Micronesian trees and waves: a new look at the lexical evidence

Lev Blumenfeld, Carleton University

COOL 13, June 2025

Outline

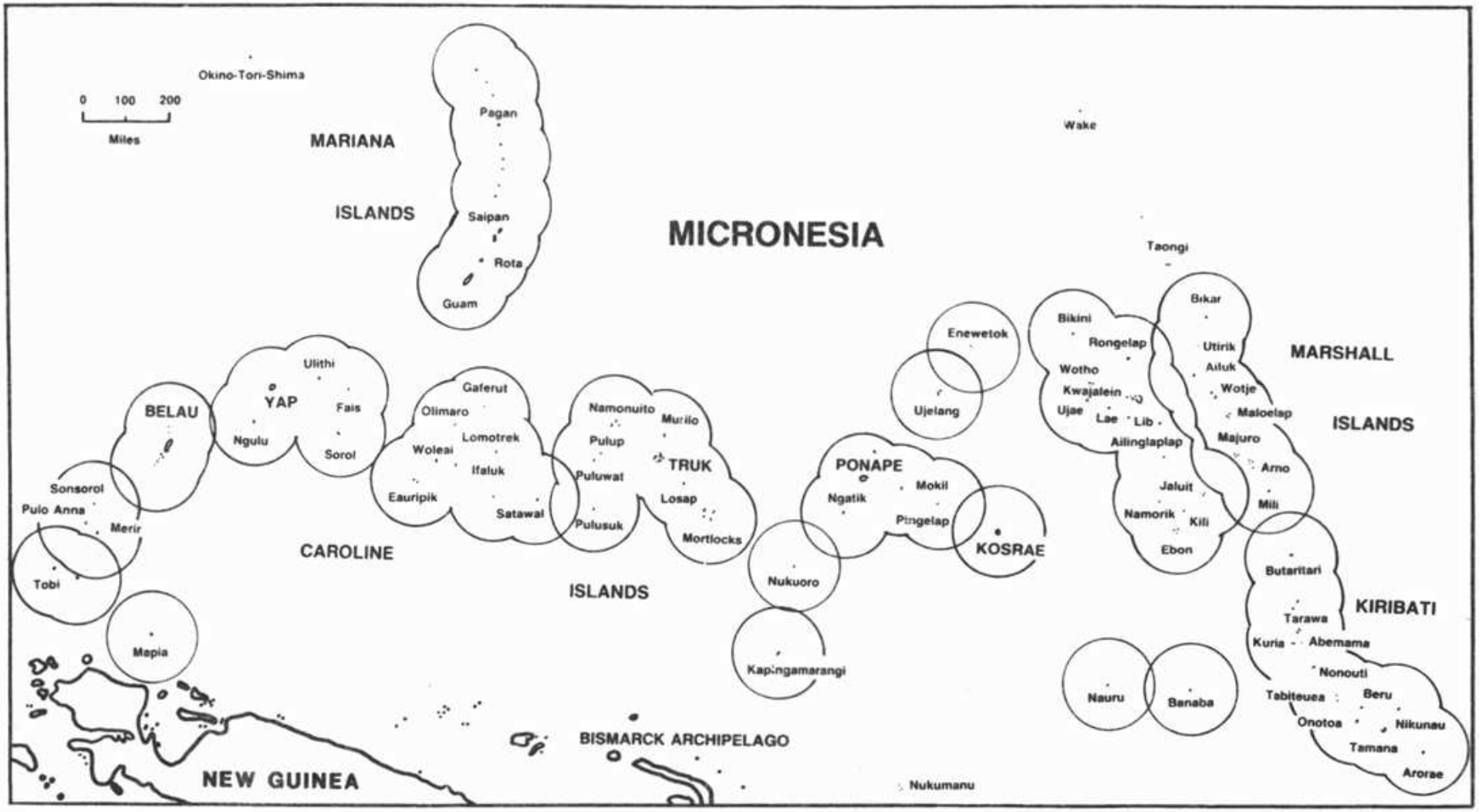


FIGURE 1.

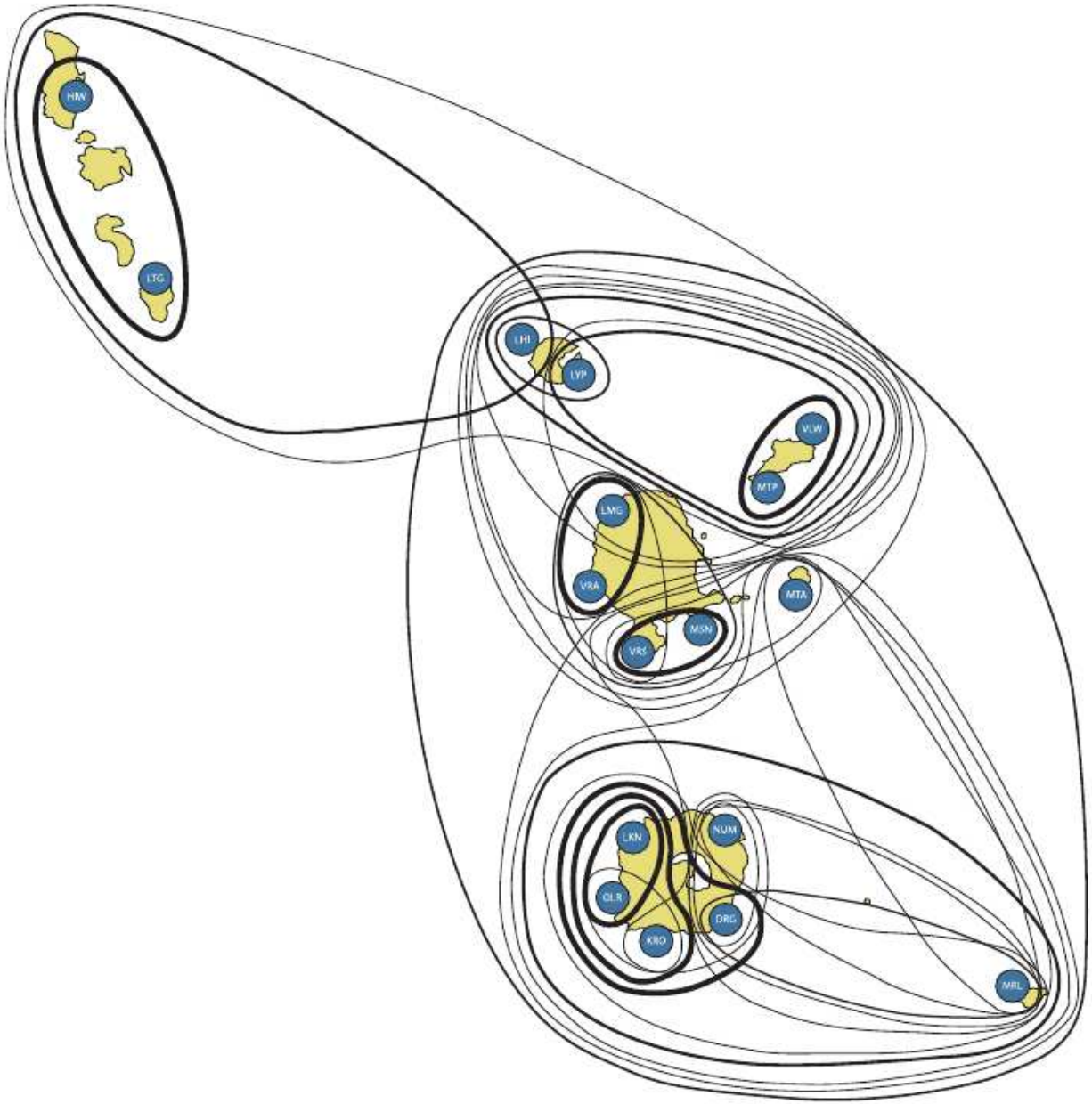
(Marck 1986)

