Internal divisions of the Flores-Lembata subgroup of Central Malayo-Polynesian

Matthew Doyle

27 August 2010

Leiden University

Supervisor: Dr. A.T.P.G. van Engelenhoven
Second reader: Dr. A.C. Schapper

TABLE OF CONTENTS

Maps Tables Appendices

1   Introduction

The Austronesian language family presently consists of over 1200 languages, which amounts to nearly a fifth of the world's living languages, making it the most numerous language family in the world (Lewis 2009). The area containing the primary concentration of Austronesian languages stretches from Taiwan in the north, where the origin of Austronesian languages can be traced, to the southern islands of Indonesia, the country with the highest number of Austronesian languages. At its geographic extremes, languages of the Austronesian family are spoken in Madagascar, Hawaii, and Easter Island. Given that in the Austronesian family there is a large number of languages with a wide and unique geographic distribution, one would expect to find difficulties when attempting to create internal classifications within the family. This difficulty may be softened when one is attempting to discover the internal divisions of a small group rather than a large one, which will be the focus of this paper.

The aim of this paper will be to improve our understanding of the historical relationships of a group of Austronesian languages spoken in eastern Indonesia referred to in Ethnologue (Lewis 2009) as the Flores-Lembata (FL) branch of the Timor group within Central Malayo-Polynesian (CMP), which is part of the Malayo-Polynesian group (see Map 1 below for a view of the Malayo-Polynesian language area). In Map 1 below, the section outlined in the lower-right, just West of Papua, is the CMP-area. The languages of the Flores-Lembata group are spoken on a chain of islands that stretches East to West from Flores Island to Alor Island (see Map 2), an area which falls within the eastern Indonesian province of Nusa Tenggara Timor (NTT). For this paper, I will be utilizing existing Swadesh-style lists that have been collected during previous research in the area. With these word lists I have created a database that allows for comparison of shared vocabulary between the languages. I used the basic vocabulary items as a base for an initial exploration into the internal divisions of the Flores-Lembata (FL) group, and their shared history and proto-forms.
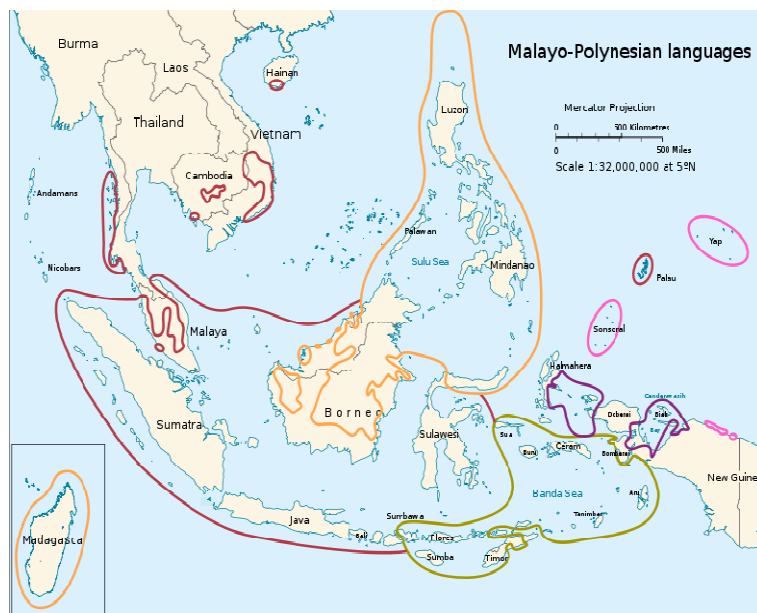
Map 1. Malayo-Polynesian Languages (Wikipedia article)

There has been much debate over the internal divisions within the Austronesian language family (see Blust 1993, 2008; Ross 1995; Adelaar 2005; Donohue and Grimes 2008; among others). I will later briefly summarize previous research, but first I will attempt to explain the steps down from the Austronesian family as whole, to Flores-Lembata; essentially, a rundown of how we got from Proto-Austronesian (PAN) to the individual languages of the FL group.

2   What is CMP, and where is Flores-Lembata?

2.1.   From Malayo-Polynesian languages to Timor languages

The Austronesian language family, like most language families, consists of several important branches, each of which contain many languages which can be further divided into groups based on shared characteristics which indicate their shared histories and 'genetic' relations. When I refer to the languages of the Austronesian family, I refer to all of the approximately 1200 Austronesian languages, which have been grouped together based on their shared history and characteristics. When I mention Flores-Lembata, I refer to a small branch all the way at the bottom of the Austronesian family tree.

So how do we get from Austronesian to Flores-Lembata? I will explain using information from Ethnologue (Lewis 2009), as well as articles by Tryon (1995) and Ross (1995), both of which are found in the Comparative Austronesian Dictionary (1995). These conclusions are said at times by Ross and Tryon to be questionable, however they appear to be the most accepted basic conclusions about the internal divisions within Austronesian present in the literature today. First, Austronesian can be split into four primary groups (see Figure 1 below), three of which, referred to collectively as the Formosan languages, are situated in Taiwan. This is the primary reason for the widely held belief that Taiwan is the site Austronesian homeland. The fourth group, Malayo-Polynesian, contains the remaining (and in fact nearly all of the) Austronesian languages. The reconstruction of Proto-Malayo-Polynesian (henceforth PMP) is generally accepted, but its purported internal divisions remain very controversial.

Figure 1. Austronesian Genetic Tree (Bellwood et al 1995)
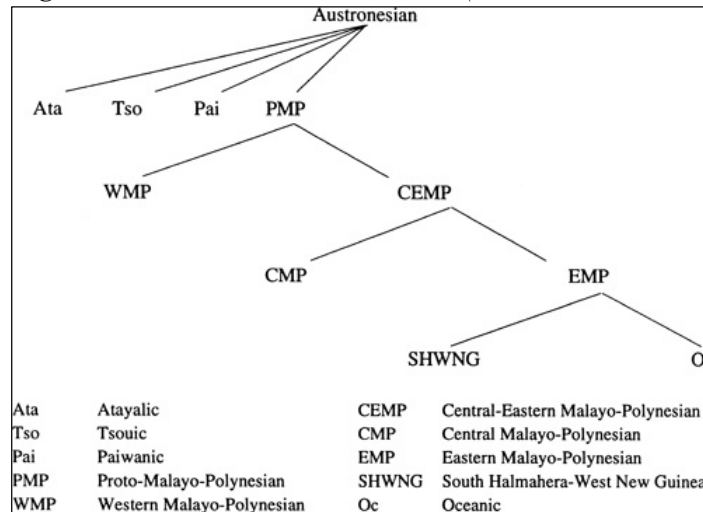
Within Malayo-Polynesian, the pattern of diversification may be seen above in Figure 1. It is believed that historically in Malayo-Polynesian language groups, migrations have taken place by segments rather than wholes of populations (Ross 1995). To get to the Flores-Lembata group, one must proceed down the tree from PMP to find the split between Western Malayo-Polynesian (WMP) and Central/Eastern Malayo-Polynesian (CEMP).

Within CEMP, there are two main branches, Central Malayo-Polynesian (CMP), and Eastern Malayo-Polynesian (EMP). EMP represents the segment of the CEMP population that separated from the group, and CMP the portion that remained 'at home', that is EMP speakers migrated and CMP speakers did not. In time, the original language spoken by the CMP population diverged into many distinct languages, totaling over 100. One branch within CMP is the Timor branch, which is

5

divided by Ethnologue (Lewis (2009) into three groups. These are the Ramelaic group, the Extra-Ramelaic group, and the Flores-Lembata group. So, to sum it up. we have this journey down the tree: Austronesian > PMP > CEMP > CMP > Timor > Flores-Lembata.

2.2.   Introducing Flores-Lembata

The Flores-Lembata group is shown (Lewis 2009) to have 13 members, and its 13 members are not yet split into any further subgroups. In fact, it is not yet all proven that all of its members are single, distinct languages. Lamaholot, for instance, may actually be a cover term for many distinct languages, three of which are featured individually in this paper. In this study I will, by using material that is available (unfortunately, reliable material is not available for all or even most the languages of the group), attempt to determine if there are any further clusters within the FL group, and investigate the legitimacy of its current purported borders.

Below is a table with basic information for all the FL languages given in Ethnologue (Lewis 2009), which at the present time must be taken as the basis for what Flores-Lembata is, and what languages it contains. In the table, the letters next to some of the language names (the ones that are in my database) represent my shorthand representation for them in this paper and in the database. The letters do not refer specifically to any properties of the languages or language names themselves, they are merely titles of convenience. As I mentioned previously (and as mentioned in Ethnologue), Lamaholot is split into several lects, many most likely distinct enough to be classified as different languages, though this has not yet to my knowledge been proven or widely accepted. Lewolema (B, Pampus 1999), Lewoingu (C, Nishiyama & Kelen 2007), and Solor (D, Klamer 2002) are by many considered to be lects of Lamaholot, but have been included as separate languages in the present study. At the very least this will help determine which of them are more closely related.

Table 1. The Flores-Lembata Languages (Lewis 2009)

FL Languages	Alternate Names	Number of Speakers	FL relational notes
Adonara	Nusa Tadon; Sagu; Vaiverang; Waiwerang	17.000	
Alor (G)	Alorese	25.000	Closely to: Lembata and Adonara.
Ile Ape	Nusa Tadon	N/A	
Kedang (F)	Dang; Kdang; Kédang; Kedangese	30.000	
Lamaholot (B, C, D) B: Lewolema lect C: Lewoingu lect D: Solor lect	Solor; Solorese	150.000 B: N/A C: N/A D: 12.000 (Klamer 2002)	Perhaps actually several distinct languages.
Lamalera (E)	Kawela; Lebatukan; Mulan	N/A	
Lamatuka	Lamatoka	N/A	Close to: Lewo Eleng.
Lembata, South		N/A	
Lembata, West	Labalekan; Mingar	N/A	
Levuka	Lembata; Lewokukun; Lewuka; Painara	N/A	
Lewo Eleng	None	N/A	Closely to: Lamatuka.
Lewotobi	SW Lamaholot	289.000	
Sika (A)	Krowe; Maumere; Sara Sikka; Sikka; Sikkanese	175.000	

Below is Map 2, which gives a view of the immediate Flores-Lembata region. On the map there are pictured six islands, which are numbered on the islands themselves and in the key in the top-right portion of the map. The area in which each language is spoken is marked with a letter corresponding to the letter used for that language's abbreviation throughout the paper. Alorese is also spoken on the coast of the Bird's Head of Alor Island. The key to the languages and their letters is found in the top-left portion of the map, on the top-right is the key to the island names.

Language			Island	
Sika	A		Flores	1
Lewolema	B		Solor	2
Lewoingu	C		Adonara	3
Solor	D		Lembata	4
Lamalera	E		Pantar	5
Kedang	F		Alor	6
Alorese	G			

LANGUAGES AND ISLANDS OF THE FLORES-LEMBATA AREA
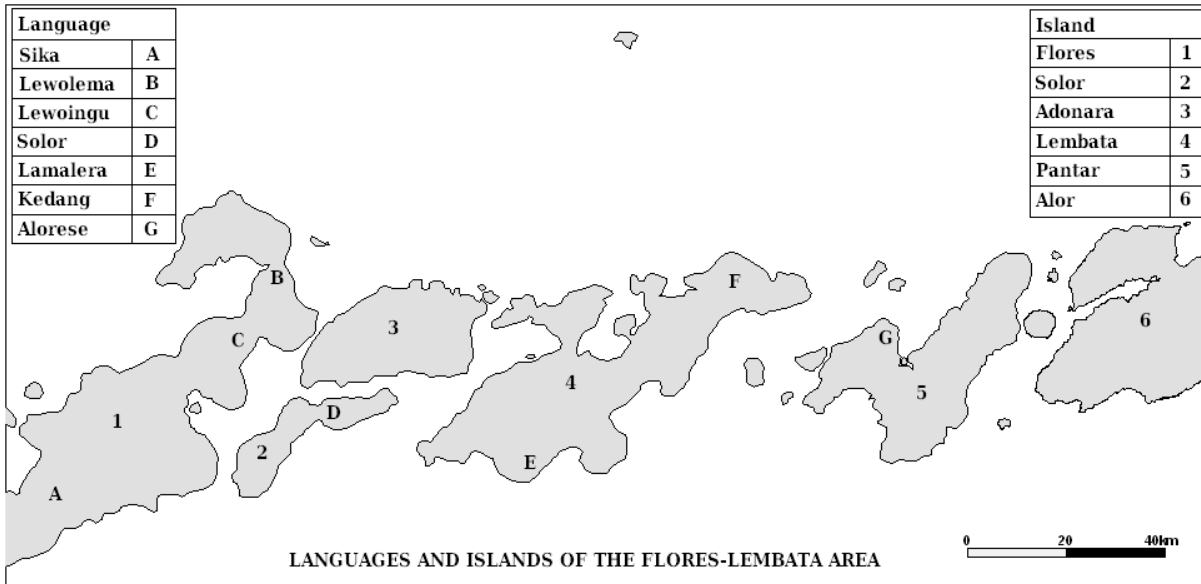
Map 2. Languages and Islands of the Flores-Lembata Area

Now that hopefully the location of the Flores-Lembata branch, both genetically and geographically, is clear, and its members have been introduced, I can proceed by attempting to summarize what has been done previously within CMP, with the goal of giving the reader an idea of what has happened in the research done in the greater region.

Later, I will be using data from the seven languages (often referred to by their corresponding letters from the map) in the above table to identify the low order relationships within what is currently classified as the Flores-Lembata subgroup. This will further the understanding of its internal divisions, and give an indication to where its borders lie. I believe pursuing bottom-up subgroupings using existing material that previously has not been used comparatively will bring interesting results to a poorly understood and understudied, yet very interesting area.

3 Previous and ongoing research related to the present study

3.1. Research within Central Malayo-Polynesian

The historical relationships of the Austronesian languages spoken in central and eastern Indonesia were first fully examined by Blust (1993). Blust claims that a group of over 100 languages spoken in the previously mentioned encircled area of Map 1 above, should be, based on shared, phonological, morphosyntactic, and semantic innovations, considered a subgroup of the Austronesian language family he calls Central Malayo-Polynesian (CMP). This viewpoint has been generally accepted in the literature, but in his paper, Blust admits that his proposed subgroup is problematic because of the uneven and overlapping distribution of the phonological and semantic features he proposes as shared innovations. The opinion (that it is problematic) is shared by Klamer

and Ewing (In press), Ross (1995), and Adelaar (2005), among others. Donohue and Grimes (2008) go further, claiming that the innovations presented by Blust do not constitute sufficient evidence for a subgroup, as well as questioning Blust's placement of the boundary between CMP and Western Malayo-Polynesian (WMP), the purported Austronesian subgroup adjacent to CMP to the West. Presently, the validity of CMP and its boundaries are in question, and a comprehensive study of its internal divisions has not been conducted.

The uneven and overlapping distribution of innovations in the CMP-area languages is the second of two general types of distribution of innovations that have occurred in Austronesian languages, as explained in Ross (1995). He outlines two ways in which Austronesian languages have diverged from each other; the first he calls *separation*, and the second *dialect differentiation*. In separation, two or more groups that share a language separate in a way that cuts off (at least almost) all contact between them (such as a migration), meaning that linguistic changes that occur in the two groups from that point on are independent of each other. In dialect differentiation, dialects of a single language gradually diverge from each other until the point where they should be considered distinct languages, which together form what Ross refers to as a *linkage*. These could also be called the 'stay-at-home' languages. The result of dialect differentiation (a linkage) is different from the result of separation because while the dialects are in the process of diverging, they also remain in sufficient contact with each other for an innovation from one dialect to spread to adjacent members of the linkage. This diffusion gives the shared innovations an overlapping pattern, instead of appearing in discrete groups, as in separation.

Returning to Blust (1993), he explains the inconsistent distribution of the CMP innovations logically by saying the languages of the area have emerged from a linkage, and not from a uniform proto-language. The extent to which this is true is under debate (see Donohue and Grimes 2008; Blust 2009), however it does appear that the languages in the CMP-area do demonstrate at least what could be called a 'linkage-style' relationship. In a linkage, (as with separation) all languages and forms may be traced back and compared to a proto-language (the language which, in dialect differentiation, initially split into the dialect chain), and this is in fact necessary because the tenability of any subgroup relies on a reconstruction of its proto-language.

Further research that takes a top-down view of the historical relationships of the NTT-East Timor languages, like the kind Blust (1993, 2009), Donohue and Grimes (2008), and others have put forth, will continue to confirm confusion. Bottom-up subgroupings, which focus on the role of shared innovations, and rule out innovations that resulted from diffusion, should be extensively investigated. Only after comprehensive low-order groupings have been accepted, may we find answers to where higher-order divisions have occurred.

3.1.1   Influence of typological data
The present manner of the distribution of CMP innovations, and the intensive contact and diffusion that caused it, has blurred the distinction between what data should be used for positing typological, genealogical, and geographical boundaries in the CMP-area. Himmelmann (2005) and Donohue (2007) have studied geographical groupings of Austronesian languages based on typological distinctions such as the order of genitives and nouns, and the order of numerals and nouns. The boundary drawn by Donohue relevant to this proposal lies between two groups he calls 'Western Austronesian' and 'Eastern Indonesian'. The location of Donohue's boundary lies in the same spot as the geographical boundary between the Austronesian branches of WMP and CMP (see Map 1). The similarity of the location of the two boundaries is not coincidental, because typological features and boundaries have silently influenced the positing of genetic groupings of the languages of eastern Indonesia. This is problematic because typological characteristics are not proper tools with which to formulate genetic relationships, and must be separated from genetic data when proposing genetic subgroupings.

