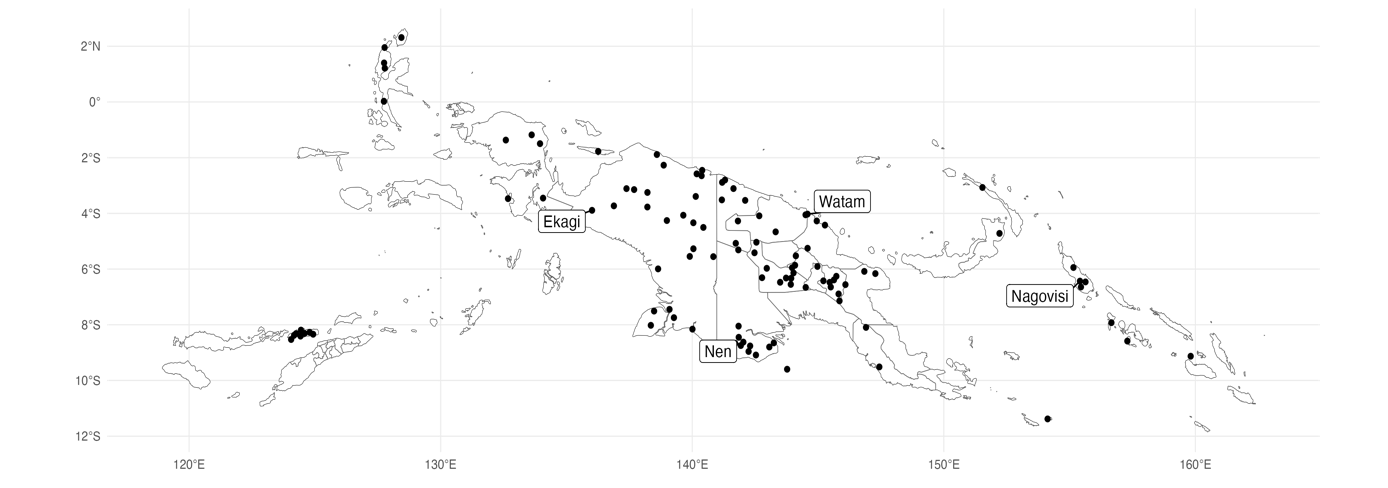
**3. Zooming out to a broader sample**

So far, we have focussed in some detail on just four Papuan systems. While this allows us to get a feel for their overall logic, it has the disadvantage of being just a tiny subsample of the diversity observed in Papuan languages. In this section we do the opposite: we draw on our sample of Papuan languages in Kinbank. Kinbank is a database containing a global sample of kinship terminology from 1,229 languages. Within this is 109 Papuan languages from 44 maximal clades (Figure 1). We use the Kinbank data, and the Papuan subset, to make broad-based statements about the diversity of kinship terminology in Papuan languages.

Kinbank is built around a set of 115 kin *types*, the genealogical categories of relatives determined through reproduction or marriage, such as elder brother, or wife’s father. No language distinguishes all kin types (typical numbers of kin terms range from 20-30 per language), so each language necessarily syncretises – uses the same term for – many kin types. The crucial point is that languages syncretise terms in different ways. For example, one language might use the same term for meB = feZ ≠ R (‘older same-sex sibling’; we use R to mean ‘rest of the kin types’, so ≠ R means ‘and no other kin type’)), one might use a rule like: meB = myB = feB = fyB ≠ R (English *brother*), but another equally plausible organisation is: meB = feB = myB = fyB ≠ R (Nen *nne* ‘older sibling’), or meB = feB ≠ R, ‘older male sibling’ (Nagovisi *tata* meB = feB ≠ R). This is just a minute sample of the syncretisms in the database, centred around some common syncretisms taking in meB (brother of a man, or literally, male speaking (m) elder (e) brother (B)). The total number of possible syncretisms for the commonly used eight sibling kin types is 4,140, and for all 115 kin types in the database, it is 1.2 x 10138. This gives a huge possibility space, mostly unpopulated, but allows us to characterise kinship systems on many dimensions. Examining all such dimensions is a vast enterprise, beyond what can be done here, so instead we focus on a restricted subset of kin and syncretisms of prima facie interest to the characterisation of Papuan systems.



