DOI : https://doi.org/10.1075/dia.22022.ski.additional

Supplementary material to:

Skirgård, Hedvig. (2024). Disentangling Ancestral State Reconstruction in historical linguistics: Comparing classic approaches and new methods using Oceanic grammar. *Diachronica*. https://doi.org/10.1075/dia.22022.ski

Contents

A Data and code availability

This study involves data from Grambank (v1.0, Skirgård et al. (2023), Skirgård et al. (2023b), D-PLACE (v2.2.1 Kirby et al. (2018) and Glottolog (v4.5, Hammarström et al. (2021). The study also involves the use of some scripts associated with the release of Grambank v1.0 which are found in the repository grambank-analysed (v1.0, Skirgård et al. (2023a)). Grambank-analysed includes in turn links to Glottolog and Grambank data. The trees from Gray et al. (2009) are stored in the D-PLACE repository, the Glottolog tree in the glottolog-cldf repository.

All the R-scripts for data wrangling, analysis and plotting are found on GitHub and Zenodo (Skirgård 2023).

Zenodo locations:

- Oceanic_computational_ASR (v1.01) https://zenodo.org/records/10390885 (Skirgård 2023)
- Grambank-analysed (v1.0) https://doi.org/10.5281/zenodo.7740822
  - Grambank (v.1.0) https://doi.org/10.5281/zenodo.7740140
  - glottolog-cldf (v4.5) https://doi.org/10.5281/zenodo.5772642
- dplace-data (v2.2.1) https://doi.org/10.5281/zenodo.5554395

GitHub locations:

- Oceanic_computational_ASR (v1.01) https://github.com/HedvigS/Oceanic_computational_ASR/tree/v1.01
- Grambank-analysed (v1.0) https://github.com/grambank/grambank-analysed/tree/v1.0 – which in turn contains submodules of:
  - Grambank (v1.0) https://github.com/grambank/grambank/tree/v1.0
  - glottolog-cldf (v4.5) https://github.com/glottolog/glottolog-cldf/tree/v4.5
- dplace-data (v2.2.1) https://github.com/D-PLACE/dplace-data/tree/v2.2.1

B Grambank features

Table 1 contains Grambank features which serves as the input to the analysis. Multistate features have been binarised. For more details, see Supplementary Material C, Skirgård et al. (2023): Materials and methods: Data and the parameters-table in the CLDF-release on Zenodo of the grambank dataset version 1 (Skirgård et al. 2023b). Documentation of the features, including procedures and examples are also found on a GitHub wiki https://github.com/grambank/grambank/wiki that is updated continuously. Release-versions are published regularly, as datasets on Zenodo.

Feature ID	Name
GB024a	Is the order of the numeral and noun Num-N?
GB024b	Is the order of the numeral and noun N-Num?
GB025a	Is the order of the adnominal demonstrative and noun Dem-N?
GB025b	Is the order of the adnominal demonstrative and noun N-Dem?
GB065a	Is the pragmatically unmarked order of adnominal possessor noun and possessed noun PSR-PSD?
GB065b	Is the pragmatically unmarked order of adnominal possessor noun and possessed noun PSD-PSR?
GB130a	Is the pragmatically unmarked order of S and V in intransitive clauses S-V?
GB130b	Is the pragmatically unmarked order of S and V in intransitive clauses V-S?
GB193a	Is the order of the adnominal property word (ANM) and noun ANM-N?
GB193b	Is the order of the adnominal property word (ANM) and noun N-ANM?
GB203a	Is the order of the adnominal collective universal quantifier (UQ) and noun UQ-N?
GB203b	Is the order of the adnominal collective universal quantifier (UQ) and noun N-QU?
GB020	Are there definite or specific articles?
GB021	Do indefinite nominals commonly have indefinite articles?
GB022	Are there prenominal articles?
GB023	Are there postnominal articles?
GB026	Can adnominal property words occur discontinuously?
GB027	Are nominal conjunction and comitative expressed by different elements?
GB028	Is there a distinction between inclusive and exclusive?
GB030	Is there a gender distinction in independent 3rd person pronouns?
GB031	Is there a dual or unit augmented form (in addition to plural or augmented) for all person categories in the pronoun system?
GB035	Are there three or more distance contrasts in demonstratives?
GB036	Do demonstratives show an elevation distinction?
GB037	Do demonstratives show a visible-nonvisible distinction?
GB038	Are there demonstrative classifiers?
GB039	Is there nonphonological allomorphy of noun number markers?
GB041	Are there several nouns (more than three) which are suppletive for number?
GB042	Is there productive overt morphological singular marking on nouns?
GB043	Is there productive morphological dual marking on nouns?

GB044	Is there productive morphological plural marking on nouns?
GB046	Is there an associative plural marker for nouns?
GB047	Is there a productive morphological pattern for deriving an action/state noun from a verb?
GB048	Is there a productive morphological pattern for deriving an agent noun from a verb?
GB049	Is there a productive morphological pattern for deriving an object noun from a verb?
GB051	Is there a gender/noun class system where sex is a factor in class assignment?
GB052	Is there a gender/noun class system where shape is a factor in class assignment?
GB053	Is there a gender/noun class system where animacy is a factor in class assignment?
GB054	Is there a gender/noun class system where plant status is a factor in class assignment?
GB057	Are there numeral classifiers?
GB058	Are there possessive classifiers?
GB059	Is the adnominal possessive construction different for alienable and inalienable nouns?
GB068	Do core adjectives (defined semantically as property concepts such as value, shape, age, dimension) act like verbs in predicative position?
GB069	Do core adjectives (defined semantically as property concepts; value, shape, age, dimension) used attributively require the same morphological treatment as verbs?
GB070	Are there morphological cases for non-pronominal core arguments (i.e. S/A/P)?
GB071	Are there morphological cases for pronominal core arguments (i.e. S/A/P)?
GB072	Are there morphological cases for oblique non-pronominal NPs (i.e. not S/A/P)?
GB073	Are there morphological cases for independent oblique personal pronominal arguments (i.e. not S/A/P)?
GB074	Are there prepositions?
GB075	Are there postpositions?
GB079	Do verbs have prefixes/proclitics, other than those that only mark A, S or P (do include portmanteau: A & S + TAM)?
GB080	Do verbs have suffixes/enclitics, other than those that only mark A, S or P (do include portmanteau: A & S + TAM)?
GB081	Is there productive infixation in verbs?
GB082	Is there overt morphological marking of present tense on verbs?
GB083	Is there overt morphological marking on the verb dedicated to past tense?

GB084	Is there overt morphological marking on the verb dedicated to future tense?
GB086	Is a morphological distinction between perfective and imperfective aspect available on verbs?
GB089	Can the S argument be indexed by a suffix/enclitic on the verb in the simple main clause?
GB090	Can the S argument be indexed by a prefix/proclitic on the verb in the simple main clause?
GB091	Can the A argument be indexed by a suffix/enclitic on the verb in the simple main clause?
GB092	Can the A argument be indexed by a prefix/proclitic on the verb in the simple main clause?
GB093	Can the P argument be indexed by a suffix/enclitic on the verb in the simple main clause?
GB094	Can the P argument be indexed by a prefix/proclitic on the verb in the simple main clause?
GB095	Are variations in marking strategies of core participants based on TAM distinctions?
GB096	Are variations in marking strategies of core participants based on verb classes?
GB098	Are variations in marking strategies of core participants based on person distinctions?
GB099	Can verb stems alter according to the person of a core participant?
GB103	Is there a benefactive applicative marker on the verb (including indexing)?
GB104	Is there an instrumental applicative marker on the verb (including indexing)?
GB105	Can the recipient in a ditransitive construction be marked like the monotransitive patient?
GB107	Can standard negation be marked by an affix, clitic or modification of the verb?
GB108	Is there directional or locative morphological marking on verbs?
GB109	Is there verb suppletion for participant number?
GB110	Is there verb suppletion for tense or aspect?
GB111	Are there conjugation classes?
GB113	Are there verbal affixes or clitics that turn intransitive verbs into transitive ones?
GB114	Is there a phonologically bound reflexive marker on the verb?
GB115	Is there a phonologically bound reciprocal marker on the verb?
GB116	Do verbs classify the shape, size or consistency of absolutive arguments by means of incorporated nouns, verbal affixes or suppletive verb stems?
GB117	Is there a copula for predicate nominals?
GB118	Are there serial verb constructions?
GB119	Can mood be marked by an inflecting word (”auxiliary verb”)?

GB120 Can aspect be marked by an inflecting word (”auxiliary verb”)?

GB121 Can tense be marked by an inflecting word (”auxiliary verb”)?

GB122 Is verb compounding a regular process?

GB123 Are there verb-adjunct (aka light-verb) constructions?

GB124 Is incorporation of nouns into verbs a productive intransitivizing process?

GB126 Is there an existential verb?

GB127 Are different posture verbs used obligatorily depending on an inanimate locatum’s shape or position (e.g. ’to lie’ vs. ’to stand’)?

GB129 Is there a notably small number, i.e. about 100 or less, of verb roots in the language?

GB131 Is a pragmatically unmarked constituent order verb-initial for transitive clauses?

GB132 Is a pragmatically unmarked constituent order verb-medial for transitive clauses?

GB133 Is a pragmatically unmarked constituent order verb-final for transitive clauses?

GB134 Is the order of constituents the same in main and subordinate clauses?

GB135 Do clausal objects usually occur in the same position as nominal objects?

GB136 Is the order of core argument (i.e. S/A/P) constituents fixed?

GB137 Can standard negation be marked clause-finally?

GB138 Can standard negation be marked clause-initially?

GB139 Is there a difference between imperative (prohibitive) and declarative negation constructions?

GB140 Is verbal predication marked by the same negator as all of the following types of predication: locational, existential and nominal?

GB146 Is there a morpho-syntactic distinction between predicates expressing controlled versus uncontrolled events or states?

GB147 Is there a morphological passive marked on the lexical verb?

GB148 Is there a morphological antipassive marked on the lexical verb?

GB149 Is there a morphologically marked inverse on verbs?

GB150 Is there clause chaining?

GB151 Is there an overt verb marker dedicated to signalling coreference or noncoreference between the subject of one clause and an argument of an adjacent clause (”switch reference”)?

GB152 Is there a morphologically marked distinction between simultaneous and sequential clauses?

GB155 Are causatives formed by affixes or clitics on verbs?

GB156 Is there a causative construction involving an element that is unmistakably grammaticalized from a verb for ’to say’?

GB158 Are verbs reduplicated?

GB159 Are nouns reduplicated?

GB160 Are elements apart from verbs or nouns reduplicated?

GB165	Is there productive morphological trial marking on nouns?
GB166	Is there productive morphological paucal marking on nouns?
GB167	Is there a logophoric pronoun?
GB170	Can an adnominal property word agree with the noun in gender/noun class?
GB171	Can an adnominal demonstrative agree with the noun in gender/noun class?
GB172	Can an article agree with the noun in gender/noun class?
GB177	Can the verb carry a marker of animacy of argument, unrelated to any gender/noun class of the argument visible in the NP domain?
GB184	Can an adnominal property word agree with the noun in number?
GB185	Can an adnominal demonstrative agree with the noun in number?
GB186	Can an article agree with the noun in number?
GB187	Is there any productive diminutive marking on the noun (exclude marking by system of nominal classification only)?
GB188	Is there any productive augmentative marking on the noun (exclude marking by system of nominal classification only)?
GB192	Is there a gender system where a noun’s phonological properties are a factor in class assignment?
GB196	Is there a male/female distinction in 2nd person independent pronouns?
GB197	Is there a male/female distinction in 1st person independent pronouns?
GB198	Can an adnominal numeral agree with the noun in gender/noun class?
GB204	Do collective (’all’) and distributive (’every’) universal quantifiers differ in their forms or their syntactic positions?
GB250	Can predicative possession be expressed with a transitive ’habeo’ verb?
GB252	Can predicative possession be expressed with an S-like possessum and a locative-coded possessor?
GB253	Can predicative possession be expressed with an S-like possessum and a dative-coded possessor?
GB254	Can predicative possession be expressed with an S-like possessum and a possessor that is coded like an adnominal possessor?
GB256	Can predicative possession be expressed with an S-like possessor and a possessum that is coded like a comitative argument?
GB257	Can polar interrogation be marked by intonation only?
GB260	Can polar interrogation be indicated by a special word order?
GB262	Is there a clause-initial polar interrogative particle?
GB263	Is there a clause-final polar interrogative particle?
GB264	Is there a polar interrogative particle that most commonly occurs neither clause-initially nor clause-finally?
GB265	Is there a comparative construction that includes a form that elsewhere means ’surpass, exceed’?

GB266 Is there a comparative construction that employs a marker of the standard which elsewhere has a locational meaning?

GB270 Can comparatives be expressed using two conjoined clauses?

GB273 Is there a comparative construction with a standard marker that elsewhere has neither a locational meaning nor a ’surpass/exceed’ meaning?

GB275 Is there a bound comparative degree marker on the property word in a comparative construction?

GB276 Is there a non-bound comparative degree marker modifying the property word in a comparative construction?

GB285 Can polar interrogation be marked by a question particle and verbal morphology?

GB286 Can polar interrogation be indicated by overt verbal morphology only?

GB291 Can polar interrogation be marked by tone?

GB296 Is there a phonologically or morphosyntactically definable class of ideophones that includes ideophones depicting imagery beyond sound?

GB297 Can polar interrogation be indicated by a V-not-V construction?

GB298 Can standard negation be marked by an inflecting word (”auxiliary verb”)?

GB299 Can standard negation be marked by a non-inflecting word (”auxiliary particle”)?

GB300 Does the verb for ’give’ have suppletive verb forms?

GB301 Is there an inclusory construction?

GB302 Is there a phonologically free passive marker (”particle” or ”auxiliary”)?

GB303 Is there a phonologically free antipassive marker (”particle” or ”auxiliary”)?

GB304 Can the agent be expressed overtly in a passive clause?

GB305 Is there a phonologically independent reflexive pronoun?

GB306 Is there a phonologically independent non-bipartite reciprocal pronoun?

GB309 Are there multiple past or multiple future tenses, distinguishing distance from Time of Reference?

GB312 Is there overt morphological marking on the verb dedicated to mood?

GB313 Are there special adnominal possessive pronouns that are not formed by an otherwise regular process?

GB314 Can augmentative meaning be expressed productively by a shift of gender/noun class?

GB315 Can diminutive meaning be expressed productively by a shift of gender/noun class?

GB316 Is singular number regularly marked in the noun phrase by a dedicated phonologically free element?

