

# Comprehensive database of sound changes reveals tree-like and wave-like processes in the Mixtec language family

Sandra Auderset

University of Bern

&

Eric W. Campbell

University of California Santa Barbara

56<sup>th</sup> Annual Meeting of the  
Societas Linguistica Europaea  
National and Kapodistrian University of Athens  
29.08.2023

# The Mixtec (Tu'un Savi) languages of southern Mexico

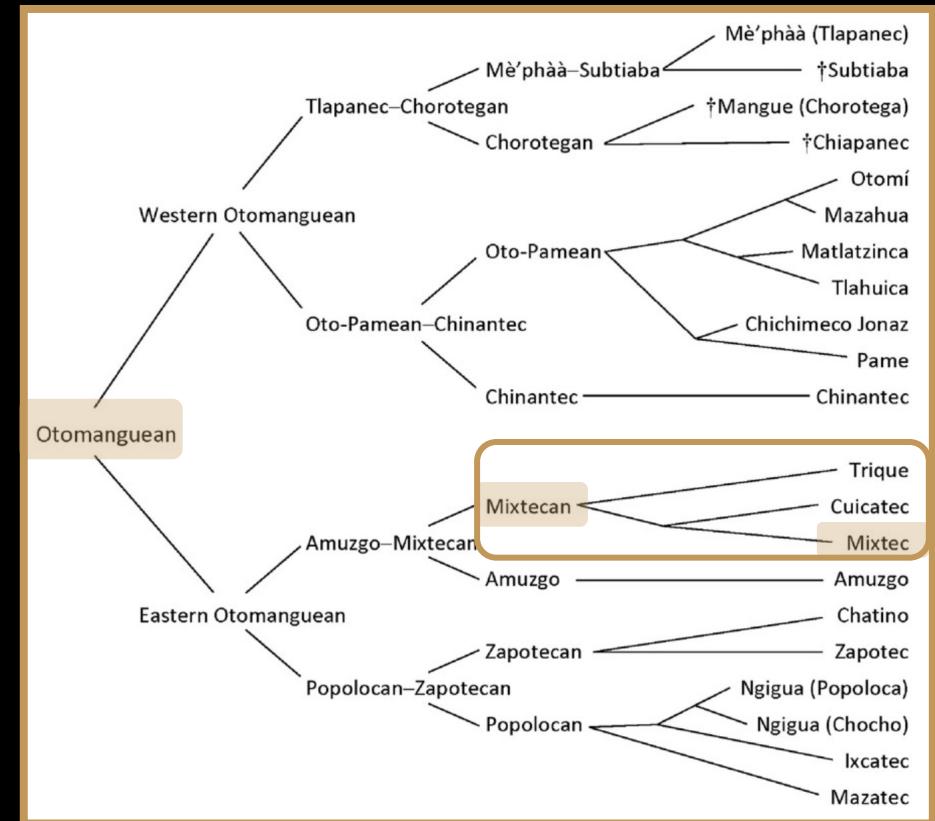
## Geography



Mapa 1. Las divisiones coloniales y contemporáneas del Nuu Savi\*.

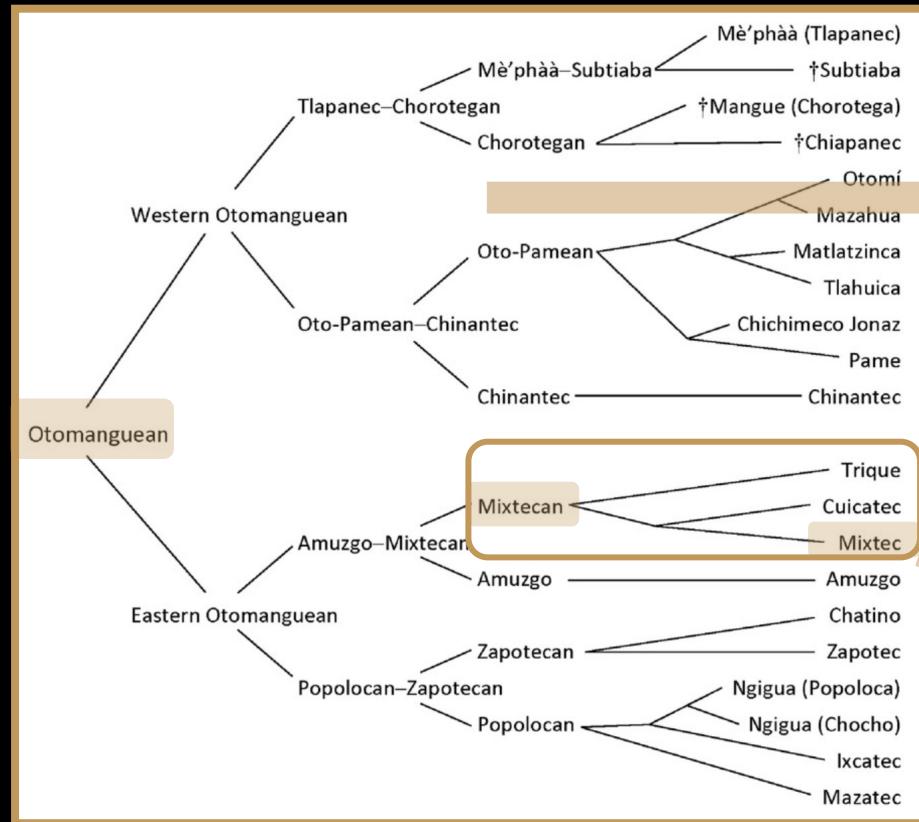
Aguilar Sánchez 2020:24. *Nuu Savi: Pasado, presente y futuro*. Diss U Leiden

## Genealogy

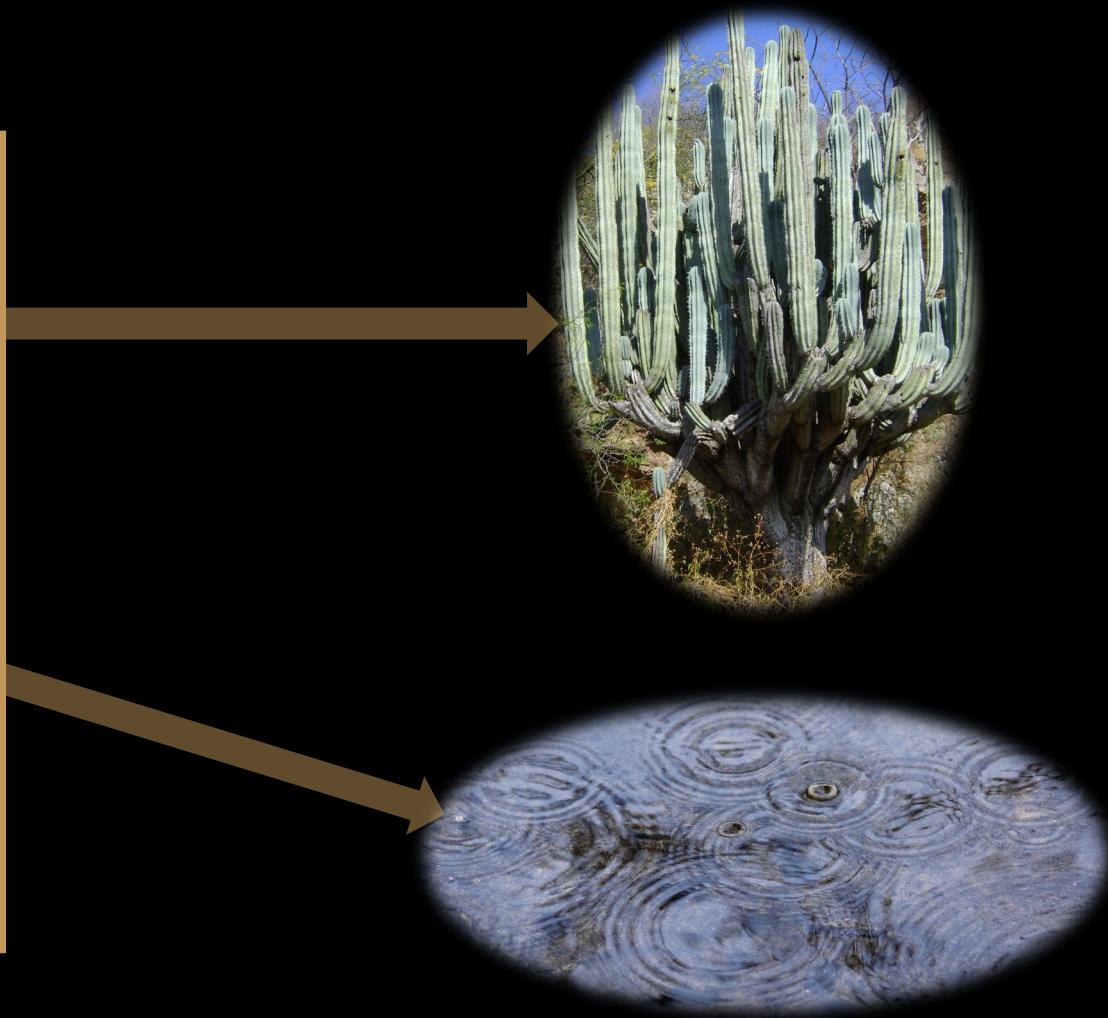


Campbell 2017 *Lang. Linguist. Compass* 11(7)  
(based on Kaufman 2006, 1988)

