

Phylogenetic approaches to culture

2.

Inferring Phylogenies

Fiona Jordan

NSF Summer Institutes for Cross-Cultural Anthropological Research

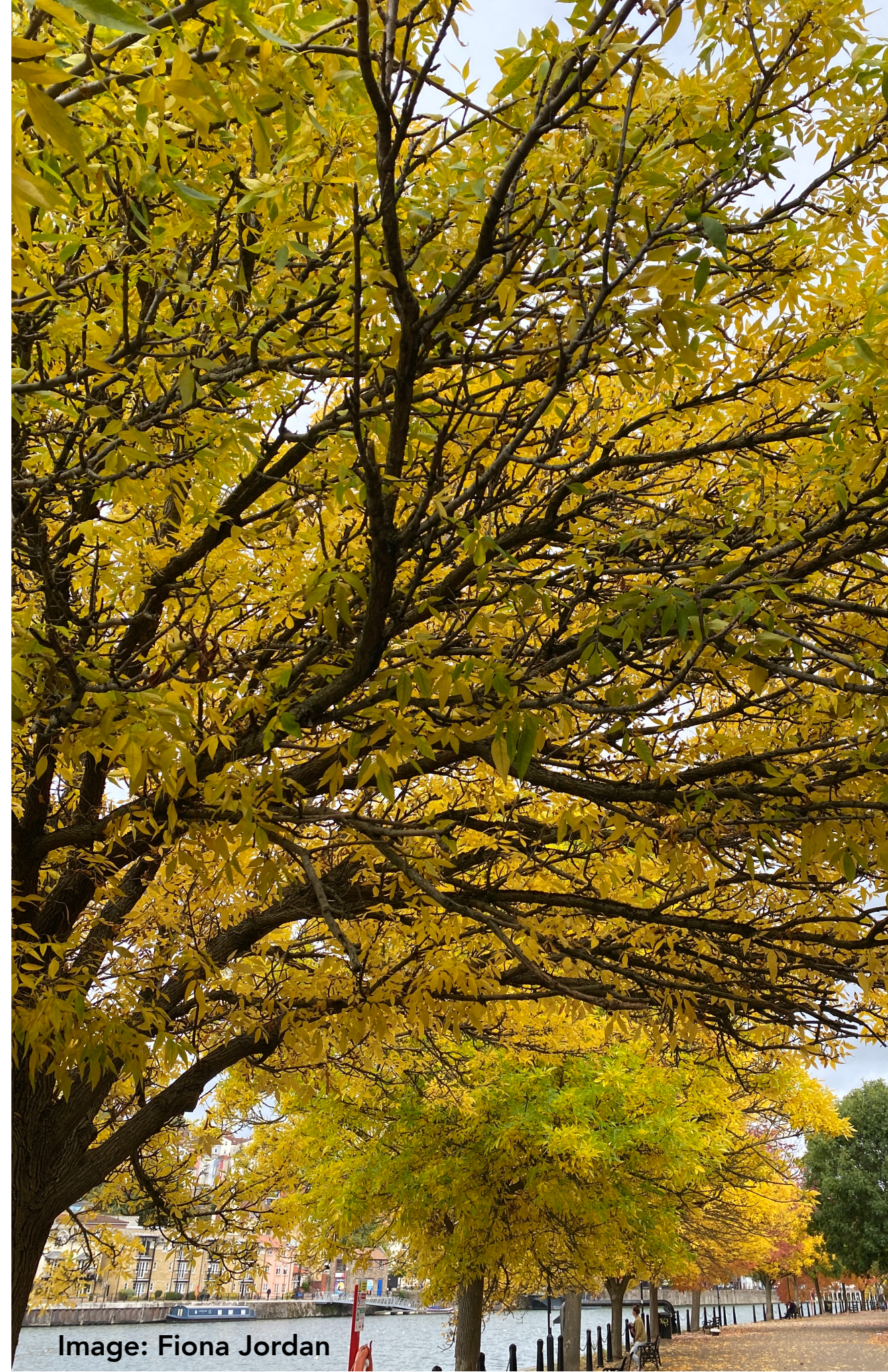
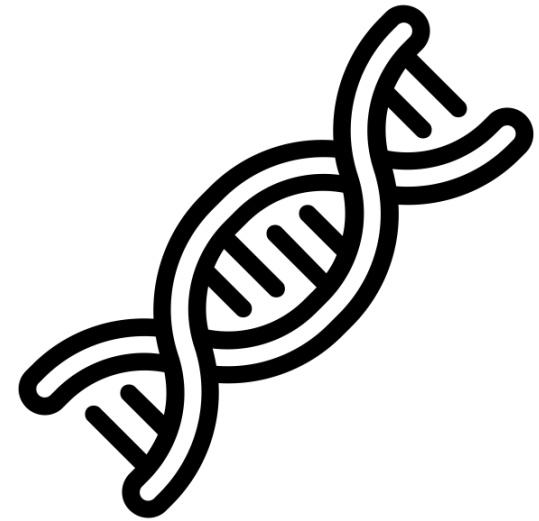


