Biology today

NCBI

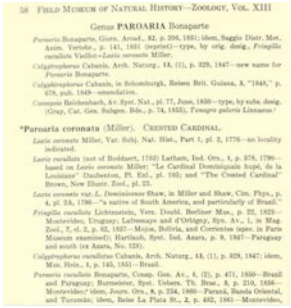
TreeBASE

Tree of Life

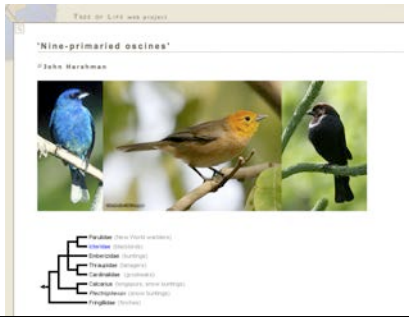
eol Encyclopedia of Life

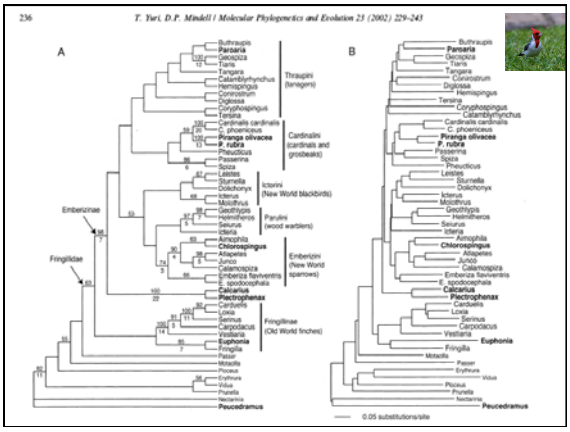
Encyclopedia of life

eol Encyclopedia of Life



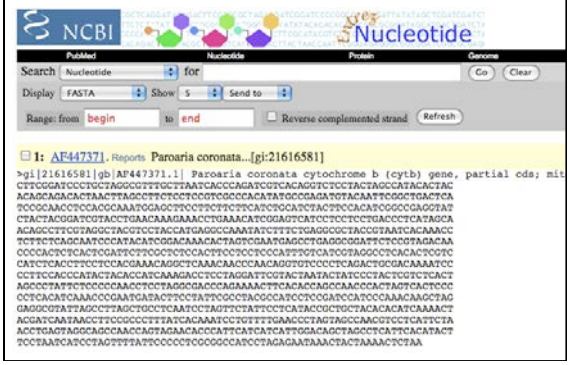
Tree of Life



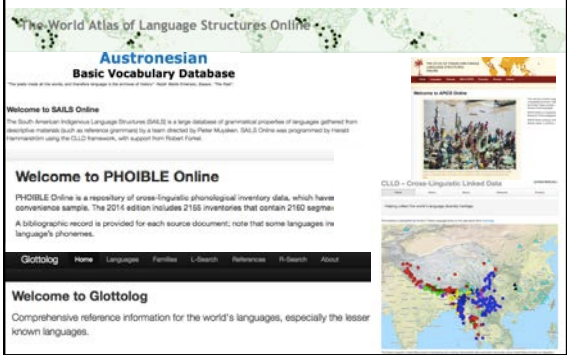


Genbank taxonomy browser

Genbank - cytochrome b sequence



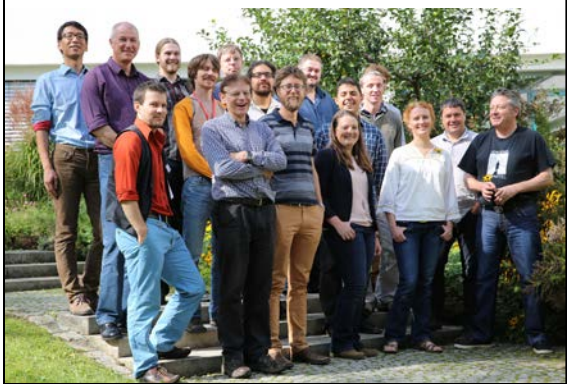
Linguistics today

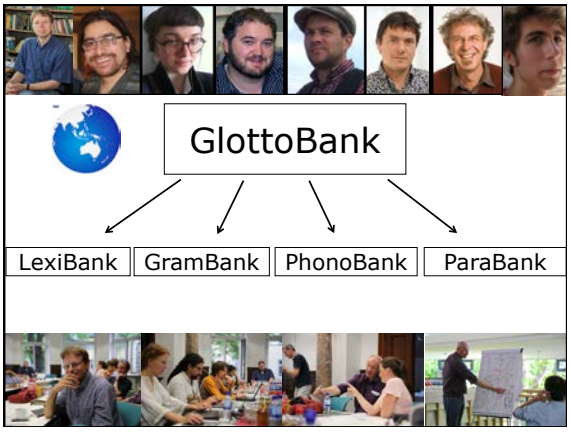


Think big – scaling up…

1. Big data
2. Big methods
3. Big questions
4. Big teams

GlottoBank





D-PLACE

a global database of cultural variation linked to language trees and ecological data