# Mixtec (Tu'un Savi) classification



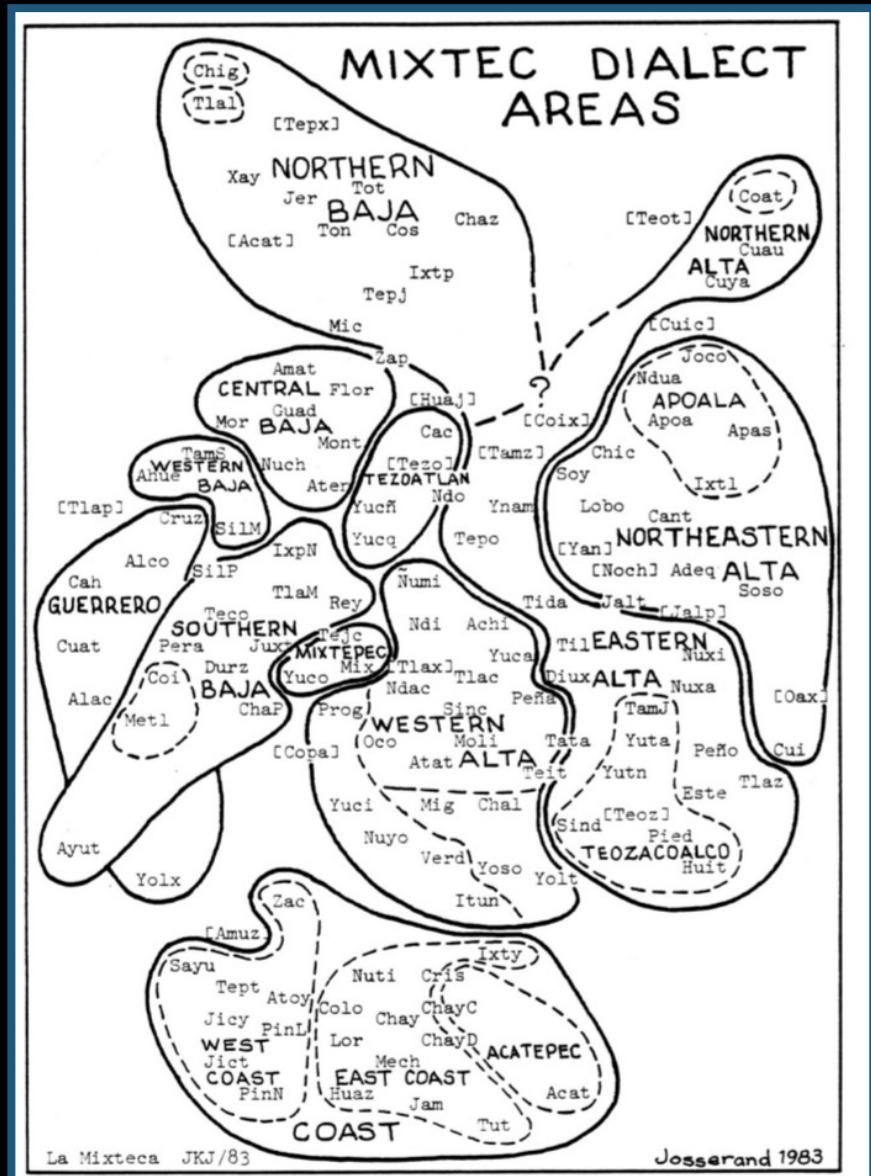
Campbell 2017. *Lang. Linguist. Compass* 11(7)  
(based on Kaufman 2006, 1988)



# The dialect area ‘problem’

*“El problema es la aparente violación o infracción de las expectativas ideales del método comparativo en la lingüística histórica (...). Pero este modelo presupone una efectiva separación de los grupos, después de un cambio que los divide. En contraste, en la Mixteca (...) vemos una situación más dinámica en el desarrollo de las familias lingüísticas.”*

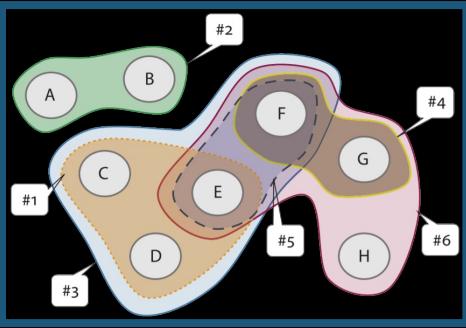
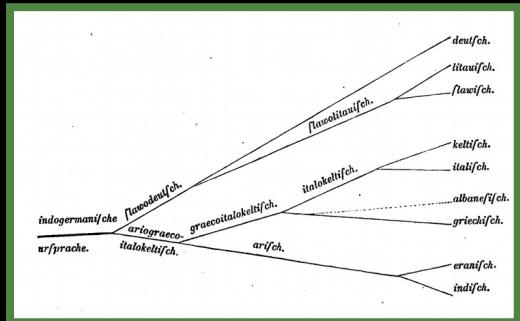
(Bradley & Josserand 1982: 303. *Anales de Antropol.* 19(2))



# Tackling sound change in Mixtec

comparative method:

- ▶ regular sound change
- ▶ language relationships & subgrouping
- ▶ inheritance vs. contact



‘dialect continuum/area/chains’:

- ▶ wave-like (not tree-like) patterns
- ▶ contact among related languages
- ▶ challenges to applying the comparative method (Ross & Durie 1996, Kalyan & François 2018)

sound change  
in Mixtec?

# Tackling sound change in Mixtec

Proto-Mixtec reconstructions  
and sound changes:

- ▶ Mak & Longacre 1960
- ▶ Bradley & Josserand 1982
- ▶ Josserand 1983
- ▶ Dürr 1987
- ▶ Swanton & Mendoza 2021
- ▶ Swanton 2021

# Tackling sound change in Mixtec

Proto-Mixtec reconstructions  
and sound changes:

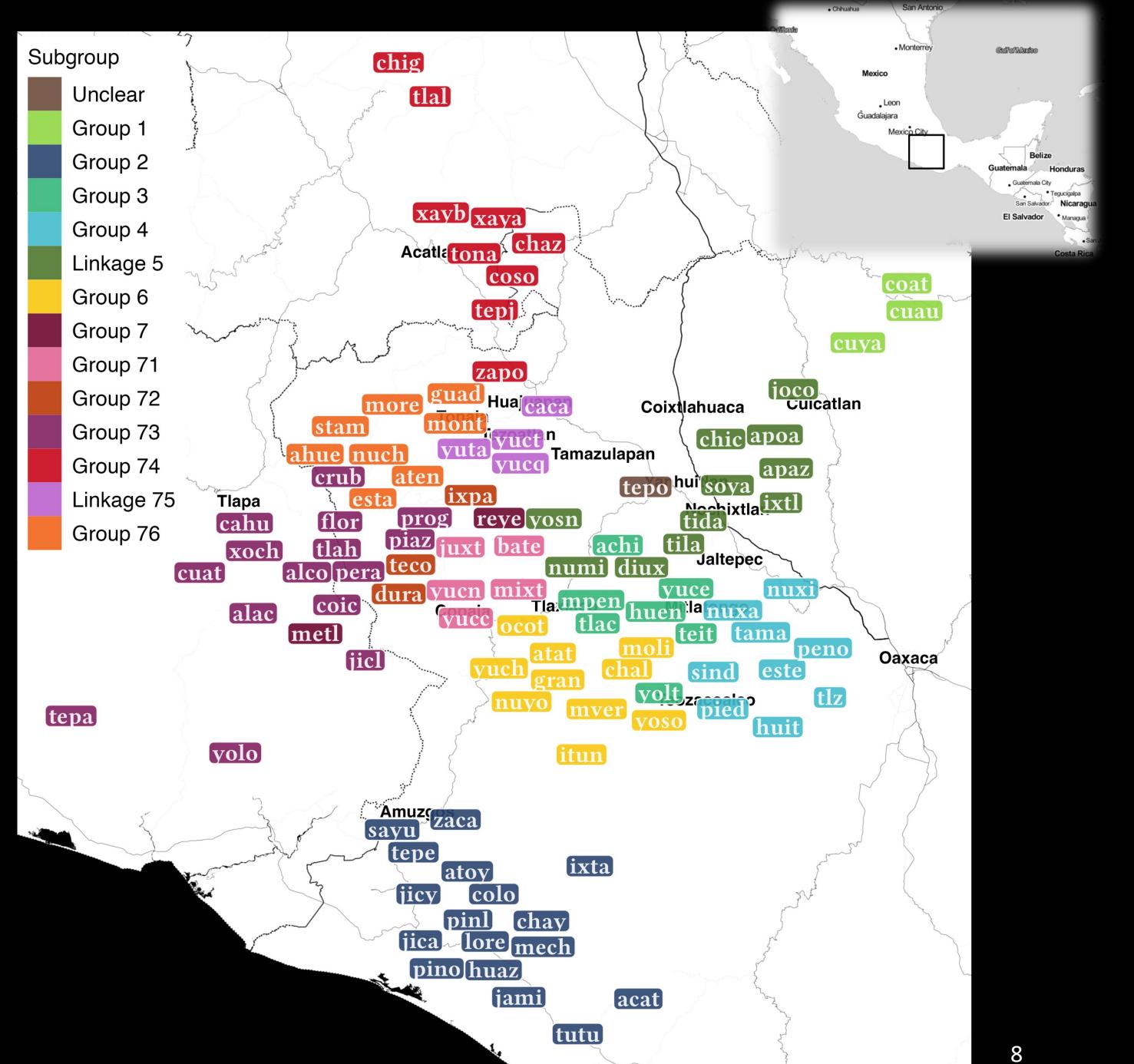
- ▶ Mak & Longacre 1960
- ▶ Bradley & Josserand 1982
- ▶ Josserand 1983
- ▶ Dürr 1987
- ▶ Swanton & Mendoza 2021
- ▶ Swanton 2021

Goal:

compile comprehensive, bottom-up data base of Mixtec sound change  
to  
investigate the patterns and distributions

# Sampled languages

105 varieties  
covering all subgroups\*



\*Auderset et al. 2023. *Journal of Lang. Evolution*

# Workflow – data base construction

AUTOTYP principles:

- ▶ autotypology
- ▶ modularity & connectivity
- ▶ separation of definition & coding files
- ▶ late aggregation

# Workflow

input: subset of annotated cognate sets in IPA from Auderset et al. 2023

ID	MEANING	DOCULECT	SOURCE_ORTHOGRAPHIC	TOKENS	COGID
2105	SALT	SanJuanDiuxiMixtec	ñii	j i 5 i 5	624
20187	SALT	SanMartinDuraznosMixtec	ñii	j i 1 i 1	624
6646	SALT	SantaMariaZacatepecMixtec	ñii	j ii	624
2127	RIVER	SanJuanDiuxiMixtec	yúté	z u 5 t e 5	607
19559	RIVER	SanMartinDuraznosMixtec	yitxa	z i 1 tç a 3	607
12814	RIVER	SantaMariaZacatepecMixtec	ju <sup>2</sup> tʃa <sup>21</sup>	j u 3 tʃ a 31	607

AUTOTYP principles:

- autotypology
- modularity & connectivity
- separation of definition & coding files
- late aggregation

cognates sets

establish sound correspondences, propose/revise proto-forms

COGID	MEANING	PMX	PMX_Josserand	JosserandID	PMX_Durr	DurrID
294	GRASS	*ite				
662	SLOW	*kʷeji	kʷeje	163		
607	RIVER	*jute	jute	23	jute	58
107	CHILI PEPPER	*jaʔa?			jaʔa?	12
624	SALT	*jiñ?	jiñ?	41		

proto-forms

# Workflow

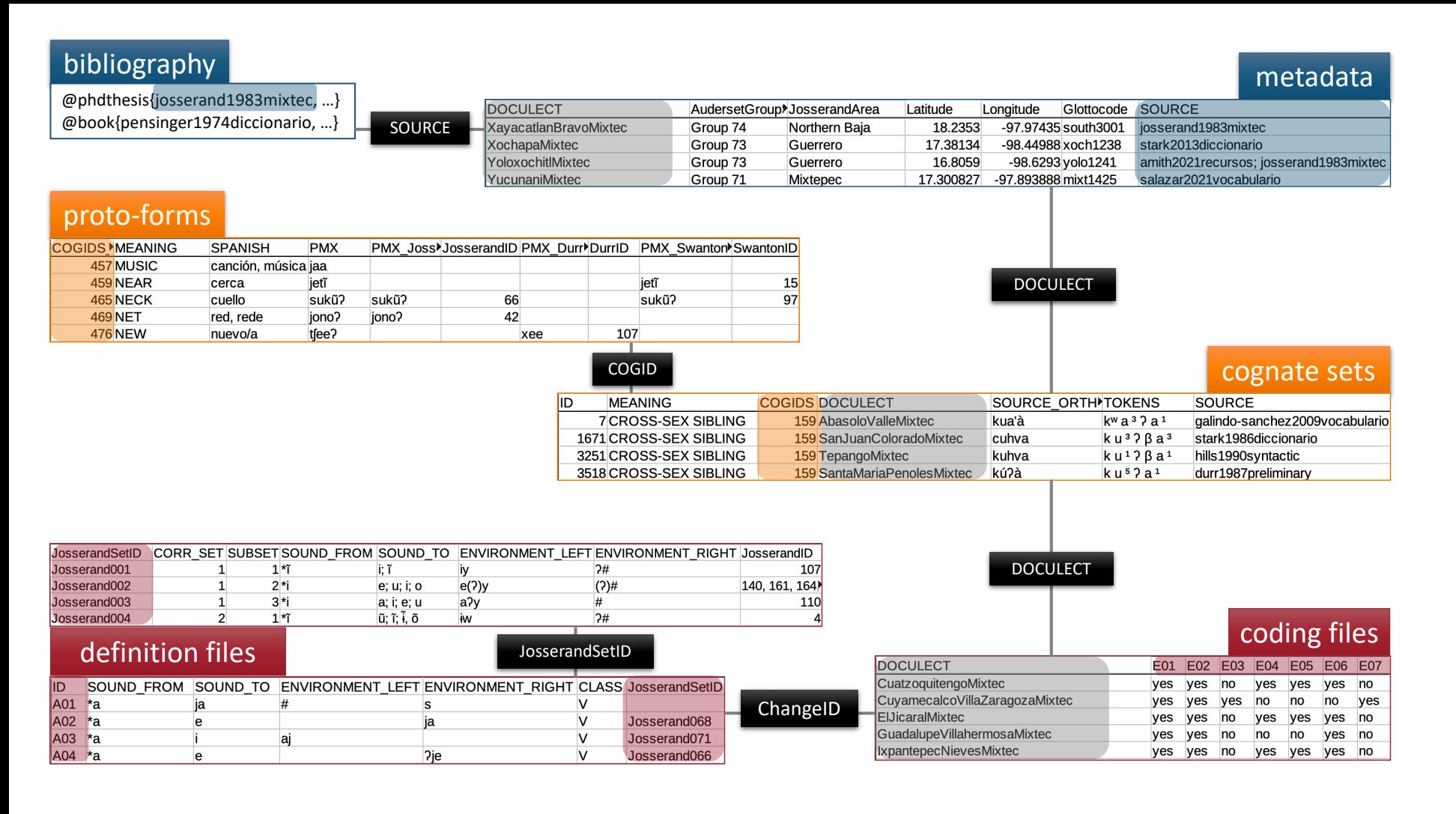
AUTOTYP principles:

- ▶ autotypology
- ▶ modularity & connectivity
- ▶ separation of definition & coding files
- ▶ late aggregation