GB317	Is dual number regularly marked in the noun phrase by a dedicated phonologically free element?
GB318	Is plural number regularly marked in the noun phrase by a dedicated phonologically free element?
GB319	Is trial number regularly marked in the noun phrase by a dedicated phonologically free element?
GB320	Is paucal number regularly marked in the noun phrase by a dedicated phonologically free element?
GB321	Is there a large class of nouns whose gender/noun class is not phonologically or semantically predictable?
GB322	Is there grammatical marking of direct evidence (perceived with the senses)?
GB323	Is there grammatical marking of indirect evidence (hearsay, inference, etc.)?
GB324	Is there an interrogative verb for content interrogatives (who?, what?, etc.)?
GB325	Is there a count/mass distinction in interrogative quantifiers?
GB326	Do (nominal) content interrogatives normally or frequently occur in situ?
GB327	Can the relative clause follow the noun?
GB328	Can the relative clause precede the noun?
GB329	Are there internally-headed relative clauses?
GB330	Are there correlative relative clauses?
GB331	Are there non-adjacent relative clauses?
GB333	Is there a decimal numeral system?
GB334	Is there synchronic evidence for any element of a quinary numeral system?
GB335	Is there synchronic evidence for any element of a vigesimal numeral system?
GB336	Is there a body-part tallying system?
GB400	Are all person categories neutralized in some voice, tense, aspect, mood and/or negation?
GB401	Is there a class of patient-labile verbs?
GB402	Does the verb for ’see’ have suppletive verb forms?
GB403	Does the verb for ’come’ have suppletive verb forms?
GB408	Is there any accusative alignment of flagging?
GB409	Is there any ergative alignment of flagging?
GB410	Is there any neutral alignment of flagging?
GB415	Is there a politeness distinction in 2nd person forms?
GB421	Is there a preposed complementizer in complements of verbs of thinking and/or knowing?
GB422	Is there a postposed complementizer in complements of verbs of thinking and/or knowing?
GB430	Can adnominal possession be marked by a prefix on the possessor?

GB431	Can adnominal possession be marked by a prefix on the possessed noun?
GB432	Can adnominal possession be marked by a suffix on the possessor?
GB433	Can adnominal possession be marked by a suffix on the possessed noun?
GB519	Can mood be marked by a non-inflecting word (”auxiliary particle”)?
GB520	Can aspect be marked by a non-inflecting word (”auxiliary particle”)?
GB521	Can tense be marked by a non-inflecting word (”auxiliary particle”)?
GB522	Can the S or A argument be omitted from a pragmatically unmarked clause when the referent is inferrable from context (”pro-drop” or ”null anaphora”)?

Table 1: Table of Grambank fetures

C Binarisation of the Grambank features

Most of the feature questions are binary (e.g. GB027: Are nominal conjunction and comitative expressed by different elements?) but a few are multi-state (e.g. GB024 What is the order of numeral and noun in the NP? 1) Num-N, 2) N-Num, 3) both). For the analysis in this study, the multi-state features have been binarised. This is because the values of the multi-state features are not independent of each other; they all contain the value “Both”. The value “Num-N” (numeral before noun) of GB024 is more similar to “Both” than it is to the other alternative “N-Num”. The relationship between the three values are not equal or independent. The table in B contains a list of all the features used in this study, including the binarised features.

D Table of historical linguistics sources surveyed

Table 2: Table of historical linguistics publications used in this dissertation for Proto-Oceanic grammar

Citation	**Title**	**Proto-Languages**	**Domains**
Pawley (1970)	Grammatical reconstruction and change on Polynesia and Fiji	Proto-Central Pacific	Verbal markers and aspect particles
Pawley (1973)	Some problems in Proto-Oceanic	Proto-Oceanic and Proto-Polynesian	Possession, noun phrase marking, negation, verbal markers, clusivity, word order
Clark (1973)	Aspects of Proto-Polynesian syntax	Proto-Oceanic and Proto-Polynesian	Alignment, negation, word order, possession, noun phrase marking, voice
Chung (1978)	Case marking and grammatical relations in Polynesian languages	Proto-Polynesian	Alignment, word order, voice, noun phrase marking
Crowley (1985)	Common noun phrase marking in Proto-Oceanic	Proto-Oceanic	noun phrase marking, clusivity
Jonsson (1998)	Det polynesiska verbmorfemet -*Cia*; om dess funktion i Samoanska	Proto-Polynesian	Verbal marker

Citation	**Title**	**Proto-Languages**	**Domains**
Marck (2000)	Polynesian languages (in Facts About the World's Languages: An encyclopaedia of the world's major languages, past and present)	Proto-Central Pacific and Proto-Polynesian	Word order, verbal markers, possession, clusivity
Evans (2001)	A study of valency-changing devices in Proto Oceanic	Proto-Oceanic	Verbal markers
Ball (2007)	On ergativity and accusativity in Proto-Polynesian and proto-Central Pacific	Proto-Polynesian	Alignment, voice
Kikusawa (2001)	Rotuman and Fijian case-marking strategies and their historical development	Proto-Oceanic	Possession, pronominal number
Kikusawa (2002)	Proto Central Pacific ergativity: Its reconstruction and development in the Fijian, Rotuman and Polynesian languages	Proto-Central Pacific	Alignment, word order
Lynch et al. (2011)	The Oceanic Languages, paper 4: Proto-Oceanic	Proto-Oceanic, Proto-Central Pacific and Proto-Polynesian	Negation, word order, verbal markers, clusivity, possession, pronominal number, polar interrogation, nominalisations and more
Ross (2004)[1]	The morphosyntactic typology of Oceanic languages	Proto-Oceanic and Proto-Polynesian	alignment, word order, verbal markers, possession, noun phrase marking

[1] This paper makes statements about "canonical" Oceanic languages, which is technically different from *reconstruction* of Proto-Oceanic. However, the author does state that the "canonic type is probably also a reflection of the morphosyntax of Proto Oceanic" (Ross 2004: 492) and has given personal approval for the paper to be included in this study in this manner.

E R packages used

All the analysis for this research project was done in the free and open source programming language R, using a multitude of packages. All code and data for this project are available in supplementary material and the locations listed in Supplementary material §A. The scripts have been written so that any user of R can execute them. Please see the bibliography for information on package versions. Below are citations for all used packages.

Jombart et al. (2022), Paradis et al. (2023), Wickham (2019), R Core Team (2023), Bååth (2018), Ottolinger (2019), Orme et al. (2023), Louca (2023), Maechler et al. (2022), Beaulieu et al. (2022), Dowle and Srinivasan (2023), Wickham et al. (2023b), Wickham (2023a), Hester et al. (2023b), Wickham et al. (2023a), Kassambara (2023), Lucas et al. (2023), Warnes et al. (2022), Firke (2023), Ooms (2023), Xie (2023), Spinu et al. (2023), Brownrigg (2022), Ripley (2023), Tierney et al. (2023), Cooper (2022), Schliep et al. (2023), et al. (2020), Revell (2023), Revelle (2023), Wickham and Henry (2023), Ching (2023), Wickham et al. (2023c), Csárdi et al. (2023), Wickham (2020), Maechler (2023), Wickham (2023b), Müller and Wickham (2023), Wickham et al. (2023d), Garnier (2023a), Garnier (2023b), Hester et al. (2023a), Ram and Wickham (2018), Dahl et al. (2019), Jombart and Dray (2010), Paradis and Schliep (2019), Louca and Doebeli (2017a), Wickham (2016), Ooms (2014), Xie (2015), Xie (2014), Grolemund and Wickham (2011), Venables and Ripley (2002), Tierney and Cook (2023), Schliep (2011), Schliep et al. (2017), Revell (2012), Wickham (2007), Garnier et al. (2023a) and Garnier et al. (2023b) .

F Technical details of ASR by Maximum Parsimony and Maximum Likelihood

For Maximum Parsimony, I am using the function `asr_max_parsimony()` from the R-package `castor` (Louca and Doebeli 2017b) (which is an instantiation of the method described in Sankoff 1975) for calculating ancestral states and stability of features. This function produces ancestral states for all nodes and reports the number of changes that was minimally required for each feature.

Ancestral state reconstruction using Maximum Likelihood Estimation involves computing each ancestral state from the tips up to the root taking into account branch lengths and the joint likelihood of states given all nodes in the tree (Wilks (1938); Pagel (1994); Cunningham et al. (1998)). The Maximum Likelihood Estimation function takes a set of observations and computes the parameter distribution that maximises the likelihood given the observed data[2]. This means that for every split in the tree – every ancestral node – the Maximum Likelihood Estimation function computes what is the most likely distribution at that point given the nature of all values in the entire tree. ML can be modified so that it allows for different rates of change. An Equal Rates (ER) model assumes that the chance of transition from state A to state B and

[2] For a gentle introduction to the concept of Maximum Likelihood Estimation, see Brooks-Bartlett (2018).

from B to A are equal. However, we as linguists are aware that certain features are more likely to be lost than gained so this is not a reasonable assumption. Therefore, I allow the model to estimate different transition rates for going from A to B and from B to A given the data. This is known as "All Rates are Different" (ARD).

When estimating ancestral states with ML, it is possible to either a) find the state at each node that maximises the likelihood (integrating over all other states at all nodes, in proportion to their probability) at that particular node (marginal reconstruction), or b) find the set of character states at all nodes that (jointly) maximize the likelihood of the entire tree (joint reconstruction). I am using marginal reconstruction in this study since it is the recommended way to deal with uncertainty in reconstruction (Revell 2014). These two methods often yield similar results, but can differ, see Felsenstein (2004: 259-260), Yang (2006: 121-126) and Joy et al. (2016: 5) for more details. For our data, a trial run of joint reconstruction did not generate drastically different outcomes.

For this study, the function `R-corHMM` from the `R`-package `corHMM` (Beaulieu et al. 2022) is used for marginal reconstruction of ancestral states and rates of change per feature.

Languages with missing data were pruned away in all analysis, no hidden state reconstruction of values at tips was performed. The match between Glottolog 4.5 and Grambank is 271, the match between Gray et al. (2009) and Grambank is 132. For both MP and ML, languages with missing data were dropped from the trees in the analysis for that feature. If after this pruning less than half of the tips remained, that analysis was not carried out.

For both Maximum Parsimony and Maximum Likelihood it is possible for a structural feature to appear and disappear several times along a lineage. This is different from cognate data where a cognate class cannot re-appear.

G Supplementary Figure: distance Scatterplot Matrix

Figure 1 shows the pairwise distances between the same tips in each of the different trees (in the case of the 100 random posterior trees it's the mean of the distances) and in addition, Gower-distances between the same languages given all Grambank features.

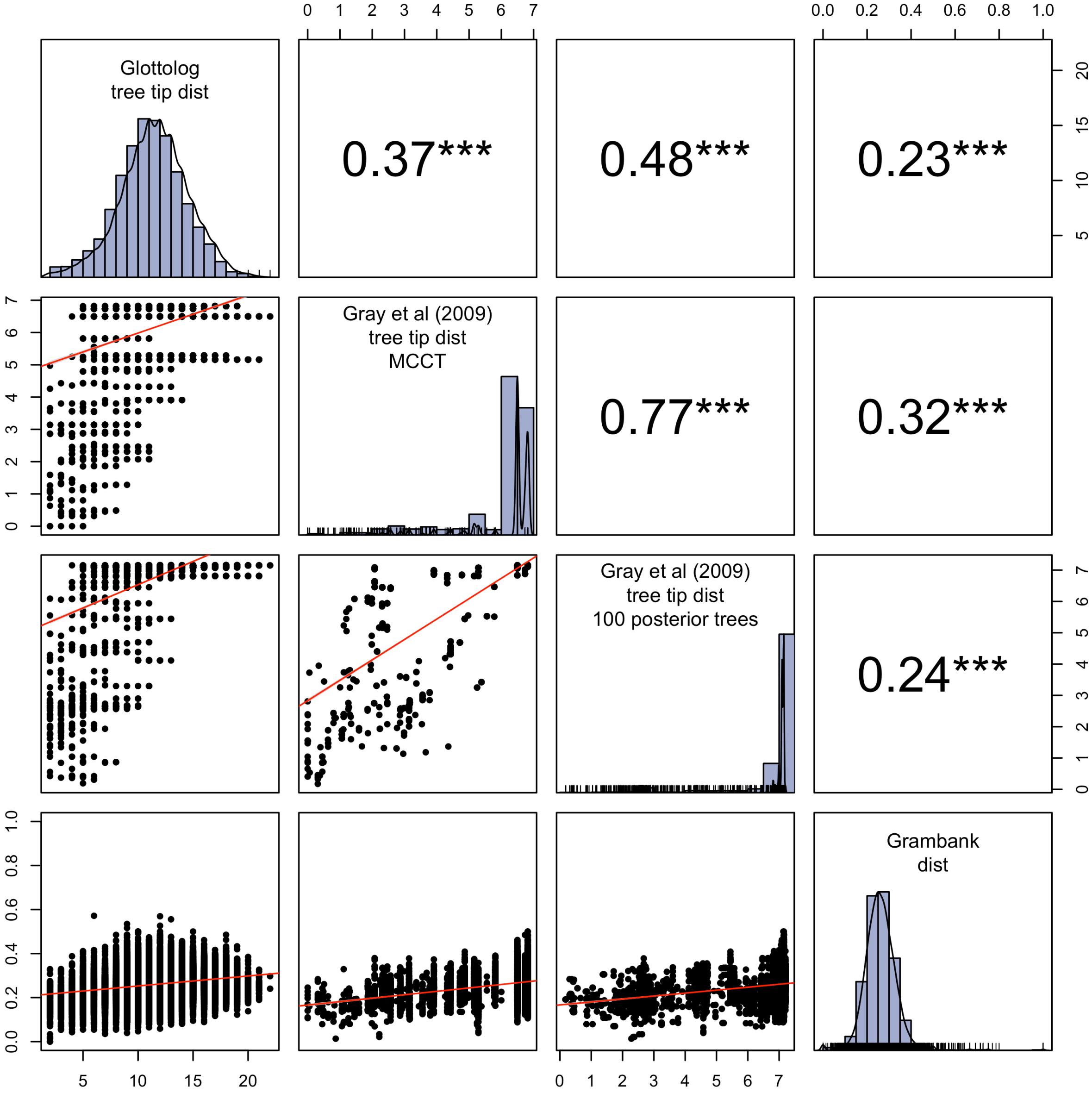


Figure 1: Comparison of distances between tips of the different trees and Grambank. Correlations are Pearson coefficients, the stars indicate the conventional p-value cut-off at 0.05.

H Supplementary Figure: tree heatmap of Gray et al (2009)-MCCT and Grambank variables

Figure 2 shows the MCC-tree from Gray et al. (2009) and a data-matrix of all 201 binarised Grambank variables. These data are the input for the ASR-analysis for this particular tree and the D-estimate calculation. Missing data are ignored in both sets of analysis.