Following the Comrie model

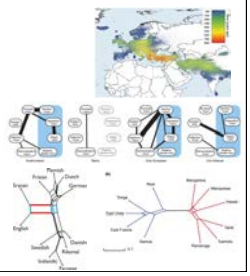
The data deluge requires computational tools

Think big – scaling up…

1. Data
2. Methods

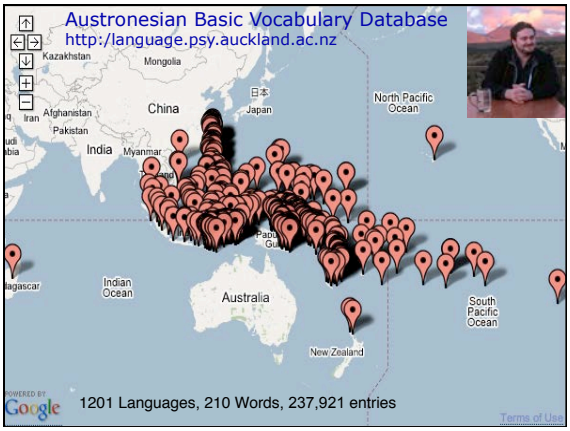
What value can computational methods add?

- Dating language divergences
- Phylogeography
- Functional dependencies
- Networks



"linguists don't do dates"

April & Robert McMahon (2006)



Cognacy


Bob Blust

John Lynch


Jeff Marck


Malcolm Ross


Laurent Sagart

Cognate coding

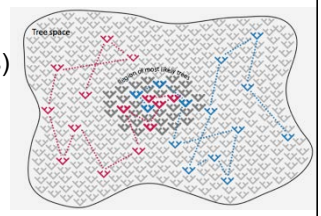
Language	*"father"*	*cognacy*	*binary*
Paiwan	tjama	1	1 0
Itbayaten	qamaq	1	1 0
Mangarrai	ema	1	1 0
Motu	tama-na	1	1 0
Fijian (Bau)	tama-na	1	1 0
Tongan	tama i	1	1 0
Rarotongan	metua	2	0 1
Maori	matua	2	0 1

Data

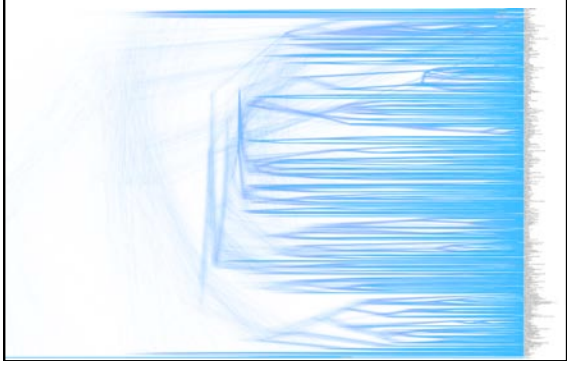
- 400 well-attested languages
  - No creoles, obvious borrowing removed
- Outgroup
  - Old Chinese (controversial)
  - Buyang (less controversial)
- Binary Coding
  - presence/absence of cognates
  - 34,440 cognate sets
  - Covarion model

