a suffix is added. Thus *undəŋ* *to treat medically* (PSan \*undam) + -en → *undajen* *will be treated*, *sueŋ* *to carry on the head* (PSan \*suən) + -aŋ → *suəŋŋən* *will be carried on the head*.<sup>31</sup>

In Snl final ɳ is replaced by n before a suffix irrespective of what the nasal was in PSan. Thus *inuŋ* *to drink* (PSan \*inum) + -aŋ → *inunaŋ* *will be drunk*, *suaŋ* *to plant* (PSan \*suan) + -en → *suanen* *will be planted*, *tubəŋ* *to fell* (PSan \*tubəŋ) + -en → *tubanen* *will be felled*.

In a systematic phonemic statement of Ban m and n would be represented word-finally in those words where they occur preceding a suffix, their surface replacement by ɳ in final position being predictable. However, for this comparative study ɳ is shown as the final nasal in all cases.

## 2.2. PROTO-SANGIRIC PHONOLOGY

### 2.2.1. Segmental phonemes

The following are the segmental phonemes reconstructed for PSan:<sup>32</sup>

p	t	k	?
b	d	g	
m	n	ŋ	
	s		R
	l		
w	y		
i		u	
e	ə	o	
	a		

The phonetic nature of \*R is not certain (see section 2.3.9.).

Voiced stops \*b, \*d and \*g probably had continuant allophones in some positions in the word (see section 2.3.13.).

The lateral \*l had two allophones: alveolar lateral [l] and retroflexed lateral flap [ɿ] (see section 2.3.10.).

In San, Tal, Ban and Rth t is dental and d is alveolar. The same points of articulation are therefore postulated for PSan.<sup>33</sup>

## 2.2.2. Stress

Stress was non-phonemic, falling on the penultimate syllable (see section 2.3.23.).

## 2.2.3. Distribution of phonemes

From a comparison of the Sangiric languages the distribution of phonemes within the morpheme can be reconstructed as follows:

All vowels occurred in initial position:

*atup	roof	*ikit	to tie
*ebeR	saliva	*olay	spadix
*əRab	to sharpen	*ulid	worm

All vowels occurred in medial position:

*tali	rope	*lima	five
*benel	deaf	*tolay	tail
*pəku	to bend	*kumi	moustache

One restriction was that \*ə did not occur in either open or closed final syllables. The other five vowels occurred finally:

*lima	five	*buno	to kill
*pile	to choose	*pitu	seven
*tali	rope		

Two vowels could occur in sequence. The only clusters reconstructed within the morpheme involving \*ə are \*aə, \*iə and \*uə. Sequences of two identical vowels did not occur. Examples of vowel sequences:

*polæn	wrist	*sai	who
*kaən	food	*tuadi	younger sibling
*taun	year	*luəm	overcast
*sie	he	*loŋ	wide
*Riud	to pull	*tuid	stump

Sequences \*eə and \*oə occurred at morpheme boundaries, i.e. where stem-final mid-vowels were followed by the suffix \*-ən (see section 2.3.1.).

All consonants except \*w, \*y and \*? occurred initially:

*paɪt	bitter	*mata	eye
*baRi	rotten	*niuR	coconut
*tau	person	*ŋuda	young
*dalen	road	*sanu	steam
*kayu	wood	*ləgøy	to laugh
*gudaŋ	old	*Ramut	root

All consonants occurred intervocally except \*?:

*lipan	centipede	*tanak	to live
*səbu	froth	*sanu	steam
*batu	stone	*tasik	sea
*gudaŋ	old	*telak	to fly
*səkol	cough	*bəRu	new
*tages	reef	*kayu	wood
*lima	five	*awak	body

All consonants occurred finally:

*əlup	to swallow	*ləbiŋ	grave
*səsub	steam	*bitis	calf
*bəRat	heavy	*sadeR	to lean against
*likud	back	*kəpal	thick
*manuk	bird	*ləgøy	to laugh
*bulag	blind	*talaw	cowardly
*inum	to drink	*leRe?	neck
*luan	to exchange		

The only consonant clusters within the morpheme were sequences of nasal plus homorganic stop or \*s:

*kumpas	tree sp.	*lansik	to bounce
*bəmbulu	feather	*kəŋkum	fistful
*pəntas	to harvest	*tingum	riddle
*dəndum	dark		

It is probable that other clusters occurred at morpheme boundaries, especially between a root and a following suffix or enclitic, but this cannot yet be established because of the different patterns in the present-day languages.

## 2.3. DIACHRONIC CHANGES

2.3.0. This section examines the phonological changes which occurred during the development of the five present-day languages from PSan. The changes are also examined for the light they shed on the historical relationships between the languages.

It will be noted that there occurs some apparently conflicting evidence for relatedness among the languages, i.e. where languages A and B share an innovation suggesting a subgroup excluding C while B and C share another innovation suggesting a subgroup excluding A. Many such features must be the result of diffusion. In particular there are a number of common innovations in San and Tal which do not occur in Snl. The very close relationship between San and Snl is shown by their large common lexicon, including a high percentage of shared vocabulary in the lexicostatistics list, and their very similar phonologies. They can rightly be regarded as dialects, or rather as dialect clusters, each with its own sub-dialects, of one language, although for the purposes of the present comparative study they are referred to as separate languages. Therefore, innovations shared by San and Tal, but not Snl, must be regarded as the result of areal spread. The same conclusion applies where Ban shares an innovation with San and Snl not found in Tal, as Tal is demonstrably closer to San and Snl than is Ban. Evidence for this is presented in a number of subsections below.

It is shown (see section 2.3.2.) that San, Snl and Tal form a subgroup, referred to as the North Sangiric languages. Their immediate parent language is called Proto-North-Sangiric (PNSan). Because of the shared history until recently of San and Snl and thus the frequent need to refer to them together in comparison with the other languages the abbreviation San/Snl is often used below. Ban and Rth are called the South Sangiric languages (see section 2.4.) and their immediate parent language is called Proto-South-Sangiric (PSSan).

### 2.3.1. Reflexes of schwa

The six-vowel system of PSan was reduced in Tal, Ban and Rth through the loss of \*ə or its replacement by some other vowel.

PSan \*ə has been retained in San in all environments. In all other languages it was lost immediately following another vowel within the same morpheme:

PSan	San	Snl	Tal	Ban	Rth	
*kaən	kaəŋ	kaŋ	anna	kaŋ	kan	to eat <sup>34</sup>
*luəm	luəŋ	-	lumma	luŋ	lum	shade

Between morphemes San retains sequences aə, eə and oə, i.e. when suffix -əŋ follows stems ending in a, e or o. According to Adriani these usually assimilate to long vowels a:, e: and o: respectively. But Steller and Aebersold give forms with ə, rather than lengthening of the preceding vowel. Thus Adriani (1893:157) pəndareno:ŋ, Steller and Aebersold pəndarenoŋ *bathing place*, from deno to bath.

In the other languages the same reduction occurs between morphemes as occurred diachronically within morphemes; suffix \*-əŋ has reduced to -ŋ in Snl and Ban, -n in Rth and -n(na) in Tal. In Snl, Ban and Rth stress shift to the final syllable (where it does not already occur on the final syllable of the stem) attests to the previous occurrence of a vowel sequence, as still reflected in San:

Snl lebá to lick, lebán licked, cf. San lebaəŋ

Tal tumpa to descend, tatumpanna place of descent

Ban baká to split, bakán will be split  
tudo to leak, tudón leaking

Rth buká to split, bukán will be split  
turo to leak, turón leaking

Schwa elsewhere remains in Snl except that in the Sarangani dialect it was lost word-initially. In Ban and Rth also \*ə was usually lost word-initially. In all three languages its reflex remains after a prefix, having undergone changes in Ban and Rth as described below. In the following examples reflexes of \*ə initially and after a prefix are illustrated for Snl, Ban and Rth.

PSan	Snl (Sarangani)	Snl (Mindanao)	Ban	Rth
*əlup	lu?	əlu?	lu?	lup <i>to swallow</i>
*məŋəlup	məŋelu?	məŋelu?	məŋulú?	məŋulúp <i>is swallowing</i>

Since the loss of \*ə word-initially did not occur in the Mindanao dialect of Snl, its loss in Sarangani dialect must have been very recent and is therefore a parallel development rather than a shared innovation with Ban and Rth. Loss of initial \*ə did not occur in Snl, Ban and Rth if it was followed by a nasal-stop cluster, although it underwent other changes in Ban and Rth. Thus Snl əndauŋ *here you are*, Ban ampaha? *underlayer* (PSan \*əmpaR), Rth ungón *to call* (cf. San əŋgo *to make a noise*).

Where not lost altogether in Ban and Rth \*ə was replaced by some other vowel. In both languages it assimilated to a high vowel in the following syllable. Assimilation is complete in Ban, \*ə always assimilating to the following vowel. However, in Rth \*ə was replaced by u if the vowel following was other than i:

PSan	Ban	Rth
*bəkis	bikisi?	bikís <i>to bind</i>
*səlet	selé?	sulé? <i>to insert</i>
*bəka	baká	buká <i>to split</i>
*ləno	lonó	lunó <i>smooth</i>
*bəndu	bundú	bundú <i>difficult</i>

Where non-initial \*ə occurred adjacent to \*h in Rth it underwent different changes from those described above. Where it preceded \*h (reflecting PSan \*R) it assimilated to a following a. Subsequent loss of \*h (see section 2.3.9.) resulted in long vowels:

PSan	Pre-Rth	Rth
*səRam	*saham	saam <i>ant</i>
*bəRat	*baha?	baa? <i>heavy</i>

This change must have occurred before \*ə was replaced by u elsewhere. If medial \*ə has become u before \*h irrespective of the following vowel then the last item above, for instance, would have become \*\*buhə?, and finally \*\*buə?. The fact that \*ə assimilated to the following vowel before \*h was lost is shown by Niemann, medial h still occurring at the time he compiled his wordlist. Thus for the items above Niemann gives <saham> *ant* and <waha> *heavy*. It is possible that \*ə assimilated to other vowels in this position but no examples have been recorded except with u; in such words replacement of \*ə by u can also be accounted for according to the regular rule outlined above, e.g. PSan \*dəRup > Rth ruup (Niemann <ruhup>) *face*.