Outline

Introduction

Mechanisms by which dialect groups arise (e.g. Ross 1988, see also Jacques and List 2019; Kalyan and François 2019):

- Separation: loss of contact between parts of a formerly unitary community, resulting in a tree-like structure
- Differentiation: gradual accumulation of differences between communities in contact, resulting in a wave-like structure, or linkage (Ross 1988).
- While Austronesian shows tree-like branching at the higher levels, low-level subgroups of Oceanic are linkages
- "[I]n early Oceanic, dialect differentiation and network-breaking were the rule rather than the exception" (Pawley and Ross 2006)
  - Fijian (Geraghty 1983), Western Melanesia (Ross 1988), Malaita (Lichtenberk 1988), Vanuatu (Clark 2009; François 2011a,b), Polynesian (Walworth 2014, Walworth and Davletshin 2019) etc.



A glottometric diagram of North Vanuatu (Kalyan and François 2019, p. 172)

Micronesian

- No reason to think that Micronesian is any different, based on archeological and (some) linguistic evidence
- Near-simultaneous settlement of major high islands about 2000BP
- Maintenance of interisland contacts for at least some time
- Eventual network-breaking and differentiation
  (Athens 2018; Kirch 2017)

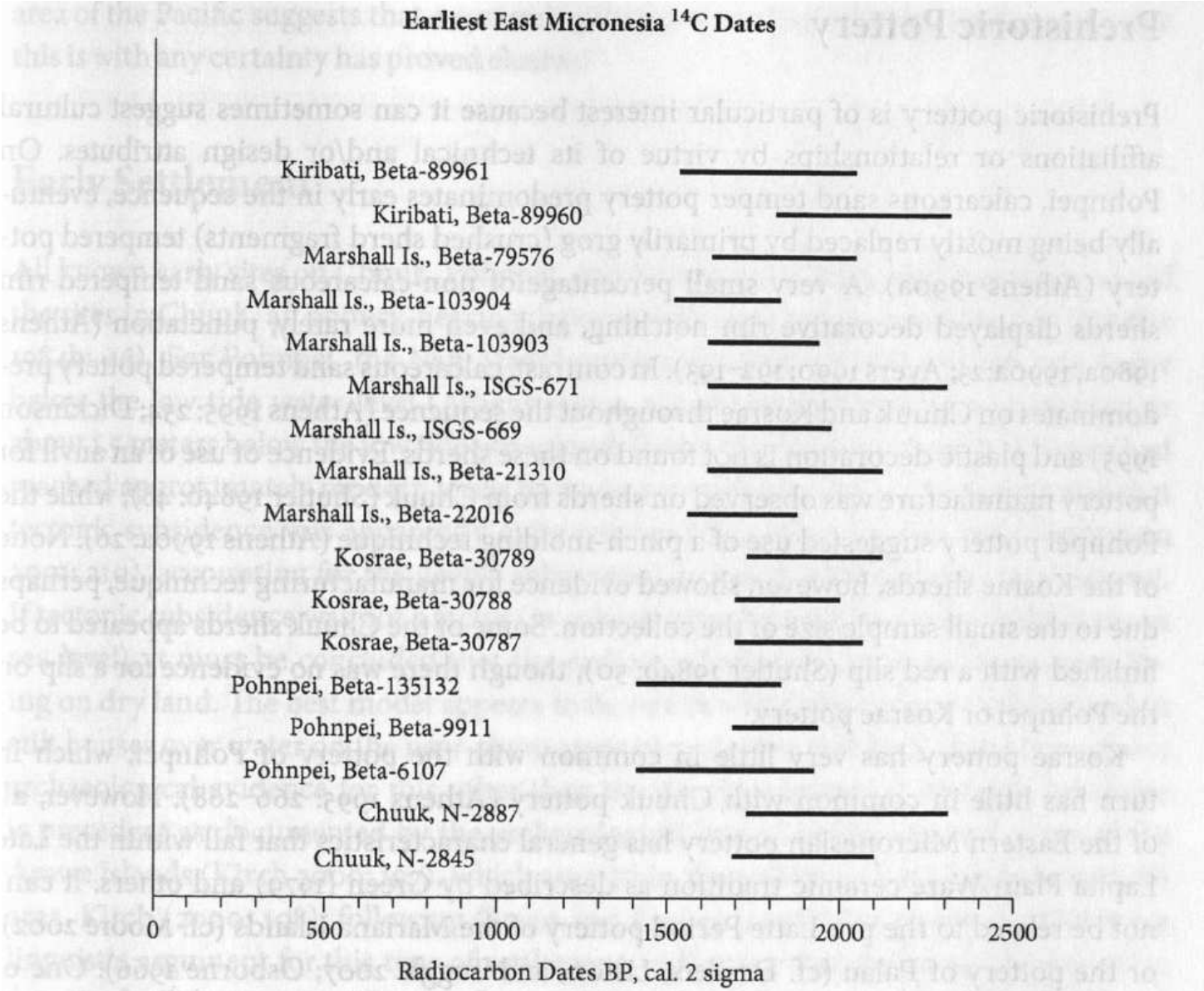
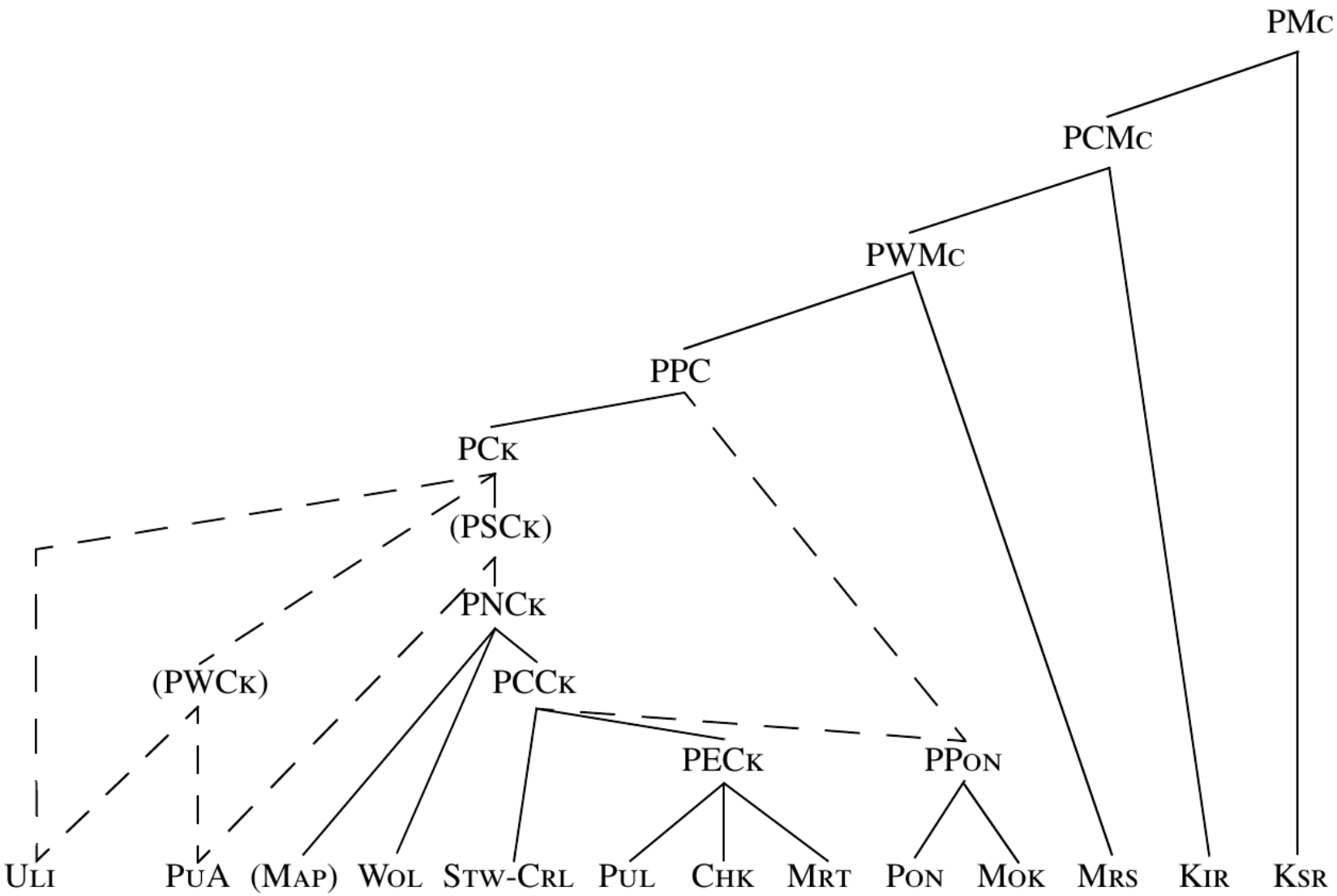


FIGURE 13.3 Graph of the two sigma calibrated ranges of the earliest archaeological radiocarbon determinations in eastern Micronesia. They collectively suggest that initial settlement over the entire region occurred between about 1,800 and 2,000 cal. B.P.

(Athens 2018:283)

