

# Phonology of East Caucasian languages

George Moroz

Linguistic Convergence Laboratory, NRU HSE

21 October 2020

part of the online course on East Caucasian languages

last version of the presentation is available here: [tinyurl.com/y2foeo9g](https://tinyurl.com/y2foeo9g)



# Outline of the talk

Introduction

Data

Methods

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

# About phonology of East Caucasian languages

- phonology
  - [Kibrik and Kodzasov 1990]
  - The indigenous languages of the Caucasus: [Job and Smeets 1994], [Smeets 1994]
  - Yaziki mira: [Alekseev et al. 2001] and predecessor projects
  - [Grawunder 2017]
  - The Oxford Handbook of Languages of the Caucasus: [Beguš 2021], [Boris 2021a], [Boris 2021b]
  - Handbook of Caucasian Languages: [Koryakov and Maisak 202]
  - and phonological part in more general descriptions of East Caucasian languages
- historical-comparative phonetics
  - S. M. Gasanova, G. Kh. Ibragimov, P. T. Magomedova, U. A. Meylanova, B. B. Talibov, S. M. Khaydakov, A. E. Alekseev, Ju. D. Desheriev, B. K. Gigineishvili, D. S. Imnaishvili, J. Nichols, S. L. Nikolayev, S. A. Starostin, A. S. Kasyan, N. A. Ardoteli, O. A. Mudrak

# Outline of the talk

Introduction

Data

Methods

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

# Languages

- Nakh
  - Chechen
  - Ingush
  - Tsova-Tush  
(Bats, Batsbi)
- Tsezic (Didoic)
  - Tsez
  - Hinuq
  - Bezhta
  - Hunzib
  - Khwarshi-Inkhoqwari
- Avar–Andic
  - Avar
  - Andi
  - Botlikh
  - Godoberi
  - Chamalal
  - Karata
  - Tindi
  - Bagvalal
  - Akhvakh
  - Dargwa
  - Lak
- Lezgic
  - Archi
  - Tabasaran
  - Aghul
  - Lezgian
  - Udi
  - Budukh
  - Kryz
  - Rutul
  - Tsakhur
  - Khinalug

# Languages not present in [Kibrik and Kodzasov 1990]

- Nakh
  - Chechen
  - Ingush
  - Tsovedi-Tush  
(Bats, Batsbi)
- Tsezic (Didoic)
  - Tsez
  - Hinuq
  - Bezhta
  - Hunzib
  - Khwarshi-Inkhoqwari
- Avar-Andic
  - Avar
  - Andi
  - Botlikh
  - Godoberi
  - Karata
  - Chamalal
  - Tindi
  - Bagvalal
  - Akhvakh
  - Dargwa
  - Lak
- Lezgic
  - Archi
  - Tabasaran
  - Aghul
  - Lezgian
  - Udi
  - Budukh
  - Kryz
  - Rutul
  - Tsakhur
  - Khinalug