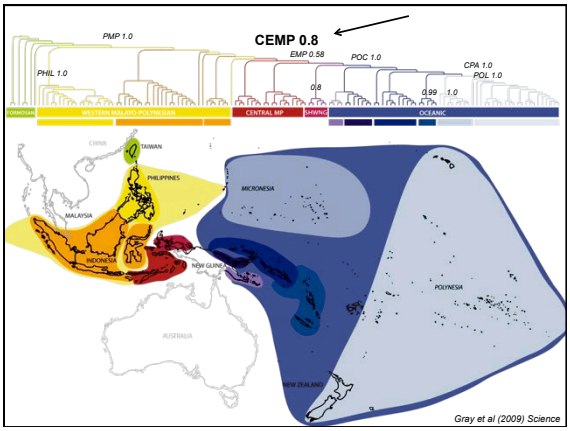
Bayesian Phylogenetic Inference

1. Data
2. Model (and priors)
3. Tree search
4. Dating (without a strict clock)

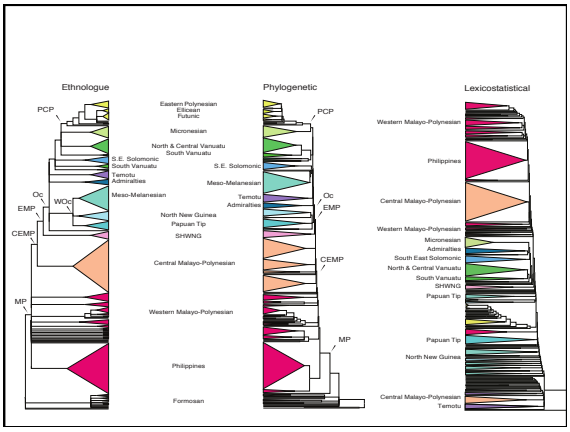


Uncertainty in tree estimation

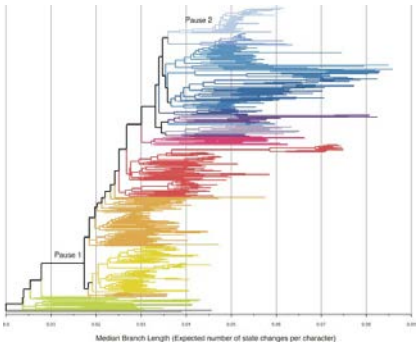




Gray et al (2009) Science

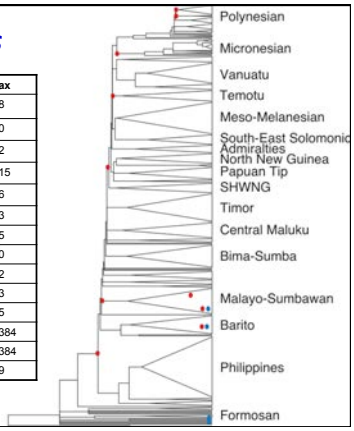


Austronesian phylogram

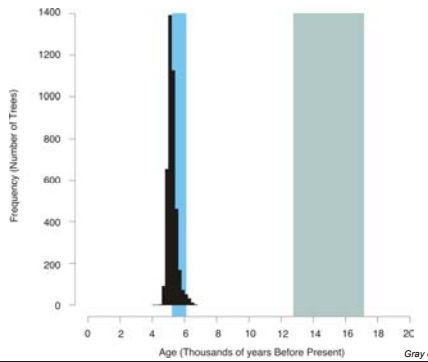


Calibrations

Constraint	Min	Max
East Polynesian	1.15	1.8
Tuvalu/Tokelau	1.0	2.0
Micronesian	1.9	2.2
Proto-Reefs Santa Cruz	3.0	3.15
Proto-Oceanic	3.2	3.6
Proto-Javanese	1.1	1.3
Proto-Chamic	1.8	2.5
Malayic-Chamic	2.0	3.0
Old Javanese	0.7	1.2
Proto-Malagasy	1.1	1.3
Proto-Malayo-Polynesian	3.6	4.5
Favorlong	0.346	0.384
Siraya	0.346	0.384
Old Chinese	2.3	2.9



Prediction 2. Age of Proto-Austronesian

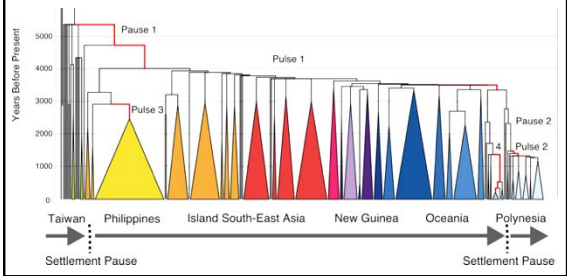


Gray et al (2009) Science

3 predictions:

Sequence, timing, pulses and pauses

Gray, Drummond & Greenhill. 2009. Science, 323, 479-483.