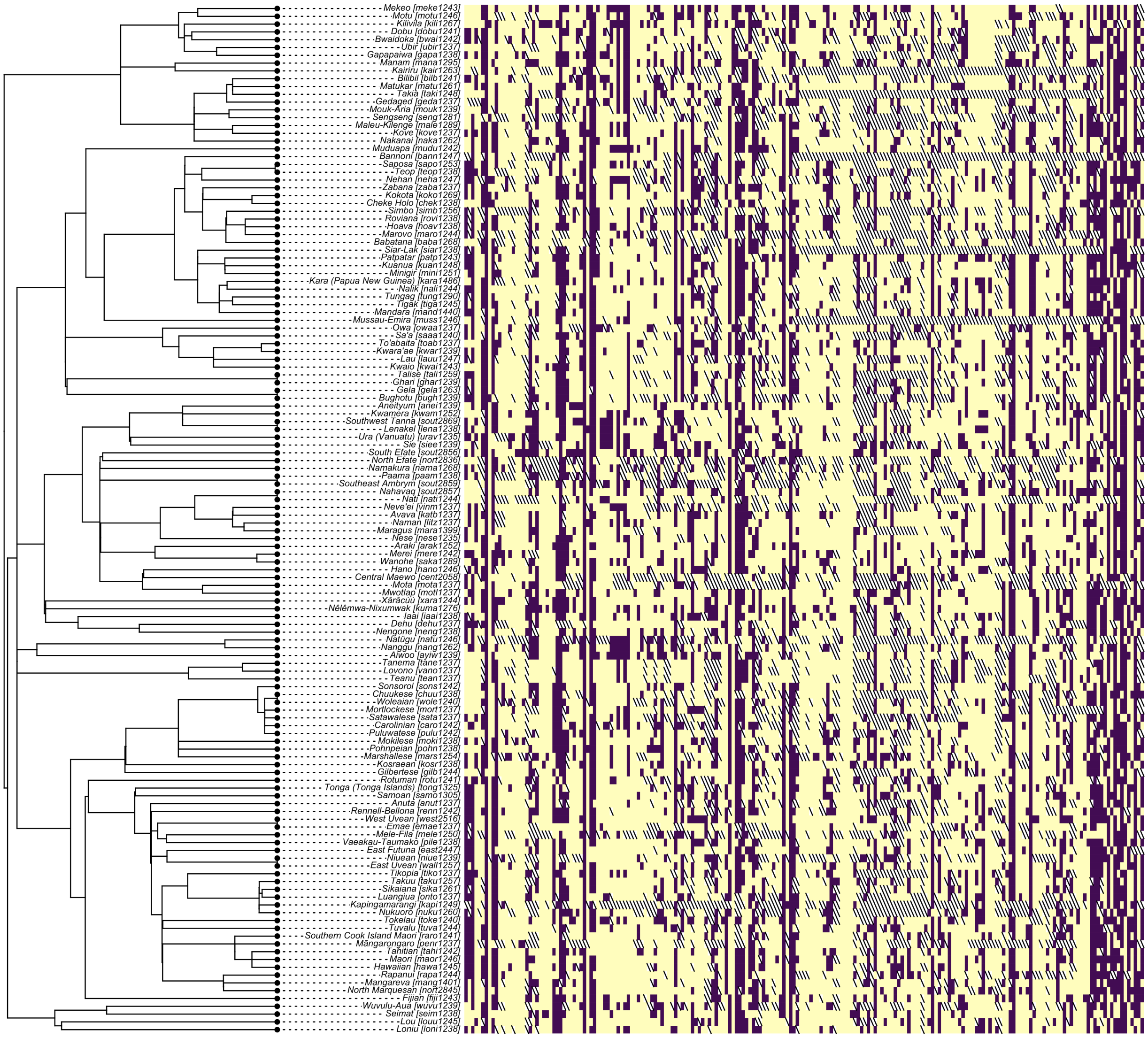


Figure 2: MCC-tree from Gray et al. (2009) with Grambank data matrix. Purple = present, yellow = absent and striped = missing.

I Technical details on D-estimation

D-estimates are a tool for measuring phylogenetic signal in a set of binary data. Phylogenetic signal can be broadly described as the degree to which the data are generated by a given tree, or whether it was generated by some other process such as randomness. This particular method was proposed by Fritz and Purvis (2010) and is implemented in the R-package caper by Fritz and Orome (Orme et al. 2013).

The method outputs three primary values per dataset and tree: i) a D-estimate, ii) a p-value that represents how similar the data are to 0 (Brownian motion) and iii) the same kind of p-value, but instead in regard to how similar the data are to a D-estimate of 1 (randomness). If the 0-p-value is large (i.e., p>0.05) that means that the D-estimate of the data are *not dissimilar* from 0, in other worse it is *similar*. If we want to find sets of data that are similar to 0, we should look for large 0-p-values (not dissimilar = similar). The same goes for the p-values relating to 1. There can be D-estimates that are similar to both 0 *and* 1 – or neither.

The method relies on generating two kinds of simulated data: a Brownian threshold process and randomness. It then measures how similar your empirical data are to the Brownian simulation in comparison to how similar the Brownian simulation is to the random simulations. A D-estimate value of 0 represents identity to the Brownian process, 1 to the random process. D-estimates can also be smaller than 0 and larger than 1, and certainly any values in between.

The results are sensitive to how many random permutations it runs for the second set of simulated data. Fritz and Purvis (2010) recommends 1,000 permutations, which is also what the default value is set to for the function phylo.d in the R-package caper. However, during the work for this paper I have found further considerations that should be taken into account when working with this method – specifically in regard to the number of random permutations and skewed distributions.

I.1 D-estimate: Sensitivity to skewed distributions

While it is true that D-estimates can be smaller than 0 and larger than 1, in my experience values lower than -7 (very strong signal) and larger than 7 (very over-dispersed) are rare in empirical data. Furthermore, we would expect that if we re-run the algorithm a second time using the same data, same tree and same settings we get a similar result to the first time. This is generally true, except in certain specific situations. When the data are such that only one datapoint has a diverging value from the rest – for example in a set of 155 tips only one of them has the value 1 for the binary trait and all others 0 – then the algorithm struggles and produces very different results on each run, and very extreme values such as -10 on one run and 10 on another. This is problematic, and was probably not discovered by Fritz and Purvis (2010) and Orme et al. (2013) because their empirical data rarely exhibited this kind of distribution (1 - 154). However, for some of the linguistic features of this study this can indeed happen.

Having identified the problem, I can also offer two solutions: a) increasing the number of random permutations and/or b) disregarding data of this kind. Many thanks to (Orme et al. 2013) for the package documentation of caper and the paper by Fritz and Purvis (2010) for providing enough methodological detail for this to be diagnosed. Stephen Mann was also invaluable to helping diagnose and address this issue mathematically.

To illustrate the problem I generated a tree with 155 tips with different distributions of binary values. The list below describes the different feature value distributions (with short names used in the plot in parenthesis) and Fig 3 shows the tree and feature value distributions

- only one tip of state 1, all other 154 tips 0 (singleton)
    - daughter with few splits from the roots (outlier)
    - in a more nested position (middle)
    - at a random position (random)
- three features with each a pair of direct sister tips of state 1, all other 153 tips 0 (sisters_a, sisters_b and sisters_c)
- two random tips with the state 1, all other 0 (two_random)
- three features with each a set of three closely related languages with the state 1, all other 152 tips 0 (triplets_a, triplets_b and triplets_c)
- three random tips with the state 1, all other 0 (three_random)
- three features with each a set of four closely related languages with the state 1, all other 151 tips 0 (quadruplets_a, quadruplets_b and quadruplets_c
- four random tips with the state 1, all other 0 (four_random)
- a cluster of 31 tips which form a clade all with 1 for the feature, all others 0 (cluster)
- 31 random tips with the same state, all others other (cluster_random)

I then proceeded to estimate the D-value for each of these 17 features, varying the number of permutations (1,000, 20,000 and 30,000). I repeated this 8 times, i.e., generating 17 * 3 * 8 D-estimates. For the entire investigation, see the script 11_phylo_d_investigation.R in the accompanying material.

The D-estimates for the singleton-features varied the most, with one iteration of the singleton outlier feature reaching a D-estimate value of 1,520 (sic). This value occurred when the number of permutations was set to 1,000. In another iteration over the same feature and the same number of iterations, the D-estimate came out as -21. While it is potentially plausible to get very small or very large values, we would expect to get *similar* values with each iteration given the same data and settings. The difference between a positive value of 1,520 and a negative of -21 is surely *unreasonably* large. When the number of permutations was increased beyond 1,000, the variance of the output with each iteration was reduced (see Fig 4), but it was still noticeably larger in cases where the distribution was heavily skewed.

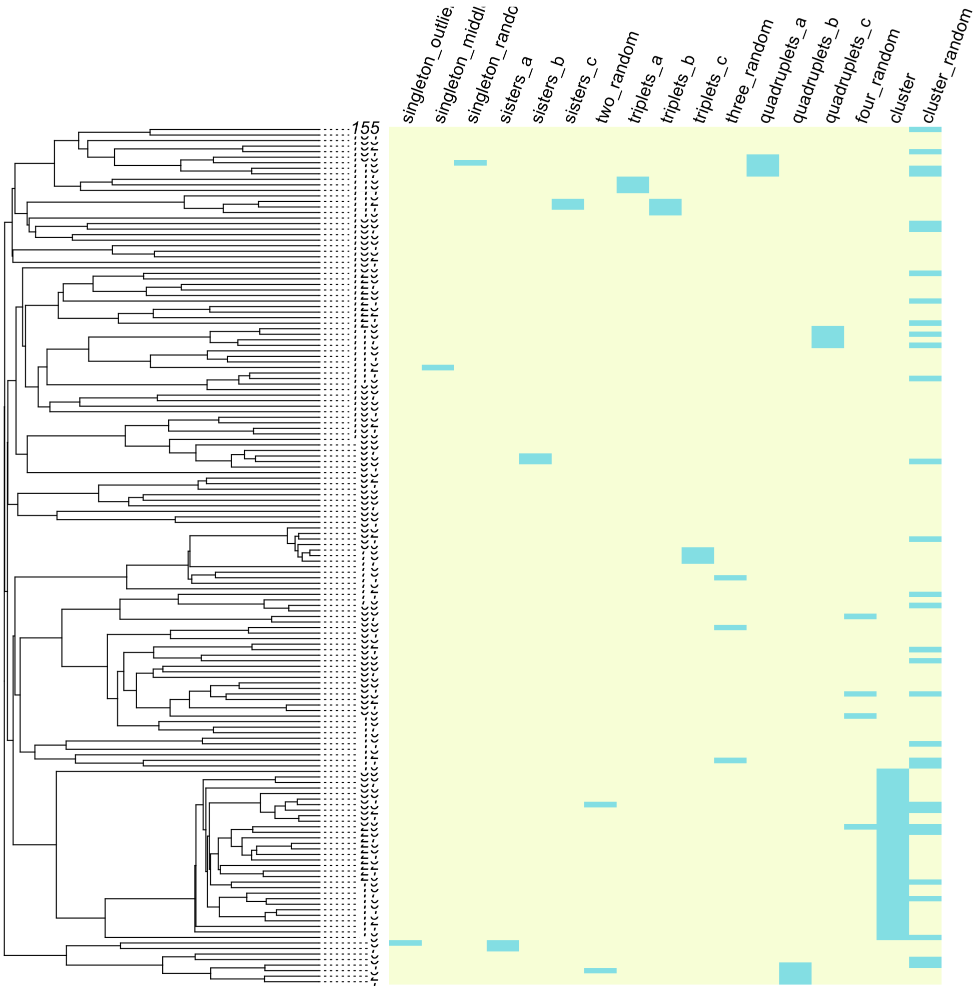


Figure 3: Tree and values heatmap for D-estimation investigation.

The cause of this issue with wildly varying D-estimates each run, especially when the feature value distribution is very skewed, has to do with the chance of generating a particular pattern of 1/154 *precisely* versus 4/151. Each time the D-estimate process is applied, a set of random and Brownian simulations are generated (the number is set by the permutations value). If the data are of the kind where 1 tip differs from all the other 154 tips (as for a few of the features in the toy example above), there is a chance that that particular position of that one value occurs in at least one of the random cases. If it does happen to occur, we would get a D-estimate that signals randomness – and conversely, if it happens to be similar to the Brownian evolutionary model. If the random and Brownian simulations end up being similar the denominator (see Eq. 1) in the formula becomes very small, which can lead to very large absolute values for the D-estimate (such as the 1,520 we saw earlier). In Eq. 1) (Fritz and Purvis 2010) r = random, b = brownian and obs = observed data.

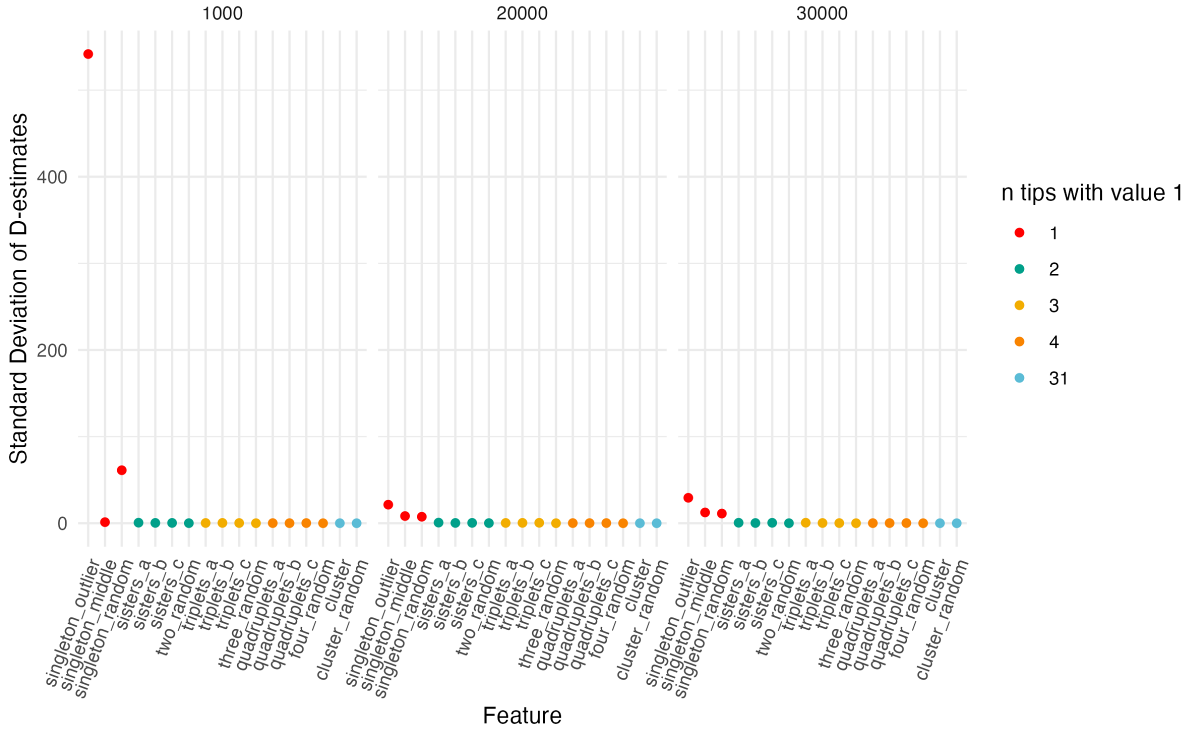


Figure 4: Scatterplot of the standard deviation of D-estimate values per feature per value of random permutations

$$\mathrm{D} = \frac{\sum d_{obs} - mean(\sum d_b)}{mean(\sum d_r) - mean(\sum d_b)} \tag{1}$$

There is less of a chance of this happening if we have more tips in each state, because those are more complicated patterns that are less likely to occur exactly in the simulated processes. Because of the possibility of this irrelevant similarity, it is necessary to increase the number of simulated permutations so that we have a larger pool of things to compare our data to. This is why the D-estimate standard deviation stabilise more in cases with skewed feature distributions if the number of permutations is increased (see Fig 4).