Micronesian



Micronesian language family, Bender et al. 2003a, based on Jackson 1983

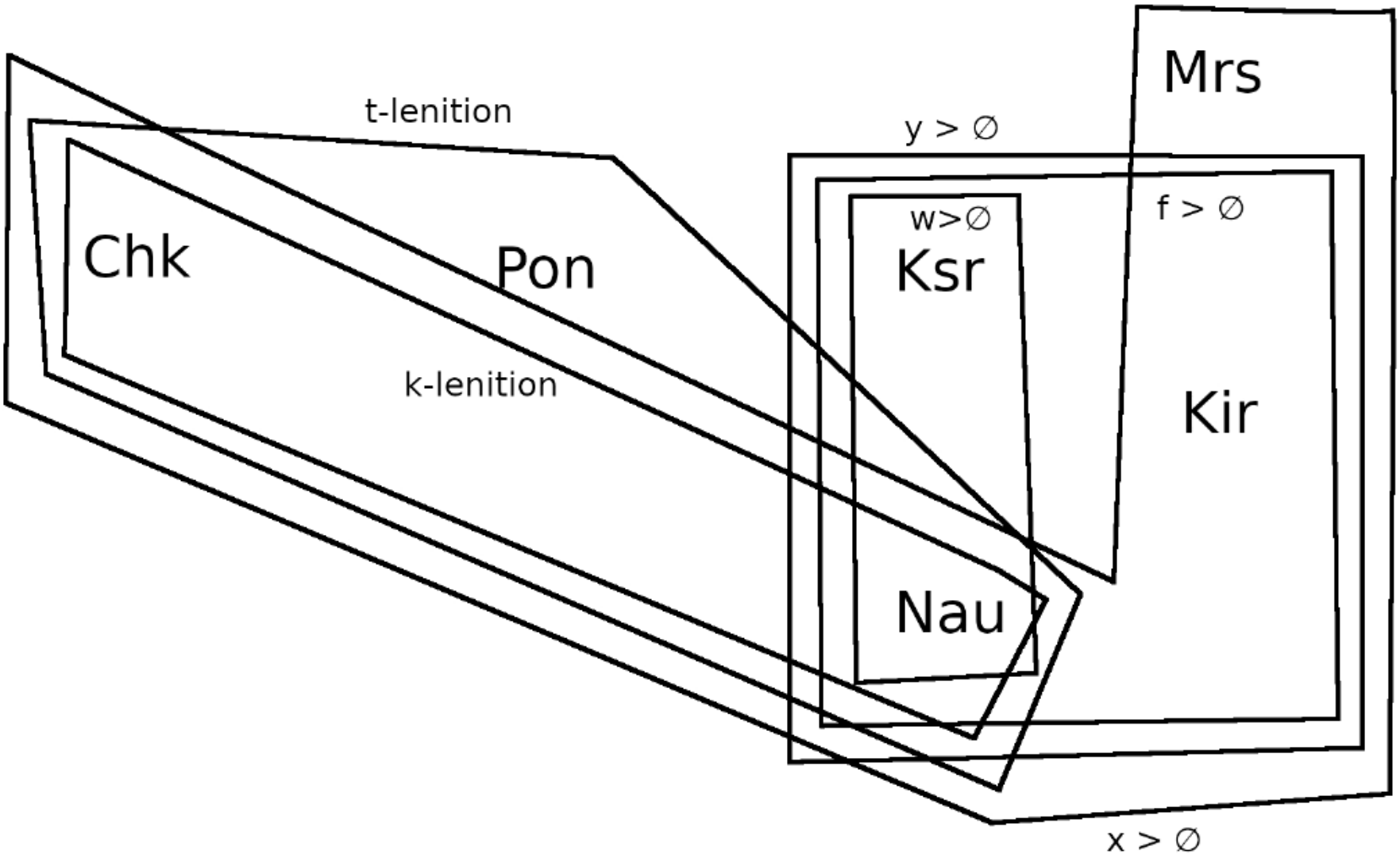
Micronesian

A literal reading of the tree implies that "Micronesia was settled by a series of discrete moves through the islands, and that at each point where we identify a subgroup, there was a pause of sufficient duration to allow a unique set of innovations to develop by which we can identify the subgroup. This seems to me to be unlikely as an account of the settlement of the entirety of this region." (Rehg 1995, emphasis mine)

Other works that critically examine the structure of the tree:

- Marck 1986, 1994, Hughes 2020a,b, Blumenfeld 2022, 2024

Evidence from sound change



Blumenfeld 2022

Outline

The MCD (Micronesian Comparative Dictionary)

- Individual language dictionaries: Pohnpeian, Woleaian, Chuukese, Carolinian, Marshallese, Kosraean, started in the 1970s, some ongoing
- Comparative work culminating in Bender et al. 2003a,b
- Steve Trussel's online version of Bender et al. 2003a
  - `https://www.trussel2.com/MCD/`
- CLDF conversion undertaken last year: Bender et al. 2023, as described in Smith et al. 2025
  - `https://zenodo.org/records/7973420`
- Work in progress: integration with the ACD (Blust et al. 2023)
- Goal for today: a preliminary look at what the MCD lexical data tells us about the internal subgrouping of Micronesian languages

Languages

- The analysis below includes the following languages: Kosraean, Kiribati (Gilbertese), Marshallese
- Pohnpeic (3 languages) and Chuukic (about 10 languages) are (relatively) well-supported subgroups within Micronesian, and are treated as single languages
- Nauruan: comparisons from Blumenfeld 2022 are included, but there are very few pairwise comparisons between Nauruan and other languages, which introduces a bias against Nauruan subgrouping with any individual language (more on these issues below)

Outline

Glottometry (Kalyan and François 2018)

For each subset of languages:

- e: exclusive isoglosses
- p: supporting (possibly non-exclusive) isoglosses
- q: conflicting isoglosses

Subgroupiness: the proportion of supporting isoglosses among all relevant isoglosses, scaled by the number of exclusive isoglosses

Subgroupiness (Kalyan and François 2018, p. 70)

$$s = e\frac{p}{p+q}$$

Calibration

- Raw counts depend on data quality: a language with good coverage will enter into more isoglosses
- Solution: compare observed distribution of reflexes to a statistically random distribution
- I.e. pretend reflexes are the same in number as in the MCD but are randomly distributed between cognate sets, and calculate the expected value for each of K&F's measures for each subset of languages: expected number of exclusive, supporting, and conflicting isoglosses
- "Subgroupiness" was obtained by applying Kalyan & François' formula to the observed/expected ratios for these measures, not raw numbers

