Universität Zürich[UZH]

Institute for the Interdisciplinary Study of Language Evolution

nccr evolving language
NATIONAL CENTRE OF COMPETENCE IN RESEARCH

Evolutionary dynamics of maximal syllable complexity

Laura Dees[1], Shelece Easterday[2], Paul Widmer[1], Balthasar Bickel[1], Chundra Cathcart[1]

1) Center for the Interdisciplinary Study of Language Evolution, University of Zurich; 2) Department of Linguistics, University of Hawai'i

New Advances in Phylolinguistics - Leipzig, June 2025

A global CV bias?

Cross-linguistic diversity in maximal syllable complexity

onset - vowel - coda

CV > …… > CCCCVCCCC

But: strong claims about a universal CV bias (Gordon 2016, Blevins 2006, Jakobson 1962)

A global CV bias?

Cross-linguistic diversity in maximal syllable complexity

But: strong claims about a universal CV bias (Gordon 2016, Blevins 2006, Jakobson 1962)

Link to language processing (Content et al. 2001, Sun & Poeppel 2023, Wedel et al. 2019)

Rooted in biomechanical constraints (MacNeilage 1998, cf. Pereira et al. 2020)

A global CV bias?

Unresolved question: evolutionary dynamics

→ global profile or family-specific dynamics?

Data & methods

Maximal onset & coda size in 5 families, 401 languages

Language	Max_onset	Max_coda
Pipil	1	1
Serrano	2	3
Tohono O'odham	4	4
Highland Puebla Nahuatl	2	1
Hopi	1	2

Phylogenetic modeling using a CTM process, inferring transition rates

All possible transitions allowed

Methods

(Jäger & Wahle 2021)

Model comparison

global: 1 CTM across all families

family-specific: 5 different CTMs

Sample of 25 trees for each family

Comparison using Bayes Factors

Results model comparison

Lineage-specific vs global

onsets: BF ≈ 8.55×10^8

codas: BF = 23708

Lineage-specific model outperforms global model

→ Idea of global trends in diachronic change is not supported

Results model comparison

Lineage-specific vs global

onsets: BF ≈ 8.55×10^8

codas: BF = 23708

Lineage-specific model outperforms global model

➞ Idea of global trends in diachronic change is not supported

What do family-level trends look like?

➞ stationary probabilities


Stationary probability - onset
Austronesian
Indo-European
Pama-Nyungan
Sino-Tibetan
Uto-Aztecan
Probability
1.00
0.75
0.50
0.25
0.00
CI
0.95
0.8
0.5


Stationary probability - coda
Austronesian
Indo-European
Pama-Nyungan
Sino-Tibetan
Uto-Aztecan
Probability
1.00
0.75
0.50
0.25
0.00
CI
0.95
0.8
0.5