If \*ə was preceded by the sequence \*ah in Rth it assimilated to the preceding \*a. It is possible that \*ə assimilated to any vowel in this position but only a has been recorded. This assimilation occurred before the loss of \*h, as shown by Niemann in, for instance, his recording <tahandum> for the first example below:

PSan	Ban	Rth	
*taRəndum	tahundún	taandúm	<i>to remember</i>
*kaRəbi	kahibí	kaabí	<i>yesterday</i>

Assimilation to preceding \*a through intervening \*h took precedence over influence of the following vowel; in the non-prefixed form of the first example above, \*Rəndum became Rth Úndúm *to remember* (see \*əndum in the wordlist). In Ban \*ə was not influenced by the preceding vowel as it was in Rth, as shown by the above examples.

There are some similarities in the way Ban and Rth reflect \*ə. In initial position \*ə was usually lost and elsewhere was replaced by another vowel. However, there are strong reasons for not regarding changes to \*ə as reflecting shared innovations. Generally \*ə became u in Rth and assimilated to the following vowel in Ban. These changes must have occurred independently with the change in each language being directly from \*ə to the present-day sounds. It is not possible that both languages underwent a common change and that after splitting one language then underwent further changes, as this would have involved other vowels as well. For instance, if \*ə first became u before all vowels except i and later assimilated in Ban there is no explanation for why other instances of u, not originally from \*ə, did not likewise assimilate.

Further, in Rth \*ə assimilated to a preceding or following a if the intervening consonant was \*h. This change in Rth must have occurred before other changes to \*ə, as its replacement by u would have prevented such assimilation. Since Ban does not reflect assimilation of \*ə to a preceding vowel through intervening h all changes to \*ə medially must have occurred after the separation of Ban and Rth. Thus changes to \*ə offer no evidence for a close link between Rth and Ban (although loss of initial \*ə might have occurred earlier, as a shared innovation).

In Tal \*ə in all environments, except where lost following a vowel (see above), was replaced by a:

PSan	San	Tal	
*əpat	əpa?	appata	<i>four</i>
*bəli	bəli	balli	<i>to buy</i>
*dəndum	dənduŋ	danduma	<i>dark</i>

The change undergone by \*ə in Tal was independent of changes in other languages. Stress shift in Ban and Rth predated changes to \*ə in those languages (see section 2.3.23.) yet stress shift did not affect Tal. On the other hand the development of long consonants predated changes to \*ə in Tal (see section 2.3.15.) but is not reflected in Ban and Rth.

### 2.3.2. Assimilation and metathesis of \*s and \*t

San, Snl and Tal reflect metathesis of word-final \*s with a preceding \*t. In San and Snl the resulting final \*t was later replaced by ? (see section 2.3.3.). In the following illustrations Ban cognates are included for comparison:

PSan	San	Snl	Tal	Ban	
*Ratus	hasu?	rasu?	žasutta	háusu?	<i>hundred</i>
*bitis	bisi?	bisi?	bisitta	bítisi?	<i>calf of leg</i>

Metathesis was regular where \*t was the consonant in the syllable immediately preceding final \*s, as in the above examples. The change sometimes also occurred in San, Snl and Tal where \*t was separated from final \*s by two syllables. Most examples recorded have been of two syllable words, with metathesis of initial \*t and final \*s.

PSan	San	SnL	Tal	Ban	
*tanis	sanj?	sanj?	sanitta	tánisi?	<i>to cry</i>
*tages	sage?	sahe?	sahatta	tágesé?	<i>reef</i>

Metathesis did not always occur in San, Snl and Tal in the environment described above. For example, San, Snl *tapisə?* *to sieve*, San *togase?*, Snl *tohasə?*, Tal *tohassa* *strong, hard*. Such forms cannot all be discounted as having been borrowed after metathesis had ceased to operate. In the file of cognates there are more items which did not undergo metathesis of initial \*t and final \*s than items which did, which raises considerable doubt about the possibility of all such items being borrowings. Further, many of them have cognates in Rth and can thus be reconstructed for PSan. The San word *togase?* is a borrowing but as cognates occur in the other Sangiric languages it is a very ancient one, clearly predating PSan, for which an etymon, \**togas*, can be reconstructed. Steller and Aebersold cite San *tapisə?* as a loan from Mal but cognates occur in Rth and Ban and borrowing need not be assumed.

Metathesis of final \*s and preceding \*t also occurred where \*t was pre-nasalised:

PSan	San	Ban	Rth	
*pentas	pensa?	pantasa?	puntás	<i>to harvest</i>

The above example is the only one recorded but it is sufficient to allow items such as San *balontase?* *tree sp.* (cf. Ttb *walontas*) to be regarded as borrowings.

Although there is only one example available it indicates that metathesis did not occur in reflexes of PPh RMS: San *tətase?* *to break (of thread); to cut loose (of seam)* reflecting PPh \**tastas* *to cut, tear; cut loose*. Cognates do not occur in Ban and Rth. PSan \**tətas* is reconstructed, presumably metathesis being prevented by influence of the initial t.<sup>35</sup>

Otherwise metathesis of final \*s and \*t in the immediately preceding syllable was regular and the few recorded items in which it is not reflected are treated as borrowings.<sup>36</sup>

In San and Snl initial \*t assimilated to a later, non-final \*s, whether in the following syllable or more distantly separated. No cognates have been recorded for Tal so it is not known if that language also reflects this assimilation:

PSan	San	SnL	Ban	
*tasik	sasi?	sasi?	tasi?	<i>sea</i>
*tasak	sasa?	sasa?	tasa?	<i>ripe</i>
*talisyay	sa ise	sa isay	talisey	<i>tree sp.</i>

Since assimilation, and not metathesis, occurred in San/Snl if \*s was originally non-final, Tal evidence is necessary to show that it was metathesis, and not assimilation, which occurred when \*s was originally in final position, the evidence in San and Snl having been obscured by later reduction of the final consonant to ?.<sup>37</sup>

Metathesis is one of the strongest pieces of phonological evidence for subgrouping. It shows that San, Snl and Tal shared a parent language not ancestral to Ban and Rth and that consequently they form a subgroup of the Sangiric languages with a common parent language, Proto-North-Sangiric (PNSan). This subgrouping is supported by lexicostatistical evidence (see section 2.4.).

Metathesis occurred before final consonant reduction as Tal split from San/Snl before that change occurred. Metathesis provides the important information that final consonant reduction in Ban, San and Snl did not occur in a shared parent language but was the result of diffusion of the innovation. It also shows that development of the paragoge in these languages was also an areal phenomenon (see section 2.3.3. for a description of final consonant reduction and paragoge development).

Assimilation of \*t to a following \*s may have occurred at the same time as metathesis but until evidence from Tal is available it cannot be shown that it was not a later development in San/Snl only.

### 2.3.3. Paragoge and final consonant reduction

In all languages but Rth a process occurred whereby an extra syllable was added word-finally following certain consonants. The nature of this paragoge varies from language to language (see section 2.1.6. for a synchronic statement).

In Tal vowel a occurs after all final consonants, usually with doubling of the preceding consonant (see section 2.1.3.1.), e.g.:

PSan	Tal
*uRas	užasa
*inum	inumma
*lanit	lanitta
*utak	uta?a

to wash  
to drink  
sky  
hair

The paragoge does not occur in words which earlier had final \*?, either because \*? resisted the addition of the paragoge or because it had already disappeared (see section 2.3.7.), leaving a final vowel, which blocked paragoge development. Where the paragoge now occurs after glottal stop, as in the last example above, ? derives from earlier \*k (see section 2.3.8.).

In San and Snl the paragoge is ø?, i.e. schwa followed by glottal stop:

PSan	San	Snl
*uRas	uhase?	urasø?
*likud	likude?	likude?

to wash  
back

In Ban the paragoge is V?, where V assimilates to the preceding vowel:

PSan	Ban
*uRas	úhasa?
*didir	dídihi?
*səkol	sokolo?

to wash  
yellow  
to cough

In Tal the paragoge occurs after all words ending in a consonant. In San, Snl and Ban it does not occur after final nasals (see section 2.3.5. for examples). Elsewhere the paragoge is word-selective in San, Snl and Ban. In words in which it was not added the final consonant was replaced by ?, unless it was a nasal or itself ?. Replacement of final consonants by ? is hereafter called final consonant reduction.<sup>38</sup>

PSan	San, Snl, Ban	Tal	
*atup	atu?	atuppa	roof
*takut	taku?	ta?utta	afraid
*manuk	manu?	manu?a	bird

Although all three languages usually underwent the same changes in the same items this was not always the case. Ban has a number of recorded words with the paragoge while corresponding words in San and Snl underwent final consonant reduction. Thus Ban kínasa?, San/Snl kina? *fish*; Ban ápuhu?, San/Snl apu? *lime*; Ban bénkolo?, San benko? *bent*. Further, Steller and Aebersold's dictionary gives numerous examples of doublets in San where one has the paragoge and the other does not. For example, salu? *river* and saluhə? *drain, channel*<sup>39</sup>; disi? and disihe? *to stand firm*; bəhi? and behisə? *line, stripe*; lintu? *to descend* and lintuhə? *degree of descent*; pongo? and pongole?. There is no reason to suppose either of the forms in each pair is a borrowing. Such doublets may result from dialect mixture. Steller and Aebersold also provide other examples where there is dialect variation. Thus Taruna lebə?, Manganitu lebhə? *young coconut*.

A very few examples have been noted where San and Snl disagree in the occurrence or not of the paragoge: San ma-ririhe?, Snl ma-didi? *yellow* (the Snl form may be irregular — see the note to \*didiR in the wordlist, section 3.2.); San hunu? *fire*, Snl lu-runuse? *bonfire*; San bisu!ə?, Snl bisu? *boil*.

It has been suggested that San forms with the paragoge may be borrowings, with only final consonant reduction occurring in inherited words (Mills 1974: 18-19). However, there are far too many basic vocabulary items with the paragoge for them all to be regarded as borrowings; most are clearly inherited words.

In the wordlist in section 3.2. there are thirty San items with the paragoge after s but only three, possibly four, reflecting s-reduction (see \*Renes, \*kəmis, \*kinas, \*nipis) and one case of doublets (see \*bəRis).<sup>40</sup> No recorded Ban words have s-reduction, the paragoge occurring in all items. Clearly paragoge addition was the usual development where PSan had final \*s.