Even when the number of permutations is increased to 30,000, the instances where there is a feature distribution of 1 - 154 (singletons) are more volatile than the rest. When using this technique, it may be necessary to set aside such cases and evaluate them separately from the rest. We may want to ask ourselves: what does it mean for something that does not even form a pair to have or not have a phylogenetic signal?

If we look at the non-singleton features (the pairs, triplets, quadruplets and larger group) in the simulation example explored here in Figure 5 we see that they behave more similarly with each iteration. Even an increase from 1 to 2 tips of the same state improves the performance of this method in terms of producing a similar value each iteration.

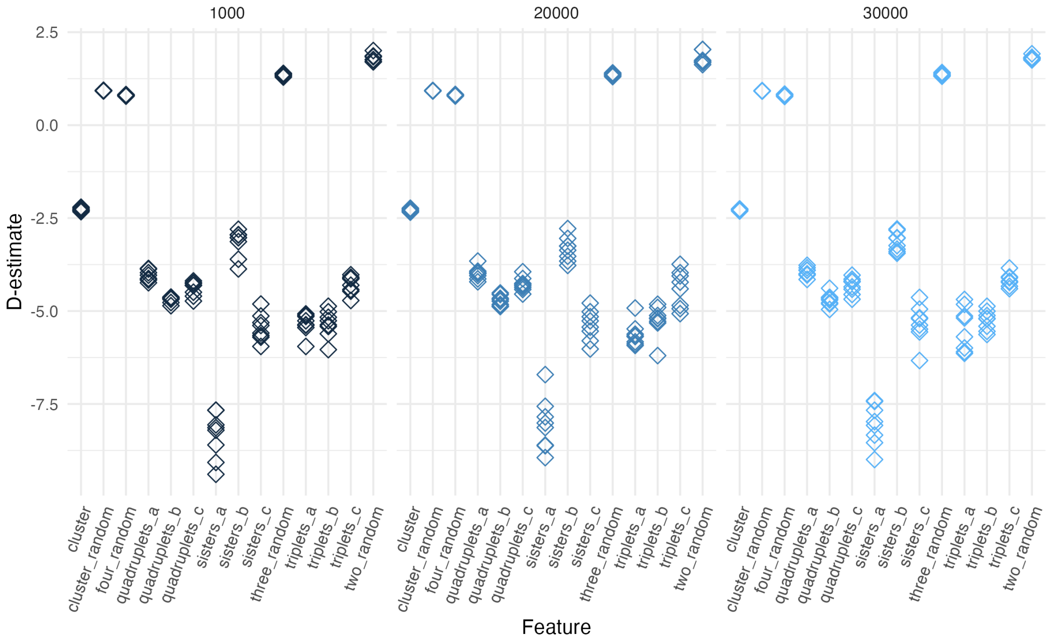


Figure 5: Scatterplot of the D-estimate values per feature per value of random permutations, for all non-singleton features. Each point represents a D-estimate value per feature, per number of permutations and per iteration.

I.2 Categories of D-estimates that do not meet the rigours of the model

In the data in this study, there were cases of inappropriate D-estimates, which were possible to diagnose both by the extremity of the D-estimates, but also by examining the p-values (dissimilarity to Brownian/clumped and random/over-dispersed).

The output is grouped into 2 groups, with 3 subgroups each. The output in the second group is not possible to include in the analysis because the conditions do not meet the model requirements, it is either impossible to conduct the analysis (all tips one state), would generate seriously unreliable results (singleton states) or shows evidence of Brownian and random being similar which also throw suspicion on the outcome. For future work, it would be desirable if the `R-function caper::phylo.d()` also output a p-value which represents the dissimilarity between the random and Brownian simulations and in addition generated a warning when the distributions are heavily skewed (for example, only 5% tips in one state).

As with the ASR-results, I also excluded output where the number of tips that had data were fewer than half of the tips in the full Oceanic-tree, i.e., for Glottolog fewer than 135.5 and for the Gray et al. (2009)-trees 66. See counts in Table 1 in §2.2 in the main text. The tables below only represent the D-estimates

The p-values that are produced by the `R-function caper::phylo.d()` represent the proportion of simulations where the observed values had a smaller sum sister-clade differences compared to the Brownian simulation, and larger than the random. Pval0 = 0 means that the observed sister-clade differences were always greater than the Brownian simulations, pval0 = 1 means they were always lower. Pval1 = 0 means that the observed sister-clade differences were always lower than the random simulations, and pval1 = 1 means that they were always greater than the random. For more details, see the source code of `caper::phylo.d()`.

- possible to include in analysis

  (i) observed values definitely on the Brownian/clumped end of the spectrum (pval0 $>$ 0.05 & pval1 $<$ 0.05)

  (ii) observed values definitely on the random/overdispersed end of the spectrum (pval0 $<$ 0.05 & pval1 $>$ 0.05)

  (iii) observed values definitely between Brownian/clumped and random/overdispersed. In all of these cases, the D-estimate is between 0 and 1. (pval0 $<$ 0.05 & pval1 $<$ 0.05)

- *not* possible to include in analysis

  (i) all tips same state (D-estimate is undefined)

  (ii) singleton (only one tip has a different state from all other tips)

  (iii) Brownian and random simulations are not sufficiently distinct from each other to get a meaningful D-estimate, observed values appear to be similar to both (pval0 $>$ 0.05 & pval1 $>$ 0.05). D-estimate can be <0, in between or >1.

Tables 3 and 4 shows the number of instances of each of these categories over the trees. There are fewer instances in the problematic categories and they have been excluded from further analysis with D-estimates. Because they represent cases with skewed distributions, it is possible to interpret them as representing very rare phenomena and one interpretation of that could be a strong phylogenetic signal – but the D-estimate test is not suitable. The values for the 100 trees from the posterior are averages.

tree	**similar to 0**	**similar to 1**	**dissimilar to both**
Glottolog	37	7	33
Gray - MCCT	39	16	12
Gray - posteriors	50	9	2

Table 3: Table of types of D-estimates per tree, data-points included.

tree	**all same**	**singleton**	**similar to both**
Glottolog	0	2	5
Gray - MCCT	1	3	13
Gray - posteriors	1	3	18

Table 4: Table of types of D-estimates per tree, data-points not included.

I.3 Correlation D-estimate and HL-concurrence

Phylogenetic signal could be an indication that it is easier to reconstruct a prior state. One may for example consider that it ought to be more difficult to reconstruct a state reliably if the pattern is a random phylogenetic signal (D-estimate similar to 1), and conversely that a strong signal may make it easier to reconstruct consistently, and therefore that the agreement between conventional historical linguistics findings and the computational methods applied in this paper would be higher if the phylogenetic signal is strong (=similar to 0, Brownian). This is however not the case in this study.

Figure 6 shows the D-estimate on the x-axis (low = strong signal, high = random) and agreement with conventional historical linguistics on the y-axis. The agreement with HL is the precise value that the method predicted for the state that HL suggests. If HL suggests that the state is present at a particular node, and the computational suggests that presence has a likelihood of 0.435, the agreement value is 0.435. This is a continuous scale, but for the parsimony results it is often 0, 0.5 or 1 because of the prevalence of binary splits in the tree and the way the method works.

The results have been grouped by method and tree. If strong phylogenetic signal (low D-estimate) predicts high agreement between conventional HL ASR and computational ASR, then the correlation would be negative but we see several positive relationships. Regardless, in no case does the correlation reach the conventional threshold for statistical significance for the Pearson correlation ($p > 0.05$).

Each point is mapped onto one prediction of one feature and one proto-language (Proto-Oceanic, Proto-Central Pacific, Proto-Polynesian or Proto-Eastern Polynesian), but the D-estimate is only taken for the entire Oceanic tree, not for each sub-clade. The predictions for ”most common” were excluded, since there is not a tree *per se* which the D-estimate can take as input to measure the phylogenetic signal. In addition, we also excluded datapoints that were ill-fitting for other reasons as discussed in the previous section.

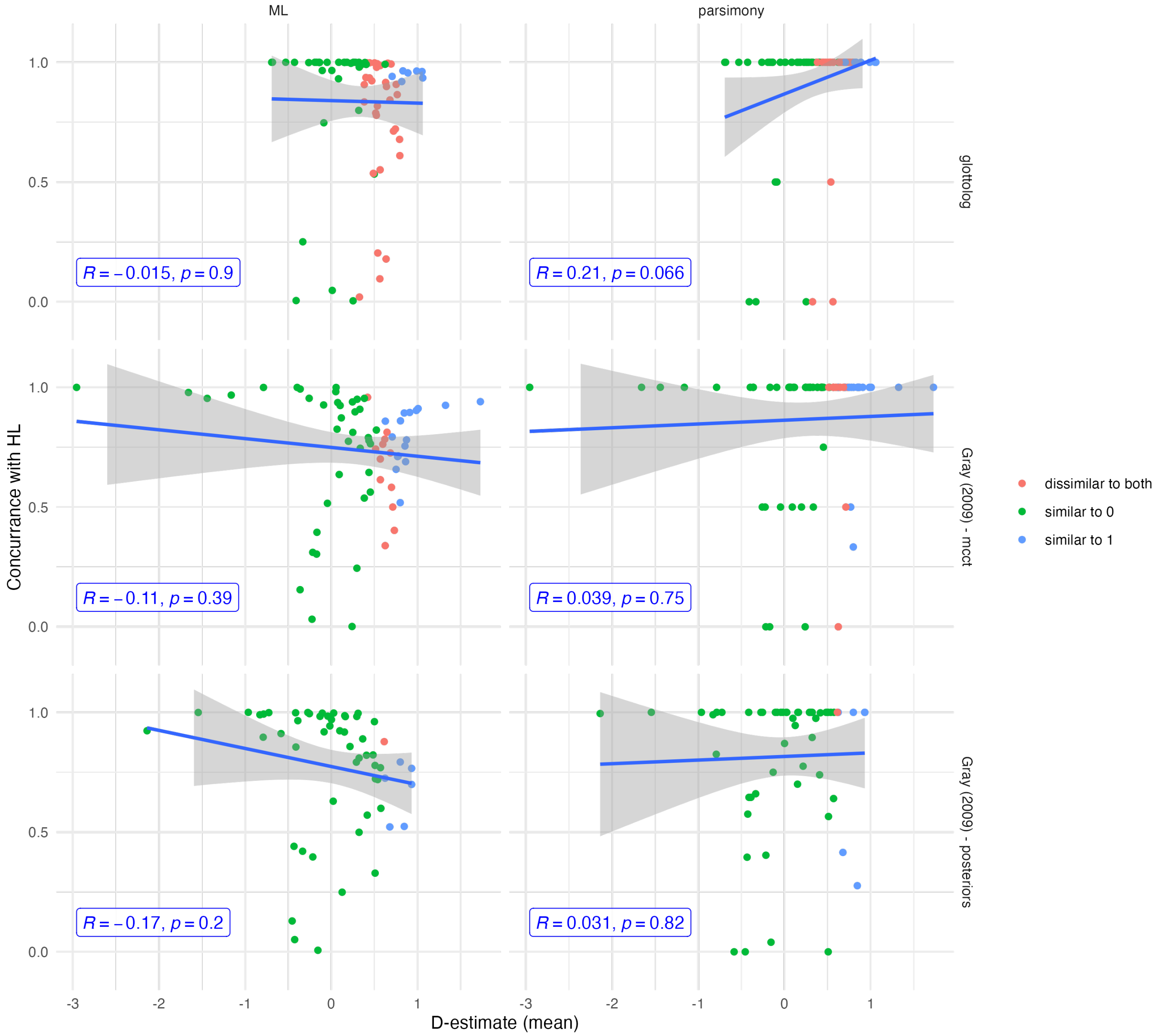


Figure 6: Scatter-plots of D-estimates (x-axis) and concurrence with conventional historical linguistics (y-axis). The points are coloured based on meeting statistical thresholds of significance for being similar to 0 (Brownian) or 1 (random). The correlation statistic in blue represents a Pearson-test.

I.4 R-function for sanity-checks

Because of the issues described here with D-estimates on certain data distributions, it is advisable to perform some sanity checks before measuring phylogenetic signal with this metric. I have written a function in R for this purpose, which is available publicly on GitHub in a package that is a work-in-progress. The package is still in alpha development, so please use it with caution.

```
library(remotes)
library(caper)
remotes::install_github("HedvigS/SH.misc@v0.1")
SH.misc::phylo.d_wrapper()
```

J Correlation value distributions and HL-concurrence

We can consider a much simpler approach to understanding what predicts agreement between the computational methods and historical linguists – the number of tips in each state. We see some strong patterns here compared to the D-estimate comparison. The idea is that if very few tips are in one state and all other tips in the other, there is little variation that can drive disagreements between the different reconstructions. If on the other hand, the states are distributed 50%/50% then it is reasonable to assume there is a greater chance for disagreement. In fig 7, the x-axis is the percentage of tips in the minority state – 0% indicates that all tips are of the same state (be that presence or absence) and 50% that half of the tips are in one state, half in another. 30% indicates that the state with the fewest tips had 30% of the tips. The y-axis represents concurrence with traditional historical linguistics. Each point is one structural feature in one of the four proto-languages.

All of the comparisons between HL-concurrence and the percentage of tips in minority state have a p-value lower than 0.05, which is a commonly used cut-off for statistical significance for Pearson correlation-tests. All correlations are negative, which is to be expected. This indicates that when tips are more evenly distributed between the two states (closer to 50% on the x-axis), there is more disagreement between the methods and traditional HL. Half of the correlations are weak (between 0.2 and 0.39), and half are of moderate strength (between 0.40 - 0.59). There are some outliers in the lower left quadrant of each plots, these represent cases where most tips are in one state and yet there is a disagreement. One of them is discussed in greater detail in the following section.