3.2.   Swadesh lists, 'basic vocabulary', and Glottochronology
The Swadesh list was made famous by the American linguist Morris Swadesh (Campbell 2004), who was investigating trends in the retention or loss of common vocabulary items over the course of time. The field of G*lottochronology* was invented by Swadesh with the aim of being able to determine a date of the historical split of a given language into its daughter languages (e.g. from Latin to Spanish, French, Italian, etc.). The name glottochronology is often used interchangeably with *Lexicostatistics*, though, while similar, actually refers in a broader sense to the use of statistical information from lexical material for historical studies of languages.

The problems with the assumption that there is a basic, cross-cultural vocabulary that is subject to a rate of constant change across languages and language families are outlined convincingly with examples by Campbell (2004). Perhaps the most important problem identified is borrowing of words across languages. One may have the intuition that core vocabulary need not be borrowed because each language should 'already' have an accepted term for (almost) every word on the Swadesh list. However, there is evidence of borrowing of even the apparently most basic words in languages. For instance, in a number of Mayan languages, the native word for 'person', *winaq,* was replaced by *kriʃtian* 'person', a loan from Spanish, after the early Spanish arrivals contrasted *crisitanos* 'Christians', i.e. the portion of the native population that had been Christianized, with the pagans, or those yet to be converted. After enough Mayans had been converted to Christianity, the term *kriʃtian* remained to express 'person', wheras *winaq* 'person' was dropped from use. In English, the word 'person' is also a borrowed term, taken from French.

The other fundamental issue with the assumption of basic vocabulary is the assumption that all languages maintain a lexical item that is a one-to-one match with each form on the Swadesh list. Examples of a language having two forms for one Swadesh 'word' comes from Spanish, where there are forms for informal/familiar 'you' *tu*, and formal/polite 'you' *usted*. This is quite commonly seen in languages elsewhere in Europe and the rest of the world. Thus many languages have more than one neutral equivalent for a word on the Swadesh list, but also many languages have no equivalent for some Swadesh list items. For instance, many languages do not make a distinction between 'man' and 'person', and many others do not have a term for 'green' or 'yellow' (Campbell 2004). In fact, I think it would be extremely difficult answer the following question: 'Does any single semantic concept have one and only one principle and neutral form across all languages?'. Answering this question 200+ times is what it takes to believe without any doubt in the concept of a universal, basic vocabulary such as the original Swadesh list.

However, in historical linguistics, a Swadesh list is the beginning point of a comparative study, and with the above concerns in mind, I will be using a comparison of 'basic' vocabulary lists to investigate the historical relationships and internal divisions within the Flores-Lembata group. In order to investigate sound changes, I will identify which words are cognate across languages, and then look for correspondences between sounds (in this case primarily consonants) that occur in similar positions in those cognates. Swadesh began with a list of 500 words which were considered to be culturally basic or universal, but later shortened the list to 200, and eventually 100 words (highlighting the difficulties discussed above). For this paper, I use a list of 292 words as a base for the comparison between the Flores-Lembata languages, and that number is essentially further reduced when making comparisons because of the need to use words that are not only present in the data of the languages used to form the database, but which are also cognates with other languages of the study.

4   Explanation of sources used

4.1.   Austronesian Basic Vocabulary Database - (Greenhill, S.J., Blust. R, & Gray, R.D. 2008)

The Austronesian Basic Vocabulary Database (ABVD) is an open, online database that has been compiled by various researchers who have contributed data to the project (Greenhill, S.J., Blust. R, & Gray, R.D. 2008). It is an extremely valuable tool that contains Swadesh vocabulary lists for about half of the known Austronesian languages, including the established proto-languages. I made use of the ABVD for the proto-languages that were relevant to the Flores-Lembata group, with hope that they could help shed some light on the past history of the FL group and its proto forms.

4.1.1   PAN - (ABVD Online)

The data I used for Proto-Austronesian (PAN) was taken from the ABVD, in which its source/author is listed as Blust (1999), and its data listed as entered by Greenhill.

4.1.2   PMP - (ABVD Online)

The items in the database for Proto-Malayo-Polynesian (PMP) were taken from the ABVD, which lists the source author as Blust (1993), and its data listed as entered by Greenhill.

4.1.3   PCMP - (ABVD Online)

The items that appear in the database for Proto-Central Malayo-Polynesian (PCMP) are taken from the ABVD, and the PCMP author is listed as Blust (1993), and its data listed as entered by Greenhill.

4.2.   Sika (Language A) - Lewis & Grimes (1995): in Tryon 1995 (Comparative Austronesian Dictionary)

Sika, also known as Sikka, or Sara Sikka (Sara means 'way; language'). The double *k* is used by many in the literature and represents both the official name of the administrative regency of Sikka in the eastern part of Flores as well as the name of the language spoken by its residents. The double *k* does not signify anything phonetically about the word. The result of this is that Sika and Sikka are used interchangeably throughout the literature.

The number of Sika speakers totals 175,000, located within the Sikka regency. Sika is classified as a member of the Flores-Lembata branch of Central Malayo-Polynesian by Wurm & Hattori (1983). According to Lewis & Grimes (1995), the speech varieties, or sub-dialects, of Sika have not been carefully studied, but nonetheless three main dialects can be recognized. They are called Sikka Natar, Sara Krowe, and Ata Tana 'Ai or Sara Tana 'Ai.

Sikka Natar is found in the Sikka Natar village, which is located on the south coast of Flores island, as well as its surrounding communities, from Lela in the west to Bola in the east, and inland north to Nita. Within the Sikka region as a whole, the dialect of Sikka Natar appears to be the most highly regarded, perhaps owing to the village's past, which included it being home to a local royal dynasty.

In the central hills of the Sika linguistic area, the *Sara Krowe* dialect is spoken. Alternatively, the people who speak the Sara Krowe dialect may be referred to as *ata ʔiβaŋ* 'hill people'. *Tana 'Ai* is the third region of the Sika area, and is the mountainous eastern part of the Sika linguistic area. The people of the region, as well as those outside of it, use the names Ata Tana 'Ai and Sara Tana 'Ai to refer to the residents and the language of the region. According to Lewis & Grimes (1995), the people of the Tana 'Ai, who number approximately 9,800, are easily distinguishable from the people

of the other Sika areas because of their ethnolinguistic identity. The word list used for the comparative purposes of this paper was compiled by Lewis during work in Sikka Natar and Tana 'Ai area beginning in 1977, though specifically the Tana 'Ai dialect is represented in the word list.

As explained by Lewis & Grimes (1995), the people of Tana 'Ai use a ritual language to convey the origins and histories of local clans, as well as during rituals and ceremonies. The ritual language of Tana 'Ai is known to share features such as semantic parallelism and elided syntax with other ritual languages in eastern Indonesia (see Fox 1988, Lewis 1982, 1988a, 1988b), and occasionally pairings of words from semantic dyads of the ritual language include a Tana 'Ai word and a Lamaholot word. Other notable resources on Sika are a sketch grammar written by Arndt (1931), and a study by a native speaker of Sika (Soge 1979) on Sika language structures.

Two consonants that are part of the phonological inventory of Sika are unique within the Flores-Lembata group. The /β/ is a 'lightly fricativised voiced bilabial fricative', and the laminal /ʝ/ is a voiced alveo-palatal affricate' (Lewis & Grimes 1995). The bilabial fricative occurs frequently within the Sika Swadesh list and therefore also features frequently in the database of this study.

4.3. Lewolema (Language B) - K.H. Pampus 1999
The information in the present study for Lewolema is taken from a dictionary created by Pampus (1999). The title of the dictionary makes mention of Lewolema as a dialect of Lamaholot, but for the purposes of this study, it will be treated as a distinct language. The information for the dictionary was recorded by Pampus during the years of 1994-1998. The village in which the fieldwork was carried out is called Belogili-Balukhering, which lies on the eastern portion of Flores island. A more precise location of Belogili-Balukhering is represented as letter 'B' on Map 2 above.

4.4. Lewoingu (Language C) - Nishiyama&Kelen 2008
Nishiyama & Kelen (2008) wrote a grammar of Lamaholot, more specifically the Lewoingu/Lewolaga dialect. Kelen is a native speaker of this dialect, and comes from the village of Leworook. Lamaholot is spoken by 150,000-200,000 people on the islands of Solor, Adonara, Lembata, the eastern portion of Flores, and parts of Pantar and Alor. In the past, Lamaholot has alternatively been known as Solor (or Solorese), including in a description of Lamaholot grammar by Arndt (1937), which incorporates features from several Lamaholot varieties. According to Keraf (1978), there are 33 dialects of Lamaholot, which Keraf groups into three branches, Western, Central, and Eastern. Keraf separates the 33 varieties of Lamaholot into 15 languages, using 80% cognacy as a baseline for classification as such. To the West of the Lamaholot linguistic area Sika is spoken. To the East, the neighboring language is Kedang (Nishiyama & Kelen 2008).

The data from this source was collected in elicitation sessions between Nishiyama and Kelen. Later, the information provided by Kelen was approved of or modified by other speakers from Leworook of varying ages and levels of education, thus perhaps 'standardizing' in a sense the data given by Kelen.

4.5. Solor (Language D) - Klamer 2002
This data was taken from field notes of Klamer (2002), which also contain detailed information about the informant. Klamer refers to the language as the Solor dialect of Lamaholot, and says it is spoken on the eastern part of Solor island. The informant estimates that approximately 12,000 speakers of the Solor dialect exist in this area. The recording sessions took place in Nieuwegein, Netherlands in March of 2002. The informant for these recordings was a male, born in 1969, who had lived in the Solor dialect speaking area for 15 years, until 1984. From 1984 onwards, the

informant lived in various locations on Flores island, where he began school at a Catholic Seminary in Larantuka, in which education was conducted in Bahasa Indonesia. Other places of residence of the informant since then have included Maumere, Bajawa, Timor island (Nenuk and Atambua), and in Nieuwegein in the Netherlands.

4.6. Lamalera (Language E) - Keraf 1978

The data included in Keraf (1978) is comprised of primarily a morphological analysis of Lamalera, coupled with information regarding the historical relations of the Lamaholot languages/dialects. In addition, Keraf (1978) includes appendices which contain Swadesh list comparisons for the 33 purported Lamaholot dialects/varieties.

4.7. Kedang (Language F) - Samely 1991

The information for this paper on Kedang is taken from Samely (1991), which includes a description of Kedang grammar, a sample text, and a Kedang-English glossary which contains the lexemes used for the database in the present study. Ethnologue (Lewis 2009) and Wurm (1983) both list Kedang as a member of the Flores-Lembata branch of the Timor group. The data from Samely (1991) is taken from field notes and audio recordings from the author's stay in the village of Lèuwayang. The Kedang speakers present in the field notes and recordings include native Kedang speakers of both sexes and various ages. In addition, one recording was made with a male informant in a recording studio in Jakarta. Kedang is reported by Samely (1991), in a lexicostatistical analysis, to have a shared cognacy rate with both Lamaholot and Alorese of 61%. The same study showed Lamaholot and Alorese to have a cognacy rate of 70%.

4.8. Alorese (Language G) - Baranusa dialect - Klamer data

The data for Alor is taken from a sketch of the Alor grammar (Klamer 2009) which contains a word list from the Baranusa dialect of Alor. The Alor language is spoken by about 25,000 people in the Alor-Pantar region of Nusa-Tenggara, Indonesia, namely on the islands of Pantar, Alor, Ternate, and Buaya. Klamer, in collecting Alorese data, worked with principally with three speakers, one of which spoke the dialect of Baranusa (West Pantar), and the other two the dialect of Alorkecil (West Alor). The speakers themselves, one man (Alorkecil dialect), and two women, all were 30-35 years of age, and all made use of many Indonesian/Malay borrowings. Though the results given by Klamer represent data from both dialects, the word list that was taken from the source for use in this study is taken from the Baranusa dialect. Klamer was careful to keep the data from the two dialects separate, as the two display notable lexical differences (Klamer 2009:12).

4.8.1 Alorese - Alorkecil dialect - personal notes

Data from the Alorkecil dialect was also considered in the present study. During my time in Alor, I collected a word list from two informants in Alorkecil, and when compared to the data from Klamer on the Baranusa dialect, no important variations that related to the present study were found. Because the word list from Klamer is more complete in regards to the words in the database, I use the Baranusa list for comparison purposes in this study. I mention my review of the Alorkecil material so that the reader is aware that with regards to the Baranusa data presented in this study, what are given as sound correspondences and sound changes should be considered to encompass what is seen in the Alorkecil dialect as well, in spite of the various lexical differences that do exist between the two dialects.

5   Methodology

5.1.   Introduction

The database for this paper was constructed using data that is already available, but had never been coordinated into single source. Much of the material that made this study possible was data that has been collected in recent years, highlighting the importance of fieldwork in the region, and the relative shortage of existing analyzable data. In fact, reliable data is only available for roughly half of the FL languages. Certainly, future study could greatly improve upon anyone's ability to produce work about the Flores-Lembata region.

Another important resource I utilize, that isn't necessarily based in the FL region, is the Austronesian Basic Vocabulary Database (ABVD), which I explained in a bit more detail in section 4.1. I have taken the existing material, and the ABVD, and built upon methods applied by Blust (2008), namely the use of of Swadesh word lists to identify cognates and examine sound correspondences across languages. Blust examined the potential validity of the purported Bima-Sumba subgroup, which is also branch within CMP. In this paper, I will be analyzing the regular sound correspondences I find in the database, with hopes of establishing low-order relationships between the Flores-Lembata languages.

The principal difference between my methodology, and that of previous studies, will be the scale of the relationships I propose. As opposed to the broad brush strokes of previous CMP studies by Blust, Donohue and Grimes, and others, my research focuses on detailed sound correspondences found in limited numbers of languages, taking a bottom-up approach to identifying relationships and shared history. I will make use of the comparative method to identify these sound correspondences, and below I will summarize the process it outlines.

5.2.   The Comparative Method

The comparative method is an indisputably important tool for historical linguistics. Its major principles outline a method for languages to be compared in such a way that it is possible to identify if they have a shared ancestor language (a proto-language), and if so, to recover data from it. The knowledge obtained from the comparative method may be used to identify relations and compare levels of shared history within a given group of languages. In this paper I will be applying the comparative method to the Flores-Lembata language with the goal of obtaining knowledge about its proto-language, Proto-Flores-Lembata (PFL), and identifying its internal divisions.

I will explain briefly now the basic principles under which I will be working in the present study, so that the reader may understand my methods. I am working with Swadesh-style vocabulary lists, the benefits and difficulties of which were discussed in chapter 3, which I will use to identify cognates. I will use the cognate sets I assemble to establish sound correspondences, which then may be used to reconstruct proto-sounds, and understand what level of shared history each FL language has with the other members of the group.

Identifying a set of cognates in the first step of the comparative method, but only true cognates may be used for comparison. Care must be taken so that words that appear similar at first sight but actually are not cognates are not selected as so. Two possibilities for similar looking words not being true cognates are (1) coincidence, and (2) borrowing (Campbell 2004). In the case of coincidence, it is expected that when comparing the vocabularies of various languages, that at some point words will be found in different languages that have very similar phonetic structures. The more space that is given semantically to forge a comparison in phonetically similar words, the more potential false cognates will be identified.

In the case of borrowing, it is expected that genetically related languages as well as genetically unrelated languages borrow lexical items from each other, either directly or by way of another language. Borrowing can be traced to linguistic contact of one form or another, and often results in words that appear very much as cognates, though they are not in the proper sense, which dictates that they must be inherited into languages from a shared ancestor. In fact, it is common for borrowed words to be assimilated into the phonological confines of the receiving language, giving them perhaps the appearance of containing relevant sound changes. However, this assimilation must be identified if the set of words being compared as cognates is to contain solely true cognates whose comparisons may shed light on shared histories of languages and the properties of their shared ancestors or proto-language.

The second step of carrying out the comparative method is to take the set of cognates that have been assembled, and to identify and establish a set of systematic sound correspondences. With these correspondences, it is possible to, based on certain principles and knowledge of trends that occur across languages, reconstruct proto-phonemes from the proto-language. I will shortly explain a few these principles, but first I will introduce the concept of sound change.

5.3. Sound change

When discussing sound changes, it is important to make note of the fundamental types of sound change that exist. The notion of regular sound change has been probably the single most important distinction made in historical linguistics. The assumption that sound change is regular, meaning that a change will recur consistently and uniformly throughout a language when the correct phonetic conditions are present, has been the driving force behind historical linguistics since the latter part of the 19th century, beginning with the 'Junggrammatiker' of Germany (Campbell 2004).

Within the context of regular sound change, an important distinction is made, that is, whether the sound change is conditioned or unconditioned. When a sound change is seen across all phonetic environments, i.e. word initially, word medially, intervocalically, etc, and is not influenced in any manner by neighboring sounds, it can be called an unconditioned change. This basically means the sound change is not conditioned or affected by other neighboring sounds. When a sound change is affected by neighboring sounds, it is referred to as a conditioned change. In sum, unconditioned changes occur in all contexts, and conditioned changes can be much more limited, only occurring in conjunction within certain phonetic environments.

An example is useful to illustrate the concept of conditioned sound changes; the following is taken from Campbell (2004). In Spanish, the original Latin *p* became *b* intervocalically (p > b/V_V), but not in other positions, such as word initially. That is, in originally Latin words which began with *p,* the initial consonant remained as *p,* and in words where p was found intervocalically in Latin, a *b* is found in Spanish. To contrast this example with an example of an unconditioned change from Spanish, I cite another example from Campbell (2004). In many dialects of Latin American Spanish, the palatalized /lʲ/ has changed to a /j/ (lʲ > j) in all phonetic environments. An example of a word containing this unconditioned change is the pronunciation of *calle* 'street' [/kalʲe/ > /kaje/].

5.4. Basic principles of reconstructing sounds

Other than the three principles that will be outlined below, it is important when reconstructing proto-forms and proto-sound systems to (1) make sure a given postulated proto-sounds fits within the overall pattern of the proto-sound system into which it is being constructed, and to (2) check that the proto-sound system being reconstructed fits within broader typological expectations (Campbell 2004).