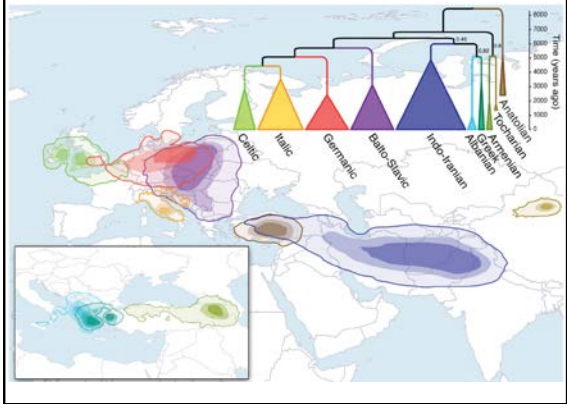
nature

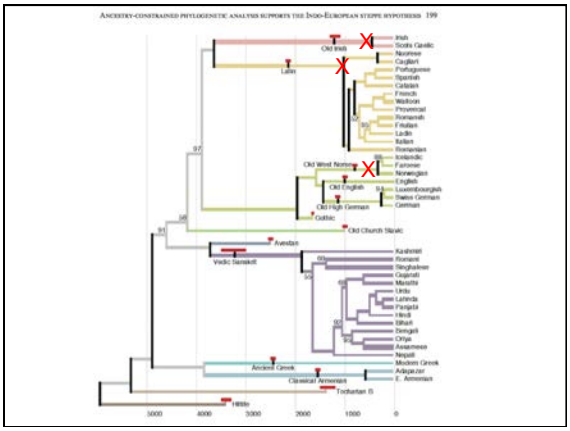
Massive migration from the steppe was a source for Indo-European languages in Europe

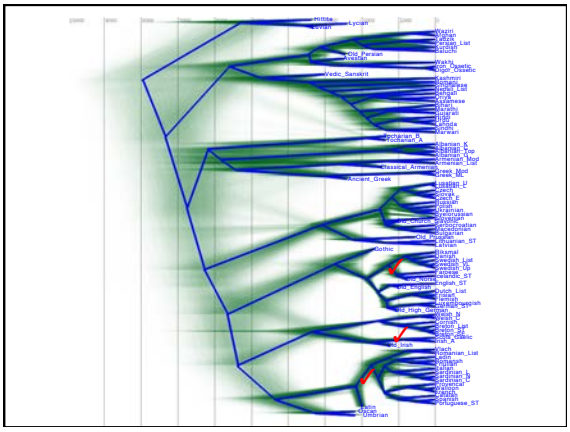
ANCESTRY-CONSTRAINED PHYLOGENETIC ANALYSIS SUPPORTS THE INDO-EUROPEAN STEPPE HYPOTHESIS

WILL CHANG, University of California, Berkeley
CHUNDRA CATHCART, University of California, Berkeley
DAVID HALL, University of California, Berkeley
ANDREW GARRETT, University of California, Berkeley

Bouckaert et al (2012) Science.







Think big – scaling up…

1. Data
2. Methods
3. Questions

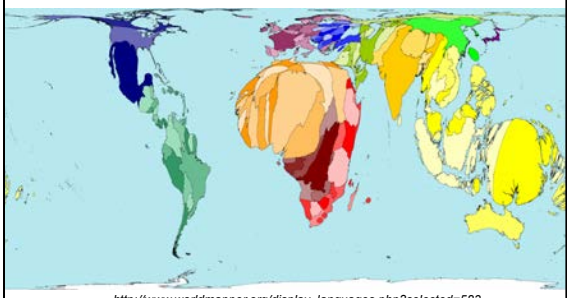
What are the Hilbert problems in linguistics?

David Hilbert

Martin Hilpert

https://www.youtube.com/watch?v=X4OaN39sNAI&feature=youtu.be

Explain this!