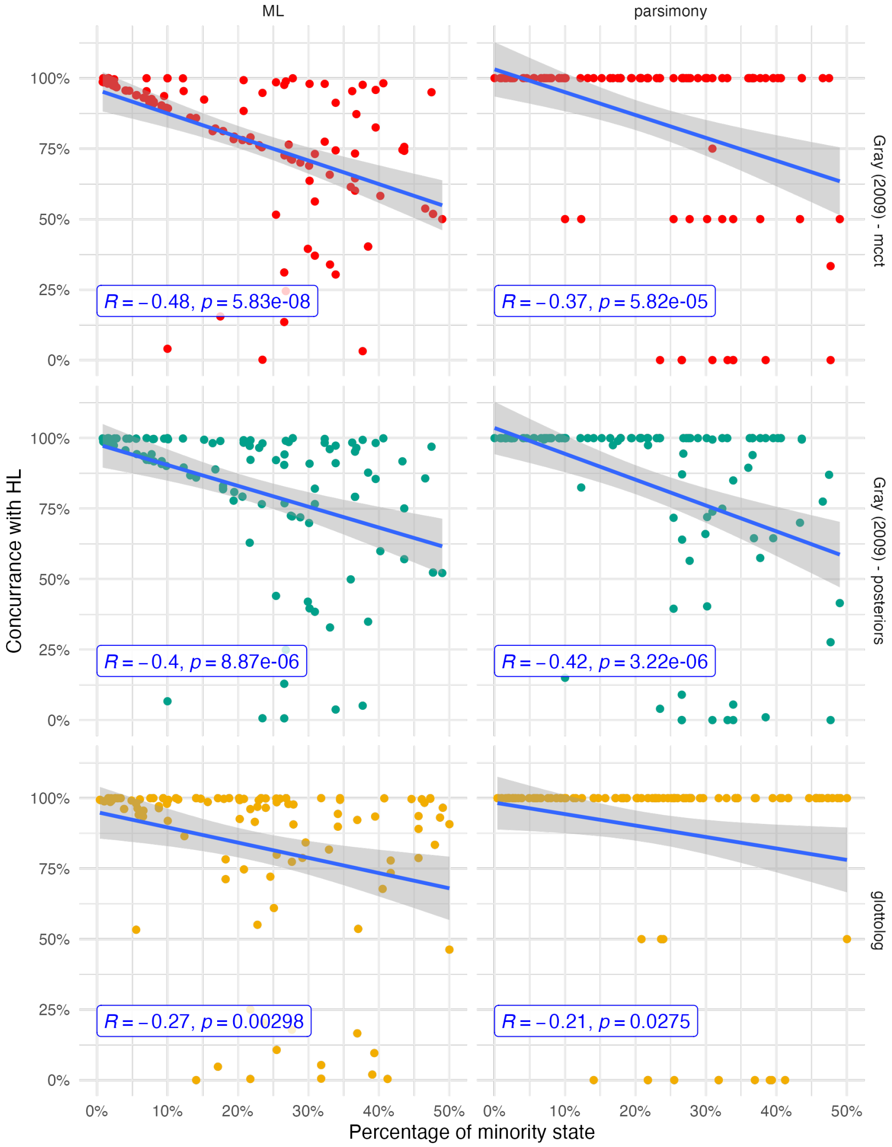


Figure 7: Scatter-plots of precentage of tips in minority state (x-axis) and concurrence with conventional historical linguistics (y-axis). The correlation statistic in blue represents a Pearson-test.

K Disagreement between methods detail

One example of disagreement between conventional HL, Maximum Parsimony and Maximum Likelihood is GB133 'Is a pragmatically unmarked constituent order verb-final for transitive clauses¿ for Proto-Oceanic. This feature has a very low concurrence with HL for the ML method (0.04) and (Gray et al. 2009)) MCC-tree, despite the tip state distribution being 14%/86% which we saw in the previous section (J) usually predicts high agreement.

Let us first consider the historical linguistics literature and the feature at hand. The coding of Proto-Oceanic as present for this feature according to conventional historical linguistics is based on the following passage from Pawley (1973):

> *Capell's suggestion that the SOV order found in many New Guinea Oceanic languages is the result of influence by Papuan (non-Austronesian) languages, almost all of which show SOV order, seems reasonable.* [..] *Still, the fact that the better-known SVO languages also tolerate certain other orders (for non- pronominal constituents) suggests that some variation occurred in POC* [Proto-Oceanic]. *In particular, occurrences of OSV and VOS order are widely distributed enough to indicate that both were possible in POC.*[3]
>
> Pawley (1973: 118)

Unlike the chapter in the World Atlas of Language Structures on order in the transitive clause (Dryer 2013), the Grambank feature questionnaire does not ask about the "dominant"-type, but has 3 different binary questions about the "pragmatically unmarked" order.

- GB131 Is a pragmatically unmarked constituent order verb-initial for transitive clauses?
- GB132 Is a pragmatically unmarked constituent order verb-medial for transitive clauses?
- GB133 Is a pragmatically unmarked constituent order verb-final for transitive clauses?

It is possible for a language to be answered "yes" for more than one question if multiple orders occur (without changing the pragmatics). However, most Oceanic languages were still coded as absent for GB133. The Maximum Parsimony and Maximum Likelihood all disagree with conventional HL regarding GB133 for Proto-Oceanic - but in different ways.

Fig 8 shows the Ancestral Nodes of GB133 on the Gray et al (2009)-MCCT with the parsimony method, and Fig 9 the same tree but with the Maximum Likelihood method. These two tree figures have the same exact topology and tip states, they only vary in the reconstruction of internal nodes (proto-languages) due to the different methods used. In each of the figures, there is a set of languages at the bottom of the

[3] Pawley does note that the "basic" word order in Proto-Oceanic is likely to be SVO (Subject-Verb-Object).

tree that are coded as ”yes” for GB133 and these are located on the island of New Guinea or nearby. Their location in the tree is such that they form a clade that is an early offshoot from the root. For the parsimony method, that means that even though most of the tips are of another state, this group carries a lot of weight. The parsimony method suggests that the state of the root, of Proto-Oceanic, is 50%/50%. However, the Maximum Likelihood method takes into account branch lengths and the overall tendency in the tree for the trait value ”absent” to be stable (because it estimates asymmetric rates, unlike MP which assumes symmetric rates). This results in a ML-estimation of proto-Oceanic as overwhelmingly absent of verb-finality.

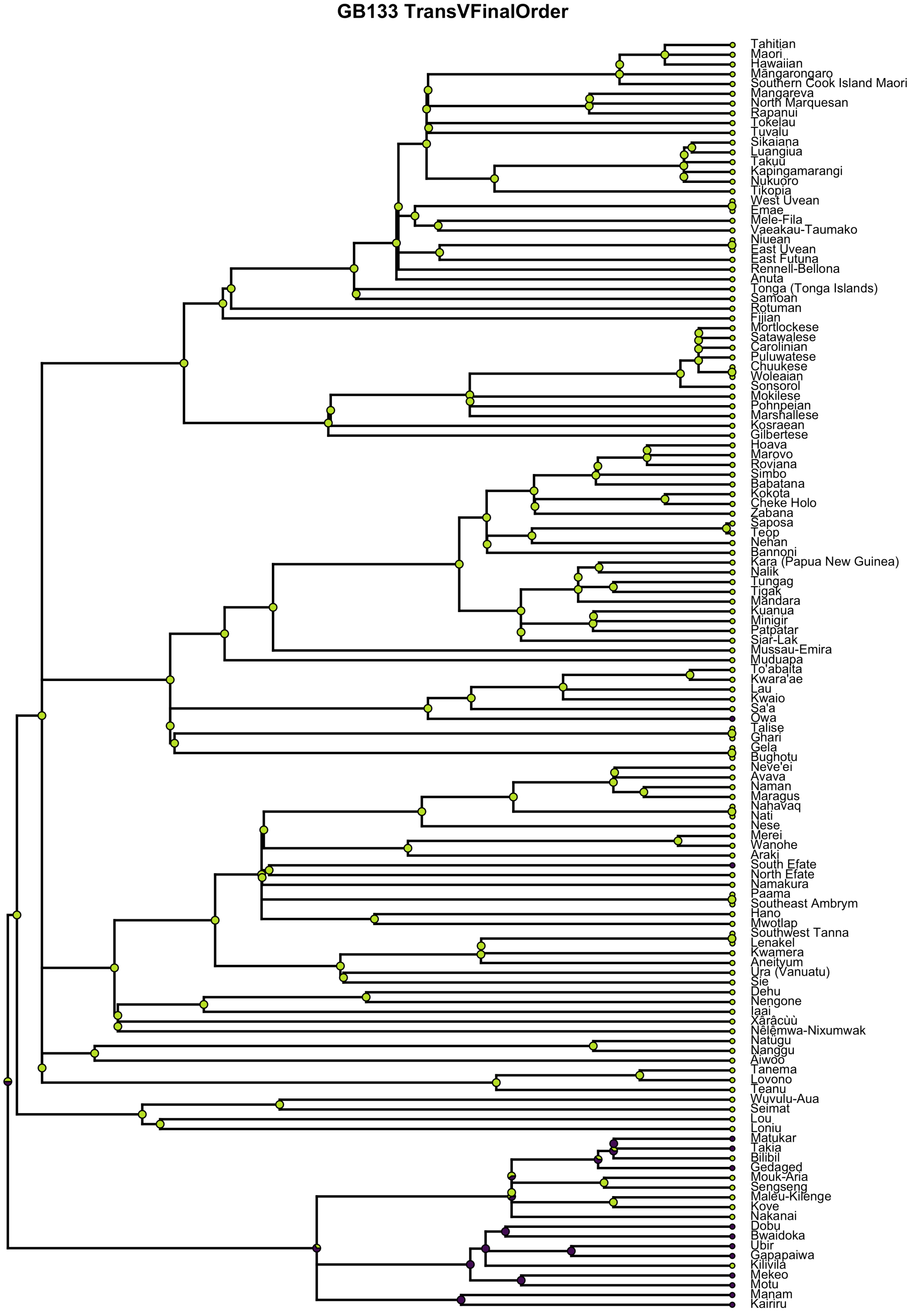


Figure 8: Gray et al 2009-tree with Maximum Parsimony method, Proto-Oceanic is reconstructed as half/half present/absent. Green = absent, purple = present. Root edge added in for visualisation purposes only.

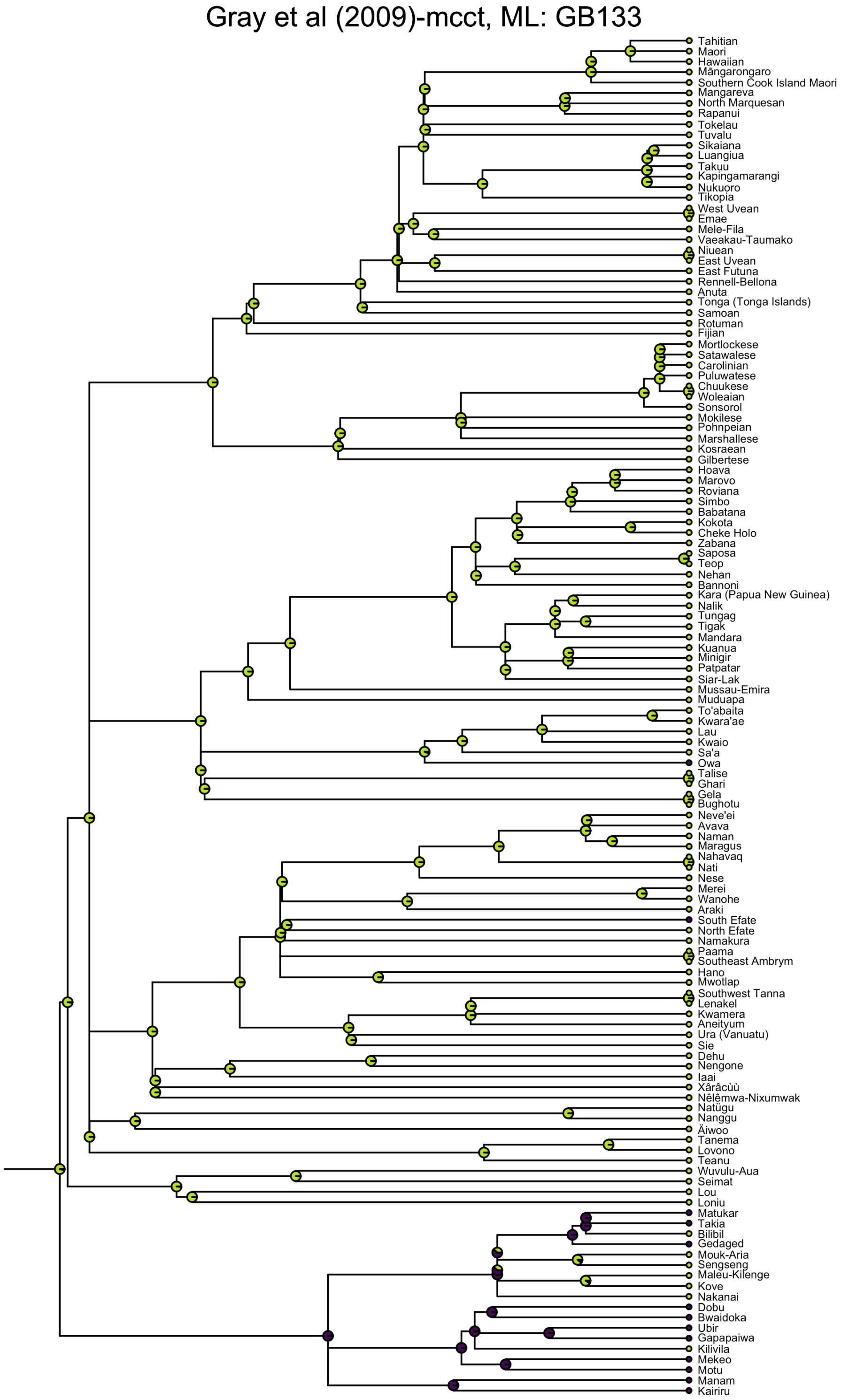


Figure 9: Gray et al 2009-tree with ML method, Proto-Oceanic is reconstructed as absent. Green = absent, purple = present. Root edge added in for visualisation purposes only.

L F1-scores

F1-scores are the harmonic mean of the precision and recall[4] (Rijsbergen 1979: 133). It is important to note that F1-scores disregard the number of True Negatives entirely, which is relevant in our case since some of the features in proto-languages are predicted to be absent. For both measures, 0 is the worst possible score and 1 the best in terms of similarity to predictions by historical linguists.

In a similar study of ancestral states of cognate classes, Jäger and List (2018) compared three different methods of ancestral state reconstruction for lexical data (cognate classes): Maximum Parsimony, Maximum Likelihood and Minimal Lateral Networks. They found that reconstructions using Maximum Likelihood performed the most like the predictions by historical linguists. However, Jäger and List (2018) describe the general performance of all the computational reconstruction methods they used as "poor". Jäger and List (2018) evaluated the methods using the F1-score. The highest F1-score was 0.79 (Austronesian language sample, Maximum Likelihood), and the worst was 0.44 (Indo-European, Minimal Lateral Networks).

