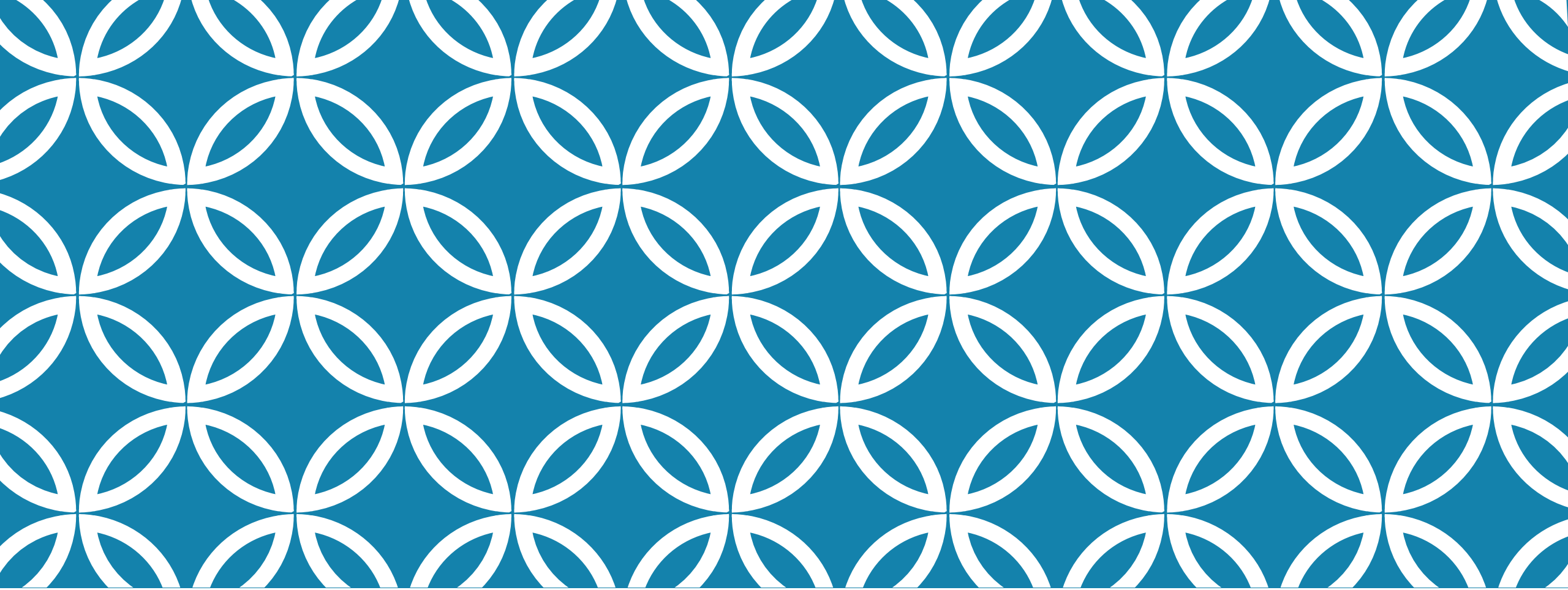
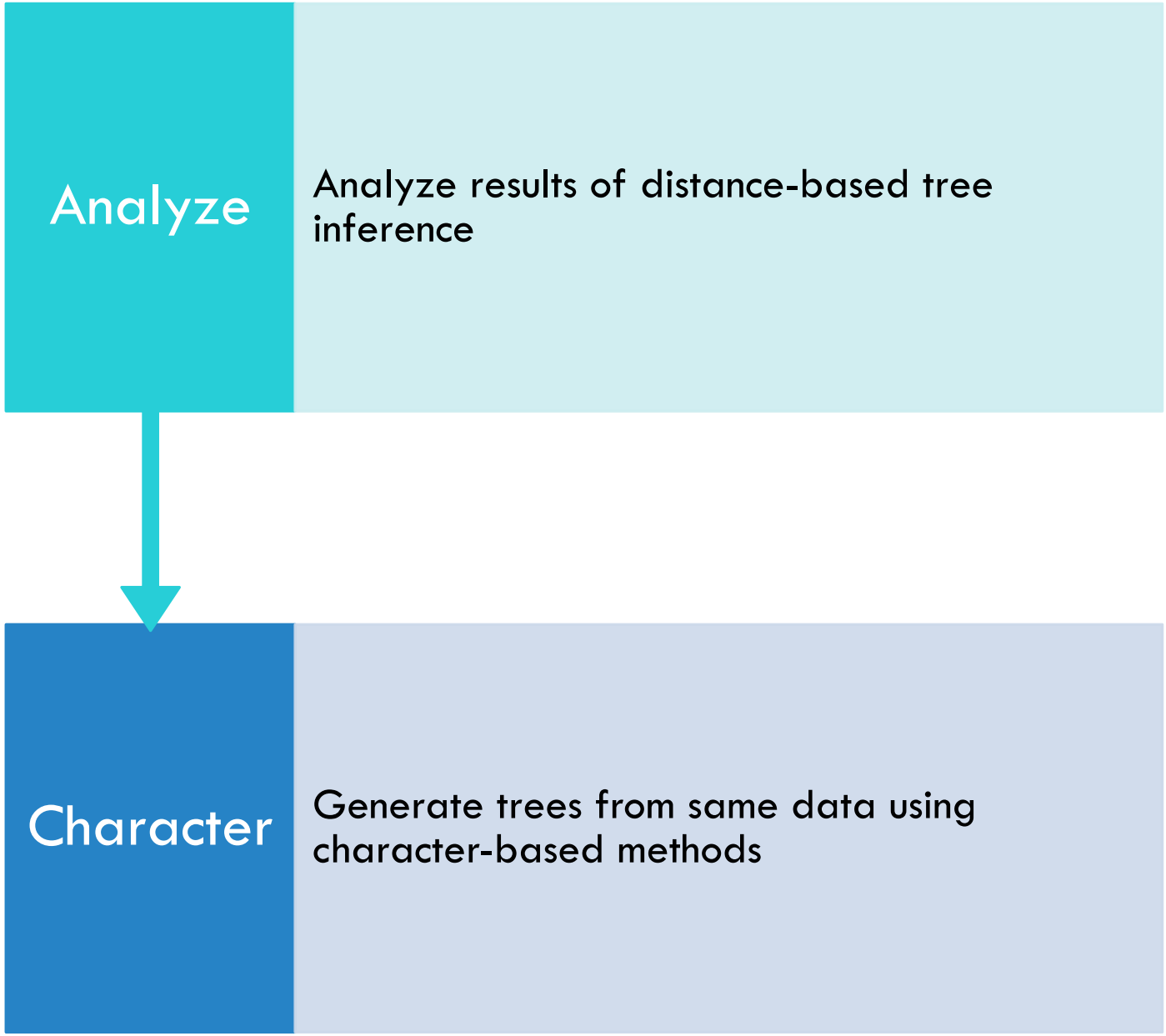