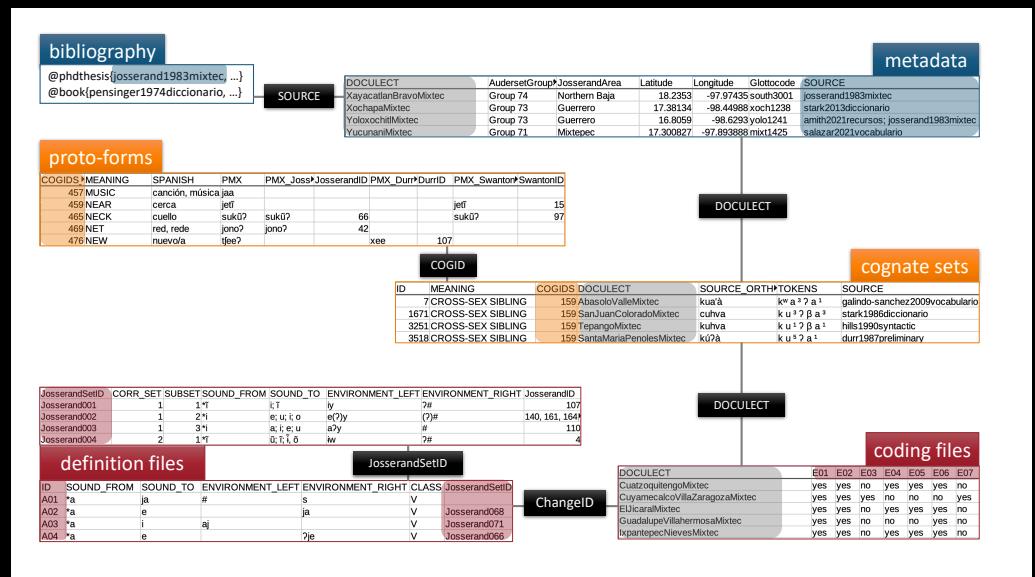
DOCULECT	E17	U32	Y01	J03	T12	T14	coding file
SanJuanDiuxiMixtec	no	no	no	yes	no	no	
SanMartinDuraznosMixtec	yes	yes	yes	yes	yes	no	
SantaMariaZacatepecMixtec	yes	no	yes	yes	no	yes	

ID	SOUND_FROM	SOUND_TO	ENVIRONMENT_LEFT	ENVIRONMENT_RIGHT	definition file
E17	*e	a	i,ut		
U32	*u	i	j	te	
Y01	*i	i		(?)#	
J03	*j	n	#	V(?)~	
T12	*t	tç		{i,e}	
T14	*t	tʃ		e	

# Structure of the data base

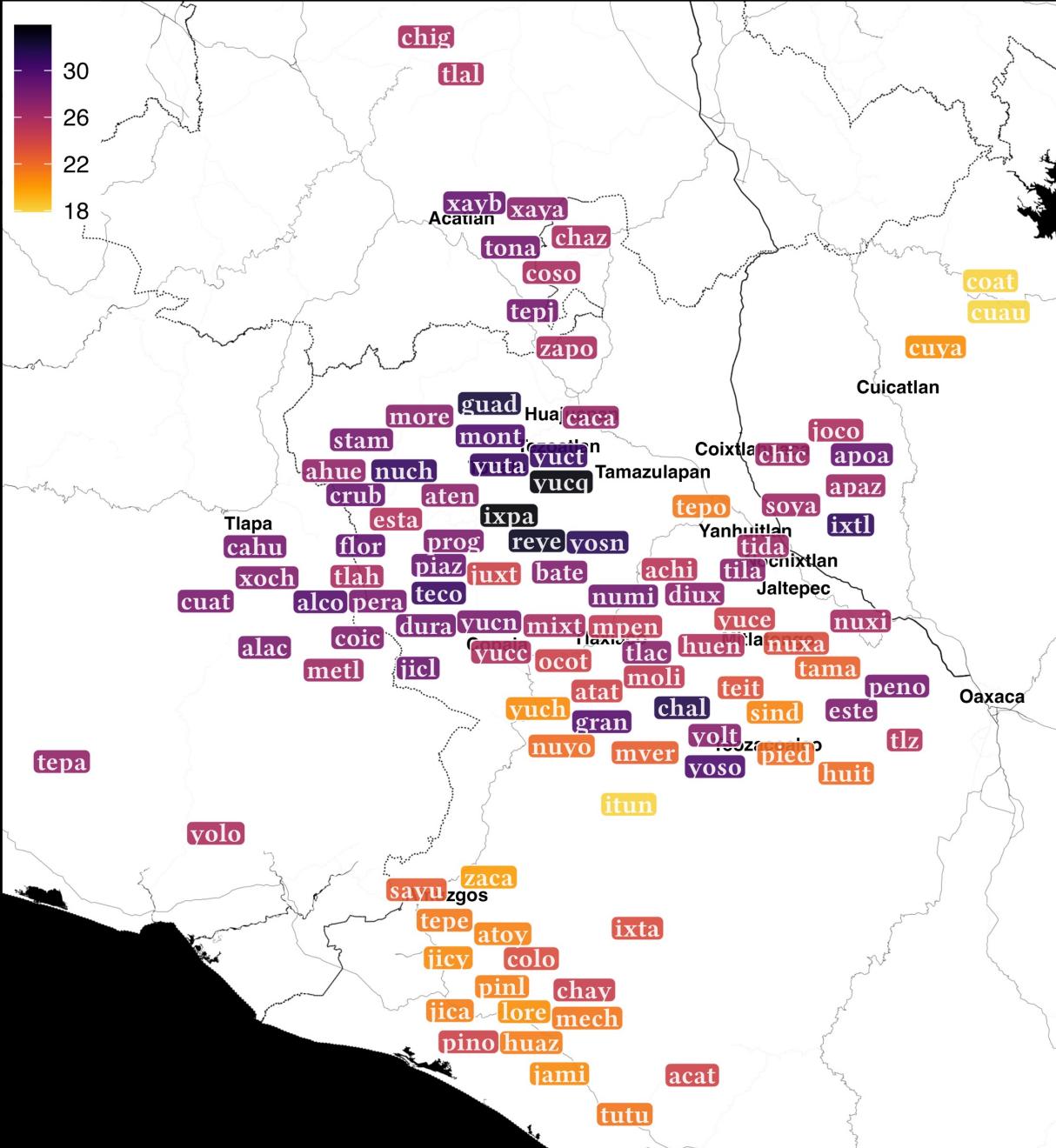


# Current state of the data base



- 105 varieties
- 256 reconstructed proto-forms
- > 14'000 lexical entries

- 245 sound changes identified
  - 227 used in analysis:
    - 110 vowel changes
    - 117 consonant changes
  - >25'000 data points (= coded sound changes)

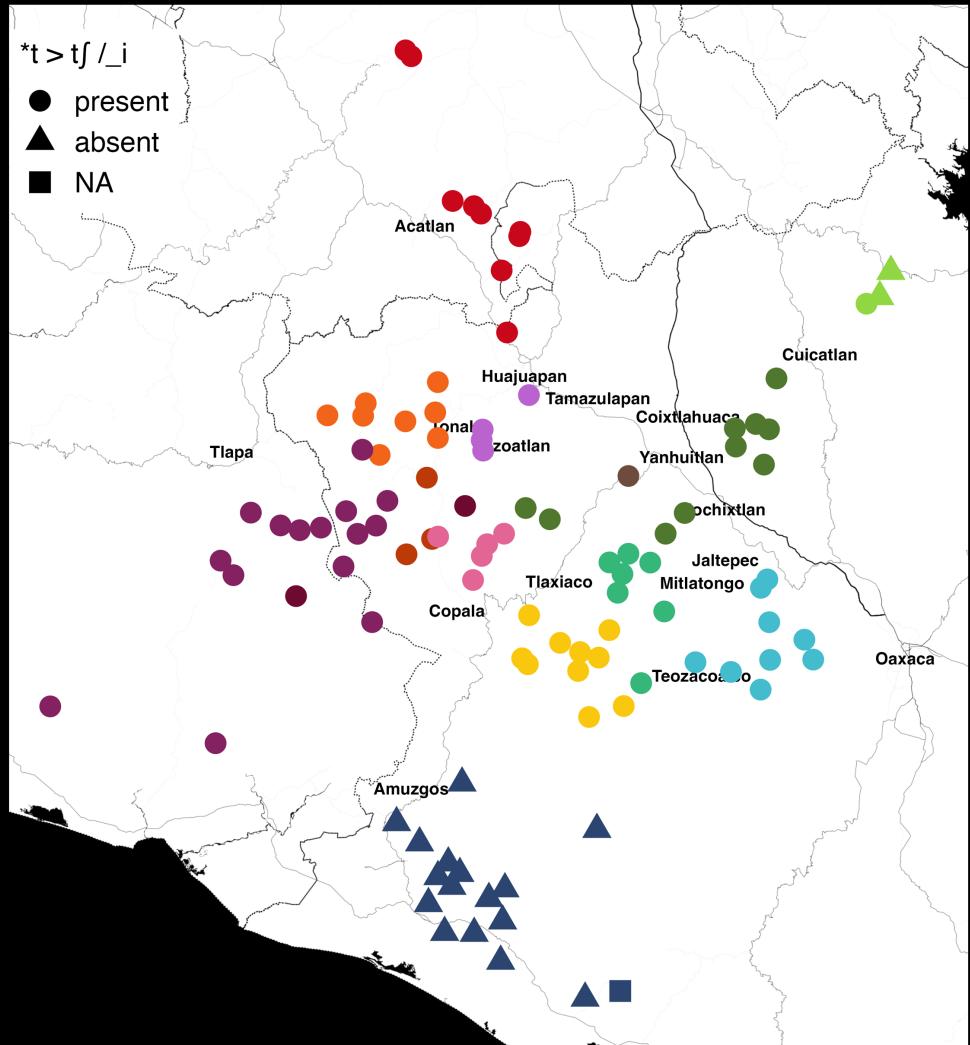


# Wave-like processes

Map displaying the percentage of changes present per variety calculated as:

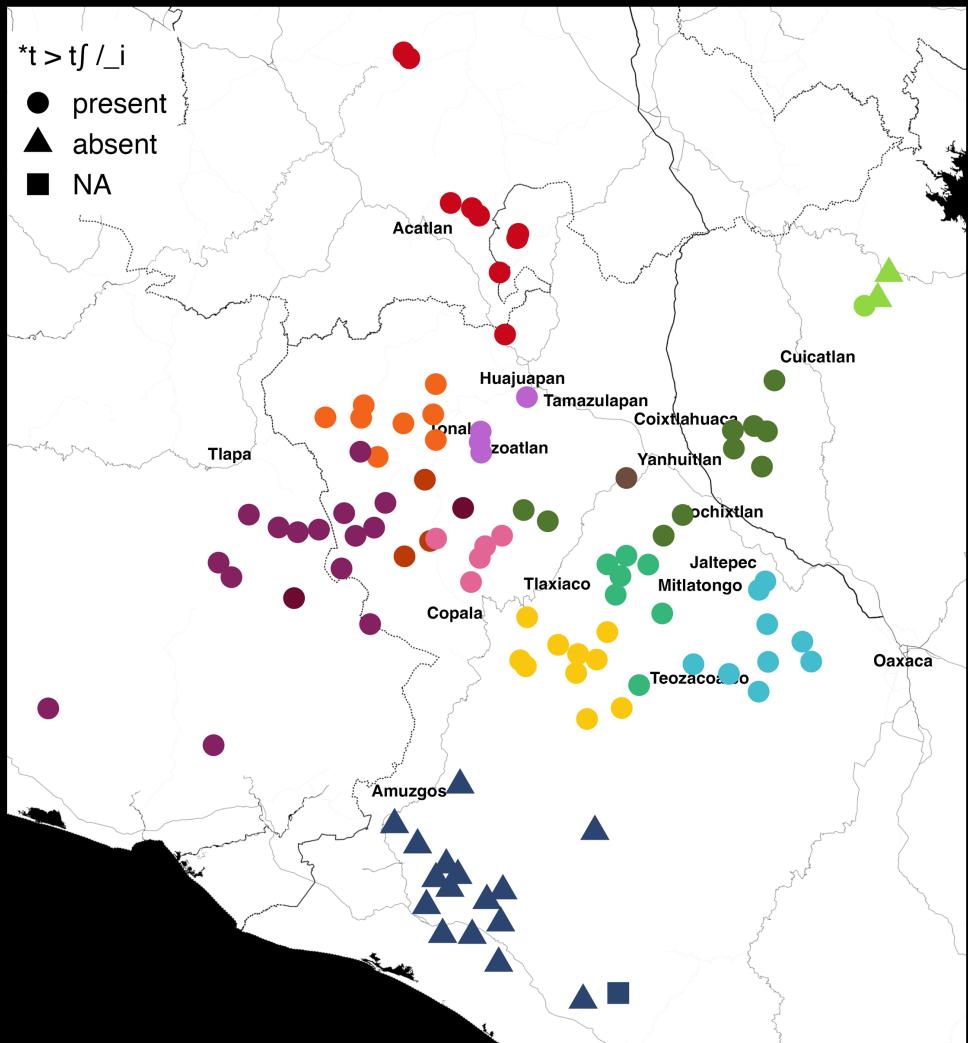
$$\frac{(\text{total number of changes}) - (\text{NA})}{(\text{changes present})}$$