One class of sounds consistently replaced by ? in San, Snl and Ban were the voiceless stops \*p, \*t and \*k. No items of basic vocabulary have paragoge and the few exceptions noted are mostly obvious borrowings. Compare, for instance, San səda? *to set (of sun)*, from PSan \*sədap, and sədape? *delicious*, from Mal sədap. San (Tabukang) otake? and Snl utuke? *brain* are borrowings, occurring beside the inherited word uta? *hair*, from PSan \*utak (see comments in section 1.6.).<sup>41</sup> Ban kúlata? *mushroom* must also be treated as a borrowing (see section 1.6.).

It is generally true that borrowings added paragoge rather than reducing final consonants in order to conform to the phonological patterns of these languages, e.g. San kapale?, Snl kape? *ship*, from Mal kapal; San malase? *lazy*, from Mal malas; San harape?, Snl halape? *hope*, from Mal harap; San poroke? *fork*, from Dutch vork, and many others. Note, however, San, Snl bebe? *duck*, with unexpected consonant reduction, beside Ban bēbeke?, presumably from Mal bebek; San/Snl uba? *monkey*, apparently from ubal in a southern Philippine language (see section 1.6.). In both the above cases medial b, instead of regular b, points to borrowing.

Why some words in San, Snl and Ban added paragoge while others reduced the final consonant to glottal stop cannot as yet be conclusively explained. The hypothesis is put forward below that paragoge addition and final consonant reduction were competing processes by which final consonants were eliminated.

As explained in section 2.3.2., metathesis of \*t and \*s occurred in PNSan after its separation from Ban and Rth. At the time metathesis took place neither paragoge addition nor consonant reduction had yet occurred; the present-day San and Snl forms can only be explained as the product of metathesis operating on words ending in a consonant.

Since Ban had separated from the other languages before metathesis took place it was already a separate language when paragoge addition and final consonant reduction occurred. Therefore the appearance of these in the modern languages reflects, at least in part, their diffusion through separate, though contiguous, speech communities.

Further evidence that Ban had already split from the other languages when final consonant reduction and paragoge development occurred comes from the fact that it undergoes certain different morphophonemic changes. Thus:

(i) In Ban replacive glottal stop remains before a suffix whereas in San and Snl it is replaced by a consonant, usually k or t.<sup>42</sup> In the following examples of reflexes of PSan items both roots and suffixed forms are given:

\*boRet > Ban bohe? to write + -aq → bohe?aq be written

\*tiap > Ban tia? to count + -en → tia?en be counted

\*boRet > San bohe? to write + -aq → bohekan be written

\*tiup > San tiu? to blow + -aq → tiukan be blown

\*seŋgot > Snl seŋgo? to sail + -aq → seŋgotan be sailed

\*tiup > Snl tiu? to blow + -aq → tiukan be blown

(ii) In Ban the paragoge remains before a genitive phrase whereas it is lost in San and Tal:<sup>43</sup>

Ban	pálede?	nu lima	palm of the hand
San	paled	u lima	palm of the hand
Tal	papaidd	u manu?a	bird's feathers

Final consonant reduction did not occur in Tal. It therefore split from San/Snl before this change began.

It is possible that the paragoge originated in PNSan. Nevertheless, the fact that it occurs after all consonants in Tal but not in San/Snl shows that the languages split before it spread to the entire lexicon. It is also possible that the paragoge first appeared in Tal after it split with San/Snl and that the innovation was then borrowed by San/Snl, as it was by Ban.

The hypothesis put forward here is that addition of the paragoge and final consonant reduction were two competing methods by which final consonants were dealt with in the Sangiric languages. According to this hypothesis the development of a paragoge probably began in Tal after it had split with San/Snl. The reduction of final consonants to ? began in the San/Snl area, affecting some consonants before others. Thus it affected final voiceless stops, spreading to Ban, which also replaced all final voiceless stops by ?. These were undoubtedly the first sounds affected as they all reduced to ?, whereas such consistent replacement did not occur with other consonant classes.

However, while consonant reduction was spreading to different classes of final consonants throughout the Ban and San/Snl areas, and at different rates in different areas, paragoge addition spread from Tal and began to affect these languages.

The paragoge became more favoured as a means of dealing with final consonants and began to operate on those words in which final consonants had not yet been replaced by ?. (Since it did not operate on words ending in glottal stop these remained as they were.) Thus, for instance, when the rule of paragoge addition spread it 'found' Pre-San \*kina? *fish* and thus did not operate on it but 'found' Pre-Ban \*kinasa?, which thus took the paragoge, reflected in modern Ban kínasa?, pre-empting the application of the consonant reduction rule in that, and other s-final words, in Pre-Ban. Similar variation in the San dialects, described above, can likewise be accounted for.

The above hypothesis appears the most satisfactory for explaining the present-day situation in Ban, San and Snl. The addition of the paragoge rather than consonant reduction to borrowed words to bring them into line with the phonological patterns of the languages is accounted for in the above hypothesis because paragoge application continued to operate after consonant reduction had ceased, it being the favoured method for 'removing' consonants from final position.

The hypothesis does not rule out the possibility that the paragoge was originally \*ə? in all languages, as regular sound changes operating in Tal and Ban could explain the present-day forms in terms of developments from \*ə?. In Tal \*ə became a in all environments and final \*? was lost.<sup>44</sup> In Ban \*ə regularly assimilated to the following vowel; if the paragoge was originally \*ə? then in this case the direction of assimilation was reversed, \*ə here assimilating to the preceding vowel.

While the paragoge was added to words ending in a nasal in Tal, final nasals were affected by neither paragoge addition nor reduction to glottal stop in Ban, San and Snl.

### 2.3.4. Reduction of final \*t in Rth

Consonant t does not occur finally in Rth, having been replaced by ?. Tal cognates are included for comparison:

PSan	Rth	Tal
*lanit	laŋe?	laŋitta
*epat	pa?	appata
*takut	taku?	ta?utta

sky  
four  
fear

Since final \*t reduced to ? in all languages except Tal it is only Tal evidence which allows its reconstruction for PSan. The reduction in Rth is dealt with separately because of the possibility that it is an entirely independent development from consonant reduction in San, Snl and Ban.

Since Rth does not reflect metathesis of final \*s and a preceding \*t it had split with the North Sangiric languages before occurrence of metathesis, which preceded final consonant reduction (see section 2.3.2.). This applies to Ban too, which also lost final \*t. It is thus possible that reduction of final \*t occurred once, in a parent language ancestral to Ban and Rth. If this is so then \*t was the first consonant to reduce to ?. Following t-reduction Rth then split from Ban and its speakers moved away from the area in which final consonant reduction was gradually spreading (see section 2.3.3.). Once Rth was isolated from the other Sangiric languages, as it is today, the process of final consonant reduction ceased to operate, leaving all other final consonants intact.

This possibility cannot be proved and it may be that the change in Rth was an independent parallel development. However, there is one piece of evidence that it was a shared innovation in a language ancestral to Ban and Rth. This is the fact that in both Ban and Rth ? reflecting \*t remains before a suffix, i.e. \*t is not recovered. In the North Sangiric languages, represented by San in the following example, final ? is replaced by another consonant (see note 42 to section 2.3.3.) before the passive suffix:

PSan	Ban	Rth	San
*ləbat	labá?, laba?en	lubá?, luba?en	ləba?, ləbaken
			to cross (river) be crossed

This is weak evidence as there are only two possibilities: retention of ? or its replacement by some other consonant, either the original one or some other, as in San. Nevertheless, the fact that Rth and Ban do behave similarly in this respect, as against the North Sangiric languages, lends weight to the possibility of a shared innovation after they split from the North Sangiric languages and before they split from each other.

### 2.3.5. Merger of final nasals

Nasals m and n do not occur word-finally in San, Snl and Ban, having been replaced by ñ. Original nasals remain in Rth and Tal and attest to their occurrence in PSan:

PSan	San, Snl, Ban	Tal	Rth
*inum	inuŋ	inumma	inum
*suan	suaŋ	suanna	suán

On the surface the innovation might appear to be good evidence for grouping Ban with San/Snl. However, since the change occurred after Tal split from San/Snl it could not have occurred in a single ancestral language because Ban and San/Snl did not exclusively share a parent language (see section 2.3.3.). Further, the change did not affect Rth and therefore occurred after Rth split from Ban, which post-dated Ban's split from San/Snl (see section 2.4.).

The change therefore must have originated somewhere in the San/Snl-Ban area and spread by diffusion. The fact that it did not affect Tal was probably because that language had already adopted the paragogic after final nasals.

### 2.3.6. Final diphthongs

The final diphthongs of PSan have been retained by Rth and Ban, although with changes, but have been replaced by single vowels in San and Tal and partially replaced by single vowels in Snl.

In San and Tal \*aw was replaced by o and \*ay by e. In Snl these changes occurred only if the vowel in the preceding syllable was a; elsewhere ay and aw remained. In Ban \*ay and \*aw are reflected as ey and ow respectively. In Rth they occur as ey and ow in most words but as ay and aw in some others, usually in free variation with ey and ow:

PSan	San	Sn1	Tal	Ban	Rth	
*balay	bale	bale	bale	baley	baley	house
*takaw	tako	tako	ta?o	-	takow	to steal
*babinay	babine	babinay	babine	babiney	babiney <b>v</b> babinay	woman
*əmay	əme	may	amme	mey	may	rice
*siaw	sio	siaw	sio	siow	siaw	nine
*tolay	tole	toay	tole	toley	toley	tail

The synchronic variation in Rth suggests the change is of recent occurrence in that language, having affected some words before others. Where ay and aw still occur, a is raised and fronted before y, e.g. may [m<sup>ā</sup>i] rice, and raised and backed before w, e.g. saw [sa<sup>ā</sup>u] (in free variation with sow [so<sup>ā</sup>u]) wet. This suggests that the original replacement of phoneme \*a by e and o in Rth and Ban developed by gradual shift in the phonetic nature of the segment in this environment, a change still in progress in Rth, having so far not affected all words. Where free variation occurs in Rth it is possibly a result of dialect mixture.

