A Computational-phylogenetic Classification of Tupí-Guaraní and its Geographical Spread

Zachary O'Hagan
University of California, Berkeley

∼

Language Variation and Change
University of Chicago
November 20, 2014

Tupí-Guaraní Comparative Project Team

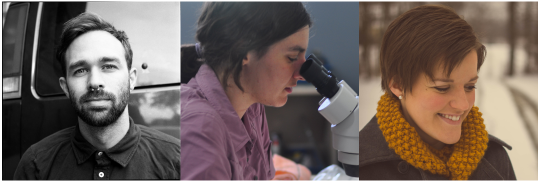


Figure 1: Bartolomei, Chousou-Polydouri, Clem

Figure 2: Donnelly, Michael, O'Hagan

Introduction

- We present a classification of Tupí-Guaraní using lexical data and computational phylogenetic methods developed in evolutionary biology and more recently applied to linguistic phylogenies
  - See Bouckaert et al. (2012); Bowern and Atkinson (2012); Forster and Toth (2003); Gray and Atkinson (2003); Gray et al. (2009); Greenhill and Gray (2005, 2009); Greenhill et al. (2010); Nakhleh et al. (2005); Ringe et al. (2002); Warnow et al. (2004)
- This new classification is based on:
  - A 596-item comparative lexical dataset for
  - 30 TG languages and 2 non-TG Tupí languages
  - organized into 4187 cognate sets
- This classification complements previous proposals based on sound change

Organization

- Introduction
  - Brief overview of previous classifications of Tupian and TG
- Data and Methodology
  - Development of the TG Comparative Lexical Database
  - Cognate set character coding
  - Overview of phylogenetic methods
- Results
  - Presentation of new classification
  - Comparison with Rodrigues and Cabral (2002)
- Geographical Spread
  - Center of Gravity Model
  - Homeland
  - Migration

Tupí Classification

- Universal agreement among specialists that TG forms a subgroup within the Tupí stock (Campbell 1997; Jensen 1999; Kaufman 1994, 2007; Rodrigues 1986, 1999; Rodrigues and Cabral 2012)
- Consensus that Awetí and Mawé are – in that order – the Tupí languages most closely related to TG (Corrêa da Silva 2007, 2010; Drude 2006, 2011; Kamaiurá 2012; Rodrigues and Dietrich 1997)
- First phylogenetic exploration of Tupí: Galúcio et al. (2013)

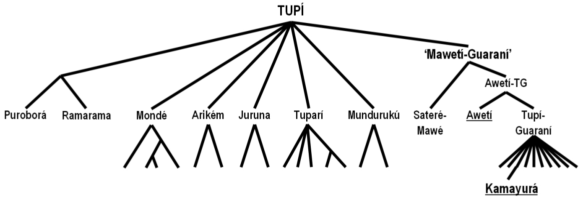


Figure 3: Tupí Classification (Drude 2011)

Proto-Tupí-Guaraní, Classification, and *Subconjuntos*

- Miriam Lemle (1971) reconstructs 221 proto-forms and proposes the following classification of Tupí-Guaraní based on shared innovation

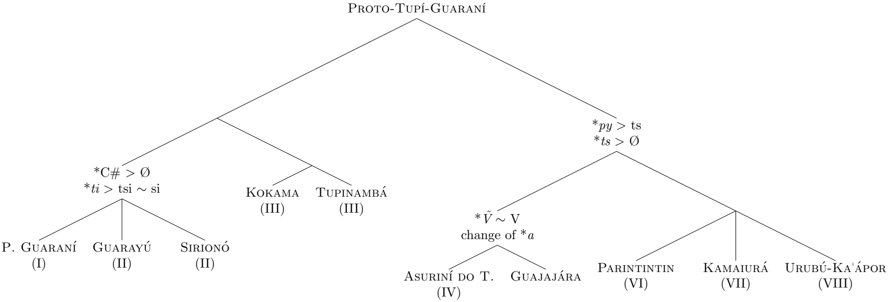


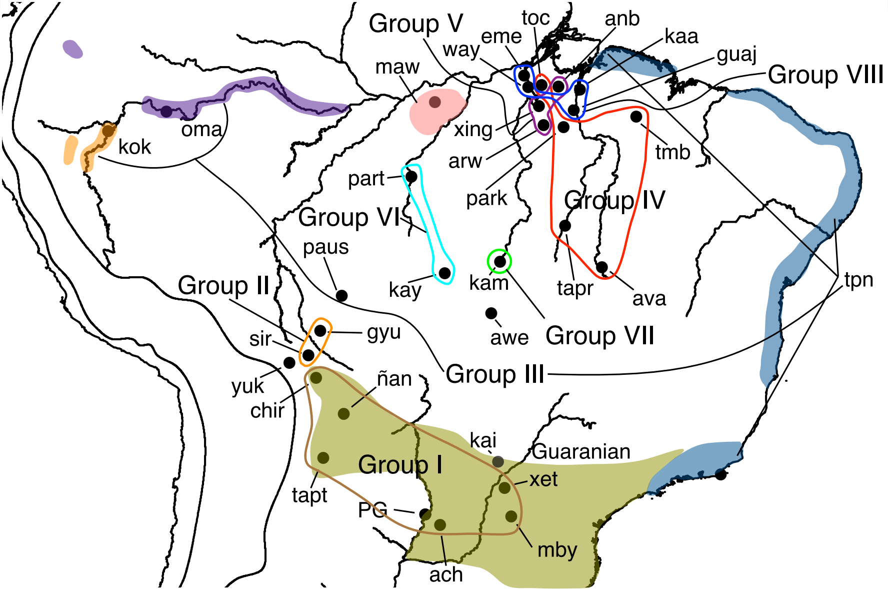
Figure 1: Tupí-Guaraní Subgrouping (Lemle 1971)

Proto-Tupí-Guaraní, Classification, and *Subconjuntos*

- The most influential grouping of 42 Tupí-Guaraní varieties are the eight *subconjuntos* proposed by Rodrigues (1984/1985)
  - Each *subconjunto* is defined by a particular **set** of sound changes attested in every member language with respect to the proto-language
  - Sound changes are often shared by several *subconjuntos*
  - Most sound changes are common and natural (e.g., palatalization, lenition), casting doubt on their utility for classification
- Rodrigues emphasizes that this classification is not a genetic one:
  - *Os subconjuntos acima delineados constituem não propriamente uma classificação interna da família ... mas antes um ensaio de discriminação de seções dessa família caracterizadas pelo compartilhamento de algumas propriedades lingüísticas, as quais podem servir para diagnosticar o desmembramento de todo o conjunto de línguas Tupi-Guaraní...* (ibid.:48)

PTG	I	II	III	IV	V	VI	VII	VIII
*C#	*C# > Ø	*C# > Ø						*C# > Ø (?)
**tʃ*	**tʃ*; **tʃ* > ts ∼ s	**tʃ* > ts ∼ s	**tʃ* > ts ∼ s	**tʃ* > h	**tʃ* > h ∼ Ø	**tʃ* > h	**tʃ* > h ∼ Ø	**tʃ* > h ∼ Ø
**ts*	**ts* > h ∼ Ø	**ts* > ts ∼ s	**ts* > ts ∼ s	**ts* > h	**ts* > h ∼ Ø	**ts* > h	**ts* > h ∼ Ø	**ts* > h ∼ Ø
**pw*	**pw* > kw ∼ k	**pw* > kw ∼ k		**pw* > kw ∼ k	**pw* > t	**pw* > kw ∼ fw ∼ f	**pw* > hw ∼ h	**pw* > kw
**pj*	**pj* > tʃ ∼ ʃ			**pj* > tʃ∼ ts	**pj* > s		**pj* > ts	**pj* > s
**j*				**j* > tʃ ∼ ts ∼ s ∼ z	**j* > dj			
STRESS		*ό# > όσ#						
LANGS.	Old Guaraní Mbyá Xetá Ñandeva Kaiowá P. Guaraní Guayakí (Aché) Tapieté Chiriguano (Avá) Izoceño (Chané)	Guarayú Sirionó Jorá	Tupinambá Tupí Austral Nheengatú Kokama Kokamilla Omagua	Tapirapé Avá-Canoeiro Asuriní do T. Suruí Parakanã Guajajára Tembé	Kayabí Asuriní do X. Araweté	Parintintin Tupí-Kawahib Apiaká	Kamaiurá	Takunhapé Wayampí Wayampipukú Emérillon Amanayé Anambé Turiwára Guajá Urubú-Ka'ápor

Figure 4: Tupí-Guaraní *Subconjuntos* (Rodrigues 1984/1985)


Group V
toc
anb
eme
kaa
way
guaj
maw
Group VIII
oma
kok
xing
arw
tmb
part
park
Group IV
Group VI
paus
kay
kam
tapr
ava
Group II
awe
Group VII
sir
gyu
yuk
Group III
ñan
chir
kai
Guaranian
Group I
xet
tapt
PG
mby
ach
tpn

Proto-Tupí-Guaraní, Classification, and *Subconjuntos*

- Mello (2000, 2002), using a larger dataset, a greater number of sound changes, and a methodology essentially equivalent to Rodrigues', proposes **nine** *subconjuntos*
  - Unlike Lemle, neither Rodrigues nor Mello propose any higher-level structure between *subconjuntos*
- Rodrigues and Cabral (2002) revise Rodrigues' previous *subconjuntos*, referring to them as a 'nova classificação' and a 'classificação interna'
  - They propose higher-level structure between *subconjuntos*, but without explicitly delineating the shared innovations of each node
- Phylogenetic work based on a large lexical dataset can usefully complement this body of prior work

Data Harvesting

- The lexical database developed for this project includes:
  - 596-item list of crosslinguistically and areally appropriate meanings in
  - 30 TG and 2 non-TG Tupí languages (Mawé and Awetí)
- Data was harvested by Keith Bartolomei, Natalia Chousou-Polydouri, Erin Donnelly, Lev Michael, Sérgio Meira, Zachary O'Hagan, Mike Roberts, and Vivian Wauters from:
  - dictionaries
  - phonological descriptions
  - grammatical descriptions
  - text collections
- Average coverage = 71%

Lexical Coverage

Aché	85%	Ñandeva	20%
Anambé	31%	Omagua	89%
Araweté	55%	Parakanã	75%
Avá-Canoeiro	51%	Paraguayan Guarani	94%
Awetí	76%	Parintintin	85%
Chiriguano	80%	Pauserna	58%
Emerillon	77%	Siriono	82%
Guajá	45%	Tapiete	84%
Guarayu	86%	Tapirapé	69%
Ka'apor	83%	Tembé	98%
Kaiowá	39%	Tocantins Asuriní	83%
Kamaiurá	75%	Tupinambá	94%
Kayabí	59%	Wayampí	89%
Kokama	89%	Xetá	33%
Mawé	80%	Xingú Asuriní	50%
Mbyá	83%	Yuki	80%

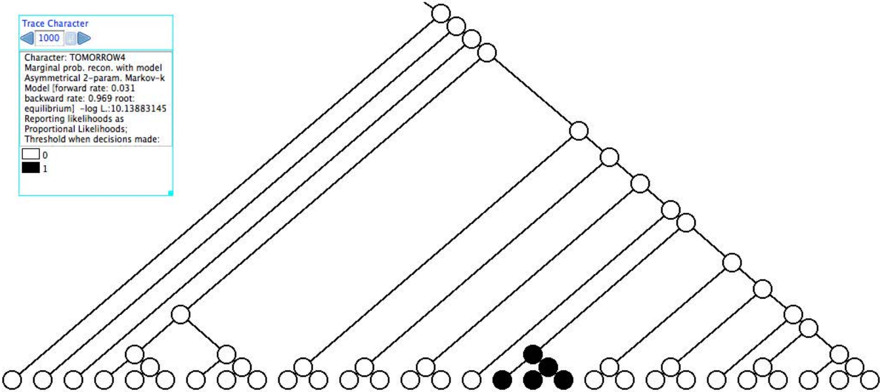
Cognate Set Construction

- Lexical list items were harvested and organized into cognate sets
- Cognacy = forms descended from a single proto-form
  - This is unlike purported "cognate" sets typically employed in phylogenetics, which additionally require identical meanings
  - This approach ignores known cognates with non-identical meanings
- Thus we consider items cognate that have undergone semantic shift
  - These include forms whose meanings shifted both to meanings included in the initial comparative list, as well as to meanings that moved outside that set of meanings
- Cognate sets were constructed collaboratively and were examined and re-evaluated in many iterations
  - Basic sound correspondences evident in the data were taken into consideration in set construction
- Intrafamily loans were in general not identified; interfamily loans were coded as independent singleton characters

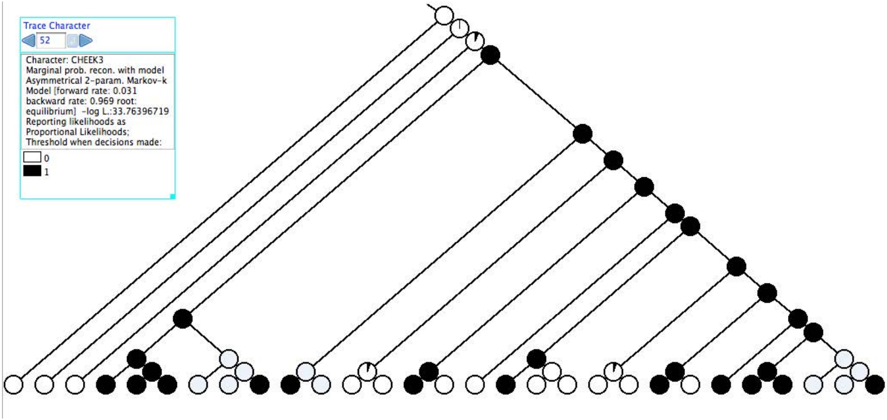
Computational Phylogenetic Methods

- Phylogenetic methods aim to find evolutionary trees that account for the distribution of traits (= 'characters') across languages
  - A character is a cognate in a particular cognate set, and it may have two states, present or absent
- Bayesian methods (see below) estimate the posterior probability of our model – i.e., tree topology, branch lengths, and rate parameters – given the data
- The phylogenetic algorithm samples the space of possible trees in proportion to their posterior probability
- The resulting tree is a majority rule consensus tree of this sampling
  - It constitutes a claim about subgrouping
  - Ancestral state reconstruction methods allow the study of the evolution of particular traits over the tree
  - Through 'rooting' the tree it distinguishes innovations from retentions

Character Evolution



Character Evolution



Binary Coding

- For purposes of phylogenetic analysis, the cognate dataset must be reduced to a numerical matrix
- A cognate set consists of binary present-absent (1/0) characters
  - Present: Language exhibits a form belonging to the cognate set
  - Absent: Language lacks a form belonging to the cognate set
- If no form corresponding to a given meaning was found in our sources, then all cognate set characters stemming from that meaning were coded as ‘unknown’ (coded as ‘?’).

Zero Confidence

- Since zeroes (0s) play a non-trivial role in the selection of an optimal tree, it is important that a '0' reflect a true absence
- A cognate was considered absent (coded as '0') for a particular language if all the following conditions were met:
  - The cognate set's associated meaning was represented in the language by a non-cognate word
  - No cognate was found when searching for similar meanings or expected forms
  - No compound word was found in our dataset containing this cognate
- However, the reason why one of these conditions may not have been met can be a mere lexicographic gap in resource material
- We carried out two procedures to reduce these accidental gaps: "Deep Drilling" and Dataset Closure

"Deep Drilling" and Dataset Closure

- **Deep Drilling:** Subsequent to initial cognate set construction, we searched again for cognates based on expected forms (given basic sound correspondences evident in the dataset)
  - For example, cognates to Omagua *yapisara* 'man' do not mean 'man' in other TG languages, but by looking for similar forms we found cognates with meanings similar to 'neighbor', e.g., Tupinambá *apiʃar* <apixara> 'próximo, semelhante' (Lemos Barbosa 1951)
- **Dataset Closure**: The initial comparative list was expanded by ∼65 additional meanings to search systematically for meanings that emerged during the search for potential cognates via deep drilling and which were outside the original set of comparative list meanings
  - For example, forms meaning 'liver' are typically cognate to *pɨʔá*, except in Kamaiurá, where 'liver' is *peré* (Seki 2000), which is cognate to forms meaning 'spleen' in most other TG languages, and so we extended our search to include the meaning 'spleen'

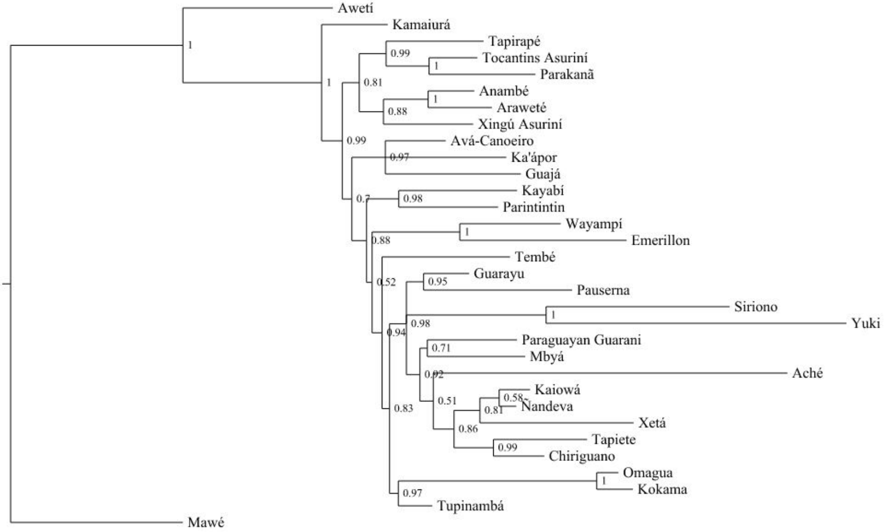
Phylogenetic Methods (MrBayes3.2)

- We used an asymmetric binary model (a.k.a. restriction site model)
  - Different rates of gain and loss for cognates
  - Uniform prior for any cognate loss/gain ratio with ratio a parameter in the search space (reached ∼31:1)
- We allowed for different rates of evolution across cognate sets
  - Gamma distributed rates
  - Gamma shape parameter had a uniform prior distribution for (0,200)
- Phylogenetic Analysis with MrBayes3.2
  - Analysis conducted with four independent runs
  - 10 million generations each, sampled every 1,000 generations
- The chains spend time in (and therefore sample) the posterior distribution of trees proportionally to their posterior probability