**Figure 1:** A map of all Papuan languages in Kinbank, with labels for the four system cameos.

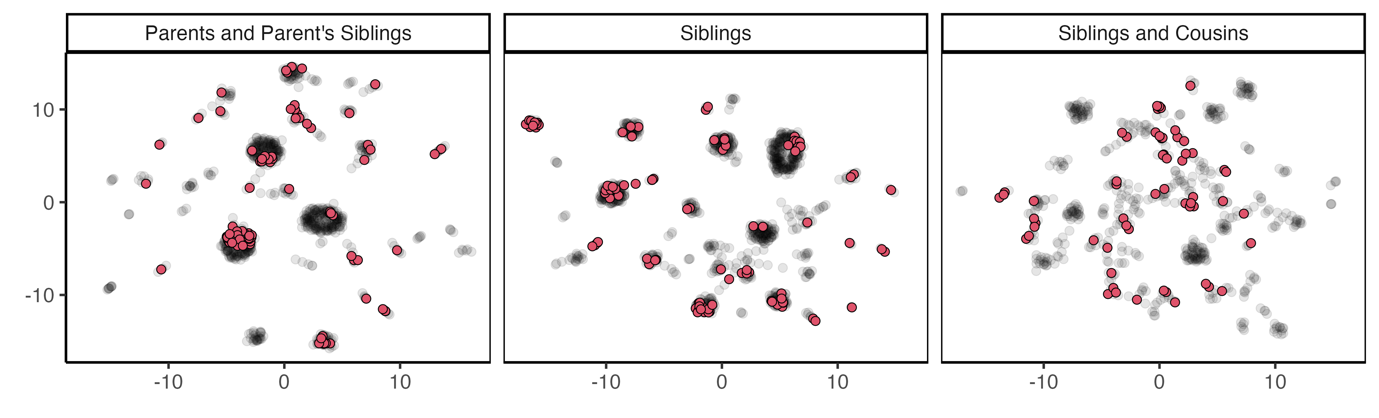
First, we look at how diverse Papuan kinship systems are relative to global variation without concerning ourselves with the specifics. Because the possible space is so vast, we restrict our analyses to three subsets of kin types: Parents and parent’s siblings, siblings, and siblings and cousins). To analyse kin term system structure, we convert the list of kinterms applied to kin types into a string of 1’s and 0’s, by comparing all kin types within our subset of interest to each other and asking if they have the same kinterm (1) or not (0). We call the binary string a *structural vector*. The structural vector is an abstract representation of a language's kinship system structure but ignores variation contained in the kinterms. Using the structural vector, we can ask: how many different structures are amongst our Papuan sample? How does that compare to the totality of global variation? Finally, how does Papuan variation compare to a random sample of languages (the same size as our sample of 109 Papuan languages)? We summarise the answers to these questions in Table 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Papuan Languages | | Kinbank | |  |
| N | Unique Structures | N | Unique Structures | Random Sample |
| Parents & Parent's Siblings | 92 | 40 | 870 | 154 | 31.9 (28.4 - 35.5) |
| Siblings | 99 | 32 | 1,011 | 118 | 28.6 (25.3 - 31.9) |
| Siblings and Cousins | 59 | 51 | 612 | 327 | 43.9 (40.6 - 47.2) |

**Table 1:** The number of languages and unique structures in Papuan languages, and within the entire sample of Kinbank. The final column shows the average number of unique structures in a random sample of languages, from 1000 random samples, with one standard deviation above and below the mean.

The most striking outcome of Table 1 is that in all our subsets, kinship terminology diversity in Papuan languages is more than one standard deviation higher than we would expect in a random sample. In fact, Papuan languages contain between 16% and 25% more diversity than a random selection of languages of the same size. Another way to look at the extent of structural diversity in Papuan kinship terminology is to look at the ratio of structures to languages. The ratio of languages to structures tells us how many languages we have per structure. The closer this ratio is to 1, the closer we are to having a unique structure for every language. In the Kinbank sample Parents and Parent’s siblings have a ratio of 5.6 languages per structure and the Papuan sample has a ratio of 2.3. Siblings have a ratio of 8.5 languages per structure but only 3 in Papuan languages, and Siblings and Cousins have a ratio of 1.9 languages per structure in Kinbank, but Papuan languages are almost completely unique, with a ratio of 1.1. Extrapolating these ratios linearly, or even logarithmically, to the total number of Papuan languages, we use 859, then we could posit that Papuan languages contain more unique kinship system structures than is currently documented in Kinbank.