5.4.1 Directionality

Certain sound changes happen frequently in languages across the world in one direction, but rarely in the opposite direction. For example, a change of *s* > *h* is common and can be expected to be found in large sets of correspondences, but a change of *h* > *s* hardly ever has been shown to occur (Campbell 2004). Knowledge of what sound changes frequently occur, and which rarely occur, can help a researcher when they are faced with a correspondence that could be ambiguous, for instance one in which data is only held for two languages.

5.4.2 Majority wins and economy

When reconstructing proto-forms for a group consisting of several languages, as I did for this study, an important principle to remember is 'majority wins'. If there is a correspondence in which, say, five languages share a *ʤ*, and one shows a *j*, then the majority sound may be picked as the proto-sound. Of course, one should be cautious with this principle (think of directionality) because some sound changes are common enough that in some correspondences, the languages which have retained the proto-sound will find themselves in the minority. This principle is most useful when a correspondence exists in which other factors have been canceled out, i.e. when all else is equal.

As common sense, and the principle of economy, would dictate, a reconstruction which involves fewer and simpler changes should be investigated before postulating a complicated and drawn out series of sound changes. The principle of economy is closely tied to the idea of majority wins, and in some ways, the two principles represent the same idea.

6 The internal divisions of Flores-Lembata

6.1. Introduction

By examining the sound changes found in the database that present patterns where divisions within FL can be seen, I have discovered patterns that demonstrate the nature of these internal divisions. Now, I will explain the method I used to determine closeness between the FL languages. I first compiled sound correspondences from the database and made groups of these correspondences. To highlight an example of one of these correspondences (more explained in 6.2), all instances where we see an intervocalic /r/ in languages ABCDEG, but an intervocalic /j/ in Kedang (language F), have been grouped together as a single sound correspondences for the purposes of the exercise. I made 8 such correspondences (which are referred to later on as C1, C2, C3, and so on), and within them, compared each language to the rest, one correspondence at a time (see Appendix 1 for tables containing these comparisons). Most correspondences (though unfortunately not all) have more than one example item in which the sound correspondence can be seen. I understand it is ideal to have as many examples as possible to demonstrate a sound change beyond doubt, but I am working within the confines of the data available to me at the present time, and I feel that even with one clear example, it becomes possible for a sound change to be identified.

I took these 8 groups and compared them to each other by making group 'family trees', as if each sound change alone was determining the internal divisions within the group. I did this to help me identify which languages share more correspondences with other languages, despite the fact that many of the correspondences overlap. In Figure 2 below there is a picture of the C1 and C6 Correspondence trees; the purpose of Figure 2 is to highlight the different types of splits seen in different correspondences. In the picture, the phoneme from the correspondence is listed below the languages which have it. In the case of C1, only one language, Kedang, separates itself from the group, but in the case of C6, we see four independent groups with different phonemes in the same position in cognates. I would propose that wider trees, such as C6, suggest sound correspondences

that are more recent than skinnier trees like C1. This is of course a general principle, but trees such as C6 display correspondences which have allowed for more time to develop; my reason being so because there are greater amounts of splits (that have already happened) to be seen amongst the languages. If the languages of Flores-Lembata descended from a single shared ancestor, then it stands to reason that they differentiated gradually over time, from then until now, when multiple distinct FL languages exist. Therefore, when we see more complicated correspondences, we are seeing correspondences that are a result of sound changes that took place in greater numbers of languages, which must have occurred later rather than earlier, because later is when we see more distinct FL languages, and earlier is when we see a single proto-language.

Figure 2. C1 and C6 Correspondence trees
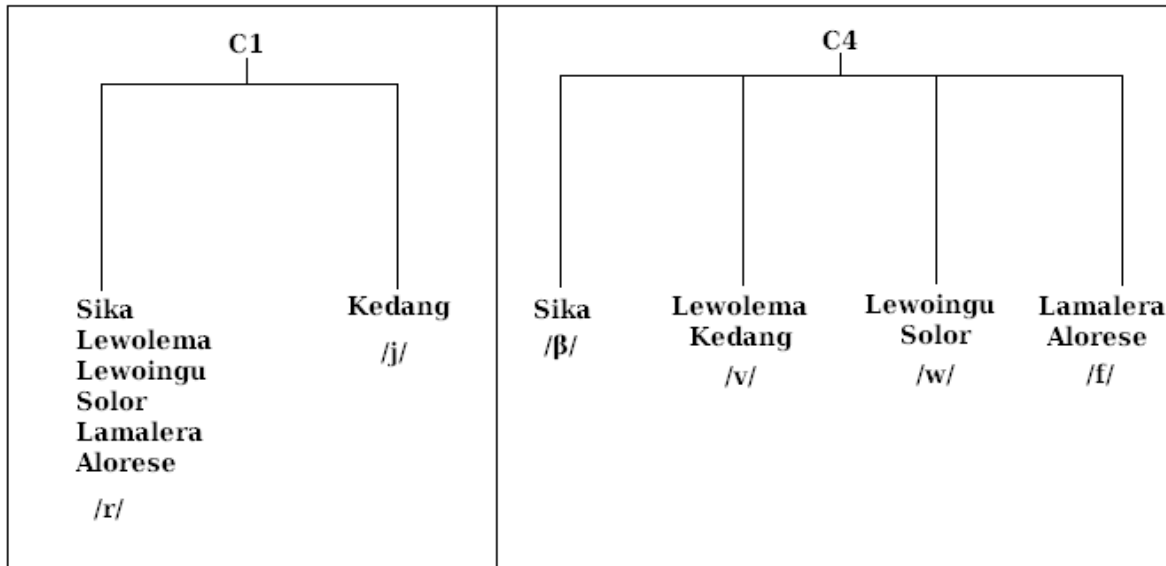
When comparing the trees created by the groups, it is possible to identify which languages are most closely related to each other, in spite of the pattern of overlapping innovations. For instance, Languages C (Lewoingu) and D (Solor) went together in 7 of a possible 8 cases, including in C4, where we see more divergence, meaning that within the database there is only one group of sound changes (apart from vowels) which demonstrates clearly a divergence between them. This suggests that C and D split more recently than Kedang (F) split from the group. Because there is only one correspondence which displays a split between C and D, I hypothesize that they split away very recently, and therefore the sound change that demonstrates their split was the most recent change that can be found in the database. Because C and D are so closely related, they very likely represent the lowest possible level relationship on the FL tree.

On the other end of the spectrum, there is Kedang, which is pictured in C1 apart from the group. Kedang, when compared in this fashion to the other six languages, shows by far the fewest amount of shared correspondences with other FL languages (refer to Appendix 1). This suggests that Kedang split early from the group, and thus has had the most time in which to differentiate itself.

6.1.1 Introduction of tables that displays sound correspondences
I have included several tables in this paper which display the sound correspondences I have found in the database. The tables all follow a basic outline, which I will describe. In these tables, there are a total of 11 languages for which words are given. After English, the next three are Proto-Austronesian, Proto-Malayo-Polynesian, and Proto-Central Malayo-Polynesian. The languages being studied that make up part of the Flores-Lembata group are the following seven. These seven languages are given the letters A through G (not seen in the tables, but found frequently elsewhere,

including Map 2), and will be referred to in the proceeding explanations by either their name, or when that becomes too cumbersome, by their corresponding letter. The rightmost column represents the reconstructed Proto-Flores-Lembata sound I postulate from each sound correspondence. The leftmost column contains the item number for each example. These numbers correspond to the items in the database, so that they may be easily referenced for further examination. Symbols are IPA format.

A note on the PAN, PMP, and PCMP forms: These forms are included for reference mostly. The principle concern of this study is the internal relations within Flores-Lembata. I have tried to select examples from the database which are interesting within FL and cognate with PAN/PMP/PCMP. Their being cognates is important for ensuring that examples are Austronesian words, and not borrowed words from neighboring Papuan languages. This gives me the best chance of reconstructing the history of FL languages correctly. However, in cases where words PAN/PMP/PCMP words do not appear to be cognate or where they contain or represent extra sound changes that have occurred between PCMP and PFL, it must be noted that my primary interest is the relationship between the correspondences in the FL languages. For instance, in C5 below I have included an item (21) which shows a different sound change from PCMP to PFL. I understand that it may be complicated to explain this change, and my including it may cause concern for those who expect that my goal is to connect PFL to PCMP. My goal is to connect FL languages to each other, from the bottom up. When that is accomplished, more research may of course be conducted with the aim of connecting PFL to PCMP. For now, however, I will be investigating solely the internal divisions within Flores-Lembata.

6.2. Sound correspondence 1 (C1)

C1 -- ABCDEG (P1) /r/ -- F /j/

Proposed sound change -- PMP *j > ABCDEG (P1) r

Example demonstrating sound change -- PMP > P1 * ŋajan 'name' > naran (item 95).

Table 2. Examples for C1

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
2	sun				-r-	-r-	-r-	-r-	-r-	-j-	-r-	*j
6	day	*qalejaw	*qalejaw *meñak; *himaʀ;	*qaləjaw	-r-	-r-		-r-	-r-	-j-	-r-	*j
66	fat	*Simaʀ	*miñak	*miñak		-r-		-r-	-r-	-j-		*j
95	name	*ŋajan	*ŋajan *ma-	*ŋajan *ma-	-r-	-r-	-r-	-r-	-r-	-j-	-r-	*j
156	dry	*ma- qaʀiw	*ma- ʀaŋaw	ʀaŋaw		-r-	-r-	-r-	-r-	-j-	-r-	*j
176	white	*ma- puNi	putiq	*burak	*-r-	-r-	-r-	-r-	-r-	-j-	-r-	*j
264	rain	*quzaN	*quzan	*quzan	-r-		-r-		-r-	-j-		*j

Note for tables 2-14: when a space is blank, the corresponding word either did not exist in the database, or was not cognate and left off the table on the basis of its irrelevancy to this exercise.

We know that Kedang was the first to split from the group, because it retains the /j/ found in the PAN, PMP, and PCMP forms. It is unlikely, considering the PAN, PMP, and PCMP knowledge, that six independent shifts of /j/ > /r/ happened. This example is a case of when not to assume proto-forms based on the 'majority rules' principle, which if I had used it would have led me to postulate a PFL /*r/. The knowledge of the PAN, PMP, and PCMP forms thus has aided my reconstruction in

17

this case. To accommodate for the fact that six independent changes did not occur, I believe that instead, Kedang split from languages ABCDEG, who later, prior to diverging themselves, experienced a single sound change of *j > r*. This correspondence shows that languages ABCDEG had yet to differentiate at the point in time when this sound change occurred, meaning also that Kedang was the first to separate from the FL group. So in sum, we know that the shift from /*j/ > /r/ happened early on, after Kedang had parted from the group (or vice versa), but before languages ABCDEG had themselves diverged. I have named this set Correspondence 1 (C1) because I believe it displays the first split within the Flores-Lembata group.

6.3.   Sound correspondence 2 (C2)
C2 -- ABCDEG (P1) /r/ -- F /s/

Proposed sound change -- F *r > s

Example demonstrating sound change -- PFL > F -- *rua 'two' > suɛ (131).

Table 3. Examples for C2

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
131	two	*duSa	*duha	*dua	r-	r-	r-	r–	r–	s-	r–	*r

Another correspondence that shows Kedang had already separated from the other six is C2. In C2, we see a sound change that occurred during the same period as C1. That is, it occurred after Kedang had split off, but before ABCDEG diverged. I used the same line of logic I explained in my discussion of C1 to conclude this.

I point out that this correspondence is unexpected, and in fact only one example showing it was found in the database. It is not within the scope of the present paper to speculate on the shift from PMP to Flores-Lembata *d > ⁺r, my purpose is to identify the correspondences I have seen within the FL group. It appears to me that item 131 is a possible indication of a shift from r > s that occurred in Kedang after it had split from the other FL languages.

C1 and C2 are both examples of sound correspondences that happened after Kedang (F) had split off from the other six, but when all other six languages were still together. C1 and C2 are distinct, however, because the former is an example of a retention by Kedang, and the latter is an example of a change by Kedang, and a retention by ABCDEG. I cannot think of a way, based on the evidence I currently have, to say surely whether C1 came before C2, or whether it was the other way around. The only thing I can say about both of them is that they occurred after Kedang had split from ABCDEG, and that C1 occurred before any of the other six had diverged. We know that ABCDEG had not yet diverged by the time of C1 because the six languages experienced the sound change as one language. I suppose this because it is a much more likely scenario than would be the separation of all six, followed by six independent but identical sound changes. The same cannot be said of C2 because it represents a shift by Kedang, not by ABCDEG. It is possible that the sound change that gives us C2 happened down the line, after ABCDEG had diverged partially or completely. This is why I have termed this example C2, and the other C1. The best I can say for C2 is that the shift occurred after the separation of the modern-day Kedang speakers from the group that once spoke Proto-Flores-Lembata.

Prior to the Kedang separation (and C1 and C2), the language being spoken by this group could be referred to as Proto-Flores-Lembata (PFL), but by the time the sound changes that produced C1 and C2 had happened, PFL had splintered into two language groups, one being Kedang, and the other

we'll call P1 (referring to Proto-ABCDEFG, see table below for details), because it was the proto-language that remained after the first split within Flores-Lembata .

Table 4. Outlining Flores-Lembata Splits

#	Splits from Group	Remaining	Proto Shorthand for Remaining
0	none yet	ABCDEFG	**PFL**
1	**F** from ABCDEFG	ABCDEG	**P1** (or Proto-ABCDEG)
2	**A** from ABCDEG	BCDEG	**P2**
3	**B** from BCDEG	CDEG	**P3** (or Proto-Lamaholot)
4	**EG** from CDEG	CD and EG	**PCD** (or P4) and **PEG** (or P5)
5	**E** from G	**G**	
6	**C** from D	**D**	

With Kedang separated early on, as is further evidenced by its lesser ties in sound correspondences with the other languages, we are down to six languages. I will attempt now to explain when and how these six diverged.

As I will explain below, I believe that Sika (A) was the next language to split from ABCDEG (i.e. the first to split from P1). It can be seen from the correspondences in the database that Sika has less in common with BCDEG than they do with each other, but more in common with them than Kedang. Sound correspondences C3-C4 can be shown to have occurred after Sika had split from ABCDEG (P1), but before they had split from each other.

6.4.   Sound correspondence 3 (C3)
C3 -- A /ʔ/ -- BCDEG (P2) /k/

Proposed sound change -- A *k > ʔ /_u, a, i

Example demonstrating sound change -- P1 > A -- *aku '1sg' > aʔu (108).

Table 5. *Examples for C3*

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
37	wing	*paNid	*kapak	*kapak	k-	k-	k-	k−	k−	(Ø)a-	k-	*k
44	louse	*kuCu	*kutu	*kutu	ʔ−	k-	k-	k−		(Ø)u-	k-	*k
47	tree				ʔ-	k-	k-	k-	k−	(Ø)-ai; or k-	k-	*k
112	1pl excl	*kami	*kami	*k-ami	ʔ-	k-	k-	k−	k−		k-	*k

In C3 we see that Sika has shifted away from the PFL /*k/, while languages BCDEG have maintained the /*k. A shift from /k/ > /ʔ/ is in fact quite common, and therefore I postulate it on the basis of directionality. I say this because a shift of /k/ > /ʔ/ is common across languages, but a shift of /ʔ/ > /k/ is very rare, and would not be expected. Therefore, the most likely scenario is a PFL *k > ʔ in Sika. This also means that the Sika shift occurred after it had split from languages BCDEG.

In addition, the principle of majority wins is another convincing factor in this case. It is far more

likely, even without considering the principle of directionality, that a shift of Sika /*k/ > /ʔ/ occurred one time rather than the shift of /*ʔ/ > /k/ occurring independently, in languages BCDEG. This would mean five shifts took place, rather than one, and that they all happened to be identical; a highly unlikely scenario, the kind one wishes to avoid in historical reconstructions.

Kedang in C2 is a separate case from either Sika or BCDEG, but based on the timing of the splits I discovered in this study, that would be expected. My main concern with regards to C3, however, is the correspondence Sika has to languages BCDEG, and therefore I do not include an analysis at the present time of what has happened in Kedang with regards to C3 (most likely a conditioned shift of *k > Ø).

Most important about C3 is that it took place after the separation of Sika (A) from ABCDEG, but before BCDEG (P2) had diverged. The separation of Sika from P1 means it is more distinct from the modern languages of B, C, D, E, and G than they are from each other, and also that it was the second language to split from the Flores-Lembata group being analyzed in the present study.

6.4.1   More on the case of the P1 (and PFL) /*k/

You may have noticed that I reconstructed *aku '1sg' for P1 (and PFL), while in fact none of the languages A-G in the database have a /k/ in item 108. Item 108 leads me to postulate that the proto-form is /*k/, even though it is not seen in any of the modern languages. Referring back to the principle of directionality discussed above, it is not likely that we would have seen a change from PFL /*g/ to Kedang and Sika /ʔ/. A change of /*k/ > /ʔ/ in Sika, and then later *k > g in BCDEG is much more likely. Confirming this, we see in PCMP that the sound was /*k/. In BCDEG we see that the initial a- was dropped, and the now initial k- became voiced (after probably a vowel shift of some sort), i.e. it became /g/. This is why when this /*k/ is word medial, it is still a /k/ in BCDEG. This can be seen in item 226, which contains for BCDEG an initial g- and an intervocalic -k-, which both have the same origin as PFL /*k/. In fact, Nishiyama & Kelen (2007) say that the /g/ and /k/ are variable between dialects of Lamaholot. I presume this variability means it is possible the dialects are currently experiencing a shift. An example of this speaker variation can be seen in Solor item 226, where the intervocalic stop may either be voiced or unvoiced.

Nishiyama & Kelen (2008:17) also claim that the first singular pronoun in Lamaholot (now /go/ in Lewoingu) once contained [k], which they say is plausible because of the existence of the /*k/ in the PMP 1sg, which became the glottal /ʔ/ in the Sika 1sg.