Family-level trends

Long-term preferences

→ no cross-family bias towards any particular state in onset or coda

→ especially onsets exhibit strong between-family variation

Family-level trends

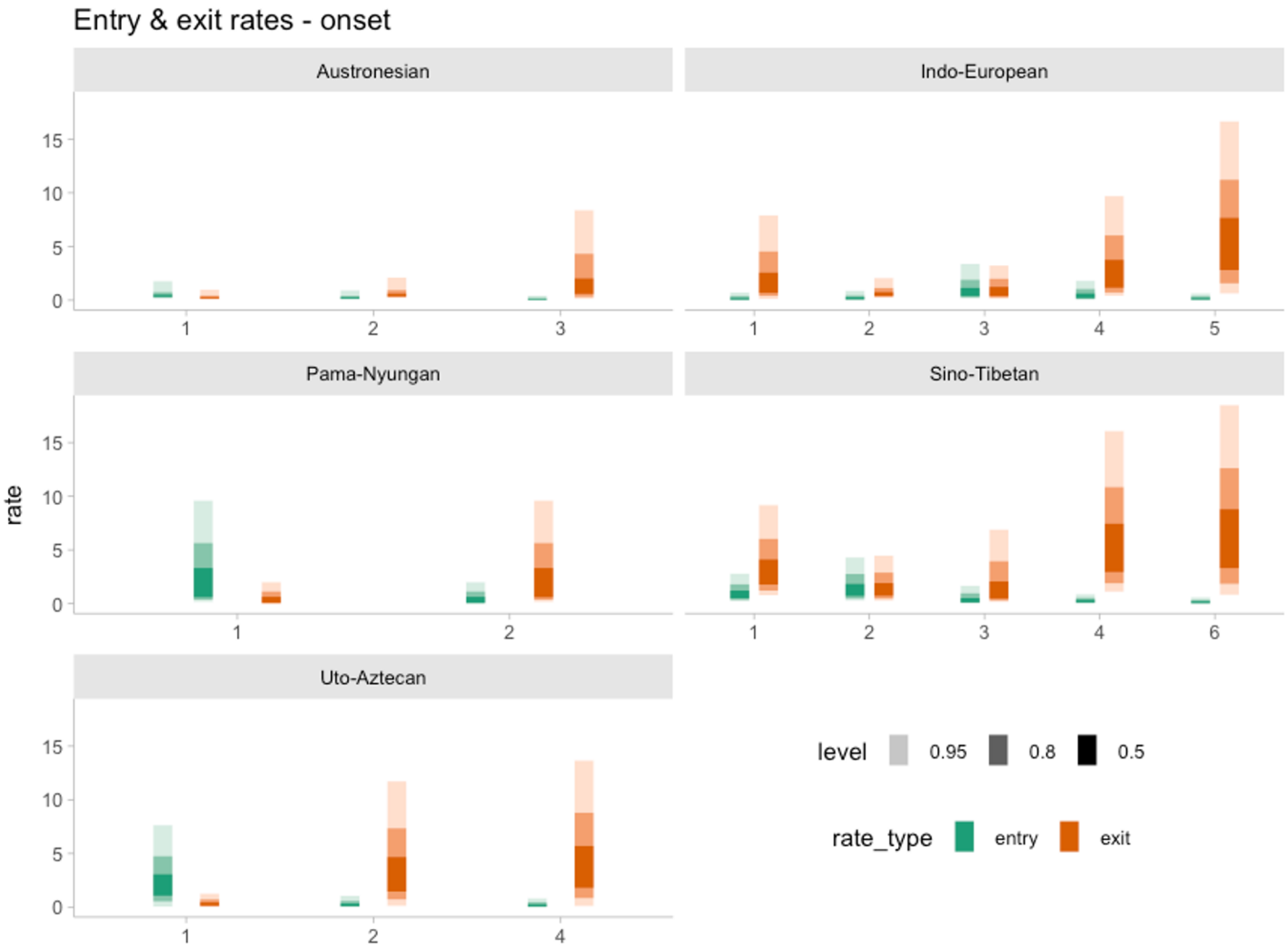
Long-term preferences

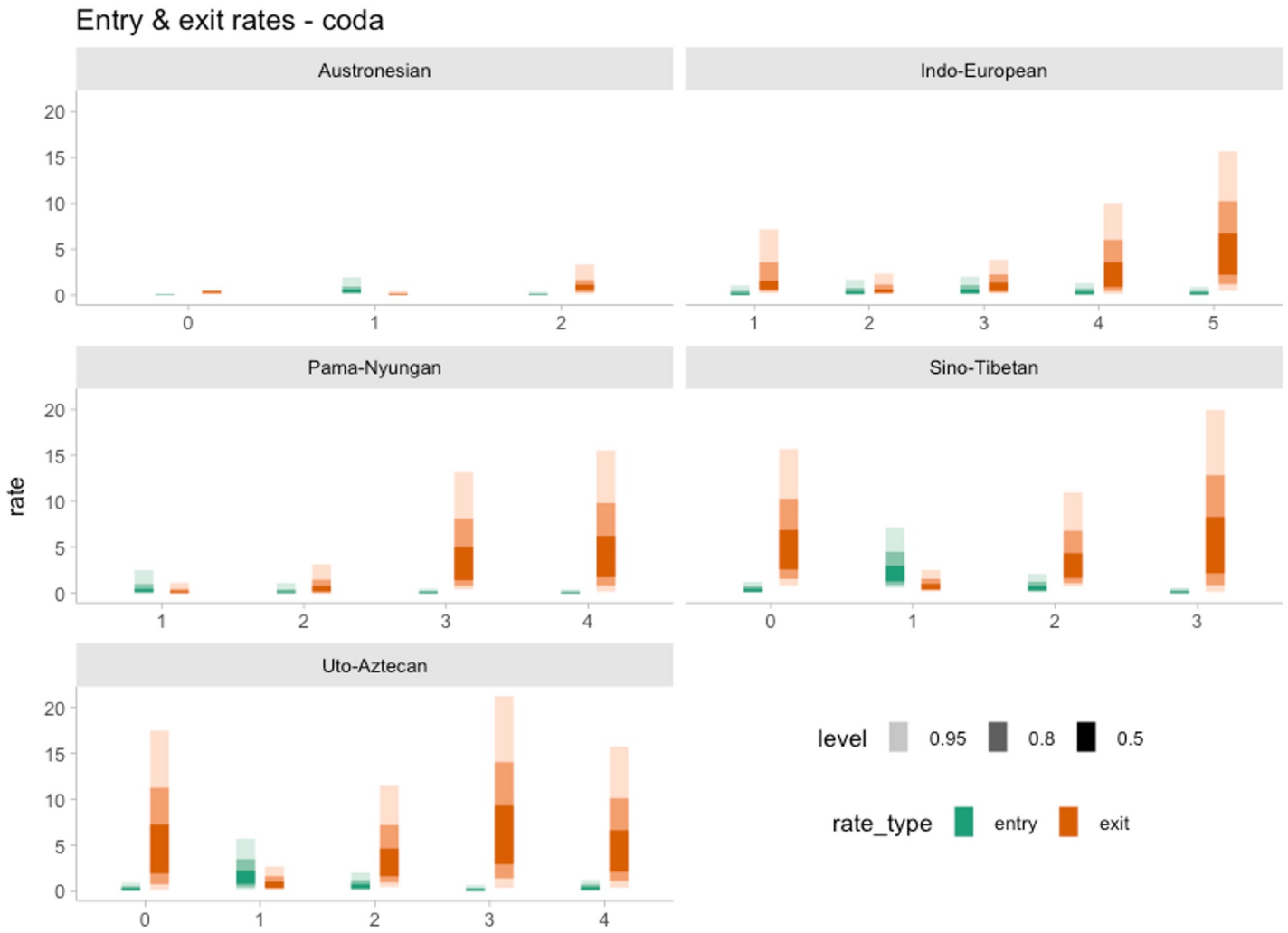
→ no cross-family bias towards any particular state in onset or coda

→ especially onsets exhibit strong between-family variation

Do pathways to complexity levels vary?

→ entry and exit rates


Entry & exit rates - onset
Austronesian
Indo-European
Pama-Nyungan
Sino-Tibetan
Uto-Aztecan
rate
level
0.95
0.8
0.5
rate_type
entry
exit


Entry & exit rates - coda
Austronesian
Indo-European
Pama-Nyungan
Sino-Tibetan
Uto-Aztecan
rate
level
0.95
0.8
0.5
rate_type
entry
exit

