Appendix

8.1 Number of missing values per language

Language	number of missing values
Abidji	41
Afrikaans	39
AkanTwi	50
Albanian	38
Amharic	25
ArabicGulf	39
ArabicMoroccan	43
ArmenianEastern	1
ArmenianWestern	41
Bafut	10
Bambara	19
Bandial	44
Bangla	45
Basaá	0

Language	number of missing values
CantoneseGuangzhou	0
Catalan	0
Chickasaw	20
ChineseWuhu	0
Croatian	0
DagaraBurkina	0
Danish	10
Dholuo	47
DidaLakotadiéko	37
Dutch	0
English	1
EnglishSingapore	39
Faroese	0
Finnish	0
FlemishWest	42
French	3

Language	number of missing values
Ga	43
Galician	39
Garifuna	32
Georgian	39
German	0
Gothic	28
Greek	0
GreekCalabria	53
GreekCappadocian	0
GreekCypriot	0
Guébie	26
GungbePorto-Novo	4
Haitian	9
Hausa	8
Hebrew	15
Hoan	37

Language	number of missing values
Hungarian	20
Ibibio	39
Icelandic	39
Igbo	15
Indonesian	48
Italian	0
ItalyNapoletanoAntico	21
ItalyNorthernCalabrese	0
ItalyReggiano	0
ItalySiciliano	0
ItalyTeramano	0
Japanese	0
Kiche	39
KiLega	22
Kiyaka	31
Kom	38

Language	number of missing values
Korean	0
Kuot	39
KurdishSorani	39
Kusaal	34
Kusunda	45
Laal	43
Mada	37
Malagasy	8
Mandarin	20
Marshallese	45
mɔ́dʒúkrù	40
MixtecSanSebastiándelMonte	68
Miya	44
Mundurukú	49
Muyang	0
NahuatlCentralHuasteca	44

Language	number of missing values
Nepali	43
Norwegian	0
Nupe	20
Nweh	21
NzimaTiapoum	22
OluwangaEji	6
Pashto	44
Polish	48
PomeranianBrazilian	3
Portuguese	1
PortugueseBrazilian	25
Qanjobal	49
QuichuaImbaburaQuichua	20
Romanian	0
Russian	0
Samoan	32

Language	number of missing values
Serbian	0
Shupamem	6
Skou	48
Spanish	0
Swedish	18
Tagalog	46
TaiwaneseSouthernMin	39
Thai	44
Tommo-So	49
TukangBesi	43
TukiTukombo	40
TupíTupinambá	0
Turkish	1
Twi	39
Vata	18
Vietnamese	1

Language	number of missing values
Wolof	23
Xhosa	43
Yémba	44

8.2 SSWL properties included in our dataset

A.01_Attr.Adj.Agr
A.02_Pred.Adj.Agr
A.03_Deg.A
A.04_A.Deg
C.01_C.Clause
C.02_Clause.C
C.03_Adv.Sub.Clause
C.04_Clause.Adv.Sub
N2.01_Num.N_.indef.
N2.02_N.Num_.indef.
N2.03_Num.N_.def.
N2.04_N.Num_.def.
N2.F1_.Def.NumN.N.Num.has.a.Def.Art.
N2.F2_Def.NumN.or.NNum.can.have.a.DEM.
N2.F3_Def.NumN.or.NNum.can.be.bare
N2.F4a_Def.NumN.or.NNum.marks.Pl.but.Indef.Num.N.lacks.Pl
N2.F5_Def.NumN.or.NNum.differs.from.Indef.NumN
N2.F6_Num.has.different.form.in.Def.and.Indef.Num.N.N.Num.contexts
N2.F7_.N.has.different.form.in.Def.Indef.Num.N.N.Num.contexts
N3.01_.Dem.A.N
N3.02_Dem.N.A
N3.03_.N.A.Dem
N3.04_N.Dem.A
N3.05_A.Dem.N
N3.06_A.N.Dem
N3.07_Dem.Num.N
N3.08_Dem.N.Num
N3.09_.N.Num.Dem
N3.10_N.Dem.Num
N3.11_Num.Dem.N
N3.12_Num.N.Dem
Neg.01_.Neg.is.Part.Verb