The formula for F1-scores is given in Eq. 2.

$$\frac{\text{True Positive}}{\text{True Positive} + \frac{1}{2} \times (\text{False Positive} + \text{False Negative})} \tag{2}$$

As stated in §3.1 in the main text, the half-results are also interesting, the formula for F1-scores including half-results is given in Eq. 3. For more on the caluclation of the F1-score including half results, see Supplementary Material M.

$$\frac{\text{True Positive} + \frac{\text{Half}}{2}}{\text{True Positive} + \frac{1}{2} \times (\text{False Positive} + \text{False Negative}) + \text{Half}} \tag{3}$$

The results of the F1-scores are shown in Fig 10, alongside the concordance scores. The result for the plain F1-score differs from the other three, this is precisely because it ignores True Negatives. While True Negatives are not included *per se* in the calculation of F1 including half-results score, the inclusion of the half-similarity still has an impact as it makes all the methods more similar.

[4] Precision is True Positives divided by True Positives + False Positives, recall is True Positives divided by False Negatives + True Positives. F1-score = 2 * ((precision*recall) / (precision + recall)) (Rijsbergen 1979).

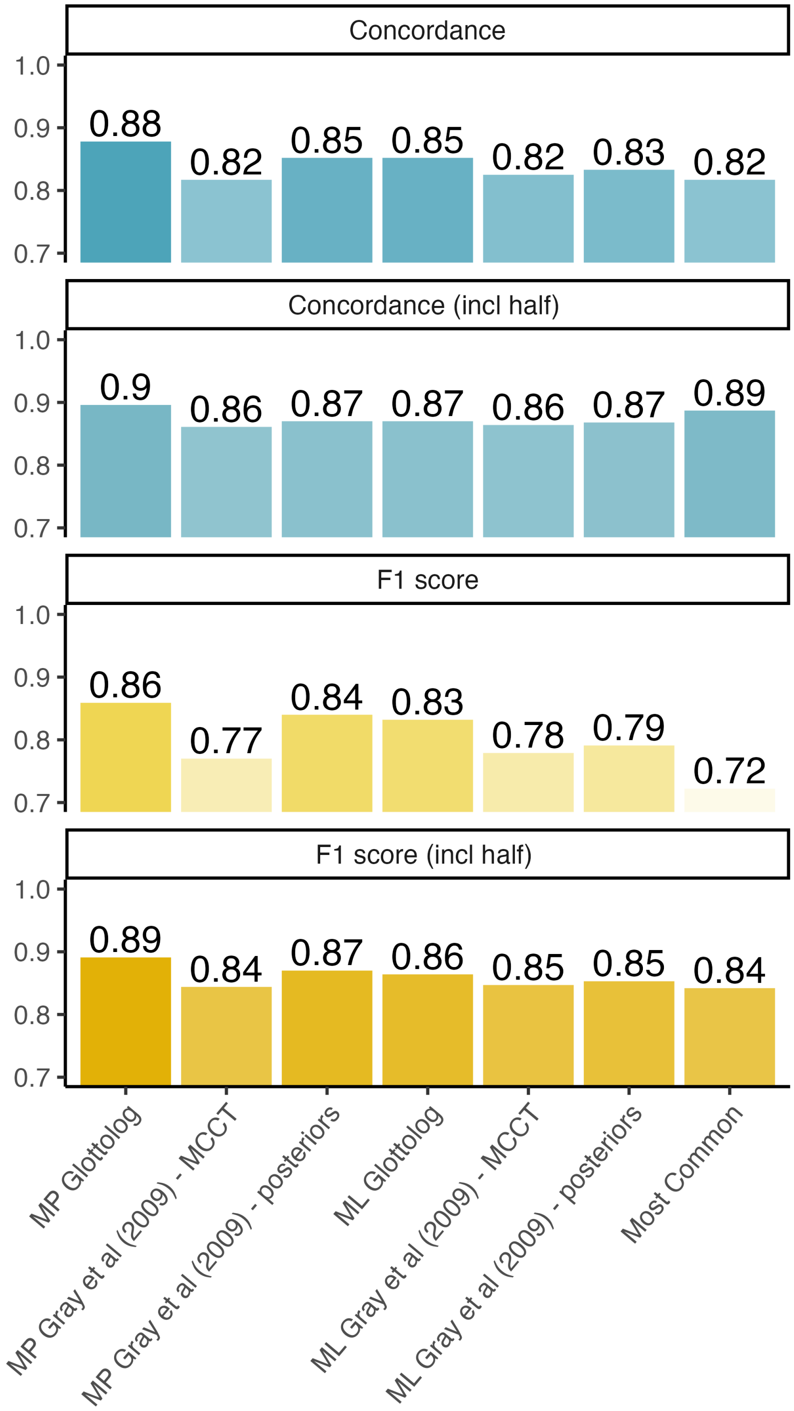


Figure 10: **Barplots of concordance and F1-scores of each method.** NB that the y-axis starts from 0.7.

Compared to the F1-scores from the lexical reconstruction of Jäger and List (2018), all of the methods achieved higher scores. The highest ("best") F1-score in Jäger and List (2018) was 0.79 (Austronesian language sample, Maximum Likelihood), and the worst was 0.44 (Indo-European, Minimal Lateral Networks). In this study, only statements about ancestral languages that could be mapped to Grambank-features were included. It is possible that the study by Jäger and List (2018) had a greater overlap between all the reconstructions made by historical linguists and the meanings that they had data for. In that case, it is possible that the features that were possible to map to Grambank data were also those that Oceanic historical linguists are the most confident about – hence the higher scores of agreement (quantified as F1-scores) compared to Jäger and List (2018).

M Mathematics of the F1-score including half-results

I am very grateful for the assistance of Stephen Mann in working out the mathematics of these scores as they incorporate the Half-results.

M.1 Standard definitions

The F1-score is the harmonic mean of precision and recall (Rijsbergen 1979).

$$\begin{aligned} F_1 &= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \\ &= \frac{\text{TP}}{\text{TP} + \frac{1}{2} \times (\text{FP} + \text{FN})} \end{aligned}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

M.2 Half-result definitions of precision and recall

The half-result-definitions of precision and recall add one half of the half-counts to the numerator, and all of the half-counts to the denominator:

$$\text{precision}_{\text{half}} = \frac{\text{TP} + \frac{\text{H}}{2}}{\text{TP} + \text{FP} + \text{H}}$$

$$\text{recall}_{\text{half}} = \frac{\text{TP} + \frac{\text{H}}{2}}{\text{TP} + \text{FN} + \text{H}}$$

M.3 The question

We want to define $F_{1,\mathrm{half}}$. A natural way to do it would be to follow the rule defined above, i.e.

$$F_{1,\mathrm{half?}} = \frac{\mathrm{TP} + \frac{\mathrm{H}}{2}}{\mathrm{TP} + \frac{1}{2} \times (\mathrm{FP} + \mathrm{FN}) + \mathrm{H}}$$

However, we want to ensure $F_{1,\mathrm{half}}$ has the same relationship with $\mathrm{precision}_{\mathrm{half}}$ and $\mathrm{recall}_{\mathrm{half}}$ as F_1 has with precision and recall. So we need to determine whether the following equation is true:

$$2 \times \frac{\mathrm{precision}_{\mathrm{half}} \times \mathrm{recall}_{\mathrm{half}}}{\mathrm{precision}_{\mathrm{half}} + \mathrm{recall}_{\mathrm{half}}} \overset{?}{=} \frac{\mathrm{TP} + \frac{\mathrm{H}}{2}}{\mathrm{TP} + \frac{1}{2} \times (\mathrm{FP} + \mathrm{FN}) + \mathrm{H}} \tag{4}$$

M.4 The proof

We will expand the left-hand side of (4) and show it is equal to the right-hand side. Let's forget about the 2× for now (we will reintroduce it at the end). Expanding the numerator gives:

$$\frac{\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)}{(\mathrm{TP} + \mathrm{FP} + \mathrm{H})(\mathrm{TP} + \mathrm{FN} + \mathrm{H})}$$

Expanding the denominator gives:

$$\begin{aligned}
&\frac{\mathrm{TP} + \frac{\mathrm{H}}{2}}{\mathrm{TP} + \mathrm{FP} + \mathrm{H}} + \frac{\mathrm{TP} + \frac{\mathrm{H}}{2}}{\mathrm{TP} + \mathrm{FN} + \mathrm{H}} \\
&= \frac{\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)(\mathrm{TP} + \mathrm{FN} + \mathrm{H})}{(\mathrm{TP} + \mathrm{FP} + \mathrm{H})(\mathrm{TP} + \mathrm{FN} + \mathrm{H})} + \frac{\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)(\mathrm{TP} + \mathrm{FP} + \mathrm{H})}{(\mathrm{TP} + \mathrm{FN} + \mathrm{H})(\mathrm{TP} + \mathrm{FP} + \mathrm{H})} \\
&= \frac{\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)(2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN} + 2 \times \mathrm{H})}{(\mathrm{TP} + \mathrm{FP} + \mathrm{H})(\mathrm{TP} + \mathrm{FN} + \mathrm{H})}
\end{aligned}$$

When we put the numerator back on top of the denominator, both of their respective denominators cancel out, because they are both (TP+FP+H)(TP+FN+H). So we end up with *the numerator of the numerator* on top of *the numerator of the denominator*, like so:

$$\begin{aligned}
&\frac{\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)}{\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)(2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN} + 2 \times \mathrm{H})} \\
&= \frac{\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN} + 2 \times \mathrm{H}}
\end{aligned}$$

Finally, we bring back the 2× from the beginning:

$$\begin{aligned} &2 \times \frac{\left(\mathrm{TP} + \frac{\mathrm{H}}{2}\right)}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN} + 2 \times \mathrm{H}} \\ &= \frac{\mathrm{TP} + \frac{\mathrm{H}}{2}}{\mathrm{TP} + \frac{1}{2} \times (\mathrm{FP} + \mathrm{FN}) + \mathrm{H}} \end{aligned}$$

And we have our suggested definition of $F_{1,\mathrm{half}}$ as required.

N Further details on the tree phylogeny

The tree from Gray et al. (2009) contains duplicates in terms of glottocodes (see for example Nakanai). This is because it is a tree of word-lists for languages (doculects) rather than languages themselves. There are also some instances where multiple dialects of one language are included. For the analysis, only one tip per language was retained, based on which had best coverage in the underlying data for the tree (i.e., the Austronesian Basic Vocabulary Database, ABVD (Greenhill et al. 2008)). This means that duplicate glottocodes were reduced to one, be it due to multiple word-lists or dialects. The specific analytical choices are found in the following three R-scripts:

- Oceanic_computational_ASR/code/01_requirements.R
- Oceanic_computational_ASR/code/analysis_scripts_gray_mcct/ 03_get_gray_tree_mcct.R
- Oceanic_computational_ASR/code/analysis_scripts_gray_all_posterior/ 03_process_gray_tree_posteriors.R

For both Maximum Parsimony and Maximum Likelihood the tree were first pruned down to only languages where there is data in Grambank for each given feature, i.e., the ASR-analysis never contains tips with missing or ambiguous data. Missing data vary with features, so each analysis per tree and method differs in number of tips.

Regarding branch lengths, most of the trees in the analysis are not ultrametric, i.e., the distances between the tips and the roots are not all the same. If we use trees to represent history and time, then an ultrametric, or near-Ultrametric, tree is a more reasonable representation of said histories when we assume that the languages at the tips existed at the same time. Fig 11 illustrates different configurations of branch lengths using only Nuclear Polynesian languages. It is reasonable to assume that the data gathered on these languages represent similar time-slices to each other, i.e., the representation of Rapa Nui is not considerably "younger" as a language than Tongan. If the tips included ancient languages, such as Sanskrit or Akkadian, it may be possible for such tips to have a shorter distance to the root than the others. However, if the languages are of a similar "age", the tree ought to be ultrametric or near-ultrametric if we understand tree length as representing time.

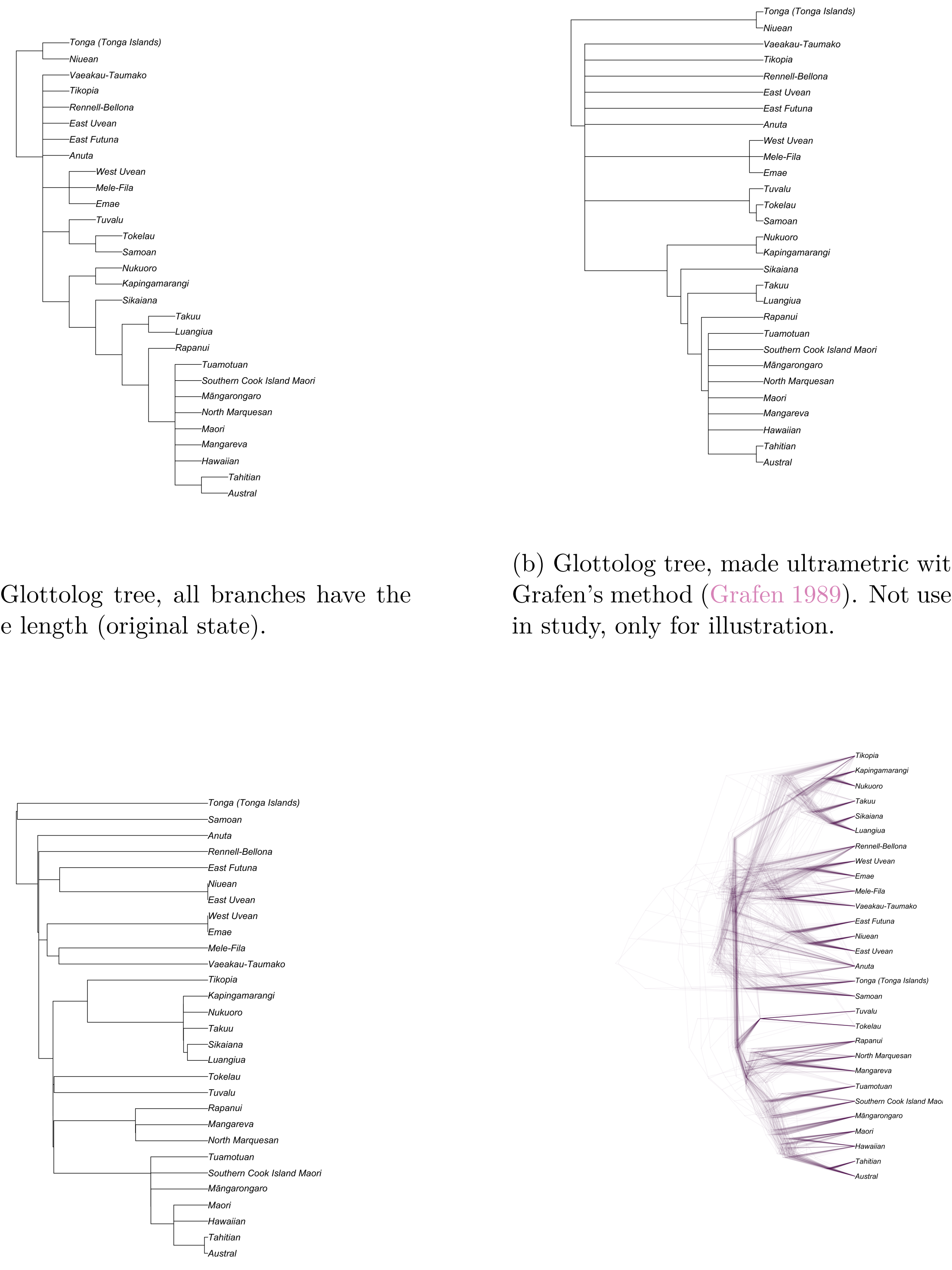
The Glottolog genealogical classification follows common principles in historical linguistics by focusing on the validity of subgrouping, not branch lengths. The Glottolog 4.5 tree does not include any information about branch lengths. This is interpreted as the same as if all branches are of the same length (they are explicitly all set to 1 in the analysis). In order to illustrate what this entails, consider the difference between Figure 11a and Figure 11b. The first is the Glottolog tree of Nuclear Polynesian as found originally, i.e., with all branches of the same length (1). The second is a transformation of the first, it has been made ultrametric by Grafen's transform (Grafen 1989), which is one of several approaches to making a tree ultrametric in lieu of branch lengths directly from the data. The second tree is *not* used in the analysis, it is included here only to illustrate how the same subgrouping of languages can be expressed when the branch lengths are changed. Transformations of branch lengths should be carried out with great care and with good reason. It is not clear what transformation of branch lengths in the Glottolog 4.5-tree is appropriate, which is why none has been carried out in the analysis of this paper. Keeping the Glottolog 4.5-tree lacking branch lengths may also be more true to historical linguistics methodology.