On the basis of reflexes of PSan diphthongs San and Tal appear to group together as against Sn1. However, this is ruled out by evidence presented in sections 2.3.2. and 2.3.3. The change may have been partly completed in PNSan in environments where it is reflected in modern Sn1. If so then the languages split before the innovation spread to environments other than after a in the preceding syllable. The reduction was then extended to all environments in San and Tal while no further change occurred in Sn1. Since San and Sn1 form a genetic subgroup the further changes in San and Tal are either the result of diffusion or parallel development.

Words such as PSan \*sai *who* and \*dau *far* are interpreted as containing a sequence of two vowels because final \*ai and \*au in such words are reflected as ai and au in San and Tal and not as e and o, the usual reflexes of diphthongs.

A different interpretation would be to recognise monosyllabic words such as \*say and \*daw in PSan with the explanation that normal changes to \*ay and \*aw did not occur in such forms. However, a distinction must be made between final \*ai and \*ay, \*au and \*aw as shown by the following examples:

PSan \*uai *mango*, with reflexes San, Tal, Ban uai [uái] but  
\*uay *rattan*, with reflexes San, Tal ue, Ban uey [úey]

PSan \*bayau *tree sp.*, with reflexes San balau, Ban bayau [bayáu]  
\*balajaw *anchor*, with reflexes San balajo, Ban bulanjow [bu]ánjow]

Since the distinction between final \*ai and \*ay, \*au and \*aw must be made then the etyma of modern sai, dau, etc. are best reconstructed with \*ai and \*au, rather than \*ay and \*aw.

Maryott (1977a) states that in Sn1 there is free variation between final ai and ay, au and aw. His convention of writing y and w finally instead of i and u is followed here for Sn1.

### 2.3.7. Loss of glottal stop in Tal

Glottal stop occurred only in morpheme-final position in PSan (see section 2.2.3.). It was subsequently lost in Tal. Because PPh final \*q or \*? was frequently lost in PSan (see section 1.4.1.(b)) the number of PSan items reconstructable with final \*? is quite limited and most of the items which can be reconstructed lack known Tal reflexes. Some items lacking known Ban and Rth

cognates are not reconstructable for PSan but can be reconstructed for PNSan; these also attest to loss of final \*? in Tal:

PSan	PNSan	San	Tal	
*kento?		kento?	ento	<i>lame</i>
*Ramu?		hamu?	zamu	<i>red</i>
	*pene?	pene?	pene	<i>full</i> <sup>45</sup>
	*singka?	singka?	singka	<i>to know</i>

In some words final syllable \*e? was lost in Tal, with subsequent addition of a paragogic vowel:

PSan	PNSan	San	Tal	
*leRe?		lehe?	ulekka	<i>neck</i>
	*kanede?	kanere?	aŋera	<i>when</i>
	*Rose?	hose?	<rɔssa>	<i>to bind</i> <sup>46</sup>

### 2.3.8. Reflexes of \*k in Talaud

PSan \*k was lost word-initially in Tal and became ? after a vowel:

PSan	Tal	
*kinas	inassa	<i>fish</i>
*kiki	i?i	<i>to bite</i>
*likud	li?udda	<i>back</i>
*manuk	manu?a	<i>bird</i>

Following a nasal k remains:

PSan	Tal	
*kəŋkum	səŋ-kankuma	<i>fistful</i>
*benkol	benkola	<i>bent</i>

Also initially k remains following the genitive marker N-: bəle nkami *our house*, cf. (i)ami *we* (nominative).

The change PSan \*k > Tal ? occurred after the change PSan \*? > Tal Ø (otherwise the reflex of \*k would be Ø) and it reintroduced ? to the language.

### 2.3.9. Reflexes of \*R

The phonetic nature of \*R in PSan is not known and the symbol <R> is chosen to represent it because it is a continuation of the phoneme represented <R> in PAN and PPh. The possible phonetic nature of this sound is further discussed at the end of this section.

\*R became h in the Manganitu and Tabukang dialects of San and also in Ban and Rth. It still occurred as h initially and medially in Rth at the time Niemann compiled his wordlist as he represents it <h> in these positions. Initially h later underwent reduction although it is still interpreted as a consonant, represented here by a stroke over the initial vowel (see section 2.1.5.1.). Medially it has since been lost. Niemann does not represent h finally and it was probably already lost in that position when he compiled his list. Below are

examples of Rth reflexes of PSan items containing \*R, as represented by Niemann and in the present-day language:

PSan	Rth (Niemann)	Rth (modern)	
*Rabun	<hawun>	ābun	<i>cloud</i>
*Rusuk	<husuk>	ūsuk	<i>thin</i>
*bəRat	<waha>	baa?	<i>heavy</i>
*duRi	<ruhi>	rui	<i>bone</i>
*bibir	<wiwi>	bibi	<i>lips</i>
*namuR	<namu>	namu	<i>dew</i>

\*R became r in Snl and the Taruna dialect of San. Maryott (1961) describes Taruna r as a retroflexed mid central non-syllabic vocoid. He describes Snl r as an alveolar flap, usually occurring with simultaneous velar friction: [ṛg] (1977). Below are illustrated reflexes in San (Manganitu) and in Snl:

PSan	San	Snl	
*Ramut	hamu?	ramu?	<i>root</i>
*duRi	duhi	duri	<i>bone</i>
*bibir	bibiḥə?	bibire?	<i>lip</i>

In Tal (Salibabu) \*R became k after \*ə, which later became a (see section 2.3.1.). Final \*R also became k, although whether this occurred before or after development of the paragoge cannot be established. Doubling of k follows the rules described in section 2.3.15.:.

PSan	Tal	
*əRe	akke	<i>grass sp.</i>
*bəRu	bakku	<i>new</i>
*bibir	bibikka	<i>lip</i>
*bəṇaR	banjaka	<i>molar</i>

In the dialect of Haines (Karakelang) \*R was replaced by t before the paragoge, rules for gemination being the same as in Salibabu. Thus for the last two items above Haines has bibitta *lip* and banjata *molar*. According to Adriani (1911:4) final \*R was replaced by different consonants in other Tal dialects.

The change of \*R to k occurred before \*ə was replaced by a, otherwise \*R would have the same reflex in this environment as it has after other vowels, as described below. The change of \*R to k occurred after the change of \*k to ? (see section 2.3.8.) as k remains unchanged where it reflects earlier \*R.

In other environments PSan \*R became Tal ž:

PSan	Tal	
*Ramut	žamutta	<i>root</i>
*daRa	daža	<i>blood</i>

Reflexes of \*R offer little evidence for subgrouping, occurring on a geographic rather than a genetic basis; a southern group: Rth, Ban and San (Manganitu and Tabukang), along with the Minahasan languages, sharing h and a northern group: San (Taruna), Snl and Tal, having r and phonetically similar ž.

In section 1.4.1.(e) it is suggested that \*R may have been an alveolar consonant in PSan. However, the widespread occurrence of a reflex h argues against this as the change from a, presumably voiced, alveolar sound to a voiceless glottal fricative seems unlikely. On the other hand the occurrence of apical reflexes argues against this sound having been [h]. A uvular trill

could have given rise to both types of reflex; the only argument against this being that PSan \*g probably had a velar fricative allophone [š] (see section 2.3.13.) and the system would have been unlikely to have sustained such phonetically similar segments (hence the changes to \*R?). Interestingly the Snl reflex r [ṛ] shares both apical and dorsal features, which raises the possibility that such simultaneous articulation may have been a characteristic of PSan \*R.

### 2.3.10. Reflexes of \*l

PSan \*l had two allophones: alveolar lateral [l] and retroflexed lateral flap [ɻ].

These merged in alveolar lateral [l] in all positions in Rth and in retroflexed lateral flap [ɻ] in all positions in Ban. In a few items \*l is reflected by h or Ø in Ban and Rth (see section 2.3.20.).

In San, Snl and Tal [l] and [ɻ] are in complementary distribution where they derive from PSan \*l. However, later sound changes have brought them into contrast, resulting in two separate phonemes. The distributions of the two sounds are quite similar in San and Tal and these languages can be described together. A number of differences occur in the distribution of the laterals in Snl and this language is described separately.

In San and Tal non-retroflexed l occurs word-initially and following front vowels. It occurs after e in San and after a in Tal where this reflects earlier \*ə (see section 2.3.1.). In both languages l is doubled in this last environment, the doubling being phonemic in Tal (see section 2.1.3.) but not in San (see section 2.1.1.):

PSan	San	Tal	
*lanit	lanj?	lanitta	sky
*bilat	bila?	bilatta	to spread out
*tatəlu	tatəlu	tatallu	three

Retroflexed l occurs after back vowels. It also occurs after a in San and in Tal it occurs after a where it is not from an earlier \*ə:

PSan	San	Tal	
*balay	bale	bale	house
*tulid	tulide?	tulidda	straight
*boleg	boleg	bolella	to pull

Tal differs from San in that l occurs (i) after prefix-final u: ulekka neck, lullage to laugh, and (ii) before the paragoge, irrespective of the preceding vowel:

PSan	San	Tal	
*kopal	kopalə?	appala	thick
*timbul	timbulə?	timbula	to rise

The two sounds came into contrast in San and Tal when \*y was replaced by l word-medially (see section 2.3.11.). Thus San bala to allow (PSan \*baya) and bala valley; Tal bala to allow and a]appa to take, alu wood (PSan \*kayu) and sajukka river.

In borrowings l is not necessarily replaced by l in environments where PSan \*l is reflected by l, probably because the change \*y > l has made l a possible

choice in such environments, e.g. San malasə? *lazy* from Mal malas. But following a, u or o and preceding the paragoge, where l does not occur in San, l replaces l in borrowings, e.g. kapalə? *ship* from Mal kapal.

In Snl l occurs word-initially, after front vowels and after morpheme-initial e. Where word-initial \*ə has been lost in the Sarangani dialect l occurs initially, as in the last example below:

PSan	SnL	
*lanit	lanj?	sky
*kilat	kila?	lightning
*pakel	pakelə?	heel
*əlup	(man)əlu?	to swallow
*əlaw	law	day

Retroflexed l occurs following a back vowel, a or non-morpheme-initial e and preceding a front vowel:

PSan	SnL	
*bəli	bəlli	to buy
*paled	paledə?	palm
*tuli	tulli	ear

PSan \*l was lost in Snl after a back vowel, a or non-morpheme-initial e and before a back or low vowel:

PSan	SnL	
*telak	tea?	to fly
*dəluk	dəu?	thunder
*balu	bau	widow
*tolay	toay	tail
*pulo	puo	ten

Loss of \*l occurred after changes to \*aw and \*ay described in section 2.3.6. Thus PSan \*balay *house* became Snl bəle; if \*ay still occurred when \*l was lost before \*a then the Snl form would have become \*\*bay, cf. l-loss where ay remained, as in Snl toay *tail*, from PSan \*tolay.