Image: Fiona Jordan

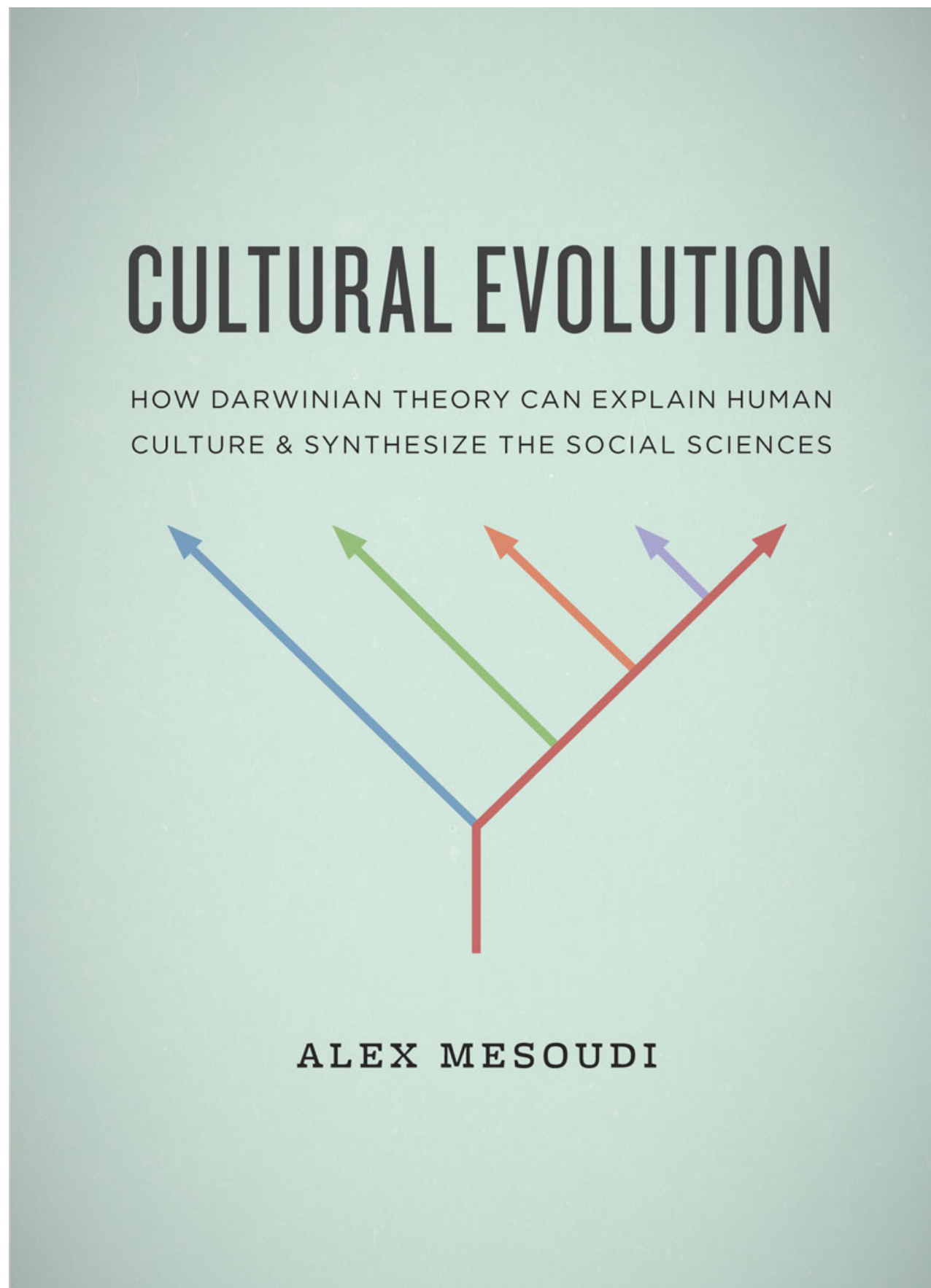
Inferring cultural phylogenies

What is the most appropriate data?

- Genes:
Slow timescale
Migrants > homogeneity
Limited diversity
Difficult & expensive
- Language:
Culturally-relevant timescale
Migrants > variation maintained
High diversity
Less difficult to source



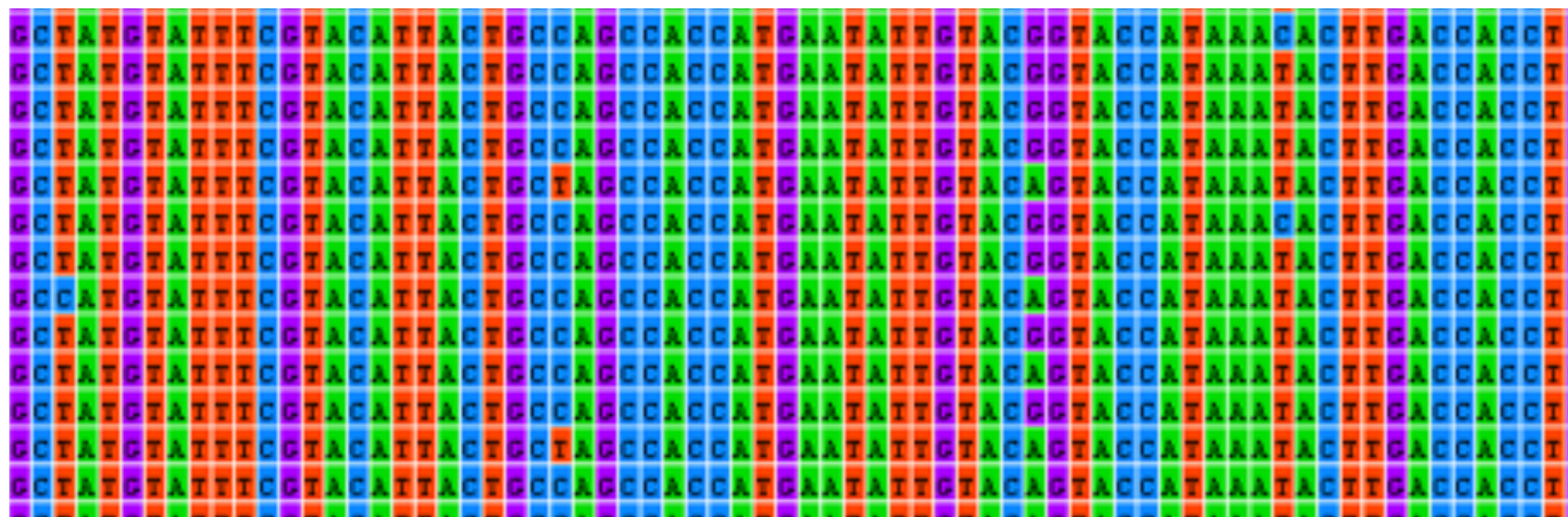
	biological	linguistic/cultural
inheritance	reproduction	social learning
transmission	vertical and horizontal	vertical, horizontal, oblique
units	genes ...	traits ...
rate of change	slow relative to generation time	variable!
accumulation	fossil record	technological sequences
forces	mutation, selection, drift	innovation, selection, drift, bias
populations	species, demes	cultures, lineages, ethnolinguistic groups



words as data

Cognate-coded vocabulary can be analysed like genetic data

Gloss	English	French	Russian	Ingush
BLOOD	blood (1)	sang (2)	krov' (3)	c'ii (4)
BONE	bone (1)	os (2)	kost' (2?)	t'exk (3)
DOG	dog (1)	chien (2)	sobaka (3)	zhwalii (4)
HEART	heart (1)	coeur (1)	serdce (1)	dog (2)
SUN	sun (1)	soleil (1)	solnce (1)	maalx (2)



I
you
we
this
that
who
what

all
many
one
two
not

red
green
yellow
white
black

man
woman
person
name

hot
cold
full
big
long
small
new
good
round
dry

hair
head
ear
eye
nose
mouth
tooth
tongue
foot
knee
hand
belly
neck
breasts
heart
liver
skin
flesh
blood
bone

drink
eat
bite
see
hear
know
sleep
die
kill
swim
burn
fly
walk
come
lie
sit
stand
give
say