Two issues with lexical methods

(A) Innovation vs. retention

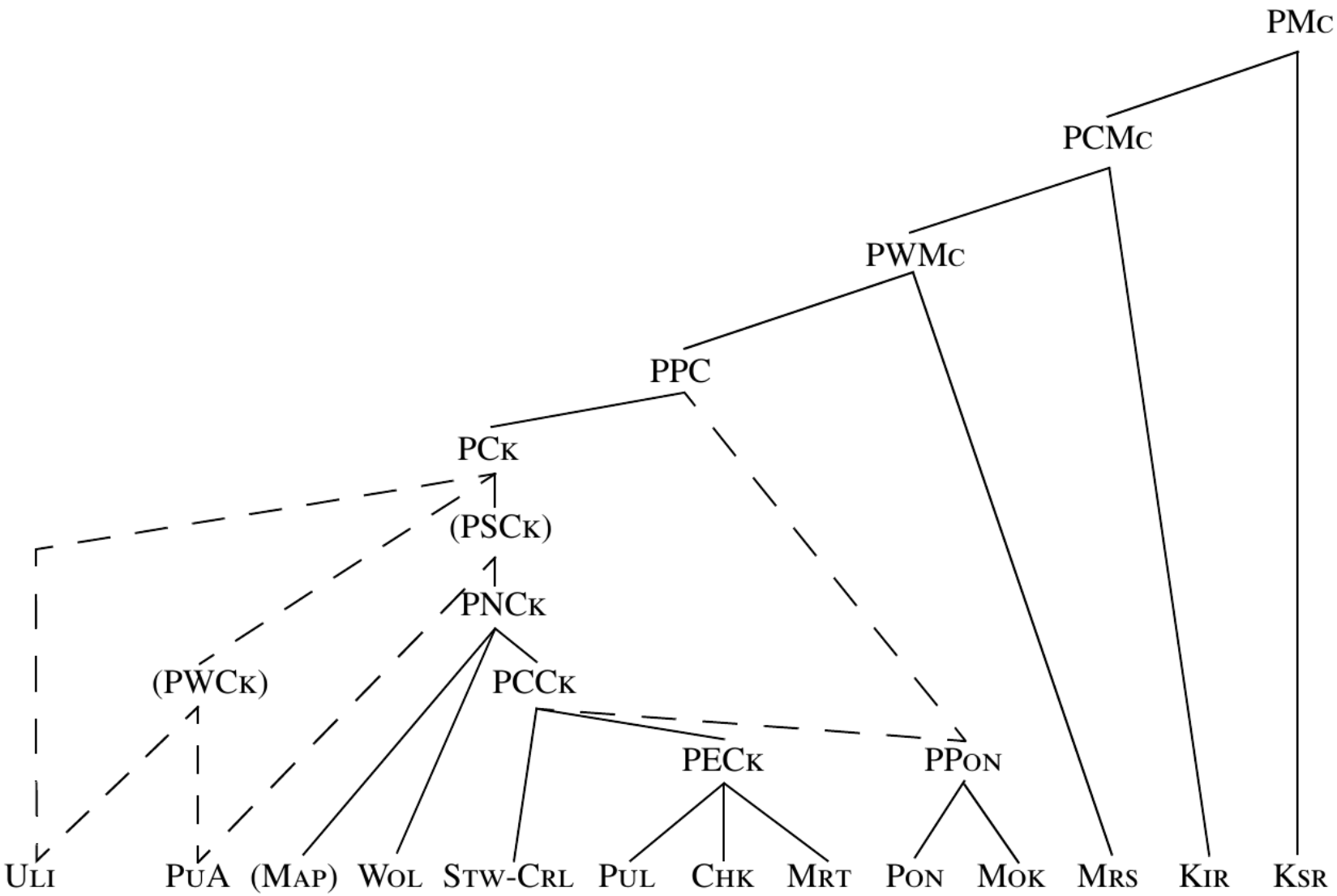
- To avoid retention-based cognate sets, we only consider MCD sets without external comparison (1125 sets out of 1723)
- This does not entirely exclude retention-based isoglosses but reduces their number by removing sets guaranteed to be retention-based

(B) Geographical contiguity

- Particularly relevant to genetic relationships are subgroups which are not geographically contiguous
- In Jackson's tree, Kosraean branches before Kiribati, but the geography would suggest the opposite order
- Thus, of special interest are the non-contiguous Central and Western Micronesian subgroups implied by Jackson's tree

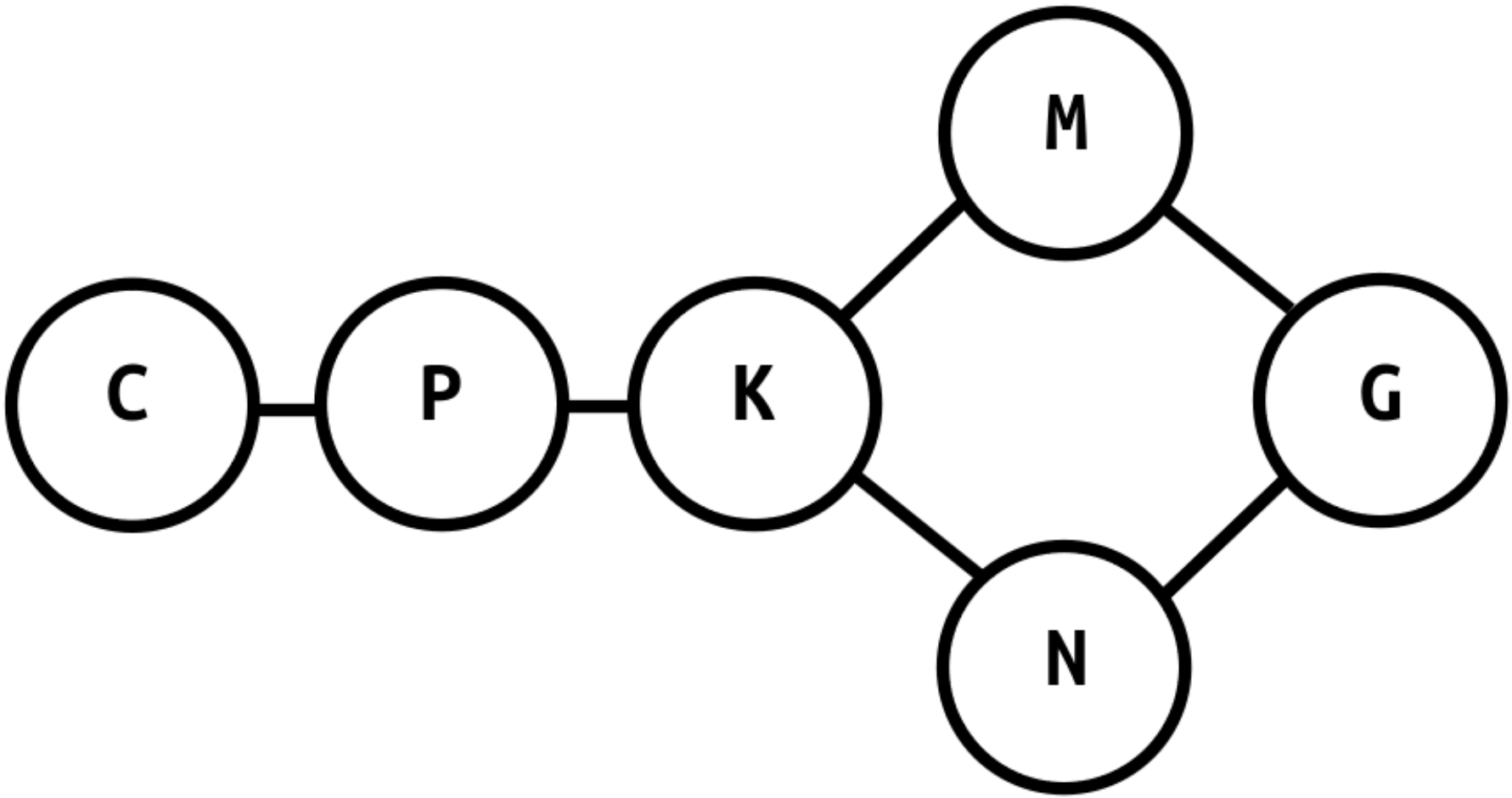
(Cf. Donohue et al. 2012; Smith 2023)