## Languages not present in [Kibrik and Kodzasov 1990]

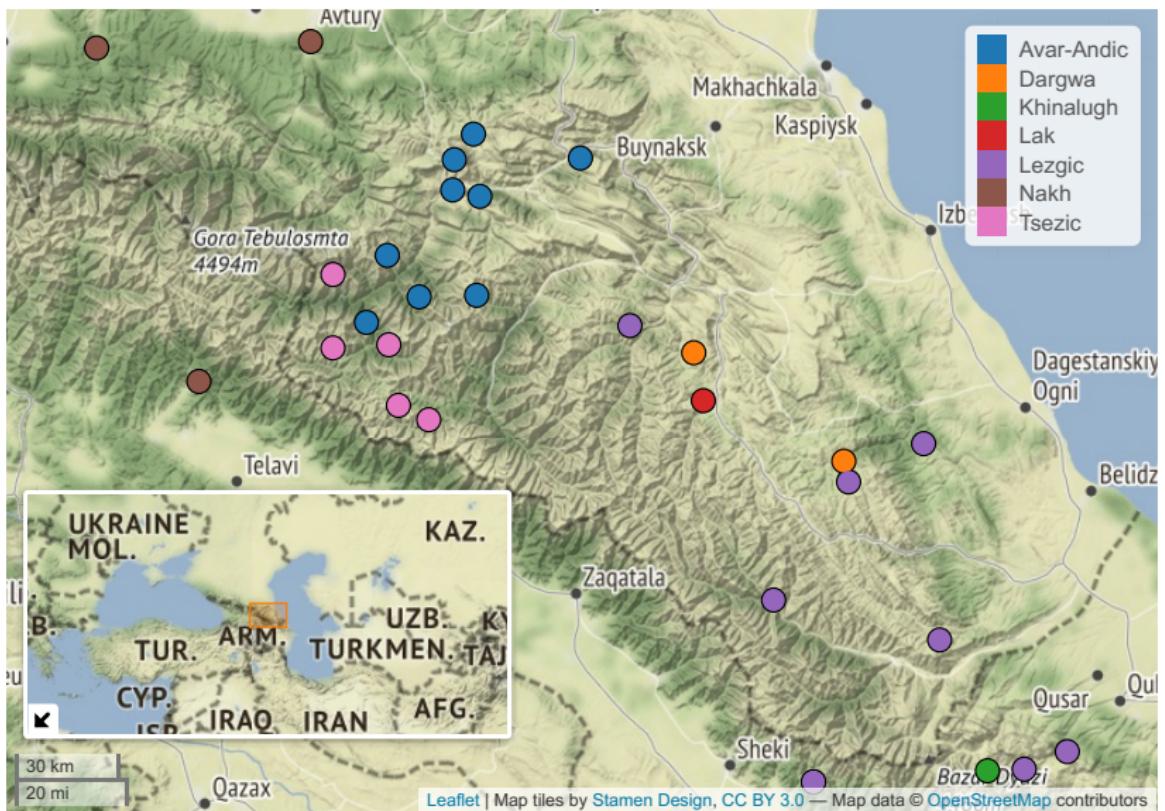
- Andic, Bagvalal [[Kibrik et al. 2001](#)]
- Andic, Botlikh [[Alexeyev and Verhees 202](#)]
- Andic, Godoberi [[Moroz et al. 202](#)]
- Andic, Karata [[Magomedbekova 1971](#)]
- Dargwa, Mehweb [[Moroz 2019](#)]
- Nakh, Chechen [[Nichols 1994a](#)]
- Nakh, Ingush [[Nichols 1994b](#)]
- Nakh, Tsoved-Tush [[Holisky and Gagua 1994](#)]

# Data

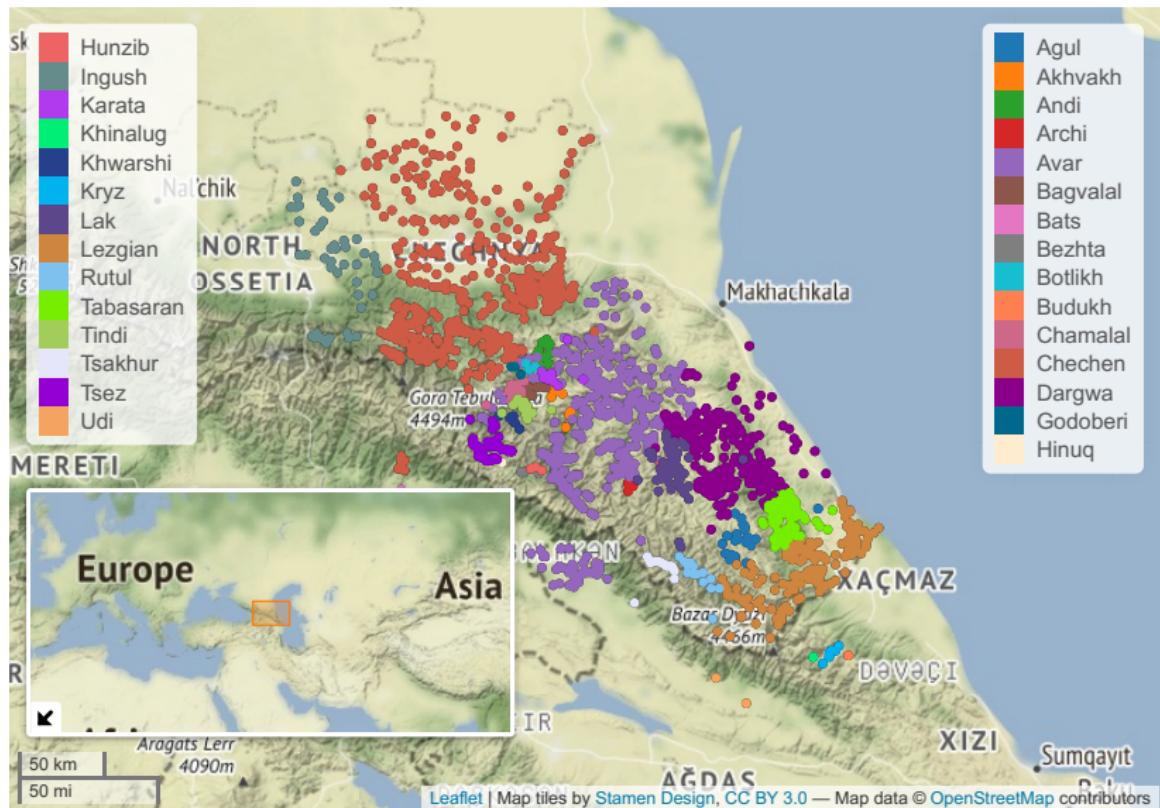
1905 × 9 **table** with the following columns:

- id
- branch
- language
- segments\_IPA
- segments\_source
- sound\_type
- source
- comments
- glottocode

# Data (language view)



# Data (village view)



# Outline of the talk

Introduction

Data

**Methods**

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

# Phonology of East Caucasian languages can be presented like this...

	Stops				Fricatives			Affricates				Son.
	Voiced	Voiceless			Voiced	Voiceless		Voiced	Voiceless			
			Gem.	Ej.			Gem.			Gem.	Ej.	
Lab.	b ɓ	p ń										w ń m ń
Den.	d ڏ	t ڻ	t: tt	t' ڻI	z ڙ	s c	s: cc		c ڦ	c: ڦڻ	c' ڻI	n ڻ
Alv.					ž ڇ	š ڻ	š: ڻڻ	ڇ ڏڇ	č ڇ	č: ڇڻ	č' ڻI	r p
Lat.						ł ڻ	ł: ڻڻ			ł ڻI	ł' ڻI	l l
Pal.												j ڻ
Vel.	g ڻ	k ڻ	k: kk	k' ڻI			x x̥					
Uv.		q x̥		q' ڻI	v g̥	χ x	χ: xx					
Ph.					f rI	h xI						

# Phonology of East Caucasian languages can be presented like this...

	Stops			Fricatives			Affricates			Son.	
			plosives			fricatives		affricates			sonorant
			asp	ej	vcd	asp	vcd	asp	ej	vcd	
Lab.	labial		p		b						w, m
	dental	-gem	t	t'	d	s	z	c	c'		n
Den.		+gem				s:		c:	c':		
alveolar	-gem				š	ž	č	č'	č̄	r	
	Alv.		+gem				š:		č:	č':	
lateral					ł (ł:)		ł	ł'		ł	
Lat.	palatal					ç, çʷ					j
	velar	-gem	k, kʷ	k', k'ʷ	g, gʷ						
Vel.		+gem	k:, k:ʷ	k':,							
Uv.	uvular	-gem		q', q'ʷ		χ, χʷ	χ, χʷ	qχ, qχʷ			
		+gem				χ:					
Ph.											

# Phonology of East Caucasian languages can be presented like this...

	Stops			Fricatives			Affricates			Son.
	plosives			fricatives		affricates			sonorant	
	asp	ej	vcd	asp	vcd	asp	ej	vcd		
Lab.	labi	labial	dental	alveolar	palatal	velar	uvular	pharyngeal	epig	
Den.	den	-lab	+lab	-lab	+lab	-lab	+lab	-lab	-lab	
Alv.	plosive	+v	b	d		g	g <sup>w</sup>		(?)	
	alvc	-v	p'	t'		k	k <sup>w</sup>	q	q <sup>w</sup>	
	ej					k'	k' <sup>w</sup>	q'	q' <sup>w</sup>	
Lat.	fricative	+v		z	ž	(y)		š	š <sup>w</sup>	
	late	-v		s	š	x	x <sup>w</sup>	χ	χ <sup>w</sup>	
								h	h <sup>w</sup>	
									(h)	
Pal.	pal:	+v		(ʒ)	(ʒ)					
	affricate	-v		c	č					
		ej		c'	č'					
Vel.	vel:	sonorant	m, w	n, l, r	j					
Uv.		---	--	--	--					
Ph.	uvular	-gem		q', q" <sup>w</sup>		χ, χ <sup>w</sup>	š, š <sup>w</sup>	qχ, qχ <sup>w</sup>		
		+gem				χ:				

This way is...