Lexicostatistics $\neq$ Phylogenetics

- Lexicostatistical Methods (e.g., NeighborNet, SplitsTree)
  - Lexicostatistical methods do not evaluate evolutionary trees
  - They instead compute a single number – e.g., % of shared cognates – for each pair of languages
  - Languages are then clustered on the basis of overall similarity, **conflating shared innovations and shared retentions**
- Phylogenetic Methods
  - All cognate sets are evaluated individually, and the specific information they bear for subgrouping is preserved
  - Thousands of trees are individually evaluated by optimizing all characters on each one
  - Only shared innovations are considered for subgrouping
  - As a result, phylogenetic methods are not fooled by shared retentions

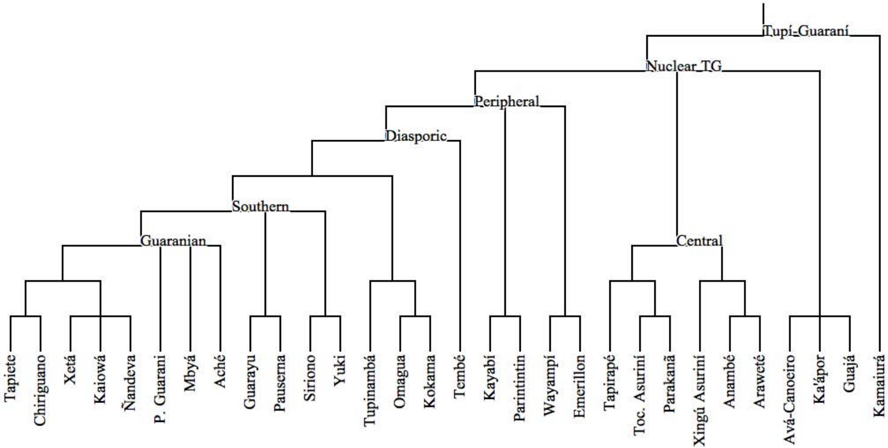
Bayesian Tupí-Guaraní Classification



Posterior Probability Cut-offs

- The ‘raw’ tree depicts all clades (i.e., subgroups) with posterior probabilities $p > 0.50$, meaning that these subgroups show up in over 50% of the posterior sample, i.e., sample of candidate trees
- We adopt a stringent requirement, conflating or ‘collapsing’ any clade of $p < 0.80$ with the superordinate clade to which it belongs
- This is an emerging consensus for well supported clades (Bowern and Atkinson 2012:829)

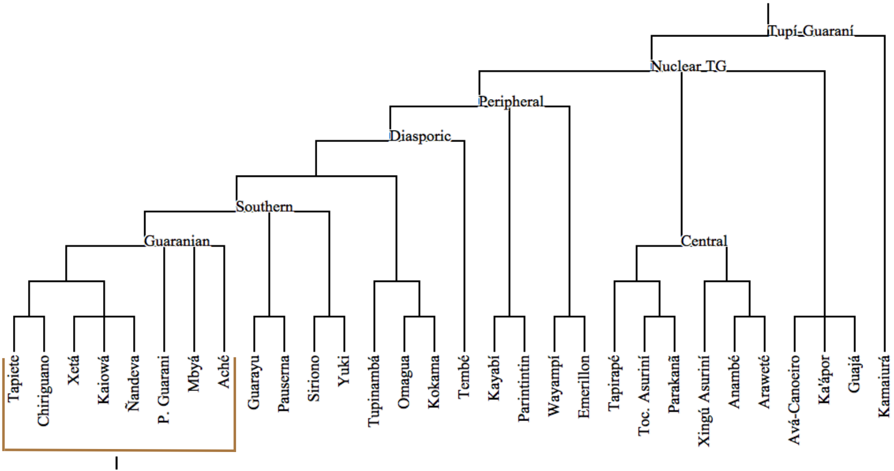
Conservative Tupí-Guaraní Classification



Comparison with Rodrigues and Cabral (2002): Subgroups

- Recovered *Subconjuntos* (Monophylies)
  - I: 98% probability
  - III: 97% probability
  - V: 81% probability
  - VI: 98% probability
  - VII: 100% probability
- Unrecovered *Subconjuntos*
  - II: paraphyletic group within Southern
  - IV: polyphyletic
  - VIII: polyphyletic

Comparison with Rodrigues and Cabral (2002): Subgroups



Monophyly, Paraphyly, and Polyphyly

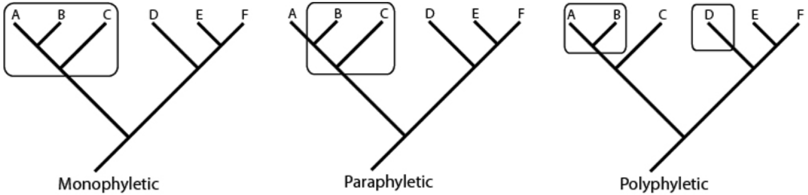
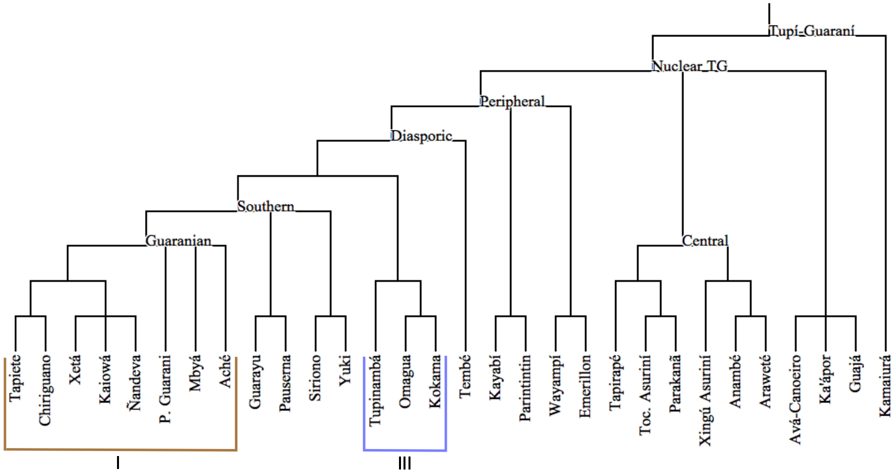
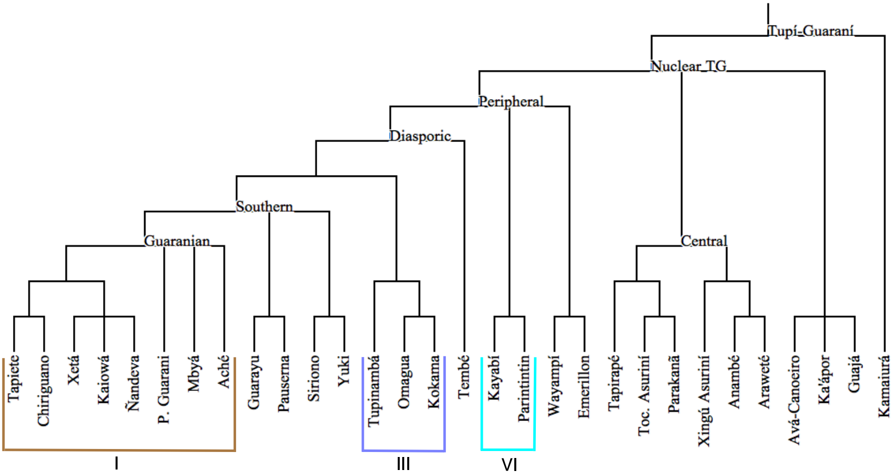


Figure 5: Cladistic Configurations

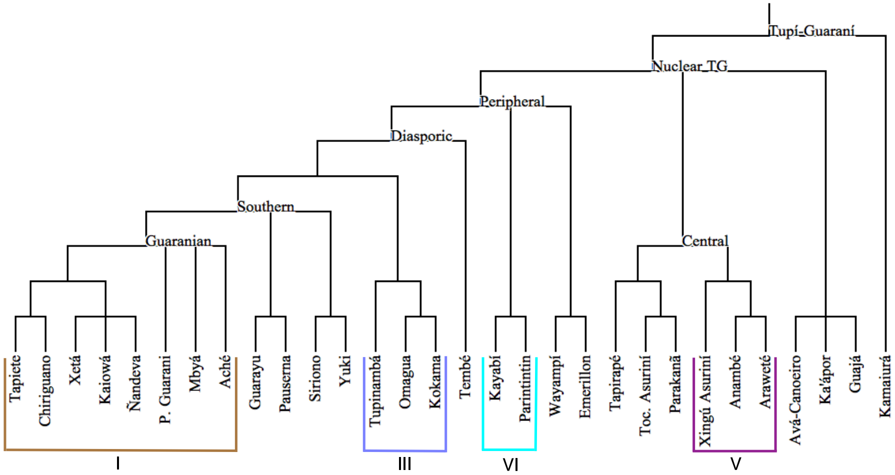
Comparison with Rodrigues and Cabral (2002): Subgroups



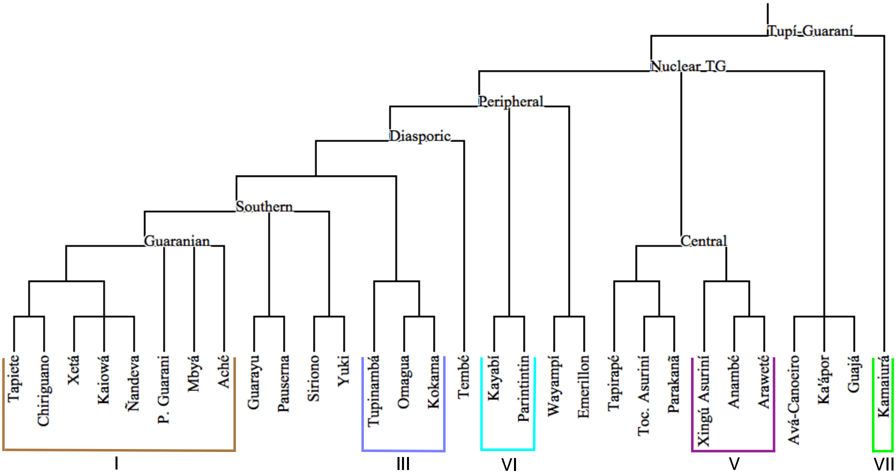
Comparison with Rodrigues and Cabral (2002): Subgroups



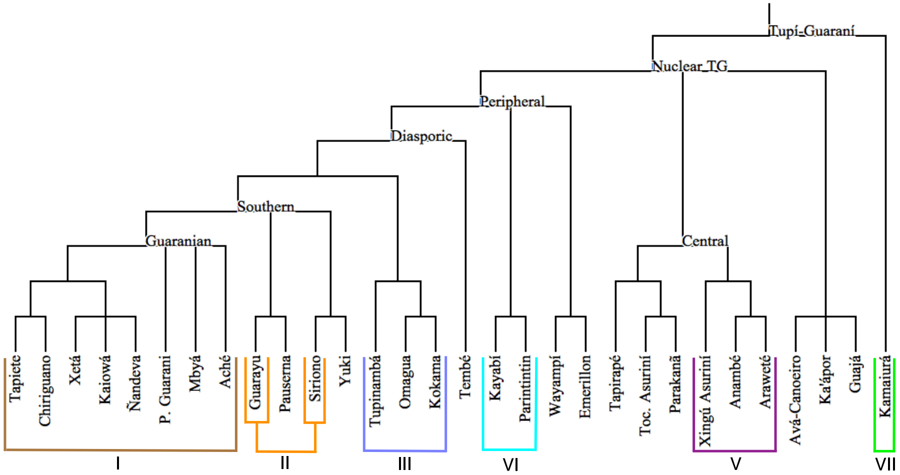
Comparison with Rodrigues and Cabral (2002): Subgroups



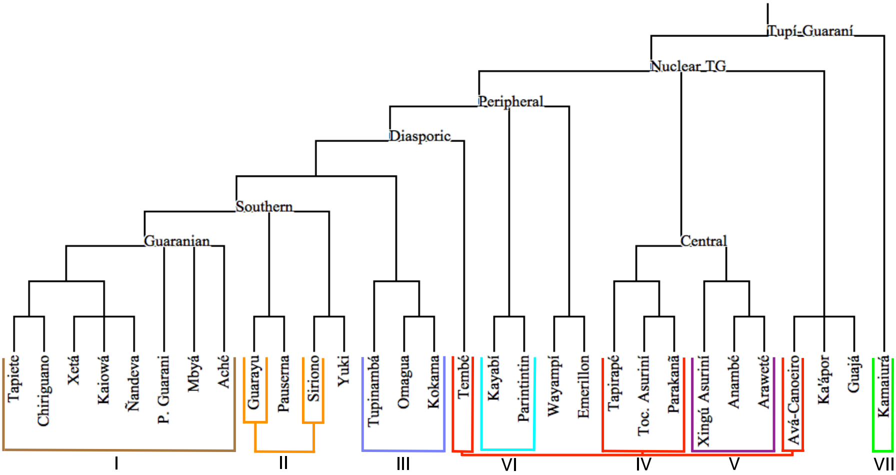
Comparison with Rodrigues and Cabral (2002): Subgroups



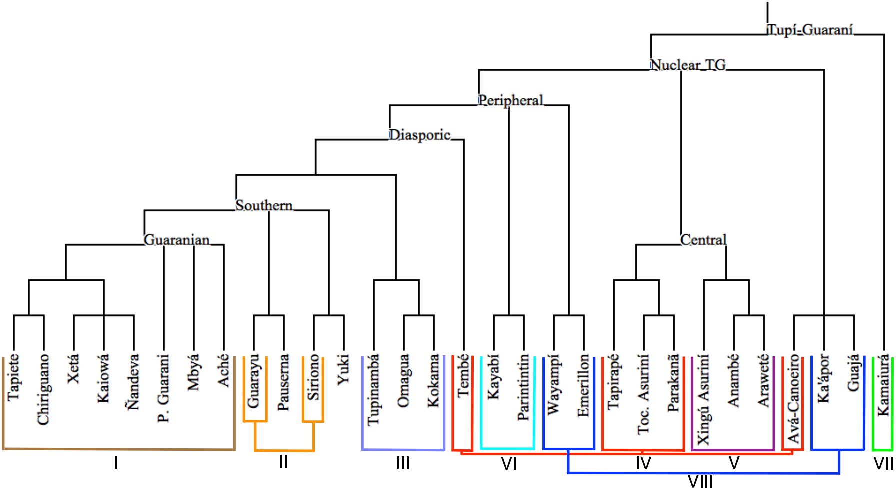
Comparison with Rodrigues and Cabral (2002): Subgroups



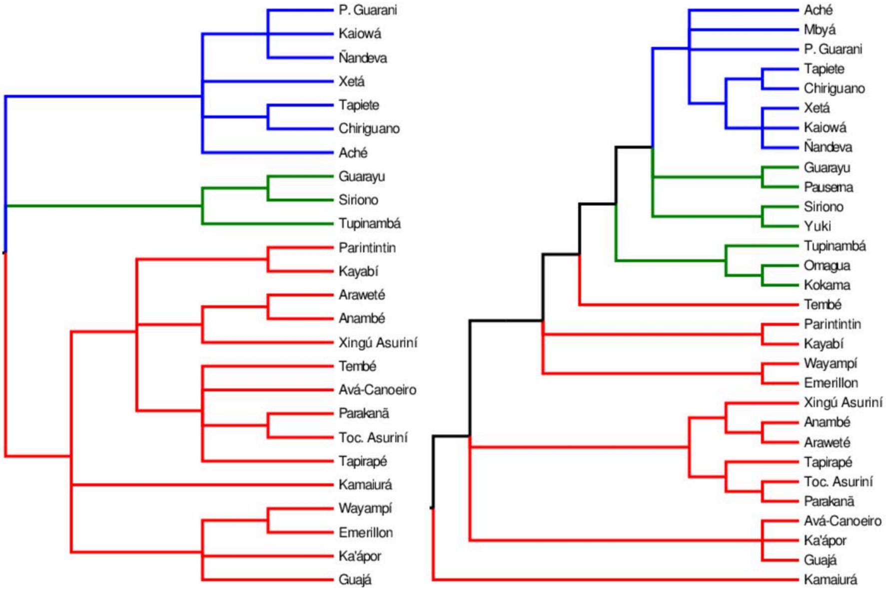
Comparison with Rodrigues and Cabral (2002): Subgroups



Comparison with Rodrigues and Cabral (2002): Subgroups



Comparison with Rodrigues and Cabral (2002): Structure



Comparison with Rodrigues and Cabral (2002): Structure

- We recover five commonly accepted subgroups (I, III, V, VI, VII), but not the others
- Higher-level structure quite different:
  - Group I (our 'Guaranian') is a first order branch for R&C, but is a deeply nested monophyletic branch for us
  - Group II/III is a first order monophyletic branch for R&C, but is a paraphyletic group embedded in the middle of the tree for us
  - R&C's large 'Amazonian group' (see Dietrich (1990)) is a first order branch, but these languages are paraphyletic in our analysis

Intermission

Introduction

- The goals of this half of the talk are twofold:
  1. Present a model for the geographic dispersal of the Tupí-Guaraní languages inferred from
     - A new internal classification of the Tupí-Guarani family (Chousou-Polydouri et al. 2014) and
     - The earliest known locations of the languages in question
  2. Contribute to dialogue between linguists, archaeologists, and human geneticists regarding the dispersal of Tupí-Guaraní languages, and identify fruitful areas of investigation in these allied fields

Center-of-gravity Inference

- The migration model we present is based on the **center of gravity** (CoG) inference heuristic developed in linguistic migration theory (Diebold Jr. 1960; Dyen 1956; Nichols 1997; Sapir [1916]1949)
- CoG infers a likely region in which the shared ancestor of a group of daughter languages was spoken, assuming, *all other things being equal*, that:
  - The ancestral language was spoken in the region occupied by the largest number of first order daughters of the proto-language
  - The homeland requires the smallest number of migratory movements to explain the modern distribution of the daughter languages
  - The homeland requires the shortest migratory movements
- This inference process can be applied both to a family as a whole and to particular branches to develop a model for the geographic dispersal of a family