When \*l was lost between identical vowels, vowel reduction resulted:

PSan	SnL	
*bulud	bude?	mountain
*sala	sa	mistake
*dalait	dai?	bad

Sometimes \*l was replaced by y, apparently unpredictably, in variation with the reduced form (see Maryott 1978a:122 for further discussion):

PSan	SnL	
*solo	so ~ soyō	lamp
*alap	a? ~ aya?	to fetch

PSan word-final \*l was lost in Snl after a back or low vowel. In such cases the paragoge occurs in Snl, showing that paragoge addition occurred before l-loss. If \*l had been lost first such words would have ended in a vowel and e? would not have been added. When \*l was lost in this position the preceding vowel was also lost if unstressed but remained if stressed:

PSan	Pre-Snl	Snl	
*səkol	*səkóle?	səkoə?	cough
*kəpal	*kəpále?	kəpaə?	thick
*pundal	*púndale?	pundə?	paddle
	*kápale?	kapa?	ship

The last item above is from Mal but was borrowed before the separation of San and Snl, cf. San *kapa?*, with later loss of \*l and the preceding vowel in Snl.

Word-initial \*l assimilated in Snl to l or r, whether in the following syllable or further separated:

PSan	San	Snl	
*lidik	liri?	lili?	garden
*ludan	luran	lujan	load
*leRe?	lehe?	rere?	neck
*linuR	linuhə?	rinurə?	earthquake

Contrast between l and l was brought about in Snl by later changes. For instance, \*d became l intervocally (see section 2.3.13.), thus producing contrast with l, e.g. sili (PSan \*sidi) to separate grain and sili to defer to. Word-initially l occurs as a result of several morphophonemic processes (discussed by Maryott, 1978a:116), e.g. latu high chieftain (from earlier \*da]atu + Ca-(intensive) + datu chieftain), which contrasts with latu ant.

Some unexplained forms occur in the limited material available on Snl.<sup>47</sup> In alula? raft l-loss did not occur. This might be due to influence of the preceding l but the word is probably a borrowing, cf. Samal alul raft. Word-initially l occurs instead of d in law distance and low thirst. These usually occur as adjectives malaw far and malow thirsty (where \*d > l intervocally) and initial l may result from analogy with the adjectival forms.

Word-initial l is usually lost in Snl if the addition of a prefix produces the environment in which it was lost historically, e.g. ma- + lukade? to watch, keep guard → maukade? morning (cf. San malukade? morning). The vowel of the prefix is lost if unstressed and followed, after l-loss, by e, as also occurred historically preceding the paragogic (see above), e.g. ma- + ləmbay → məmbay remaining, ka- + ləndi? → kəndi? to catch attention.

### 2.3.11. Reflexes of \*y

In San and Tal y does not occur. Final diphthongs reduced to mid vowels in these languages (see section 2.3.6.). Word-medially \*y was replaced by l:

PSan	San	Tal	
*baya	bala	bala	to allow
*kayu	kalu	alu	wood

This change has resulted in contrast between l and l (see section 2.3.10.). The change of \*y to l is not reflected in Snl and thus its occurrence in San and Tal must be the result of diffusion after the separation of the languages.

In Snl, Ban and Rth y occurs finally in diphthongs (see section 2.3.6.). It remains medially in the three languages except that in Snl it has been lost irregularly between identical vowels. In a few such words it remains while in others it remains in variation with its loss or is lost altogether. All three

possibilities are exemplified below. Where \*y was lost subsequent vowel contraction occurred.

PSan	Snl	
*payan̩	payan̩	<i>thigh</i>
*nayan̩	nayan̩ ~ nan̩	<i>to play</i>
*daya	daya ~ da	<i>inland</i>
*maRuaya	maruá	<i>girl</i>

Loss of \*y in Snl followed changes to diphthongs described in section 2.3.6. Thus PSan \*dayaw *to praise* > Snl dayo; with y then bracketed by non-identical vowels the environment in which it was lost no longer occurred.

Word-initially high front non-syllabic vocoid [y] does not occur in San, Tal or Ban. It occurs in Rth as the manifestation before a vowel of the personal noun and pronoun marker i. Thus Rth iá? [yá?] I, cf. San, Ban iá? [iá?], Tal ia?u [iá?u]. The only recorded instance of initial [y] in Rth where it is not the personal marker is in iúr [yúṛ] *to pull*, from PSan \*Riud.

Initial [y] occurs in Snl before vowels. Maryott (1978a:124) recognises it as a phoneme y, reflecting previous \*y (which he thus treats as having become i in San). In most of his examples it is the reflex of personal noun and pronoun marker \*i: yakəŋ *eldest sibling* (< i + akan̩), yupuŋ *grandparent* (< i + upuŋ). His other example is yuta?, San iuta? *innumerable*, a borrowing from Sanskrit (via Mal juta *million*). Word-initially the sound appears to be in complementary distribution with i in Snl, as it is in Rth, and it is here recognised as a non-syllabic allophone of i.<sup>48</sup>

Since PSan initial \*i before a vowel is reflected as [i] in San, Tal and Ban and as [y] in Snl and Rth it is possible that the two phones occurred in free variation in that position in PSan.

### 2.3.12. PSan \*e > Tal \*a

PSan \*e in final closed syllables assimilated to a preceding a in Tal:

PSan	Tal	
*aden	aranna	<i>name</i>
*paled	paladda	<i>palm</i>
*saleR	sa akka	<i>nest</i>

In most recorded examples, including those above, PPh etyma have \*a in the final syllable. Thus PPh \*-ajan *name*, \*palaj *palm*, \*sala[R] *nest*. It might be argued that Tal did not undergo the innovation whereby PPh \*a in this environment became PSan \*e, as reflected in the other four languages (see section 1.4.1.(e)), and that this is thus strong evidence that Tal does not descend from PSan.

The evidence that Tal is a daughter language of PSan and that the change \*e > a occurred subsequent to its separation from the other languages is as follows:

It is not only in words where PPh had \*a that Tal has a corresponding to e in the other PSan languages. Two examples have been recorded where this is a reflex of PPh \*ə: PPh \*saNdəR, PSan \*sandeR, Tal sandaka *to lean*;<sup>49</sup> PPh \*dakəl, PSan \*dakel *big, many*, Tal <da'ala> (from Steller and Aebersold, presumably da?alla).

It could be argued that these are merely instances of earlier \*ə becoming Tal a by the regular rule described in section 2.3.1. and that it offers no evidence for a close relationship between Tal and the other Sangiric languages. However, PPh \*ə in the word-final syllable did not always become PSan \*e; for instance, it became \*u if followed by \*m or \*p, e.g. PPh \*itm̥, PSan \*itum *black*; PPh \*atəp, PSan \*atup *roof*. It became \*i in some other words, e.g. PPh \*ikət, PSan \*ikit *tie* (see further discussion in section 1.4.1.(c)).

If Tal were not a daughter language of PSan but instead directly reflected all instances of PPh \*ə by a then the examples above would have the expected Tal reflexes \*\*itamma *black*, \*\*atappa *roof*, \*\*i?atta *to tie*. However, where PSan reflects PPh ultimate \*ə by a vowel other than \*ə then Tal identically reflects the PSan innovation. Thus its reflexes for the abovementioned items are itumma, atuppa, i?itta.

This clearly establishes that Tal is a descendant of PSan. Consequently (discounting later borrowing) Tal sandaka and <da'ala> must be continuations of PSan \*sandeR and \*dakel respectively and thus they establish that PSan \*e in final syllables assimilated to a preceding a in Tal, after its separation from the other Sangiric languages. This assimilation also affected \*e in the suffix \*-en, as this only occurred in PSan following a closed syllable containing \*a (see section 1.4.1.(f)). Thus PSan \*suanen (<- \*suan + \*-en>) > Tal suananna (<- suanna + -anna>) *will be planted*.

### 2.3.12. Reflexes of \*b, \*d and \*g

With the exception of Ban all languages have non-plosive reflexes of \*b, \*d and \*g in some environments and it is most likely that continuant allophones also occurred in PSan.

The distribution of stops and continuants is set out in the following chart, where S = stop and C = continuant. Five positions in the word need to be distinguished:

	initial	post-nasal	post-ə	intervocalic	final
San	S	S	S	C	S
Snl	S,C	S	S	C	S
Tal	S,C	S	S	C (S)	S
Ban	S	S	S	S	S
Rth	C	S	C	C (S)	C

The following points should be noted:

(i) The position 'post-ə' includes other vowels which are reflexes of PSan \*ə. Thus it includes 'after a where it reflects earlier \*ə' in Tal, etc.

(ii) The position 'intervocalic' does not include a preceding ə or any reflex of PSan \*ə.

(iii) The position 'final' ignores the presence of a paragoge.

(iv) Parentheses indicate marginal occurrence, which is explained below.

The reflexes of \*b, \*d and \*g are summarised as follows:

## (a) \*b

In all languages S is realised as [b]. C is a voiced bilabial fricative, in some languages in variation with a labiodental fricative [b ~ v], its exact nature depending on the phonological environment. Word-finally in Rth it is voiceless [ɸ]. Where [ɸ] occurs in the present-day language Niemann writes <w> or <u>, probably representing [b], e.g. <ahaw>, modern aab [a:b] fern sp.; <kanehau>, modern kaneab [kanéap] *yesterday*, so the devoicing in this position may be recent.

## (b) \*d

In all languages S is realised as [d]. C is usually realised as [l] in the Sarangani dialect of Snl and as [ɻ] in the Mindanao dialect and in all other languages.

In Sarangani the development of l as the reflex of intervocalic \*d post-dates the rule of l-loss in certain environments (see section 2.3.10.) as l from \*d was not lost in these positions. Thus PSan \*sala *wrong* > San sala, Snl sa but PSan \*mada *dry* > San mara, Snl ma|a.