- boring,

## This way is...

- boring,
- mechanical,

## This way is...

- boring,
- mechanical,
- it is hard to make generalisations (at least, on the fly)

# Phonology of East Caucasian languages can be presented like this...

Nakh-Daghestanian: Essential background, what's distinctive... (Johanna Nichols)



## The N-D phonological signature

Complex consonant inventories. Typically:

3 or more manners. t t' d tt /t/ usually [t<sup>h</sup>]

Uvular, sometimes lateral, sometimes pharyngeal series

Vowel systems vary in size:

3 or 4 vowel qualities: Lak, several Dargwa

5 Avar, most Andic

6-8 most Tsezic, most Lezgian; Nakh

9-10 Kryz, Budukh, Khinalug



± length, nasalization

Pharyngealization

Epiglottalization vs. true pharyngealization

Segment? syllabic feature? consonant feature? vowel feature?

Syllable structure fairly simple: mostly CVRC

Tones ...

▶ ⏴ 🔊 37:24 / 1:30:04



# Phonology of East Caucasian languages can be presented like this...

Nakh-Daghestanian: Essential background, what's distinctive... (Johanna Nichols)



## The N-D phonological signature

Complex consonant inventories. Typically:

3 or more manners. t t' d tt /t/ usually [t<sup>h</sup>]

Uvular, sometimes lateral, sometimes pharyngeal series

Vowel systems vary in size:

3 or 4 vowel qualities: Lak, several Dargwa

5 Avar, most Andic

6-8 most Tsezic, most Lezgian; Nakh

9-10 Kryz, Budukh, Khinalug



± length, nasalization

Pharyngealization

Epiglottalization vs. true pharyngealization

Segment? syllabic feature? consonant feature? vowel feature?

Syllable structure fairly simple: mostly CVRC

Tones ...

▶ ⏸ 🔊 37:24 / 1:30:04



(If you have not seen it, go and watch our previous lecture!)

## These ways are...

- To list all inventories
  - boring,
  - mechanical,
  - hard to make generalisations (at least, on the fly)
- To present generalisations
  - restricted to peculiarities,

## These ways are...

- To list all inventories
  - boring,
  - mechanical,
  - hard to make generalisations (at least, on the fly)
- To present generalisations
  - restricted to peculiarities,
  - you need to be a specialist

## These ways are...

- To list all inventories
  - boring,
  - mechanical,
  - hard to make generalisations (at least, on the fly)
- To present generalisations
  - restricted to peculiarities,
  - you need to be a specialist
- May be these conclusions could be data driven?

## These ways are...

- To list all inventories
  - boring,
  - mechanical,
  - hard to make generalisations (at least, on the fly)
- To present generalisations
  - restricted to peculiarities,
  - you need to be a specialist
- May be these conclusions could be data driven?
  - some of them can be visualised with a map
  - some of them can be gathered via frequency and combinatorics analysis