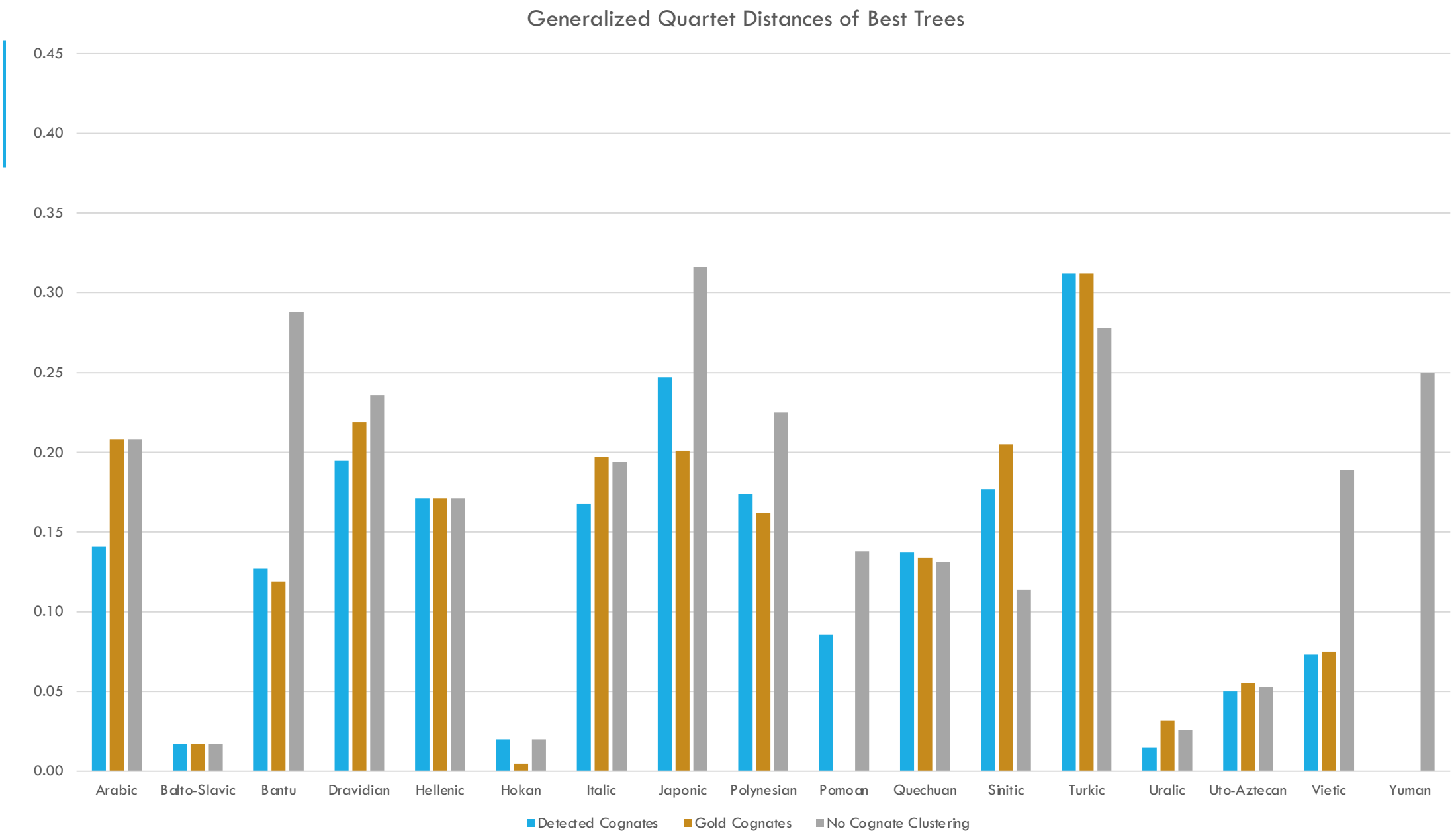
MASTER'S THESIS MEETING

Philip Georgis

November 15, 2021

CURRENT TASKS

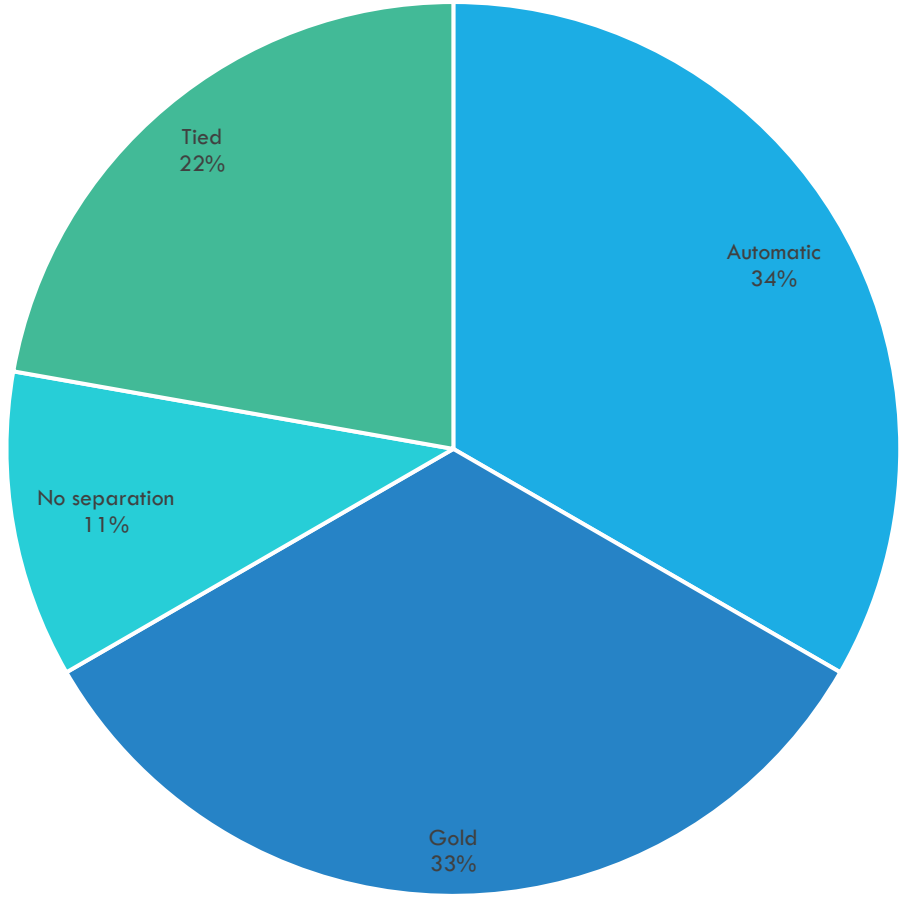



Generalized Quartet Distances of Best Trees
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Arabic
Balto-Slavic
Bantu
Dravidian
Hellenic
Hokan
Italic
Japonic
Polynesian
Pomoan
Quechuan
Sinitic
Turkic
Uralic
Uto-Aztecan
Vietic
Yuman
Detected Cognates
Gold Cognates
No Cognate Clustering

COGNATE SET BASIS OF BEST TREES

Family	Cognate Set
Arabic	Automatic
Balto-Slavic	*Tied*
Bantu	Gold
Dravidian	Automatic
Hellenic	*Tied*
Hokan	Gold
Italic	Automatic
Japonic	Gold
Polynesian	Gold

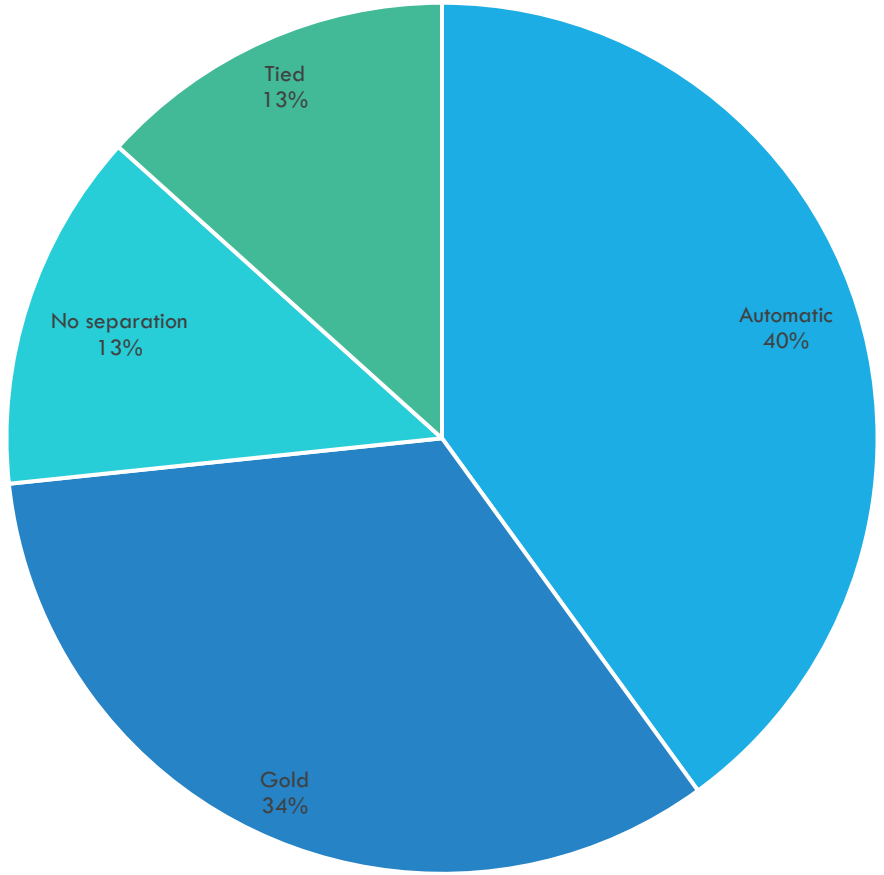
Family	Cognate Set
Pomoan	Gold
Quechuan	Gold
Sinitic	No separation
Turkic	No separation
Uralic	Automatic
Uto-Aztecan	Automatic
Vietic	Automatic
Yana	*Tied*
Yuman	*Tied*



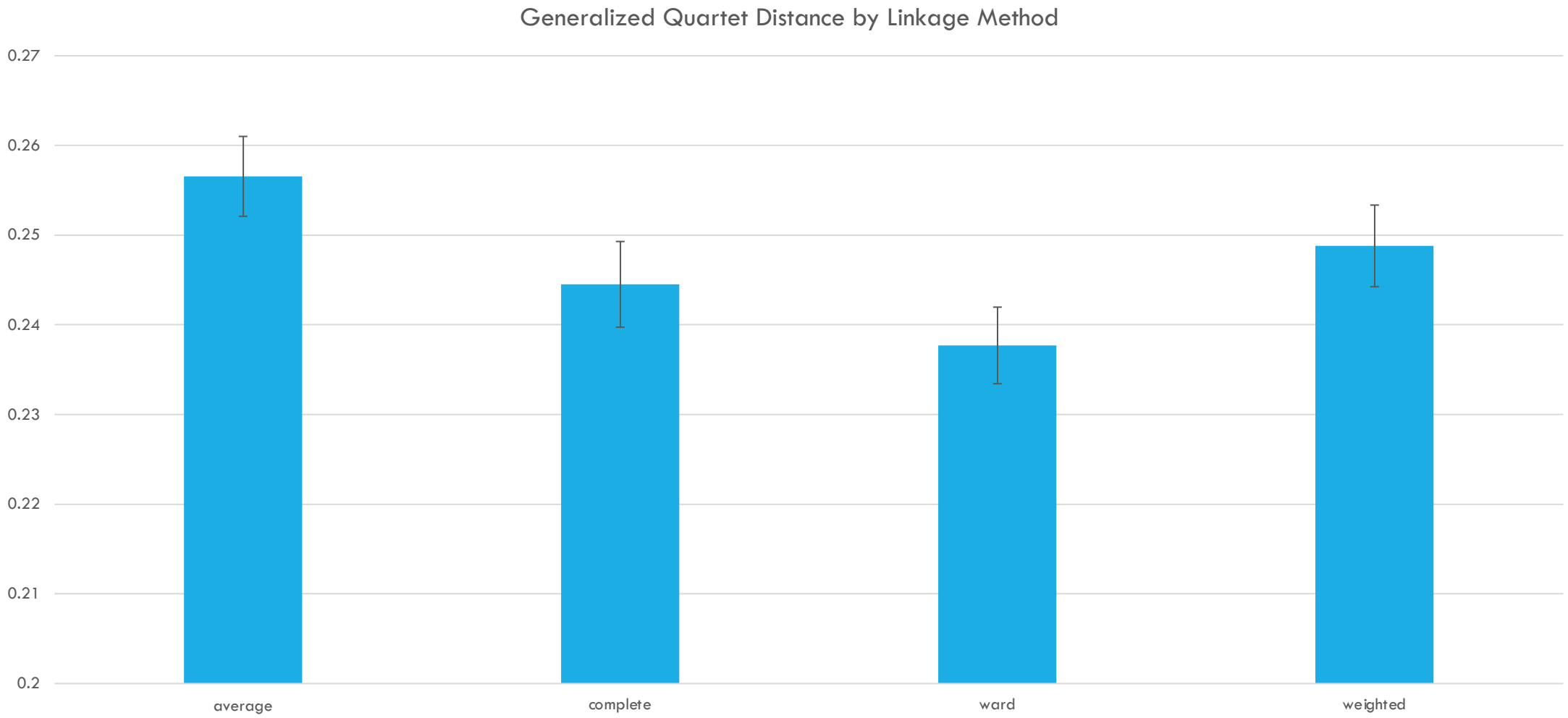
COGNATE SET BASIS OF BEST TREES

Family	Cognate Set
Arabic	Automatic
Balto-Slavic	*Tied*
Bantu	Gold
Dravidian	Automatic
Hellenic	*Tied*
Hokan	Gold
Italic	Automatic
Japonic	Gold
Polynesian	Gold

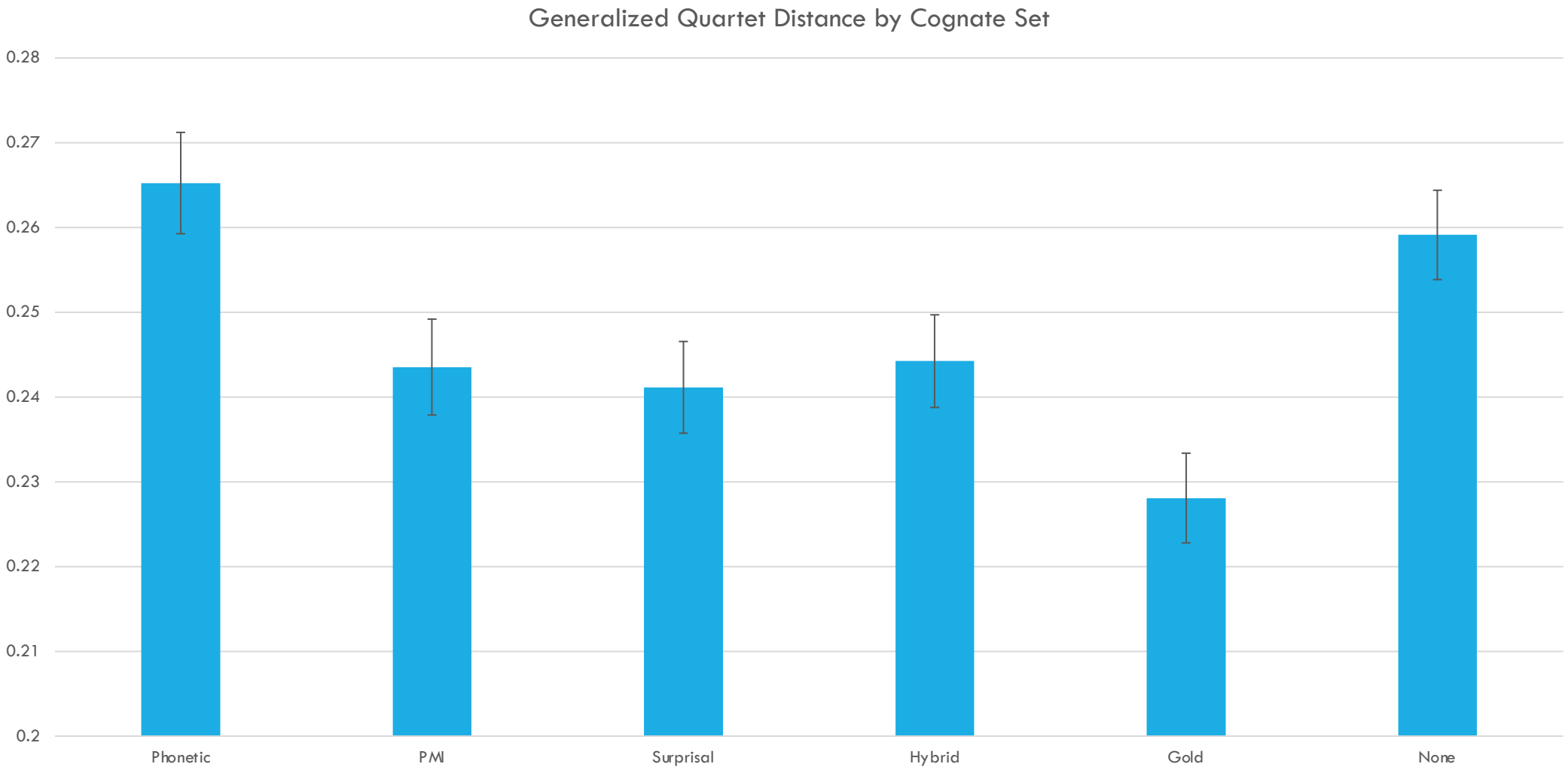
Family	Cognate Set
Pomoan	Gold
Quechuan	Gold
Sinitic	No separation
Turkic	No separation
Uralic	Automatic
Uto-Aztecan	Automatic
Vietic	Automatic
Yana	*Tied*
Yuman	*Tied*



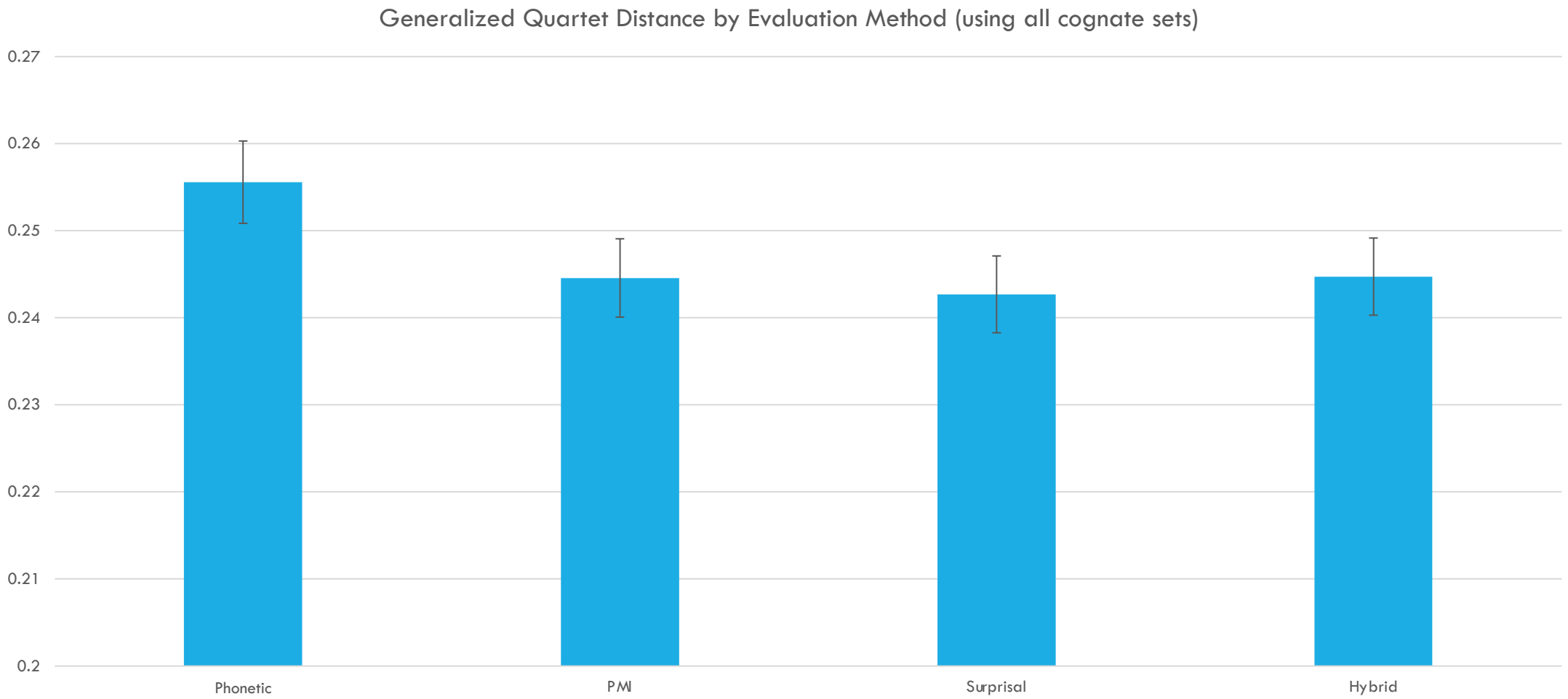
DISTANCE-BASED TREE ANALYSIS



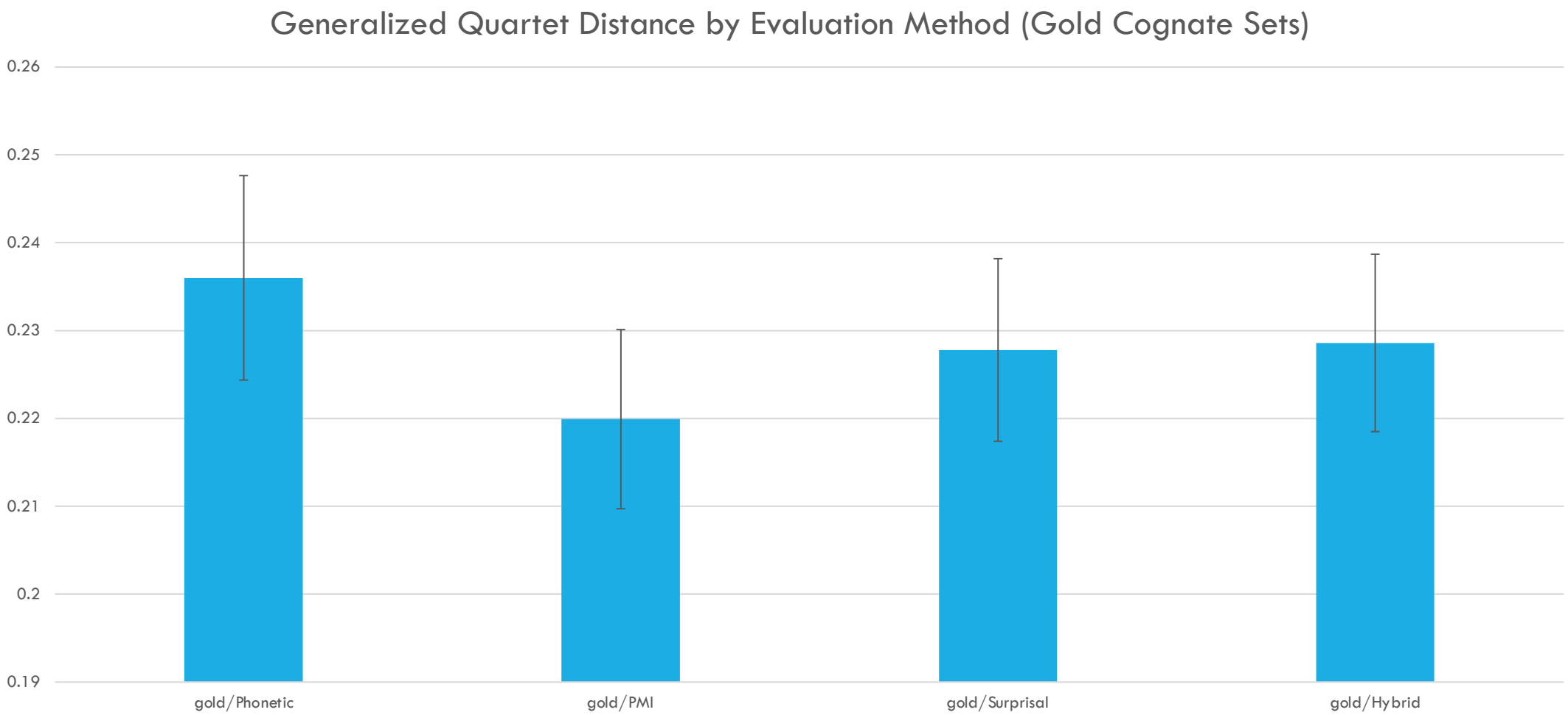
DISTANCE-BASED TREE ANALYSIS



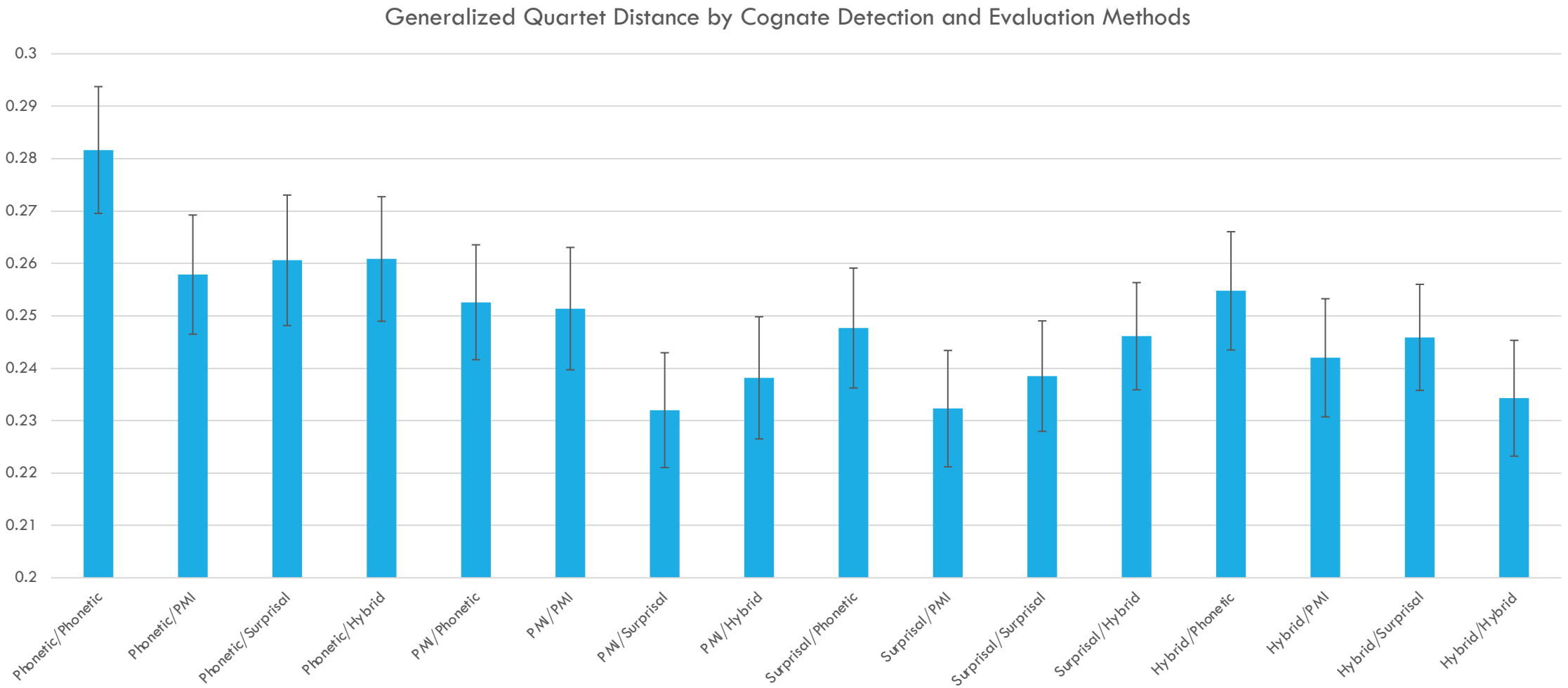
DISTANCE-BASED TREE ANALYSIS



DISTANCE-BASED TREE ANALYSIS



DISTANCE-BASED TREE ANALYSIS



BEST TREE SCORES

by automatic cognate set used

Phonetic only:
Arabic, Dravidian, Japonic, Uralic

PMI only:
Italic, Sinitic

Surprisal only:
Bantu, Turkic, Uto-Aztecan, Vietic

Hybrid only:
Polynesian, Pomoan, Quechuan

Tied:
Balto-Slavic, Hellenic, Yana, Yuman

Family	Phonetic	PMI	Surprisal	Hybrid
Arabic	**0.141**	0.250	0.208	0.208
Balto-Slavic	0.033	**0.017**	**0.017**	**0.017**
Bantu	0.218	0.288	**0.127**	0.148
Dravidian	**0.233**	0.236	0.248	0.239
Hellenic	**0.171**	**0.171**	**0.171**	**0.171**
Hokan	**0.023**	0.056	0.038	**0.023**
Italic	0.194	**0.168**	0.216	0.257
Japonic	**0.247**	0.254	0.255	0.272
Polynesian	0.190	0.182	0.189	**0.174**
Pomoan	0.293	0.293	0.138	**0.086**
Quechuan	0.140	0.140	0.140	**0.137**
Sinitic	0.200	**0.177**	0.224	0.241
Turkic	0.439	0.333	**0.295**	0.350
Uralic	**0.036**	0.042	0.046	0.042
Uto-Aztecan	0.054	0.061	**0.050**	0.090
Vietic	0.220	0.210	**0.074**	0.245
Yana	**0.000**	**0.000**	**0.000**	**0.000**
Yuman	0.250	**0.000**	**0.000**	**0.000**

BEST TREE SCORES

by word form evaluation method

Phonetic only:
Bantu, Japonic, Uto-Aztecan

PMI only:
Dravidian, Polynesian

Surprisal only:
Arabic, Quechuan, Turkic

Hybrid only:
Italic, Sinitic

Tied:
Balto-Slavic, Hellenic, Hokan, Pomoan, Uralic, Vietic, Yana, Yuman

Family	Phonetic	PMI	Surprisal	Hybrid
Arabic	0.261	0.202	**0.141**	0.202
Balto-Slavic	**0.017**	0.029	**0.017**	**0.017**
Bantu	**0.127**	0.218	0.261	0.217
Dravidian	0.239	**0.233**	0.238	0.236
Hellenic	**0.171**	**0.171**	**0.171**	**0.171**
Hokan	**0.023**	**0.023**	0.038	**0.023**
Italic	0.228	0.255	0.234	**0.168**
Japonic	**0.247**	0.254	0.302	0.289
Polynesian	0.180	**0.174**	0.190	0.201
Pomoan	0.224	**0.086**	**0.086**	0.224
Quechuan	0.148	0.142	**0.137**	0.141
Sinitic	0.207	0.217	0.189	**0.177**
Turkic	0.379	0.354	**0.295**	0.361
Uralic	**0.036**	**0.036**	0.046	0.042
Uto-Aztecan	**0.050**	0.055	0.056	0.054
Vietic	0.245	**0.074**	**0.074**	0.098
Yana	**0.000**	**0.000**	**0.000**	**0.000**
Yuman	**0.000**	**0.000**	**0.000**	**0.000**

Family	Phon-Phon	Phon-PMI	Phon-Surp	Phon-Hybr	PMI-Phon	PMI-PMI	PMI-Surp	PMI-Hybr	Surp-Phon	Surp-PMI	Surp-Surp	Surp-Hybr	Hybr-Phon	Hybr-PMI	Hybr-Surp	Hybr-Hybr
Arabic	0.295	0.202	**0.141**	0.202	0.268	0.323	0.250	0.382	0.280	0.356	0.208	0.208	0.261	0.265	0.208	0.208
Balto-Slavic	0.045	0.045	0.033	0.045	0.045	0.045	**0.017**	**0.017**	**0.017**	0.029	**0.017**	**0.017**	0.033	0.029	**0.017**	**0.017**
Bantu	0.298	0.218	0.290	0.275	0.295	0.328	0.288	0.315	**0.127**	0.234	0.315	0.217	0.148	0.266	0.261	0.243
Dravidian	0.277	**0.233**	0.238	0.252	0.255	0.255	0.248	0.236	0.248	0.257	0.263	0.258	0.239	0.242	0.249	0.249
Hellenic	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**	**0.171**
Hokan	0.040	**0.023**	0.081	**0.023**	0.061	0.071	0.061	0.056	0.038	0.038	0.038	0.038	**0.023**	0.038	0.052	0.038
Italic	0.309	0.284	0.263	0.194	0.308	0.274	0.234	**0.168**	0.228	0.255	0.238	0.216	0.277	0.257	0.258	0.265
Japonic	**0.247**	0.291	0.336	0.289	0.263	0.254	0.310	0.301	0.255	0.338	0.318	0.305	0.272	0.293	0.302	0.320
Polynesian	0.237	0.236	0.190	0.255	0.182	0.190	0.204	0.206	0.189	0.217	0.206	0.234	0.180	**0.174**	0.198	0.201
Pomoan	0.293	0.293	0.293	0.293	0.362	0.362	0.293	0.362	0.276	0.138	0.224	0.224	0.224	**0.086**	**0.086**	0.224
Quechuan	0.148	0.153	0.140	0.354	0.152	0.157	0.140	0.144	0.163	0.171	0.140	0.360	0.166	0.142	**0.137**	0.141
Sinitic	0.207	0.219	0.203	0.200	0.250	0.217	0.189	**0.177**	0.246	0.276	0.224	0.249	0.246	0.283	0.241	0.246
Turkic	0.489	0.480	0.439	0.441	0.379	0.419	0.333	0.361	0.380	0.358	**0.295**	0.375	0.411	0.354	0.350	0.375
Uralic	**0.036**	**0.036**	0.046	0.042	0.042	0.070	0.058	0.055	0.046	0.070	0.058	0.058	0.042	0.055	0.058	0.058
Uto-Aztecan	0.059	0.055	0.056	0.054	0.061	0.086	0.063	0.113	**0.050**	0.120	0.084	0.084	0.090	0.196	0.108	0.192
Vietic	0.246	0.220	0.220	0.220	0.245	0.220	0.210	0.220	0.245	**0.074**	**0.074**	0.098	0.245	0.267	0.245	0.267
Yana	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Yuman	0.250	0.250	0.250	0.250	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.250	0.250	**0.000**	**0.000**	0.250	**0.000**
AVERAGE	0.203	0.189	0.188	0.198	0.186	0.191	0.171	0.182	**0.164**	0.172	0.174	0.187	0.168	0.173	0.177	0.179
AVERAGE w/o Hokan subgroups	0.207	0.191	0.190	0.201	0.198	0.205	0.185	0.195	0.179	0.198	**0.177**	0.193	0.187	0.202	0.190	0.199

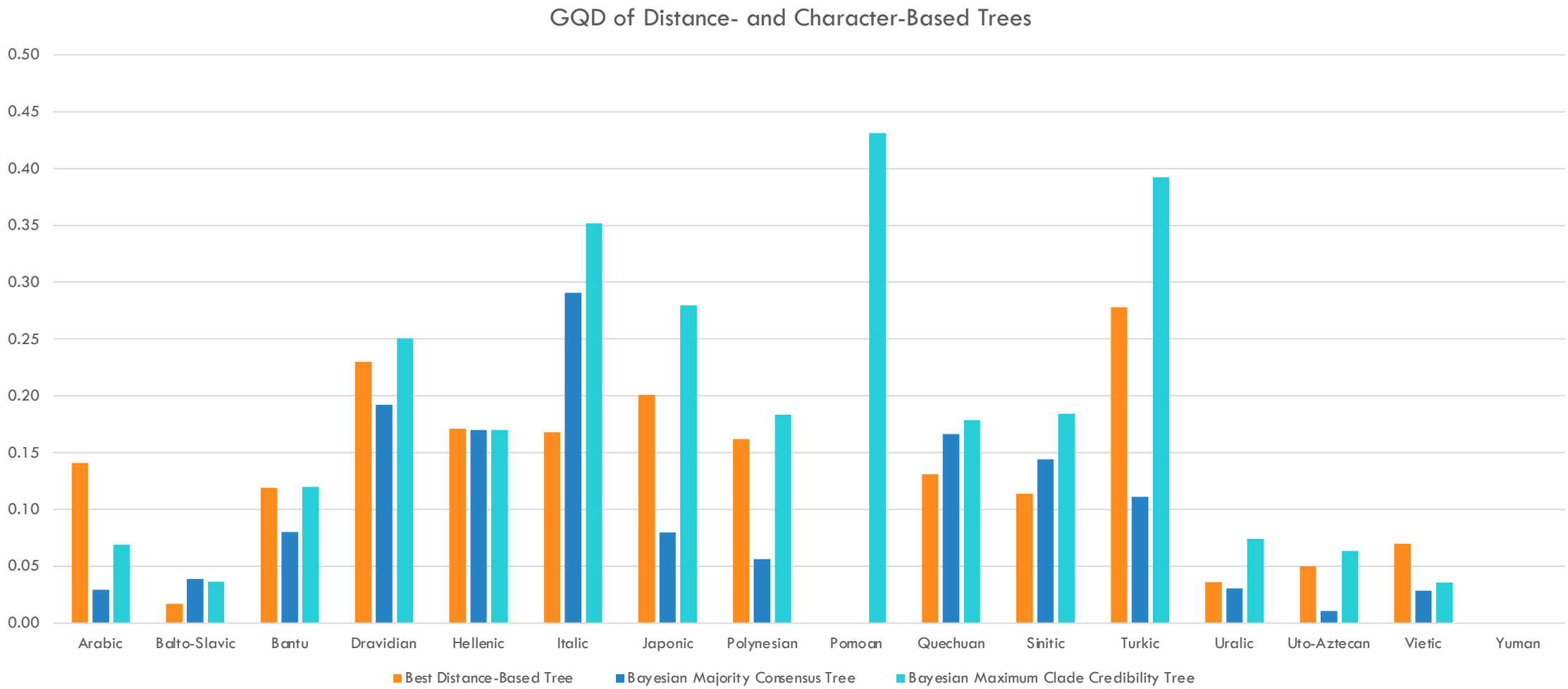
BAYESIAN TREE INFERENCE

- Used BEASTling package to prepare input for BEAST 2
- BEAST 2 settings (basic settings from tutorial)
  - Model: covarion
  - Chain length = 2 million
  - Rate variation = TRUE
  - Clock: relaxed

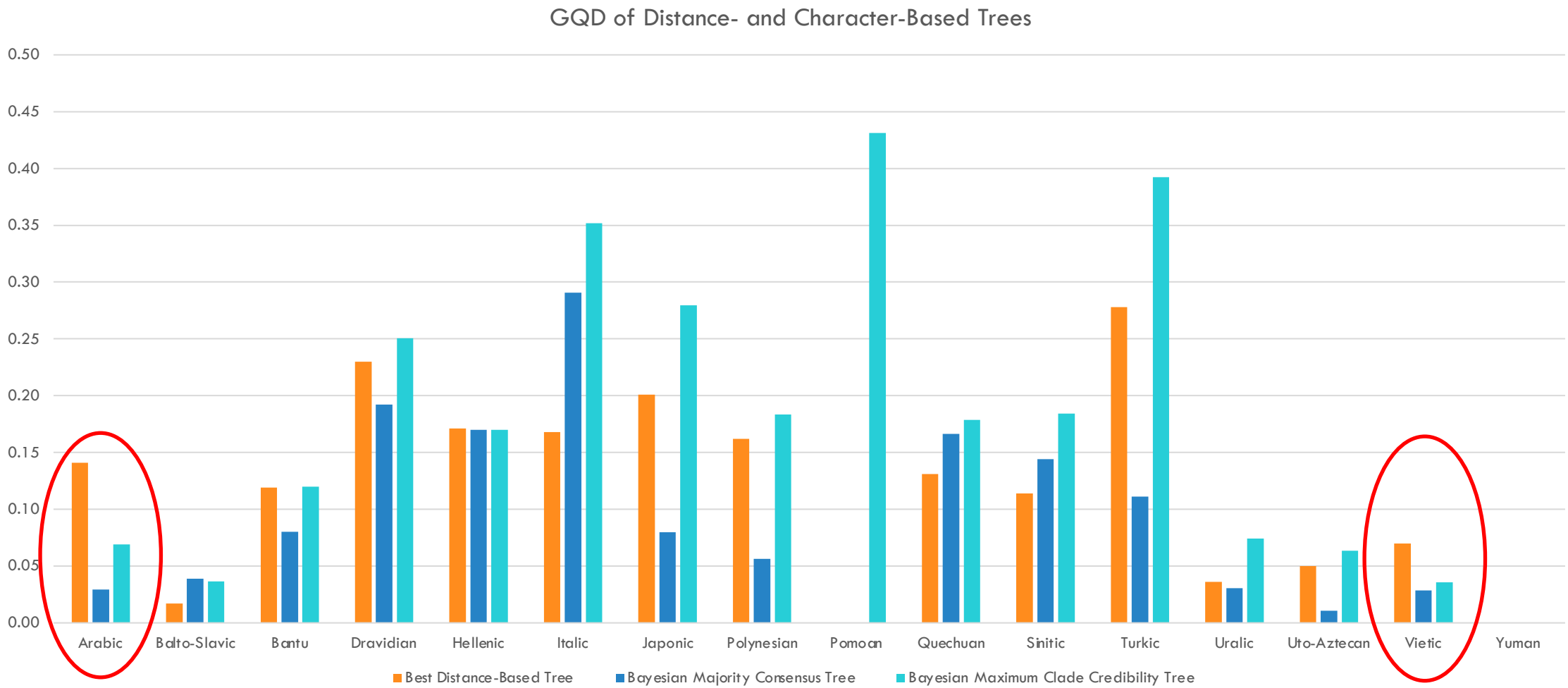
BAYESIAN TREE INFERENCE

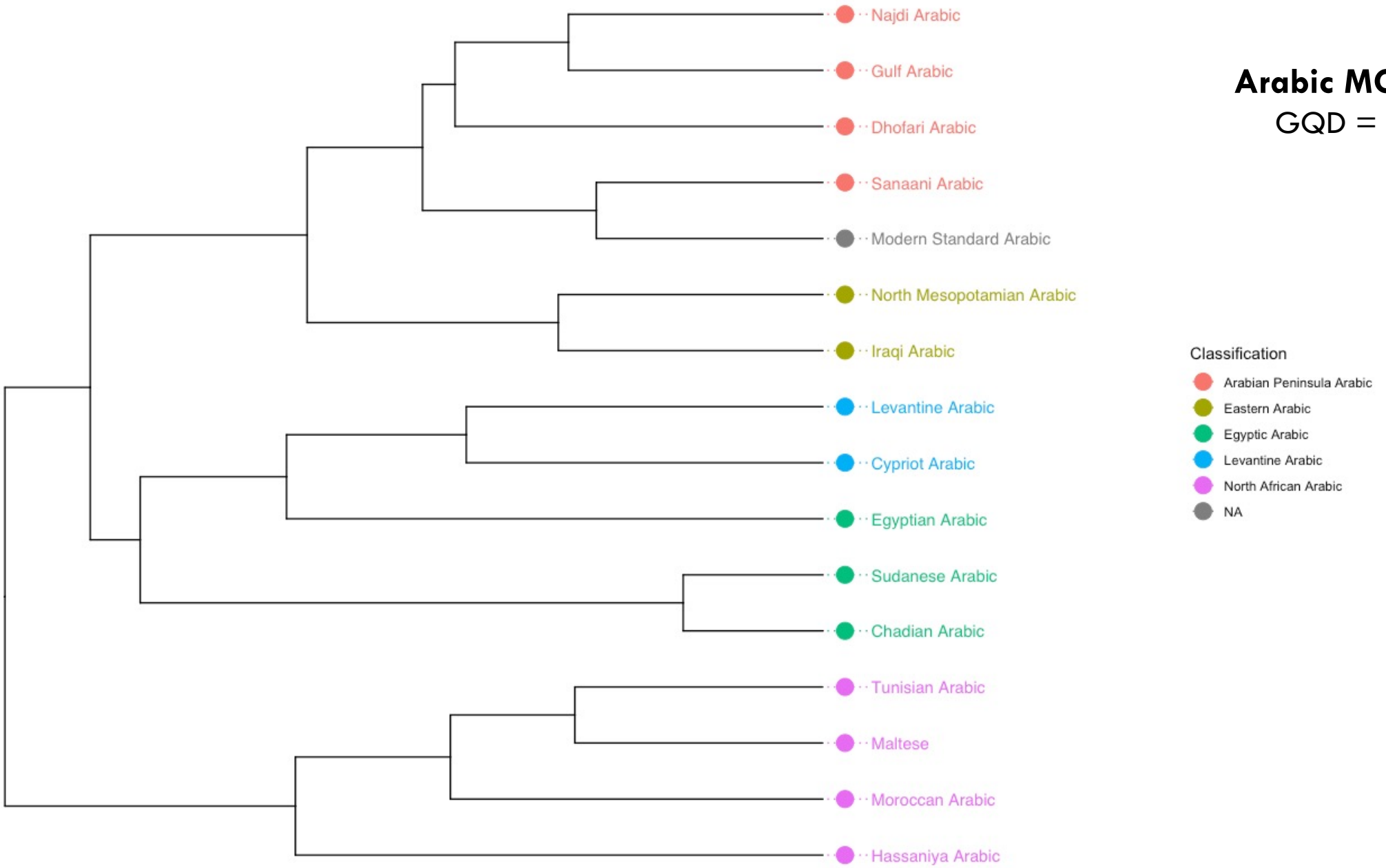
- Used BEASTling package to prepare input for BEAST 2
- BEAST 2 settings (basic settings from tutorial)
  - Model: covarion
  - Chain length = 2 million
  - Rate variation = TRUE
  - Clock: relaxed
- Ran on all families using gold cognate sets
  - Except Hokan, wouldn't work → ran on Pomoan and Yuman groups individually
  - >6 hours to run for all families with just these specific model settings
  - Extracted majority consensus and maximum clade credibility trees for each family
  - *How to get maximum likelihood or maximum parsimony trees?*

DISTANCE- VS. CHARACTER-BASED TREES



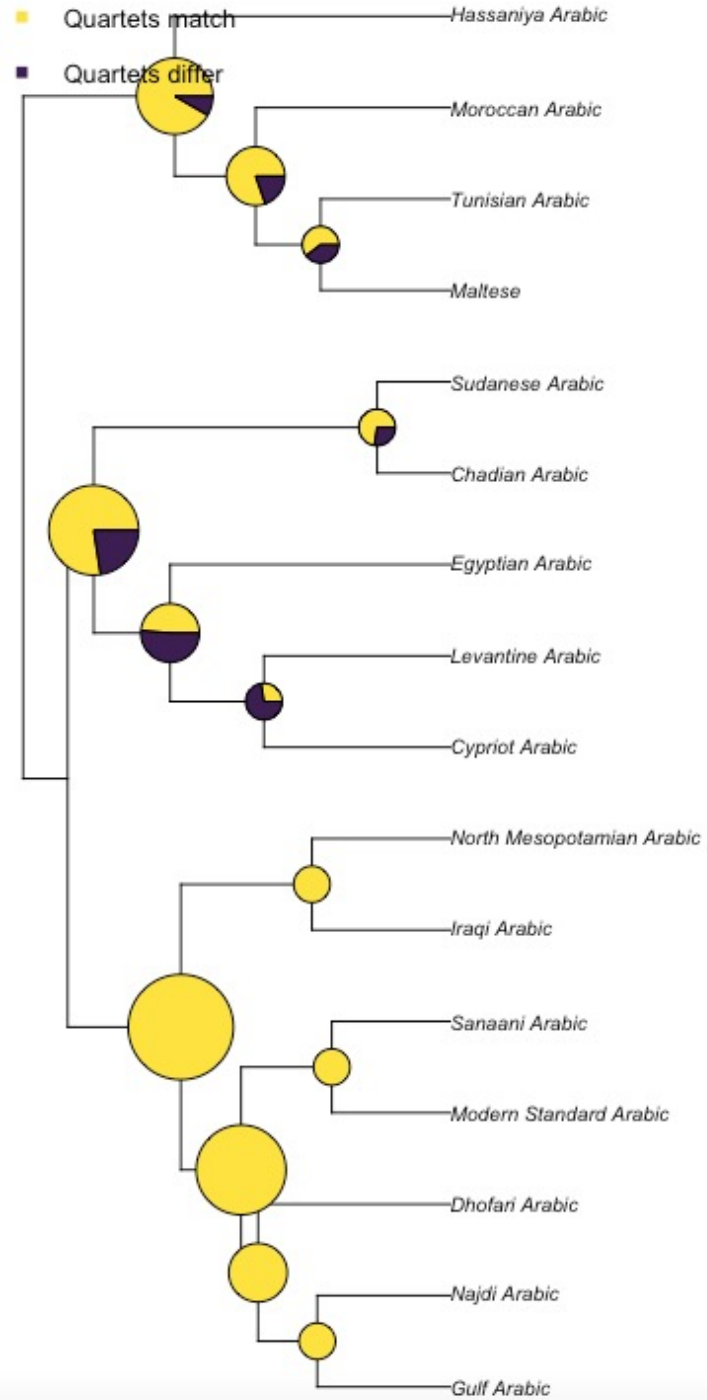
DISTANCE- VS. CHARACTER-BASED TREES

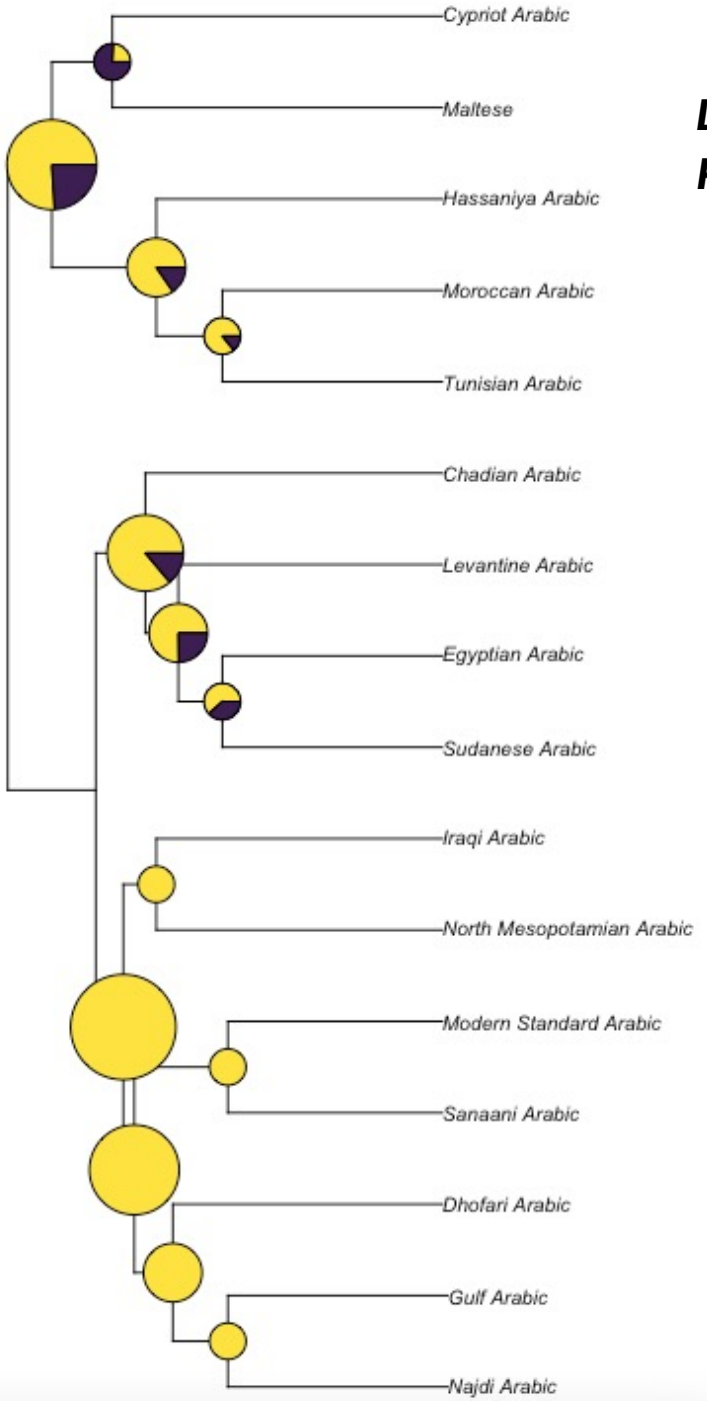




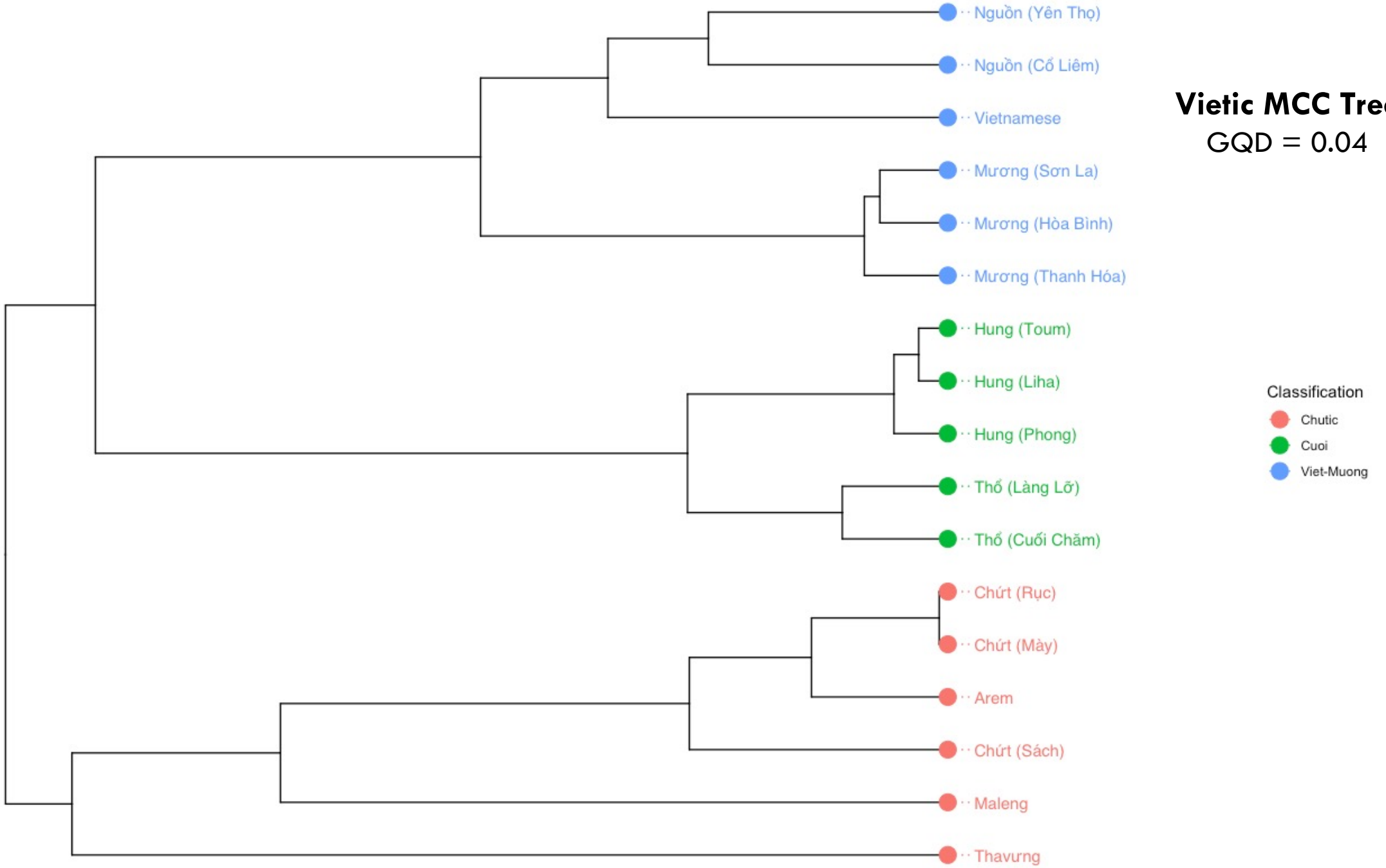
Arabic MCC Tree

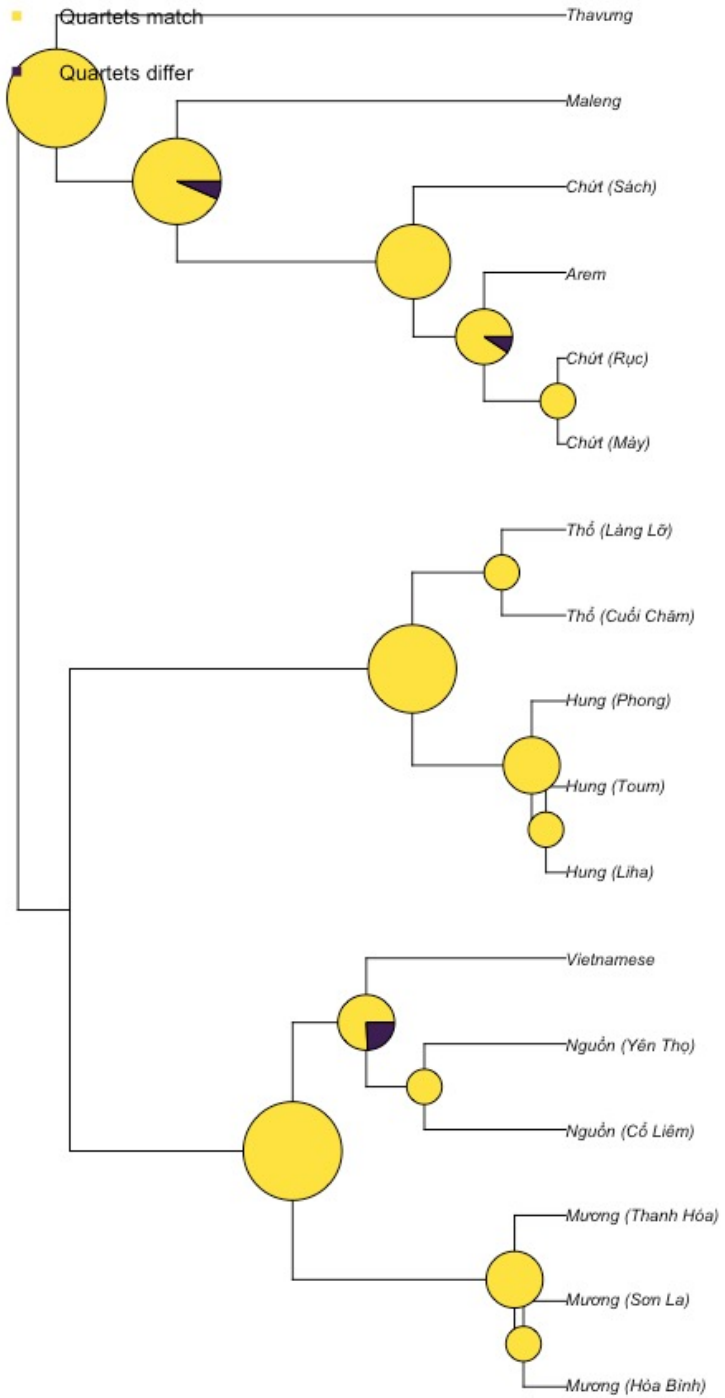
GQD = 0.07

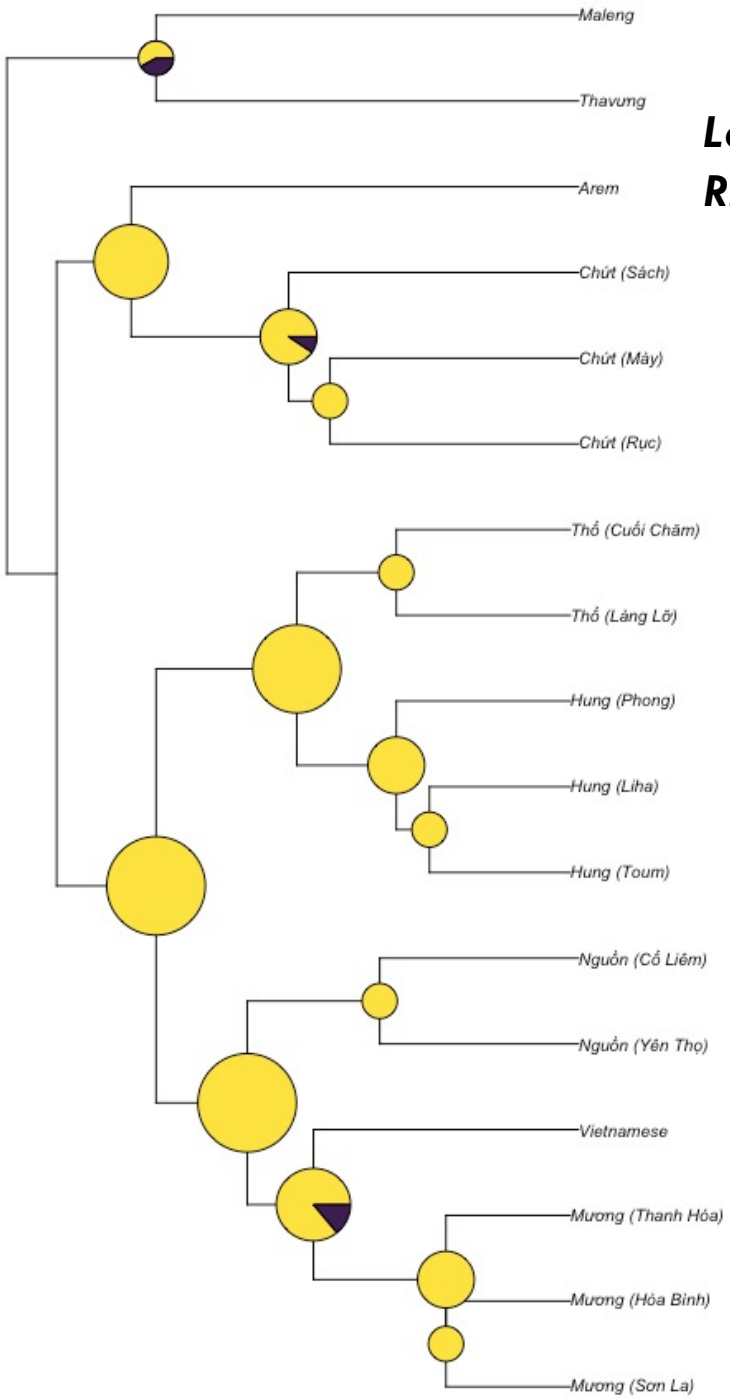




Left: *Bayesian MCC tree*
Right: *best distance-based tree*


Nguồn (Yên Thọ)
Nguồn (Cổ Liêm)
Vietnamese
Mương (Sơn La)
Mương (Hòa Bình)
Mương (Thanh Hóa)
Hung (Toum)
Hung (Liha)
Hung (Phong)
Thổ (Làng Lỡ)
Thổ (Cuối Chăm)
Chứt (Rục)
Chứt (Mày)
Arem
Chứt (Sách)
Maleng
Thavưng
Vietic MCC Tree
GQD = 0.04
Classification
Chutic
Cuoi
Viet-Muong





Left: *Bayesian MCC tree*
Right: *best distance-based tree*