Micronesian



Micronesian language family, Bender et al. 2003a, based on Jackson 1983

Contiguity assumptions



Questions

- (A) Do subgroups implied by this tree show up in the data? Particularly the geographically non-contiguous ones?
  - Central Micronesian
  - Western Micronesian
  - Pohnpeic-Chuukic
- (B) More generally, is there *any* tree-like signal in the data?
  - Do some non-overlapping subsets of languages stand out as having high subgroupiness, and lower-than-expected number of conflicting isoglosses?
  - Is there a negative correlation between conflicting and supporting isoglosses?
- Spoiler: (A) probably not, and (B) maybe

Outline

Is Central Micronesian a strong subgroup?

SUBG	*e*	*p*	*q*	*s*	contiguity
CPMK	4.13	7.33	1.14	3.57	y
CPMG	3.08	6.56	1.16	2.62	n
CPGK	1.75	6.55	1.18	1.48	n
CMGK	1.6	7.88	1.22	1.38	n
PMGK	1.33	10.17	1.31	1.18	y

C=Chuukic; P=Pohnpeic; M=Marshallese; G=Gilbertese; K=Kosraean

e: exclusive

p: supporting

q: conflicting

s: subgroupiness

Is Western Micronesian a strong subgroup?

SUBG	*e*	*p*	*q*	*s*	contiguity
PMK	1.63	4.63	1.15	1.3	y
CPK	1.24	3.14	1.15	0.91	y
CPM	1.26	2.97	1.19	0.9	n
CMG	1.2	3.29	1.18	0.88	n
CMK	0.92	3.52	1.27	0.68	n
MGK	0.8	4.29	1.11	0.64	y
CPG	0.9	2.66	1.13	0.63	n
PMG	0.67	3.64	1.29	0.49	n
CGK	0.5	2.97	1.19	0.36	n
PGK	0.25	3.48	1.26	0.18	n

C=Chuukic; P=Pohnpeic; M=Marshallese; G=Gilbertese; K=Kosraean

e: exclusive

p: supporting

q: conflicting

s: subgroupiness

Two-member subgroups

SUBG	*e*	*p*	*q*	*s*	contiguity
CG	1.32	1.78	1.42	0.73	n
CP	1.31	1.74	1.5	0.71	y
CM	0.63	1.66	1.57	0.32	n
CK	0.56	1.64	1.49	0.29	n
PM	0.29	1.84	1.43	0.17	n
MG	0.25	1.93	1.32	0.15	y
PK	0.24	1.95	1.34	0.14	y
MK	0.15	2.19	1.23	0.098	y
GK	0.17	1.69	1.31	0.094	n
PG	0.13	1.53	1.52	0.065	n

C=Chuukic; P=Pohnpeic; M=Marshallese; G=Gilbertese; K=Kosraean

e: exclusive

p: supporting

q: conflicting

s: subgroupiness

Nauruan

(Too little data for anything conclusive)

SUBG	*e*	*p*	*q*	*s*	contiguity
CNG	2	4.56	1.12	1.6	n
PKN	1	14	0.98	0.83	y
CNM	0.67	3.4	1.24	0.49	n
CPN	0.6	3.38	1.1	0.45	n
PNG	1	12.7	1.04	0.41	n
CKN	0.33	3.83	0.94	0.36	n