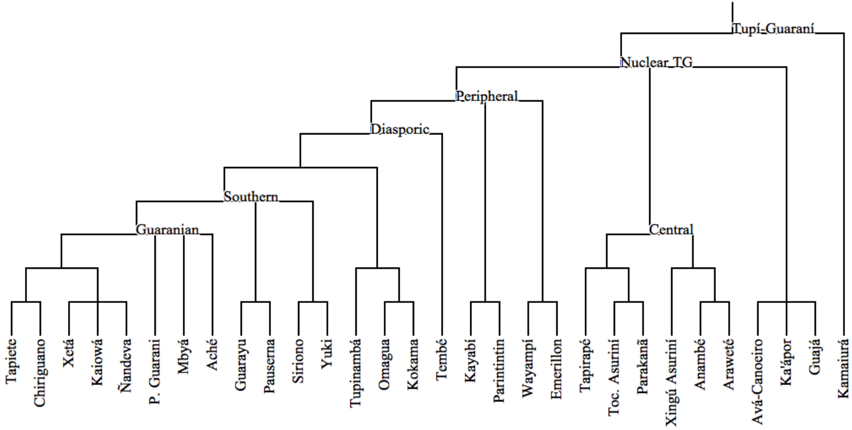
Refining CoG

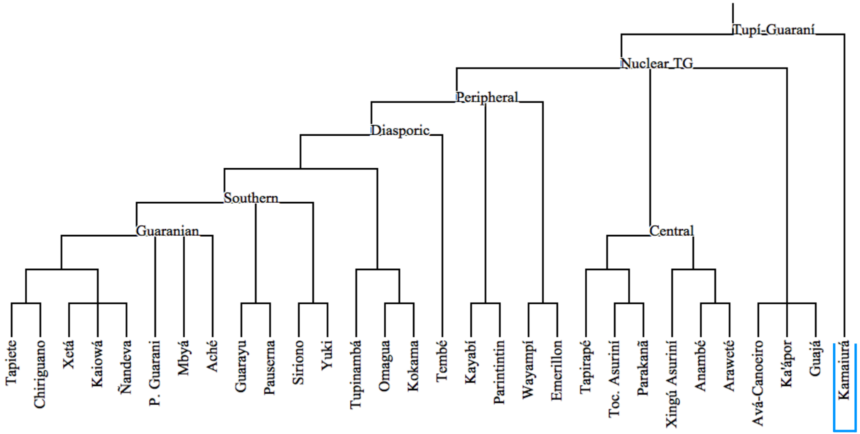
- **Geographical features** make certain movement trajectories typically more or less costly than others
  - Riverine movement is typically less costly than overland movement (provided the groups in question have water craft)
- **Ecological factors** likewise affect movement
  - Remaining within ecologically similar zones allows for continuity in subsistence practices
    - For exampe, movements that allow riverine groups to retain riverine subsistence practices are more probable than ones that require such groups to develop interfluvial practices.
- Assessing the effect of geographical features and ecological factors is facilitated by knowledge of **cultural practices** and **subsistence practices**

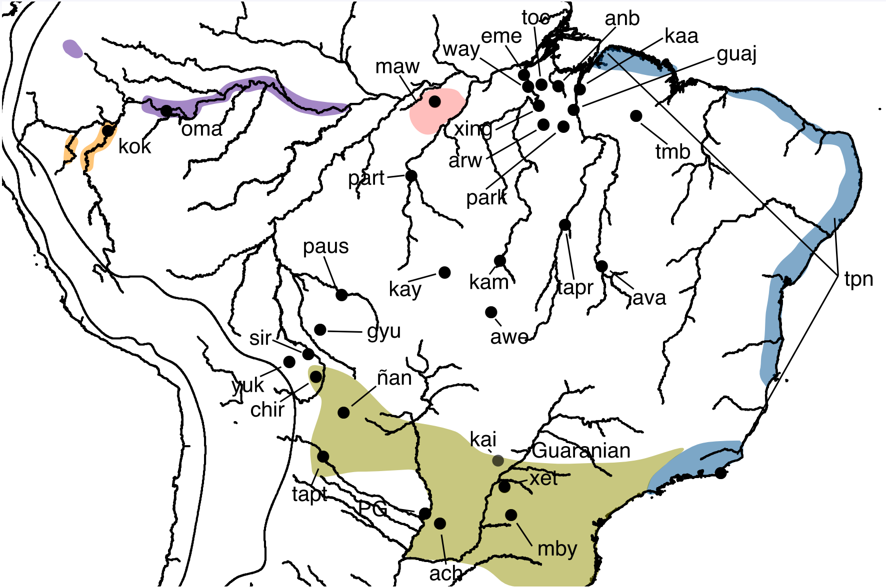
Language and Subgroup Locations

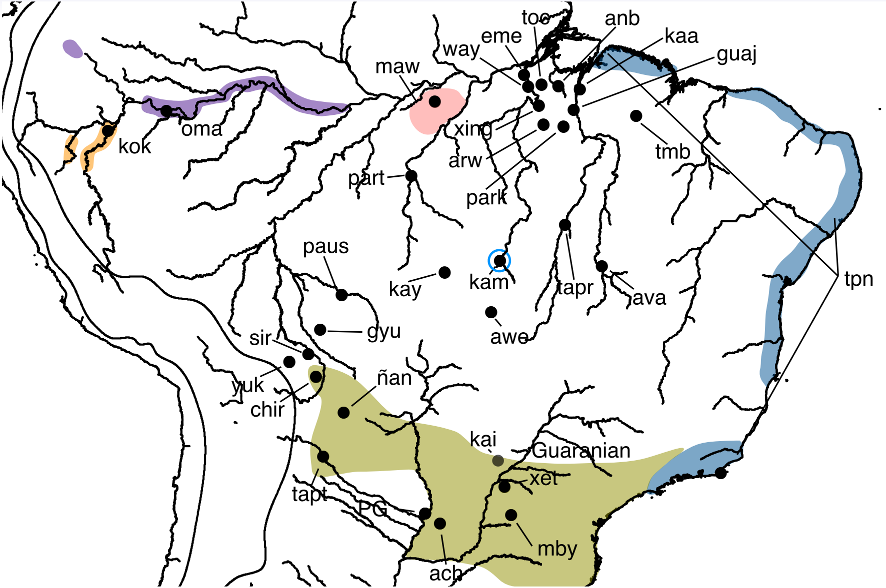
- In carrying out CoG inference, locations attributed to attested languages plays an important role
- We increase the accuracy of the inference process by using the earliest known language locations (generally at "time of contact")
  - For example, Emerillon and Wayampí, now spoken in French Guiana and northern Amapá, respectively, were both spoken on the lower Xingú in the early colonial period (Grenand 1982)
  - Guajá and Ka'ápor were probably spoken on the lower Tocantins not long before the arrival of Europeans (Balée 1994)
- We now consider the location of the various subgroups proposed by Chousou-Polydouri et al. (2014) in order to develop a sense of how the proposed classification maps onto the geography

Classification (Chousou-Polydouri et al. 2014)

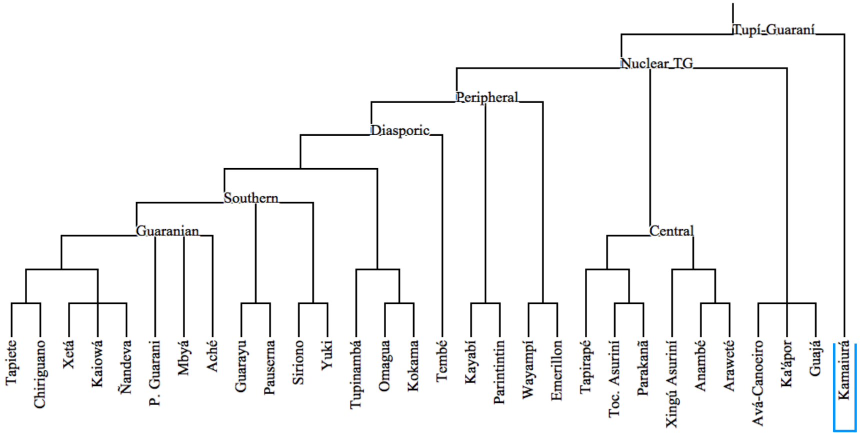


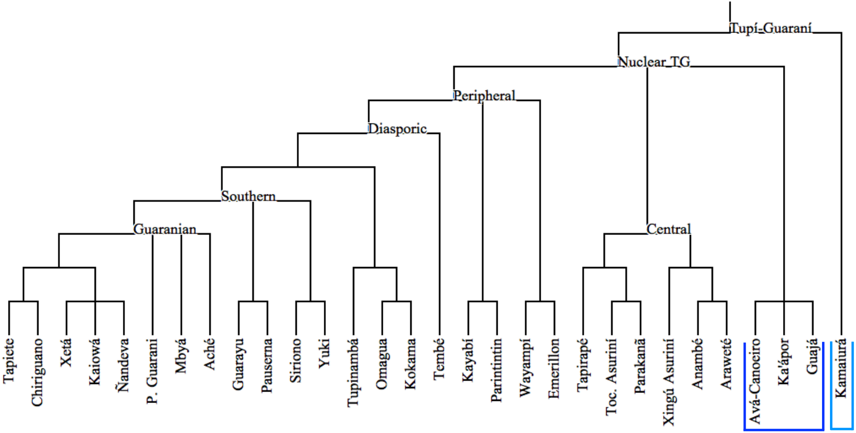

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampí
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá

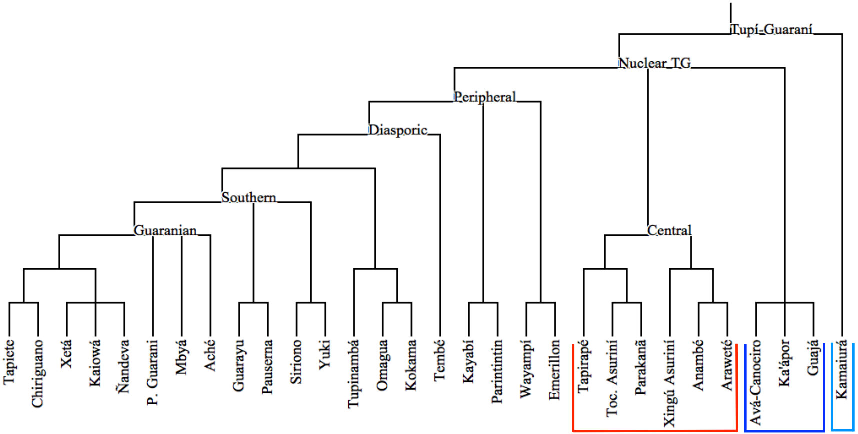

toc
anb
eme
kaa
way
guaj
maw
oma
kok
xing
arw
tmb
part
park
paus
kay
kam
tapr
ava
sir
gyu
awe
yuk
chir
ñan
kai
Guaranian
xet
tapt
PG
mby
ach
tpn



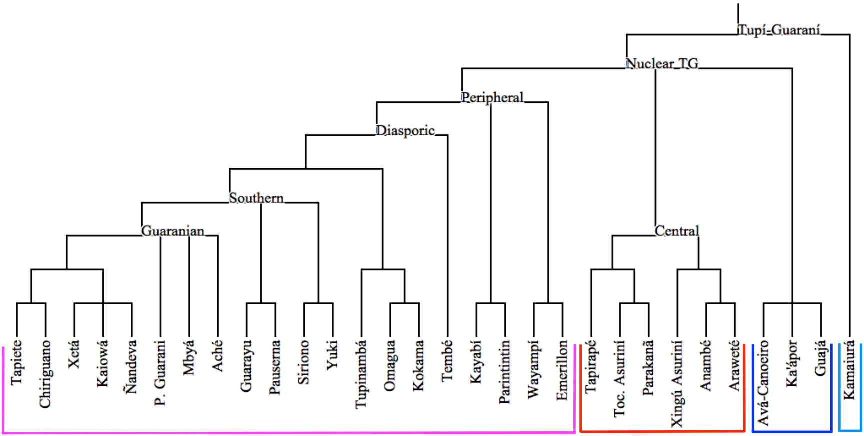
toc
anb
eme
kaa
way
guaj
maw
oma
kok
xing
arw
tmb
part
park
paus
kam
kay
tapr
ava
tpn
sir
gyu
awe
yuk
ñan
chir
kai
Guaranian
xet
tapt
PG
mby
ach

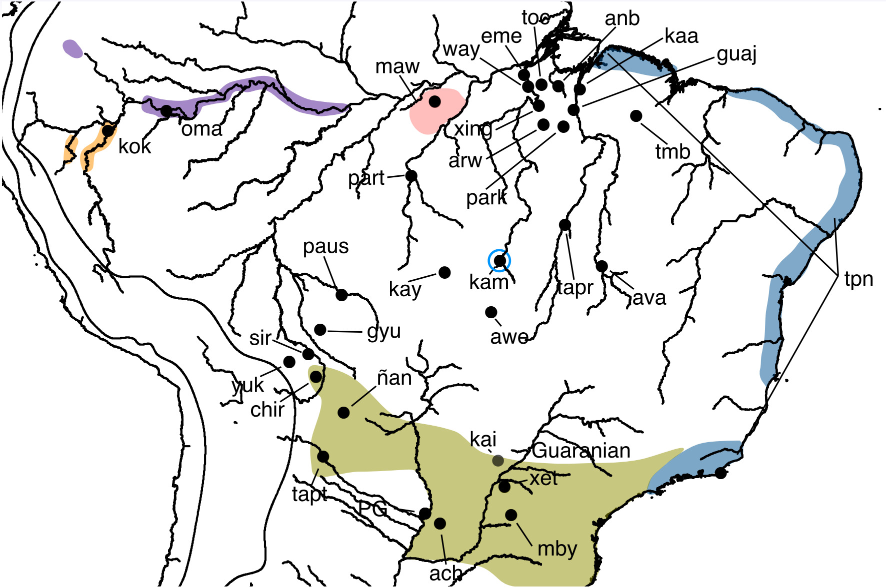

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampí
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá

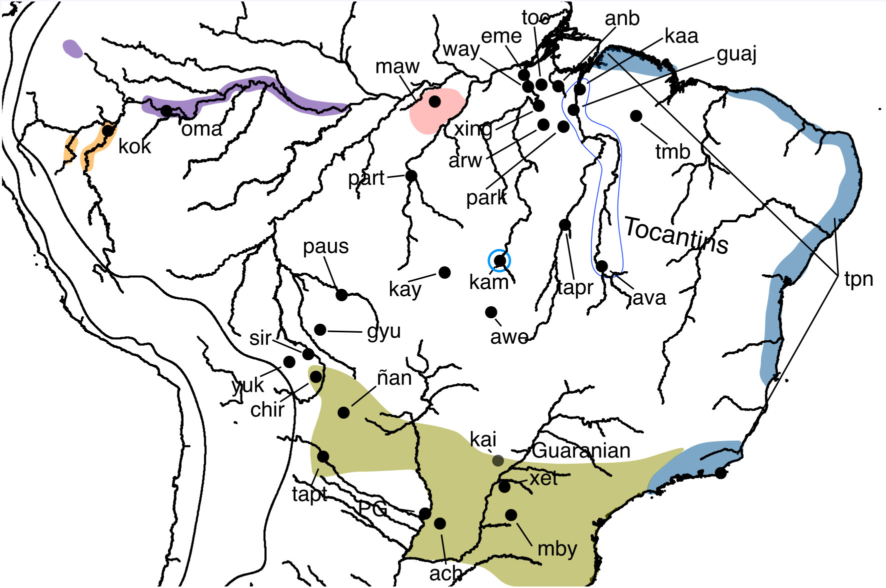

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampí
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá

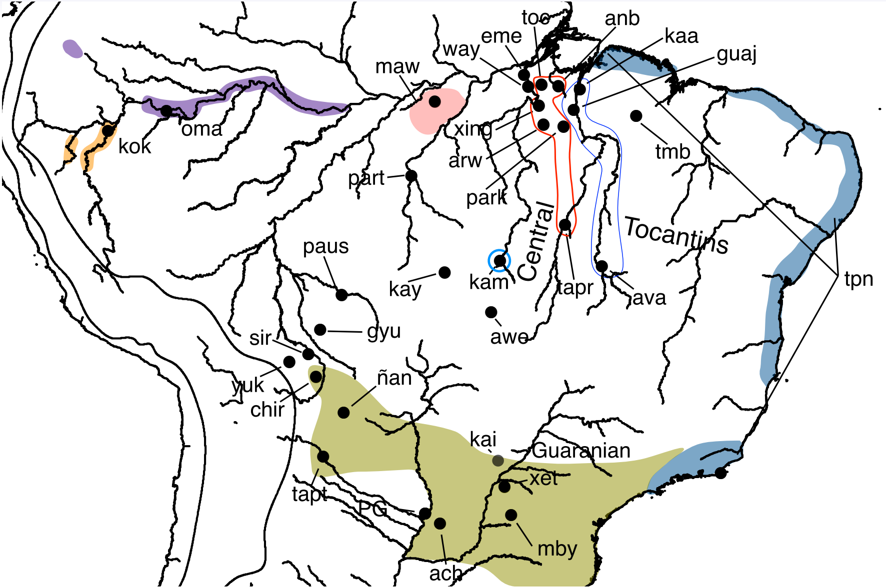


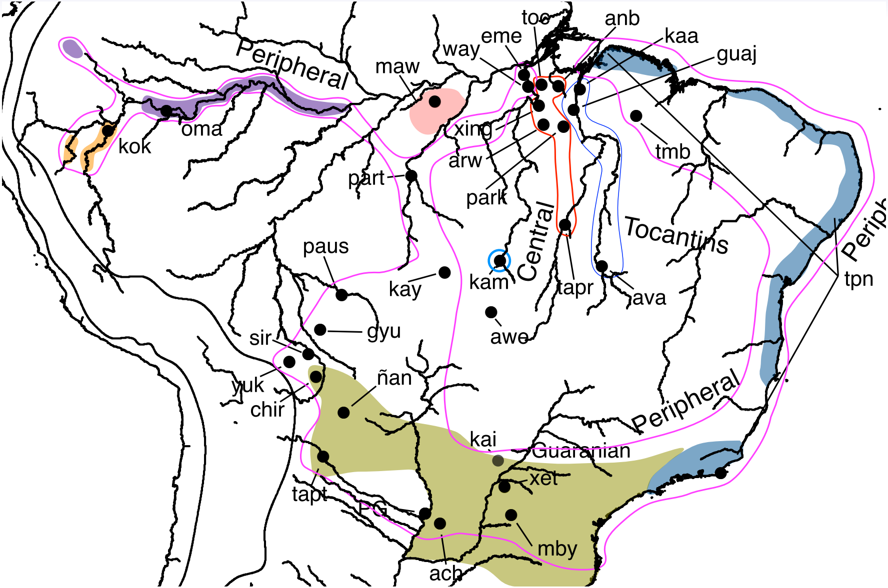
Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampí
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá

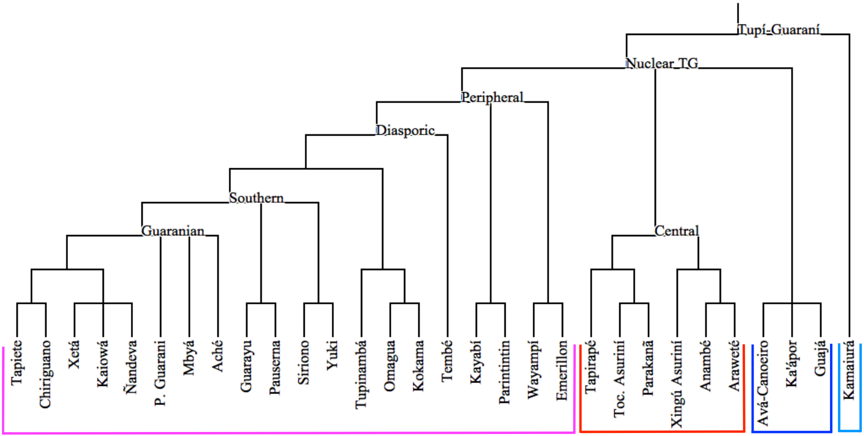

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabi
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá

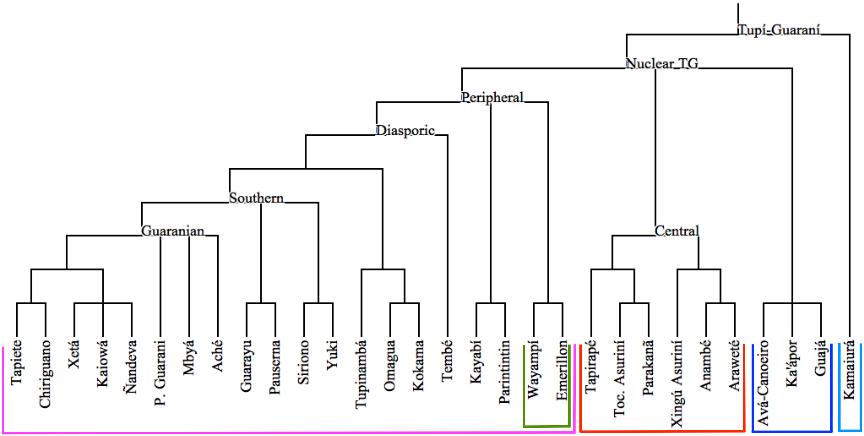

toc
anb
eme
kaa
way
guaj
maw
oma
kok
xing
arw
tmb
part
park
paus
kam
tapr
kay
ava
gyu
awe
sir
yuk
ñan
chir
kai
Guaranian
xet
tapt
PG
mby
ach
tpn

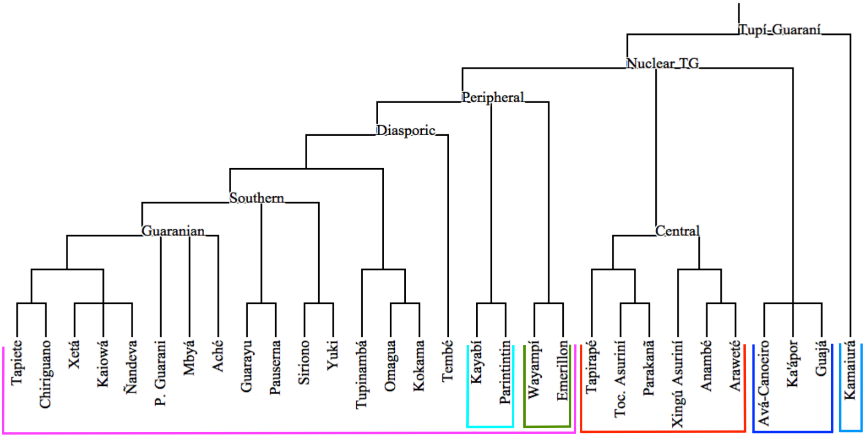

toc
anb
eme
kaa
way
guaj
maw
oma
kok
xing
arw
tmb
part
park
Tocantins
paus
kay
kam
tapr
ava
tpn
gyu
awe
sir
yuk
ñan
chir
kai
Guaranian
xet
tapt
PG
mby
ach

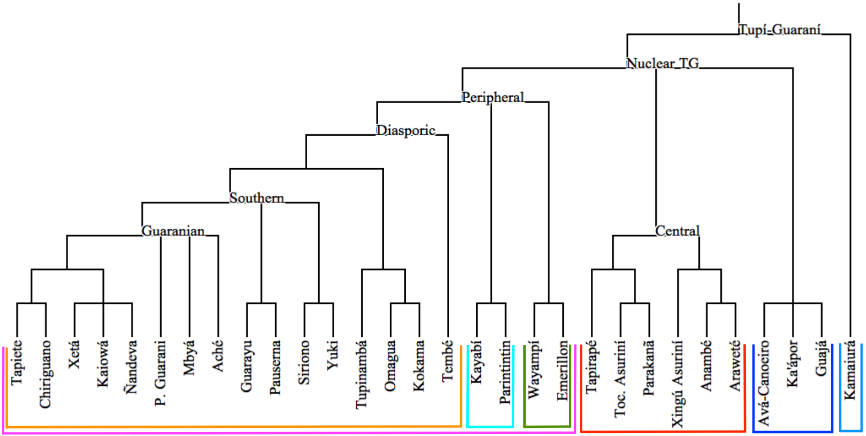

toc
anb
eme
kaa
way
guaj
maw
oma
kok
xing
arw
tmb
part
park
Central
Tocantins
paus
kay
kam
tapr
ava
tpn
gyu
awe
sir
yuk
ñan
chir
kai
Guaranian
xet
tapt
PG
mby
ach

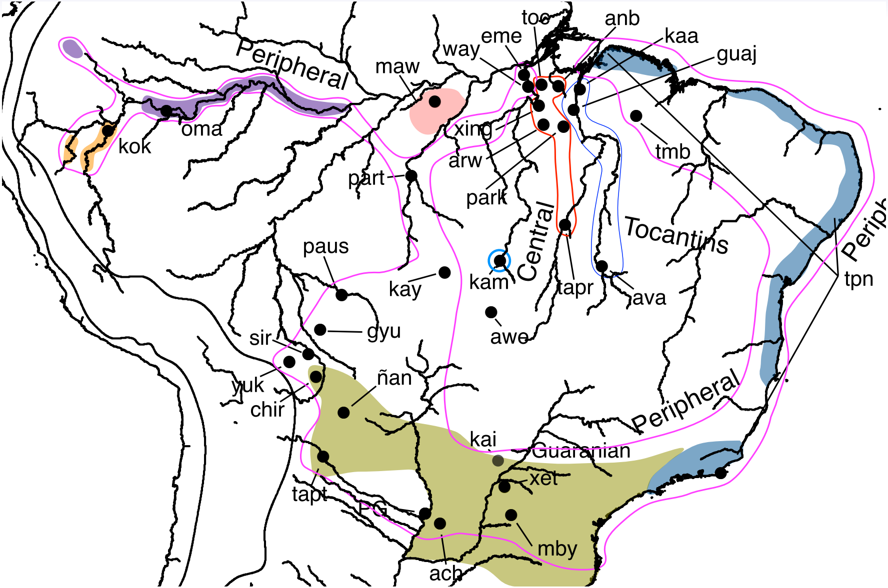

Peripheral
maw
way
eme
toc
anb
kaa
guaj
oma
kok
xing
arw
park
part
tmb
Central
Tocantins
Periph
paus
kay
kam
tapr
ava
tpn
gyu
sir
awe
yuk
chir
ñan
Peripheral
kai
Guaranian
xet
tapt
ach
mby

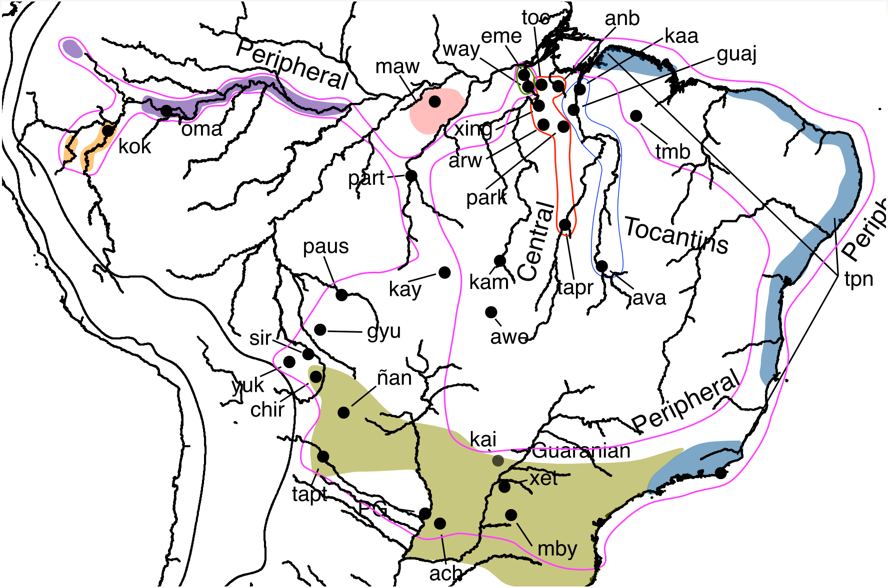

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabi
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'apor
Guajá
Kamaiurá

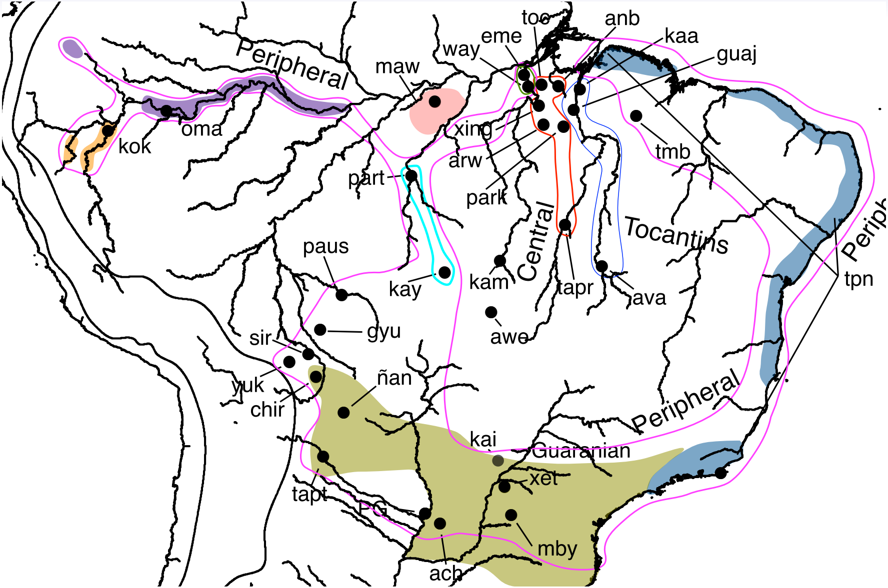

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabi
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá

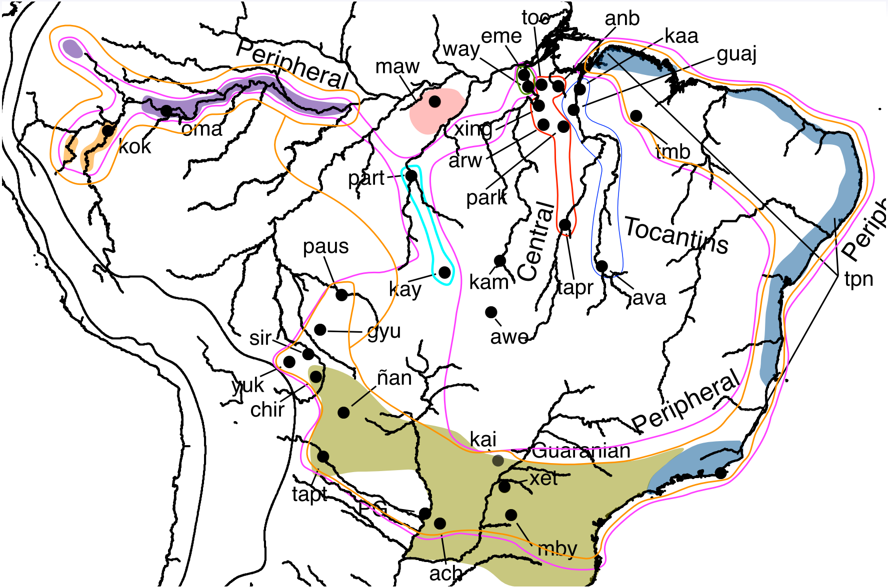

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Arawetê
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá

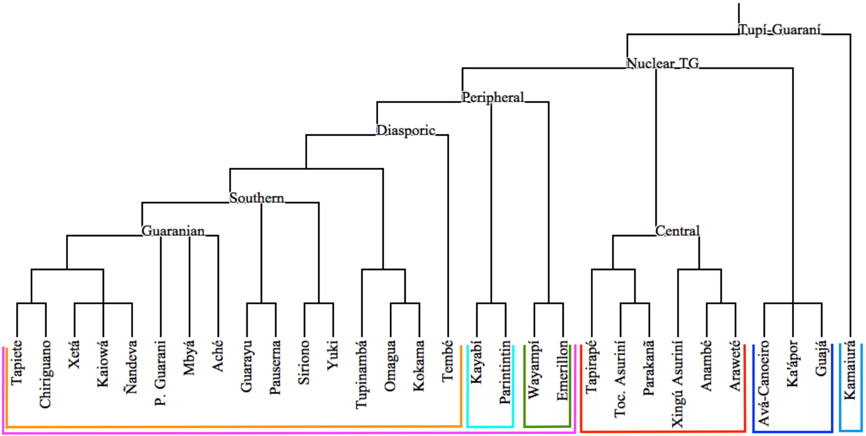

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'àpor
Guajá
Kamaiurá

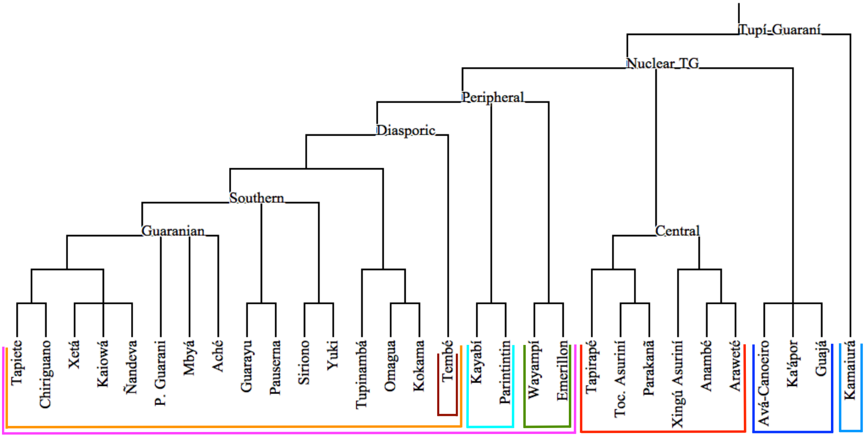

Peripheral
maw
way
eme
toc
anb
kaa
guaj
oma
kok
xing
arw
park
part
tmb
Central
Tocantins
paus
kay
kam
tapr
ava
tpn
gyu
awe
sir
yuk
chir
ñan
kai
Guaranian
xet
tapt
mby
ach
Peripheral

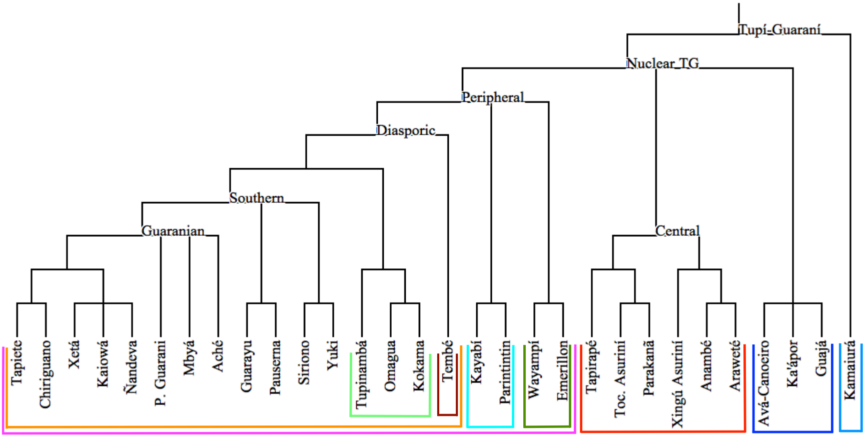

Peripheral
maw
way
eme
toc
anb
kaa
guaj
oma
kok
xing
arw
park
part
tmb
Central
Tocantins
paus
kay
kam
tapr
ava
Periph
tpn
gyu
awe
sir
yuk
chir
ñan
Peripheral
kai
Guaranian
xet
tapt
PG
mby
ach

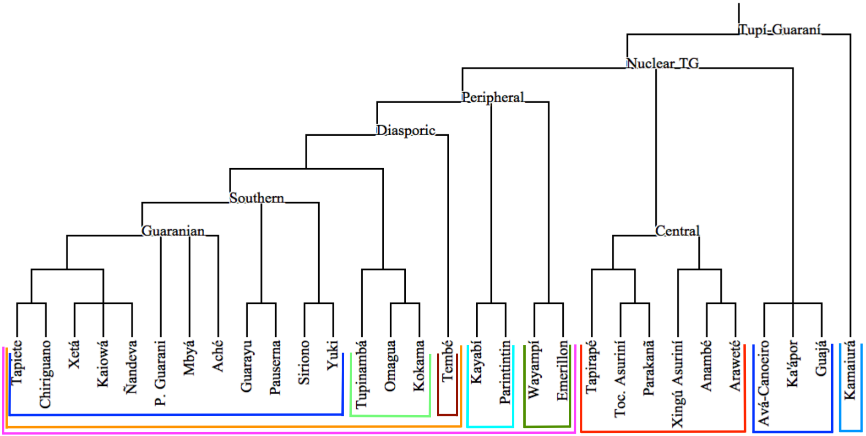

Peripheral
maw
way
eme
toc
anb
kaa
guaj
oma
kok
xing
arw
park
part
tmb
Central
Tocantins
paus
kay
kam
tapr
ava
tpn
Periph
gyu
awe
sir
yuk
chir
ñan
Peripheral
kai
Guaranian
xet
tapt
FG
mby
ach