Further evidence for the reconstruction of a PFL /*k/ is the fact that a /g/ in any position is seen only rarely in Sika, and only once in the database as a cognate with other languages (item 235). In addition to this, item 235 for Sika is a so-called 'reconstructed' form (there are in fact many of these Sika reconstructed items in the Comparative Austronesian Dictionary (Lewis & Grimes 1995)). They are the Sika items preceded by an '*' in the tables/database). Because of the nature of reconstructed items, I approach them more cautiously, and therefore am hesitant to assume or fear that this is a counterexample that jeopardizes my argument for the PFL /*k/.

As for Kedang, none of the primary words for items in the database contain a /g/. According to Samely (1991), /g/ is in fact part of the consonant inventory of modern Kedang, but occurs with extreme infrequency. Specifically, the frequency of occurrence of the consonant phoneme /g/ in Kedang stands at 0.37%, and its appearances are limited to syllable initial position in polysyllabic words.

Items in Sika which contain a /k/, and which are cognate to languages BCDEG, do however exist. Items 37, 125, 143, 169, 228, are 247 are all cases in which the Sika word contains a /k/ in place of

a /ʔ/. Item 37 is included in Table 5 above, but the others are not because they display the same pattern as 37. They may be found in the database for further reference. Something may be said about all the /k/ items from Sika. When we see a word or syllable beginning with a /k/ in Sika, it is always followed by a schwa /ə/ (except, puzzlingly enough, the name Sika). This is why the sound change from C3 is considered a conditioned change, meaning that the correct phonemic environment must occur for it to take place. In this case, the /*k/ only shifted (in examples found in the database) to a /ʔ/ when followed by a /u/, /a/, or /i/, but not when followed by a schwa.

6.5.   Sound correspondence 4 (C4)
C4 -- A /n/ -- BCDEG (P2) /ŋ/

Proposed sound change -- PMP *ŋ > Sika n /V_V

Example demonstrating sound change -- PMP > A -- *haŋin 'wind' > anin (8).

Table 6. Examples for C4

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
8	wind	*bali	*haŋin *laŋuy;	*haŋin	-n-	-ŋ-	-ŋ-	-ŋ-	-ŋ-	-ŋ-	-ŋ-	*ŋ
200	swim	*Naŋuy	*naŋuy	*naŋuy	*-n-	-ŋ-	-ŋ-	–ŋ–	–ŋ–	-ŋ-	–ŋg–	*ŋ

C4 is another correspondence which signals the separation between Sika and languages BCDEG. My reason for postulating a P2 /*ŋ/ is the principle of majority wins. It is far more likely that a change of *ŋ > n would have happened one time in Sika, rather than several times in the other languages.

Another reason I postulate a /*ŋ/ is because of the evidence in PAN, PMP, and PCMP in the database. In item 8, the word-medial consonant is /*ŋ/ in PMP and PCMP, and in item 200 the word-medial consonant is /*ŋ/ in PAN, PMP, and PCMP. The existence of the /*ŋ/ in PAN, PMP, and PCMP signals its link from those three proto-languages to PFL. The FL language where it does not currently exist is Sika (A), which confirms that Sika was the language to undergo the change.

Whereas we can be sure that C3 took place in the not too recent past, owing to the later change from *k > g we can see that occurred later on in BCDEG (presumably before those five diverged), we cannot be sure of exactly when C4 took place. It is possible even that it was a very recent change. This is because the change that brought about C4 was a change by Sika, not P2 (Proto-BCDEG). If it had been the opposite, as in C3, then we would know that the change would have had to happen after the split of A from BCDEG, but before the divergence of Proto-BCDEG (P2), because the result would have   The only thing I can say for sure was that it took place after the split of Sika from languages BCDEG, which left us with P2.

As a side development unrelated to the present focus, I wish to note in item 200 a case of strengthening in Alorese, for which I postulate the creation of a consonant cluster in Alorese as the type of strengthening seen in this sound change. The other languages all have a single consonant in this position, therefore the principle of majority wins also lead me to postulate this sound change. Further investigation of this shift in Alorese has not been investigated because it is not presently relevant, though it could be something useful for future study.

6.6.   Sound correspondence 5 (C5)

C5 -- B /r/ -- CDEG (P3) /l/

Proposed sound change -- B *l > r /#_

Example demonstrating sound change -- P2 > B *ləra 'day' > rəro (6).
Table 7. Examples for C5

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
2	sun				l-	r-	l-	l–	l–	l-	l–	*l
6	day	*qalejaw	*qalejaw	*qaləjaw	l-	r-		l–	l–	l-	l–	*l
21	road		*zalan	*zalan	l-	r-	l-	l–	l–			*l

C5 is good evidence of a split that occurred after B broke away from CDEG, though because B experienced the sound change, I cannot be sure if it happened before or after the eventual split of CDEG. In spite of this, I can see that B split by itself from CDEG before CDEG diverged to become C, D, E, and G.

I realize that items 2 and 6 are basically the same word, and both most likely descend from the same Malay lexical item, but I have included them separately here because they are in fact different words in the modern languages. The point of this correspondence (C5) is to demonstrate the divergence between Lewolema (B) and the rest of the group. This is also why item 21 is included. If I were arguing the change of the PMP or PCMP sounds into the FL sounds, then item 21 would represent a different sound change than items 2 and 6, are therefore not be appropriately placed in table 7. However, in my opinion item 21 should be in table 7 because it is another example of the l > r correspondence in the modern languages. This change appears to be quite recent, and the time gap between it and any shifts from PCMP into PFL is thus large enough to contain changes which

resulted in what appears to be a merger of the PCMP *l and *r into a PFL ⁺l. Because of the nature of its distribution, we know this correspondence has occurred after a split of Lewolema from languages CDEG.

For PFL, I postulate a /*l/ in this case because of the principle of majority wins, and because of its appearance in PAN, PMP, and PCMP in item 6. In fact, in items 2 and 6, where C5 can be seen, Kedang, which was the first to split from the group, and Sika (also in item 21), which was the second language to do so, both have cognates that contain a word-initial /l/. Thus, I can be quite confident based on the principle of majority wins alone that the PFL sound was /*l/, and that Lewolema underwent a change of *l > r /#_ after splitting from what remained of the group.

For the above stated reasons, I reconstruct a /*l/ for PFL, and state that C5 is evidence that Lewolema separated from CDEG before it underwent the change from *l > r /#_.

6.7.   Sound correspondence 6 (C6)

C6 -- A /β/ -- B /v/ -- CD /w/ -- EG /f/ -- F /v/, Ø, /b/

Initial sound change -- PMP *w, *b > PFL ⁺w
Proposed sound change --   A *w > β -- B *w > v -- EG *w > f
                          F *w > v /#_V; *w > Ø /#_>V

Example demonstrating sound change -- *batu 'stone' > βatu (A); > vato (B); > wato (CD)
> fato (EG); > vaʔ (F) (15).
*bulan 'moon' > >ula (F) (3).

*Table 8. PCMP *b examples for C6*

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
3	moon	*bulaN, *qiNas	*bulan	*bulan	β-	v-	w-	w–		(Ø)>u-	f-	*w
15	stone	*batu	*batu	*batu	β-	v-	w-	w–	f–	v- a	f-	*w
51	fruit	*buaq	*buaq	*buaq	β-	v-	w-	(Ø)u–	f–	(Ø)>u-		*w
79	mouth	*ŋusu	*baqbaq	*babaq		v-	w-	w–	f–	v- o	f-	*w
145	new (house)	*ma-baqeʀu	*ma-baqeʀu	*baqəʀu *bəqəʀu	β-	v-	w-	w–	f–	v- u	f-	*w
98	woman	*bahi	*bahi	*bai, *b-in-ay	β-	(-)v-	-w-	w–	–f–		–f–	*w

*Table 9. PCMP *w examples for C6*

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
123	left	*ka-wiʀi	*ka-wiʀi	*ka-wiʀi	β-					v-		*w
124	right (side)	*ka-wanaN	*ka-wanan	*ka-wanan	β-	v-	w-	w–	f–	v-		*w
285	water (fresh)	*daNum	*wahiʀ	*waiʀ	β-		w-	w–	f–	v-		*w

The examples from Tables 8 and 9 demonstrate that in the cases of the PMP and PCMP *b and *w, the consonant correspondences seen within the FL group are the same. This suggests that there was a merger of PMP *b and *w in PFL, meaning the resulting correspondences all descend from the same proto-form, the PFL *w. I believe the PMP *b and *w merged into a *w for two reasons. First, it is more likely that we see a case of lenition, or weakening in the PMP *b, because if it were the other way, we would have a scenario which involves strengthening followed by lenition (for instance: PMP *w > PFL ⁺b > Solor w), which in unlikely. Secondly, we see two of the languages, Lewoingu and Solor, have a /w/ in the correspondence. It seems more likely that these are cases of retention rather than a shift of PMP *w > PFL ⁺b > CD w. It is much more reasonable to postulate zero sound changes in place of two sound changes. For the above reasons I postulate a PFL *w for C6.

Flatter trees of correspondence are newer, because they show a sound correspondence that is the result of a sound change that occurred after greater degrees of divergence within a group (see Figure 2). C6 is the best example in the database of this type of correspondence. We have already seen in C1-C5 examples of correspondences that occurred after Kedang (C1-C2), Sika (C3-C4), and Lewolema (C5) had split away from the group of PFL speakers. In C1-C5, however, CDEG have remained together throughout (CDEG could be part of what may be called Proto-Lamaholot). C6 is the first correspondence in which we see a division within the final four languages that had remained together to this point. The difference between C6 and C1-C5, however, is that in it we see a group of four languages (CDEG) split into two groups, each containing what will become two languages.

In C6, Sika is alone in its correspondence, but that is not a surprise because Sika had already split from the group by the time the sound change(s) that resulted in C6 had taken place. I hypothesize

that the sound change seen in Sika of *w > β is a result of fricativization, based on the principle of directionality. Fricativization is said by Campbell (2004) to be a relatively common sound change, and it appears highly likely that a change from PFL to Sika of *w > β occurred. The timing of this specific change cannot be exactly estimated, however I can say that it occurred after Sika had separated from BCDEG. I cannot be sure at what point in time after that event, though, because theoretically it is possible that after splitting from the group, Sika maintained the PFL *w until only very recently while languages such as EG began to move away from the PFL *w.

Lewolema and Kedang appear at first to group together in this sound correspondence, but the patterns of sounds occurring in the items used in C6 are in fact slightly different, which suggest that their individual changes happened independently. This would be expected, as Kedang was the first language to split from the FL languages, meaning it did not share a period exclusivity with Lewolema. The principle difference between Lewolema and Kedang in C6 is that the sound change for Lewolema is unconditioned, and the Kedang sound change is conditioned, meaning that in Kedang certain phonemic conditions produce different results in the correspondence. In Lewolema, each instance of PFL /*w/ has changed into a /v/ (*w > v), whereas in Kedang this is true only when the proceeding vowel is not a breathy vowel variant (represented by the symbol: >) (*w > v /_V; and *w > Ø /#_>V).

The Kedang vowel inventory can be divided into two sets of six (Samely 1991). The six vowels that each set contains correspond to each other, but one set could be described as a breathy variant of the other. These Kedang breathy vowels do not occur word medially or finally, as do their non-breathy counterparts, but only word-initially, as do their non-breathy counterparts as well. When the first vowel of a word in the database that is part of C6 is a breathy variant, apparently its restriction to word-initial position takes ultimate precedence and the consonant (in these cases PFL *w) is deleted. This describes what has taken place in Kedang examples for C6 which contain breathy vowels. Those examples in Kedang for C6 which do not contain breathy vowels appear with a /v/ in word-initial position.

It is the conditioned nature of the Kedang sound change seen in C6 which leads me to maintain my belief that the changes that occurred in C6 in Lewolema and Kedang are separate. Speaking of them together but as separate changes, I postulate a change for both (keeping the fact of their independent nature in mind) of *w > v on the principle of directionality.

In C6 we see a correspondence that appears to be the first that is a result of a change that occurred after the split of languages CDEG. I believe that CDEG (P3) had split into two groups, CD and EG, by the time this change had happened. After that shift had occurred is when P3 split into PCD and PEG. After CD and EG split, a shift from *w > f occurred in EG. This change is likely because of the continuing process of lenition seen in C6. In this case the voiced consonant *w became unvoiced, and became a /f/ in EG.

The split of CD and EG, therefore, happened relatively late because of the other splits seen before it. C6 is a good representation of the all the splits that had already occurred, and it is the earliest correspondence in the database that represents a split between CD and EG. Because the earliest correspondence showing EG and CD had split also contains correspondences demonstrating the split of A, B, and F (but not vice versa), I can say that the split of EG from CD happened after A, B, and F separated from the PFL group. Because in C6 E groups with G and C groups with D, I can say that EG and CD represent the lowest level pairings within the Flores-Lembata subgroup.

6.8. Sound correspondence 7 (C7)

C7 -- E /Ø/ -- G /h/

Proposed sound change -- E *h > Ø

Example demonstrating sound change -- PEG > E -- *aho 'dog' > ao (32).

Table 10. Examples for C7

#	English	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
32	dog	-h-	-h-	-h-	–h–	a–(Ø)–o	>a-(Ø)-u	–h–	*h
53	flower	-h-	-h-	-h-		'pu–(Ø)	-h-		*h
153	near		-h-		–h–	–a(Ø)e		–h–	*h

In C7 the words for PAN, PMP, and PCMP are not included in Table 10 because they do not contain sounds relevant to the correspondence. The point of C7 is to show the divergence of Lamalera from Alorese. What happened before regarding sound shifts from PMP > PFL is beyond the present focus of this exercise, and I do not make attempt to make any claims regarding it. The change seen in C7 occurred relatively recently because it represents a later split within the FL group. The time gap between the change seen in C7 and any changes between PMP and PFL is great enough so that I will not be attempting a review of it in this paper. I am concerned here with the shift of PFL *h > Lamalera Ø, and Table 10 is designed to display that.

In C7 we see that the PFL /*h/ was dropped in Lamalera (E), but retained in Alor (G), as it has been in all the other languages. Languages E and G share all of the sound correspondences C1-C6, so not only can we say the two languages very closely related to each other, their split occurred relatively recently, compared to when the other languages of the database split. Because the difference we see in C7 from E is not seen in other languages, we know that it happened after EG had split away from CD, and in fact after E and G had split. This means the split we see in C7 is a split taking us to the lowest level of the Flores-Lembata tree.

As for the change itself, it is quite straightforward to presume, based on the presence of the /h/ in the cognates of the other languages of the database, and based on the principle of directionality, that the shift seen in C7 is a change from *h > Ø, rather than the opposite. In fact it is much more common across languages to see a shift of h > Ø, while it is rare to see a shift of Ø > h. For those reasons I postulate a PFL (and PEG) *h.

I can say that the EG split came before the CD split (explained in C8) because in the database when looking at C7, it can be seen that while EG have split, C and D are still sharing a sound in the correspondence. In C8, C and D are shown to have already split, but so are E and G. This means at the time of the EG split, C and D were still unified, but at the time of the CD split, E and G were separated. Logic then dictates that the change that resulted in C7 happened before the change that resulted in C8, meaning E and G split before C and D split.

To what degree I can give a timetable for any of the changes seen in the present study is difficult to say, but it is worth noting that Klamer (2010) also claims that Alorese split off from Lamaholot about 600-700 years ago and headed East, to where the speakers are found today.

6.9. Sound correspondence 8 (C8)
C8 -- C /n/ -- D /Ø/

Proposed sound change -- PCD *Vn > D Ṽ (note: previous symbol represents 'nasalised vowel')

Example demonstrating sound change -- PCD > D ikan 'fish' > ikã̃ (35).

Table 11. Examples for C8 (for more see database)

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
3	moon	* bulaN, *qiNas	*bulan	*bulan	-ŋ	-ã(Ø)	-n	−a(Ø) −ã(Ø)		-a(Ø)	−ŋ	*n
21	road		*zalan	*zalan	-ŋ	-ã(Ø)	-n	/ −aŋ	-ã(Ø)			*n
35	fish	*Sikan	*hikan	*hikan	-ŋ	-ã(Ø)	-n	−ã(Ø)	-ã(Ø)	-a(Ø)	−ŋ	*n
49	leaf	*biʀaq	*dahun	*daun			-n	−ã(Ø)				*n
95	name	*ŋajan	*ŋajan	*ŋajan	*-ŋ	-ã(Ø)	-n	−ã(Ø)	−ŋ	−a(Ø)	−ŋ	*n
124	right (side)	*ka– wanaN	*ka- wanan	*ka- wanan	−a(Ø)	-ã(Ø)	-n	−a(Ø)	−a(Ø)	−a(Ø)		*n

Because Lewoingu (C) and Solor (D) appear to be the two languages of this study that are mostclosely related to each other, most likely any items that display divergence between the two are the product of relatively recent sound changes. There are various examples in the database which show a divergence between Lewoingu and Solor in consonants of word initial and word medial position, but they are all singular examples that do not correspond well with each other. Because these divergences are not well-represented enough in the database for me to claim anything of substance from them, I investigated the word-final correspondences from the database. C8 is a word-final correspondence, but it demonstrates a clear difference between Lewoingu and Solor, and it occurs with great frequency in the database.

The sound correspondence I refer to is, as is outlined above: Lewoingu *-n*; Solor *-(Ø)*. Speaking to the frequency of occurrence of C8, it should be noted that the close nature and recent divergence of Lewoingu and Solor causes more cognates to appear in the database between them than an average pairing of FL languages, which leads to more available examples for comparison. In spite of this, it is still interesting to note that C8 is by far the most common consonant correspondence found in the database. I understand there are limitations, especially in this part of the Austronesian world, with regards to using word final correspondences, but in this case we see that over 10% percent of the items in the database consistently display the correspondence seen in C8.