The branch lengths in the trees from Gray et al. (2009) are derived from the dynamics of the data – the word-lists – and certain priors regarding island settlement. The MCC-tree of Gray et al. (2009) is not ultrametric, but very close to it (see Fig 11c. After pruning to the subset that overlaps with Oceanic languages in Grambank (as described above), the difference between the tip with the largest distance to the root and the smallest is tiny (3.408377 - 3.408362). In the random sample of 100 from the 4,200 posteriors trees, the case is the same as with the MCCT – they are not perfectly ultrametric but very close (see Fig 11d. The trees from Gray et al. (2009) may not be perfectly ultrametric, but they are much more closer to ultrametric than the Glottolog tree, as can be seen by comparing the visualisations in Fig 11.

Concerning binary splits, there are non-binary splits in the Glottolog 4.5 tree, the Gray et al. (2009) MCCT and in 64 of the 100 posterior trees. For this analysis, I have chosen to not resolve these polytomies into binary splits in order to stay as true as possible to the original phylogeny. There are branches of length 0 in the MCCT and posterior trees. It is not possible to collapse these into polytomies as this may in cases introduce basal polytomies. Instead, 0.00011 length was added to all branches. Doing this removes branches of length 0 while maintaining the relative lengths of all branches in the tree.

In some cases, pruning a given posterior tree to the relevant tips resulted in the tree becoming unrooted. In such cases, the tree was re-rooted using midpoint rooting `castor::root_at_midpoint()` Louca (2023). There were 4 such cases in the random sample of posteriors trees (random seed = 147).

All of the wrangling of the trees is found in the data analysis R-scripts that accompany this paper.



(a) Glottolog tree, all branches have the same length (original state).

(b) Glottolog tree, made ultrametric with Grafen's method (Grafen 1989). Not used in study, only for illustration.

(c) Gray et al. (2009) MCCT.

(d) Gray et al. (2009) posteriors tree (random sample of 100). The densitree-visualisation is in phylogram-style.

Figure 11: Four trees of Nuclear Polynesian, demonstrating branch lengths.

O Further details on the Grambank coding of proto-languages

Example of how information in the publications was turned into Grambank feature coding relating to verbal markers encoding subjects and objects, as proposed by Lynch et al. (2011) among others. In their book, there is a paper on reconstructions of grammar for Proto-Oceanic and in the section on the basic verb phrase we find the statement below:

> *Attached to the verb root were a subject proclitic and, if the verb had a non-generic object, an object enclitic.*
>
> Lynch et al. (2011: 83)

This statement, together with a verb schema provided in the section, support the notion that Proto-Oceanic had subject proclitics and object enclitics. We can also infer from this publication as a whole that the authors believe Proto-Oceanic in fact did *not* have subject *en*clitics and object *pro*clitics. This second prediction relies on the absence of evidence and is less strong than the first, but given that the whole paper is void of any description of object proclitics or subject enclitics being a possibility (including the verb schema) and argument structure is well-discussed, we may dare to make this leap. This information can be translated into the Grambank questionnaire by positing absence and presence for the six relevant features that concern argument marking on the verb (where S stands for subject of intransitive, A for subject of transitive and O for object; see table 5).

Table 5: Example of predictions from historical linguistics as rendered in Grambank features.

Grambank ID	**Question**	**Proto-language**	**Expert prediction**	**Reference**
GB089	Can the S argument be indexed by a suffix/enclitic on the verb in the simple main clause?	Proto-Oceanic	Absent	Ross (2004: 498-499), Lynch et al. (2011: 83)
GB090	Can the S argument be indexed by a prefix/proclitic on the verb in the simple main clause?	Proto-Oceanic	Present	Ross (2004: 498-499), Lynch et al. (2011: 83)
GB091	Can the A argument be indexed by a suffix/enclitic on the verb in the simple main clause?	Proto-Oceanic	Absent	Ross (2004: 498-499), Lynch et al. (2011: 83)
GB092	Can the A argument be indexed by a prefix/proclitic on the verb in the simple main clause?	Proto-Oceanic	Present	Ross (2004: 498-499), Lynch et al. (2011: 83)
GB093	Can the P argument be indexed by a suffix/enclitic on the verb in the simple main clause?	Proto-Oceanic	Present	Ross (2004: 498-499), Lynch et al. (2011: 83)
GB094	Can the P argument be indexed by a prefix/proclitic on the verb in the simple main clause?	Proto-Oceanic	Absent	Ross (2004: 498-499), Lynch et al. (2011: 83)

P Supplementary Material bibliography

References

et al., R Hackathon. 2020. *phylobase: Base package for phylogenetic structures and comparative data.* https://github.com/fmichonneau/phylobase. R package version 0.8.10.

Ball, Douglas. 2007. On ergativity and accusativity in proto-polynesian and proto-central pacific. *Oceanic Linguistics* 128–153. doi:10.1353/ol.2007.0014.

Beaulieu, Jeremy, Brian O'Meara, Jeffrey Oliver and James Boyko. 2022. *corhmm: Hidden markov models of character evolution.* https://CRAN.R-project.org/package=corHMM. R package version 2.8.

Brooks-Bartlett, Jonny. 2018. Probability concepts explained: Maximum likelihood estimation. https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1.

Brownrigg, Ray. 2022. *maps: Draw geographical maps.* https://CRAN.R-project.org/package=maps. R package version 3.4.1.

Bååth, Rasmus. 2018. *beepr: Easily play notification sounds on any platform.* https://CRAN.R-project.org/package=beepr. R package version 1.3.

Ching, Travers. 2023. *qs: Quick serialization of r objects.* https://github.com/traversc/qs. R package version 0.25.5.

Chung, Sandra. 1978. *Case marking and grammatical relations in polynesian languages.* Austin: University of Texas. doi:10.7560/710511.

Clark, D. Ross. 1973. *Aspects of proto-polynesian syntax*: University of California San Diego dissertation.

Cooper, Nicholas. 2022. *Ncmisc: Miscellaneous functions for creating adaptive functions and scripts.* https://CRAN.R-project.org/package=NCmisc. R package version 1.2.0.

Crowley, Terry. 1985. Common noun phrase marking in proto-oceanic. *Oceanic Linguistics* 24(1/2). 135–193.

Csárdi, Gábor, Jim Hester, Hadley Wickham, Winston Chang, Martin Morgan and Dan Tenenbaum. 2023. *remotes: R package installation from remote repositories, including github.* https://CRAN.R-project.org/package=remotes. R package version 2.4.2.1.

Cunningham, Clifford W., Kevin E. Omland and Todd H. Oakley. 1998. Reconstructing ancestral character states: a critical reappraisal. *Trends in Ecology & Evolution* 13(9). 361–366. doi:10.1016/s0169-5347(98)01382-2.

Dahl, David B., David Scott, Charles Roosen, Arni Magnusson and Jonathan Swinton. 2019. *xtable: Export tables to latex or html*. http://xtable.r-forge.r-project.org/. R package version 1.8-4.

Dowle, Matt and Arun Srinivasan. 2023. *data.table: Extension of 'data.frame'*. https://CRAN.R-project.org/package=data.table. R package version 1.14.8.

Dryer, Matthew S. 2013. Order of subject, object and verb (v2020.3). In Matthew S. Dryer and Martin Haspelmath (eds.), *The world atlas of language structures online*, Zenodo. doi:10.5281/zenodo.7385533.

Evans, Bethwyn. 2001. *A study of valency-changing devices in proto oceanic*: Research School of Pacific and Asian Studies, The Australian National University Phd thesis.

Felsenstein, Joseph. 2004. *Inferring phylogenies*, vol. 2. Sunderland, MA: Sinauer Associates.

Firke, Sam. 2023. *janitor: Simple tools for examining and cleaning dirty data*. https://CRAN.R-project.org/package=janitor. R package version 2.2.0.

Fritz, Susanne A. and Andy Purvis. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* 24(4). 1042–1051. doi:10.1111/j.1523-1739.2010.01455.x.

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Antônio Pedro, Sciaini, Marco, Scherer and Cédric. 2023a. *viridis(Lite) - colorblind-friendly color maps for r*. doi:10.5281/zenodo.4679424. https://sjmgarnier.github.io/viridis/. Viridis package version 0.6.3.

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Antônio Pedro, Sciaini, Marco, Scherer and Cédric. 2023b. *viridis(Lite) - colorblind-friendly color maps for r*. doi:10.5281/zenodo.4678327. https://sjmgarnier.github.io/viridis/. ViridisLite package version 0.4.2.

Garnier, Simon. 2023a. *viridis: Colorblind-friendly color maps for r*. https://CRAN.R-project.org/package=viridis. R package version 0.6.3.

Garnier, Simon. 2023b. *viridislite: Colorblind-friendly color maps (lite version)*. https://CRAN.R-project.org/package=viridisLite. R package version 0.4.2.

Grafen, Alan. 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 326(1233). 119–157. doi:10.1098/rstb.1989.0106.

Gray, R. D., A. J. Drummond and S. J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in pacific settlement. *Science* 323(5913). 479–483. doi:10.1126/science.1166858.

Greenhill, Simon J., Robert Andrew Blust and Russell D. Gray. 2008. The austronesian basic vocabulary database: From bioinformatics to lexomics. *Evolutionary Bioinformatics* 4. 271–283. doi:10.4137/ebo.s893.

Grolemund, Garrett and Hadley Wickham. 2011. Dates and times made easy with lubridate. *Journal of Statistical Software* 40(3). 1–25. https://www.jstatsoft.org/v40/i03/.

Hammarström, Harald, Robert Forkel, Martin Haspelmath and Sebastian Bank. 2021. Glottolog/glottolog: Glottolog database 4.5. doi:10.5281/zenodo.5772642.

Hester, Jim, Hadley Wickham and Jennifer Bryan. 2023a. *vroom: Read and write rectangular text data quickly*. https://CRAN.R-project.org/package=vroom. R package version 1.6.5.

Hester, Jim, Hadley Wickham and Gábor Csárdi. 2023b. *fs: Cross-platform file system operations based on libuv*. https://CRAN.R-project.org/package=fs. R package version 1.6.3.

Jäger, Gerhard and Johann-Mattis List. 2018. Using ancestral state reconstruction methods for onomasiological reconstruction in multilingual word lists. *Language Dynamics and Change* 8(1). 22–54. doi:10.1163/22105832-00801002.

Jombart, T. and S. Dray. 2010. adephylo: exploratory analyses for the phylogenetic comparative method. *Bioinformatics* 26. 1907–1909. doi:10.1093/bioinformatics/btq292.

Jombart, Thibaut, Stéphane Dray and Anders Ellern Bilgrau. 2022. *adephylo: Exploratory analyses for the phylogenetic comparative method*. https://CRAN.R-project.org/package=adephylo. R package version 1.1-13.

Jonsson, Niklas. 1998. Det polynesiska verbmorfemet - cia; om dess funktion i samoanska [the polynesian verbal morphene -cia; about its function in samoan].

Joy, Jeffrey B., Richard H. Liang, Rosemary M. McCloskey, T. Nguyen and Art F.Y. Poon. 2016. Ancestral reconstruction. *PLoS Computational Biology* 12(7). e1004763. doi:10.1371/journal.pcbi.1004763.

Kassambara, Alboukadel. 2023. *ggpubr: ggplot2 based publication ready plots*. https://rpkgs.datanovia.com/ggpubr/. R package version 0.6.0.

Kikusawa, Ritsuko. 2001. Rotuman and fijian case-marking strategies and their historical development. *Oceanic Linguistics* 40(1). 85–111. doi:10.1353/ol.2001.0008.

Kikusawa, Ritsuko. 2002. *Proto central pacific ergativity: Its reconstruction and development in the fijian, rotuman and polynesian languages*. Canberra: Pacific Linguistics.

Kirby, K.R., R.D. Gray, S.J. Greenhill, F.M. Jordan, S. Gomes-Ng, H.J. Bibiko, Damián E. Blasi, Carlos A. Botero, Claire Bowern, Carol R. Ember, Dan Leehr, Bobbi S. Low, Joe McCarter, William Divale and Michael C. Gavin. 2018. D-place/dplace-data: D-place – the database of places, language, culture and environment (version v2.0.1). doi:10.5281/zenodo.1466634.

Louca, Stilianos. 2023. *castor: Efficient phylogenetics on large trees.* https://CRAN.R-project.org/package=castor. R package version 1.7.10.

Louca, Stilianos and Michael Doebeli. 2017a. Efficient comparative phylogenetics on large trees. *Bioinformatics* doi:10.1093/bioinformatics/btx701.

Louca, Stilianos and Michael Doebeli. 2017b. Efficient comparative phylogenetics on large trees. *Bioinformatics* 34(6). 1053–1055. doi:10.1093/bioinformatics/btx701.