# Outline of the talk

Introduction

Data

Methods

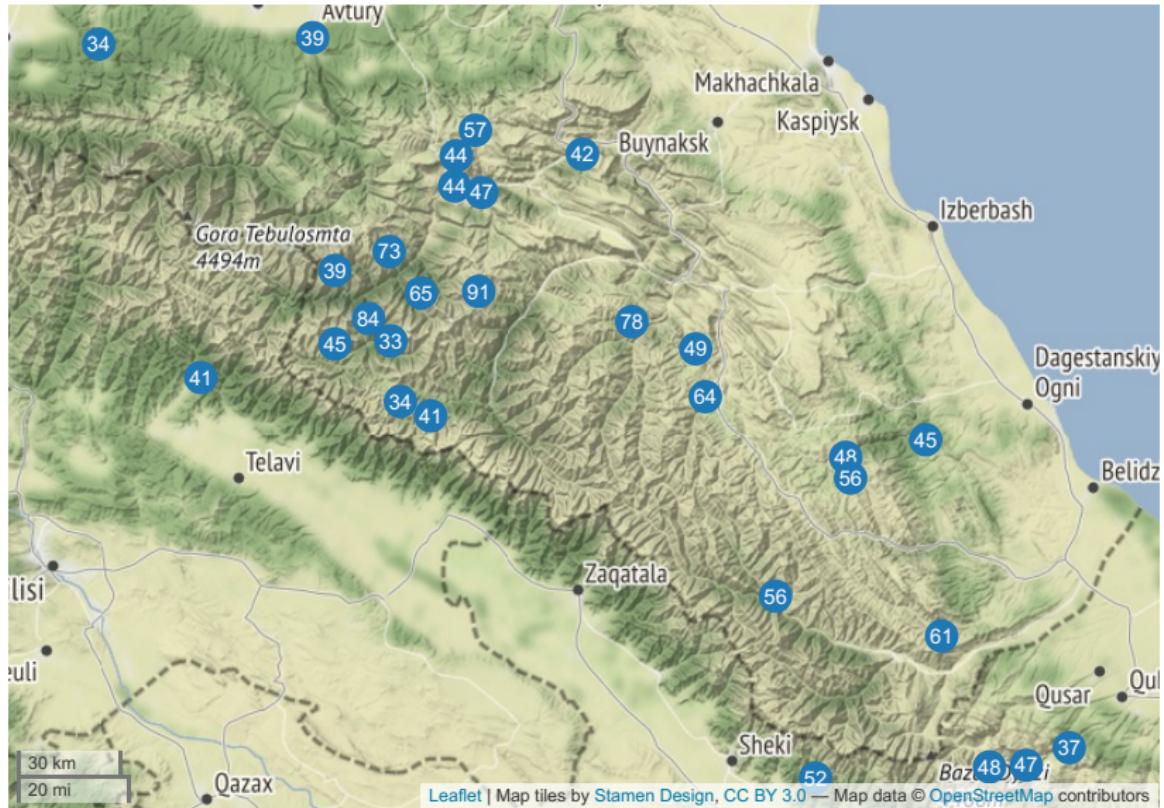
**Inventory size**

Frequency analysis

Compare to other databases

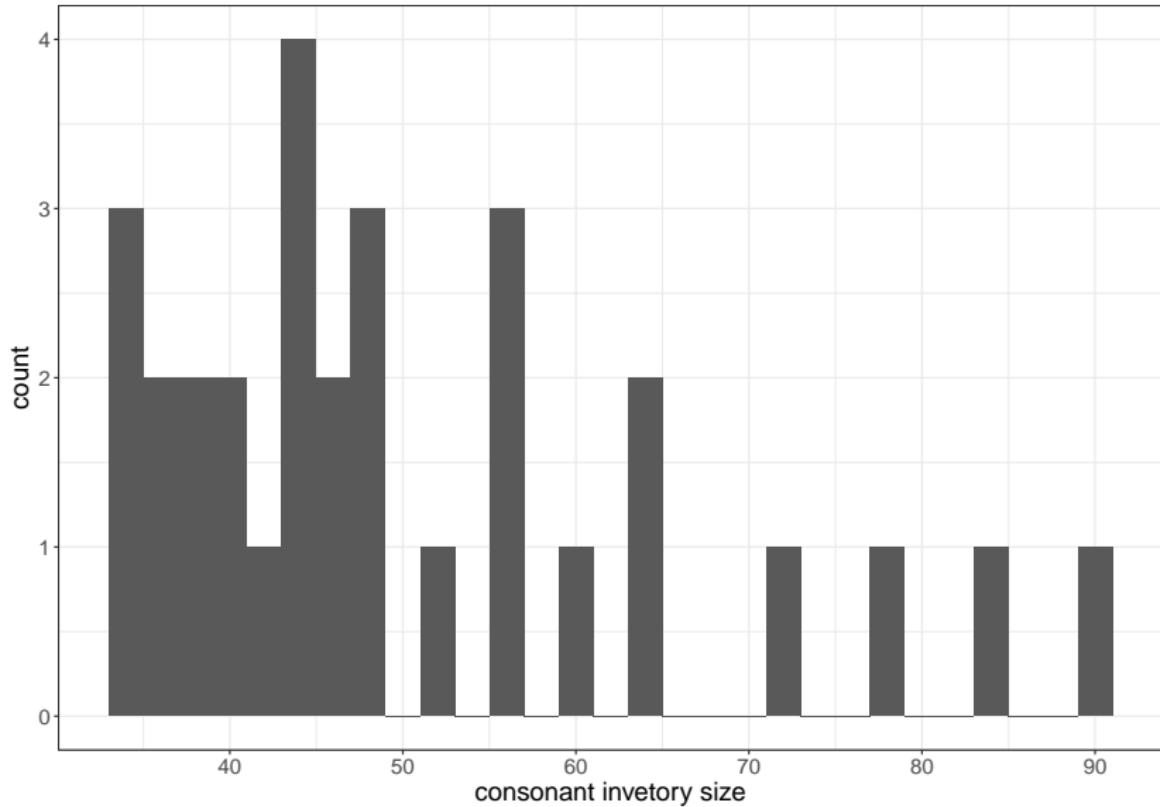
Suprasegmental features