There are only two exceptions to the rule found in C8, in contrast to the over thirty items following the pattern. These are items 162 and 259, which both have Lewoingu Solor cognates that both contain a final *-n*, rather than a Ø in Solor. It should also be noted that three other items have a word in Solor (D) has a final *-n*. Unfortunately, for two of these, items 146 and 205, there are no items present for comparison in Lewoingu, and in the third, item 233, the Lewoingu word in the database is not cognate with the Solor word. These items therefore are not valid examples for the present study. These five Solor words certainly are the exception to the rule, however, with more than thirty other items providing direct correspondence of cognates with a final *-n* in Lewoingu, and a final *-Ø* in Solor. I believe that this correspondence is the best evidence present in the database, and sufficient enough evidence, to make a case for the timing of the split between the Lewoingu and Solor.

The lack of a consistently occurring final -*n* in the other six languages caused me to at first think that the sound change that occurred was an innovation into Lewoingu, which would have been the addition of a final -*n*. However, the unlikelihood of this change in terms of the principle of directionality, and the presence of this final -*n* in many cognates in PAN, PMP, and PCMP, makes this possibility quite dubious. I say that because in my original scenario a change in Lewoingu of **n > Ø > n* would have occurred (meaning all the other languages would have lost the final -*n* at some point as well), a sequence that seems unlikely to say the least when compared to the possibility that Lewoingu simply retained the final -*n*, while the other languages lost it.

A change of n > Ø is in fact much more common and expected than a change of Ø > n, which is one reason to believe C8 demonstrates a Lewoingu retention. Another is that certain morphological processes, specifically various strategies of suffixation (i.e. marking of the genitive) employed by the FL languages, may have played an important role in the shifts away from final -*n* seen in the other six languages. Of all the sound changes I have postulated in the present study, C8 is the most difficult to reconstruct with confidence. However, the presence of the PAN, PMP, and PCMP final -*n* in cognates, and the greater likelihood of a final -*n* retention rather than creation, makes me believe that Lewoingu is simply the final FL language to retain the PFL **n* in word-final position. Solor is presumably the latest language to lose the PFL **n*, which happened after a split between Lewoingu and Solor. Because Lewoingu and Solor pair in C1-C7, I believe that they were the last to two FL languages to split apart.

The Lewoingu Solor split is thus the final split to occur in Flores-Lembata languages, and is what brought the FL branch to its current structure.

6.10.   Other correspondences

6.10.1 PFL **s > z > r* --   PFL **s > h*

Table 12. PFL *s > ⁺z > r -- PFL *s > h

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
246	rub				-r-	-s-	–r–	–r–	-r-	–h–	*s

An /*s/ is reconstructed in in PFL for reasons of directionality. Different sound changes have occurred in this example, because multiple resulting sounds are seen. In the Alorese case, the principle of directionality is key, *s > h* is a common change, but the reverse is very rare. A change directly from *s > r* is not likely, but because the /s/ is intervocalic it would be quite normal for the PFL **s* to have become a *z* at one point. A common change, which subsequently happened, is *z > r*.

Therefore, by way of *z*, we can see in languages BDEF, the change of PFL **s > ⁺z > r*.

6.10.2 PFL **j > ʤ*

Table 13. PFL *j > ʤ

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese	Proto-sound
47	tree				–(Ø)	–ʤ–	-j-	–j–	- ɟ -	–(Ø)	–ʤ–	*j
127	many					–ʤ–	-j-	–j–	–j–			*j
178	green	*mataq	*mataq	*mataq		–ʤ–		–j–				*j

The items from Table 13 above demonstrate a case of strengthening in Lewolema and Alorese. Strengthening refers to a change in which the resulting consonant is in one way or another articulated in a 'stronger' manner. We have already seen in C1 the reconstruction of a PFL **j, and

therefore, for reasons of majority wins and economy, I have postulated a PFL *j for these items. Additionally, we can see in C4 another case of strengthening in Alorese.

6.11. Metathesis in Flores-Lembata

Table 14. Metathesis in Flores-Lembata

#	English	PAN	PMP	PCMP	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese
5	night	*beʀŋi	*beʀŋi	*bəŋin			rəman		'remã		ma'reŋ
33	rat	*labaw	*labaw	*labaw		kərome	kərome				ka'more
56	betel nut					vuaʔ		'ʔuwa			'ʔufa
68	bone	*CuqelaN	*tuqelan	*zuʀi	luri-ŋ	riʔuk	ri'uk	ri'ʔuk	'riuk	lurin	ru'ʔiŋ
241	to spit		*luzaq	*zulaq							

In the database there are four examples of potential metathesis that have occurred within the Flores-Lembata group. I included item 241 as well, in spite of the lack of cognacy it shares with FL languages. I will not attempt to analyze it, it is included only to highlight a potential display of metathesis occurring farther back in Malayo-Polynesian languages.

(1)   reman -- mareŋ

The is a difference between the Lewoingu and Lamalera words and the Alorese word (reman > mareŋ). There has been a transposition of the /re/ and /ma/. In this case the syllables metathesized as whole parts, which is slightly different from the other examples below.

(2)   kərome -- kamore

In the case of (2) the transposition only occurs in the consonants, rather than a transposition of the entire syllable, as was seen above in (1). Additionally, the metathesis occurs word-internally, which may have some influence to why the vowels did not transpose along with the consonants.

(3)   vuaʔ -- ʔufa -- ʔuwa

In this example of metathesis (3), we see that the labiodental consonant has shifted from intervocalic position to word-initial position. When in intervocalic position, the consonant appears both in a voiced (Solor) and unvoiced (Alorese) variant.

(4)   riʔuk > ruʔiŋ

This metathesis example (4) is unique from the other examples because it is an example where metathesis of the vowels is seen, but where the consonants do not change positions.

The four cases of metathesis are the only three I identified in the database, and Alorese is the only language to participate in all four examples. The four cases present a different types of metathesis, which suggests that perhaps many more examples may be found in Flores-Lembata with further investigation.

7 Results and conclusions

7.1. Proto-Flores-Lembata consonant inventory

Table 15 below displays the Proto-Flores-Lembata consonant inventory, as reconstructed based on sound correspondences identified in the database.

Table 15. Proto-Flores-Lembata Consonant Inventory

PFL CONSONANTS	Bilabial		Alveolar		Palatal	Velar		Glottal
Plosive	*p	*b	*t	*d		*k	*g	*ʔ
Nasal		*m		*n			*ŋ	
Trill				*r				
Fricative			*s					*h
Approximate		*w			*j			
Lateral Approximate				*l				

The consonant inventory appears to be uncontroversial, with no real unexpected consonants, or places or manners of articulation. Because of this, I would not be surprised, if with further study, one, or maybe two more consonants may appear in reconstructions. This study does not make an investigation of Flores-Lembata vowels, so I cannot offer a guess as to what the PFL vowel inventory may be. The vowel information exists in the database, and could be analyzed at a later date.

It is natural that working with small word lists will produce only common and expected consonants in reconstructions, because the chances are fewer to encounter rare phonemes in lexical items. For instance, I reconstruct a *g in PFL because of the likelihood of its existence, despite the fact that it was not the reconstructed sound in any of correspondences from the database. I reconstruct *g because it appears in the phonological systems of all seven modern languages, and also because not having it would be a curious empty spot in the PFL system. This reasoning is upheld by principles outlined by Campbell (2004), explained briefly in section 5.4.

For the unanimous consonant correspondences found in the database that complete the proto-sound system, please refer to Appendix 3. The phoneme present in each unanimous sound correspondence has been reconstructed as the PFL sound, and for that reason these correspondences are not discussed further in this paper.

7.2. Flores-Lembata internal divisions

Below stands Figure 3, which contains the Flores-Lembata tree. At the top is Proto-Flores-Lembata, and below are its daughter languages, all seven that were analyzed as a part of this study. I do not make any claims for the rest of the languages currently classified as FL. The lack of materials available fore the other languages precluded their inclusion in the present study.

Figure 3. Flores-Lembata Tree
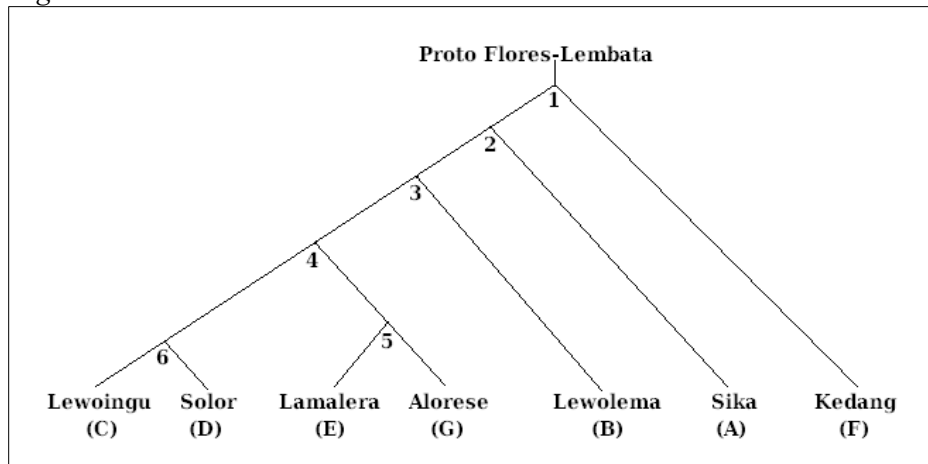
The splits I will summarize below are numbered above in the order I believe they occurred. Split 1, for instance, was the separation of Kedang from the other six languages. Split 2 was the separation of Sika from BCDEG, and so on. When a language separated the group, its split is represented by a divergence from the main line proceeding from the top of the tree to its bottom-left, with the exception of Split 5, which represents E splitting off from EG, which occurred after EG had split off from CDEG.

As can be seen in FL tree, the languages are positioned to represent their earlier or later divergence from the group and each other. To recap the splits that were outlined in Chapter 6, I will explain the tree briefly from top to bottom. The earlier splits are represented higher up on the tree, meaning Kedang was the first FL language to split from PFL. The reasons for this argument are outlined in sections 6.2-6.3, which cover C1 and C2. Split 2 was explained with C3 and C4 (Sections 6.4-6.5), where it can be seen that Sika was the next language to separate. Sika, and Kedang to an even greater extent, are the two languages which have the least in common with the rest of the group. They are the languages at the geographic 'borders' of the Flores-Lembata area, and, I would argue, they stand at the genealogical borders as well.

Lewolema is much more closely related to the languages from the left side of the tree than Kedang or Sika, but it still was the next to split off, after Sika. C5 shows an innovation from *l > r* in Lewolema, one which none of the remaining languages share, and which happened after Lewolema separated.

After the separation of Lewolema, the four languages remaining at the left side of the tree most likely were part of what may be called Proto-Lamaholot. They are the four most closely related languages within FL (with respect to each other), based on the frequency of shared innovations between them. The frequency of those shared innovations also suggests they diverged much later than did the other three languages in the FL group. C6 demonstrates the first divide within the Proto-Lamaholot group, with EG splitting from the CD, which left two languages, Proto-EG and Proto-CD. The split of E and G is classified as Split 5 because in the correspondence that displays it, C7, Lewoingu and Solor remain together. In C8, where Lewoingu and Solor are shown to have split, Lamalera and Alorese had also split, suggesting that E and G split before C and D.

7.3. Implications for future study and limitations of the present study

The present study focus principally on consonant correspondences seen in seven Flores-Lembata languages, and used them to make conclusions about the shared innovations and history of the group as a whole. Future study could of course make use of vowel correspondences to confirm or deny findings from this study related to the structure and timeline of the internal divisions of Flores-Lembata. Future study on this subject may also lead to profitable attempts at reconstruction of PFL lexical items. With additional study and reconstruction of the vowel system of PFL, a short list of PFL lexical items could also be created without too much trouble. Additionally, an attempt to include morphological data could be added to give us a more complete view of what may have happened in Flores-Lembata history.

The present study was limited by the lack of resources presently available for the FL group. I simply could not find or did not have access to information from many of the FL languages, which severely limited the potential scope of my study. As time goes on, hopefully more fieldwork will be carried out in the FL area, which should lead to more analyzable data, and thereby a more complete analysis of the internal divisions of the Flores-Lembata group. I hope that this paper offers a sound beginning to the exploration of the history of this interesting, complex, and poorly understood region, and that its implications and results may offer clues and ideas to other researchers in related areas as well.

References

Adelaar, Alexander. 2005. The Austronesian languages of Asia and Madagascar: A historical perspective. In: Alexander Adelaar and Nikolaus P. Himmelmann, *The Austronesian languages of Asia and Madagascar.* London/New York: Routledge.

Arndt, Paul. 1931. *Grammatik der Sika-Sprache.* Ende, Flores: Arnoldus.

Bellwood, Peter S., James J. Fox, and Darrell T. Tryon. 1995. The Austronesians: historical and comparative perspectives. Canberra, Australia: ANU E Press.

Blust, Robert. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241-93.

Blust, Robert. 2008. Is there a Bima-Sumba Subgroup? *Oceanic Linguistics* 40:45-113.

Blust, Robert. 2009. The position of the languages of eastern Indonesia: A reply to Donohue and Grimes. *Oceanic Linguistics* 48:36-77.

Campbell, Lyle. 2004. *Historical Linguistics: an introduction.* Cambridge, Massachusetts: MIT Press

Donohue, Mark. 2007. Word order in Austronesian from north to south and west to east. *Linguistic Typology.* 11:349-391.

Donohue, Mark, and Charles E. Grimes. 2008. Yet more on the position of the languages of eastern Indonesia and East Timor. *Oceanic Linguistics* 47:114-58.

Fox, James J., (ed.). 1988. *To speak in pairs: essays on the ritual languages of Eastern Indonesia* (CUP).

Greenhill, S.J., Blust. R, & Gray, R.D. (2008). The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexinomics. *Evolutionary Bioinformatics*, 4:271-283. Online: http://language.psy.auckland.ac.nz/austronesian/

Haspelmath, Martin , Matthew S. Dryer, David Gil, & Bernard Comrie (eds.) (2005). *The World Atlas of Language Structures.* Oxford: Oxford University Press.

Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In: Alexander Adelaar and Nikolaus P. Himmelmann, *The Austronesian languages of Asia and Madagascar.* London/New York: Routledge.

Keraf, Gregorius. 1978. *Morfologi Dialek Lamalera.* PhD dissertation, Universitas Indonesia. Ende, Flores: Arnoldus.

Klamer & Ewing. n.d. The languages of East Nusantara: an introduction. In *Typological and areal analyses: Contributions from East Nusantara.* Canberra: Pacific Linguistics (In press).

Klamer, Marian. 2009. Alorese: An Austronesian language with a Papuan substrate. Draft: Leiden.

Klamer, Marian. 2010. Lamaholot to Alorese: Simplification without Shift. Draft: Leiden.

Lewis, E. Douglas. 1982. "The metaphorical expression of gender and dual classification in Tana Ai ritual language", *Canberra Anthropology* 5/1: 47-59.

Lewis, E. Douglas. 1988a. *People of the source: the social and ceremonial order of Tana Wai Brama on Flores* (VKI 135) (Dordrecht: Foris).

Lewis, E. Douglas. 1988b. "A quest for the source: the ontogenesis of a creation myth of the Ata Tana Ai", in: Fox (ed.), 246-281.

Lewis, E. Douglas and Charles E. Grimes 1995. Sika. In Comparative Austronesian dictionary, ed. by Darrell Tryon, 601-609. Berlin: Mouton de Gruyter.

Lewis, M. Paul (ed.). 2009. Ethnologue:Languages of the World, Sixteenth edition. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/.

Nishiyama, Kunio and Herman Kelen. 2007. *A grammar of Lamaholot, Eastern Indonesia. The Morphology and Syntax of the Lewoingu Dialect*. [Languages of the World/Materials 467]. Muenchen: Lincom Europa.

Pampus, Karl-Heinz. 1999. Koda Kiwa: Dreisprachiges Worterbuch des Lamaholot (Dialekt von Lewolema). Abhandlungen fur die Kunde des Morgenlandes, no. 52.4. Stuttgart: Franz Steiner Verlag.

Ross, Malcolm. 1995. Some current issues in Austronesian linguistics. In *Comparative Austronesian dictionary*, ed. by Darrell Tryon, 45-120. Berlin: Mouton de Gruyter.

Samely, Ursula. 1991. *Kedang (Eastern Indonesia), Some Aspects of its Grammar*. Hamburg: Helmut Buske Verlag.

Soge, Paulinus. 1979. The structure of Sikka (Thesis, Department of English Language and Liturature, Sanata Dharma Teachers' Training Institute, Yogyakarta).

Stokhof, W.A.L. 1975. *Preliminary notes on the Alor and Pantar languages (East Indonesia)*. Canberra: Pacific Linguistics.

Tryon, Darrell T. (ed.). 1995. *Comparative Austronesian Dictionary*. Berlin: Mouton de Gruyter.

Wikipedia article: http://en.wikipedia.org/wiki/Austronesian_languages. Retrieved: 07 July 2010.

Wurm, S. A. and S. Hattori (eds). 1983. *Language Atlas of the Pacific Area. Part 2. Japan Area, Taiwan (Formosa), Philippines, Mainland and Insular South East Asia*. Pacific Linguistics, Series C, No. 67. Canberra: The Australian Academy of the Humanities.