Neg.02_Neg.is.V.Part

Neg.03_Dist.Neg.of.Class.Property

Neg.04_Dist.Neg.of.identity

Neg.05_Dist.Neg.of.Location

Neg.06_Dist.Neg.of.Existence

Neg.F1_Neg.is.NegAux

Neg.F2_.Neg.is.lexical.V

Neg.F3_Neg.is.Tonal

Neg.F4_Neg.is.Tone.and.other.modification

Neg.F5_Neg.is.Reduplication

Neg.F6_Neg.is.Infix

O.01.1_Indef.Mass_Can.be.bare

O.01.2_Indef_Mass_Can.have.an.article

O.01.3_Indef_Mass_Must.have.an.article

O.02.1_Def_Mass._Can.be.bare

O.02.2_Def_Mass._Can.have.an.article

O.02.3_Def_Mass._Must.have.an.article

O.02.4.1_DefMass_Art.N

O.02.4.2_DefMass_N.Art

O.03.1_Indef_Sg_Can.be.bare

O.03.2_Indef_Sg_Can.have.an.article

O.03.3_Indef_Sg_Must.have.an.article

O.03.4_Indef_Sg_Can.have.a.plural.reading

O.04.1_Def_Sg_Can.be.bare

O.04.2_Def_Sg_Can.have.an.article

O.04.3_Def_Sg_Must.have.an.article

O.04.4.1_DefSg_Art.N

O.04.4.2_DefSg_N.Art

O.05.1_Indef_Pl_Can.be.bare

O.05.2_Indef_Pl_Can.have.an.article

O.05.3_Indef_Pl_Must.have.an.article

O.06.1_Def_Pl_Can.be.bare

O.06.2_Def_Pl_Can.have.an.article

O.06.3_Def_Pl_Must.have.an.article

O.06.4.1_DefPl_Art.N

O.06.4.2_DefPl_N.Art

O.07.1_Unique.Ref_Can.be.bare

O.07.2_Unique.Ref_Can.have.an.article

O.07.3_Unique.Ref_Must.have.an.article

O.07.4_Unique.Ref_Must.have.a.special.article.

O.08.1_PN_Can.be.bare

O.08.2_PN_Can.have.an.article

O.08.3_PN_Must.have.an.article

O.08.4_PN_Must.have.a.special.article.

O.08.5.1_Art.PN

O.08.5.1_PN.Art

O.09.1_PN.A_Can.be.bare

O.09.2_PN.A_Can.have.an.article

O.09.3_PN.A_Must.have.an.article

O.09.5.2_.PN.Adj

Q01_Initial.polar.Q.marker

Q02_Final.polar.Q.marker

Q03_Clause.internal.polar.Q.marker
Q04_Polar.question.by.intonation
Q05_Polar.question.by.V.fronting
Q06_Polar.question.by.Aux.fronting
Q07_Q.marker.follows.narrow.focus
Q08_Affirmative.answer.by.particle
Q09_Affirmative.answer.by.V
Q10_Affirmative.answer.by.aux
Q11_Affirmative.answer.by.auxiliary.and.verb
Q12_Affirmative.answer.by.particle.and.verb
Q13_Affirmative.answer.by.predicative.be.
Q14_Verb.answer.to.indefinite.subject.question
Q15_Particle.and.verb.answer.to.indefinite.subject.question
Q16_NEGA_Negative.answer.by.special.neg.particle
Q17_NEGA_Neg.answer.by.bare.sentential.negation
Q18_NEGA_Neg.answer.by.sentential.neg.plus.bare.verb
Q19_ANegQ_Truth.based.confirmation
Q20_ANegQ_Polarity.based.confirmation
Q21_ANegQ_Polarity.reversing.particle
Q22_ANegQ_Polarity.reversing.answer.by.affirmative.and.special.particle
V_Aux.01_Aux.Selection
V2.01_Declarative.Verb.Second
V2.02_Interrogative.Verb.Second
X01_SV
X02_VS
X03_VO
X04_OV
X05_SVO
X06_SOV
X07_VSO
X08_VOS
X09_OSV
X10_OVS
X11_P.NP
X12_NP.P
X13_A.N
X14_N.A
X15_Num.N
X16_N.Num
X17_Dem.N
X18_N.Dem
X19_Poss.N
X20_N.Poss
X21_Pron.Poss.N
X22_N.Pron.Poss

8.3 The influence of missing values

The languages in the dataset exhibit differing numbers of missing values, with some languages having no missing values; the highest number of missing values (54 in total) is found in Calabrian Greek. These missing values, although imputed in the dataset, could potentially diminish the posterior support of individual nodes. As a result, we need to determine whether the low clade support values found throughout the tree are due to this missingness. To gauge the influence of missing values on the topology of the consensus tree, we used the maximum clade credibility tree to extract the per-node support along with the missing values of the tip nodes descended from each node (except the root). As a result, we have a dataset with all ancestor nodes and their support aligned with the missing values in the daughter nodes. Afterwards, we ran a Bayesian beta-regression model using the R-package *brms* (Bürkner 2017) with per-node support values as the outcome and the number of missing values of each tip node as the predictor. The priors were chosen to be $Intercept \sim Normal(0, 1.5)$, and $Missing \sim Normal(0, 1)$. Table 3 shows the posterior effects of this model, along with their 89% HDI compatibility intervals.

Effect	Estimate	Est.Error	Lower-89CI	Upper-89CI
Intercept	-0.958	0.061	-1.057	-0.862
Missing	-0.007	0.001	-0.010	-0.003

Table 3: Posterior coefficients of the beta-regression model

The models shows that there is a reduction in clade support with increasing missingness. This effect, however, is fairly small: when we calculate the expected reduction of missingness for the tip node with the most missing values (54), the posterior expectation is a reduction of 0.06 with a 89-CI of [0.001, 0.08]. This means that even for clades with the maximum number of missing values, we would expect only a decrease in clade support of 0.06, which is too small to distort the tree in general to a strong degree.