Family-level trends

Entry and exit rates

→ most variation in onset entry rates between states *within* families

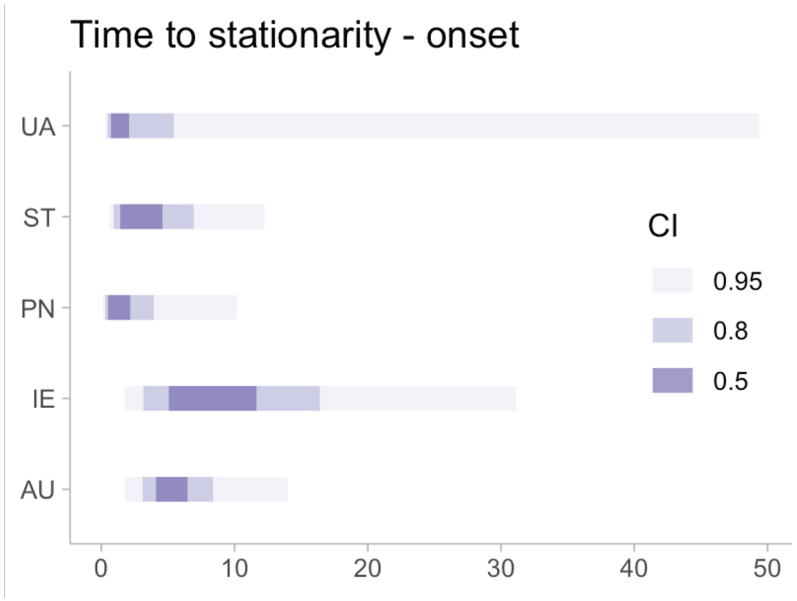
→ moderate degrees of between-family variation

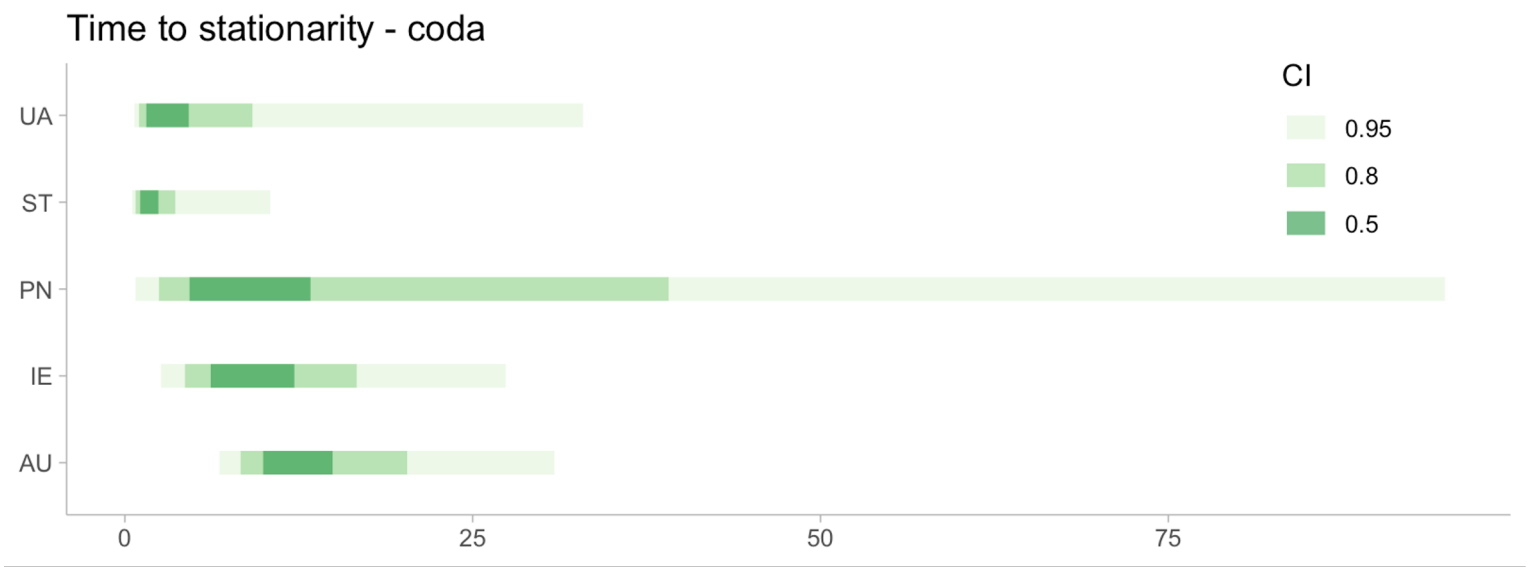
→ does not seem to heavily mediate long-term preferences

Family-level trends

How fast do systems change?

→ time to stationarity


Time to stationarity - onset
UA
ST
PN
IE
AU
0
10
20
30
40
50
CI
0.95
0.8
0.5


Time to stationarity - coda
CI
0.95
0.8
0.5
UA
ST
PN
IE
AU
0
25
50
75

Family-level trends

Speed of change

→ little substantial variation; overall slow change

Discussion

Maximal syllable complexity evolves in an overwhelmingly family-specific fashion

- stationary probabilities
- entry/exit rates
- time to stationarity

Independently from universal biomechanical constraints

References

Blevins J. 2006 Syllable typology. In Brown K, editor, *Elsevier encyclopedia of language and linguistics*, pp. 333–337. Oxford: Elsevier.

Content A, Kearns R, Frauenfelder U. 2001 Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language* 45, 177–199.

Easterday, S. 2019. *Highly complex syllable structure: a typological and diachronic study*. Language Science Press.

Fenk-Oczlon, G. and Pilz, J. 2021. Linguistic complexity: Relationships between phoneme inventory size, syllable complexity, word and clause length, and population size. *Frontiers in Communication*, 6.

Gordon MK. 2016. *Phonological typology*. Oxford University Press.

Jäger G, Wahle J. 2021 Phylogenetic typology. *Frontiers in Psychology* 12.

Jakobson R. 1962 *Selected writings 1: Phonological studies*. the Hague: Mouton.

MacNeilage P. 1998 The frame/content theory of evolution of speech production. *Behavioral and Brain Science*s 21, 499–511.

Maddieson, I. 2005. Issues of phonological complexity: Statistical analysis of the relationship between syllable structures, segment inventories and tone contrasts. *UC Berkeley PhonLab Annual Report*, 1:259–268.

Pereira A, Kavanagh E, Hobaiter C, Slocombe K, Lameira A. 2020 Chimpanzee lip-smacks confirm primate continuity for speech-rhythm evolution. *Biology Letters* 16.