fish
bird
dog
louse
egg
horn
tail
feather
claw

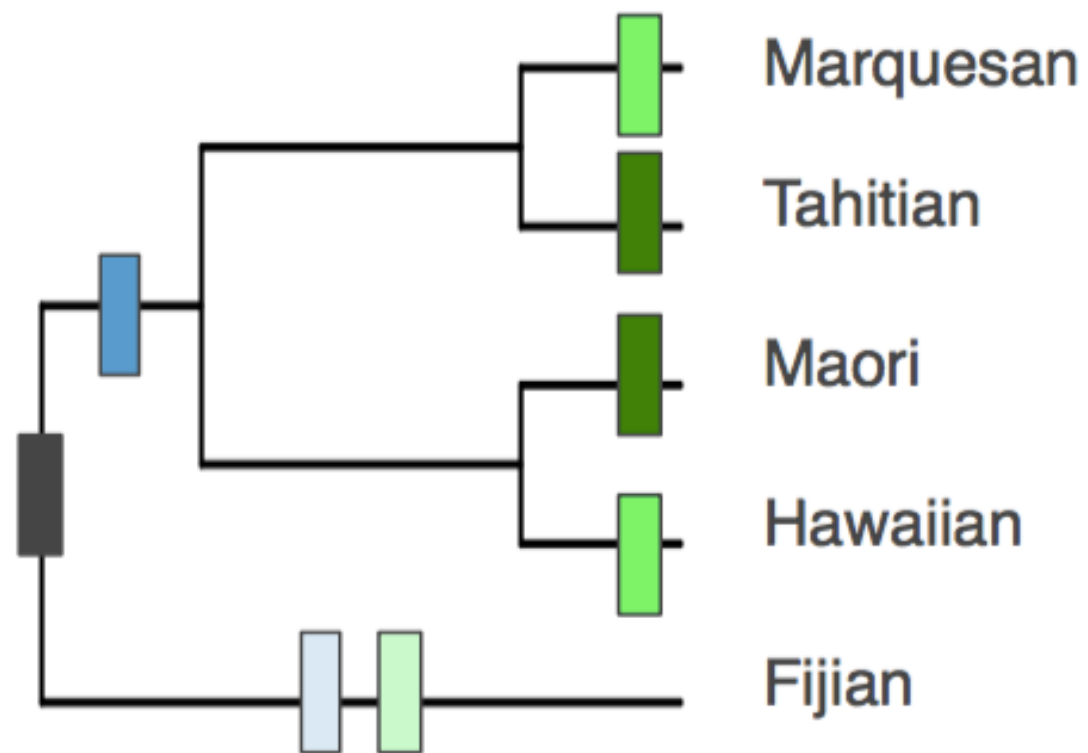
sun
moon
star
water
rain
cloud
night

stone
sand
earth
tree
seed
leaf
root
bark
grease
smoke
path
mountain
fire
ash

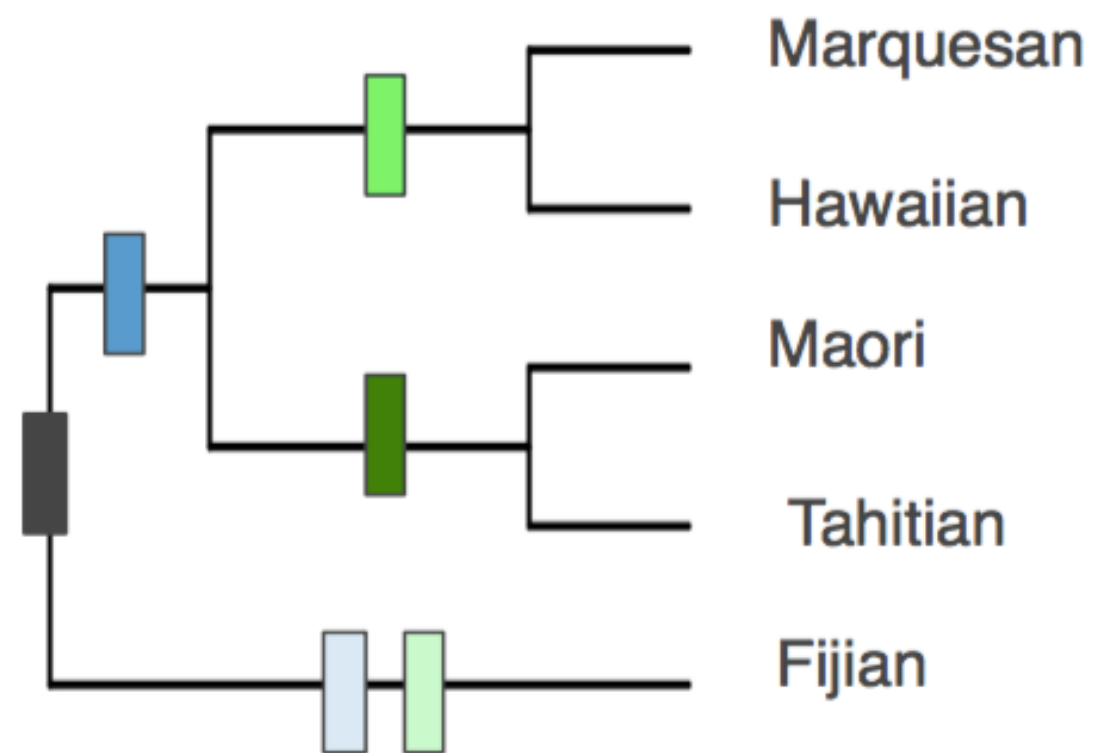
Swadesh
100-meaning list

	Taboo	Blood	To Suck
Fijian	tabu	drā	sucu-ma
Tahitian	tapu	toto	ngote
Maori	tapu	toto	ngote
Hawaiian	kapu	koko	omo
Marquesan	tapu	toto	omo

Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1



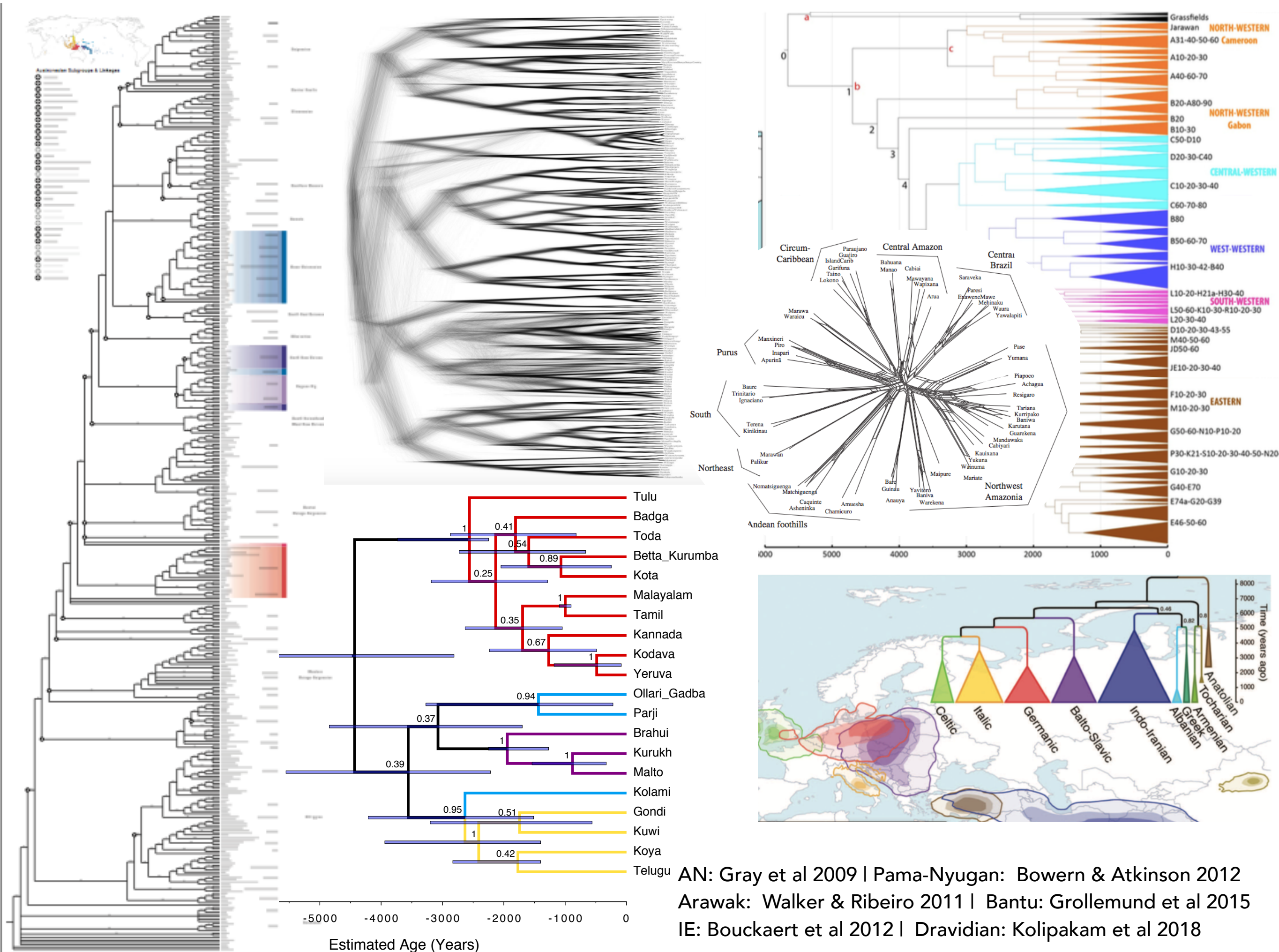
Length=8



Length=6

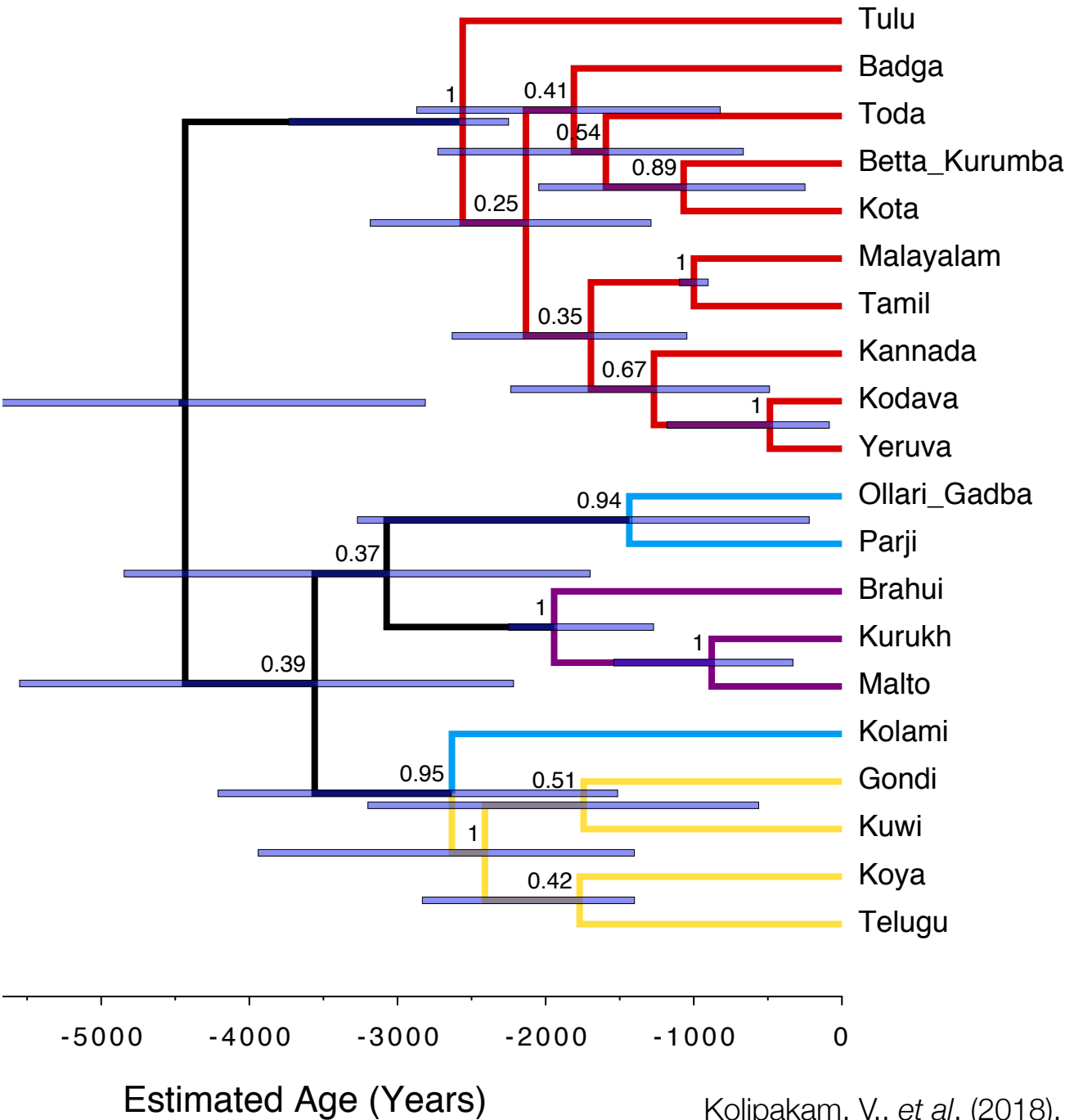


Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1



AN: Gray et al 2009 | Pama-Nyugan: Bown & Atkinson 2012
Arawak: Walker & Ribeiro 2011 | Bantu: Grollemund et al 2015
IE: Bouckaert et al 2012 | Dravidian: Kolipakam et al 2018

Dravidian language family is 4,500 years old, finds international study



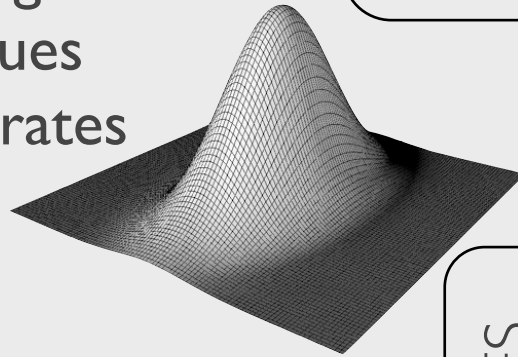
Michael Dunn Vishnu Kolipakam
Annemarie Verkerk

A Bayesian phylogenetic study of the Dravidian language family

Vishnupriya Kolipakam^{1,2}, Fiona M. Jordan^{2,3,4},
Michael Dunn^{2,5}, Simon J. Greenhill^{4,6},
Remco Bouckaert^{4,7}, Russell D. Gray⁴ and
Annemarie Verkerk^{2,4}

PARAMETERS

topology
branch lengths
node values
transition rates



MODELS

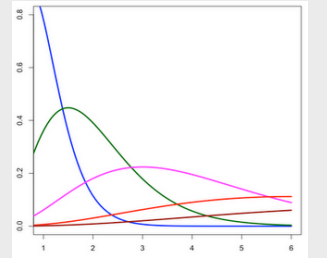
discrete characters
rate matrix
gamma rate heterogeneity
heterotachy: covarion

$$q_{01} = \alpha$$
$$q_{10} = \beta$$

PRIORS

information as a distribution

uniform
exponential
gamma



BAYES
THEOREM

likelihood

prior

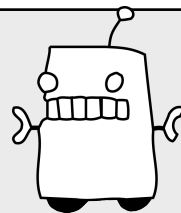
$$p(\theta|X) = \frac{p(X|\theta) p(\theta)}{p(X)}$$

posterior
probability

marginal probability

MCMC(MC)