Appendix 1. Comparison of shared correspondences

A Sika

	A	B	C	D	E	F	G
C1	**A**	B	C	D	E		G
C2	**A**	B	C	D	E		G
C3	**A**						
C4	**A**						
C5	**A**		C	D	E	F	G
C6	**A**						
C7	**A**	B	C	D			G
C8	**A**						G
		3	4	4	3	1	5 **20**

D Solor

	A	B	C	D	E	F	G
C1	A	B	C	**D**	E		G
C2	A	B	C	**D**	E		G
C3		B	C	**D**	E		G
C4		B	C	**D**	E	F	G
C5	A		C	**D**	E	F	G
C6			C	**D**			
C7	A	B	C	**D**			G
C8		B		**D**	E		
	4	6	7		6	2	6 **31**

G Alorese

	A	B	C	D	E	F	G
C1	A	B	C	D	E		**G**
C2	A	B	C	D	E		**G**
C3		B	C	D	E		**G**
C4		B	C	D	E	F	**G**
C5	A		C	D	E	F	**G**
C6					E		**G**
C7	A	B	C	D			**G**
C8	A						**G**
	5	8	6	6	6	2	**33**

B Lewolema

	A	B	C	D	E	F	G
C1	A	**B**	C	D	E		G
C2	A	**B**	C	D	E		G
C3		**B**	C	D	E		G
C4		**B**	C	D	E	F	G
C5		**B**					
C6		**B**				F	
C7	A	**B**	C	D			G
C8		**B**		D	E		
	3		5	6	5	2	5 **26**

E Lamalera

	A	B	C	D	E	F	G
C1	A	B	C	D	**E**		G
C2	A	B	C	D	**E**		G
C3		B	C	D	**E**		G
C4		B	C	D	**E**	F	G
C5	A		C	D	**E**	F	G
C6					**E**		G
C7					**E**	F	
C8		B		D	**E**		
	3	5	5	6		3	6 **28**

C Lewoingu

	A	B	C	D	E	F	G
C1	A	B	**C**	D	E		G
C2	A	B	**C**	D	E		G
C3		B	**C**	D	E		G
C4		B	**C**	D	E	F	G
C5	A		**C**	D	E	F	G
C6			**C**	D			
C7	A	B	**C**	D			G
C8			**C**				
	4	5		7	5	2	6 **29**

F Kedang

	A	B	C	D	E	F	G
C1						**F**	
C2						**F**	
C3						**F**	
C4		B	C	D	E	**F**	G
C5	A		C	D	E	**F**	G
C6		B				**F**	
C7					E	**F**	
C8						**F**	
	1	2	2	2	3		2 **12**

Appendix 2. Word lists from Flores-Lembata languages

Column sources: PAn/PMP/PCMP — ABVD online; A Sika — IPA; * = reconstructed form; *[ ] = ungrammatical; Lewis & Grimes 1995 (in: Tryon 1995); B Lewolema — IPA, Pampus (1999); C Lewoingu — IPA, Nishiyama and Kelen (2008); D Solor — IPA, Klamer (2002); E Lamalera — IPA, Keraf (1978); F Kedang — IPA; >V = breathy alternate; Samely (1991); G Alorese (Baranusa) — IPA, Klamer (2003). FLORES-LEMBATA orthography notes / Source notes.

#	English	PAn	PMP	PCMP	Sika (A)	Lewolema (B)	Lewoingu (C)	Solor (D)	Lamalera (E)	Kedang (F)	Alorese (Baranusa) (G)
1	sky	*laŋiC	*laŋit	*laŋit	βula βutu	kəlõ	kələn	'kələ̃	'kelã	ælæŋ ləjɔ; ləjɔ	'laŋi
2	sun				ləro	rəra	ləra	'lə'ra	lə'ra	matan	la'ra
3	moon	* bulaN, *qiNas	*bulan	*bulan	βula-ŋ	vulã	wulan	'wulɑ		>ula ləjɔ malæ manuʔ; malæ **tala**;	'fulaŋ
4	star	*bituqen	*bituqen	*bituqən	dala	pətala nokõʔ (on paper – nɔkõʔ)	pətala	pə'tala	'tona	navun	ta'mala
5	night	*beʀŋi	*beʀŋi	*bəŋin	*ʔβau-ŋ	ara, ('siang':)	rəman	'hɔkɔ̃	'remã	ubɛn	ma'reŋ
6	day	*qalejaw	*qalejaw	*qaləjaw	ləro-ŋ	rərõ		la'rɔŋ	'lɛro	>ɛʔa; haraʔ; ləjɔ	la'ra:
7	year	*kawaS	*taqun	*taqun	ʔliβa-ŋ	sũ, tũ	ləron	tũ:	'tʊŋ	tun	'tu:ŋ
8	wind	*bali	*haŋin	*haŋin	ani-n	aŋ ĩ	aŋi	'ɑŋĩ	'aŋi	>aŋin vɛnin	'aŋi
9	ice										'ɛs
10	snow										'salʤu
11	fog		*kabut	*kabut	*koβa / bhs rit bavo milã		tun/meto			habɔʔ	

#	gloss										
12	river				*naŋa ba; βair ba (river, stream, brook)	suŋe		su'ŋe:	'suŋe	>ubɛ?; ævæŋ	'kali
13	lake	*danaw	*danaw	*danaw	*rano	livo dʒonẽ			'lifo		'danɑu
14	earth	*daʀeq	*tanaq; *taneq	*tanaq	tana	tana	tana	'tana	'tana	au?	'tana
15	stone	*batu	*batu	*batu	βatu	vato	wato	'wato	'fato	tavan; uar; ulu?; va? ai; va? ɛra; va? luli; va? nima?	'fato
16	mountain			*halas	*ili-n (mountain, hill)	ile	ile	i'le:	'ile	>ili	'foto
17	forest	*kaSiw-kaSiw-an				neva, ruka?	kajo	bərə'ga?	'eka pə'nət	>ɛ?a kain; >utaŋ; kain laraŋ	u'taŋ
18	fire	*Sapuy	*hapuy	*hapuy	api	ape	ape	a'pe?	'ape	>api padu; hima?	a'pe:
19	smoke	*CebuN, *qebel	*qebel; *qasu	*masu	nuhi-n	(ape) nuhũ	nuhun	nuhũ	sənə'gʊr	dumɛ?; maliraŋ; in bakɔ; rubu	,ape 'nɑhiŋ
20	dust	*likeS	*qabuk; *qapuk	*qabu	ʔreβu	ape ulã, kəʔavuk	kəawuk			læbu? boŋ; rubu	a'fo:
21	road		*zalan	*zalan	lala-ŋ	rarã	laran	la'rã / la'raŋ	'larã		'tɔ:r
22	house	*ʀumaq	*ʀumaq	*ʀumaq		laŋo? (on paper – laŋɔ?)		'laŋo	'laŋo		ʔuma
23	rope	*CaliS	*talih	*tali; *waʀəj	*tali	taleʔ, korã (on paper – kɔrã)	tale	ta'le:	'tale	vadɛ?	'tale
24	canoe					tɛna, (sampan:) sapã, nelĩ	tena	tena			'tɛna

#	gloss										
25	machete					kənumɛʔ, *ar-kais*: pɛda soru, vaðõ, *tradi-sional*:					'pɛda
26	axe					badõ	soru		hepe		bali
27	spear					gala	gala				
28	knife					hɛpɛʔ, *utk meng-iris*: mərɛʔ	hepe		hepe		du'ri
29	comb					kiriʔ					ki'ri
30	mat					kɛka, ohã					sa'fae
31	clothing					ale lolõ/pake, lipaʔ labu aho(ʔ), *ritual* asu		'agopa'kei	ale'lolo		'konʤo
32	dog	*asu, *wasu	*asu	*asu	ahu		aho	'aho	'ao	>au vavi	a'ho
33	rat	*labaw	*labaw	*labaw		kərome	kərome				ka'more
34	tail	*ikuʀ	*ikuʀ	*ikuʀ	*iʔur	ikuʔ	ikun	'ikũ	'iku	ɛbɔ	'ikuk
35	fish	*Sikan	*hikan	*hikan	iʔa-ŋ	ikã	ikan	'ika~	'ikã	>iʔa >ɛlaŋ	'ikaŋ
36	bird	*qayam	*manuk; *qayam	*manuk	horo-ŋ	kolõ	kolon		'kolo	udaŋ	'koloŋ
37	wing	*paNid	*kapak; *panid	*kapak; *panij	kəpik	kəpiʔ	kəpi	ka'pĩ	kə'pɪk	apiʔ	'kapik
38	claw					təmuʔi(t)		tə'muʔi			'limaŋ ta'nuŋgul
39	feather		*bulu	*bulu	βulu-ŋ	ravuk	rawun	'rawuk	ra'fʊk		wu'luk
40	horn				tara-ŋ	tarã		ta'ra~			ru'ha:
41	snake	*Sulaʀ	*nipay; *hulaʀ	*ulaʀ	*[ular]	ulaʔ	ula	'ula	'ula	>ular pupun	ʔu'la
42	egg	*qiCeluʀ	*qatelur; *qiteluʀ	*qatəluʀ	*təlo	təluk	təlu	tə'lɔŋ	tə'lu	manuʔ tɔlɔr	'taluk

43	worm	*kulay	*kalati; *qali-wati	*kalati	doho	(ulaʔ) valaʔ 'cacing tanah'		kəˈbɛŋə; məˈneo	təlu	hala mukɛl	,ʔula ˈfale
44	louse	*kuCu	*kutu	*kutu	ʔutu	kuto (on paper - kutɔ)	kuto	kutɔ	təlu	utu (head louse)	kuˈtːɔ
45	crocodile					kobu, vadʒaʔ					ˈbapa
46	grass	*Cemel	*baliji	*udu	*βaʔa-ŋ	kahã, kənumak, kəlikət, kərəmət, ('alang2':) luʔo	kərəmək		kərəˈmɛt	>uru; ruʔ vadɛʔ	laˈduŋ
47	tree	*(kayu) Aone			ʔai	kadʒo(')	kajo/pukən	kajo	ˈkaɟo	ai vaʔ; kain laraŋ	ˈkadʒu ˈlɔlɔŋ
48	bark (of tree)				ʔai ʔuli-t	kamak = kadʒo kamã?		kajo kama	ˈkaɟo ˈkãmã	ai amaʔ	ˈkadʒu ˈkamaŋ
49	leaf	*biʀaq	*dahun	*daun	*ʔrou	(kadʒo) lolõ	ləpan	laˈpã	ˈlolo	ai lɔlɔn; lɛpaʔ; lɔlɔn	ˈlɔlɔŋ
50	root	*ʀameC, *ʀamiS	*uʀat	*wakaʀ	ʔramut	(r)amut 'tunggul' , kəlipa [dlm tanah]	ramuk/ramukən	ˈamut	raˈmʊt	ramuʔ	ˈramuk
51	fruit	*buaq	*buaq	*buaq	βua-ŋ	vuã	a wuan	ˈuã / ˈuwã	ˈfuã	>uan	ˈkuluŋ
52	seed				βini	kuluk, kənulũʔ	əra	əˈraʔ	ˈupu	tavan; uluʔ	ˈkuluŋ
53	flower		*buŋa	*buŋa	puhu-ŋ	puhũ	puhun	ˈbuŋa	ˈpu	puhun	ˈbuŋa
54	uncooked rice					lamak, tahã		bəˈras			ˈapa
55	cooked rice					baʔit, lamak (mə-nihũ)		ˈnukɑ; wata təˈnaha			ˈfata

56	betel nut					vua? (malu?)		'ʔuwa			'ʔufa
57	betel vine					malu?		mɑ'lu:			'malu
58	lime					apu?		'ʔapu			'ʔapu
59	chew betel		*mamaq	*mamaq		gã₍ₖⱼ₎ vua? malu?		'ʔuwa / 'ʔua mɑ'lu			
60	sweet potato					ue dʒavã		'ʔuwe			,kurːaː 'ʔutaŋ
61	cassava					ue kadʒo					'kurːa
62	taro					./.		ke'lɑdi (Ind)			
63	sago					./.		kətə'buk			
64	milk					tuho waʔĩ (on paper - tuhɔ waʔi)		'susu (Ind)			'susu
65	salt	*qasiʀa	*qasiʀa; *timus	*qasiʀa; *tasik	*hini	si?a	si?a	'si?a	'sia	tæ?u >apur	'si?a
66	fat	*Simaʀ	*meñak; *himaʀ; *miñak	*miñak		vorã? [pd daging] (on paper - vɔrã?)	mɘlu	wɔ'ra˜ dagĩ (B Kupang)	'forã	hækær tɛdal; ɔpɔ? dækær; vɔja? >aval	
67	flesh, meat	*Sesi, *isi	*hesi; *isi	*isi	*ʔlui-ŋ (flesh)	ihik	mɘnaken	'hik/'elã		>ihiŋ	'ʔihik
68	bone	*CuqelaN	*tuqelan	*zuʀi	luri-ŋ	ri?uk	ri?uk	ri'?uk	'riuk	hara? lurin	ru'?iŋ
69	blood	*daʀaq	*daʀaq	*daʀaq	*mei	mei, rit ra?a	mei	mei	'mei	væi?	'raː
70	liver	*qaCay	*qatay	*qatay	*ʔβate	ate, kɛ?ɛk	aten	a'tɛ˜	'onã		'hati
71	heart				puhu-ŋ βua-ŋ	vuak		pu'hũ	'puo	>ɔnɛ?; muti?	
72	guts	*Cinaqi	*tinaqi (intestines)	*taqi (intestines)	ta?i-ŋ	taɛ kɘbote	tahi onɘn	tãĩ	kə'boti		'teiŋ 'ɔnɔŋ

No.	gloss										
73	stomach		*tian (belly)	*tian (belly); *kampuŋ (belly)	taʔi-ŋ (stomach)	ale, luvo		'alɛ̃	'lufu	bɔtin	'teiŋ
74	head	*quluh	*qulu	*qulu	ala-ŋ	kotə, koto (on paper - kɔtə, kɔtɔ)	kotən	'kotõ	'kotã	tubar	'kɔtɔŋ
75	hair	*bukeS	*buhek	*buk; *qulu; *daun ni qulu	*ala	rataʔ	rata(n)	'ratã	'rata	uha	'rata
76	eye	*maCA	*mata	*mata	*mata	mata	mata	'matã	'matã	matɔ	'mataŋ
77	ear	*Caliŋa	*taliŋa	*taliŋa	*tilu	tilu	tilun	'tilũ	'tilu	til	'til:uŋ
78	nose	*mujiŋ	*ijuŋ; *ujuŋ	*ijuŋ	iru-ŋ	iru	irun	'irũ	(n)i'rʊŋ	niŋ	'ir:uŋ
79	mouth	*ŋusu	*baqbaq	*babaq	məmu	nuhu, vəva	wəwa(n)	wə'wã	fə'fã	nunu vɔvɔ	'fofaŋ
80	lips					nuhu elã		nə'hũ	'nu		'fifiŋ
81	teeth	*nipen	*ipen; *nipen	*ipən; *nipən	niu-ŋ (tooth)	ipə		'ipẽ	'ipã	avar (tooth)	'ʔulɔŋ
82	tongue	*Sema	*dilaq; *hema	*ləma; *lama; *maya	*ma	veve(r) (on paper - vɛvɛ(r))	wewel	wɛwɛl	e'fɛl	æbæl	'fɛfɛl
83	neck	*liqeʀ	*liqeʀ		*ʔəru-ŋ	vuliʔ	wuli(n)	'wulĩ	'fuli	>uli; aduŋ bɔkɔ; bɔŋan aduŋ	'fuliŋ
84	arm					lima	lima(n)	ke'palik			'limaŋ
85	hand	*(qa)lima	*[qa]lima	*lima	lima-ŋ	lima [lima anã? 'jari..']	lima(n)	lima	'lima	liŋ	'limaŋ ka'lumak
86	fingernail				ʔunur	təmuʔi(t)				liŋ urun	ta'nuŋgul
87	breast	*susu	*susu	*susu	uhu-ŋ	tuho, rit tuso (on paper - tuhɔ)	tuho	'tuhũ			tu'hɔ:
88	navel					kəpuhu(r)	(kə)puhurə, puhurən	kə'puhər			ka'pu,hɔr

			*qaqay (leg/foot)	*wai (leg/foot)							
89	leg	*qaqay	*qaqay (leg/foot)	*wai (leg/foot)	*βaʔi	lei	lein	lēī:	'lei		'leiŋ
90	foot		*qaqay (leg/foot)	*wai (leg/foot)	*βaʔi	[lei anãʔ ʻjari ...']		lēī: 'makat		læi	'leiŋ ka'lumak
91	knee				*tur	loto(r) (on paper - btɔ(r))	lotor(ən)	lɔ'tɔr		udul	,leiŋ 'kudul
92	back	*likud	*likud	*mudi	*leʔar	kolaʔ, uhuk	kola(ʔan)	'kolã		>uhur; ɔbi	'ʔalɛŋ
93	skin	*qaNiC	*kulit	*kulit	ʔuli-t	kamak	kuliʔ	'kama	'kãmã	amaʔ	'kamaŋ
94	person	*Cau	*tau	*tau, *taumataq		ata (dikɔ̃ʔ)		'ʔata	ata'dikã		'ʔata
95	name	*ŋajan	*ŋajan	*ŋajan	*nara-ŋ	narã	naran	'nara̰	na'rɛŋ	naja	'naraŋ
96	child	*aNak	*anak	*anak	me	anaʔ	anaʔ	'ʔana	'ana	>anaʔ	'ʔanaŋ / bi're ka'ri:
97	man	*ma-ʀuqanay	*laki; *ma-ʀuqanay	*laki, *ma-ʀuqanay	ata laʔi (laʔi = male)	ama lake; kəlake (yg sdh kawin)	ianməlake	bə'lakĩ	kəbae'lake	>anaʔ >abɛ; >ata diʔɛn; >atɛ diʔɛn; >atan	ka'lake
98	woman	*bahi	*bahi	*bai, *b-in-ay	ata βai (βai = female)	ina (ke)vae	inawae	barɛ' kuaĩ; bər 'waĩ	kəba'rafae	inaʔ; inɛ; >inɔ	ka'fae
99	husband		*bana; *qasawa	*qasawa	ata laʔi	ama lake; kəlake	lake	'lakḛ	kə'lake	>atɛ rian	ka'lake
100	wife		*qasawa	*qasawa	*ata duʔa βai	(ata) (kə)vae	kəwae	kə'wai	kə'fae	væʔ rian	ka'fae
101	mother	*t-ina	*t-ina	*ina	ina	əmaʔ, rit ina	inawae	'əma; 'ina	əma		'ʔina
102	father	*t-ama	*t-ama	*ama	*ama	ama, bapaʔ, tata	ama	'ama	bapa	amɛ; >amɔ	'ʔama
103	older sibling					kaka, tata		ʔɔa (female sibling); nɔ̃ (male sibling)	'tata		'tataŋ