C=Chuukic; P=Pohnpeic; M=Marshallese; G=Gilbertese; N=Nauruan

e: exclusive

p: supporting

q: conflicting

s: subgroupiness

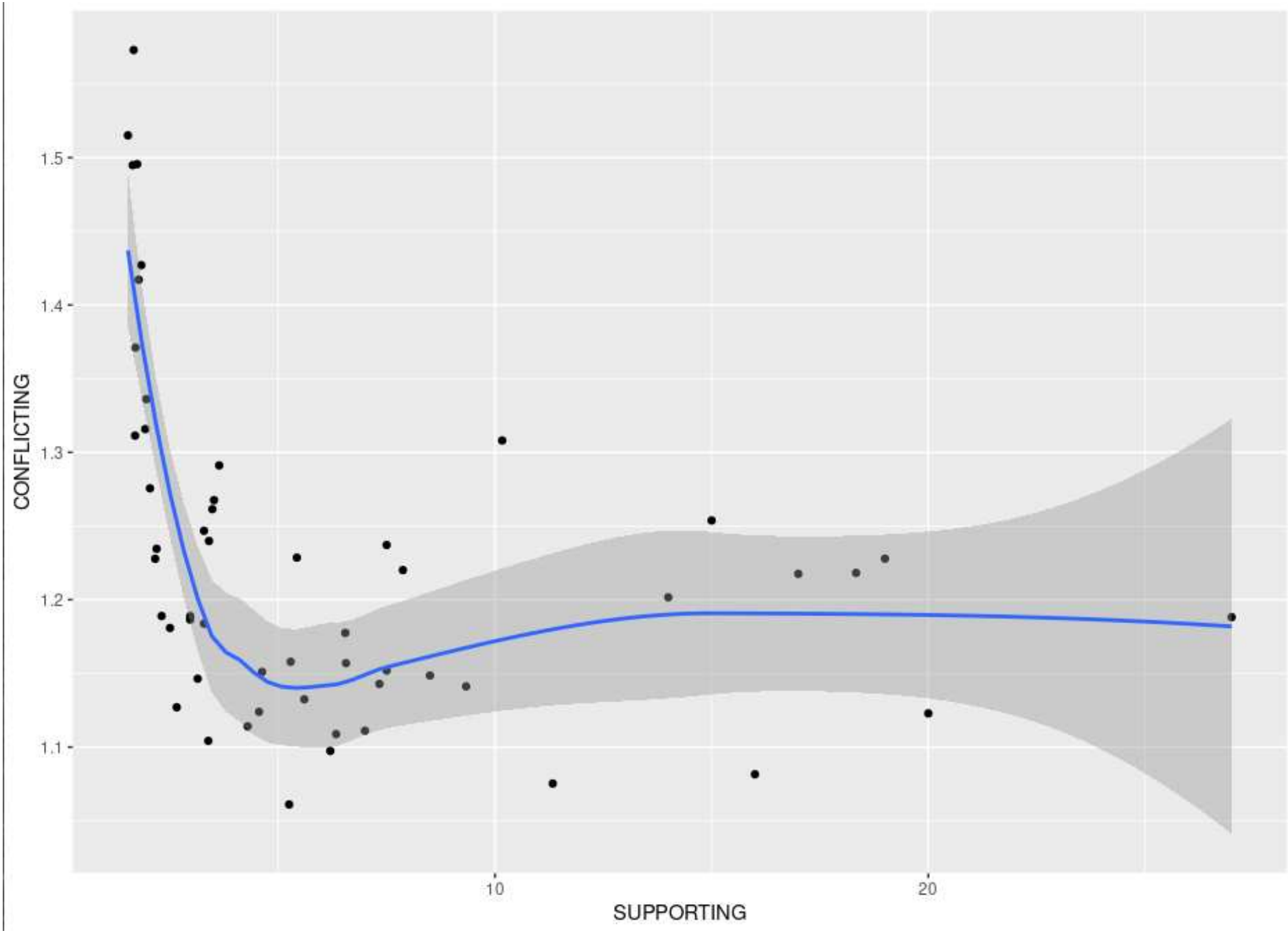
Summary

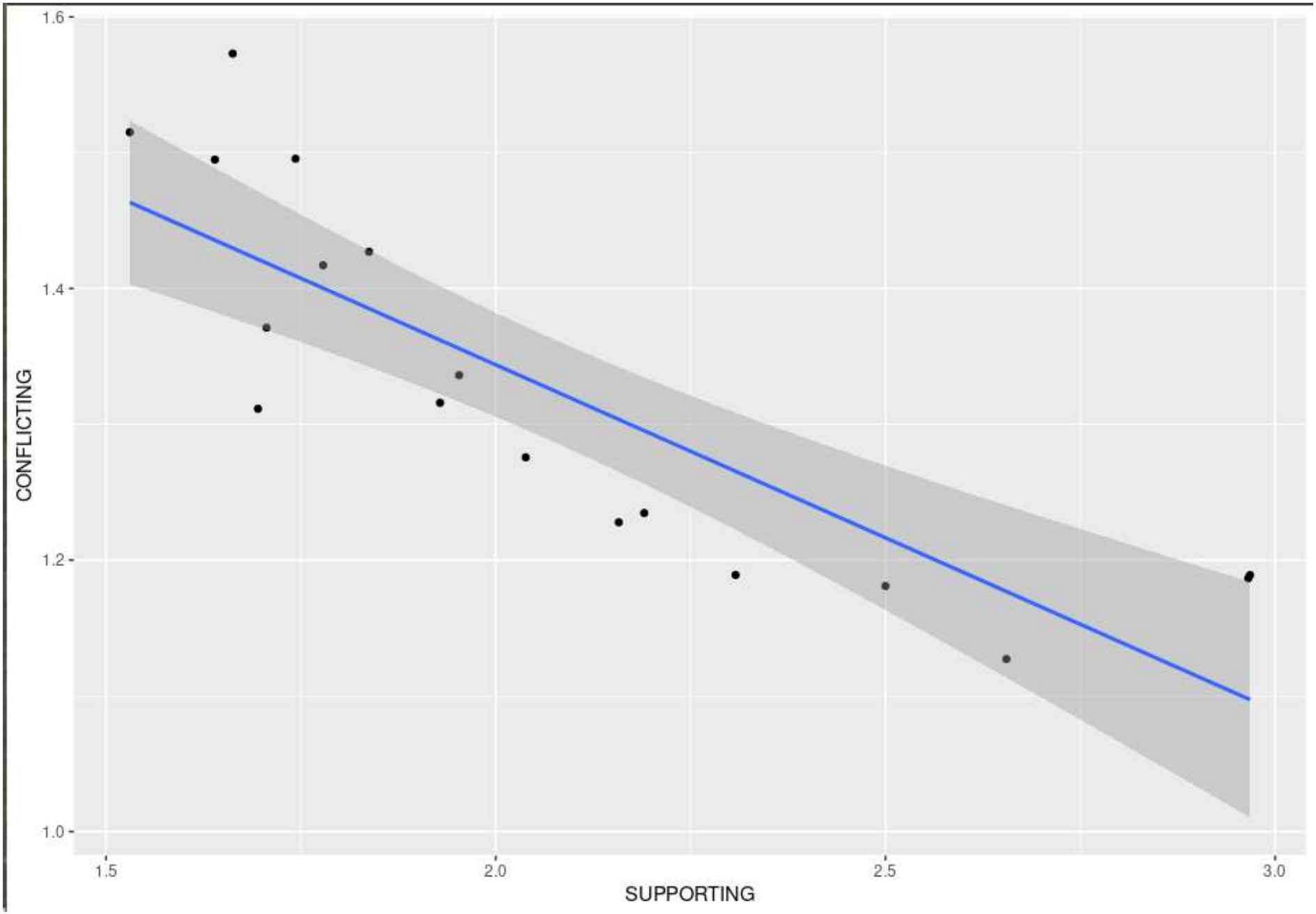
(A) Are subgroups implied by Jackson's tree observed?

- No.

(B) Is there any tree-like signal in the data?

- No subset of languages is a clear "winner" subgroup
- Contiguous sets of languages generally, but not always, form stronger subgroups, but...
- Chuukic groups with some Eastern languages (Gilbertese and Nauruan)
  - Chuukic-Gilbertese is the strongest two-language subgroup, while Pohnpeic-Gilbertese is the weakest
- There is a negative correlation between supporting and conflicting isogloss measures


CONFLICTING
1.5
1.4
1.3
1.2
1.1
10
20
SUPPORTING


1.6
1.4
1.2
1.0
CONFLICTING
1.5
2.0
2.5
3.0
SUPPORTING

Negative correlation between supporting and conflicting isoglosses

- It appeas that this correlation holds only for "the worst" subgroups (where $p_r \leq 3$; Pearson's correlation test: $r = -0.84, p < 0.001$)
- There appears to be a floor in the number of conflicting isoglosses

Some discussion

Notes on calibration

- If raw numbers rather than O/E numbers are used, Central and Western Micronesian come out as better subgroups (though still not blow-out winners!)
- It appears that the compilers of the MCD were guided by their subgrouping assumptions and specifically looked for cognatesets supporting them

Questions for future work

- Can both things be true at the same time: a clean(er) tree defined by innovative sound changes, a wave structure defined by lexical isoglosses?
- Work in progress: Bayesian phylogenetic analysis

But we can make an interim conclusion:

Micronesian appears to be a typical Oceanic linkage

Outline

References I

Athens, J. Stephen (2018). "Archeology of the Eastern Caroline Islands, Micronesia". In: *The Oxford handbook of prehistoric Oceania*. Ed. by Ethan E. Cochrane and Terry L. Hunt. Oxford: Oxford University Press, pp. 271–301.