# Tree-like processes

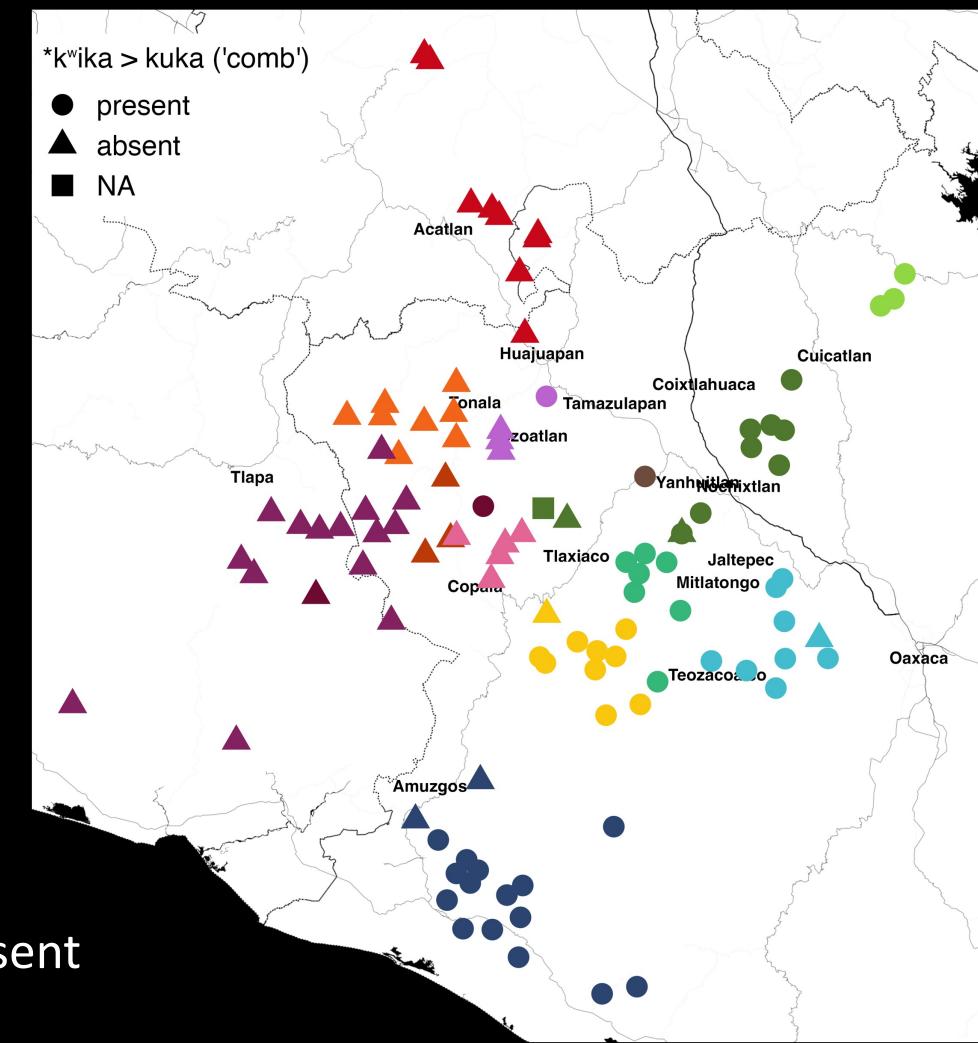


colors represent  
subgroups

# Tree-like processes



colors represent  
subgroups

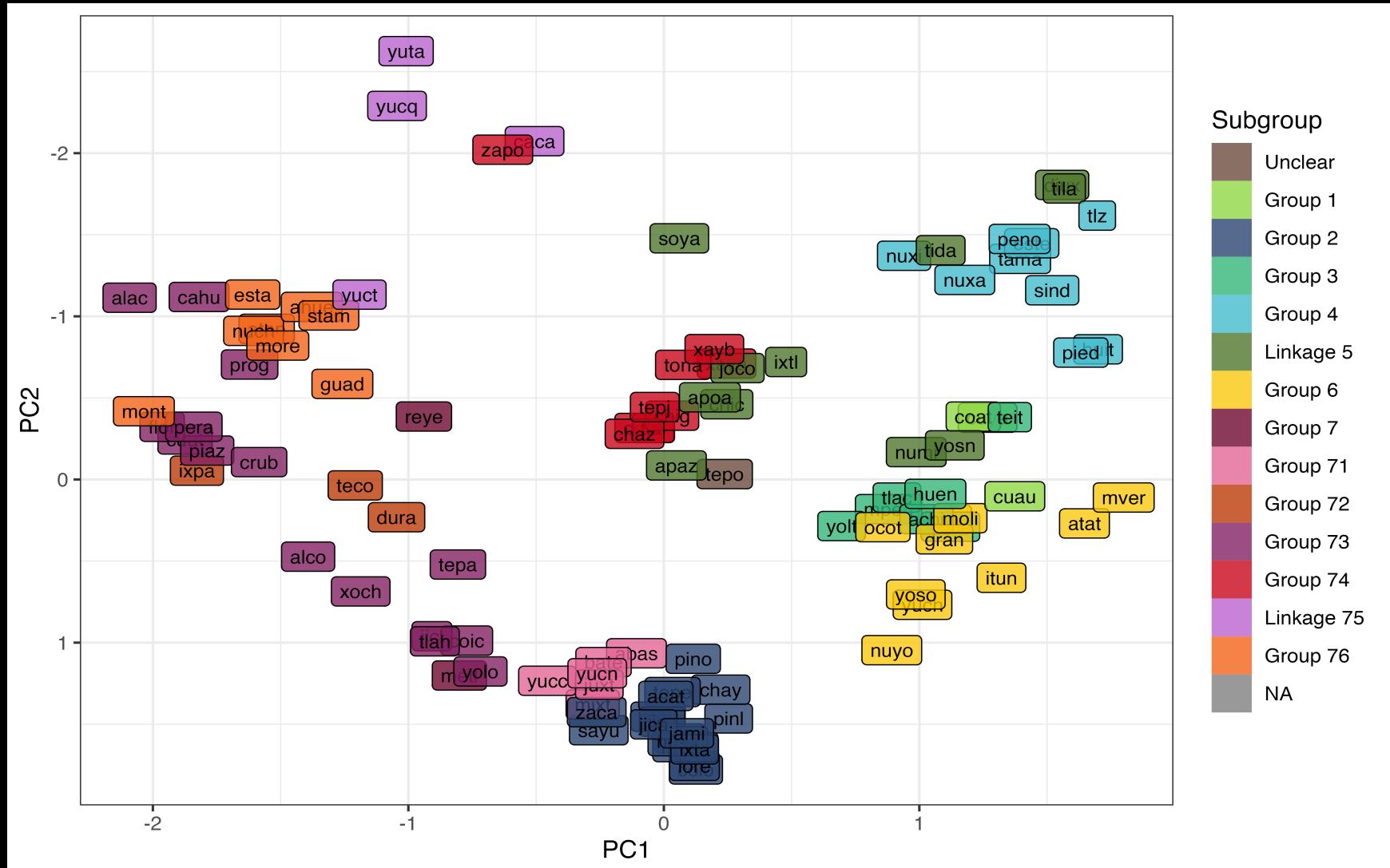


# Beyond trees and waves: PCA

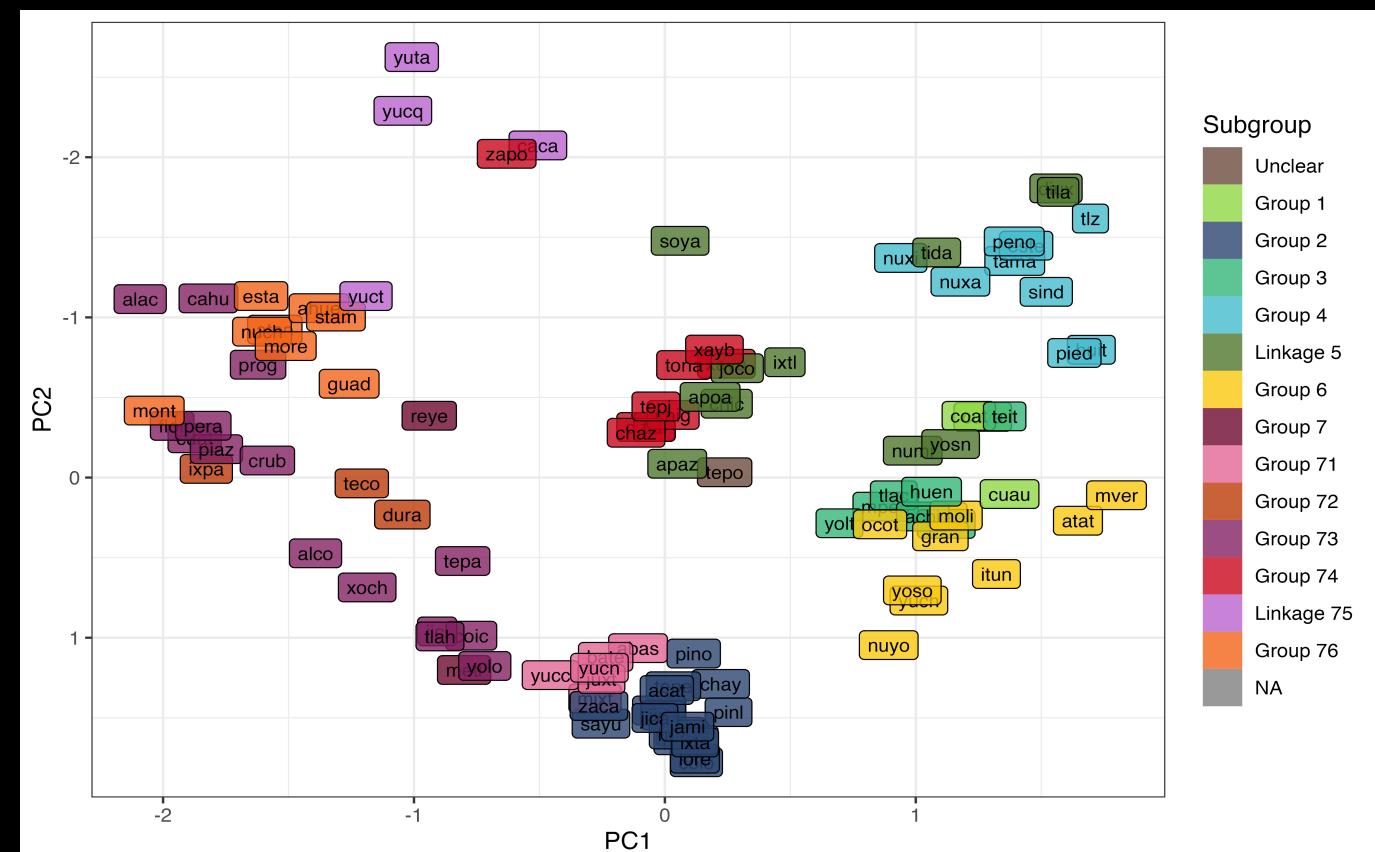
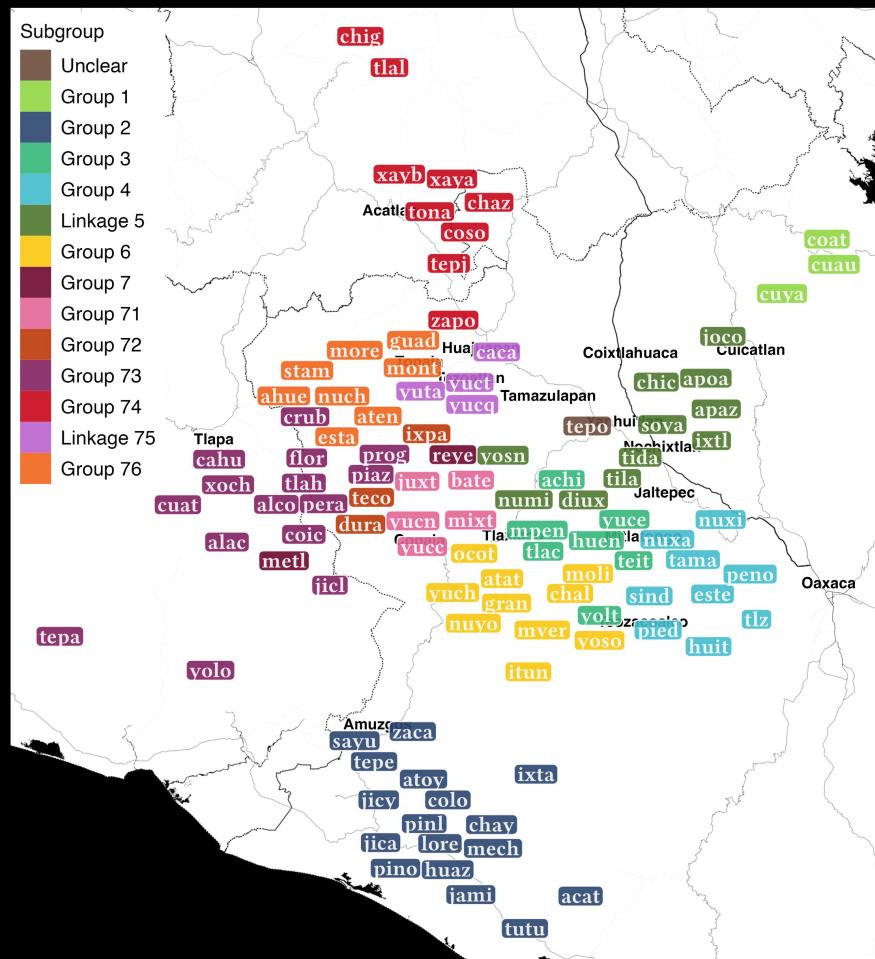
```

bpca calculated PCA
Importance of component(s):
          PC1    PC2
R2        0.1807 0.1205
Cumulative R2 0.1807 0.3011
227    Variables
105    Samples
1431   NAs ( 6.004 %)