Sarangani l as a reflex of \*d was almost certainly via \*r and r still occurs if there is also r (reflecting earlier \*R) in the preceding syllable:

PSan	Snl	
*Rodaw	roro	sharp
*Rado	raro	tame

## (c) \*g

S is realised as [g]. In the Manganitu dialect of San C is realised as [g], although other variants occur in other dialects. It is realised as [h] in Snl and in the Salibabu dialect of Tal, although in some Tal dialects it is [g]. In Rth voiceless velar fricative [χ] occurs word-finally and [h] elsewhere. Rth probably had [g] medially and initially until quite recently as [h] could only have developed following changes to phoneme h, within the last one hundred years (see section 2.3.9.). Niemann writes <g> or <gh> in these environments, probably representing [g], e.g. <lugai> *laugh* and <lughai> *laughter*, for present-day luháy.

The question of whether to recognise allophones or separate phonemes in the present-day languages arises here but this must be decided separately for each language and sometimes for each set of reflexes within a language. The question does not arise in Ban where only stops occur.

In San numerous examples occur of contrast between b and ɸ, d and r, g and χ intervocally. It is probable that borrowing is an important factor here. In many cases words with medial voiced stops are borrowings from Mal, e.g. diadi *to create* from Mal jadi, diabatan *sceptre* from Mal jabatan, diaga? *to keep guard* from Mal jaga. Each of the above examples is identified as a borrowing from Mal because of, among other things, initial di from Mal j, instead of d, which occurs in direct correspondences, e.g. San daleñ, Mal jalan *road*; San dahami *unused ricefield*, Mal jérämi *rice stalk*. Nevertheless, since they are assimilated words they establish contrast between stops and corresponding continuants and hence these must be treated as separate phonemes.

Word-initially in Snl stop reflexes occur for \*d and \*g but a continuant reflex for \*b. Thus dara *blood*, gelí? *to give*, bale *house*. Stop b occurs initially in some words. None of these appear to be reflexes of PSan words and many are clearly borrowings, e.g. bila *if* from Mal. The word baga? *lungs*

resembles PPh \*baRa? but contains two irregularities apart from initial b, i.e. medial g, which is not a correct reflex of \*R, and occurrence of final ?, where PPh final ? is usually, although not always, lost in PSan, cf. San lum-baha *lungs*, which shows correct reflexes of PPh \*baRa?.

Contrast between former variants of the same phoneme has also come about medially either through borrowing, as in the case of b and þ, e.g. *tubu* *sugarcane* and *kubu?* *grave* (from Mal *kubur*), or through merger with another phoneme; thus the non-stop variant of \*d has become ! in Sarangani, merging with a consonant already in contrast with d (see section 2.3.10.).<sup>50</sup> Thus stop and non-stop reflexes of PSan \*b, \*d and \*g must be treated as phonemically separate in present-day Snl.

The phonemic status of voiced stops and corresponding non-stops has not been clearly established for Tal but more data would almost certainly show them to be phonemically separate (see section 2.1.3.1.).

In Rth [d] and [ȝ] are in complementary distribution as are [g], [h] and [x]. [b], [þ] and [p] are also in complementary distribution except that [b] sometimes occurs intervocally. For the purposes of this comparative study stops and continuants are treated as phonemically separate; thus phonemes b, d, g, þ [þ, p], r and h [h, x] are recognised. (See section 2.1.5.1. for a fuller discussion.)

Because of the variation in the present-day languages it is not possible to state positively the distribution of allophones in PSan. But it appears that \*b, \*d and \*g had continuant allophones intervocally except following \*ə and that voiced stops occurred in all other environments. From the evidence of the present-day languages it is possible that some degree of variation, free or conditioned, occurred word-initially in PSan. The continuant allophones were probably [þ], [ȝ] and [g].

While each language has undergone some unique changes since separating from the other languages, San, Snl and Tal have basically similar reflexes. Ban and Rth are the most different, Ban having only stops and Rth being the only language to have continuants finally and following the reflexes of \*ə, although continuants occur in this last environment also in the Tahulandang dialect of San.

The evidence suggests a grouping of San, Snl and Tal, supporting other stronger evidence (see sections 2.3.2. and 2.4.) but it gives no support for a grouping of Ban and Rth.

### 2.3.14. Reflexes of \*ŋg

Medial \*ŋg is reflected in San, Snl and Tal as ŋg and in Ban and Rth as nk:

PSan	San	Ban	Rth	
*tengoR	tengohə?	téŋkoho?	tenko	<i>to strike</i>
*tingum	tinguŋ	tíŋkuŋ	tíŋkum	<i>riddle</i>
*səŋgap	səŋga?	səŋká?	sunkap-án	<i>group</i>

Two unexplained occurrences of ŋg have been recorded for Rth: ungón *to call out*, cf. San əŋgo *to make a noise to attract attention*; ŋgonguán *throat*, cf. San təŋgołəŋ. The first pair above can be treated as cognates but the second pair cannot despite the partial phonological resemblance. In ŋgonguán Rth has possibly retained medial ŋg under influence of the identical word-initial cluster.

In Ban also  $\eta g$  occurs in two recorded items: *púŋgutu?* *stunted*, cf. San *pungu?* *tailless*, Rth *punku?* *low*; Ban *sagéŋgele?* *to carry on strap*, cf. San *sagéŋgele?*, Rth *saŋkel*. In the first item above Ban has irregular paragogic instead of t-reduction, which indicates a borrowing (see section 2.3.3.). In the second item the preceding  $\eta$  may have been an influencing factor.

In Ban and Rth the cluster  $\eta g$  occurs intervocally at a morpheme boundary, e.g. Ban *maN-* + *gau* → *mangau* *to lie*, Rth *muN-* + *ha?* (from \**gaet*) → *munga?* *to lessen*. The cluster  $\eta g$  also occurs word-initially in Rth (see section 2.3.19.).

The replacement of \* $\eta g$  by  $\eta k$  within the morpheme is an innovation uniquely shared by Ban and Rth within the Sangiric group but the change also occurred among the Minahasan languages and may be an areal phonomenon (see section 2.4.).

### 2.3.15. Gemination of consonants

Consonants became doubled in Tal following \* $\theta$ , which later became a (see sections 2.3.1. and 2.3.9. for examples). They also became doubled preceding the paragoge unless the preceding syllable contained \*R (which later became k or ź), a doubled consonant or a nasal-stop cluster (see sections 2.1.3.1. and 2.3.9. for examples).

Gemination following PSan \* $\theta$  must have occurred before \* $\theta$  was replaced by a as it does not occur after a where this is from PSan \*a, e.g. PSan \*ləmis > Tal lammisa *to drown* but PSan \*Ramis > Tal źamissa *to knead*. It is unlikely that gemination occurred word-finally so it may be assumed that it developed after addition of the paragoge (see also note 44 to section 2.3.3.).

In San single consonants were either doubled or preglottalised after \* $\theta$  (see section 2.1.1.1.) and in Snl doubling of l and nasals occurred in this environment (see section 2.1.2.1.), in both languages non-phonemically.

It is possible that doubling of consonants began to develop in PNSan.

### 2.3.16. Reflexes of \*w

PSan \*w occurred finally in the diphthong \*aw. Reflexes of this in the present-day languages are described in section 2.3.6.

It also occurred word-medially. Evidence from Ban is necessary for reconstructing \*w in this position. In all other Sangiric languages \*w merged with b, while in Ban the sequence \*wa became o:

PSan	San/Snl	Tal	Ban	Rth
*awak	aþa?	aþa?a	ao?	aþak <i>body</i>
*bulawan		bu aþanna	bulaon	gold

These reconstructions are supported by evidence from PPh: \*hawak *body*, \*bulaw-an *gold (coloured)*. For the last item above related forms occur in the other Sangiric languages: San bu|aeŋ, Snl þuaen *gold*, Rth þulaun *coloured stones used in rings*. However, these show unexplained differences in the final syllable and the PSan form is reconstructed on the basis of the Tal and Ban evidence alone (see \*bulawan in the wordlist).

The two examples above are the only items for which Ban evidence has been recorded. Nevertheless this is sufficient to establish that \*w occurred

intervocalically in PSan and that it was phonetically different from the inter-vocalic allophone of \*b, which is reflected as b in Ban; both sounds being reflected as b in all other Sangiric languages. Some other items can be reconstructed for PSan with medial \*w even though Ban evidence is lacking, if external evidence supports the reconstruction. Thus from San, Snl, Tal, Rth saba and PPh \*sawa spouse can be reconstructed PSan \*sawa. Where external evidence is lacking, however, \*w cannot be reconstructed unless a Ban item is known.<sup>51</sup>

In Ban and Rth the high back non-syllabic vocoid [w] occurs word-initially preceding a stressed vowel, e.g. Ban [wá]a], Rth [wála] tusk, cf. Ban [úa] to disembowel, Rth [úa] to rummage. In initial position [w] is best treated as an unstressed variant of u; thus Ban, Rth uala tusk.

In Snl also [w] occurs in initial position. Maryott (1977a:265) treats this as a consonant, which contrasts with b. Thus baw gong and waw eight. In this work Maryott's Snl phoneme w in initial position is treated as a non-syllabic variant of u.

In San and Tal [w] does not occur initially, syllabic [u] corresponding to [w] in the other languages. There is no contrast between [w] and [u] in Ban and Rth and this may also be the case in Snl.

On the above evidence \*w is not reconstructed word-initially for PSan. Phoneme \*u may have had a variant [w] in unstressed initial position, possibly in free variation with [u].

### 2.3.17. Vowel reduction

There have been a number of innovations whereby vowel sequences have reduced to single vowels. Some of these changes are mentioned elsewhere. For instance, within the morpheme \*ə was lost immediately following another vowel except in San (see section 2.3.1.). Sequences of identical vowels, resulting from loss of \*l or \*y, underwent coalescence in Snl (see sections 2.3.10. and 2.3.11.). In some environments it cannot be determined whether vowel reduction is regular, because of the limited number of examples recorded. In such cases a comment is made in a note to each individual item in the wordlist. Some other reductions are the following:

(a) Sequences of \*a followed by another vowel sometimes reduced to single mid vowels in closed final syllables in Tal. The sequences \*ai and \*ae reduced to e and \*au reduced to o. No examples of PSan \*ao with Tal reflexes have been recorded:

PSan	Tal
*paɪt	petta bitter
*laed	ledda foot
*daʊŋ	donja leaf

This change was not regular and exceptions have been recorded, e.g. papaidda wing,<sup>52</sup> baekka to pay (PSan \*baeR), <laida> (Adriani) full, satiated, cf. San laedə?.