An assumption implicit in the analyses of structural vectors is that having a different structural organisation is indicative of a fundamental difference, but we might not always think this is the case. For example: we often think of *cousin* in English, as being conceptually similar to *cousine* and *cousin* in French, despite having different syntactic structures. It could be the case that Papuan languages contain only minor differences, accumulated over a long period, but are all cut from similar cloths. We look at the distance between systems structures, calculated from our structural vectors to explore this possibility. We calculate the Manhattan distance between all pairs of structural vectors. Manhattan distance is the sum of the difference between the two vectors. If a language has a 1 in the position where another language has 0, that increases the distance between them by 1. To visualise these distances, we use an algorithm called Uniform Manifold Approximation and Projection (UMAP; (McInnes et al., 2018)). UMAP projects the distances into a two-dimensional space, using manifold learning techniques and logic drawn from topological data analysis (Figure 2).



**Figure 2:** A UMAP projection of global kinship system diversity for Parent and Parent’s siblings, Siblings, and Siblings and Cousins. Red dots are Papuan languages, grey dots are all other languages.

Figure 2 provides us with a visual description of global kinship terminology, and how Papuan kinship systems (red) are distributed across the observed design space. Visually although we see some clusters of Papuan systems, Papuan languages have far-reaching diversity. Average Manhattan distances between languages can quantify this observation. Within Parents and Parent’s siblings, the average structural distance between all languages is 32 (that is 32 different structural properties on average). Papuan languages are slightly more similar, with an average distance of 24. This might be a result of the Papuan sample containing more closely related languages, which are more likely to have similar kinship systems. Indo-European (n = 107) languages are comparably homogenous, with 16 differences on average. Pama-Nyungan languages (105) showed 24 differences on average, but Austronesian languages (377) contained the most differences on average, around 40. Across the other two subsets, siblings, and siblings and cousins, Indo-European kinship systems show the least diversity (Siblings = 4 differences on average; Siblings and cousins[[1]](#footnote-1) = 200), and Austronesian shows the most diversity (Siblings = 12; Siblings and cousins = 336). Papuan languages show more diversity in sibling systems than Pama-Nyungan languages (10 differences on average in Papuan languages, 4 in Pama-Nyungan), but show less diversity in cousin systems (213 differences versus 293).

A broad-scale approach has shown us that Papuan kinship is more diverse than it is within Indo-European contexts but has similar levels of diversity to neighbouring families, in Austronesian and Pama-Nyungan. Although the level of diversity between language families is similar, this does not mean the syncretisms within these families are the same between them. We now explore the patterns of syncretism that characterise Papuan kinship.