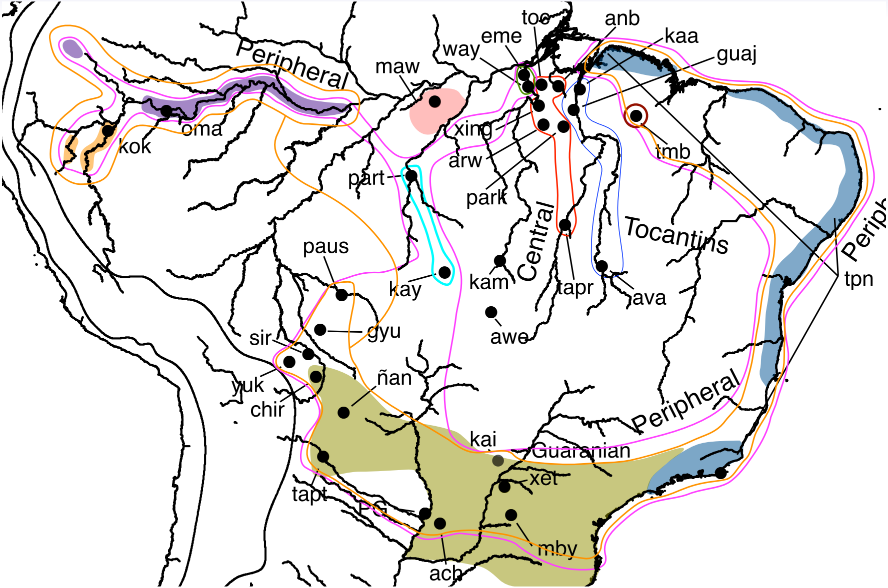

Peripheral
maw
way
eme
toc
anb
kaa
guaj
oma
kok
xing
arw
park
part
tmb
Central
Tocantins
Periph
paus
kay
kam
tapr
ava
tpn
gyu
awe
sir
yuk
chir
ñan
Peripheral
kai
Guaranian
xet
tapt
ach
mby

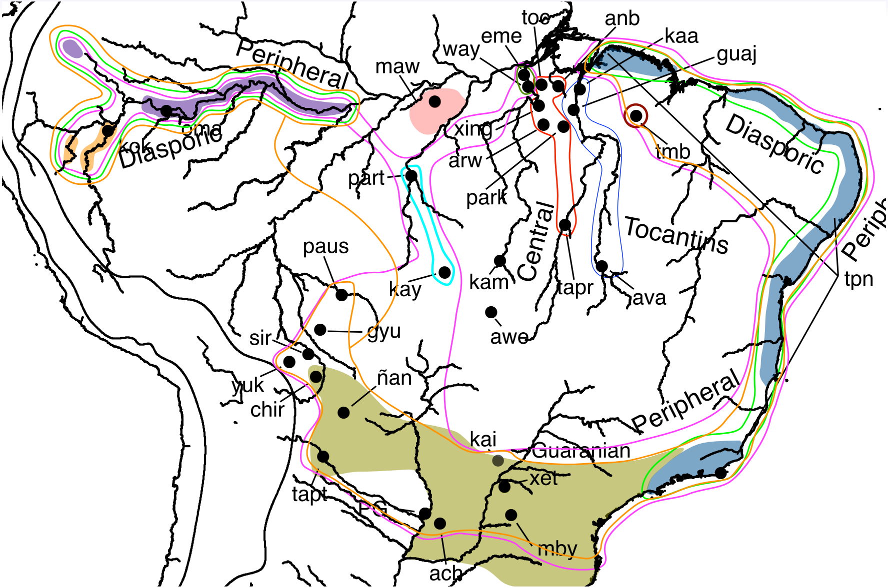

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá

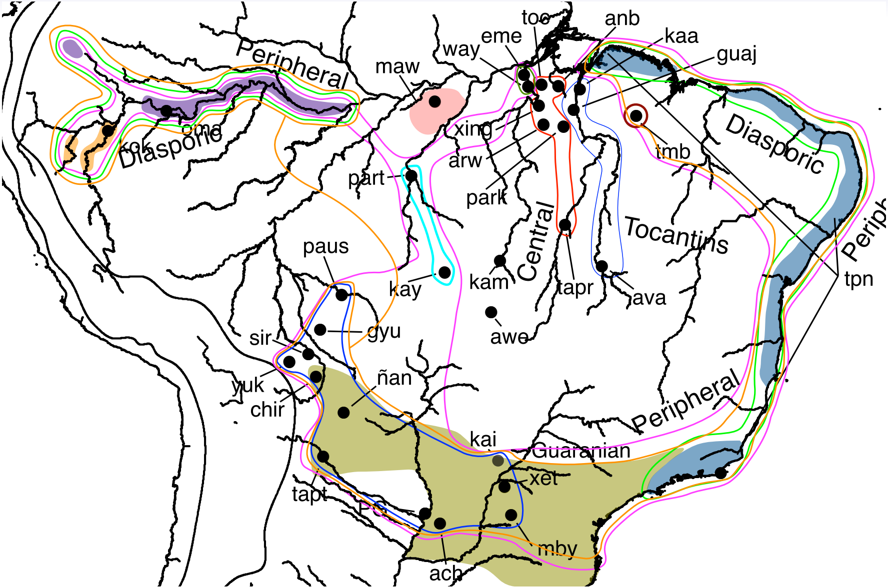

Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá


Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá


Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'àpor
Guajá
Kamaiurá

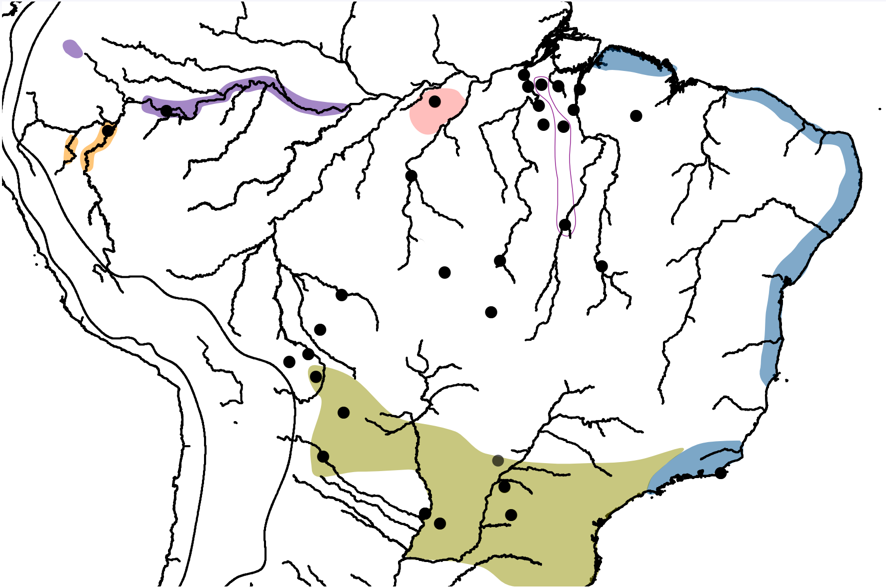

Peripheral
maw
way
eme
toc
anb
kaa
guaj
oma
kok
xing
arw
park
part
tmb
Tocantins
Central
paus
kay
kam
tapr
ava
Periph
tpn
awe
gyu
sir
yuk
chir
ñan
Peripheral
kai
Guaranian
xet
tapt
mby
ach

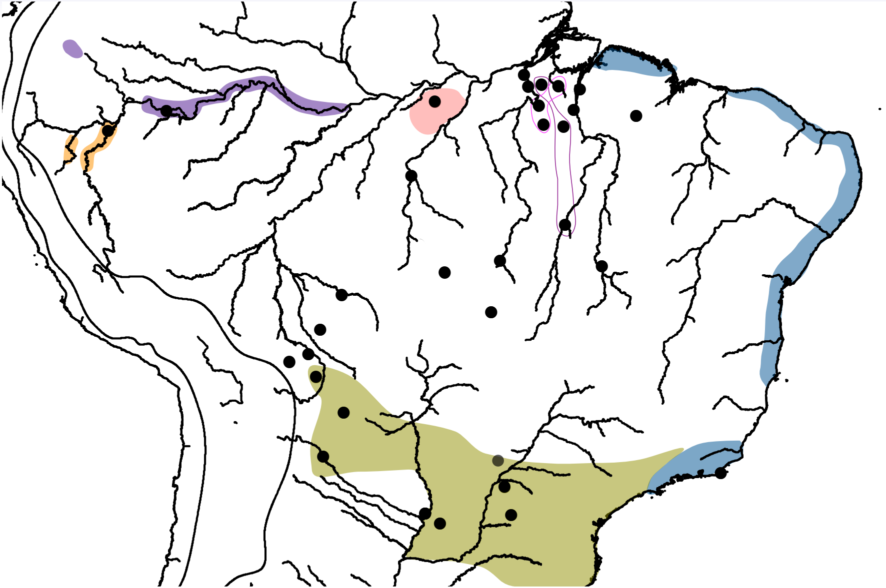

Peripheral
Diasporic
kok
oma
maw
way
eme
toc
anb
kaa
guaj
xing
arw
park
part
tmb
Diasporic
Central
Tocantins
Periph
tpn
paus
kay
kam
tapr
ava
awe
gyu
sir
yuk
chir
ñan
kai
Guaranian
Peripheral
xet
tapt
mby
ach

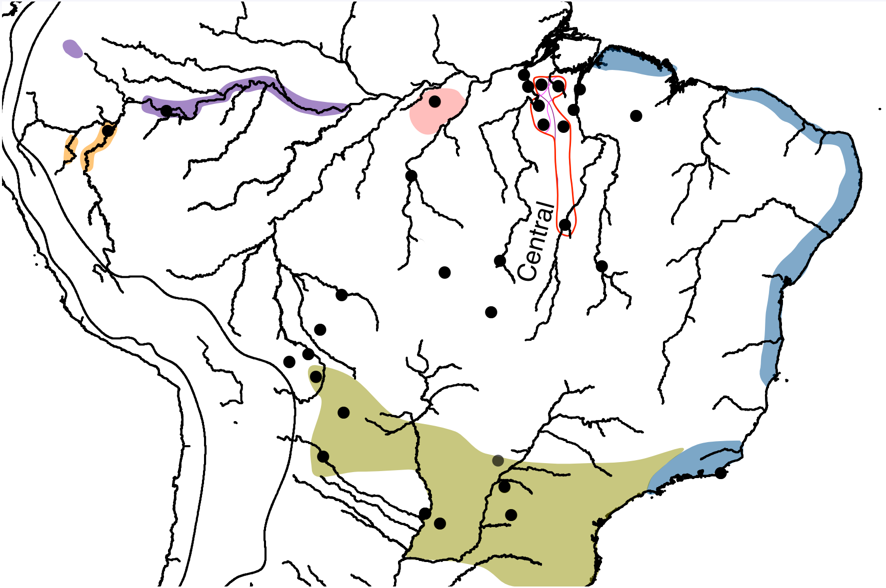

Peripheral
Diasporic
kok
cma
maw
way
eme
toc
anb
kaa
guaj
tmb
xing
arw
park
part
Diasporic
Central
Tocantins
Periph
tpn
paus
kay
kam
tapr
ava
awe
sir
gyu
yuk
chir
ñan
Peripheral
kai
Guaranian
xet
tapt
mby
ach

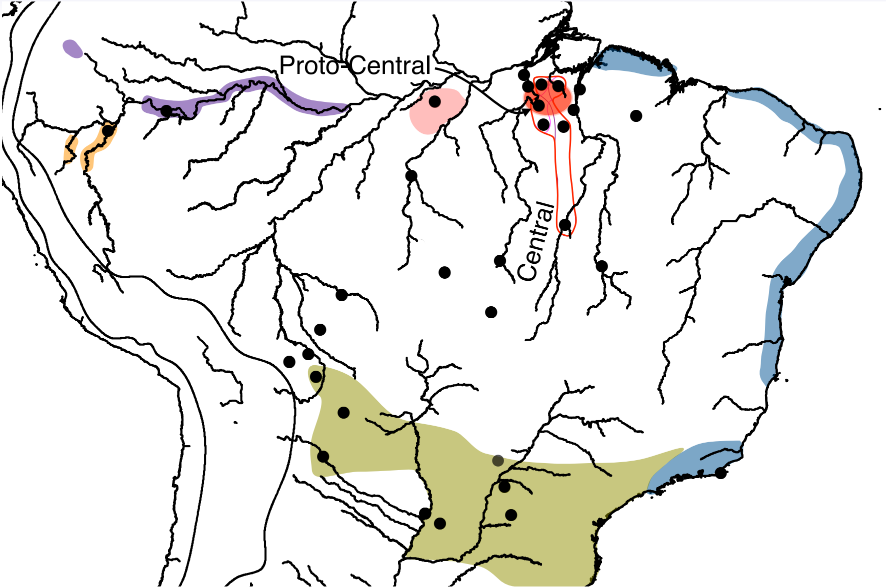
Inferring the PTG Homeland

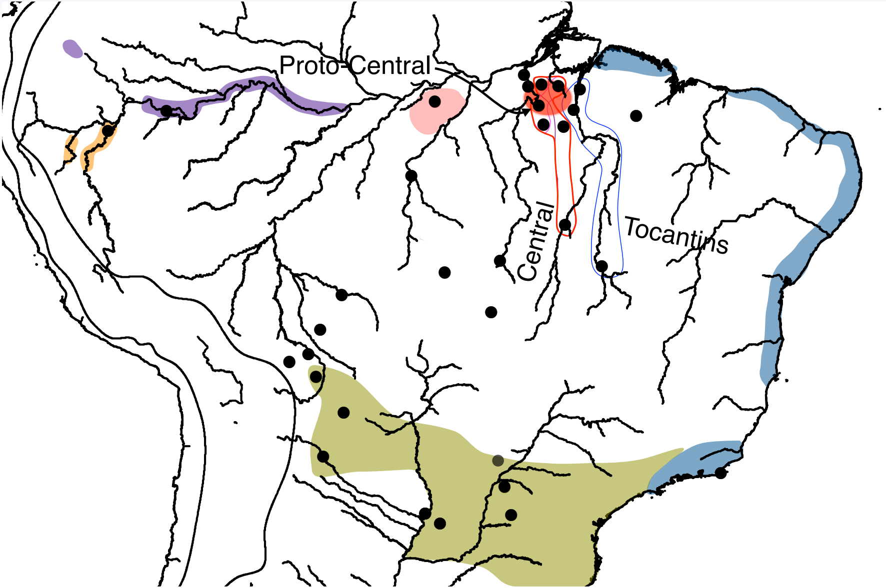
- We now turn to inference of the PTG homeland
- According to our classification, PTG split into Kamaiurá and the much larger Nuclear TG (NTG) branch
- We will temporarily set aside the question of the PTG homeland as such and focus on the Proto-NTG (PNTG) homeland
- Inference of the PNTG homeland depends on the location that we attribute to its three first order daughters: Proto-Central, Proto-Tocantins, and Proto-Peripheral
- Inferring the location of the Proto-Central and Proto-Tocantins homelands using CoG is relatively straightforward
- The Proto-Peripheral homeland is somewhat less obvious. . .