Sun Y, Poeppel D. 2023 Syllables and their beginnings have a special role in the mental lexicon. *Proceedings of the National Academy of Sciences* 120.

Wedel A, Ussishkin A, King A. 2019 Incremental word processing influences the evolution of phonotactic patterns. *Folia Linguistica* 53, 231–248.

Thanks to:

Shelece Easterday

Paul Widmer

Balthasar Bickel

Chundra Cathcart

Covariates of maximal syllable complexity

Correlation with consonant inventory size observed (e.g. Maddieson 2005, Gordon 2016, Easterday 2019, Fenk-Oczlon & Pilz 2021)

→ Control for relatedness

Data

Syllable complexity

- SE
- WALS
- ASJP

Consonant inventories

- PHOIBLE

Phylogenies

- 5 focal families
- World tree

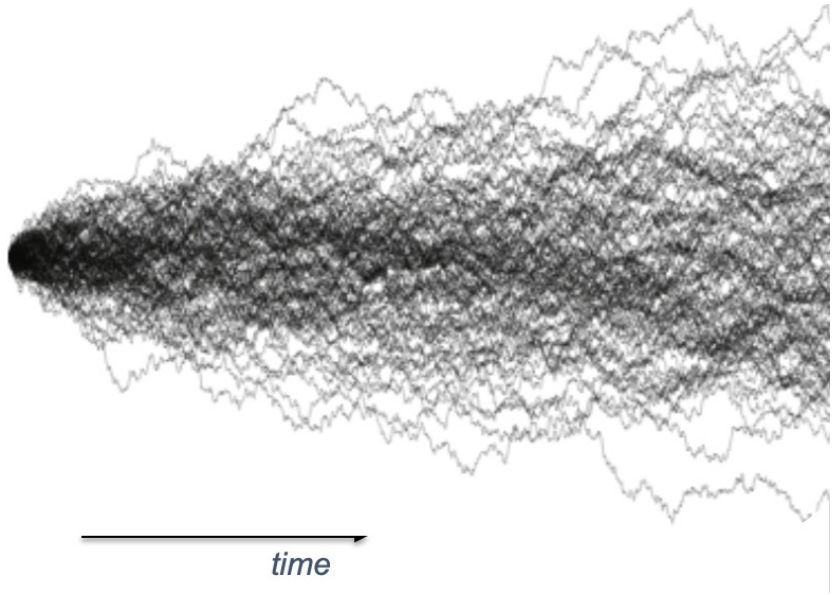
Language	Max_onset	Max_coda
Pipil	1	1
Serrano	2	3
Tohono O'odham	4	4
Highland Puebla Nahuatl	2	1
Hopi	1	2

Abipón	Complex
Abkhaz	Complex
Aché	Simple
Acoma	Moderately complex
Aghem	Moderately complex
Ahtna	Complex

I	I	any~i7i
I	I	a5i7i
you	THOU	api7i
we	WE	aCim
one	ONE	h3mako
two	TWO	gok
person	PERSON	o7odh~am
fish	FISH	watopi
dog	DOG	gogs
louse	LOUSE	a7aC
tree	TREE	us
leaf	LEAF	hahag

Modeling phylogenetic correlation

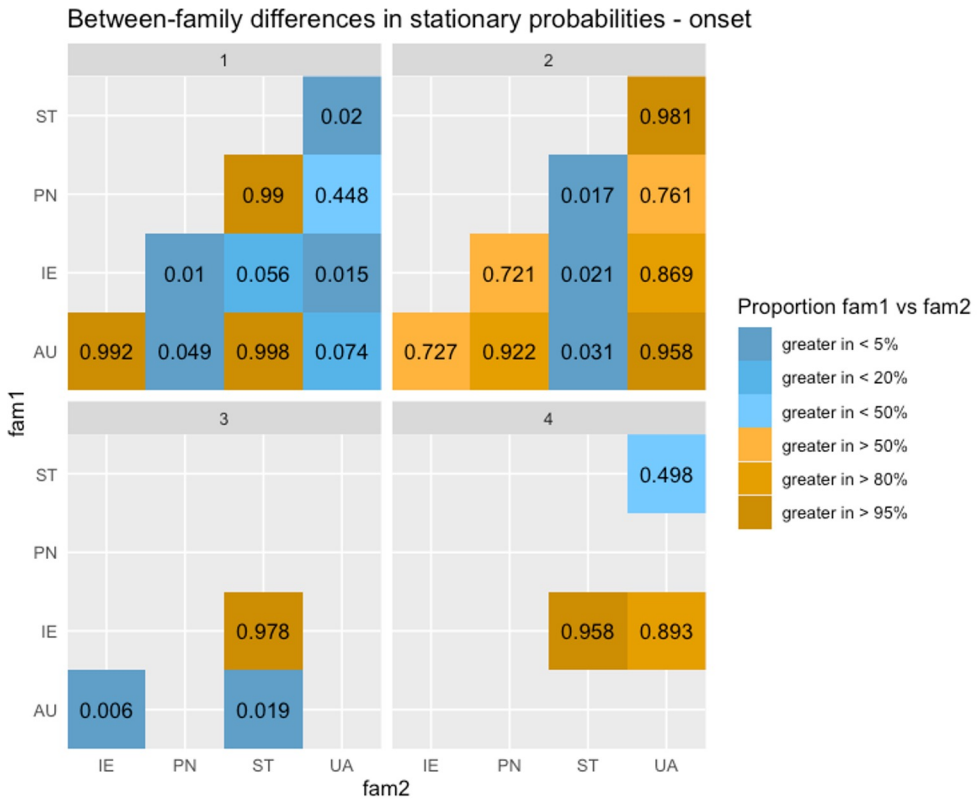
Brownian motion

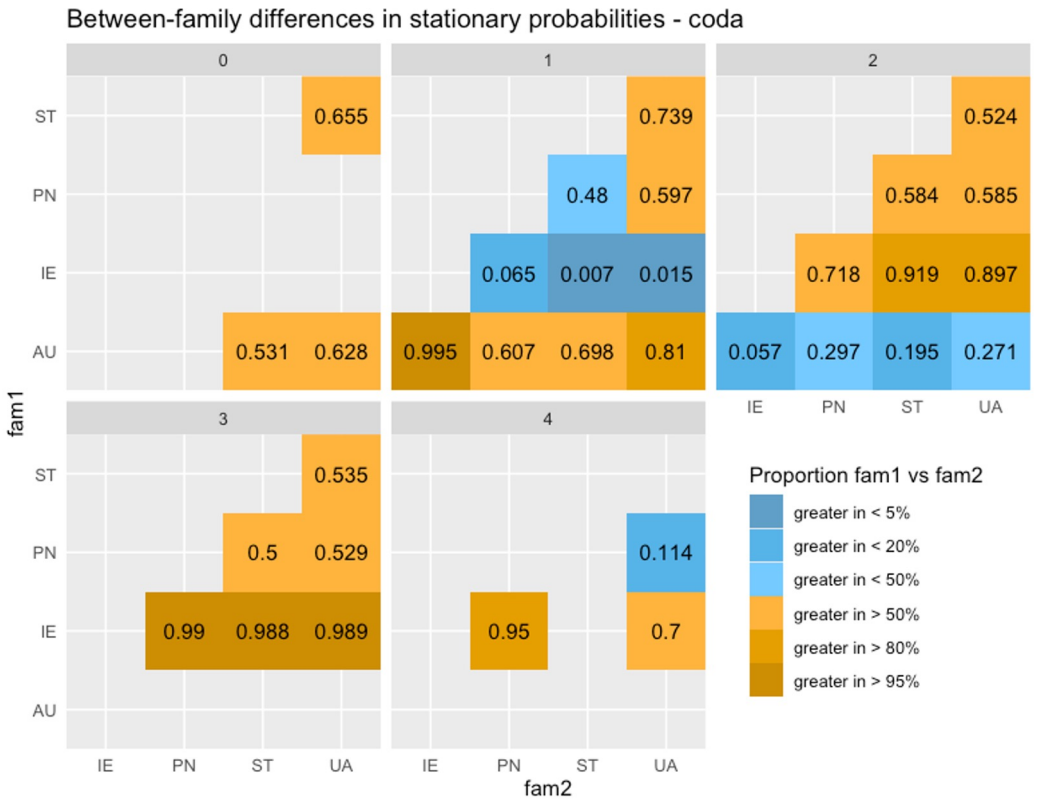


Modeling phylogenetic correlation

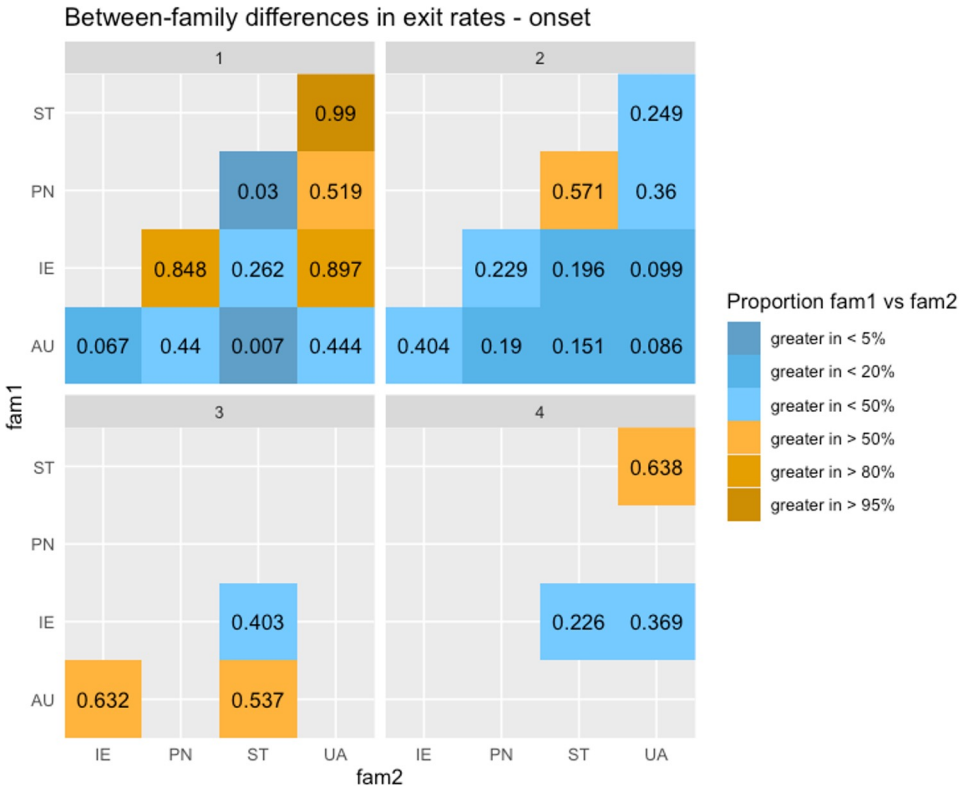
Brownian motion

\+ Ornstein-Uhlenbeck

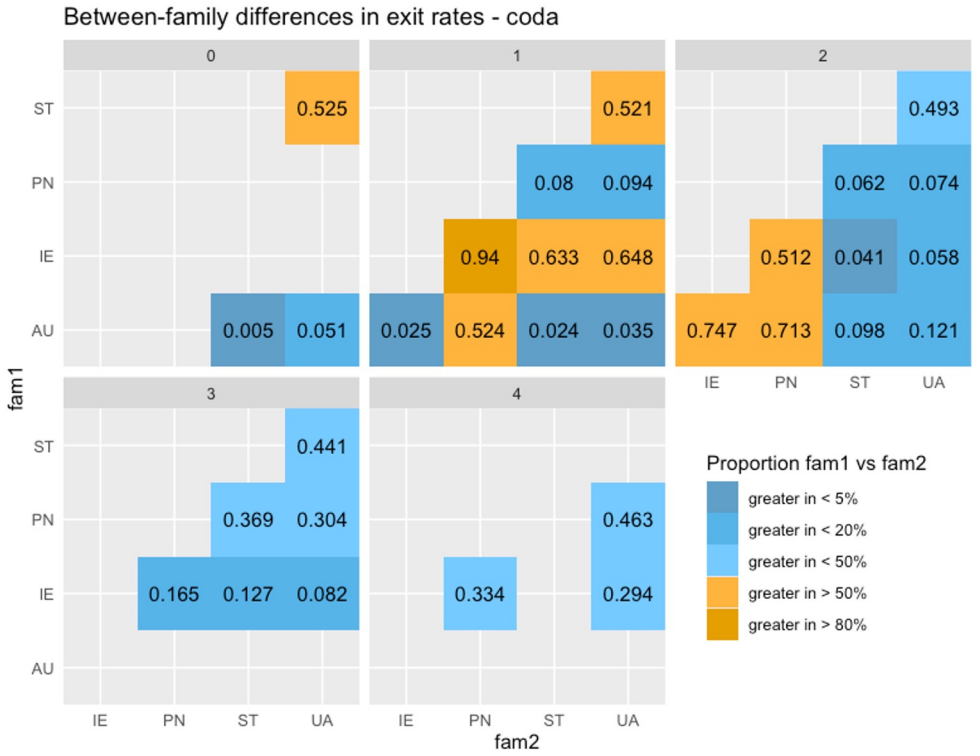

Between-family differences in stationary probabilities - onset
1
2
3
4
ST
PN
IE
AU
0.02
0.981
0.99
0.448
0.017
0.761
0.01
0.056
0.015
0.721
0.021
0.869
0.992
0.049
0.998
0.074
0.727
0.922
0.031
0.958
0.498
0.978
0.958
0.893
0.006
0.019
IE
PN
ST
UA
fam1
fam2
Proportion fam1 vs fam2
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%
greater in > 95%

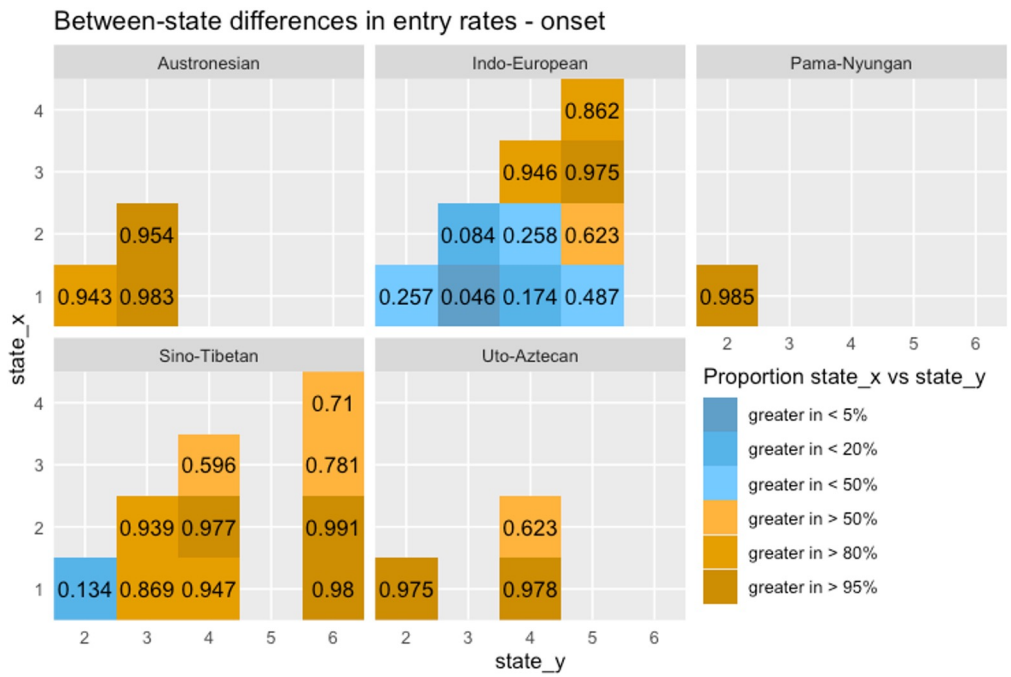

Between-family differences in stationary probabilities - coda
0
1
2
3
4
ST
PN
IE
AU
0.655
0.739
0.524
0.48
0.597
0.584
0.585
0.065
0.007
0.015
0.718
0.919
0.897
0.531
0.628
0.995
0.607
0.698
0.81
0.057
0.297
0.195
0.271
0.535
0.5
0.529
0.114
0.99
0.988
0.989
0.95
0.7
IE
PN
ST
UA
fam1
fam2
Proportion fam1 vs fam2
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%
greater in > 95%

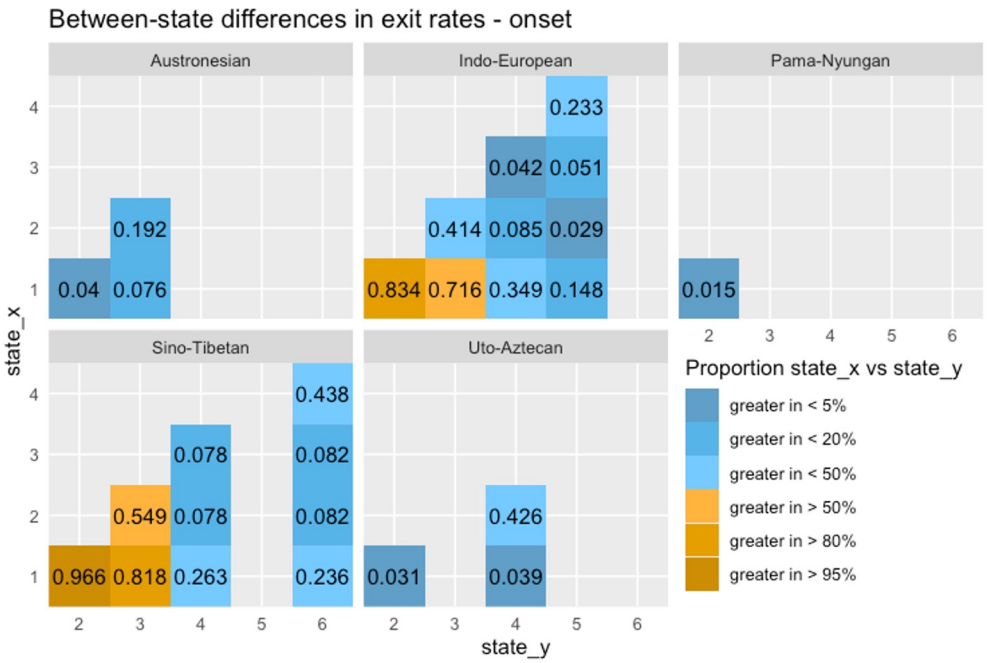

Between-family differences in entry rates - onset
1
2
3
4
ST
PN
IE
AU
0.224
0.947
0.684
0.461
0.139
0.588
0.042
0.041
0.041
0.374
0.032
0.443
0.892
0.147
0.198
0.103
0.62
0.425
0.035
0.54
0.63
0.786
0.675
0.767
0.025
0.127
IE
PN
ST
UA
fam1
fam2
Proportion fam1 vs fam2
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%

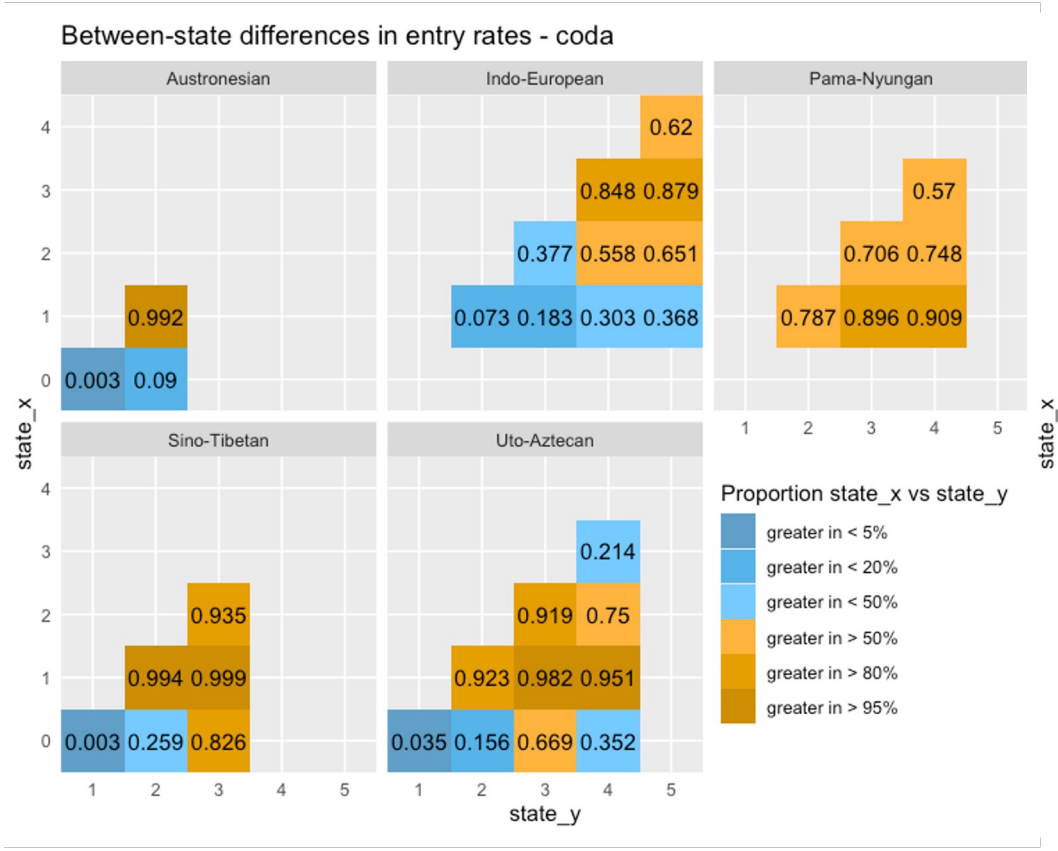

Between-family differences in exit rates - onset
1
2
3
4
ST
PN
IE
AU
0.99
0.249
0.03
0.519
0.571
0.36
0.848
0.262
0.897
0.229
0.196
0.099
0.067
0.44
0.007
0.444
0.404
0.19
0.151
0.086
0.638
0.403
0.226
0.369
0.632
0.537
IE
PN
ST
UA
fam1
fam2
Proportion fam1 vs fam2
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%
greater in > 95%


Between-family differences in entry rates - coda
0
1
2
3
4
ST
PN
IE
AU
0.666
0.653
0.516
0.055
0.122
0.105
0.116
0.285
0.016
0.048
0.767
0.268
0.286
0.021
0.1
0.851
0.662
0.054
0.154
0.161
0.497
0.038
0.052
0.491
0.413
0.404
0.907
0.86
0.845
0.123
0.849
0.447
IE
PN
ST
UA
fam1
fam2
Proportion fam1 vs fam2
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%

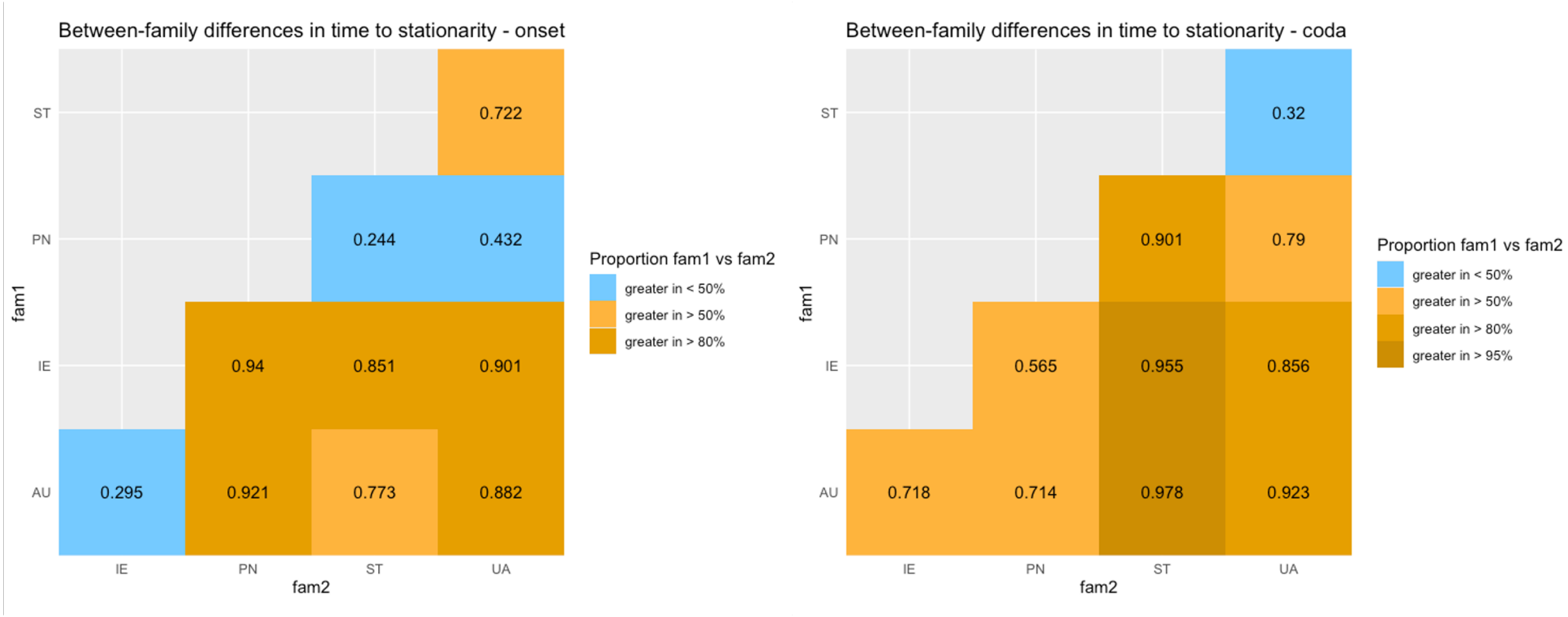

Between-family differences in exit rates - coda
0
1
2
3
4
ST
PN
IE
AU
0.525
0.521
0.493
0.08
0.094
0.062
0.074
0.94
0.633
0.648
0.512
0.041
0.058
0.005
0.051
0.025
0.524
0.024
0.035
0.747
0.713
0.098
0.121
0.441
0.369
0.304
0.463
0.165
0.127
0.082
0.334
0.294
IE
PN
ST
UA
fam1
fam2
Proportion fam1 vs fam2
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%


Between-state differences in entry rates - onset
Austronesian
Indo-European
Pama-Nyungan
Sino-Tibetan
Uto-Aztecan
0.943 0.983
0.954
0.257 0.046 0.174 0.487
0.084 0.258 0.623
0.946 0.975
0.862
0.985
0.134 0.869 0.947
0.939 0.977
0.596
0.98
0.991
0.781
0.71
0.975
0.978
0.623
state_x
state_y
Proportion state_x vs state_y
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%
greater in > 95%


Between-state differences in exit rates - onset
Austronesian
Indo-European
Pama-Nyungan
Sino-Tibetan
Uto-Aztecan
0.04 0.076
0.192
0.834 0.716 0.349 0.148
0.414 0.085 0.029
0.042 0.051
0.233
0.015
0.966 0.818 0.263
0.549 0.078
0.078
0.236
0.082 0.082
0.438
0.031
0.039
0.426
state_x
state_y
Proportion state_x vs state_y
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%
greater in > 95%


Between-state differences in entry rates - coda
Austronesian
Indo-European
Pama-Nyungan
Sino-Tibetan
Uto-Aztecan
state_x
state_y
0.003 0.09 0.992
0.073 0.183 0.303 0.368 0.377 0.558 0.651 0.848 0.879 0.62
0.787 0.896 0.909 0.706 0.748 0.57
0.003 0.259 0.826 0.994 0.999 0.935
0.035 0.156 0.669 0.352 0.923 0.982 0.951 0.919 0.75 0.214
Proportion state_x vs state_y
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%
greater in > 95%


Between-state differences in exit rates - coda
Austronesian
Indo-European
Pama-Nyungan
Sino-Tibetan
Uto-Aztecan
state_x
state_y
0.897 0.142 0.009
0.907 0.565 0.322 0.14 0.257 0.133 0.04 0.178 0.071 0.253
0.227 0.014 0.024 0.048 0.07 0.432
0.967 0.702 0.479 0.017 0.097 0.348
0.896 0.616 0.399 0.51 0.035 0.058 0.062 0.287 0.373 0.614
Proportion state_x vs state_y
greater in < 5%
greater in < 20%
greater in < 50%
greater in > 50%
greater in > 80%
greater in > 95%


Between-family differences in time to stationarity - onset
ST
PN
IE
AU
fam1
0.722
0.244
0.432
0.94
0.851
0.901
0.295
0.921
0.773
0.882
IE
PN
ST
UA
fam2
Proportion fam1 vs fam2
greater in < 50%
greater in > 50%
greater in > 80%
Between-family differences in time to stationarity - coda
ST
PN
IE
AU
fam1
0.32
0.901
0.79
0.565
0.955
0.856
0.718
0.714
0.978
0.923
IE
PN
ST
UA
fam2
Proportion fam1 vs fam2
greater in < 50%
greater in > 50%
greater in > 80%
greater in > 95%