#	Gloss							ʔɔa (female sibling); nɔ̃ (male sibling)			
104	younger sibling					adeʔ, *rit* ariʔ		ʔɔa (female sibling); nɔ̃ (male sibling)	a'riɪk		'ʔar~i
105	granparent					baʔ / dadi					'ʔufa 'beiŋ / gina
106	grandchild					anaʔ susu					
107	slave					naba, ala dʒati					
108	1sg	*i-aku	*i-aku	*i-aku	aʔu	goʔε, *colloquial* go (on paper - gɔʔε)	go	'goʔe	'goe	>εʔi (obj); >εi (subj)	go:
109	2sg (informal)	*i-kaSu	*i-kamu; *kamiu	*imi, *miu	ʔau	moʔε, *coll* mo (on paper - mɔʔε)	mo	'moʔe	'moe /' mio	ɔ	'mɔ:
110	2sg (polite)					moʔε (on paper – mɔʔε)		'moʔe			'mɔ:
111	3sg	*si-ia	*si-ia	*s-ia	nimu	naʔε	na	'naʔε	'nae	nuɔ; ni	'nɔ:
112	1pl excl	*kami	*kami	*k-ami	ʔami	kamε	kame	'kamε	'kame	tε	'kame
113	1pl incl	*i-kita	*i-kita	*k-ita	ʔita	titε	tite	'tite	'tite	ε; kε	,ʔi'tε
114	2pl	*i-kamu			miu	mio (vaoke)	mio	'mio	'mio	mε	,mi sa'kali
115	3pl		*si-ida	*sida	rimu	raʔε (vaoka)	ra	'raʔε	'rae	suɔ	'fe: sa'kali
116	this	*i-ni	*i ni	*-ni		pi(a), piʔ ĩ (mĩn)	pi, piʔin	pi	'pi	nɔʔ	,ha'ʔ~a
117	that		*i na	*-na		pe, peʔε̃ (mε̃)	pe, peʔen	pe	'pe	sɔbε	,ka'te
118	at	*i, *di	*i; *di	*di		tə(ka)	pe	ʔia	'lau, 'rae		'ʔunuŋ
119	here					tə pi, təʔi	pi(a)	pia, pi	'dipi	bæʔ; di nɔʔ	'handʒa'fa

#	gloss										
120	there					pəli, tə pe, te	pe	pə'rai, rai, pe:	'depe	di mæ; di nɔbɛ; nɔbɛ	'alika'le:
121	above	*i babaw, *i taqas	*i-taqas; *i-babaw	*atas		lolõ		tɛti; tɛti ... wutu			'lɔlɔŋ
122	below		*i babaq	*babaq		laũ [tempat/ daerah !?]		lali; lali ... 'lɛreŋ			'la:uŋ
123	left	*ka-wiʀi	*ka-wiʀi	*ka-wiʀi	βiri	nek ĩ	nekin	nei'ki	'meki	vɛri	'mekiŋ
124	right (side)	*ka-wanaN	*ka-wanan	*ka-wanan	βana	vanã	wanan	'wana	'fana	vana	di'kɛ~:
125	a few				*kəsik ha	ata bərua		bə'rua	'usi	uraŋ	'ʔata 'ʔusu
126	all	*amin	*amin		*haha	ata vəkən kaẽʔ	wəkən kaen	wɔkɔ'kai	fa'kahae	ræiʔ	'ʔata sa'kali
127	many				gaβa-ŋ	kadʒak [< adʒaʔ]	aja	'ajaˇ	'ajã	rai varan	'ʔata la'bi:
128	some				*pira ko pira	toʔu rua, bərua (on paper - tɔʔu)	bua	bə'rua	bo'pira	vai	
129	to count	*Sipuʀ	*ihap	*iap	*[rekeŋ]	gasit, hivək	gasik	ga'sik		karɛʔ kataʔ	'rɛkiŋ
130	one	*esa, *isa	*esa; *isa	*əsa, *isa	ha	toʔu (on paper - tɔʔu)	toʔu	tɔ'ʔu	'tou	>udɛʔ	'tɔ:
131	two	*duSa	*duha	*dua	rua	rua	rua	'rua	'rua	suɛ	'rua
132	three	*telu	*telu	*təlu	təlu	təlo	təlo	tə'lo:	'telo	tælu	'tal:au
133	four	*sepat	*epat	*[ə]pat, *pati, *pani	hutu	pat	pak	pa:	'pa	>apaʔ	pa:
134	five	*lima			lima	lɛma	lema	le'ma:	'lema	lɛmɛ	'lɛm:a
135	six					nəm(ə)	nəmən	nə'mũ	'nemu		'nam:u
136	seven					pito	pito	'pitu / pito	'pito		'pit:ɔ
137	eight					buto	buto	'wutu:	'buto		'but:ɔ

#	gloss										
138	nine					hiva	hiwa	hiwa	'hifa		'hif:a
139	ten					pulo	pulo	puloʔ; pulok¬	'pulo		'kartɔ
140	twenty					pulu rua	pulu rua	pulu 'rua			ka:'rua
141	one hundred					ratu toʔu, təratu (on paper - tɔʔu)	ratu toʔu, təratu	tɛ'ratu			'ratu
142	big	*ma-ʀaya	*ma-ʀaya	*ʀaya	gəte	belə?	belə / bapan	'bɛlə̄:	'belã	kɛdɛŋ kɛnaʔ; ria baraʔ; tiri bɔrɛʔ	bī:
143	small	*kedi	*dikiq; *kedi	*dikiq, *kedi	*kəsik	kəne/kəniʔ/kəsiʔ	kəne	a'nə̄:	'keni	kulɛŋ kaiʔ; mæʔæ; utu kulɛŋ; utu ɔlɔr	'an:aŋ
144	old (house)				*blupur	(laŋoʔ) nolõ (on paper - (laŋɔʔ) nɔlõ)	okin			tua	'nɔlɔ, 'uma 'nɔlɔ
145	new (house)	*ma-baqeʀu	*ma-baqeʀu	*baqeʀu; *bəqəʀu	βəru	(laŋoʔ) wuʔũ (on paper - (laŋɔʔ))	wuʔun	'wũʔũ	'fu	vɛrun taʔɛn	'funɔ, 'uma 'funɔ
146	old (people)	*ma-tuqaS	*ma-tuqah	*tuqa		tua belək, tənuẽ		ɔ'kī /ɔ'kɪn	'magu		'gambɛ-'gambɛ, 'ina-'ina
147	young (people)					nubũ, bəlia					bɛ 'lamɛŋ, 'kafaɛ
148	hot	*ma-panas (warm)	*ma-panas (warm)	*gahu (warm)		pəlate		pə'lɛte:	pə'late	daja (warm) mi ælæʔ; æmi pana; æmi ruku	pa'latiŋ
149	cold		*ma-diŋdiŋ	*ma-dindiŋ	*blata-ŋ	bəriŋ̃ĩ	gələtə	gə'lətə mə'laŋ;	gə'lətã		ka'luaŋ
150	good		*ma-pia	*ma-pia	ʔepa-ŋ	dikəʔ, maɛ, sarɛ?	əre	kele'mur/se narɛ	səna'rɛŋ	diʔɛn hɛrun	'dik:ɛ

151	bad	*kuya	*zaqat	*zaqat	*goʔis	da(tə), əvə da	məko	'datə̃:	afada'tən	buruʔ balaŋ; datɛn ɔhaʔ; iŋin bɔaŋ; kɛhɛ >alɛ	'datɛ
152	full				bənu	mənu, ('sesak':) hugiʔ		mə'nu:		>ihi; >ihiʔ; bɔlɔr mapaʔ; pænu mænu	'pan:ɔŋ
153	near	*ma-azaNih	*ma-azani	*hazani; *raŋi	*groʔo (near adv)	dahɛʔ		dahɛ̃	'dae		'dah:ɛ
154	far	*ma-dawiN	*ma-zauq	*zauq	blaβir	doã		dɔ̃e	'doe	dɔa tɛvɛl	'dʒuaŋ
155	wet		*ma-baseq	*ma-basəq	gəma	dəmɛʔ	dəman	də'mɛ̃	sə'nəbe	baha; dɔrɔʔ vɛrɔʔ; ritaʔ dɔrɔʔ	'ɛlɛ
156	dry	*ma-qaʀiw	*ma-ʀaŋaw	*ma-ʀaŋaw	*duʔur	maraʔ	mara	ma'rã	'marã	majaʔ midɛ; pariʔ; turi rɔka	'mara
157	long	*inaduq	*anaduq	*anaduq	bloŋ	bəlaha	bəlola	bə'lahã	'blã	lava dɛkar (distance); lɛla dɔaʔ (time)	ba'lah:a
158	short		*ma-babaq	*babaq	buluk	kəruʔ, kubar, boso (on paper - bɔsɔ)	kəsuʔ	kə'rõ	kə'ru	ɔhaʔ lɛla	'mak:u
159	thin	*ma-NiSepiS	*ma-nipis	*ma-nipis	*blelər	kərogo, komoʔ (on paper - kərɔgɔ, komɔʔ)	kərogon, mənipi	tipis (Ind)	mə'nipi	mipi tɛʔul; piki rikan	'kar:i
160	round				*guər	mogo		bə'lopor	belo'pɔr	ɔpɔl kɔdɔʔ; ɔpaŋ sɛlɛ	ga'lɔkɔ
161	dark					bəruhu		'mitə̃:			'kuiŋ
162	dirty		*cemed	*ma-qetəm	miʔak	kotoʔ, milã, (air:) pəmu (on paper - kɔtɔʔ)	milan	'milã̄	'milã	buhaŋ ræmuʔ; mitɛŋ kajɔʔ; ulaʔ kɔpiŋ	ka'lit:a

#	gloss										
163	sharp	*Cazem	*ma-tazem; *tazim	*ma-tazim	dira-ŋ	dəkət	bərəkə	də'kət	bərə'kət	dæjæʔ; æruʔ; pahɛʔ	'dak:ɛ
164	blunt (dull)		*pundul; *dumpul	*dumpul	*bou-ŋ	gã halaʔ 'tdk ma-kan', mənəŋõ		bə'da:t	'moso	>umal	'kumbu
165	heavy		*ma-baseq	*ma-basəq	*[bərat]	baʔat	baʔa	ba'ʔa	'batã	baraʔ bɛtɛ	ba:
166	straight				dəlor	mopaʔ (on paper - mɔpaʔ)	mopaʔ	mo'pa˜	'mūlu	ɔlɔr ælɔr	'mallɔŋ
167	wide	*ma-lawas	*ma-labeʀ	*labeʀ	*kləβa-ŋ	nasar, vakoʔ varik (on paper - vakɔʔ)		lɛbar (Ind)	pela'fɛt	kɛdɛŋ kɛnaʔ; ria baraʔ	'bɛa
168	narrow		*kepit		*ipot	(h)ipət		kə'təkə	hi'pɛt		ki'p:ɛ
169	correct		*ma-bener		*akə-t	dikəʔ, murə̃ʔ, vanã		'murə˜	'murã		'mallɔŋ
170	ripe					tahak		tənə'hə	'tã		'tah:a
171	rotten	*ma-buʀaq	*ma-buʀuk; *ma-busuk	*buʀuk; *busuk	*βau-ŋ	da(tə)	məko	'waũ	'fãu	buruʔ balaŋ; iŋin bɔaŋ	'datɛ
172	flat, smooth				*ʔalus	əbaʔ, nasar, rana	kəloho	kə'lehɔk	'lere		'hama
173	thick	*ma-kaSepal	*ma-kapal	*kapal, *telu	ʔapar	batɛʔ, budʒet	bəsiʔ	təbal (Ind)	pə'fəre	kapal vahɔʔ; batuʔ nukɛl	'gapa
174	other	*duma	*duma; *laqin	*liqan		gehak, ikər	ikərən	'ikər	'geak	>ahin; palan; vai; vɛʔɛn	'hama la'hɛ
175	black	*CeŋeN	*ma-qitem	*ma-qetəm	*mita-ŋ	mitə̃	mitən	mi'tə˜	mi'tãŋ	mitɛŋ kajɔʔ	'mit:ɛŋ
176	white	*ma-puNi	*ma-putiq	*burak	*bura	buraʔ	bura	bu'rã	'burã	bujaʔ tapɛ	'bur:a
177	red	*ma-taNah	*ma-iʀaq	*meʀaq	*mera-ŋ	mɛʔa	meʔan	'meʔə˜	'meã	kɔrəŋ; putuʔ	'mɛã

#	gloss										
178	green	*mataq	*mataq	*mataq	daʔa-ŋ	idʒo, taŋə̃ (on paper - idʒo)	bətən	'ijɔ̃	'keor	taŋɛ; taŋɛn dɔlɔr	lɔ~:
179	blue					aŋat			pe'lɛŋ		ba'lapã
180	yellow		*ma-kunij	*kunij	herə-t	kumã	kuman	'kumaˉ	'kũmã	ɛɔr; uma katɛʔ	'kumɔ~:
181	blind					kisa, kəbulõ	matan buta				
182	deaf					kəbekeʔ, kələtũʔ	kəbeke				ka'muk:ɛ
183	to see	*kita	*kita	*kita	*ʔita	loŋo(t), tədəʔ, təŋət, noi_kj1 (baĩ) (on paper - lɔŋɔ(t)) (tilu) baĩ, dəŋəʔ,	təŋə		tə'gəl	bɔraʔ hɛraʔ; ɛbɛŋ bɔraʔ; ui	'hik:i, 'sɛru
184	to hear	*tumaNa	*deŋeʀ	*dəŋəʀ	*rəna	vɛŋɛ	bain	baĩ	də'ŋa	baŋæer; dæŋɛr bæiŋ	'daŋ:a
185	to smell	*Sajek	*hajek (sniff/smell)	*hajək (sniff/smell)	βau-ŋ (vb intr.); sino-ng (vb trans.)	pənu, siõ	sion	'siɔ̃		miʔɛn; naæʔ pɔvɔn	'siɔ~:
186	to think	*nemnem, *-ajem	*demdem		*huk	reã	pikir	pikir (Ind)	'petã	kælæn kauʔ	'pikir
187	to know	*bajaq	*taqu		raʔinta-ŋ	noi_kj1	koiro, moiro	'moiro	'toi	lalaŋ; ui	
188	to say		*kaʀi; *tutur	*tutur	*βeta	mañ	marin	'marĩ	'mari	tælæ	ma'rĩ:
189	to speak					koda, mañ, tutu(ʔ)	koda, tutu	pə'tutuk; pə'kɔdak			'tut:u
190	to lie down		*qinep	*qenəp; *qinəp	tuʔə	gola (on paper - gɔla)	tobo	'turu		padɛ akal	'tur:u
191	to sleep	*tuduʀ	*tiduʀ; *tuduʀ	*tuduʀ	tuʔə	turu(ʔ)	turu	'turu	'turu	butɛ; tæʔæl tɛbɛʔ	'tur:u

No.	gloss										
192	to wake up					hogo (on paper - hɔgɔ)					'baũ
193	to wake someone up					iũ, tobẽ					'guɔ 'baũ
194	to bathe					həbo(ʔ)(kj2)			'hebo		'habːɔ
195	to bathe a child					həbo anaʔ					'habːɔ 'anaŋ
196	sit		*tudan	*todan	təri	tobo (on paper - tɔbɔ)	tobo	'tobo; 'toboh	'tobo	tɛbɛʔ	'tɔbɔ
197	stand	*diʀi	*diʀi; *tuqud	*diʀi	gəra	dɛʔi, sedət	deʔin	'de'ʔĩ	'dei	madɛr hɛdaŋ; madɛr tɛbɛʔ	'tidːɛ
198	to walk	*Nakaw	*lakaw; *paNaw	*lakaw; *panaw	pano	pana	pana	'pana(h)	'pana	pan lɛdɔ	'pana
199	run					pəlaʔe(ʔ)(kj2)	pəlae				pa'laɛ
200	swim	*Naŋuy	*laŋuy; *naŋuy	*naŋuy	*nani	naŋe	naŋe	'naŋe	'naŋe	naŋi >ɔjaŋ	'naŋɛ
201	fly	*layap	*ʀebek; *layap		horo	bəka, dadʒuk	kənəpun	bə'ka(h)	bə'ka	baʔa baæʔ; ubur	'bakːaŋ
202	fall		*nabuq	*ka-nabuq; *ma-nabuq	*ʔela	goka, hoat, loʔuk (on paper - gɔka, hɔat)	dəka	dəkak; lɔrã	'goka	hubaʔ hɔkɔ; hubaʔ kɔliʔ; kɔʔal; lɔduŋ; mɔruʔ hɔkɔ	'gɔka
203	to drop					ləŋat					'nĩː 'gɔka
204	to drink	*mimah	*inum	*inum	ʔ-inu	nenuŋ(kj1)		'tɛnũ	'tenu	in a	
205	to eat	*kaen	*kaen	*kan	*ʔoa	buʔa, (dia makan:) gã(kj)		tə'kən / tə'kə̃	'kã, 'go	a; hiduʔ; kæmæŋ	