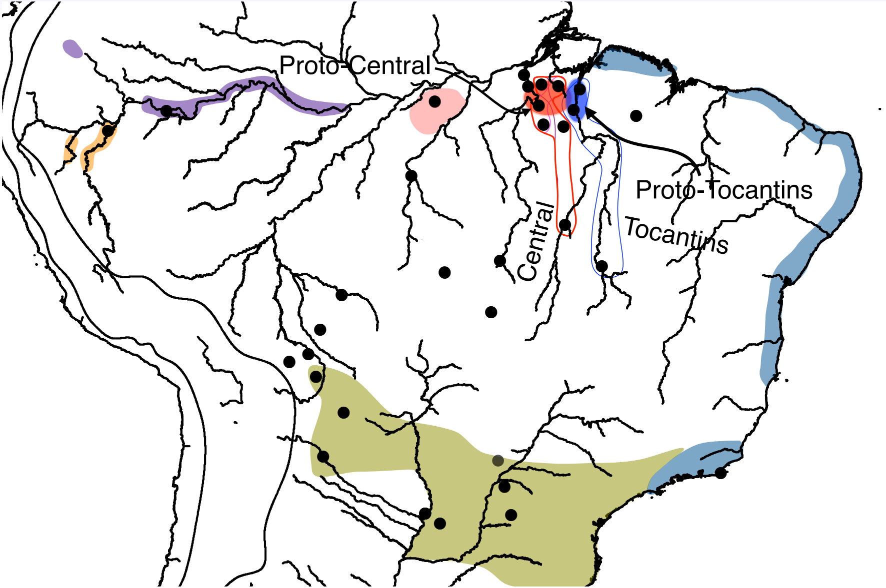


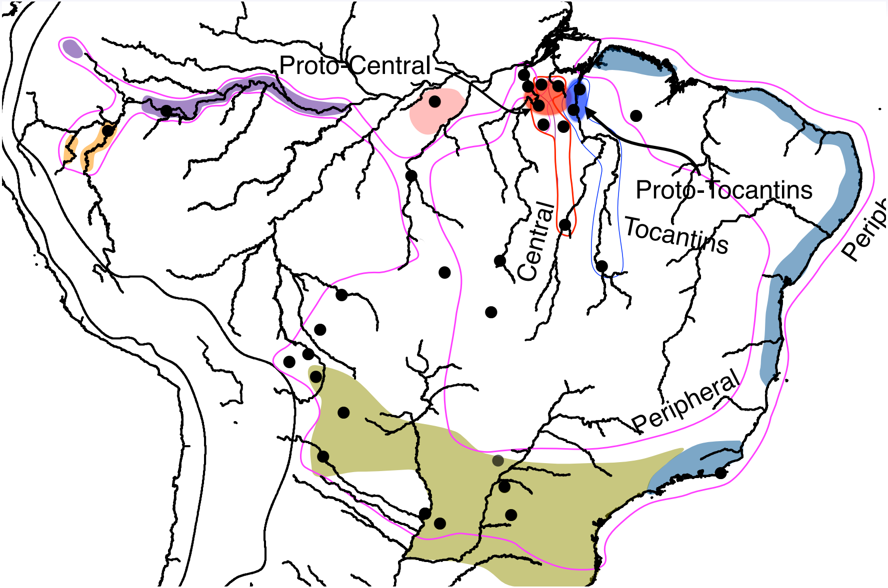



Central


Proto-Central
Central

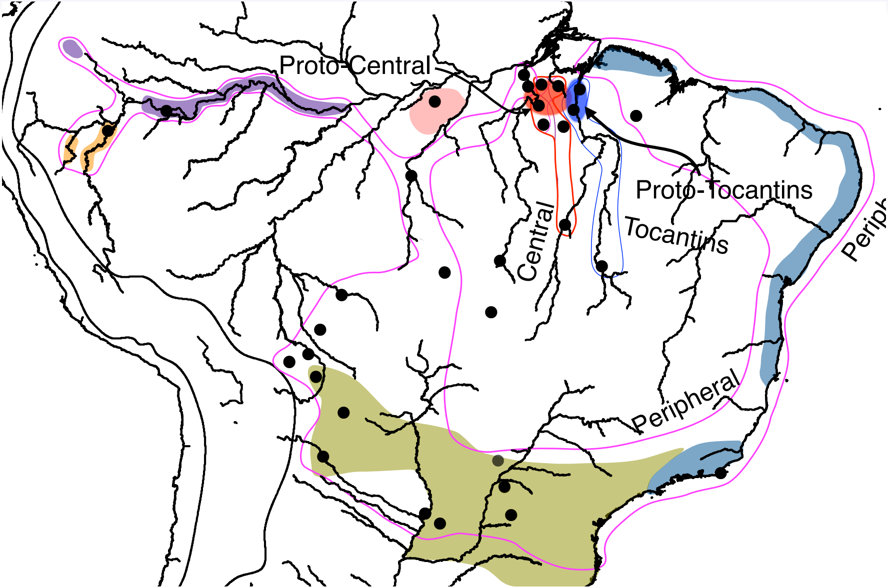

Proto-Central
Central
Tocantins

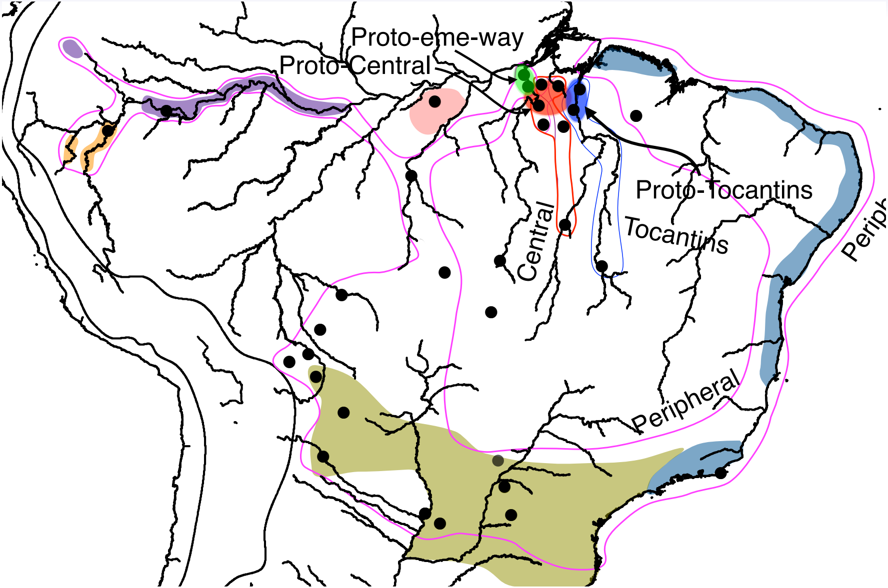

Proto-Central
Proto-Tocantins
Central
Tocantins

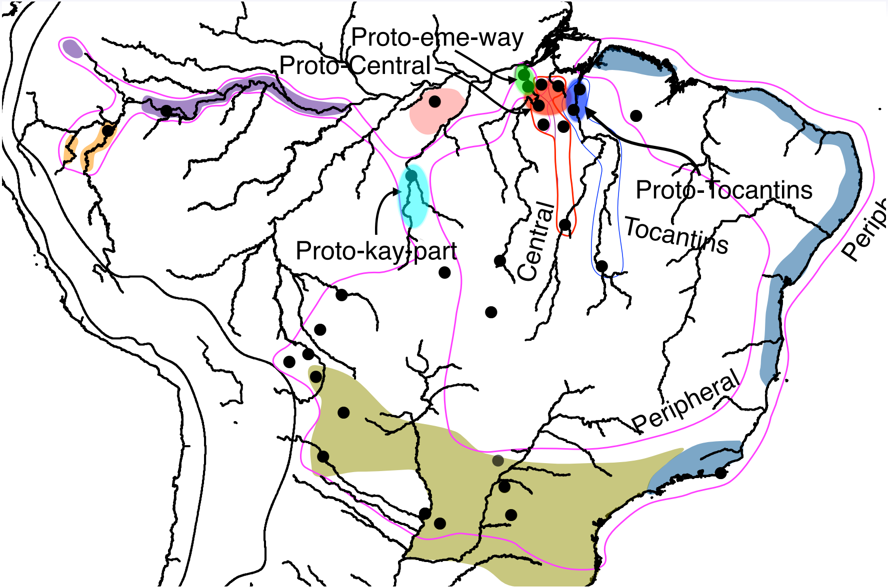

Proto-Central
Proto-Tocantins
Central
Tocantins
Peripheral
Periph

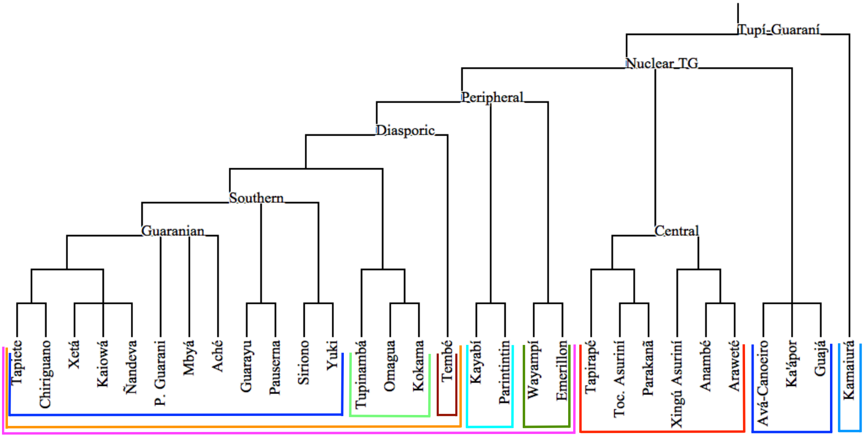
Inferring the Proto-Peripheral Homeland

- To infer the Proto-Peripheral homeland, it is helpful to consider the homelands associated with its three first-order branches: eme-way, kay-part, and Diasporic
- The Proto-eme-way and Proto-kay-part homelands can be inferred straightforwardly
- The Proto-Diasporic homeland is less clear, but based on the proximity of Tembé and Tupinambá (of the Omagua-Kokama-Tupinambá branch), we infer a region on the southern banks of the mouth of the Amazon

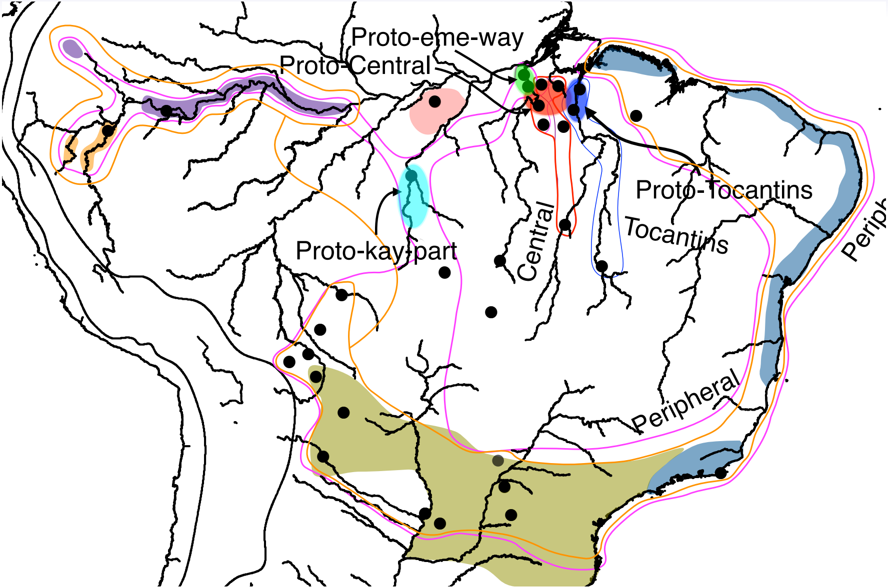

Proto-Central
Proto-Tocantins
Central
Tocantins
Peripheral
Periph

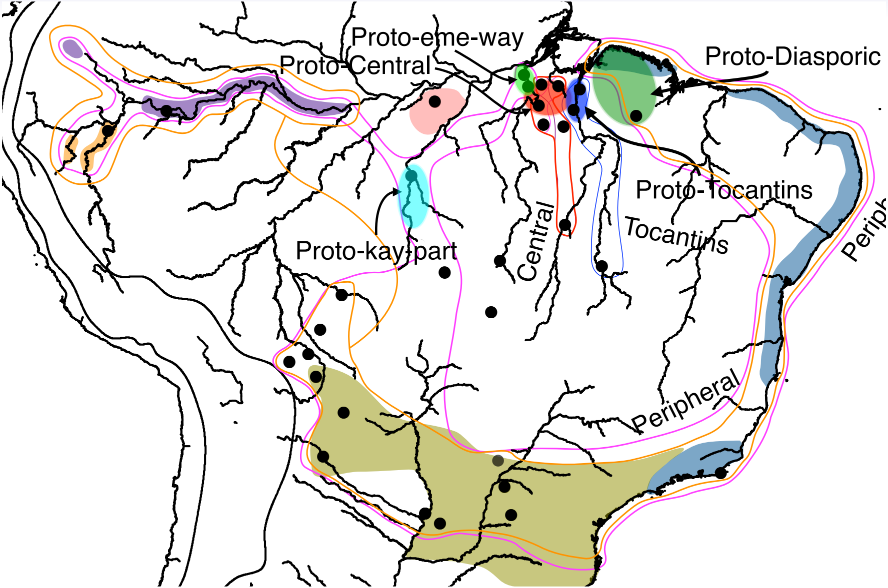

Proto-eme-way
Proto-Central
Proto-Tocantins
Central
Tocantins
Peripheral
Periph


Proto-eme-way
Proto-Central
Proto-Tocantins
Proto-kay-part
Central
Tocantins
Peripheral
Periph



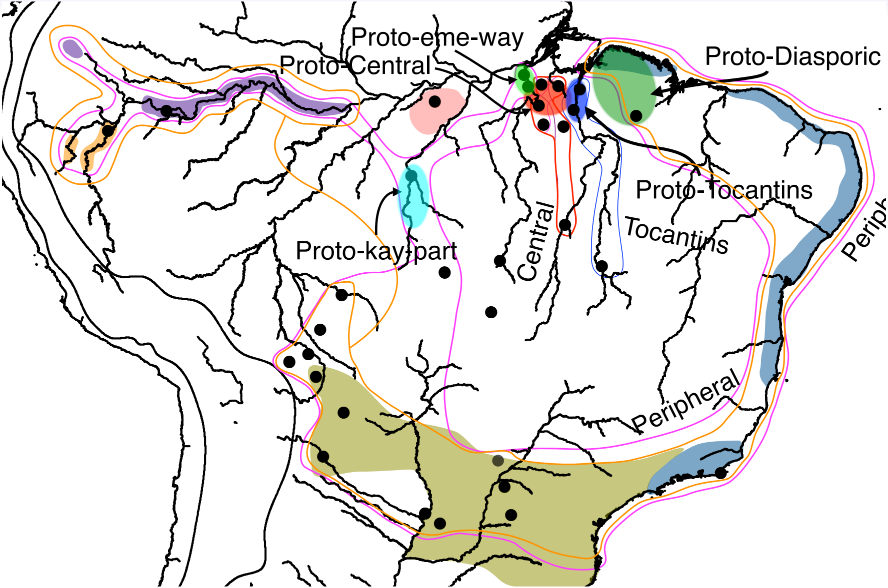
Tupí-Guaraní
Nuclear TG
Peripheral
Diasporic
Southern
Guaranian
Central
Tapiete
Chiriguano
Xetá
Kaiowá
Ñandeva
P. Guarani
Mbyá
Aché
Guarayu
Pauserna
Siriono
Yuki
Tupinambá
Omagua
Kokama
Tembé
Kayabí
Parintintin
Wayampi
Emerillon
Tapirapé
Toc. Asuriní
Parakanã
Xingú Asuriní
Anambé
Araweté
Avá-Canoeiro
Ka'ápor
Guajá
Kamaiurá

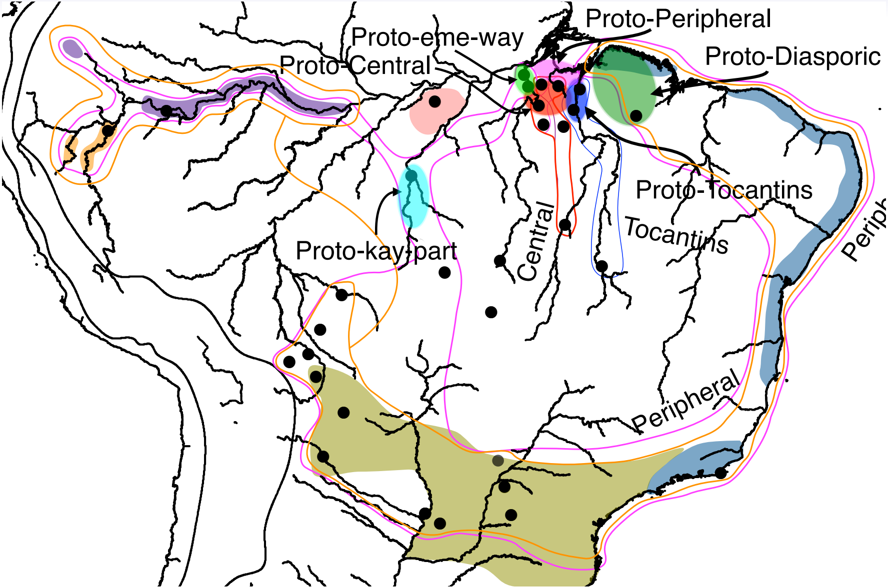

Proto-eme-way
Proto-Central
Proto-Tocantins
Proto-kay-part
Central
Tocantins
Peripheral
Periph


Proto-eme-way
Proto-Central
Proto-Diasporic
Proto-Tocantins
Tocantins
Central
Proto-kay-part
Peripheral
Periph

Inferring the Proto-Peripheral Homeland

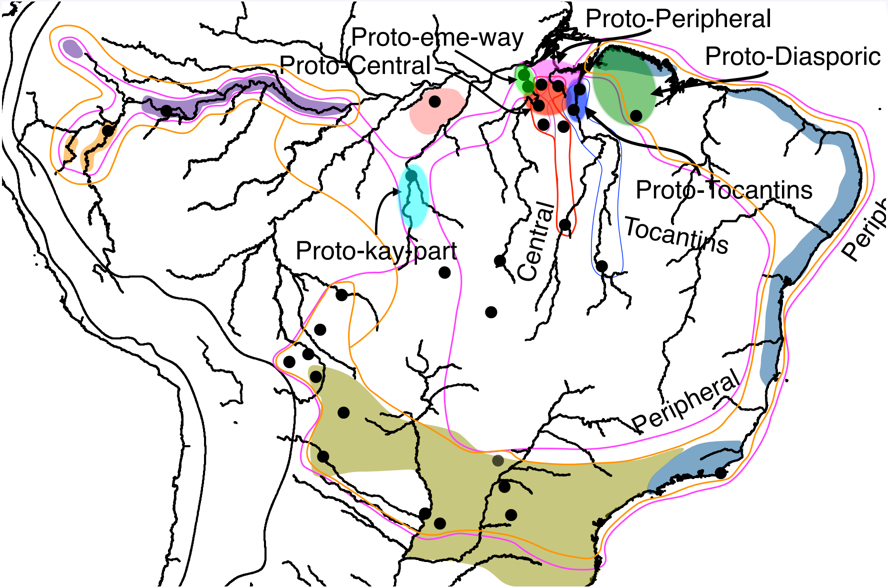
- Having posited homelands for the three branches of Proto-Peripheral (i.e., Proto-eme-way, Proto-Kay-part, and Proto-Diasporic), we can infer a homeland for Proto-Peripheral itself
- The most compact area straddling more than one branch of Peripheral stretches from the western bank of the Xingú to east of the Tocantins, leading us to place the Proto-Peripheral homeland there

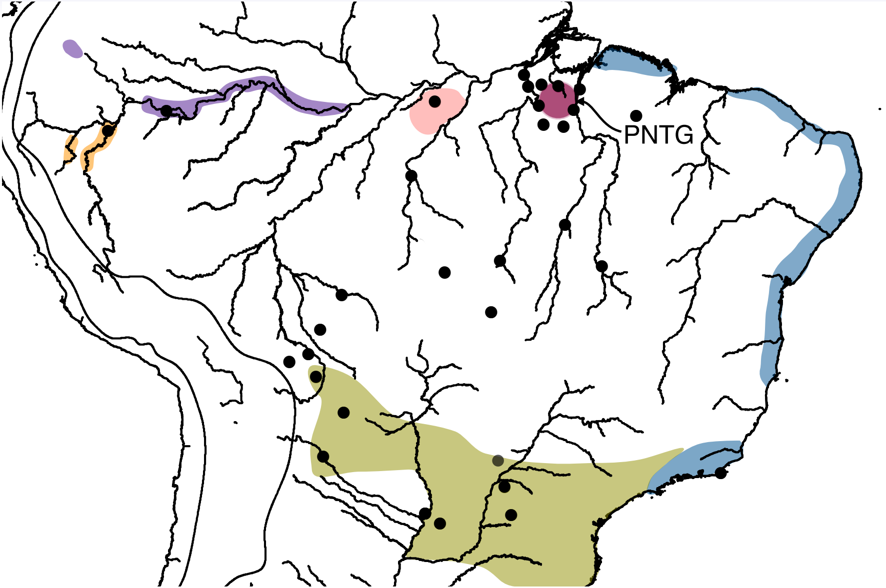

Proto-eme-way
Proto-Central
Proto-Diasporic
Proto-Tocantins
Tocantins
Central
Proto-kay-part
Peripheral
Periph


Proto-Peripheral
Proto-eme-way
Proto-Diasporic
Proto-Central
Proto-Tocantins
Central
Tocantins
Proto-kay-part
Peripheral
Periph