“robot sampler”
Markov chain

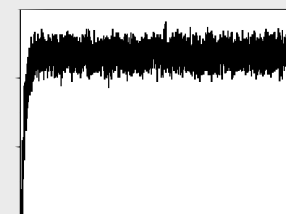


stationarity $\approx$ posterior distribution
Metropolis coupling

Bayesian phylogenetic methods

DIAGNOSTICS

convergence
ESS/autocorrelation
chains
trace plots

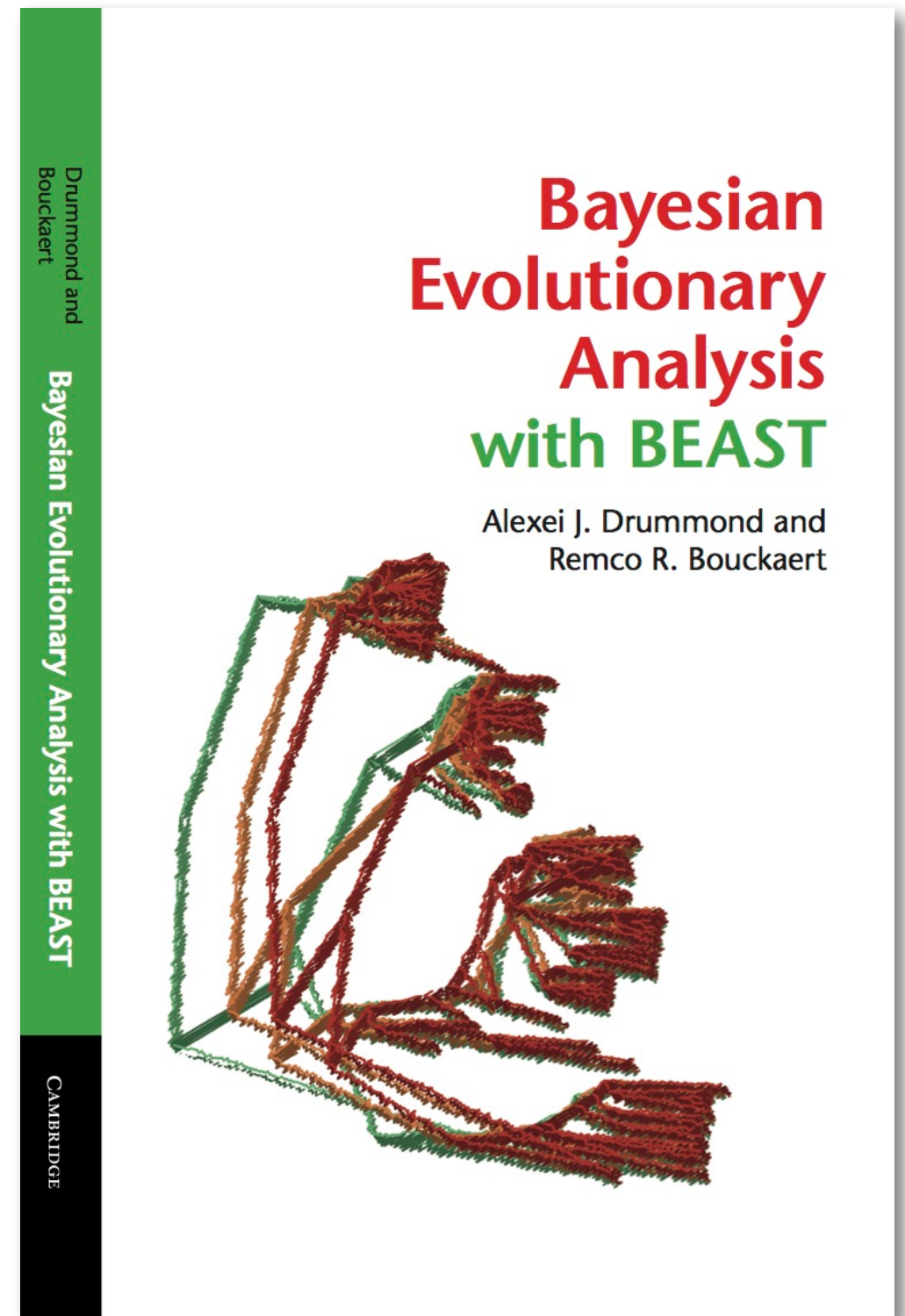


SUMMARY

histograms
credibility interval
marginal Lh
Bayes factors
model plots

Phylogenies and the BEAST

- Flexible probabilistic inference
- Test different models
- Models for language evolution
- Incorporate priors
 - Archaeology, history, geography
- Incorporate uncertainty
- Multiple data types



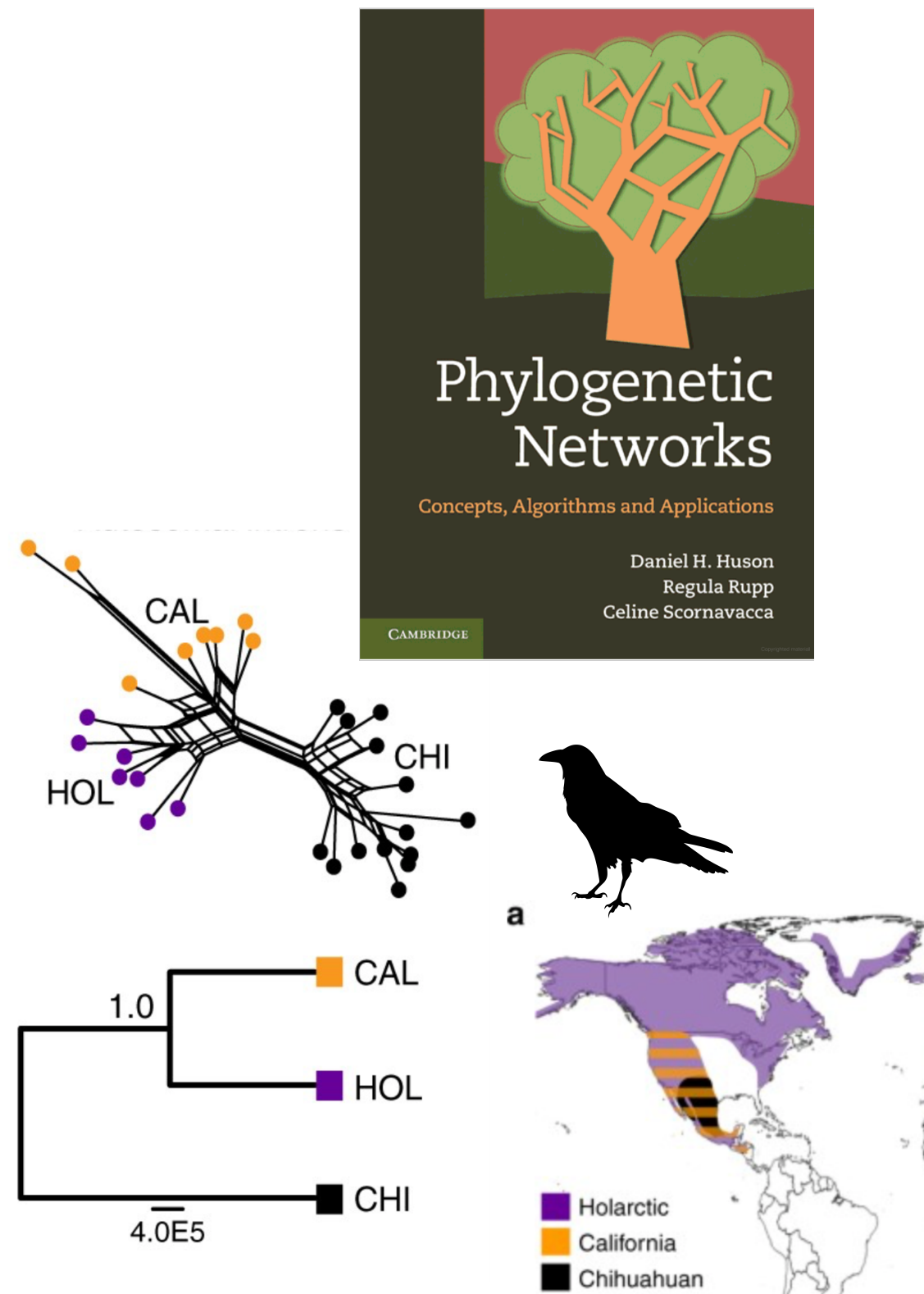
Networks

“The tree of life is eternally branching, and never doing anything fundamental but branching, except for the dying-away of branches. The tree of human history, on the contrary, is constantly branching and at the same time having its branches grow together again. Its plan is therefore much more complex and difficult to trace.”