206	to cook	*tanek	*tanek; *zakan (both also to boil food)	*tanək; *zakan (both also to boil food)		biho (behĩ), patã		'biho			'dakaŋ
207	to wash				*popo	baha?, hue, pu?u	baha	ba'hak	'ba / 'pu	hidaŋ	'huɛ, 'bɛmɛ, 'lamiŋ
208	to sew	*taSiq	*tahiq; *zaqit	*zaqit	ra?it	havu, (dng jarum:) dau	agi?	hɑu	'hau		'haur
209	to live	*ma-qudip	*ma-qudip	*maqudip	morə-t	morit(kj2)	i?a	'mori	'mori	bita; bita matɛ	'mɔri
210	to breathe		*mañawa	*ñawa	aiŋ	hañĩ (nañĩ)	nahin	'napɑs	ə'rã 'nãi	>ipɔ; dujɛŋ	'tar:ɛ 'nah:ĩ:
211	to work	*qumah	*quma			gəriã/kəriã, (di kebun:) ola mã	kərian	'kriã	kəri'ã		ka'radʒaŋ
212	to die	*m-aCay	*m-atay	*matay	mate	mata	mata	'mata(h)	'mata		'matɛ
213	to give				βəli	nẽĩ	nein	'sɔrɔ̃	'soro	bɔ?; hɔta?; lɔbɔ; ni; sɛra? sara?; sɔrɔŋ natɛŋ	'nɛĩ
214	wipe				*bloso (rub or wipe)	gi?e, (badan:) hamu	təru	'so?o		dɔru; pɔhɔ?	'hapɔ
215	to come		*maʀi	*mai	*mai	bəso, be?o, beto, həvo, səga (on paper - bəsɔ, həvɔ)	səga	be'go; se'ga	bə'so		'bɛta, 'nɔu
216	to laugh	*Cawa	*tawa	*malip; *tawa	*to	geka (on paper - gɛka)	geka	'gɛka	'geka	tavɛ hɛkɔ	'gɛki

217	to cry	*Caɲis	*taŋis	*taŋis		tani	tanin	'tani	'tani		'tanĩ
218	to dance					soka, selɛ̃, (tarian adat:) hamã (on paper - sɔka)		'sɔka; 'selɛŋ	'soka		'tam:ɔ
219	to sing				*[kantar]	opak (belũ), kan-tar bəpana, hodeʔ anaʔ, iɲ giɲ) (on paper - ɔpak)	kanta	kantar (Port)	kantar		'pantɔ, 'bɔtɛ 'liaŋ
220	burn		*tunu	*tunu	*holo (vb trans); olor (vb intrans)	tunoʔ (on paper - tunɔʔ)	tuno, buko	sə'rũ	'papi	hɛiʔ; paʔ bɛɔʔ	'tun:ɔ
221	dry in sun					paʔĩ					'paɛ~:
222	to blow	*Siup	*hiup	*upi		bu	bu	bu:	'die		'pui
223	search					beã, genak, hulɔ̃, saeʔ, səbaʔ, vaeʔ					'gɛna
224	to hit	*palu	*palu	*palu	*tola	baheʔ, bəɲ, guar, gəlolak (on paper - gəblak)	bərin	bə'rĩ	'teka	paluʔ; (others see Samely 1991:214)	'bɛh:ɛ
225	shoot	*panaq	*panaq (an arrow)	*panaq (an arrow)		pasak (bədi)		'leʔɔ̃; 'pasak	'pasa		'pasa
226	bite	*karat	*kaʀat	*kaʀat	ʔiʔi	gakaʔ, gike	gike	gi'ge / gi'ke	'goki	kiʔi ka; kiʔi avar	'gaki

227	to cut	*taʀaq, *tektek	*taʀaq; *tektek	*taʀaq (cut wood); *tətək; (cut wood)	*poro	gəto, pətə, poʔok (on paper - pɔʔɔk)		pɔ'ʔɔk	'belo	bɛl baʔ; bɛluŋ; ɔhɔr; patɛ >ihu; pæriʔ piʔ	'pak:u
228	to split		*belaq; *silaq	*bəlaq	kəlak	apik, giaʔ, gikaʔ, hikaʔ, ləgaʔ, pakar, seʔek, tikaʔ	bia	'gikaʔ	lə'ka	baʔ; likɔʔ mapaʔ	'bat:a
229	stab		*suksuk	*susuk	*rohuk	robak, rohoʔ, tu-bak, səgat (on paper - rɔhɔʔ)	robo	'rɔhɔk	'tuba	bakɔʔ; bælɔʔ baʔ; bælɔʔ bɛhar; bælɔʔ ɔhɔr; tuhuʔ bakɔʔ	'sik:a
230	to fight				*punu βiʔi-ŋ	gəñi vəkik(a), pəvuno (on paper - pəvunɔ)	gənin	gə'ni	'uno	awɛ >unu; bajuʔ bɛkɛ; halaʔ; miʔɛr rɛŋa; sara bɛl ka'laĩ	'fak:ĩ
231	scared	*ma-takut				kotə bou, takut(Ki2) (on paper - kɔtə)		ta'kutɑ	kə'ruit		'taku
232	to throw		*tudaq	*tudaq	*roga	dəkaʔ, gebaʔ, hitõ, tadaʔ, valuʔ, veleʔ	geba	ge'bɑʔ		bɛʔ; kakaʔ; vɔtɔʔ; tidaʔ; vidaʔ	dɛĩ:
233	to hunt	*qaNup	*qanup		raka-ŋ	hevak/sevak (on paper - hɛvak/sɛvak)	pəreha	ba'tin	'bati	dɛruŋ dæŋ	'tut:ɛ

#	gloss									
234	to kill	*p–aCay	*bunuq	*bunuq	*dola mate	belo, buno, naʔã mata (on paper - bɛlo)	maan mata	'bɛlo	təba'jɛk	bælɔʔ baʔ; bælɔʔ bɛhar; bælɔʔ ɔhɔr; avaŋ	'bun:ɔ
235	to dig	*kalih	*kali	*kali; *keli	*goʔi	gəliʔ, guʔit, robək/ robok (on paper - rɔbək/ rɔbɔk)	gali	'gũʔĩ	'gui	kaæl kɔrɔʔ; kaliŋ kɔrɔʔ	'gali
236	to suck	*sepsep	*sepsep	*susu	*hiruk	dəmuʔ, isək	isək	'isək		duhu; dumɛʔ; dupɛŋ	'dum:ɔ
237	to flow	*qaluʀ	*qaliʀ; *qaluʀ; *saliʀ	*saliʀ	ba	ba	ba	ba:ʔ	'bã	kavaŋ kɔŋ	'pana
238	to freeze					./.					'fatɔ
239	to hold	*gemgem	*gemgem		toe	aʋut, hukũ, napũ, pehẽ	pehen	'pehɛ̃	'pe	pænæ; tɔbɔl	pa'ha:
240	to tie		*hiket	*hikət	pəte	gahã, gorã, gudut, pũĩ, seget, soɠĩ, ʋã, ʋido, ʋihik (on paper - gɔrã)	puin	wi'do:ʔ	'hoŋã	aviʔ; hɔŋɛn; læra; pætiŋ	pɔ'hi:
241	to spit		*luzaq	*zulaq	*ilur	pərino, putə̃ iluʔ _rit_ pitũ	bage	'pino	tə'miro	>ijuʔ; mijuʔ; pɔtaʔ mijuʔ	'buh:u 'ill:u
242	to vomit	*utaq	*utaq	*mutaq	*muta	luʔuk, muta	muta	mu'ta	'muta	>ɔkaʔ; mutɛ >ɔkaʔ	'mut:a
243	to play				ʔləbe	mura, gənəkuʔ	gənəku	mə'ŋər	gə'la	huaŋ æbæl	ka'nɔku

#	gloss										
244	to pull				*gide	bəreta, deru, ge-hã, gideʔ, tamĩ odo, (t)odʒon/to-dʒok (on paper - all o = ɔ)	odoʔ	ge'hɛ̃; rə'dɔk		bɛʔ; dɛra bɛkɛ; kɔliʔ	'tar:ɛ
245	to push				*rusik (push, shove)		gehan	bo'gɔ̃	'odo / 'uruk	>ænæʔ; >ɔbaŋ; >uruʔ; hɔtaʔ; tæʔæ	'tɔbaŋ
246	rub				*bloso (rub or wipe)	doruʔ, purit, rosuk, teʔu palé	dosuʔ	do'rɔʔ	'doru	dɔru	dɔ'h:ɔ
247	scratch	*kaʀaw	*kaʀaw	*kaʀaw	*kəru (scrape)	gaʔu, raguʔ	ragu	'ragu	'rago	karɔ	'gaɔ
248	float				*baβak	bao (on paper - baɔ)	bao	'bau		baɔ tata	'nɛpi
249	swell	*ʀibawa, *baʀeq	*baʀeq	*baʀəq	boβo (swelling)	baʔa		bə'ʔah	kəba'ras		'baɔ
250	to turn	*Niku	*biliŋ; *puter	*putər	*pleur (turn around); βalər (t. over)	sɔlɛut, véu	peko	'balik		balɛ kɔlɔʔ; batiŋ; bɛlɔʔ balɛ; bɔkɔ iɔʔ; ɛkuʔ; kɔlɔʔ balɛ	'lakɔŋ
251	no/not	*ini	*qazi; *diaq	*diaq, *ta, *ta-i	eʔo-ŋ	eka, halaʔ noʔõ (on paper - noʔõ)	take/halaʔ	'take; 'amu = "no", hala = negator	'take	ɔhaʔ	la'h:ɛ
252	and	*ka, *mah	*ka; *ma	*ma	mole			nə	'nã		nɔ~:
253	because				loni-ŋ	halaʔ, vaŋũ, turũ	dari/pəkən	lə'kũ	pu'kãŋ	ɛlɛ	'karɛna, sɛ'bab
254	if	*ka, *nu	*ka; *nu	*ma	raʔik	kalo, kalu	kalo	kalau	'kalu	ɛrɛ	'kalau

#											
255	with					noʔõ (on paper - nɔʔõ)	noʔon	mə̃	'na~	dapɛʔ; nɔrɛ	nɔ~:
256	what?	*n-anu	*apa; *-anu	*apa, *sapa		a	a, aʔa	a:	a'laka	>apɛ	pei
257	who?	*si-ima	*i-sai	*sai, *sei	hai	hege	hege	hɛ'gei	'heku	siɔ	'haf:a
258	where?	*i-nu	*i nu	*be, *pae		tə ga	gaʔe		'diga	dɛnita; dita	na'ŋga 'ɔrɔ
259	when?	*ija-n	*p-ijan	*p-ijan	*nora pae	ara/ərə̃ pira	ərən pəwia (past), ərən pira, ərən pia (future)	ərən 'pira	ara'pira	ɛrɛ; væŋ piɛ	ɛrpɛ'hɛlɛ
260	how much/many?					pukə̃ a	dari a, pukən a				'peinã:
261						pira	pira				pir:a
262	how?	*kuja, *numa	*kuja; *kua		ganupae		nənən gənai, nən gaʔe		ara'pira	nara bɔnɛ	na'mɔnaŋga
263	way, path	*zalan	*zalan (road/path)	*zalan (road/path)	lala-ŋ				larã		
264	rain	*quzaN	*quzan	*quzan	ura-n		uran		'urã	>uja	
265	shoulder	*qabaʀa	*qabaʀa	*qabaʀa			hanan				
266	chew		*mamaq (also chew betel)	*mamaq (also chew betel)				nənə'gaʔe			
267	yawn	*ma-Suab, *ma-Suaw	*ma-huab	*mawap; *moap							
268	dream	*Sepi	*hipi	*nipi							
269	thatch	*Cawali	*qatep (thatch/roof)	*qatəp (thatch/roof)					u'fun		
270	needle	*zaʀum	*zaʀum	*zaʀum					'luhir		
271	steal	*Cakaw	*takaw	*takaw					'tāka		

#	gloss								
272	living, alive		*ma-qudip (to be alive)	*maqudip (to be alive)					
273	wood	*kaSiw	*kahiw (stick/wood)	*kayu (stick/wood)				'pukā	
274	to plant	*mula	*mula	*tanəm				'mula tu 'bak	
275	choose	*piliq	*piliq	*piliq				'pile	
276	grow		*tu[m]buq	*tumbuq					
277	squeeze	*pereq	*pereq; *peʀes	*pəʀəs; *ʀaməs (as in juice from fruit, for both)	pəra	pe?uk	pi'?uk	'piuk	>ɛrɛʔ; dipɛʔ; kimuʔ kamuʔ
278	buy	*baliw, *beli	*beli	*bəli					
279	to open		*buka (open/uncover)	*buka (open/uncover)					
280	fowl								
281	mosquito	*likeS	*ñamuk	*ñamuk		kənamu			
282	spider	*kakaCu	*lawaq	*lawaq					
283	branch		*daqan	*daqan; *saŋan					
284	sand	*bunaj, *qenay	*qenay	*qənay	*ne	wəra		'botā	buta; >ɛnɛ
285	water (fresh)	* daNum	*danum (fresh); *wahiʀ (fresh)	*waiʀ (fresh water)	βair	wai	wai	'fai	væi ai
286	sea/salt water	*tenem	*tasik; *laud	*tasik	tahi	tani		'lefa	lɛva; tahiʔ >aɲin

287	thunder	*deʀuŋ	*gurgur; *kudug; *deruŋ	*gugur	pələdə
288	lightning	*likaC	*kilat; *qusilaq	*kilat	kila
289	sick, painful		*ma-sakit	*sakit	bəlara
290	shy, ashamed	*ma–Siaq	*ma-hiaq	*mayaq	'meaŋ
291	hide		*buni	*buni	
292	climb	*dakiS	*pa-nahik; *dakih	*panaik; *sakay	

Appendix 3. Unanimous word-initial consonant correspondences

FLORES-LEMBATA	orthography notes	IPA; * = reconstructed form; *[ ] = ungrammatical Lewis & Grimes 1995 (in: Tryon 1995)	IPA Pampus (2005)	IPA Nishiyama and Kelen (2008)	IPA Klamer (2002)	IPA Keraf (1978)	IPA; >V = breathy alternate Samely (1991)	IPA Klamer (2003)	
Source notes		A	B	C	D	E	F	G	
#	English	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese (Baranusa)	Proto-sound
7	year		t–		t–	t–	t-	t–	*t
14	earth	t-	t-	t-	t-	t-		t–	*t
33	rat		k-	k-				k–	*k
39	feather		r-	r-	r–	r-			
46	grass		k-	k-		k–			*k
57	betel vine		m-		m–			m–	*m
65	salt		s-	s-	s–	s–		s–	*s
69	blood	*m-	m-, *rit* r-	m-	m–	m–		r–	*m; *r
74	head		k-	k-	k–	k–		k–	*k
76	eye	*m-	m-	m-	m–	m–	m-	m–	*m
85	hand	l-	l-	l-	l–	l–	l-	l–	*l
88	navel		k-	(k)-	k–			k–	*k
95	name	n-	n-	n-	n–	n–	n-	n–	*n
110	2sg (polite)		m-		m–			m–	*m
114	2pl	m-	m-	m-	m–	m–	m-	m–	*m
132	three	t-	t-	t-	t–	t–	t-	t–	*t
133	four	hutu	p-	p-	p–	p–	>ap-	p–	*p
134	five	l-	l–	l-	l–	l–	l-	l–	*l
138	nine		h-	h-	h–	h–		h–	*h

156	dry		m–	m-	m–	m–	m-	m–	*m
157	long	b-	b-	b-	b–	b–		b–	*b
175	black	m-	m-	m-	m–	m–	m-	m–	*m
176	white	b-	b-	b-	b–	b–	b-	b–	*b
177	red	m-	m-	m-	m–	m–		m–	*m
184	to hear	***r-**	b-; **d-**	b-	b–	**d–**	**d-** b-	**d–**	*b
185	to smell	s-	s-	s-	s–			s–	*s
188	to say		m-	m-	m–	m–		m–	*m
191	to sleep	t-	t-	t-	t–	t–		t–	*t
196	sit	t-	t-	t-	t–	t–	t-	t–	*t
209	to live	m-	m-		m–	m–	m-	m–	*m
212	to die	m-	m-	m-	m–	m–		m–	*m
242	to vomit	*m-	m-	m-	m–	m–	m-	m–	*m
246	rub		d-	d-	d–	d–	d-	d–	*d
254	if		k-	k-	k–	k–		k–	*k

Appendix 4. Unanimous word-medial consonant correspondences

FLORES-LEMBATA	Source notes	Lewis & Grimes 1995 (in: Tryon 1995) A	Pampus (2005) B	Nishiyama and Kelen (2008) C	Klamer (2002) D	Keraf (1978) E	Samely (1991) F	Klamer (2003) G	
#	English	Sika	Lewolema	Lewoingu	Solor	Lamalera	Kedang	Alorese (Baranusa)	Proto-sound
19	smoke	-h-	-h-	-h-	–h–			–h–	*h
83	neck	**-r-**	-l-	-l-	–l–	–l–	**-d-**	–l–	*l
87	breast	-h-	-h-	-h-	–h–			–h–	*h
88	navel		-h-	-h-	–h–			–h–	*h
88	navel		-p-	-p-	–p–			–p–	*p
93	skin	**-l-**	-m-	**-l-**	–m–	–m–	-m-	–m–	
102	father	*-m-	-m-	-m-	-m-			-m-	-m-	*m
112	1pl excl	-m-	-m-	-m-	-m-	-m-	tɛ	-m-	*m
123	left	**-r-**	-k-	-k-	–k–	–k–	**-r-**	–k–	*r; *k
126	all		-k-	-k-	–k–	–k–			*k
134	five	-m-	-m-	-m-	–m–	–m–	-m-	–m–	*m
135	six		-m-	-m-	–m–	–m–		–m–	*m
157	long		-l-	-l-	–l–	–l–		–l–	
180	yellow		-m-	-m-	–m–	–m–	-m-	–m–	*m
216	to laugh		-k-	-k-	-k-	-k-	-k-	-k-	*k
225	shoot		-s-		-s-	-s-		-s-	*s