Lucas, Antoine, Immanuel Scholz, Rainer Boehme, Sylvain Jasson and Martin Maechler. 2023. *gmp: Multiple precision arithmetic.* https://forgemia.inra.fr/sylvain.jasson/gmp. R package version 0.7-2.

Lynch, John, Malcolm Ross and Terry Crowley. 2011. Proto oceanic. In John Lynch, Malcolm Ross and Terry Crowley (eds.), *The oceanic languages* Curzon Language Family Series, 54–91. Richmond: Curzon 2nd edition.

Maechler, Martin. 2023. *Rmpfr: R mpfr - multiple precision floating-point reliable.* https://rmpfr.r-forge.r-project.org/. R package version 0.9-2.

Maechler, Martin, Peter Rousseeuw, Anja Struyf and Mia Hubert. 2022. *cluster: "finding groups in data": Cluster analysis extended rousseeuw et al.* https://svn.r-project.org/R-packages/trunk/cluster/. R package version 2.1.4.

Marck, Jeffrey C. 2000. Polynesian languages. In J. Garry and C. Rubino (eds.), *Facts about the world's languages: An encyclopaedia of the world's major languages, past and present*, 560–567. New York: H.W. Wilson.

Müller, Kirill and Hadley Wickham. 2023. *tibble: Simple data frames.* https://CRAN.R-project.org/package=tibble. R package version 3.2.1.

Ooms, Jeroen. 2014. The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv:1403.2805 [stat.CO]* https://arxiv.org/abs/1403.2805.

Ooms, Jeroen. 2023. *jsonlite: A simple and robust json parser and generator for r.* https://jeroen.r-universe.dev/jsonlitehttps://arxiv.org/abs/1403.2805. R package version 1.8.7.

Orme, David, R. Freckleton, G. Thomas, Thomas Petzoldt, Susanne Fritz, Nick Isaac and Will Pearse. 2013. The caper package: comparative analysis of phylogenetics and evolution in r. *R package version* 5(2). 1–36.

Orme, David, Rob Freckleton, Gavin Thomas, Thomas Petzoldt, Susanne Fritz, Nick Isaac and Will Pearse. 2023. *caper: Comparative analyses of phylogenetics and evolution in r*. https://CRAN.R-project.org/package=caper. R package version 1.0.2.

Ottolinger, Philipp. 2019. *bib2df: Parse a bibtex file to a data frame*. https://github.com/ropensci/bib2df. R package version 1.1.1.

Pagel, Mark. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 255(1342). 37–45. doi:10.1098/rspb.1994.0006.

Paradis, Emmanuel, Simon Blomberg, Ben Bolker, Joseph Brown, Santiago Claramunt, Julien Claude, Hoa Sien Cuong, Richard Desper, Gilles Didier, Benoit Durand, Julien Dutheil, RJ Ewing, Olivier Gascuel, Thomas Guillerme, Christoph Heibl, Anthony Ives, Bradley Jones, Franz Krah, Daniel Lawson, Vincent Lefort, Pierre Legendre, Jim Lemon, Guillaume Louvel, Eric Marcon, Rosemary McCloskey, Johan Nylander, Rainer Opgen-Rhein, Andrei-Alin Popescu, Manuela Royer-Carenzi, Klaus Schliep, Korbinian Strimmer and Damien de Vienne. 2023. *ape: Analyses of phylogenetics and evolution*. https://CRAN.R-project.org/package=ape. R package version 5.7-1.

Paradis, Emmanuel and Klaus Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35. 526–528. doi:10.1093/bioinformatics/bty633.

Pawley, Andrew. 1970. Grammatical reconstruction and change in polynesia and fiji. In S.A. Wurm and D.C. Laycock (eds.), *Studies in honour of arthur capell*, 301–368. Canberra: Pacific Linguistics. https://openresearch-repository.anu.edu.au/bitstream/1885/253824/1/PL-C13.301.pdf.

Pawley, Andrew. 1973. Some problems in proto-oceanic grammar. *Oceanic Linguistics* 12(1/2). 103–188. doi:10.2307/3622854.

R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Ram, Karthik and Hadley Wickham. 2018. *wesanderson: A wes anderson palette generator*. https://github.com/karthik/wesanderson. R package version 0.3.6.

Revell, Liam J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3. 217–223. doi:10.1111/j.2041-210X.2011.00169.x.

Revell, Liam J. 2014. Ancestral state reconstruction. In *Presentation at the anthrotree workshop*, Duke University, Durham, NC. http://www.phytools.org/anthrotree/ancestral-states.pdf.

Revell, Liam J. 2023. *phytools: Phylogenetic tools for comparative biology (and other things)*. https://github.com/liamrevell/phytools. R package version 1.9-16.

Revelle, William. 2023. *psych: Procedures for psychological, psychometric, and personality research*. https://personality-project.org/r/psych/https://personality-project.org/r/psych-manual.pdf. R package version 2.3.6.

Rijsbergen, Cornelis Joost van. 1979. *Information retrieval*. Butterworths.

Ripley, Brian. 2023. *Mass: Support functions and datasets for venables and ripley's mass*. http://www.stats.ox.ac.uk/pub/MASS4/. R package version 7.3-60.

Ross, Malcolm D. 2004. The morphosyntactic typology of oceanic languages. *Language and Linguistics* 5(2). 491–541. http://hdl.handle.net/1885/87569.

Sankoff, David. 1975. Minimal mutation trees of sequences. *SIAM Journal on Applied Mathematics* 28(1). 35–42. doi:10.1137/0128004.

Schliep, Klaus, Potts, Alastair J., Morrison, David A., Grimm and Guido W. 2017. Intertwining phylogenetic trees and networks. *Methods in Ecology and Evolution* 8(10). 1212–1220.

Schliep, Klaus, Emmanuel Paradis, Leonardo de Oliveira Martins, Alastair Potts and Iris Bardel-Kahr. 2023. *phangorn: Phylogenetic reconstruction and analysis*. https://CRAN.R-project.org/package=phangorn. R package version 2.11.1.

Schliep, K.P. 2011. phangorn: phylogenetic analysis in r. *Bioinformatics* 27(4). 592–593. doi:10.1093/bioinformatics/btq706.

Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latarche, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore, Angela Chira, Luke Maurits, Russell Dinnage, Michael Dunn, Ger Reesink, Ruth Singer, Claire Bowern, Patience Epps, Jane Hill, Outi Vesakoski, Martine Robbeets, Noor Karolin Abbas, Daniel Auer, Nancy A. Bakker, Giulia Barbos, Robert D. Borges, Swintha Danielsen, Luise Dorenbusch, Ella Dorn, John Elliott, Giada Falcone, Jana Fischer, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima A. Goodall, Victoria Gruner, Andrew Harvey, Rebekah Hayes, Leonard Heer, Roberto E. Herrera Miranda, Nataliia Hübler, Biu Huntington-Rainey, Jessica K. Ivani, Marilen Johns, Erika Just, Eri Kashima, Carolina Kipf, Janina V. Klingenberg, Nikita König, Aikaterina Koti, Richard G. A. Kowalik, Olga Krasnoukhova, Nora L.M. Lindvall, Mandy Lorenzen, Hannah Lutzenberger, Tônia R.A. Martins, Celia Mata German, Suzanne van der Meer, Jaime Montoya Samamé, Michael Müller, Saliha Muradoglu, Kelsey Neely, Johanna Nickel, Miina Norvik, Cheryl Akinyi Oluoch, Jesse Peacock, India O.C. Pearey, Naomi Peck, Stephanie Petit, Sören Pieper, Mariana Poblete, Daniel Prestipino, Linda Raabe, Amna Raja, Janis Reimringer, Sydney C. Rey, Julia Rizaew, Eloisa Ruppert, Kim K. Salmon, Jill Sammet, Rhiannon Schembri, Lars Schlabbach, Frederick W.P. Schmidt, Amalia Skilton, Wikaliler Daniel Smith, Hilário de Sousa, Kristin

Sverredal, Daniel Valle, Javier Vera, Judith Voß, Tim Witte, Henry Wu, Stephanie Yam, Jingting Ye 葉婧婷, Maisie Yong, Tessa Yuditha, Roberto Zariquiey, Robert Forkel, Nicholas Evans, Stephen C. Levinson, Martin Haspelmath, Simon J. Greenhill, Quentin D. Atkinson and Russell D. Gray. 2023. Grambank reveals global patterns in the structural diversity of the world's languages. *Science Advances* 9. doi:10.1126/sciadv.adg6175.

Skirgård, Hedvig. 2023. Hedvigs/oceanic_computational_asr: v1.0. doi:10.5281/zenodo.8370386.

Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi, Sam Passmore, Angela Chira, Luke Maurits, Russell Dinnage, Robert Forkel, Simon J. Greenhill and Johannes Englisch. 2023a. Grambank-analysed v1.0. doi:https://doi.org/10.5281/zenodo.7740822. R-scripts associated with the release of Gramabnk data v1.0.

Skirgård, Hedvig, Hannah J. Haynie, Harald Hammarström, Damián E. Blasi, Jeremy Collins, Jay Latarche, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Michael Dunn, Ger Reesink, Ruth Singer, Claire Bowern, Patience Epps, Jane Hill, Outi Vesakoski, Noor Karolin Abbas, Sunny Ananth, Daniel Auer, Nancy A. Bakker, Giulia Barbos, Anina Bolls, Robert D. Borges, Mitchell Browen, Lennart Chevallier, Swintha Danielsen, Sinoël Dohlen, Luise Dorenbusch, Ella Dorn, Marie Duhamel, Farah El Haj Ali, John Elliott, Giada Falcone, Anna-Maria Fehn, Jana Fischer, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima A. Goodall, Victoria Gruner, Andrew Harvey, Rebekah Hayes, Leonard Heer, Roberto E. Herrera Miranda, Nataliia Hübler, Biu H. Huntington-Rainey, Guglielmo Inglese, Jessica K. Ivani, Marilen Johns, Erika Just, Ivan Kapitonov, Eri Kashima, Carolina Kipf, Janina V. Klingenberg, Nikita König, Aikaterina Koti, Richard G. A. Kowalik, Olga Krasnoukhova, Kate Lynn Lindsey, Nora L. M. Lindvall, Mandy Lorenzen, Hannah Lutzenberger, Alexandra Marley, Tânia R. A. Martins, Celia Mata German, Suzanne van der Meer, Jaime Montoya, Michael Müller, Saliha Muradoglu, Hunter-Gatherer, David Nash, Kelsey Neely, Johanna Nickel, Miina Norvik, Bruno Olsson, Cheryl Akinyi Oluoch, David Osgarby, Jesse Peacock, India O.C. Pearey, Naomi Peck, Jana Peter, Stephanie Petit, Sören Pieper, Mariana Poblete, Daniel Prestipino, Linda Raabe, Amna Raja, Janis Reimringer, Sydney C. Rey, Julia Rizaew, Eloisa Ruppert, Kim K. Salmon, Jill Sammet, Rhiannon Schembri, Lars Schlabbach, Frederick W. P. Schmidt, Dineke Schokkin, Jeff Siegel, Amalia Skilton, Hilário de Sousa, Kristin Sverredal, Daniel Valle, Javier Vera, Judith Voß, Daniel Wikalier Smith, Tim Witte, Henry Wu, Stephanie Yam, Jingting Ye 葉婧婷, Maisie Yong, Tessa Yuditha, Roberto Zariquiey, Robert Forkel, Nicholas Evans, Stephen C. Levinson, Martin Haspelmath, Simon J. Greenhill, Quentin D. Atkinson and Russell D. Gray. 2023b. Grambank v1.0. doi:10.5281/zenodo.7740140. Dataset.

Spinu, Vitalie, Garrett Grolemund and Hadley Wickham. 2023. *lubridate: Make dealing with dates a little easier*. https://CRAN.R-project.org/package=lubridate. R package version 1.9.2.

Tierney, Nicholas, Di Cook, Miles McBain and Colin Fay. 2023. *naniar: Data structures, summaries, and visualisations for missing data*. https://github.com/njtierney/naniar. R package version 1.0.0.

Tierney, Nicholas and Dianne Cook. 2023. Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. *Journal of Statistical Software* 105(7). 1–31. doi:10.18637/jss.v105.i07.

Venables, W. N. and B. D. Ripley. 2002. *Modern applied statistics with s*. New York: Springer 4th edition. https://www.stats.ox.ac.uk/pub/MASS4/. ISBN 0-387-95457-0.

Warnes, Gregory R., Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz and Bill Venables. 2022. *gplots: Various r programming tools for plotting data*. https://github.com/talgalili/gplots. R package version 3.1.3.

Wickham, Hadley. 2007. Reshaping data with the reshape package. *Journal of Statistical Software* 21(12). 1–20. http://www.jstatsoft.org/v21/i12/.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley. 2019. *assertthat: Easy pre and post assertions*. https://CRAN.R-project.org/package=assertthat. R package version 0.2.1.

Wickham, Hadley. 2020. *reshape2: Flexibly reshape data: A reboot of the reshape package*. https://github.com/hadley/reshape. R package version 1.4.4.

Wickham, Hadley. 2023a. *forcats: Tools for working with categorical variables (factors)*. https://CRAN.R-project.org/package=forcats. R package version 1.0.0.

Wickham, Hadley. 2023b. *stringr: Simple, consistent wrappers for common string operations*. https://CRAN.R-project.org/package=stringr. R package version 1.5.1.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani and Dewey Dunnington. 2023a. *ggplot2: Create elegant data visualisations using the grammar of graphics*. https://CRAN.R-project.org/package=ggplot2. R package version 3.4.4.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller and Davis Vaughan. 2023b. *dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr. R package version 1.1.4.

Wickham, Hadley and Lionel Henry. 2023. *purrr: Functional programming tools*. https://CRAN.R-project.org/package=purrr. R package version 1.0.2.

Wickham, Hadley, Jim Hester and Jennifer Bryan. 2023c. *readr: Read rectangular text data*. https://CRAN.R-project.org/package=readr. R package version 2.1.4.

Wickham, Hadley, Davis Vaughan and Maximilian Girlich. 2023d. *tidyr: Tidy messy data*. https://CRAN.R-project.org/package=tidyr. R package version 1.3.0.

Wilks, S. S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* 9(1). 60–62. doi:10.1214/aoms/1177732360.

Xie, Yihui. 2014. knitr: A comprehensive tool for reproducible research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng (eds.), *Implementing reproducible computational research*, Chapman and Hall/CRC. ISBN 978-1466561595.

Xie, Yihui. 2015. *Dynamic documents with R and knitr*. Boca Raton, Florida: Chapman and Hall/CRC 2nd edition. https://yihui.org/knitr/. ISBN 978-1498716963.

Xie, Yihui. 2023. *knitr: A general-purpose package for dynamic report generation in r*. https://yihui.org/knitr/. R package version 1.43.

Yang, Ziheng. 2006. *Computational molecular evolution*. Oxford University PressOxford. doi:10.1093/acprof:oso/9780198567028.001.0001.