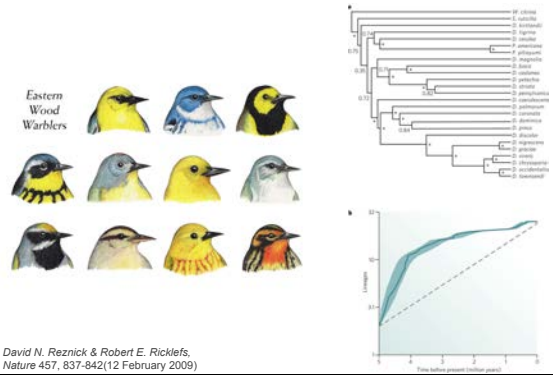


http://www.worldmapper.org/display_languages.php?selected=583

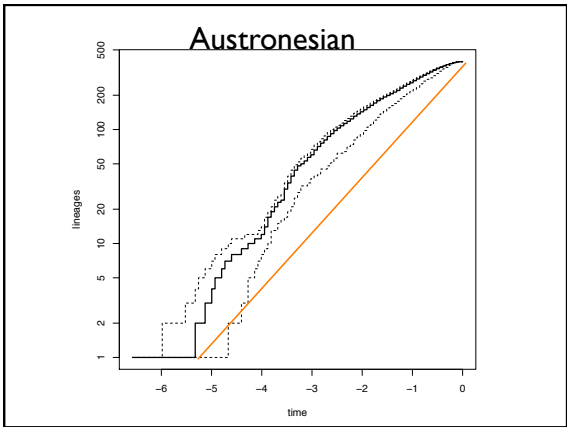
Some suggestions…

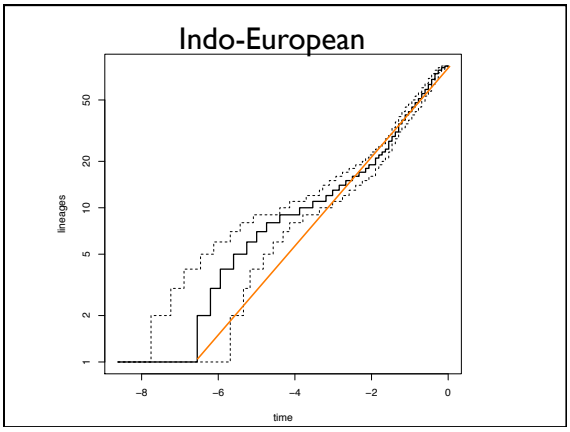
1. Why are there approximately 7000 languages?
2. Why is language diversity distributed so patchily?
3. What drives the evolution of linguistic disparity?
4. When did spoken language evolve?
5. How far back can we push the time barrier for detecting language relationships?

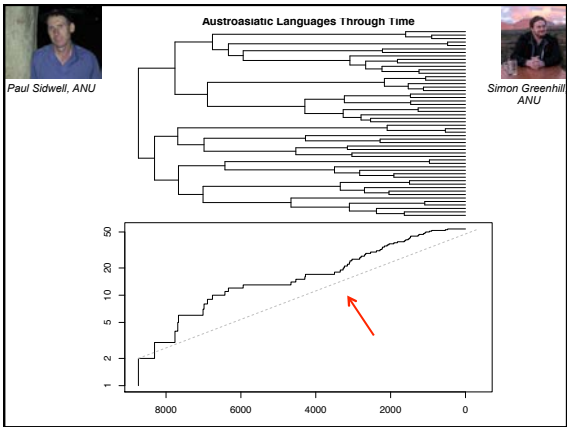
Lineage through time plots

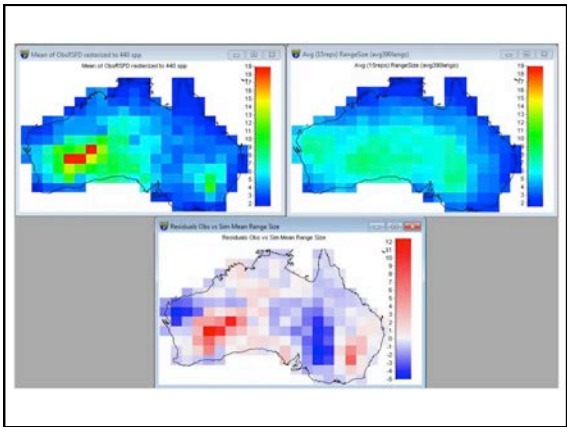


David N. Reznick & Robert E. Ricklefs,
Nature 457, 837-842(12 February 2009)









Think big – scaling up…

1. Data
2. Methods
3. Questions
4. Teams

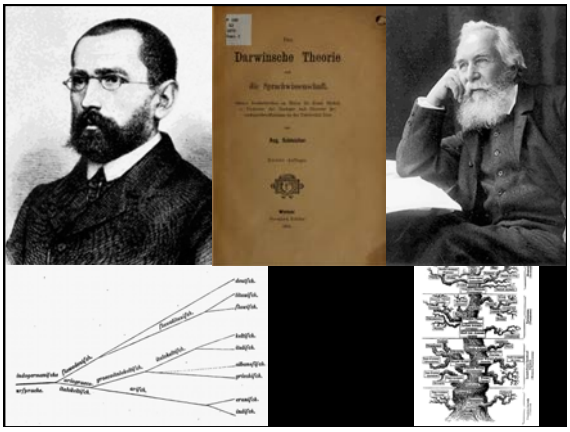
nature International weekly journal of science