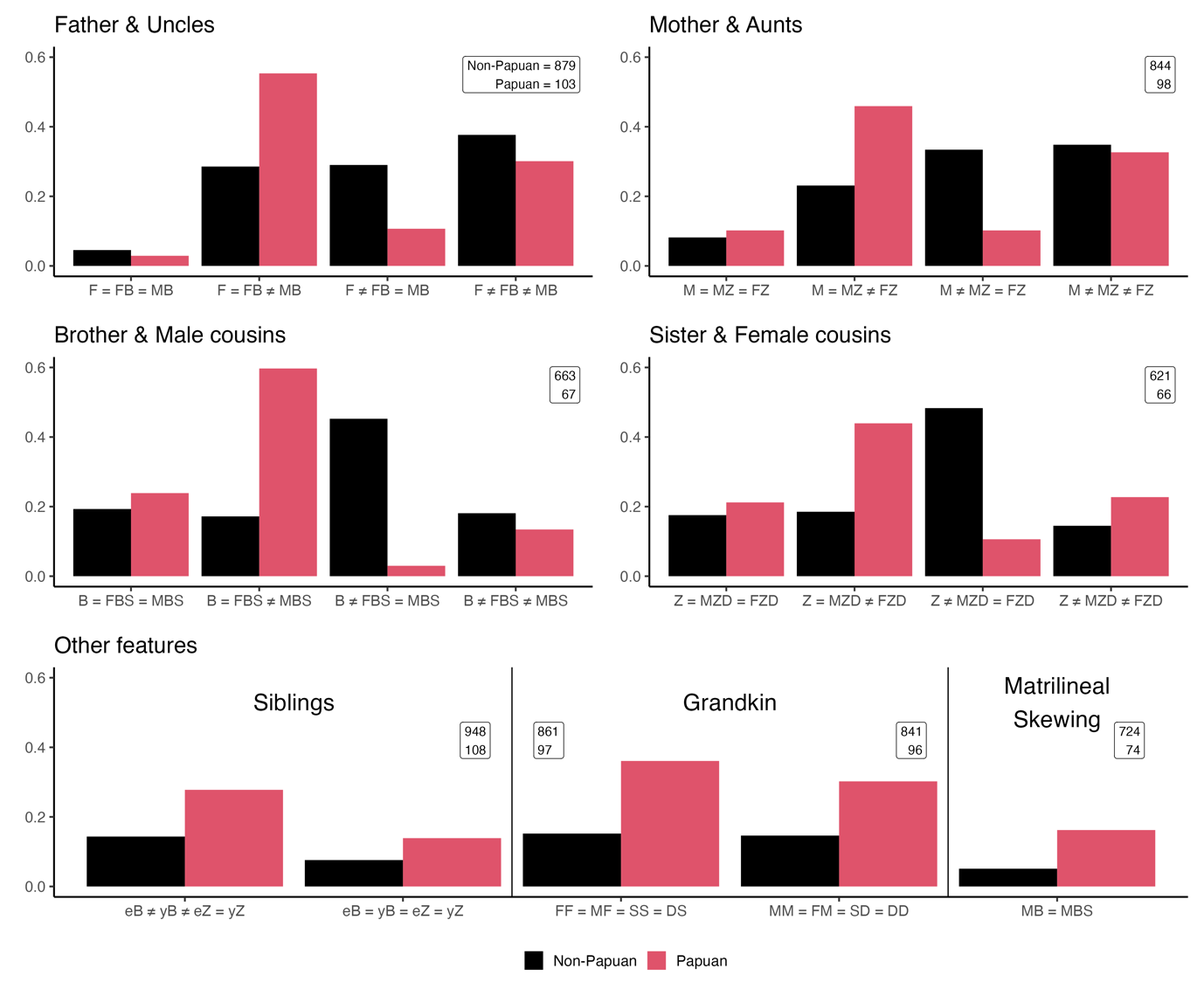
The most prominent feature in Papuan kinship is the bifurcate merging rule. In general, bifurcate merging kinship systems splits (*bifurcates*) a group of kin into two groups based on the gender of the connecting relative, and *merges* lineal and collateral relatives together. For example, in a father/nuncle paradigm we consider three relatives, father, father’s brother, and mother’s brother. In a bifurcate merging system, the mother’s brother is separated from the father and father’s brother because the relative that connects the speaker (ego) to the mother’s brother is a woman (mother) but is connected through a man for father’s brother (father) or is the father. The system then also *merges* the father (a direct lineal relative) with the father’s brother (a collateral relative). A system that does not merge these two relatives is called bifurcate collateral. The pattern of separating relatives connected through opposite genders and merging collateral relatives with lineal relatives can be applied throughout the kinship system. We look at the father/nuncle set, mother/aunt set, brother and male siblings, and sister and female siblings. [[2]](#footnote-2) For each of these three relative sets, there are only five possible organisations, of which we always observe four in both Papuan in non-Papuan languages. [[3]](#footnote-3)

First, we look at the parent/nuncle set, specifically: F = FB ≠ MB. Papuan languages show a disproportionate frequency of bifurcate merging terminology compared to the other three possible systems (Figure 3). The bifurcate merging system makes up 55% of the sample of Papuan languages (n = 57), but only 28% of non-Papuan (n = 251). This system is most prominent in Nuclear Trans New Guinea (n = 21) but occurs in 27 of the 44 language clades for which we have data. Papuan languages show a similar preference for bi-furcate merging in aunt terms (45%; n = 45), which only occurs in 23% of non-Papuan languages (n = 189; Figure 3), and 24 of 44 language clades.

Within sibling and cousin terms, a bifurcate merging pattern occurs in the form of either B = FBS ≠ MBS or Z = MZD ≠ FZD. Focusing on male relatives, 60% of Papuan languages contain this pattern (n = 39), compared to only 17% of the general sample (n = 108). The bifurcate merging structure for female-kin is similarly common, occurring in 43% of Papuan languages (n = 29), but only 18% of non-Papuan languages (n = 108). Bifurcate merging terminology is seen slightly less broadly than the equivalent structures in the generation above, occurring in 18 language clades for the male kin, and 16 for the female, from a total of 44, but as with the parental structures, most languages displaying bifurcate merging terminology are from the Nuclear Trans-New Guinea family.

Despite the prevalence of the bifurcate merging pattern, it does not appear to be a contemporary organising principle of Papuan kinship diversity. Of the 62 Papuan languages, we have data for, only 16 have a bifurcate merging system in all subsets. Of the 97 languages for which we have father/nuncle and mother/aunt data, 38 have bifurcate merging patterns in both systems. Of the 64 languages for which we have cousin data, only 27 have the bifurcate merging pattern in both male and female cousins. We see similar patterns of heterogeneity when assessing the relationship through gender, father / nuncle bifurcate merging systems co-occur with male cousins bifurcate merging systems in 28 of 66 languages, and mother /aunt and female cousin systems in 19 of 62 languages.

Linguistic anthropologists have traditionally aligned bifurcate merging systems with patterns of marriage exchange, or gender-biased migration. By having brothers and sisters in different places, the relationships an individual has with geographically separated kin are likely to be systematically different and therefore linguistically marked. The prevalence of the bifurcate merging system through kinship terminology in Papuan languages asks the question of whether the Proto-Papuan kinship system contained a pattern of exchange or migration that structured the kinship system and whether the lack of coherency we observe is a result of languages drifting away from a homogenous kinship system. Future work might explore the relationship of Papuan languages to social organisation to further this line of enquiry.



**Figure 3:** Bar plots showing the proportion of languages that show particular kinship syncretisms. Numbers in boxes show the total number of non-Papuan and Papuan languages used in each calculation. In the bottom row of graphs, only select structures are shown but the numbers reflect total counts.

Asides from bifurcate merging systems, Papuan languages stand out in other ways. First, Papuan languages show an increased frequency of two different sibling systems, relative to the global sample: One system contains terms for elder brother, younger brother, and a single sister term (27 Papuan languages or 28% of the sample), and the other is a single kinterm for all siblings (14; 14%). Both these structures occur most commonly in Nuclear Trans-New Guinea languages, however, this is also the largest clade. Small counts of languages in Papuan language clade make comparisons difficult. The single-term sibling system has as many languages in the Timor-Alor-Pantar clade as it does in the Trans-New-Guinea, but this count makes up 71% of Timor-Alor-Pantar languages. These systems are found in 13 and 6 Papuan language clades, respectively.

Another common feature of Papuan kinship terminology is reciprocity in grandkin terms, as described in Watam, Nen and Komnzo, and Nagovisi. In taking a broader view, we see that around 34% of our Papuan sample contains reciprocity between grandfathers and grandsons, compared to only 15% of our non-Papuan sample. We see a similar pattern between grandmothers and granddaughters (36% in Papuan languages, 15% in non-Papuan). Although this feature is proportionally more common in Papuan languages than non-Papuan, it is not the most common organisation of grandkin – with most Papuan languages (57%) having separate words for grandfather and grandson. Grandkin reciprocity occurs in 18 languages clades.

Finally, Papuan kinship contains a disproportionate occurrence of matrilineal skewing languages, another feature that is commonly affiliated with patterns of exchange. 16% of Papuan languages contain this feature (n = 12), compared to 5% of non-Papuan languages. Seven of these 12 languages are from the Nuclear Trans-New Guinea clade, largely in Central West Papuan, with a small cluster of languages in the Trans-fly region of Papua New Guinea.

1. Note that the number of possible systems in the set of 40 sibling and cousin kin types is 1.57 x 1035. In Siblings, the total number is only 4,140 and in Parents and Parent’s siblings, it is 5.17 x 1013. Since the possible number of systems is much higher in siblings and cousins, so to are the number of possible differences. [↑](#footnote-ref-1)
2. We could also look at the relative age of parent’s siblings, but these are not important distinctions in Papuan languages. [↑](#footnote-ref-2)
3. The fifth possibility is F = MB ≠ FB which violates the rule of collaterality and has never been observed. [↑](#footnote-ref-3)