8.4 Model details

8.4.1 Graphical model summary

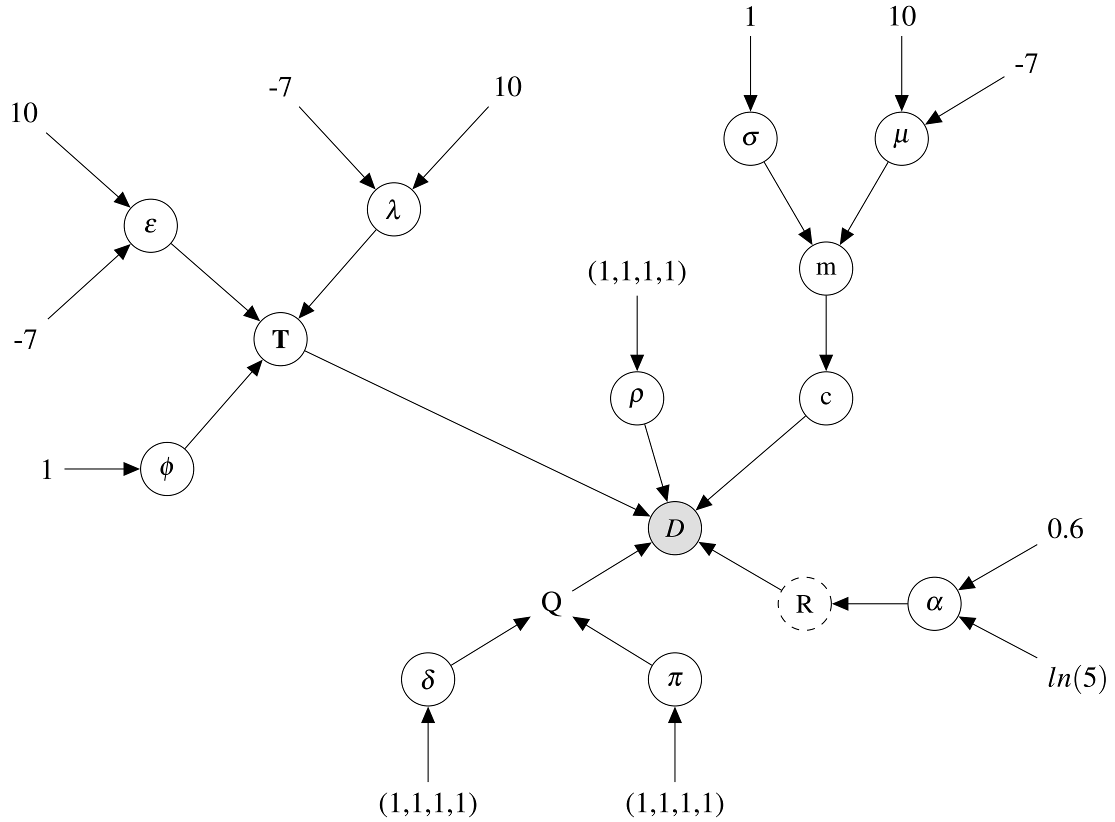

Figure 2: Graphical representation of the phylogenetic model

8.4.2 Model priors

	Node	Prior	Specification
Tree model	ϕ	1	root age
	ε	LogNormal(-7, 10)	extinction rate
	λ	LogNormal(-7, 10)	speciation rate
Substitution model	δ	Dirichlet(1, 1, 1, 1)	substitution rate
Root frequencies	ρ	Dirichlet(1, 1, 1, 1)	
Clock model	c_{branch}	Exponential($\frac{1}{m}$)	branch rates
	m	LogNormal(μ, σ)	branch rates prior
	μ	Normal(-7, 10)	
	σ	Exponential(1)	
Site rate model	R	DiscretizedGamma(α, α, 4)	site rates
	α	LogNormal($ln(5), 0.6$)	site rates prior

Table 4: Prior summary

8.4.3 Posterior draws of model parameters

parameter	mean	lower.89CI	higher.89CI
alpha	2.64	1.99	3.35
extinction rate	0.02	0.00	0.02
speciation rate	3.64	2.74	4.49
root frequencies 1	0.36	0.27	0.45
root frequencies 2	0.26	0.13	0.38
root frequencies 3	0.37	0.16	0.51

Table 5: Posterior estimates of select model parameters

The alpha parameter around 2.5 implies that there was notable site heterogeneity detectable. This means that the change rates of the characters in the dataset are very different for some characters than for others. It should be kept in mind, however, that extinction and speciation rates, as well as the root frequencies, are less relevant for the goal of this investigation. In other words, these parameters give insights into the model from a purely computational and inferential viewpoint but are not relevant to the question of the strength of the phylogenetic signal in this paper.