Massive migration from the steppe was a source for Indo-European languages in Europe

Wolfgang Haak, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, Susanne Nordenfelt, Eadaoin Harney, Kristin Stewardson, Qiaomei Fu, Alissa Mittnik, Eszter Bánffy, Christos Economou, Michael Francken, Susanne Friederich, Rafael Garrido Pena, Fredrik Hallgren, Valery Khartanovich, Aleksandr Khokhlov, Michael Kunst, Pavel Kuznetsov, Harald Meller, Oleg Mochalov, Vayacheslav Moiseyev + *et al.*

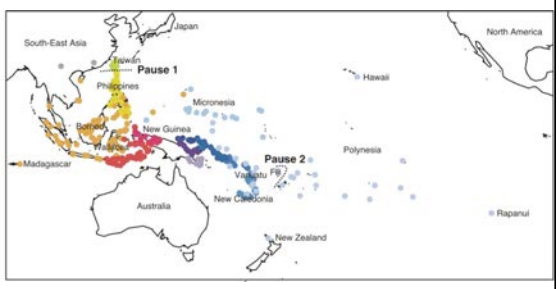
Nature (2015) | doi:10.1038/nature14317
Received 29 December 2014 | Accepted 12 February 2015 | Published online 02 March 2015

Big interdisciplinary teams



Vanuatu – the Galapagos of language evolution

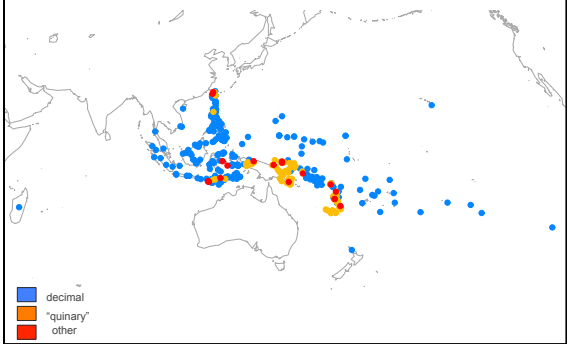
Why do "Remote Melanesians" not look like Polynesians?



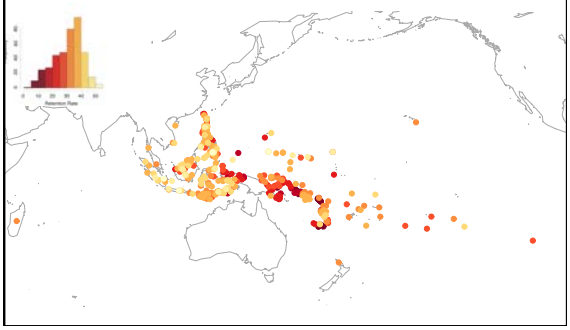
Blust (2005, 2008)

1. Phenotypic differences in “Remote Melanesia”
2. Cultural similarities
3. Language typology - serial verb constructions
4. Loss of decimal system - switch to various quinary systems
5. Large amount of sound change

Distribution of numeral systems



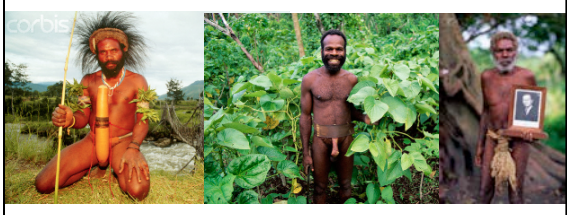
Circumstantial evidence for extended contact
Distribution of retention rates



Variation in rates of sound change

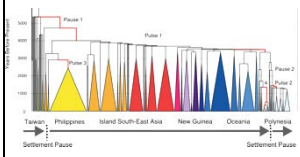
Number	*Paiwan*	*Cebuano*	*Maori*	*Nengone*
1	ita	usa	tahi	sa
2	dusa	duhá	rua	rewè
3	tjelu	tulo	toru	tini
4	sepatj	upát	whaa	ece
5	rimáʔ	lima	rima	sɛduŋ

Chimbu Valley, New Guinea *PNG highlands* *Vanuatu*



Mek warrior, Irian Jaya *Pentecost Island, Vanuatu* *Tanna, Vanuatu*

Vanuatu – the Galapagos of language evolution



Max Planck Institute for the Science of Human History

Leipzig linguistics library
Jena prize for historical linguistics