```



# Beyond trees and waves



comprehensive database of sound changes coded in a bottom-up, fine-grained way reveals both tree-like and wave-like processes in the Mixtec language family

#### Limitations and future work:

- ▶ tone change
- ▶ relative chronology
- ▶ correlations between changes ~ chain shifts
- ▶ generalizations
- ▶ revision and expansion of the data base with new cognate sets

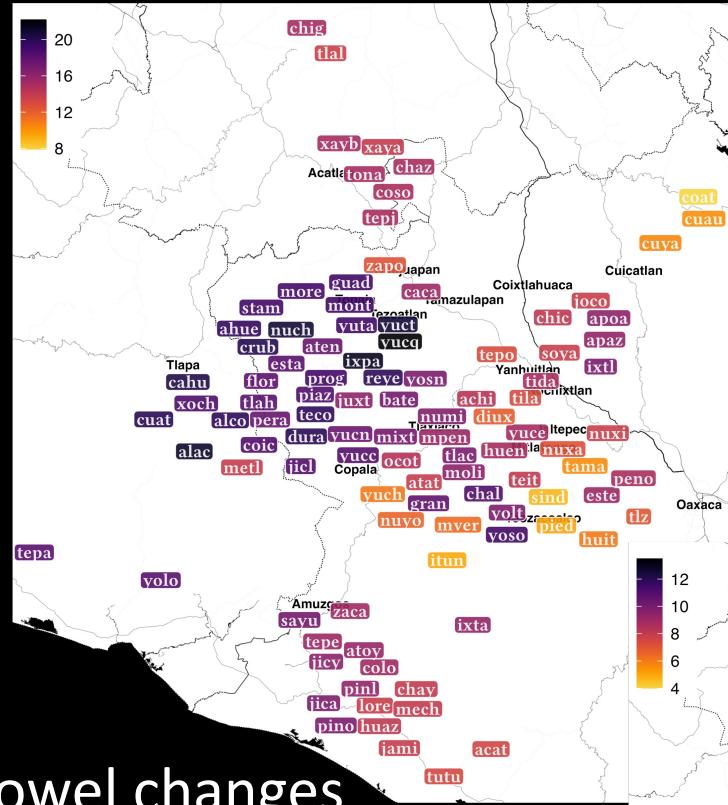
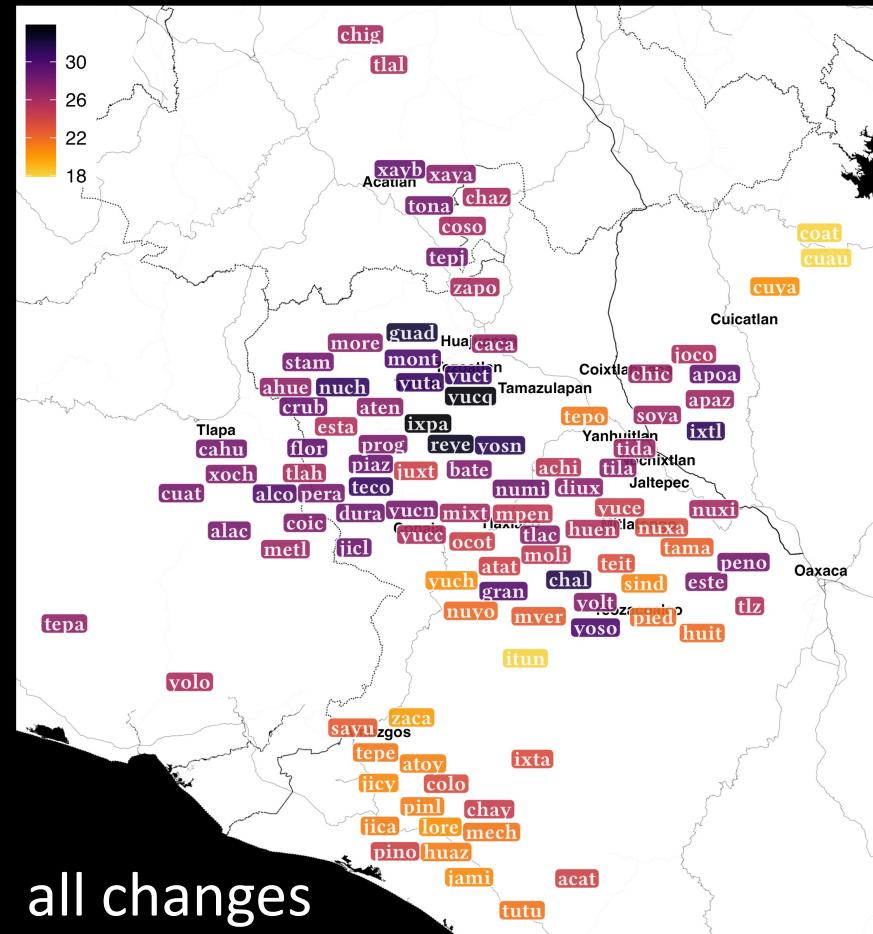
comprehensive database of sound changes coded in a bottom-up, fine-grained way  
reveals both tree-like and wave-like processes in the Mixtec language family

#### Limitations and future work:

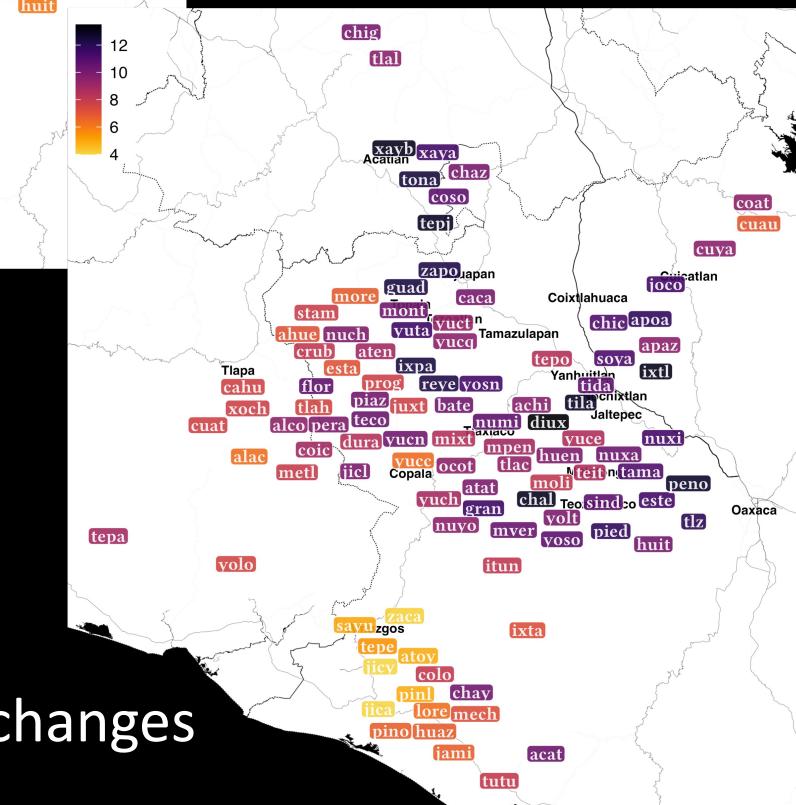
- ▶ tone change
- ▶ relative chronology
- ▶ correlations between changes ~ chain shifts
- ▶ generalizations
- ▶ revision and expansion of the data base with new cognate sets