8.5 Full tree cladograms

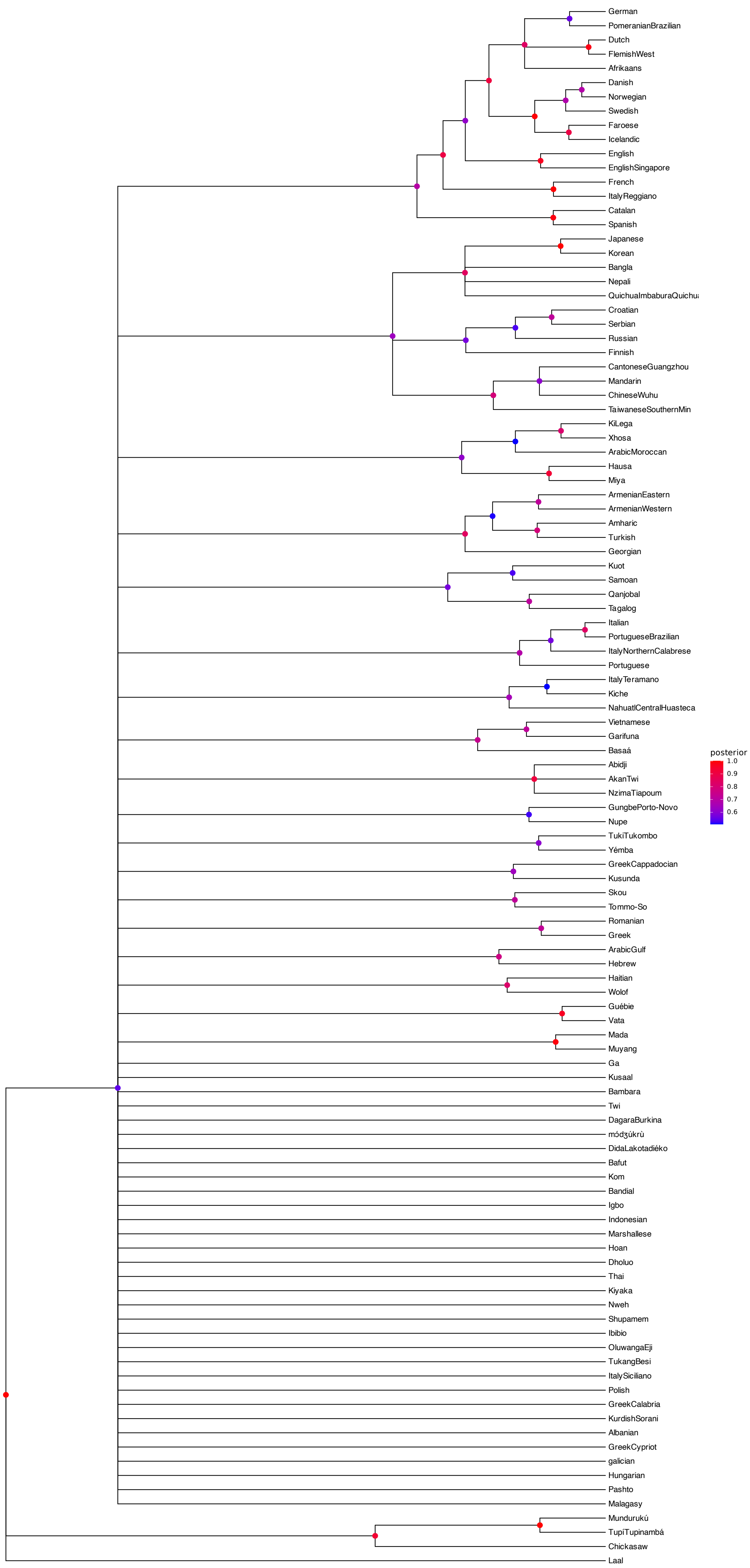

Figure 3: Posterior consensus tree of all languages in the dataset.