Inferring the PNTG Homeland

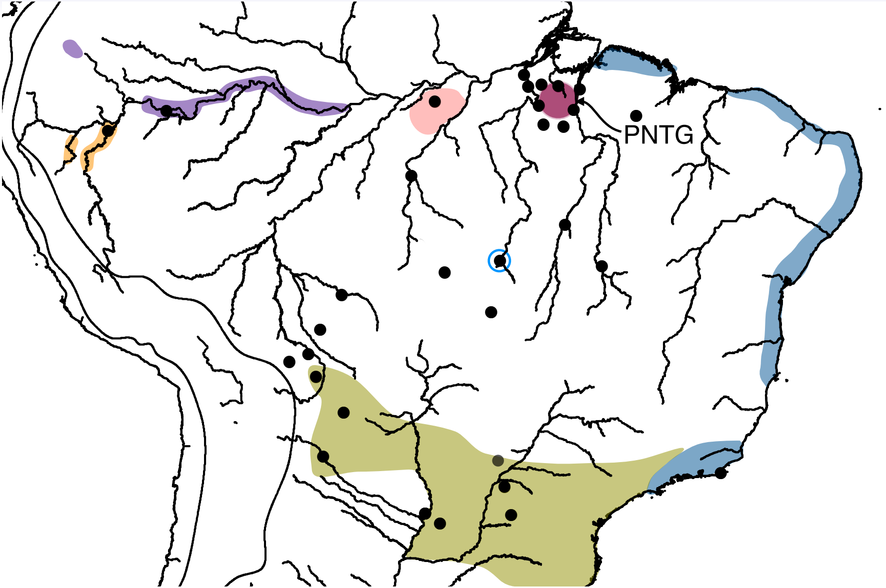
- Having located the homelands for the three first-order branches of PNTG (i.e. Proto-Central, Proto-Tocantins, and Proto-Peripheral), the inference of the PNTG homeland is straightforward
- The locus of genetic diversity is clearly located in a region extending from the Xingú to Tocantins, some small distance upriver from the mouths of these rivers

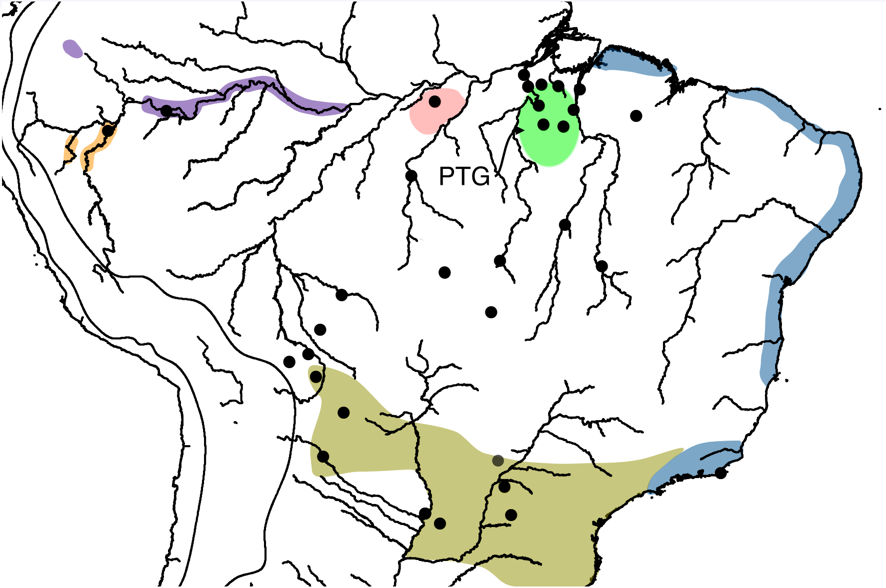

Proto-Peripheral
Proto-eme-way
Proto-Central
Proto-Diasporic
Proto-Tocantins
Central
Tocantins
Proto-kay-part
Peripheral
Periph


PNTG

Inferring the PTG Homeland

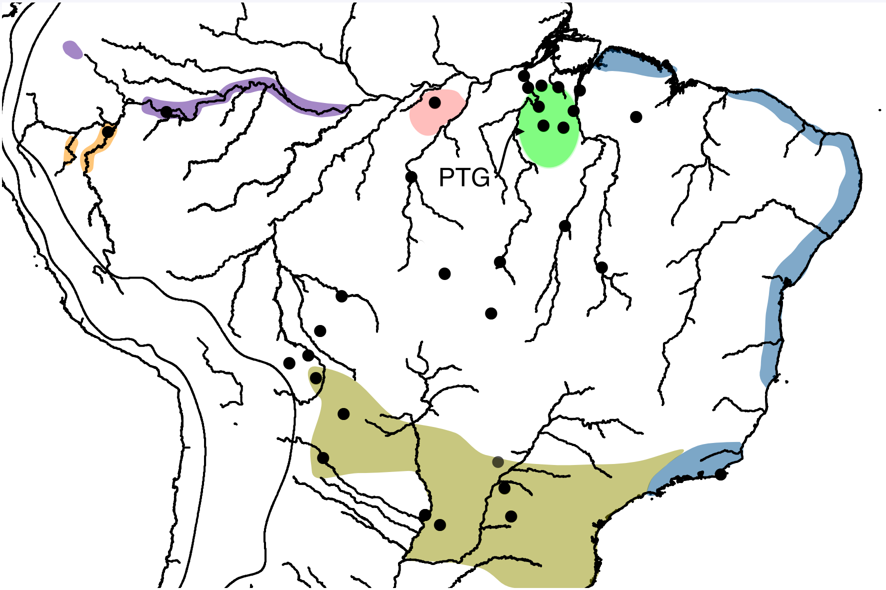
- The two first order daughters of PTG are PNTG and Kamaiurá, which are relatively distant from one another
- Given that:
  1. Kamaiurá is located upriver of the posited PNTG homeland
  2. We have seen a general trend for upriver dispersals in the diversification of TG (e.g. Proto-kay-part, Tapirapé, Avá-Canoeiro)
- . . . we hypothesize that it was, to a greater degree, Kamaiurá that migrated upriver than PNTG that migrated downriver
- This leads us to posit that PTG was spoken in a region similar to that of PNTG, but with modestly greater upriver extension

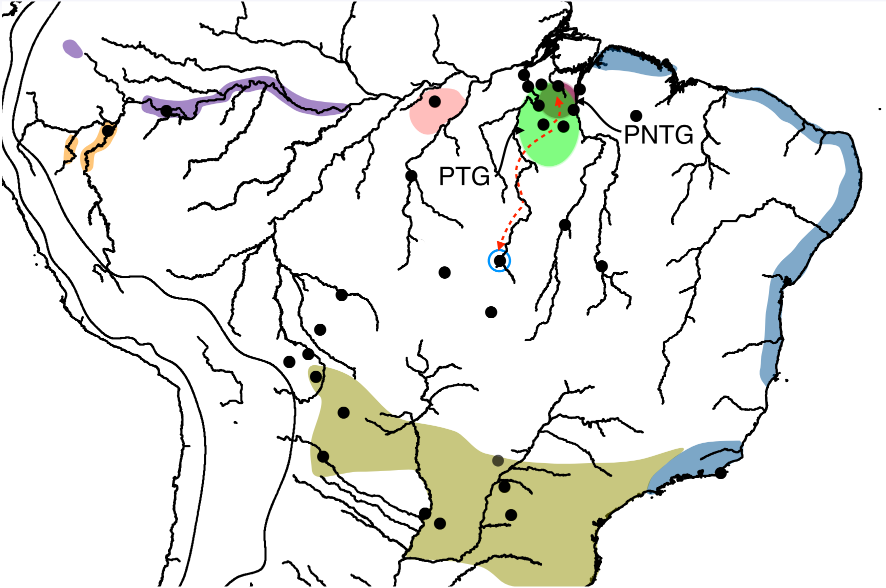

PNTG

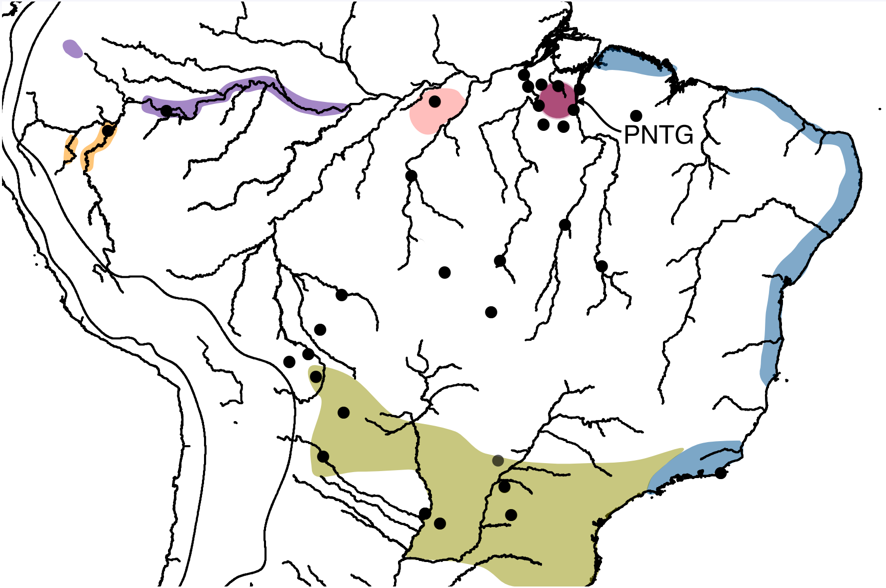

PTG

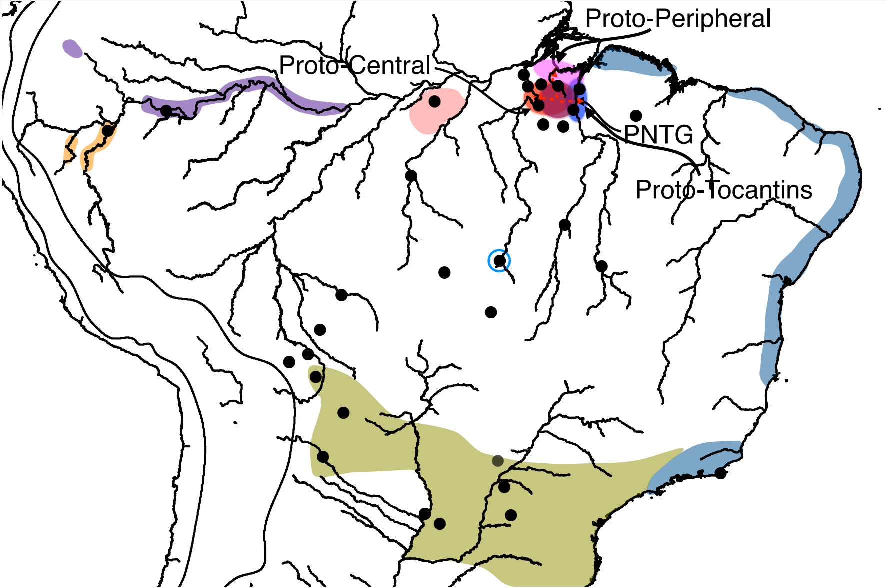
Migratory Model

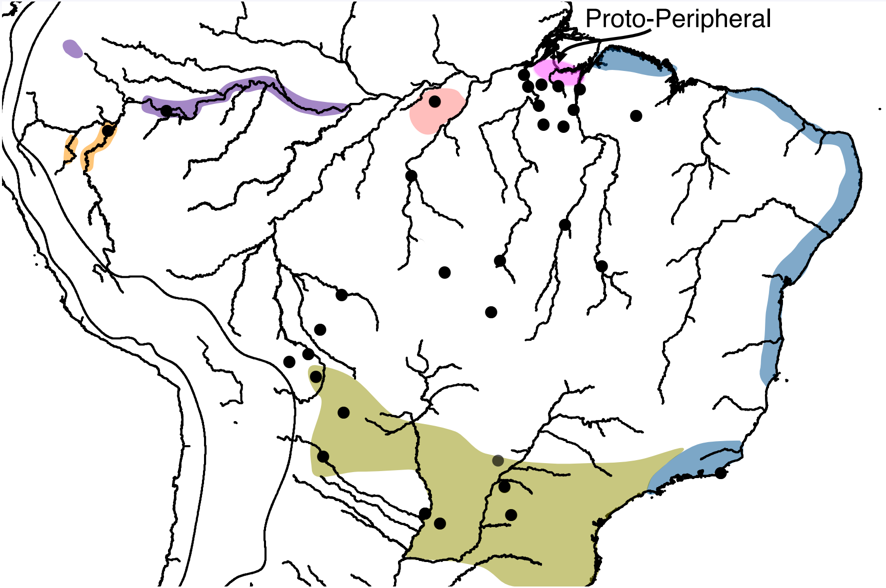
- Having inferred the PTG homeland, as well as homelands for several important daughter nodes, we can reverse our account to yield a migratory model:
  1. PTG $\rightarrow$ PNTG + kam
  2. PNTG $\rightarrow$ Proto-Central + Proto-Tocantins + Proto-Peripheral
  3. Proto-Peripheral $\rightarrow$ Proto-eme-way + Proto-kay-part + Proto-Diasporic

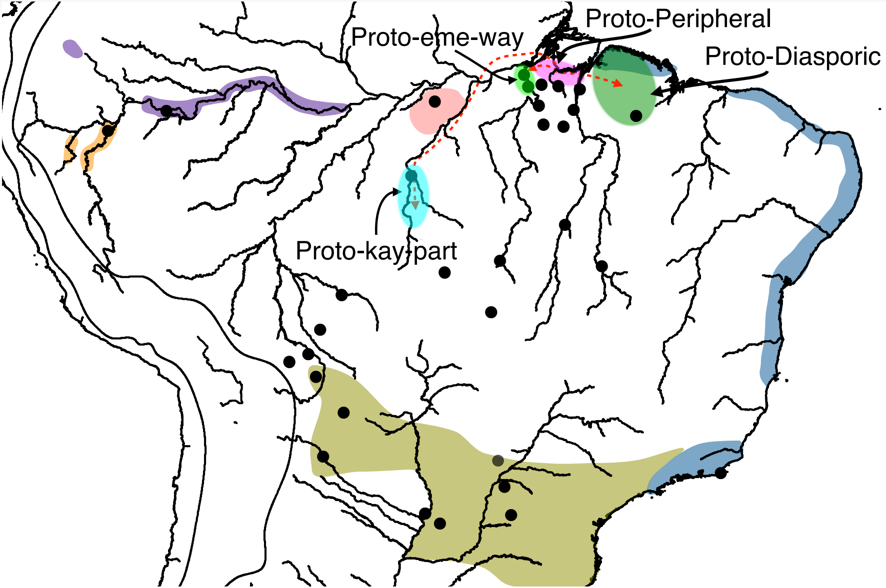

PTG

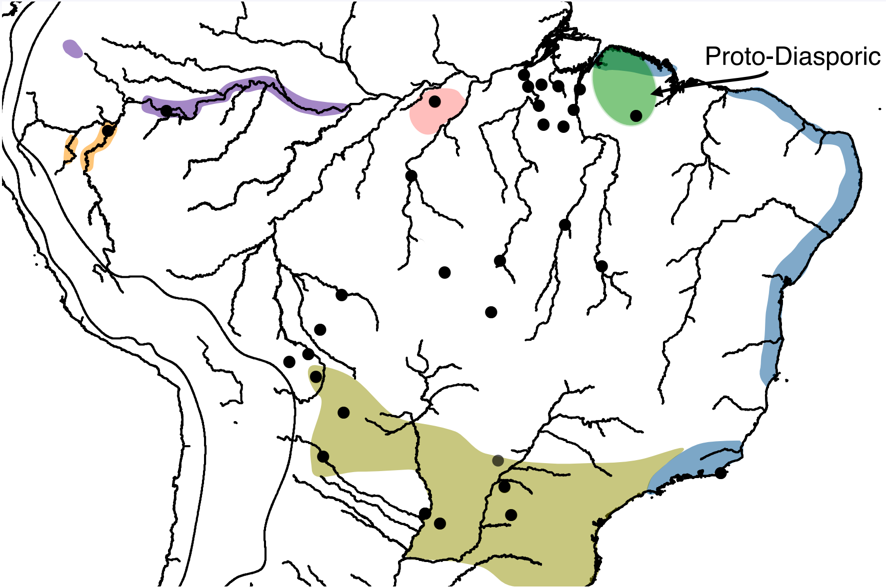

PNTG
PTG

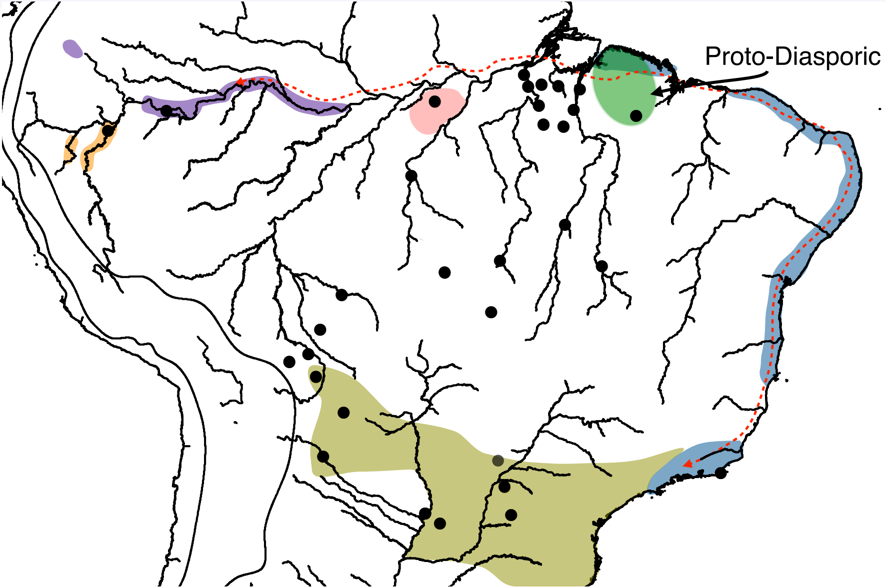

PNTG

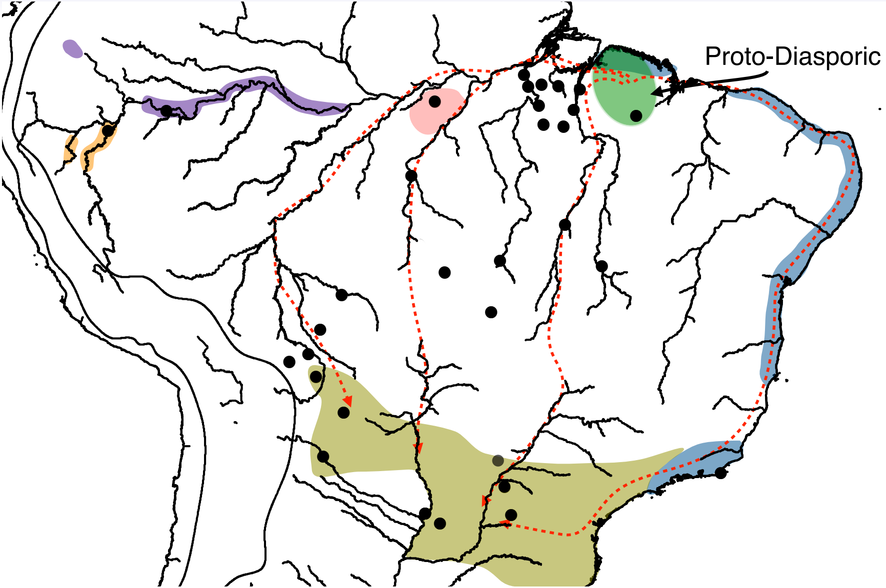

Proto-Peripheral
Proto-Central
PNTG
Proto-Tocantins


Proto-Peripheral


Proto-Peripheral
Proto-eme-way
Proto-Diasporic
Proto-kay-part


Proto-Diasporic


Proto-Diasporic


Proto-Diasporic

Proto-Southern Migrations

- It is unclear which route or routes were taken by Proto-Southern to arrive in the greater Paraná drainage, where we assume it diversified
- Three routes are in principle possible:
    1. Tocantins/Araguaia: this is the route of shortest distance
    2. Tapajós
    3. Madeira: this is appealing based on the presence of Southern languages in what is now Bolivia
- It is noteworthy that migrations up the Tapajós or Madeira require Proto-Southern to traverse territory previously traversed by speakers of Proto-Omagua-Kokama