#### Outlook:

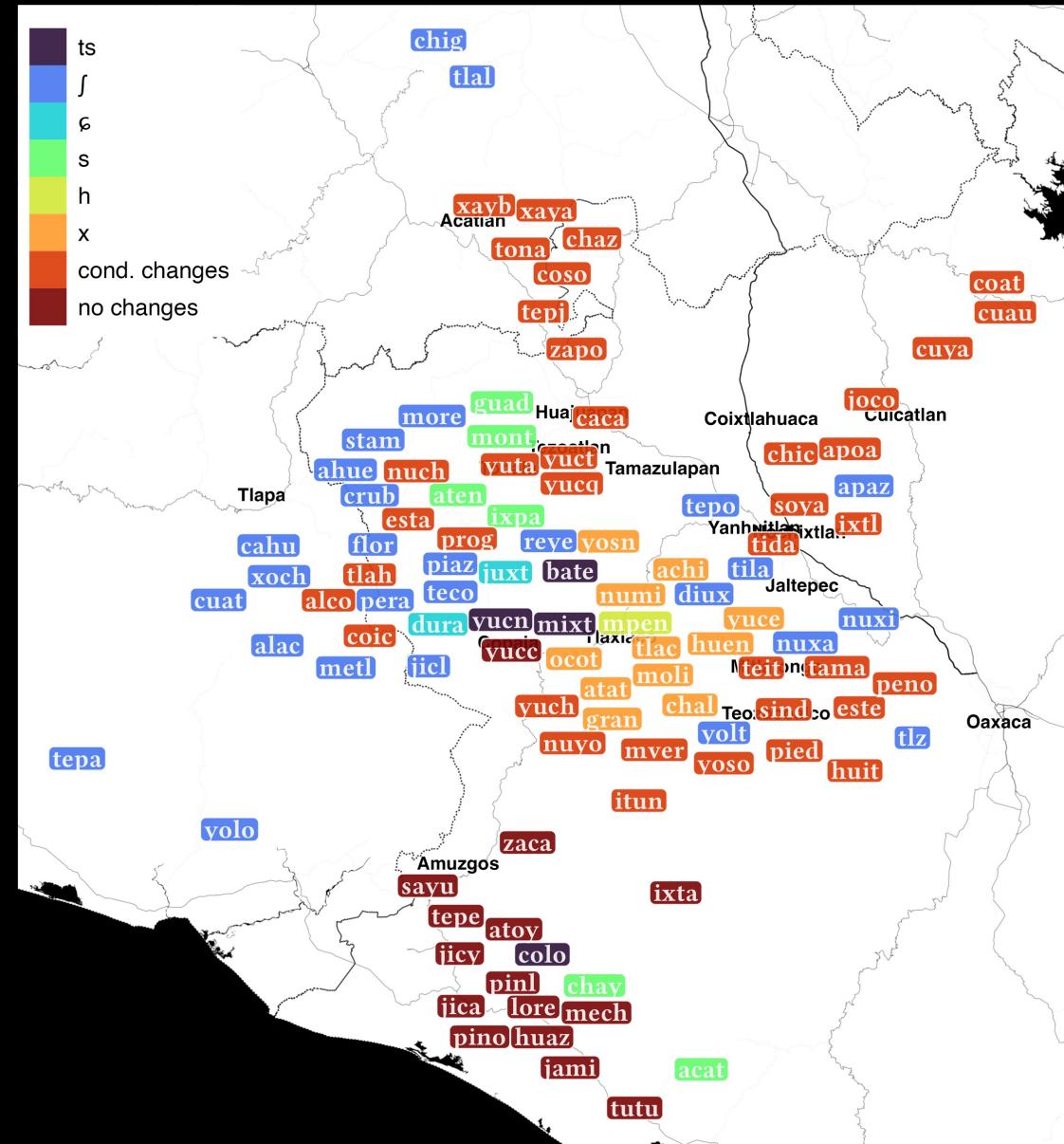
- ▶ making sound change & reconstruction more explicit, accessible, and re-usable
- ▶ application to other language families
- ▶ long run: typological data base of sound change but bottom-up



# vowel changes

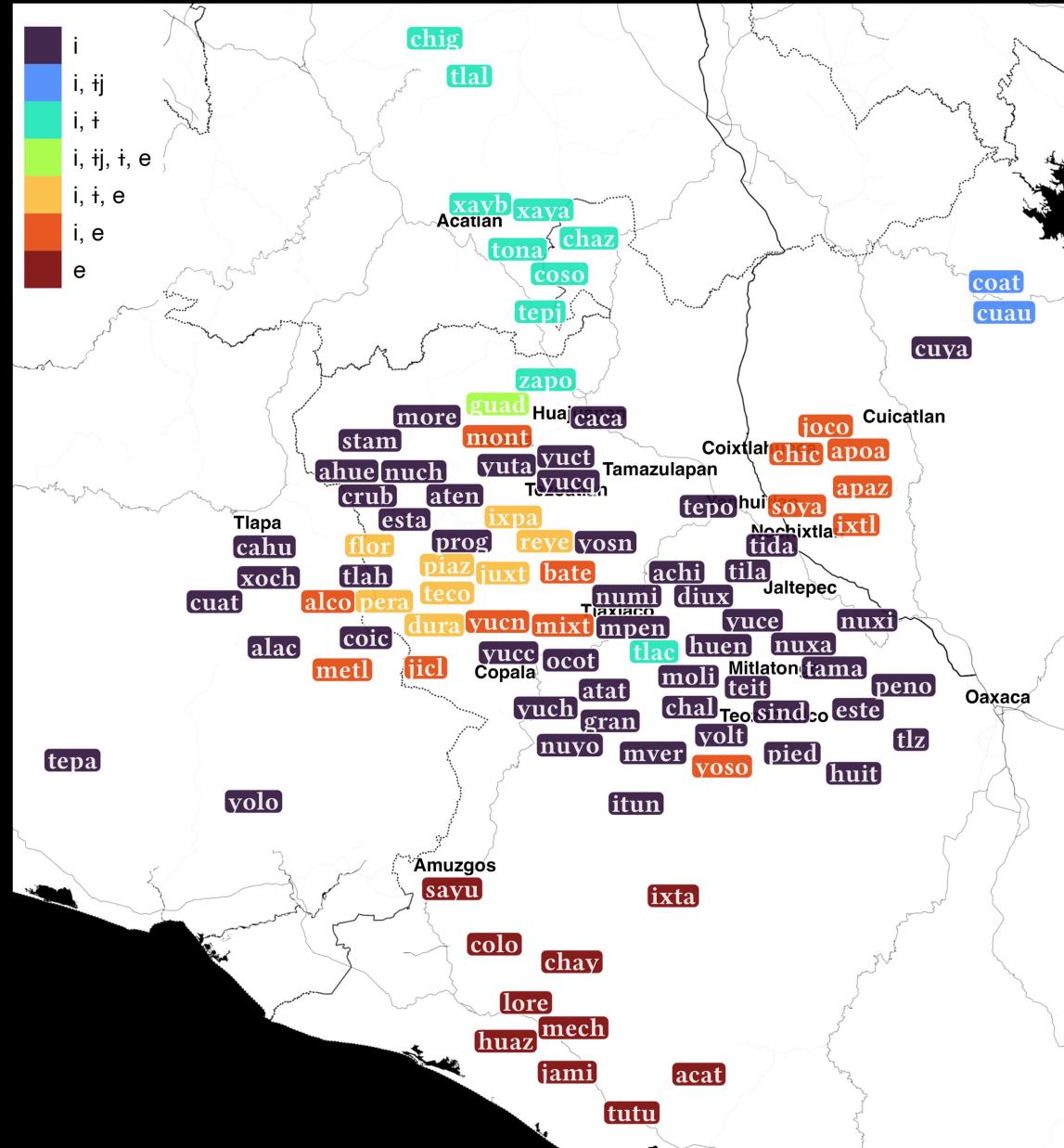


# Global reflexes of PM \*tʃ



# Wave-like processes

## environments that condition palatalization of proto-Mixtec \*t



# Beyond trees and waves: Correlations

each point = sound change

teal = correlation with subgroups by  
Auderset et al. 2023

purple = correlation with dialect areas  
by Josserand 1983

Cramer's V (correlation measure)  
> 0.6      strongly correlated  
0.3-0.6    moderately correlated  
< 0.3     weakly/not correlated