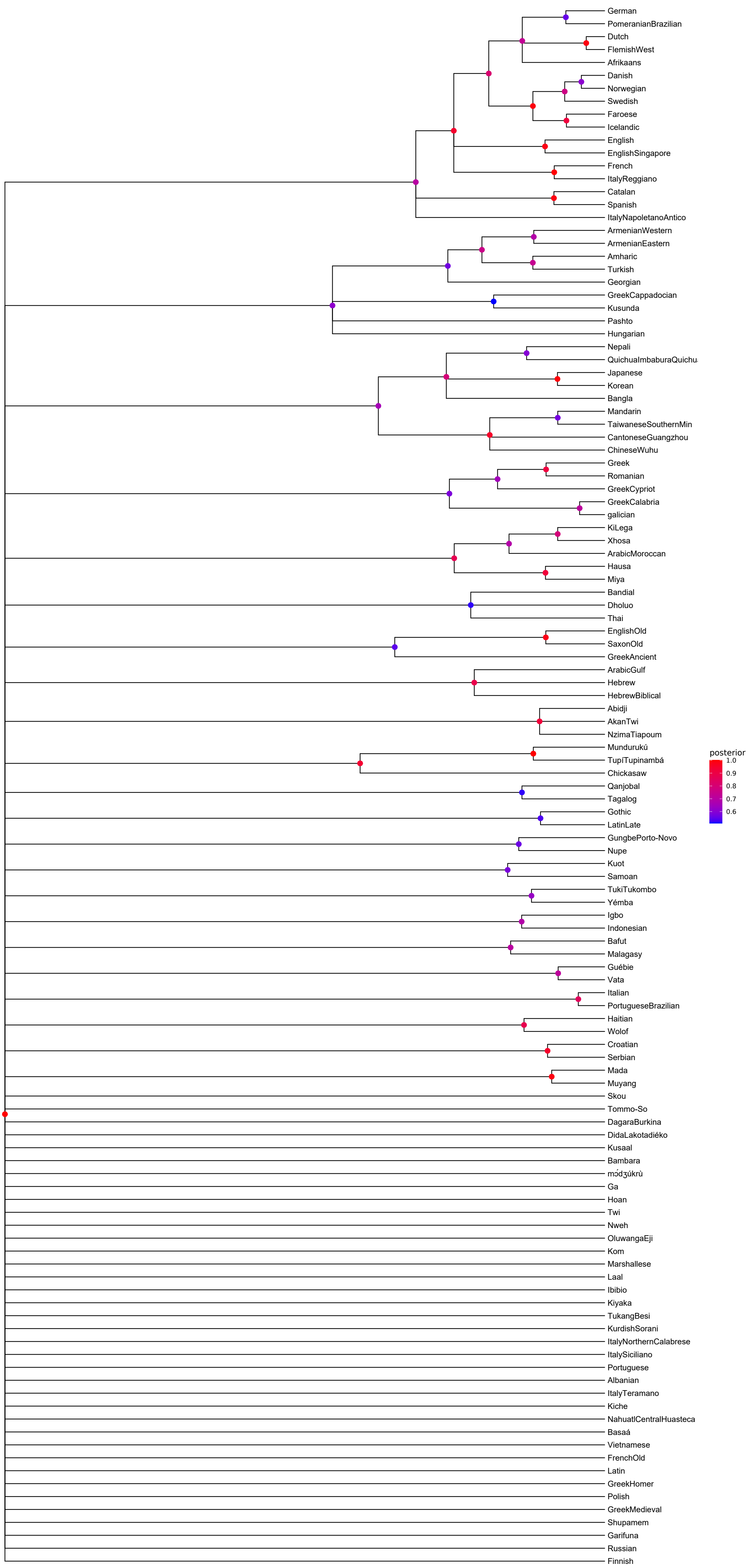

Figure 4: Posterior consensus tree of all languages in the dataset with the two highly correlated features removed.

8.6 UPGMA and Neighbor Joining results

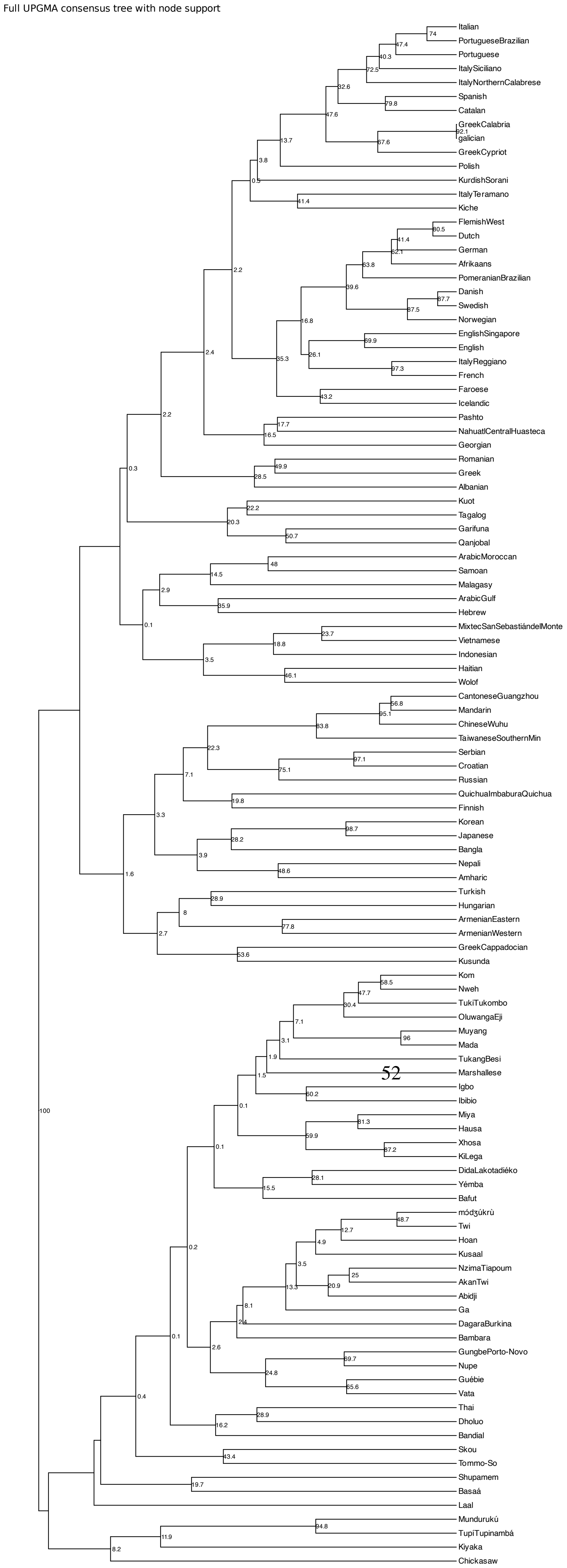

Figure 5: Full UPGMA tree. Node labels indicate clade support, from 0 (little support) to 100 (high support), according to the bootstrapping analysis.