Kroeber 1931:86

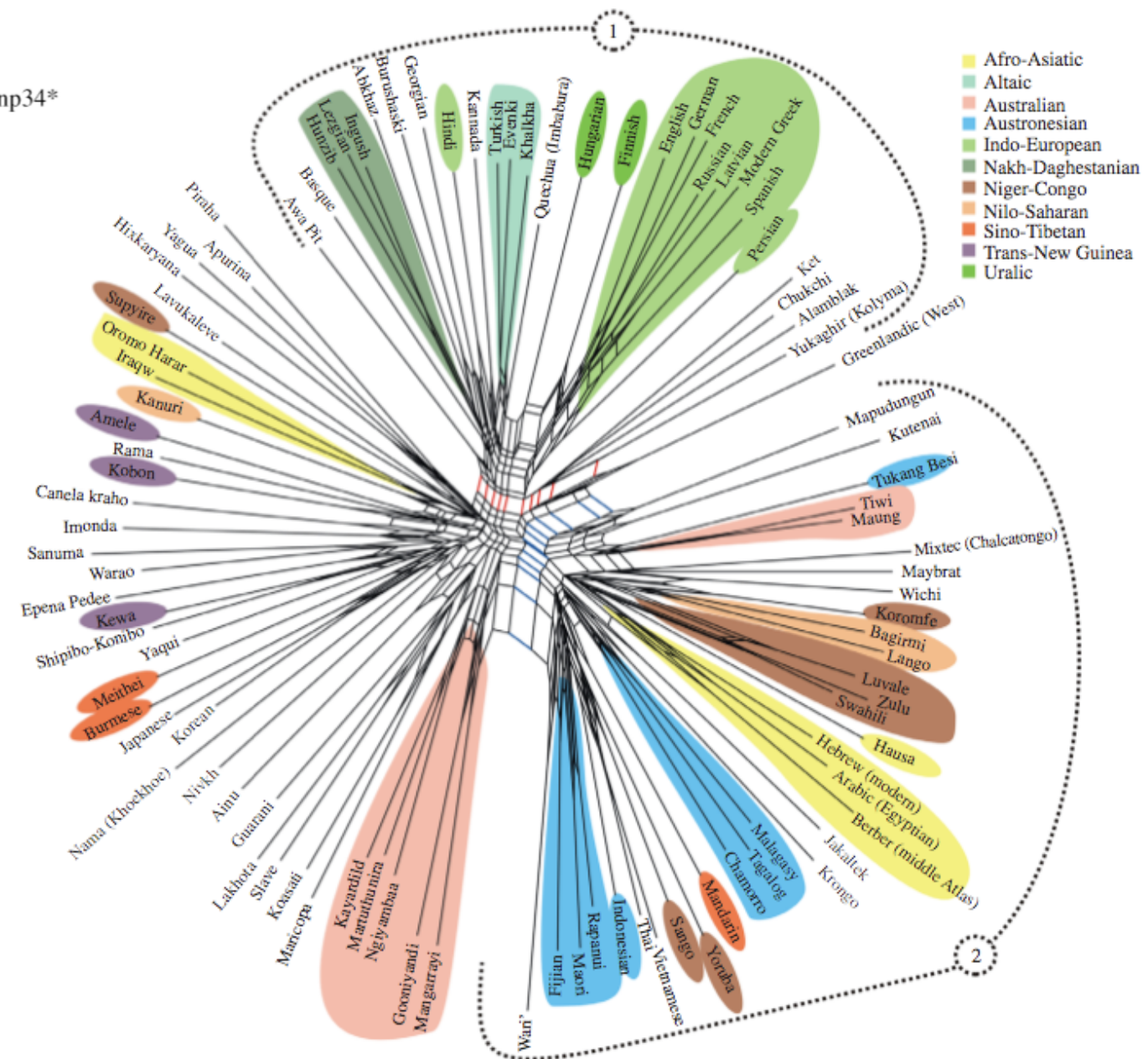
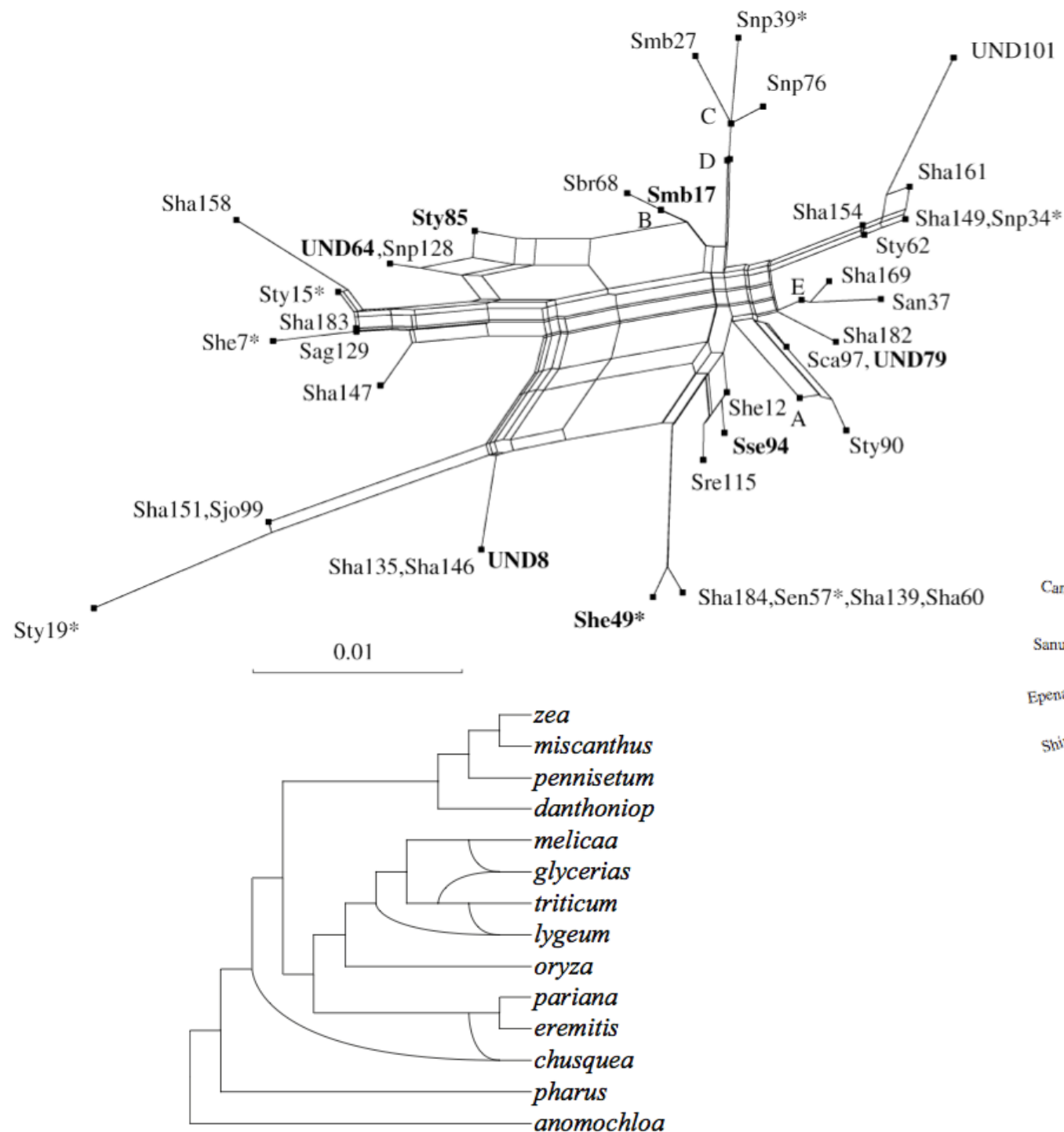
Representing complex evolutionary relationships

- A single tree may not adequately model:
 - Gene recombination, lateral gene transfer, hybridisation
 - Contact, borrowing, elite dominance, dialect chains...
- Different elements may have different histories
 - How do we reconcile alternative evolutionary histories?