(b) In Rth the sequence \*ua reduced to o preceding final i and y. This change did not occur where the two vowels were earlier separated by \*R, subsequently lost in Rth (see section 2.3.9.), as in the last example below:

PSan	Rth	
*suay	soy	<i>finished, used up</i>
*uay	oy	<i>mango</i>
*uai	oy	<i>rattan</i>
*puRay	puay	<i>shell fish</i>

### 2.3.18. First syllable loss in Rth

Many PSan words with the structure C<sub>1</sub>ə(N)C<sub>1</sub>..., i.e. with initial syllable Cə followed by a consonant identical to the first consonant, with or without a homorganic nasal, lost the first syllable in Rth:

PSan	Rth	
*kəŋkum	ŋkum	<i>handful</i>
*kəkud	kur	<i>to dig</i>
*tətuR	tu	<i>hot coals</i>
*səsub	süb	<i>smoke</i>
*bəmbulu	mbulu	<i>feather</i>
*dəndipaR	ndipa	<i>rainbow</i>

In most cases recorded the PSan forms reflect PPh repeated monosyllables. Thus the first two items above reflect PPh \*kəmkəm and \*kuDkuD respectively. Although PPh etyma are not known for the second two items in the list it is very likely that they reflect earlier RMs.<sup>53</sup> They have the same structure as reflexes of RMs and the only other recorded words in which initial syllable loss occurred are three syllable words in which a following nasal-stop cluster occurs, as in the last two items in the list above.

Loss of the first syllable did not occur in all items reflecting PPh RMs:

PPh	PSan	Rth	
*bukbuk	*bəbuk	bubúk	<i>borer</i>
*dəmdəm	*dəndum	rundúm	<i>dark</i>
*panpan	*pəmpañ	pumpáñ	<i>river bank</i>

### 2.3.19. Prenasalisation in Rth

Voiced stops word-initially became continuants in Rth (see section 2.3.13.) except that in a small number of words stops were preserved by the addition of a preceding homorganic nasal, as in the first two examples below. Loss of first syllable Cə before a nasal-voiced stop cluster (see section 2.3.18.) left the cluster word-initially, as in the last two examples below:

PSan	Rth	
*gipu	ŋipu	<i>soot</i>
*bud/lalak	mbulalak	<i>tree sp.</i>
*bəmbulu	mbulu	<i>feather</i>
*dəndeRoŋ	ndeoŋ	<i>tree sp.</i>

In some cases where initial syllable Cə was lost from PSan reflexes of earlier RMs an initial nasal was added to the remaining monosyllable:

PSan	Rth	
*dədap	ndap	<i>tree sp.</i>
*tətul	ntul	<i>tree sp.</i>

Many of the words undergoing prenasalisation are the names of plants and trees; the reason for this connection is not known.<sup>54</sup>

### 2.3.20. Irregular correspondence of \*d and \*l

In a number of cases irregular correspondences occur among the languages involving the sounds d (or r) and l. In some items the North Sangiric languages reflect \*l while Ban and Rth reflect \*d:

San	Snl	Tal	Ban	Rth	
linuhə?	rinurə?	linukka	dīnuhə?	rinu	<i>earthquake</i>
lonj?	lonj?	-	donj?	ronj?	<i>gums</i>
kulilase?	-	-	kudilas	kurilas	<i>spleen</i>

Initial r in the first Snl word above reflects earlier \*l with regular assimilation to the following r (see section 2.3.10.).

In a few cases Rth reflects \*l while San reflects \*d, the reverse of the pattern illustrated above (cognates in other languages have not been recorded):

San	Rth	
buraʃa?	mbulalak	<i>tree sp.</i>
dulunj	lulun	<i>to roll up</i>
daʃajiran	laliniran	<i>tree sp.</i>

In a few recorded items the correspondences show even more irregularity:

San	Snl	Tal	Ban	Rth	
dila, lila	dila	lila	dila	rila	<i>tongue</i>

In all these cases the explanation is probably that there were competing forms in the proto-language. In most of these words the consonant following the consonant in question is l and the forms with d may have arisen through dissimilation. Where PPh etyma occur they have \*l, with the exception of PPh \*dila *tongue*.

In a number of cases the competing forms both still occur in San. Thus: dila, lila *tongue*; linuhə?, Siau dinuhə?, Tahulandang dinuhi? *earthquake*.

### 2.3.21. Development of final glottal stop in San and Snl

In a number of words San and Snl have a final glottal stop where cognates in the other languages end in a vowel:

San	Snl	Ban	Rth	
ake? <sup>55</sup>	ake?	ake	ake	<i>water</i>
deno?	deno?	deno	deno	<i>to bathe</i>
deso?	deso?	deso	deso	<i>to store</i>
kite?	kite? <sup>56</sup>	kite	kite	<i>we</i>
kapuna? <sup>57</sup>	-	kapuna	kapuna	<i>dog</i>

Where cognates are known in other languages they do not have final glottal stop. Thus Ternate aki, Galelarese ake *water*; Mal kita, PPh \*kita we. Further, San sometimes adds final ? in borrowings, although no cognates of such items have been recorded for Snl. Thus San diaga? *to guard* from Mal jaga, kunsí? *key* from Mal kunci, piso? *knife* from Mal pisau.

Thus the occurrence of final ? in San/Snl in such items is clearly a case of addition rather than of loss in the other Sangiric languages (Tal does not enter into the discussion as it has no final glottal stop). The addition of ? in such words is irregular, affecting only a small number of words previously ending in a vowel.<sup>58</sup>

### 2.3.22. Further changes to \*l in Ban and Rth

Where a PSan word had the final sequence \*luR, Ban and Rth reflexes have h (later lost in Rth) instead of l. Only one PSan item has been positively reconstructed with final sequence \*liR and the change occurred here also.<sup>59</sup> The change may have occurred before changes to \*R and probably represents assimilation. In a very few items PSan final \*l was also replaced by h in Ban, as in the last two examples below. No Rth cognates of such forms have been recorded:

PSan	San	Tal	Ban	Rth
*uluR	u!uhə?	-	úhuhu?	uu <i>to lower</i>
*kuluR	kulu?	-	kúhuhu?	kuu <i>breadfruit</i>
*salíR	ensalí?	-	sáhihi?	sai <i>floor</i>
*pankul	pankulə?	-	páŋkuhu?	- <i>to beat</i>
*ongol	ongo?	ongola	óŋkoho?	- <i>to give</i>

Only one example has been noted where medial \*l was not replaced in Rth, and no examples in Ban:

PSan	San	Rth
*silaR	silahə?	sila <i>palm sp.</i>

This is the only recorded example where original \*R was not preceded by a high vowel and consequently no general rules can be drawn.

Where PSan had initial \*l and medial or final \*R the initial \*l was usually lost in Ban, although it remains in a few items. Examples in Rth are very limited but they show that initial \*l was retained, replaced by h or lost. All these possibilities are illustrated below:

PSan	San	Ban	Rth
*leRe?	lehe?	ehe?	- <i>neck</i>
*leneR	lenehə?	énehe?	- <i>calm</i>
*lətuR	lətuhə?	lutuhu?	útú <i>bang, thud</i>
*lintuR	lintuhə?	íntuhu?	Tntu <i>to descend</i>
*liRa	liha	-	ia <i>ant sp.</i>
*le/inoR	lenohə?	-	lino <i>insane</i>

Although the changes to \*l described above appear to be irregular, except where medial \*l occurred before final sequence \*uR, they nevertheless reflect a common innovation in Ban and Rth and this constitutes evidence for a period of uniquely shared ancestry.

### 2.3.23. Stress

Word stress is predictable in San, falling on the penult (not counting the paragoge) even if this contains  $\emptyset$  (see section 2.1.1.2.). Although stress placement has not been fully determined for Tal it is probably predictable and can occur on a where this reflects a penultimate \* $\emptyset$  (see section 2.1.3.2.). Thus in San and Tal there is evidence that PSan stress also fell always on the penult:

PSan	San	Tal	
*télu	télu	tállu	three

Stress is largely predictable in Snl, although in a small number of words phonological changes have resulted in stress contrast. Thus contrast occurs between kápú *to rub* and kapú *desire*, the latter from earlier \*kapúyu, with loss of \*y and vowel contraction (see section 2.3.11.). The word pasáŋ *to carry on the shoulder*, from PSan \*pasaən, exemplifies occurrence of final stress resulting from  $\emptyset$ -loss.

In Snl, Ban and Rth stress shifted from penultimate schwa to the following syllable. Schwa was later replaced by another vowel in Ban and Rth (see section 2.3.1.) resulting in unpredictable word stress, e.g. Ban búku, Rth búku *to kneel*, from PSan \*buku, and Ban, Rth pukú *to bend*, from PSan \*pəku.

It is possible that in PSan stress did not fall on \* $\emptyset$  but later shifted to penultimate schwa in San and Tal. This is, however, most unlikely. The doubling or preglottalising of consonants after schwa in San and Tal was almost certainly a device to retain penultimate stress.<sup>60</sup>

Stress changes in Snl post-dated its split with San and so are recent. Thus, although the changes are largely identical to those in Ban and Rth, they can only have been parallel, and not shared, developments.

Although changes to \* $\emptyset$  must have occurred separately in Ban and Rth (see section 2.3.1.), stress shifts occurred before changes to \* $\emptyset$  and could have been a common innovation. That stress shift in Ban and Rth pre-dated changes to \* $\emptyset$  is established by evidence such as that given above: if \* $\emptyset$  in \*pəku *to bend* had become u before stress shift then no explanation could be given as to why stress did not change in \*buku *to kneel*. Consequently, stress shift must have occurred before \* $\emptyset$  was replaced by another vowel.