Bender, Byron W. et al. (2003a). "Proto-Micronesian reconstructions—1". In: *Oceanic Linguistics* 42.1, pp. 1–110.

— (2003b). "Proto-Micronesian reconstructions—2". In: *Oceanic linguistics* 42.2, pp. 271–358.

— (2023). *Micronesian Comparative Dictionary (v1.0) [Data set]*. Zenodo. DOI: 10.5281/zenodo.7973420.

Blumenfeld, Lev (2022). "Notes on the diachronic phonology of Nauruan". In: *Oceanic linguistics* 61.2, pp. 684–720. DOI: 10.1353/ol.2022.0023.

— (2024). "Retention and loss of plant terms between Proto-Oceanic and Micronesian". In: *Waka Kuaka: The Journal of the Polynesian Society* 133.2, pp. 197–180. DOI: 10.15286/jps.133.2.137-180.

Blust, Robert et al. (2023). *CLDF dataset derived from Blust's "Austronesian Comparative Dictionary" (v1.2) [Data set]*. Zenodo. DOI: 10.5281/zenodo.7741197.

Clark, Ross (2009). **Leo Tuai: A comparative lexical study of North and Central Vanuatu languages*. Pacific Linguistics. Canberra: The Australian National University.

References II

Donohue, Mark et al. (2012). "New methodologies for historical linguistics? Calibrating a lexicon-based methodology for diffusion vs. subgrouping". In: *Diachronica* 29.4, pp. 505–522. DOI: 10.1075/dia.29.4.04don.

François, Alexandre (2011a). "Social ecology and language history in the northern Vanuatu linkage: a tale of divergence and convergence". In: *Journal of Historical Linguistics* 1.2, pp. 175–246. DOI: 10.1075/jhl.1.2.03fra.

— (2011b). "Where *R they all? The geography and history of *R-loss in Southern Oceanic languages". In: *Oceanic linguistics* 50.1, pp. 140–197. DOI: 10.1353/ol.2011.0009.

Geraghty, Paul (1983). *The history of the Fijian languages*. Oceanic Linguistics Special Publication 19. Honolulu: University of Hawaii Press.

Hughes, Kevin (2020a). "Nauruan classification". In: *Proceedings of the linguistic society of America* 5.1, pp. 257–269. DOI: 10.3765/plsa.v5i1.4717.

— (2020b). "The synchronic and diachronic phonology of Nauruan: towards a definitive classification of an understudied Micronesian language". PhD thesis. CUNY Graduate Center.

Jackson, Frederick H. (1983). "The internal and external relationships of the Trukic languages of Micronesia". PhD thesis. University of Hawai'i.

Jacques, Guillaume and Johann-Mattis List (2019). "Save the trees: why we need tree models in linguistic reconstruction (and when we should apply them)". In: *Journal of historical linguistics* 9.1, pp. 128–166. DOI: 10.1075/jhl.17008.mat.

References III

Kalyan, Siva and Alexandre François (2018). "Freeing the comparative method from the tree model: a framework for historical glottometry". In: *Let's talk about trees: genetic relationships of languages and their phylogenetic representations*. Ed. by Kikusawa Ritsuko and Lawrence A. Reid. Vol. 98. Senri Ethnological Studies. National Museum of Ethnology, Osaka, pp. 59–89.

— (2019). "When the waves meet the trees: a response to Jacques and List". In: *Journal of historical linguistics* 9.1, pp. 167–176.

Kirch, Patrick Vinton (2017). *On the road of the winds: an archaeological history of the Pacific islands before European contact*. Revised and expanded edition. Oakland: University of California Press.

Lichtenberk, František (1988). "The Cristobal-Malaitan subgroup of Southeast Solomonic". In: *Oceanic linguistics* 27.1/2, pp. 24–62.

Marck, Jeffrey (1986). "Micronesian dialects and the overnight voyage". In: *Journal of the Polynesian Society* 95.2, pp. 253–258.

— (1994). "Proto Micronesian terms for the physical environment". In: *Austronesian terminologies: continuity and change*. Ed. by A.K. Pawley and M.D. Ross. Pacific linguistics Series C – No. 127. Canberra: The Australian National University, pp. 301–328.

Pawley, Andrew and Malcolm Ross (2006). "The prehistory of Oceanic languages: a current view". In: *Austronesians: historical and comparative perspectives*. Ed. by Peter Bellwood et al. Canberra: ANU Press, pp. 43–80. DOI: 10.22459/A.09.2006.03.

References IV

Rehg, Kenneth L. (1995). "The significance of linguistic interaction spheres in reconstructing Micronesian prehistory". In: *Oceanic linguistics* 34.2, pp. 305–326.

Ross, Malcolm (1988). *Proto Oceanic and the Austronesian languages of Western Melanesia*. Pacific Linguistics Series C 98. Canberra: Australian National University.

Smith, Alexander D. (2023). "Evidence and Models of Linguistic Relations: Subgroups, Linkages, Lexical Innovations, and Borneo". In: *Oceanic linguistics* 62.2, pp. 324–365. DOI: 10.1353/ol.2023.a913564.

Smith, Alexander D. et al. (2025). "The Austronesian and the Micronesian Comparative Dictionaries as CLDF datasets". In: *Scientific data* 12, p. 1015. DOI: 10.1038/s41597-025-05301-4.

Walworth, Mary (2014). "Eastern Polynesian: the linguistic evidence revisited". In: *Oceanic linguistics* 53.2, pp. 256–272.

Walworth, Mary and Albert Davletshin (2019). *New perspectives on Eastern Polynesian subgrouping*. COOL 11, University of New Caledonia.