Networks

represent complex evolutionary relationships



Huson D et al 2010 *Phylogenetic Networks*

Greenhill, SJ et al 2010 The shape and tempo of language evolution. *Proceedings B*, 277(1693), 2443-2450.

Networks represent complex relationships

“Splits” methods partition taxa into groups

Data matrix:

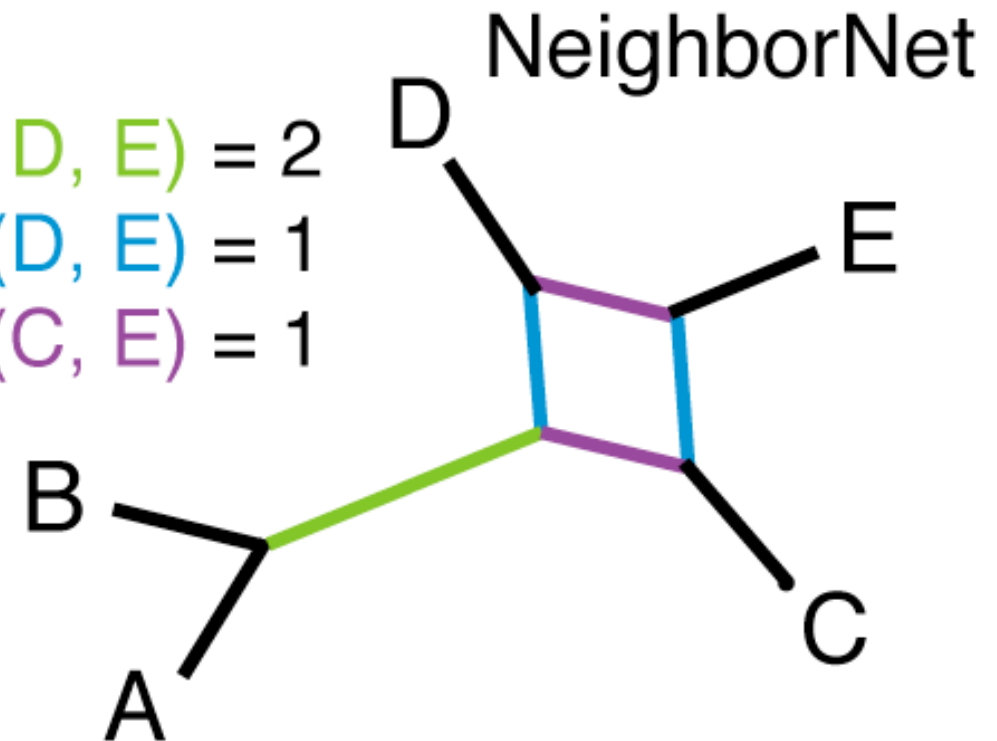
	Cogset 1	Cogset 2	Cogset 3	Cogset 4
Lang A	1	0	1	1
Lang B	1	0	1	1
Lang C	0	1	1	0
Lang D	0	1	0	1
Lang E	0	1	0	0

Splits:

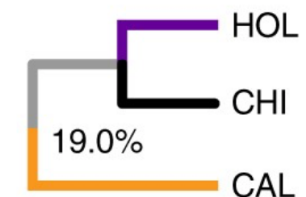
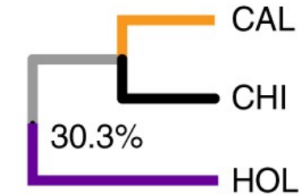
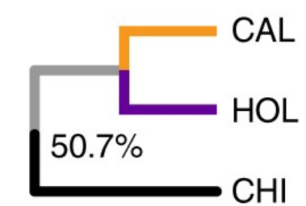
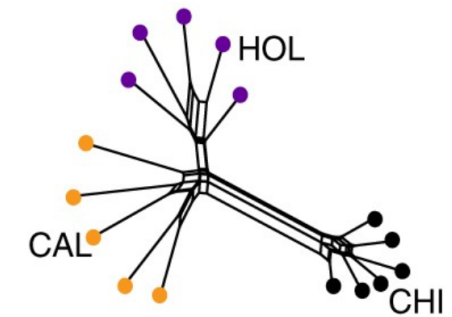
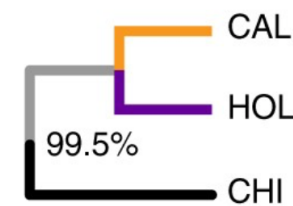
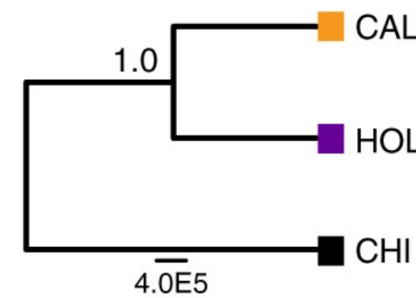
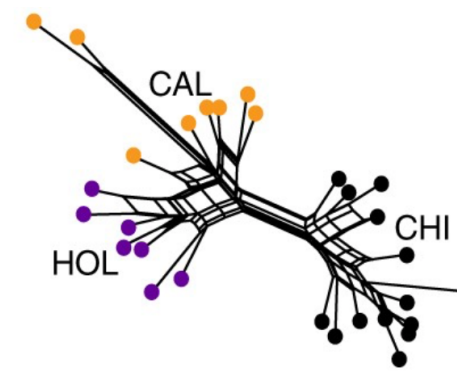
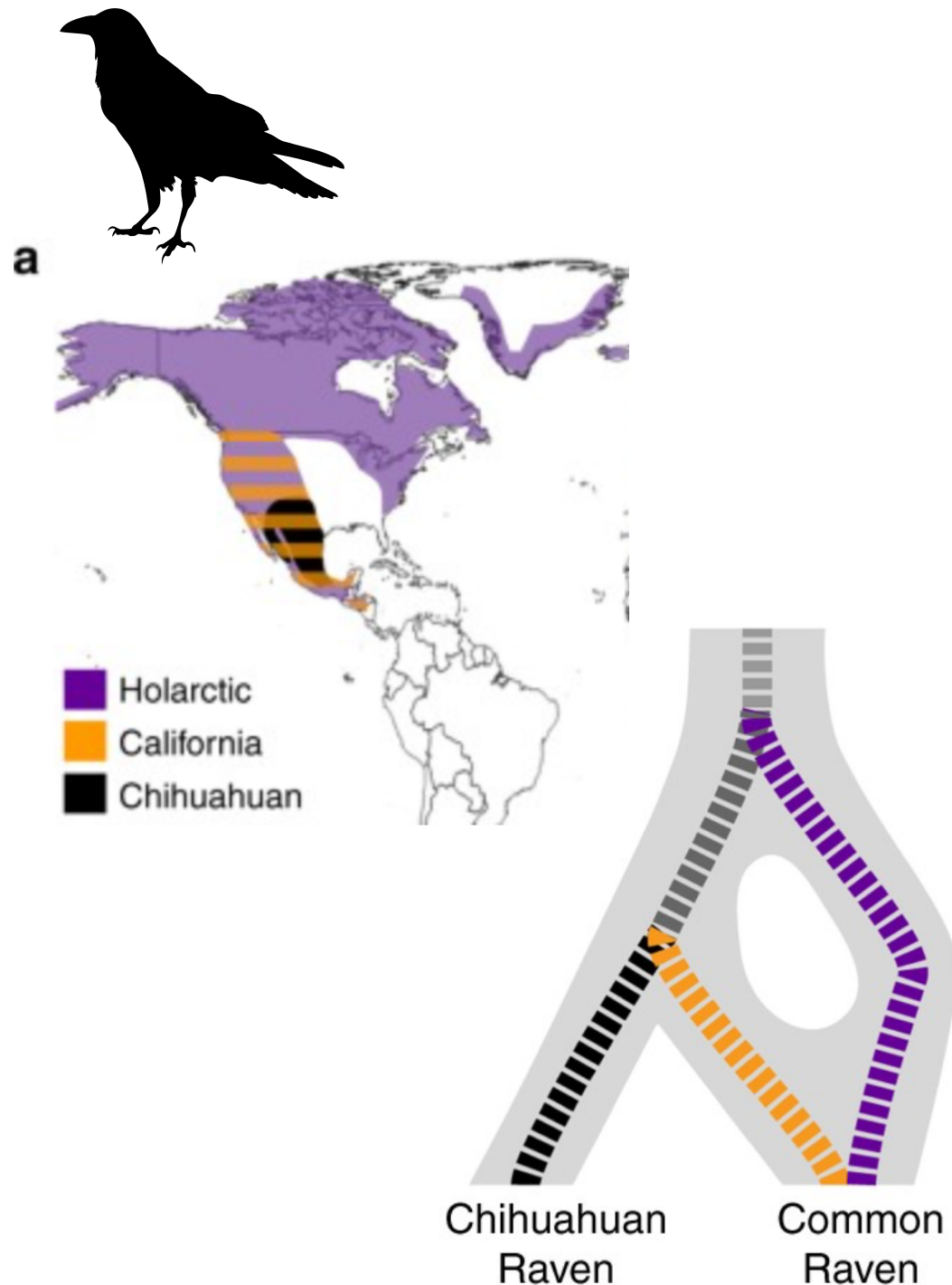
$$(A, B) \mid (C, D, E) = 2$$