In Rth stress shifted from a penultimate high vowel to an immediately following vowel in a final closed syllable:

PSan	Rth
*Ríud	júr <i>to pull</i>
*tían	tián <i>belly</i>
*súan	suán <i>to plant</i>
*túid	tuír <i>stump</i>

This stress shift operated before the loss of \*h, reflecting PSan \*R, in Rth. Subsequent loss of \*h resulted in stress contrasts:

PSan	Pre-Rth	Rth
*níuR	*niúh	níu <i>coconut</i>
*níRu	*níhu	níu <i>winnowing pan</i>
*búa?	*búa?	búa? <i>to stand up</i>
*súRa?	*súha?	súa? <i>fish poison</i>

## 2.4. RELATIONSHIPS WITHIN THE SANGIRIC GROUP

The Sangiric languages are not all equally closely related to each other but belong to several subgroups.

An examination of shared innovations shows clearly that San, Snl and Tal form one branch of Sangiric languages; the North Sangiric branch. The strongest evidence for this grouping is metathesis of t and s, discussed in section 2.3.2.

Phonological evidence for the place of Ban and Rth within the Sangiric group is limited. However, from all the evidence available they must form a subgroup, being more closely related to each other than to any of the other Sangiric languages.

Ban and Rth share the change of \*g to k after η morpheme-medially (see section 2.3.14.), an innovation not shared by the North Sangiric languages. This may appear to be good evidence for subgrouping. However, the change also occurred in the Minahasan languages:

PSan	Ban	Ttb, Tbl
*sengot	senko?	senkot      to sail
*tengoR	téñkoho?	tenkor      to beat (gong)

Thus it is quite likely that the change spread by diffusion to Ban and Rth from adjacent Minahasan languages and if so it does not reflect an innovation in an exclusively shared parent language.

In a small number of items Ban and Rth reflect initial or medial \*d while the North Sangiric group reflect \*l (see section 2.3.20.). Whatever the origin of the difference the agreement between Ban and Rth points to a period of uniquely shared ancestry.

Ban and Rth share the change of \*l to h under some conditions (see section 2.3.22.). While this was mostly sporadic it appears to have been regular in reflexes of PSan words ending in the sequence \*luR. This change is not known to have occurred in other nearby languages and it is among the strongest evidence for a grouping of Ban and Rth.

In the wordlist Ban and Rth each have more cognates with San than with each other. This is a reflection of the limited opportunity available for collecting data on Ban and Rth. For any Ban or Rth item recorded related San forms could be sought at any time in Steller and Aebersold's dictionary but time did not permit the seeking of cognates in Rth and Ban for every item recorded in the other language.

Further, because of other priorities in the limited time available for work with informants, there was usually no attempt to find a cognate in the other language for any Ban or Rth item for which there had not already been located a cognate in San or another North Sangiric language.

Thus the limited number of uniquely shared lexical items recorded for Rth and Ban is, at least in part, a result of lack of opportunity during field work and it therefore does not pose such a problem to a Ban-Rth grouping as might at first appear.

Items which are apparently unique to Ban and Rth or which have uniquely shared phonological or semantic features are listed below. Where items have cognates in San but with different meanings and no outside cognates are known it cannot be established whether the semantic change occurred in Ban and Rth or in San. However, clear cases of uniquely shared retention are excluded here. The

list is not exhaustive; for one thing forms which are possibly uniquely shared but about which doubt of any kind exists are omitted. The items below are also given in the wordlist in section 3.2. with further details.

Ban bahudin, Rth baurin *tree sp.*

Ban bebe?, Rth bebe? *to carry by rope.* These are cognate with San bobe? *carry by handle*, reflecting PPh \*bitbit. Ban and Rth share irregular change of \*ə to e. Although \*ə regularly assimilated to the following vowel in Ban the occurrence of stress on the penult in this item shows that the change here occurred prior to stress shift, not after, as occurred regularly (see section 2.3.23.). Since the independent occurrence of such an irregularity in the two languages can be discounted it must have occurred in their shared parent language.

Ban bokan, Rth bokan *maize.* San bokan *cataract of the eye* may be related.

Ban bule, Rth bule *to release.* San bule *to forget* is related but has a different meaning.

Ban dádehe? *to depend,* Rth rare *to lean.* San darehə? *fish with diagonal stripes* may be related. This item may reflect PPh \*zazar in rows but if so has undergone semantic change.

Ban dahumú, Rth raamú *nest.*

Ban duhú?, Rth ruup *face.* San dəhu? *forehead* is related but has a different meaning.

Ban hemban, Rth ēmban *flame.* San hiamban *flame* is related but has vowel sequence ia where Ban and Rth have e.

Ban kanehaba?, Rth kaneab *yesterday.*

Ban kajeden, Rth kaneren *when?.* The cognate in the North Sangiric languages has final ? instead of ə.

Ban, Rth kumu *you (pl).* This reflects \*kemu as do Siau kemu and Tahulandang kumu. The forms in these southern San dialects may be borrowings from Ban (see section 1.6.).

Ban, Rth kembo? *back of the neck.*

Ban kóyaba?, Rth koyab *to yawn.* San kiolabə?, Tal <iolaba> are related but have vowel sequence io where Ban and Rth have o.

Ban, Rth lasia nit, louse egg. The languages share irregular replacement of \*ə by a, cf. San ləsia, PMin \*ləse?a nit.

Ban, Rth maya all. PMin \*baya all is related but has a different initial consonant.

Ban, Rth nai? long.

Ban polaen, Rth polán *upper arm.* San polaen *wrist* is related but has a different meaning.

Ban timbonan, Rth timbonan *head.*

Ban tingkahia, Rth tonkayá *ear.* A parallel to the first syllable occurs in Ban timpunú, Rth tomponú *tortoise*, which are innovations, cf. PPh \*pəñu *tortoise.* The words for 'ear' and 'tortoise' cannot be chance similarities and must reflect a shared innovation in each case.

Ban, Rth tondo to push. San tondo to shove off (a boat) is cognate but has a different meaning.

There is also a limited amount of grammatical evidence available to support the lexical and phonological evidence for a grouping of Ban and Rth. The two differ from the other languages, and agree with each other, in genitive markers. The non-personal marker in Ban and Rth is nu, for some speakers ñu in Rth, in free variation with a homorganic nasal before stops and s:

Ban tilo? nu babi, tilo? mbabi pig's foot.

Rth lar nu ñabi, lar mbabi pig's foot.

In the North Sangiric languages the genitive marker is a nasal after a vowel and u after a consonant:

San apen u rano banks of the lake, tipu nsoso? cigarette smoke.

SnI kan u katoan snake's food, bisa ñkatoan snake's poison.

Tal sakaen u ratu chief's boat, bale ndatu chief's house.

The personal genitive marker in Ban and Rth is ni, for some speakers ñi in Rth:

Ban baley ni jon John's house.

Rth lar ni leksi Lexy's foot.

In San and Tal it is i after a consonant (the paragogic being lost) and a nasal after a vowel:

San ahus i dabid David's son, bale nsimone Simon's house.

Tal arann i tuanna name of the Lord, bale njon John's house.

In Ban and Rth the genitive marker with plural pronouns is a prenasal (singular genitive pronouns being enclitic):

Ban niala? nside fetched by them, baley nside their house.

Rth niala? ñkami fetched by us, baley ñkami our house.

In San and Tal the genitive marker with pronouns, as with personal nouns, is i after a consonant and a nasal after a vowel:

San laed i kami our foot, bale ñkami our house.

Tal nisuann i ami planted by us, bale ñkami our house.

Lexicostatistics provides supporting evidence for subgrouping the Sangiric languages. A lexicostatistical comparison of the five languages yields the percentages given in the table below. The list used is the same as that presented in Sneddon 1970, drawn from the lists of Swadesh (1955) and Gudschinsky (1956).<sup>61</sup>

Lists were obtained directly from informants for San, Tal, Ban and Rth. This was not possible for SnI, for which the list in Reid 1971 was used. This omits a number of items on the list used for the other languages. Considering the percentages obtained it is doubtful if a fuller SnI list would have produced other than a negligible difference in the scores.<sup>62</sup> The estimated percentages of shared basic vocabulary for the five languages are as follows, percentages having been taken to the nearest whole number:

	Sn1	San	Tal	Ban
Rth	54	54	47	59
Ban	58	58	51	
Tal	66	66		
San	82			

San and Snl, with 82% cognates, clearly form a subgroup of closely related speech forms.<sup>63</sup> These link with Tal at 66% in the North Sangiric subgroup. The percentages clearly suggest Ban and Rth are excluded from this grouping. But on the surface the figures for these two languages do not suggest they form a subgroup; Ban's score with Rth is not significantly higher than its score with San and Snl. However, the following interpretation is suggested. Just as the geographical proximity of Ban, San, Snl and Tal is reflected in numerous shared phonological characteristics spread by diffusion it must also be assumed that Ban's proximity to San has influenced its vocabulary as well, through borrowing and shared retention. Thus it could be expected that Ban's percentage with the other Sangiric languages would be somewhat inflated. On the other hand, Rth's isolation from the other four languages would have prevented its participation in borrowing and shared retention just as it has excluded Rth from phonological innovations spread by diffusion. Ban's percentage with each of the three North Sangiric languages is four points higher than is Rth's and this could represent inflation resulting from geographical proximity, just as its percentages with Tbl and Tse, with which it shares its borders, are higher than its scores with the other Minahasan languages (see Sneddon 1970). Thus geographical factors could have had a masking effect on the Ban-Rth relationship, which would show up more clearly in the lexicostatistical percentages if Ban had also been isolated from the North Sangiric languages.

In section 1.6. it was suggested Ban and Rth split from each other shortly after the original dispersal of PSan. The exclusion of Rth from probably all of the innovations which Ban shares with the North Sangiric group points to this as do important phonological differences between the two such as differences in their reflexes of voiced stops (see section 2.3.13.). Also, if Rth's lexicostatistical percentages with the other languages are taken as more reliable than Ban's then, although the Ban-Rth percentage of 59, as against 54 with San/Snl, indicates a Ban-Rth grouping, the difference of only five percentage points nevertheless supports the suggestion of only a short period of uniquely shared ancestry.

San, Snl and Tal form the North Sangiric group, as previously discussed. The first split was between San/Snl and Tal. San and Snl link closely, probably representing a string of dialects, stretching from the islands just north of the Minahasan coast to the Sarangani Islands and adjacent Mindanao mainland areas. Rth and Ban form the South Sangiric group, splitting from each other soon after the original dispersal of PSan.

The relationships of the Sangiric languages are represented in the following diagram:

