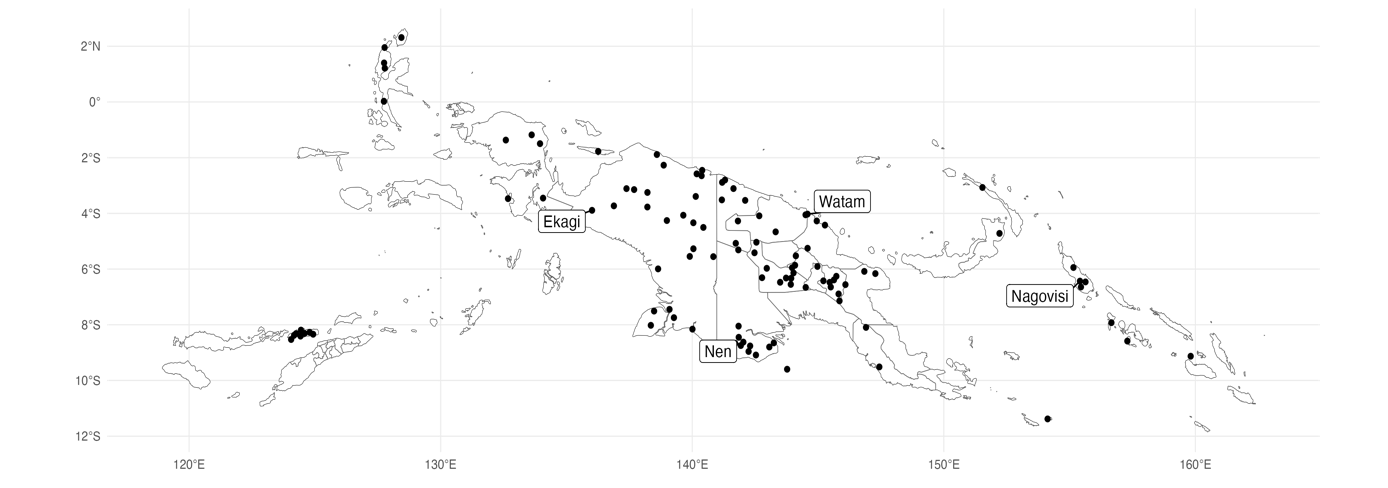
**3. Zooming out to a broader sample**

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 and contrast their diversity against global patterns. Kinbank is a database containing a global sample of kinship terminology from 1,229 languages, built around 115 kin types. Within this dataset are 112 Papuan languages from 36 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 relative to global patterns.

The theoretical bounds of kinship terminology diversity create a vast design space. For example, of our four cameo systems, we see four possible sibling terminology, but these are only four of 4,140 theoretically possible sibling terminology when considering the commonly used eight sibling kin types (Nerlove & Romney, 1967).[[1]](#footnote-1) If we consider all 115 kin types in Kinbank, then there are 1.2 x 10138 possible kinship terminology. This huge possibility space is 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. Instead, we examine subsets of kin and syncretisms of prima facie interest in the characterisation of Papuan systems.



**Figure 1:** A map of the 112 Papuan languages from 36 clades stored 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 parents’ siblings, siblings, and siblings and cousins. For each of these subsets, we consider genealogical categories that are differentiated by relative age (e.g., elder brother vs. younger brother) or by the relative age of connecting kin (father’s elder brother vs. father’s younger brother) and gender of speaker (a man talking to their sister vs. a woman talking to their sister). To analyse kinship terminology 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 the language-specific formal instantiations contained in the kinterms. For example, English grandfather and Russian dedushka, and English grandmother and Russian babushka, have unrelated forms but cover identical sets of kin types, i.e., they have the same patterns of syncretism, so they would receive the same structural vector. 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 the variation from a random worldwide sample of languages (the same size as our sample of Papuan languages)? We summarise the answers to these questions in Table 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| [OLD TABLE KEPT FOR COMPARISON] | ~~Papuan Languages~~ | | ~~Kinbank~~ | |  |
| ~~N~~ | ~~Distinct 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)~~ |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Parents &  Parent's Siblings | Siblings | Siblings & Cousins |
| Papuan | N | 94 | 102 | 61 |
| Distinct Structures | 40 | 33 | 53 |
| Indo-European | N | 88 | 99 | 81 |
| Distinct Structures | 12 | 7 | 22 |
| Austronesian | N | 263 | 329 | 191 |
| Distinct Structures | 50 | 52 | 99 |
| Pama-Nyungan | N | 68 | 98 | 26 |
| Distinct Structures | 36 | 18 | 20 |
| Kinbank | N | 870 | 1,011 | 612 |
| Distinct Structures | 154 | 118 | 327 |
| Rand. Sample | 32.37  (28.93 - 35.82) | 29.26  (25.94 - 32.59) | 45.04  (41.63 -48.45) |

**Table 1:** For each subset of kin types (columns), we show the number of languages and unique structures for our sample of Papuan languages, for Indo-European, Austronesian, and Pama-Nyungan languages held in Kinbank, and within the entire sample of Kinbank. The final row shows the average number of unique structures in a random worldwide sample of languages, drawn from 1000 random samples, with one standard deviation above and below the mean.

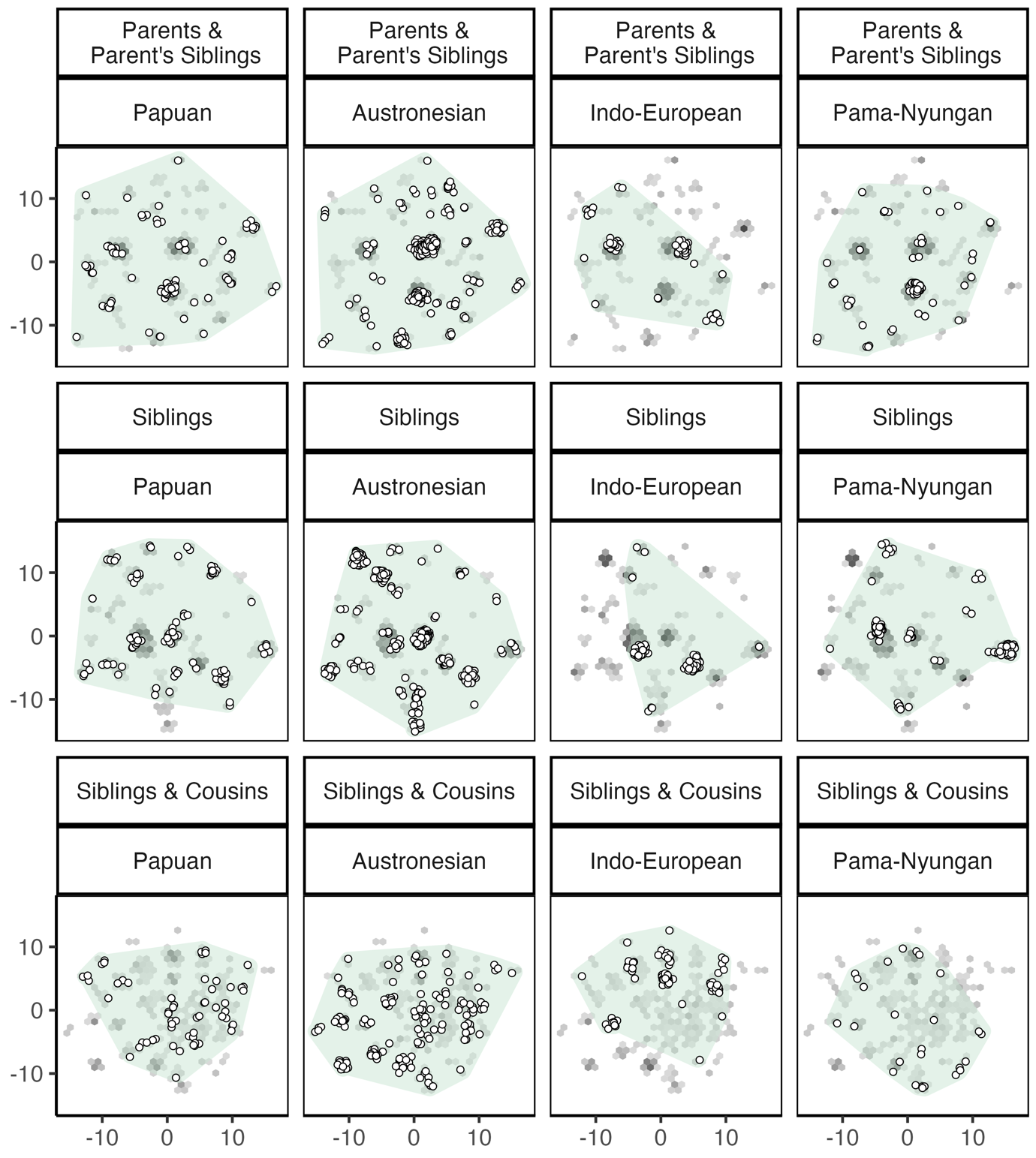
The most striking outcome of Table 1 is that in all our subsets, kinship terminology diversity in Papuan languages is more than one standard deviation higher than we would expect in a random sample. In fact, Papuan languages contain between 16% and 25% more diversity than a random selection of languages of the same size.

To compare the level of diversity in Papuan languages to the global pattern, and to other families, we use the ratio of structures to languages. The ratio of languages to structures tells us how many languages we have per distinct kinship terminology structure and allows us to compare subsets of languages of different sizes.[[2]](#footnote-2) The closer this ratio is to 1, the closer we are to having a unique structure for every language. In the entire 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. Papuan languages show more diversity than the general sample. This effect is greater when compared to Indo-European,with a ratio of 7.3. Austronesian languages show comparable diversity to the global sample, a ratio of 5.3. Pama-Nyungan shows the lowest ratio, 1.9, and the highest level of diversity.

Siblings show a ratio of 8.5 languages per structure in the global sample, but only 3.0 in Papuan languages. Indo-European again shows limited diversity with a larger of 14.1, Austronesian has a ratio of 6.3, and Pama-Nyungan has a ratio of 5.4. Finally, Siblings and Cousins have a ratio of 1.9 languages per structure in Kinbank, but Papuan languages are almost unique, with a ratio of 1.1. Again, Indo-European shows the most limited diversity, with a ratio of 3.7, Austronesian has a ratio of 1.9, and Pama-Nyungan contains a comparable level of diversity, a ratio of 1.3. In general, Papuan kinship shows a higher level of diversity than global, Indo-European, and Austronesian samples, and comparable levels of diversity to Pama-Nyungan languages.

An assumption implicit in the analyses of structural vectors is that each formal difference, even if small, counts as an equally important difference in semantic structure, but we might not always think this is the case. As an example, consider the four possible kin types referring to male siblings (elder, younger, and whether a male or female is speaking), and the systems used in English, Nagovisi, and Watam. In English, there is one term that glosses over all four possible categories*.* In Nagovisi there are two terms, one for elder brother and one for younger brother. Finally, In Watam there are three terms elder same-sex sibling, younger same-sex sibling, and opposite-sex sibling. These are three distinct sibling systems, but we also might think that a system with a term for brother is more similar to a system with a word for elder and younger brother than it is to a system with terms based on relative sex. This example helps us think of kinship terminology systems existing within a space of possible designs where the distance between systems is indicative of similarity. This is sometimes thought of as a *design space.*

Papuan kinship contains more distinct systems than we would expect, given the number of languages, but do Papuan kinship systems cover more of the kinship *design space* than we might expect? A metric that approximates this idea in ecology is called “functional richness”, which quantifies the area occupied by a particular species defined by a set of measured features (Villéger et al., 2008), and has been used in linguistics to quantify the amount of grammatical diversity at risk from languages no longer being spoken (Skirgård et al., 2023). We calculate the functional richness of Papuan languages by first calculating the Manhattan distance between all pairs of languages. Manhattan distance is the sum of the difference between the two vectors and thus gives us an idea of the degree of formal difference between any pair of terms. If a language has a 1 in the position where another language has 0, that increases the distance between them by 1. We then summarise the distance matrix into two dimensions using 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). We use the projection of the distance matrix to calculate the area covered by each language group, in the form of convex hulls. The area of the convex hulls (green) allows us to quantify how much of the total space is covered by a subset of languages, compared to the total diversity. Accounting for all languages, we would have a volume of 1. Scores less than one can be considered a proportion of the total area covered.

****

**Figure 2:** An MDS projection of global kinship system diversity for Parent and Parent’s Siblings, Siblings, and Siblings and Cousins. Grey hexagons show the frequency of languages across the whole database in a particular area, white circles show the distribution of Papuan languages. The green convex hull shows the totality of the space covered by language groupings.

Figure 2 provides us with a visual description of global kinship terminology diversity, and how Papuan kinship systems (white circles) are distributed across the observed design space compared to other language families. Although there are clusters of Papuan systems, Papuan languages are scattered widely across the observed space. In Parents and Parent’s siblings, Papuan languages cover approximately 0.74 of the space, which is comparable to the coverage seen in Pama-Nyungan (0.75). Austronesian languages cover almost the entire diversity space (area = 0.91), whereas Indo-European languages show the lowest level of diversity (0.53). Papuan sibling diversity covers a large portion of observed diversity (0.96), as does Austronesian sibling diversity (0.92). Pama-Nyungan sibling terminology has a functional diversity score of 0.52. Indo-European sibling terminology shows the lowest level of coverage again, with a score of 0.30. Expanding the sibling category to include cousins creates a much larger theoretical and observed possibility space, and consequently reduces the diversity seen in Papuan languages (0.55). Pama-Nyungan languages show the same level of diversity (0.55). Indo-European languages cover a larger portion of the sibling and cousin space (0.69). This high functional diversity score is attributed to the two extremes of cousin organisation occurring in Indo-European, but as shown in Figure 2, the coverage is sparse. In Western Europe (English, French) it is common to see a lineal pattern, separating lineal relatives (mother, father) from collateral relatives (parent’s siblings, cousins), but in Indo-European languages east of Europe, we observe more diversity in cousin organisation, and a higher prevalence of cross-cousin marriage (Schulz et al., 2019). Finally, Austronesian sibling and cousin terminology, again, covers a large portion of the observed diversity (0.90).

Testing the functional richness of kinship has shown that Papuan kinship is more diverse than Indo-European contexts but has similar levels of diversity to neighbouring language families, in Austronesian and Pama-Nyungan. Since Austronesian and Pama-Nyungan are clearly defined language families and the Papuan category contains many unrelated language groups, the diversity in kinship systems within Papuan languages is lower than we might expect if we assumed kinship terminology variation was unbounded.

Anthropologists and cognitive scientists point to the influence of social and cognitive constraints on kinship terminology diversity (Kemp & Regier, 2012; Murdock, 1949; Passmore & Jordan, 2020). These theories can explain why it is unlikely that we will ever observe all theoretical possibilities of kinship organisation and give reason to believe that kinship variation is not unlimited. However, it leaves open the question of why Papuan languages show an unusually high ratio of languages to distinct structures, but moderate levels of functional diversity. These are seemingly contrasting conclusions. In part, conclusions are limited by the available samples. We analyse a small subset of Papuan languages (~100 languages from a total of ~850), whereas we have a large sample of Austronesian languages (~~330 languages from a total of ~1,200). Sample sizes between Papuan, Pama-Nyungan, and Indo-European languages were comparable. Sampling more Papuan languages could introduce languages that inhabit areas of the functional space that our current sample does not. It could also be that social organisation diversity constrains the functional diversity in kinship terminology amongst Papuan languages. Assuming that social organisation is a limiting factor in kinship systems, and therefore kinship terminology, if Papuan societies contain less variable kinship systems, we would also expect them to contain less terminological diversity. Without paired social data it is not possible to tell whether diversity in social organisation is correlated with terminological diversity, however, speaker populations for Papuan languages are smaller and vary less than in Austronesian societies.

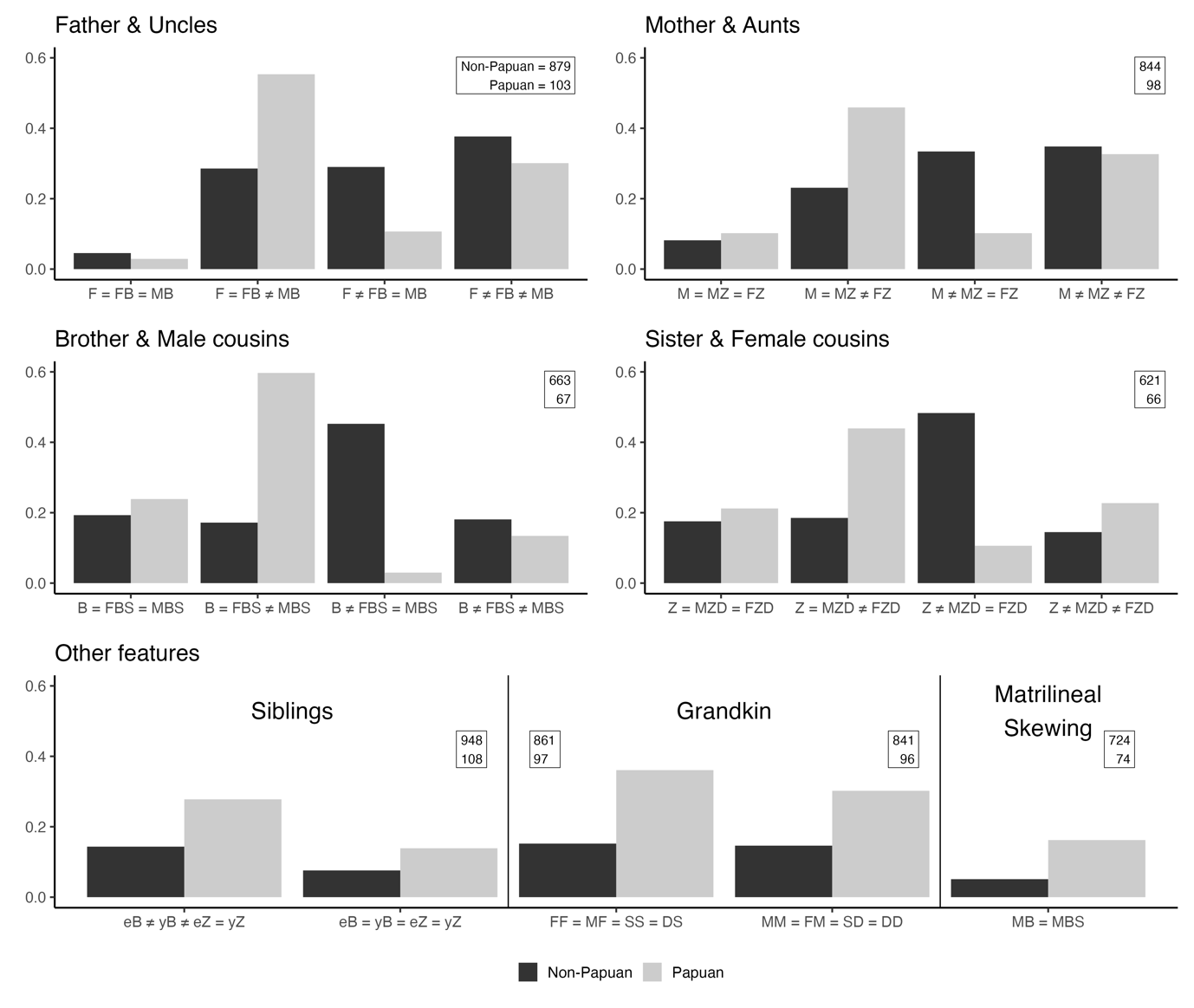
The analysis of high-level diversity has shown that Papuan systems contain many distinct systems, relative to sample size, but moderate levels of functional diversity. We now explore the specific patterns of syncretism that characterise Papuan kinship.

The most prominent feature in Papuan kinship is the bifurcate merging rule we mentioned in the introduction. 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. [[3]](#footnote-3) 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. [[4]](#footnote-4)

First, we look at the parent/nuncle set, specifically: F = FB ≠ MB. Papuan languages show a heavy predominance 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), compared to 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, the corresponding pattern to a bifurcate merging parent terminology 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 coherent organising principle of Papuan kinship diversity. Of the 62 Papuan languages, we have data for, only 16 have a bifurcate merging system in the four subsets of kin analysed (father/nuncles, mothers/aunts, brothers/male cousins, and sisters/female cousins). Of the 97 languages for which we have father/nuncle and mother/aunt data, 38 have bifurcate merging patterns in both subsets of kin, and 21 show the pattern in either male or female relatives. Of the 64 languages for which we have cousin data, 27 have the bifurcate merging pattern in both male and female cousins, and 13 show the pattern in either male or female cousins. We see similar patterns of heterogeneity when assessing the relationship across generations, but within genders: 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. As was discussed in the Nen/Komnzo case, and as we have discussed elsewhere, bifurcate merging patterns do not have to be coherent throughout the system for a system to be functional (Passmore et al., 2021). The inconsistency of merging patterns appears to be a feature of Papuan kinship.



**Figure 3:** Bar plots showing the proportion of languages that show 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 a higher frequency of two particular sibling systems, relative to the global sample. One system contains three terms, one each for elder brother, younger brother, and a single sister term (27 Papuan languages or 28% of the sample). The other is a single kinterm for all siblings (14; 14%). The three-term system is most common in the Trans-New-Guinea family (7 languages), with a secondary cluster in North Halmaheran languages (4 languages). In general, this pattern is spread widely across the Papuasphere. In total, this system is found in 15 clades. The single-term sibling system mostly occurs between the Timor-Alor-Pantar clade (6 languages) and the Trans-New-Guinea (5 languages). The Timor-Alor-Pantar clade is considerably smaller than Trans-New-Guinea, so these counts make up 71% of sampled Timor-Alor-Pantar languages and only 5% of sampled Trans-New-Guinea languages. Timor-Alor-Pantar family has recently been shown to be a part of the Trans-New-Guinea family (REF). In total, this sibling pattern is found in six clades.

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 13 Papuan languages clades.

Finally, Papuan kinship contains a disproportionate occurrence of matrilineal skewing (MB = MBS), another feature that is commonly affiliated with patterns of exchange and with bifurcate merging patterns. 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.

**References**

Kemp, C., & Regier, T. (2012). Kinship Categories Across Languages Reflect General Communicative Principles. *Science*, *336*(6084), 1049–1054. https://doi.org/10.1126/science.1218811

McInnes, L., Healy, J., Saul, N., & Grossberger, L. (2018). UMAP: uniform manifold approximation and projection. *The Journal of Open Source Software*, *3*(29), 861.

Murdock, G. P. (1949). *Social structure* (Vol. xvii). Macmillan.

Nerlove, S., & Romney, A. K. (1967). Sibling terminology and cross-sex behavior. *American Anthropologist*, *69*(2), 176–187.

Passmore, S., Barth, W., Quinn, K., Greenhill, S. J., Evans, N., & Jordan, F. M. (2021). Kin Against Kin: Internal Co-selection and the Coherence of Kinship Typologies. *Biological Theory*. https://doi.org/10.1007/s13752-021-00379-6

Passmore, S., & Jordan, F. M. (2020). No universals in the cultural evolution of kinship terminology. *Evolutionary Human Sciences*, *2*. https://doi.org/10.1017/ehs.2020.41

Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science*, *366*(6466), eaau5141. https://doi.org/10.1126/science.aau5141

Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latarche, J. J., Lesage, J., Weber, T., Witzlack-Makarevich, A., Passmore, S., Chira, A., Maurits, L., Dinnage, R., Dunn, M., Reesink, G., Singer, R., Bowern, C., Epps, P., Hill, J., … Gray, R. D. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances*, *9*(16), eadg6175. https://doi.org/10.1126/sciadv.adg6175

Villéger, S., Mason, N. W. H., & Mouillot, D. (2008). New Multidimensional Functional Diversity Indices for a Multifaceted Framework in Functional Ecology. *Ecology*, *89*(8), 2290–2301. https://doi.org/10.1890/07-1206.1

1. Elder brother, elder sister, younger brother, younger sister, and, for each term, whether it is a man or woman speaking. [↑](#footnote-ref-1)
2. The influence of sample size is not completely mitigated because there is a finite number of possible systems. As the number of distinct structures increases, the probability a new structure arises logically decreases. [↑](#footnote-ref-2)
3. We could also look at the relative age of parents’ siblings, but these are not important distinctions in Papuan languages. [↑](#footnote-ref-3)
4. The fifth possibility is F = MB ≠ FB which violates the rule of colaterality and has never been observed. [↑](#footnote-ref-4)