$$(A, B, C) \mid (D, E) = 1$$

$$(A, B, D) \mid (C, E) = 1$$



Networks represent complex relationships



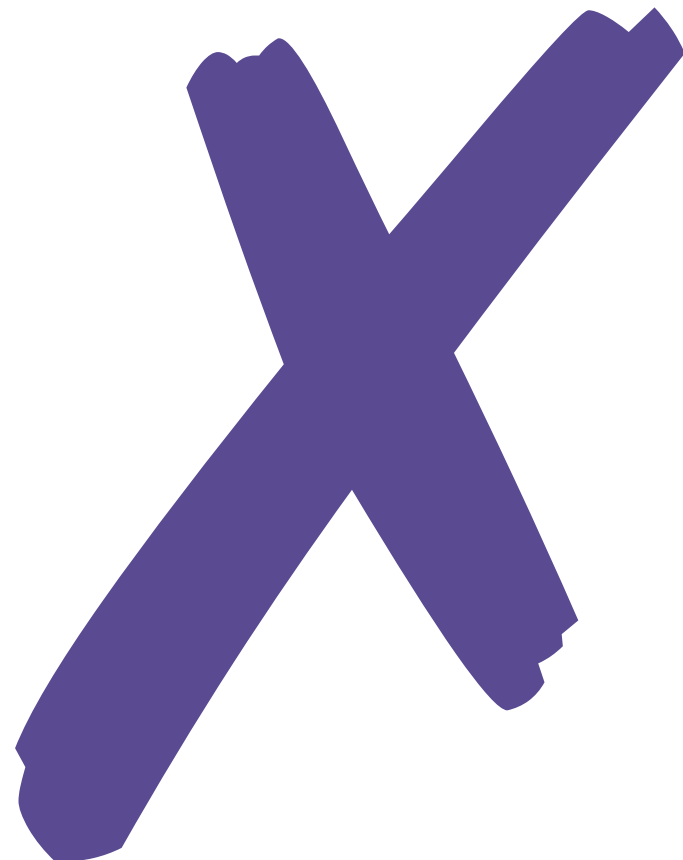
The good and not-so-good of networks

Represent non-tree-like signal in the data

Excellent intuitive exploratory tool

Some ability to quantify tree-like-ness

Can identify taxa/characters that cause conflict



NOT an explicit history of character history

Many algorithms lose information

Interpretation is somewhat arbitrary

No comparative methods available (yet)

References and credits

**All images from academic publications used in an educational setting for non-commercial fair use.
Unless listed, other images/figures by Fiona Jordan and licensed under [CC BY-NC-SA 4.0](#)**

- 1) Images: The Noun Project. *Gene* by Adrien Coquet. *Speaking* by Unknown (CC BY 3.0)
- 3) Image: Cultural Evolution Society poster by Annette Guenzel / Mesoudi, A. (2011). *Cultural evolution: How Darwinian theory can explain human culture and synthesize the social sciences*. University of Chicago Press.
- 7) Images: Greenhill S.J. (2023). Language Phylogenies. *The Oxford Handbook of Cultural Evolution*, C61P1-C61P248.
- 8) Images: Gray, R. D., Drummond, A. J., & Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 323(5913), 479-483. / Bower, C., & Atkinson, Q. (2012). Computational phylogenetics and the internal structure of Pama-Nyungan. *Language*, 817-845. / Kolipakam, V., et al. (2018). A Bayesian phylogenetic study of the Dravidian language family. *Royal Society Open Science*, 5(3), 171504. / Walker, R. S., & Ribeiro, L. A. (2011). Bayesian phylogeography of the Arawak expansion in lowland South America. *Proceedings B: Biological Sciences*, 278(1718), 2562-2567. / Grollemund, R., et al. (2015). Bantu expansion shows that habitat alters the route and pace of human dispersals. *PNAS*, 112(43), 13296-13301. / Bouckaert, R. et al. (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, 337(6097), 957-960.
- 11) Drummond, A. J., & Bouckaert, R. R. (2015). *Bayesian evolutionary analysis with BEAST*. Cambridge University Press.
- 12) Image: Piet Mondrian, 'Gray Tree' in Public Domain via Wikimedia Commons
- 13) Images: Huson, D. H., Rupp, R., & Scornavacca, C. (2010). *Phylogenetic Networks: concepts, algorithms and applications*. Cambridge U Press. / Kearns, A. M., et al (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, 9(1), 906.
- 14) Images: Huson, D. H., Rupp, R., & Scornavacca, C. (2010). *Phylogenetic Networks: concepts, algorithms and applications*. Cambridge U Press. / Greenhill, S. J., et al. (2010). The shape and tempo of language evolution. *Proceedings B: Biological Sciences*, 277(1693), 2443-2450.
- 15) Adapted from Bryant, D., & Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21(2), 255-265.
- 16) Images: Kearns, A. M., et al (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, 9(1), 906.