Previous Proposals for PT(G) Homelands and Migrations

- Previous proposals for Proto-Tupí and Proto-Tupí-Guaraní homelands and migrations are numerous (Noelli 1996:11-25)
- We review a set of prominent and more recent proposals, by archaeologists, anthropologists, and linguists alike
  - Lathrap (1970)
  - Brochado (1984)
  - Urban (1992)
  - Rodrigues (2000)
- Archaeological claims rely **heavily** on pottery traditions
- Much work only considers the geographical spread and pottery traditions of the Tupinambá and Guaraní, ignoring other Tupí-Guaraní groups

T(G) Homeland and Migration Proposal

- Lathrap (1970:78-79): PTG spoken at mouth of Amazon
  - Spread began ∼500BC, up Madeira, Xingú, Tocantins, and down the Atlantic coast
- Brochado (1984): 'Two-Pronged Hypothesis' (Urban 1996:62)
  - PTG spoken on the Amazon proper
  - Guaraní migrate up the Madeira (∼200BC) and reach the Paraná-Paraguay basin by ∼100AD
  - Tupinambá migrate down the Atlantic coast by ∼800AD
- Comparison: Homeland and spread broadly compatible with the model presented here

T(G) Homeland and Migration Proposal (Urban 1992)

- PTG spoken in Madeira-Xingú headwaters, where it diversified
- Wave 1: Linguistically most divergent groups split off first
  - Omagua and Kokama-Kokamilla migrated towards the Amazon
  - Aché migrated southward into Paraguay
  - Siriono migrated to the southwest into Bolivia
- Wave 2: Amazonian TG languages split off
  - Pauserna and Kawahib migrate west
  - Kayabí and Kamaiurá migrate to the Xingú
  - Xetá migrate to southern Brazil
  - Tapirapé and Tenetehara migrate to the Tocantins and descend to near the mouth of the Amazon
  - Wayampí precede the Tapirapé and Tenetehara, crossing the Amazon into French Guiana (known to not be a prehistoric migration)

T(G) Homeland and Migration Proposal (Urban 1992)

- Wave 3 (∼1000AD): remaining non-Amazonian languages split off
  - Chiriguano and Bolivia in Bolivia
  - Tapiete and Guaraní in Paraguay
  - Kaiowá in Argentine-Brazilian-Paraguayan border region
  - Tupinambá along Atlantic coast
- Comparison: Homeland and migration model significantly at odds with the homeland, internal classification, and migration model presented here

Archaeological Observations

- Noelli (1998:656; see also Noelli (1996, 2008)):

  . . . [W]here occupation sequences are known, confronting the archaeological publications will rule out Paraguay, southern Bolivia, Mato Grosso do Sul, Goiás, southern, southeastern and northeastern Brazil as a centre of origin. In the upper and main course of the Xingu, in the Araguaia and in the upper and main course of the Tocantins, ... no archaeological evidence identifies an origin there...

- Leaves viable the lower Tocantins and Xingú and their associated interfluvial zone (our suggested homeland)

T(G) Homeland and Migration Proposal (Rodrigues 2000)

- Proto-Tupí-Guaraní diversifies in the Juruena-Arinos interfluvium
- Wave 1: II & III split off, migrating southward
  - II maintains contact with I
  - II & III then each split in two, with one branch of each remaining in contact with each other
  - II heads (north)west into Bolivia in two migrations
  - III heads (north)east to the Atlantic in two migrations
- Wave 2: I splits off, migrating further southward than did II & III
- Comparison: The early migrations of II and III, and later, I, are difficult to reconcile with their deep position in our proposed tree

Conclusion

- Our model posits a PTG homeland that spans the lower Tocantins and Xingú Rivers, with out-migrations from this region
- Major migrations are associated with the Diasporic branch:
  - Proto-Omagua-Kokama up the Amazon, Proto-Tupinambá south along the Atlantic coast
  - Southern towards the Paraná River basin, up along the Tocantins/Araguaia
- The PTG homeland we propose
  - Largely coincides with the homeland near the mouth of the Amazon discussed by Lathrap (1970) and Brochado (1984) and is not contradicted by available archaeological evidence
  - But is placed much further north and east that the homelands proposed by Urban (1992) Rodrigues (2000)
- The classification of Chousou-Polydouri et al. (2014) poses significant challenges for Urban (1992) and Rodrigues (2000)

Acknowledgements

- Diamantis Sellis for crucial computational assistance:
  - automated binary coding
  - cognate set completeness and consistency checking scripts
- The following colleagues for generously sharing primary data:
  - Sebastian Drude (Awetí)
  - Sérgio Meira (Mawé, Tembé)
  - Françoise Rose (Emerillon)
  - Eva-Maria Rößler (Aché)
  - Rosa Vallejos (Kokama-Kokamilla)
- Tammy Stark for GIS assistance
- Noé Gasparini for access to Anambé and Yuki data
- And the following Berkeley TG group alumni:
  - Mike Roberts
  - Vivian Wauters
- NSF DEL Award #0966499
- UC Berkeley Social Science Matrix 2013-2014 Grant

References I

Balée, William. 1994. *Footprints of the Forest: Ka'apor Ethnobotany – The Historical Ecology of Plant Utilization by an Amazonian People*. Columbia University Press.

Bouckaert, Remco; Philippe Lemey; Michael Dunn; Simon J. Greenhill; Alexander V. Alekseyenko; Alexei J. Drummond; Russell D. Gray; Marc A. Suchard; and Quentin D. Atkinson. 2012. Mapping the Origins and Expansion of the Indo-European Language Family. *Science* 337(6097):957–960.

Bowern, Claire and Quentin D. Atkinson. 2012. Computational Phylogenetics and the Internal Structure of Pama-Nyungan. *Language* 88(4):817–845.

Brochado, José P. 1984. *An Ecological Model of the Spread of Pottery and Agriculture into Eastern South America*. PhD dissertation, University of Illinois Urbana-Champaign.

Campbell, Lyle. 1997. *American Indian Languages: The Historical Linguistics of Native America*. New York: Oxford University Press.

Chousou-Polydouri, Natalia; Zachary O'Hagan; Keith Bartolomei; Erin Donnelly; and Lev Michael. 2014. *An Internal Classification of Tupí-Guaraní Using Computational Phylogenetic Methods*. Belém: Talk at AMAZONICAS V, May 28.

Corrêa da Silva, Beatriz Carretta. 2007. Mais fundamentos para a hipótese de Rodrigues (1984/1985) de um proto-awetí-tupí-guaraní. *Línguas e Culturas Tupí*, edited by Ana Suelly Arruda Câmara Cabral and Aryon Dall'Igna Rodrigues, Campinas: Editora Curt Nimuendajú, 219–239.

Corrêa da Silva, Beatriz Carretta. 2010. *Mawé/awetí/tupí-guaraní: Relações lingüísticas e implicações históricas*. PhD dissertation, Universidade de Brasília.

References II

Diebold Jr., A. Richard. 1960. Determining the Centers of Dispersal of Language Groups. *International Journal of American Linguistics* 26(1):1–10.

Dietrich, Wolf. 1990. *More Evidence for an Internal Classification of Tupí-Guaraní Languages*. Berlin: Gebr. Mann Verlag.

Drude, Sebastian. 2006. On the Position of the Awetí Language in the Tupí Family. *Guaraní y Mawetí-Tupí-Guaraní: Estudos históricos y descriptivos sobre una familia lingüística de América del Sur*, edited by Wolf Dietrich and Haralambos Symeonidis, Berlin: LIT Verlag, 11–45.

Drude, Sebastian. 2011. Awetí in Relation with Kamayurá: The Two Tupian Languages of the Upper Xingu. *Alto Xingu: Uma Sociedade Multilíngue*, edited by Bruna Franchetto, Rio de Janeiro: Museu do Índio; Fundação Nacional do Índio (FUNAI), 155–191.

Dyen, Isidore. 1956. Language Distribution and Migration Theory. *Language* 32(4):611–626.

Forster, Peter and Alfred Toth. 2003. Toward a Phylogenetic Chronology of Ancient Gaulish, Celtic, and Indo-European. *Proceedings of the National Academy of Sciences*, vol. 100, vol. 100, 9079–9084.

Galúcio, Ana Vilacy; Sérgio Meira; Sebastian Drude; Nilson Gabas Jr.; Denny Moore; Gessiane Picanço; Carmen Reis Rodrigues; and Luciana Storto. 2013. Genetic Relationship and Degree of Relatedness within the Tupi Linguistic Family: A Lexicostatistical and Phylogenetic Approach. *ms.* .

Gray, Russell D. and Quentin D. Atkinson. 2003. Language-tree divergence time support the Anatolian theory of Indo-European origin. *Nature* 426:435–439.

References III

Gray, Russell D.; Alexei J. Drummond; and Simon J. Greenhill. 2009. Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science* 323(5913):479–483.

Greenhill, Simon J. and Russell D. Gray. 2005. Testing Population Dispersal Hypotheses: Pacific Settlement, Phylogenetic Trees and Austronesian Languages. *The Evolution of Cultural Diversity: Phylogenetic Approaches*, edited by Ruth Mace; Clare J. Holden; and Stephen Shennan, London: UCL Press, 31–52.

Greenhill, Simon J. and Russell D. Gray. 2009. Austronesian Language Phylogenies: Myths and Misconceptions about Bayesian Computational Methods. *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, edited by Alexander Adelaar and Andrew Pawley, Canberra: Pacific Linguistics, 1–23.

Greenhill, Simon J.; Alexei J. Drummond; and Russell D. Gray. 2010. How accurate and robust are the phylogenetic estimates of Austronesian language relationships? *PLoS ONE* 5(3):e9573.

Grenand, Pierre. 1982. *Ainsi parlaient nos ancêtres*. Paris: ORSTOM.

Jensen, Cheryl. 1999. Tupí-Guaraní. *The Amazonian Languages*, edited by R. M. W. Dixon and Alexandra Y. Aikhenvald, Cambridge: Cambridge University Press, 125–163.

Kamaiurá, Warý. 2012. *Awetí e tupí-guaraní: Relações genéticas e contato lingüístico*. MA thesis, Universidade de Brasília.

Kaufman, Terrence. 1994. The Native Languages of South America. *Atlas of the World's Languages*, edited by C. Mosley and R.E. Asher, London/New York: Routledge, 46–76.

References IV

Kaufman, Terrence. 2007. South America. *Atlas of the World's Languages*, edited by R. E. Asher and Christopher Moseley, London/New York: Routledge, 61–93, 2 edn.

Lathrap, Donald W. 1970. *The Upper Amazon*. New York: Praeger.

Lemle, Miriam. 1971. Internal Classification of the Tupí-Guaraní Linguistic Family. *Tupí Studies I*, edited by David Bendor-Samuel, Summer Institute of Linguistics (SIL), 107–129.

Lemos Barbosa, Antônio. 1951. *Pequeno vocabulário tupí-portugues*. Rio de Janeiro: Livraria São José.

Mello, Antônio Augusto Souza. 2000. *Estudo histórico da família lingüística tupí-guaraní: Aspectos fonológicos e lexicais*. PhD dissertation, Universidade Federal de Santa Catarina.

Mello, Antônio Augusto Souza. 2002. Evidencias fonológicas e lexicais para o sub-agrupamento interno tupí-guaraní. *Línguas Indígenas Brasileiras: Fonologia, Gramática e História*, vol. 1, edited by Ana Suelly Arruda Câmara Cabral and Aryon Dall'Igna Rodrigues, Belém: Editora Universitária, Universidade Federal do Pará, 338–342.

Nakhleh, Luay; Don Ringe; and Tandy Warnow. 2005. Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. *Language* 81(2):382–420.

Nichols, Johanna. 1997. Modeling Ancient Population Structures and Movement in Linguistics. *Annual Review of Anthropology* 26:359–384.

Noelli, Francisco S. 1996. As hipóteses sobre o centro de origem e rotas de expansao dos Tupi. *Revista de Antropologia* 39(2):7–53.

References V

NOELLI, FRANCISCO S. 1998. The Tupi: Explaining origin and expansions in terms of archaeology and of historical linguistics. *Antiquity* 72(277):648–663.

NOELLI, FRANCISCO S. 2008. The Tupi Expansion. *The Handbook of South American Archeology*, edited by Helaine Silverman and William H. Isbell, New York: Springer, 659–670.

RINGE, DON; TANDY WARNOW; and ANN TAYLOR. 2002. Indo-European and Computational Cladistics. *Transactions of the Philological Society* 100(1):59–129.

RODRIGUES, ARYON DALL'IGNA. 1984/1985. Relações internas na família lingüística tupí-guaraní. *Revista de Antropologia* 27/28:33–53.

RODRIGUES, ARYON DALL'IGNA. 1986. *Línguas brasileiras: Para o conhecimento das línguas indígenas*. São Paulo: Edições Loyola.

RODRIGUES, ARYON DALL'IGNA. 1999. Tupí. *The Amazonian Languages*, edited by R. M. W. Dixon and Alexandra Y. Aikhenvald, Cambridge: Cambridge University Press, 107–124.

RODRIGUES, ARYON DALL'IGNA. 2000. Hipótese sobre as migrações dos três subconjuntos meriodionais da família tupí-guaraní. *II Congresso Nacional da ABRALIN e XIV Instituto Lingüístico*, Florianópolis: Associação Brasileira de Lingüística, 1596–1605.

RODRIGUES, ARYON DALL'IGNA and ANA SUELLY ARRUDA CÂMARA CABRAL. 2002. Revendo a classificação interna da família tupí-guaraní. *Línguas Indígenas Brasileiras: Fonologia, Gramática e História*, edited by Ana Suelly Arruda Câmara Cabral and Aryon Dall'Igna Rodrigues, Belém: Editora Universitária, Universidade Federal do Pará, 327–337.

References VI

RODRIGUES, ARYON DALL'IGNA and ANA SUELLY ARRUDA CÂMARA CABRAL. 2012. Tupian. *The Indigenous Languages of South America: A Comprehensive Guide*, Berlin: De Gruyter Mouton, 495–574.

RODRIGUES, ARYON DALL'IGNA and WOLF DIETRICH. 1997. On the Linguistic Relationship Between Mawé and Tupí-Guaraní. *Diachronica* 14(2):265–304.

SAPIR, EDWARD. [1916]1949. Time Perspective in Aboriginal American Culture: A Study in Method. *Selected Writings of Edward Sapir*, edited by David G. Mandelbaum, Berkeley: University of California Press, 389–467.

SEKI, LUCY. 2000. *Gramática do kamaiurá: Língua tupí-guaraní do alto Xingu*. Campinas: Editora da UNICAMP.

URBAN, GREG. 1992. A história da cultura brasileira segundo as línguas nativas. *História dos Índios no Brasil*, edited by Carneiro da Cunha, São Paulo: FAPESP/SMC/Cia das Letras, 87–102.

URBAN, GREG. 1996. On the geographical origins and dispersion of Tupian languages. *Revista de Antropologia* 39(2):61–104.

WALKER, ROBERT S.; SØREN WICHMANN; THOMAS MAILUND; and CURTIS J. ATKISSON. 2012. Cultural Phylogenetics of the Tupi Language Family in Lowland South America. *PLoS ONE* 7(4):e35,025.

WARNOW, TANDY; STEVEN N. EVANS; DON RINGE; and LUAY NAKHLEH. 2004. Stochastic Models of Language Evolution and an Application to the Indo-European Family of Languages. *Technical Report, Department of Statistics, University of California, Berkeley*.

Walker et al. (2012)

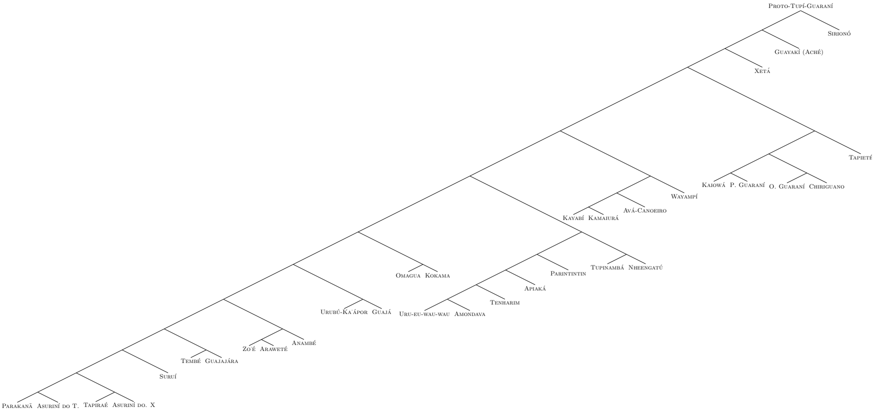


Figure 1: Tupí-Guaraní Subgrouping (Walker et al. 2012)