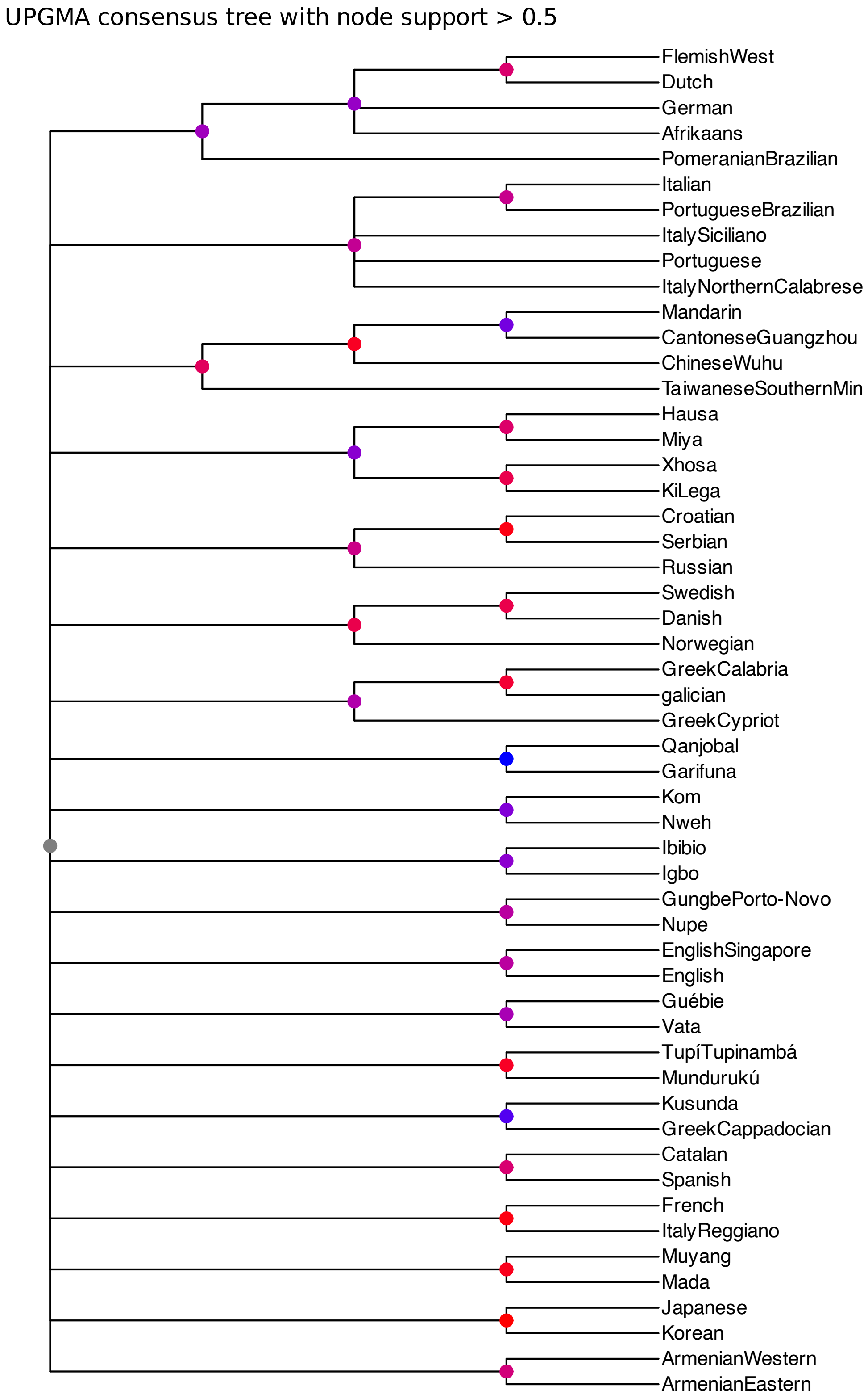

Figure 6: Truncated UPGMA consensus tree obtained from the bootstrap analysis. Node colors indicate clade support, from blue (little support) to red (high support), according to the bootstrapping analysis.

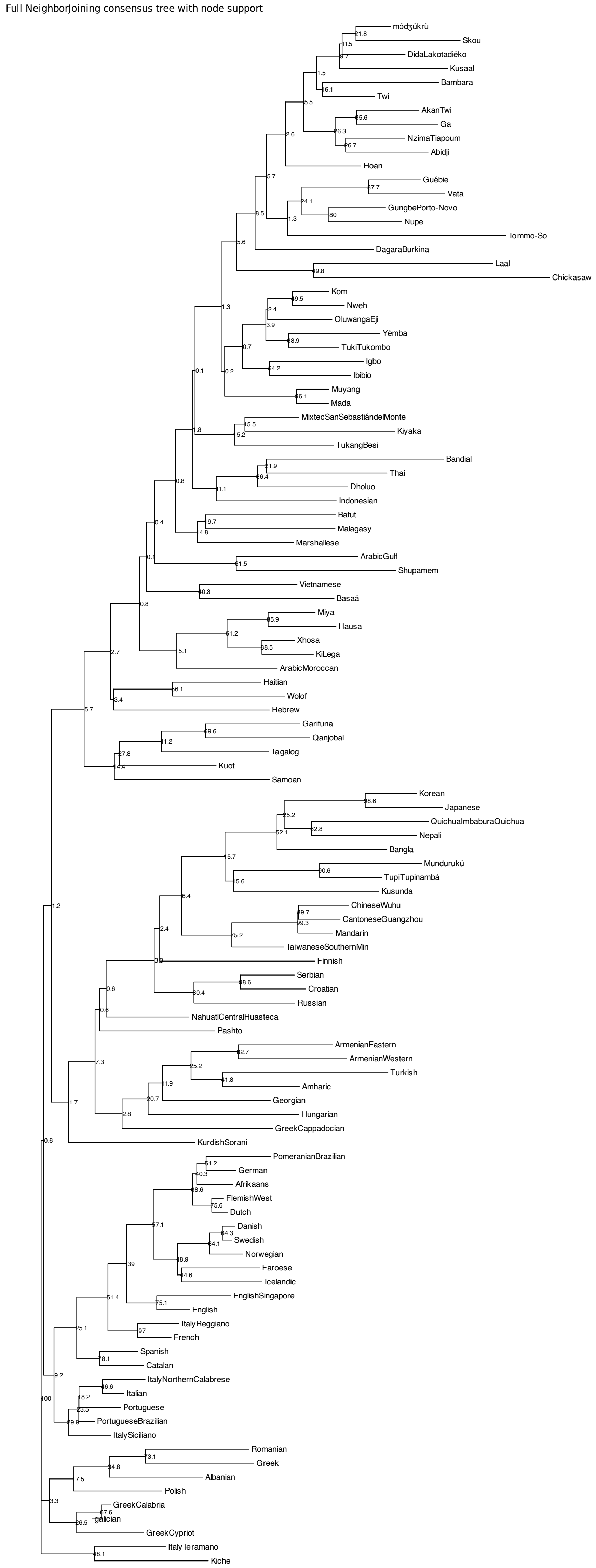

Figure 7: Full Neighbor Joining tree. Node labels indicate clade support, from 0 (little support) to 100 (high support), according to the bootstrapping analysis.

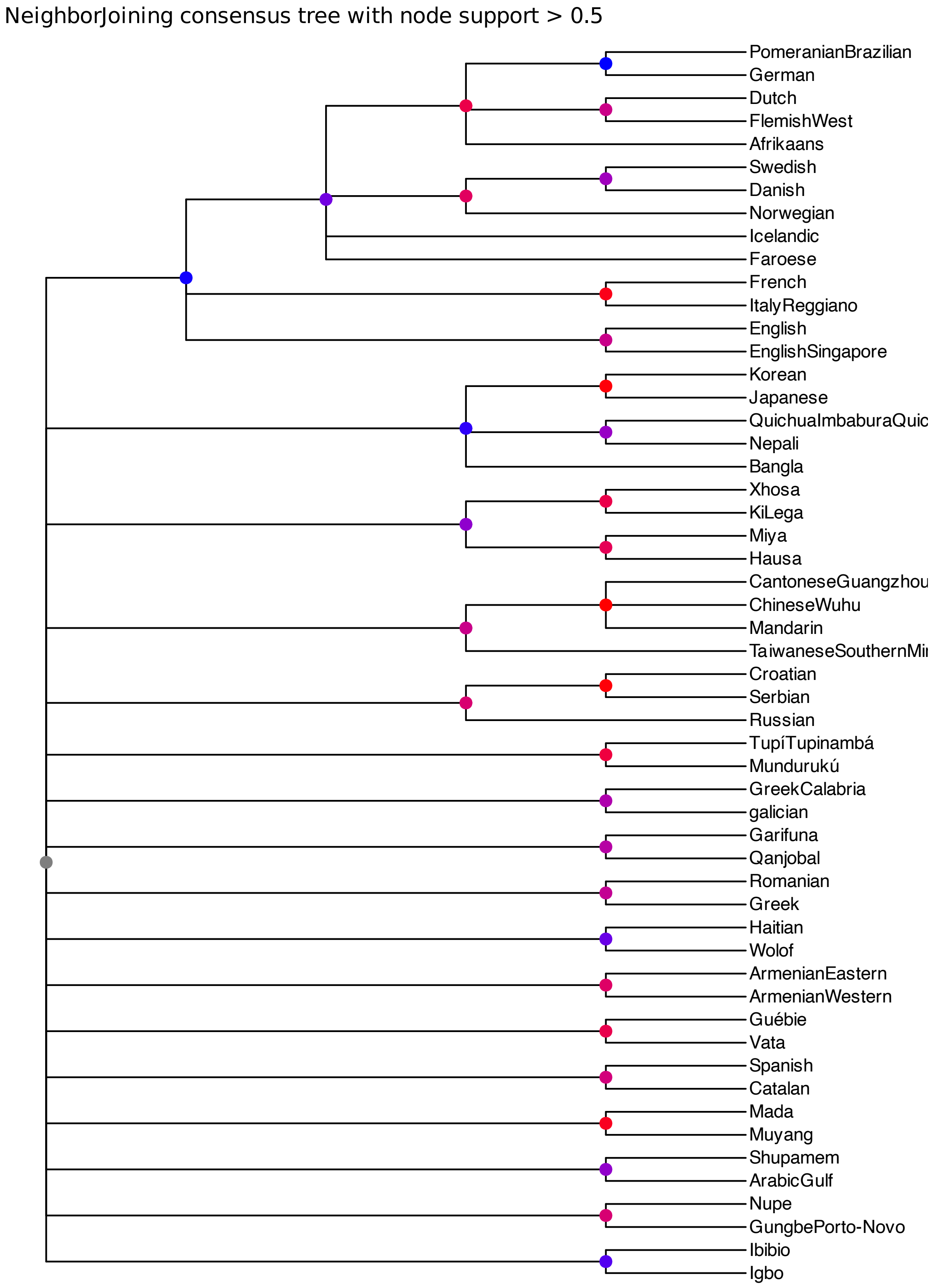

Figure 8: Truncated Neighbor Joining consensus tree obtained from the bootstrap analysis. Node colors indicate clade support, from blue (little support) to red (high support), according to the bootstrapping analysis.

8.7 Reduced model results

To compare our results with a reduced model with much stricter assumptions and fewer parameters, we ran a model on the same data (only extant languages, no feature dependencies) with the following properties: (1) Uniform time tree (i.e, no birth death) with a (2) global molecular clock (clock rate distributed as $Gamma(2.0, 4.0)$) and (3) a Jukes-Cantor substitution model. Figure 9 shows the posterior consensus tree of the results.

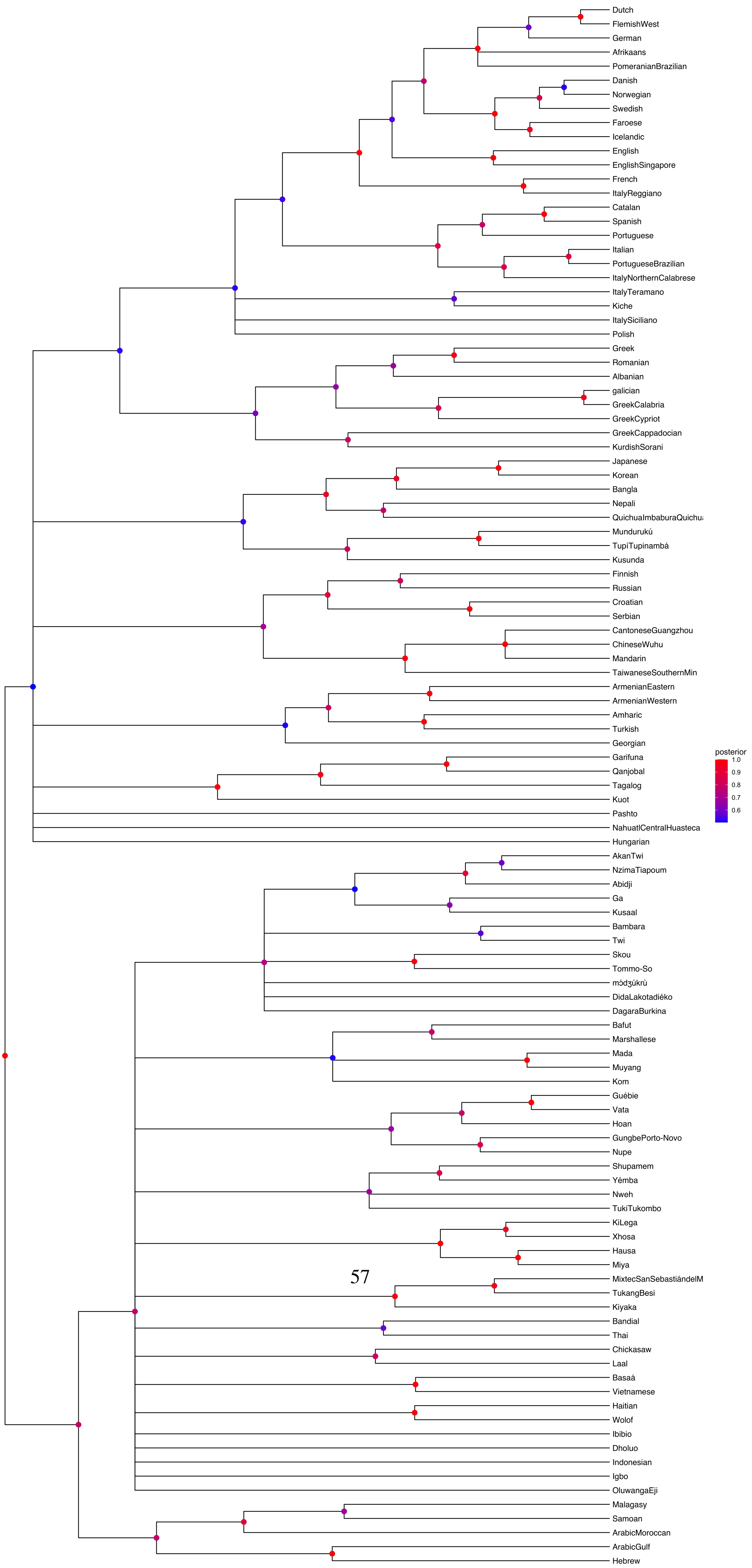

Figure 9: Posterior consensus tree of the reduced model.

Here, we can see that it shows similar patterns as the main model, not being able to recover the main families and being affected by potential areal effects. It has more clade support overall and more uniformly distributed splits (akin to the UPGMA and NJ models) which is likely the result of the more constrained parametrization of the model. However, this also means that a model with more flexibility is less likely to assume these groupings to be robust.