# Consonants



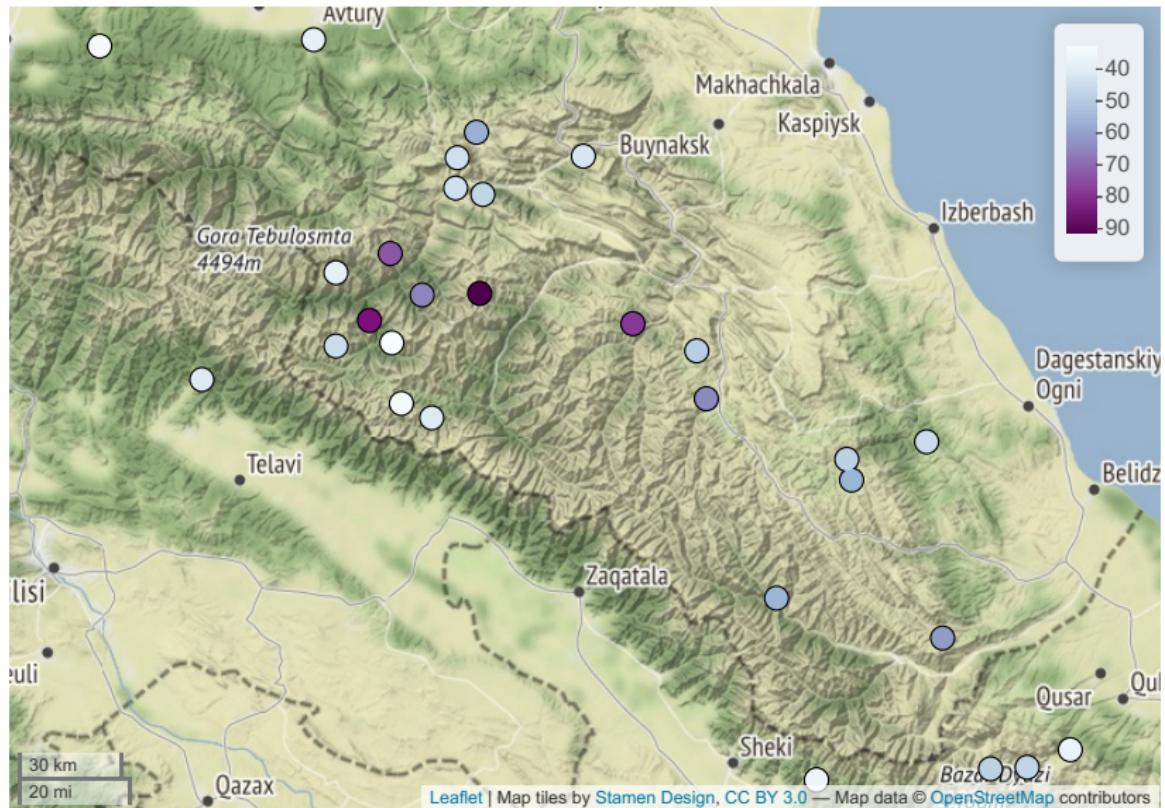
Languages ranges from 33 to 91 consonants.

# Consonants



Most of the languages have  $42 \pm 5$  consonants.

# Consonants



Most of the huge systems are presented in Andic and grouped together.

# Consonants

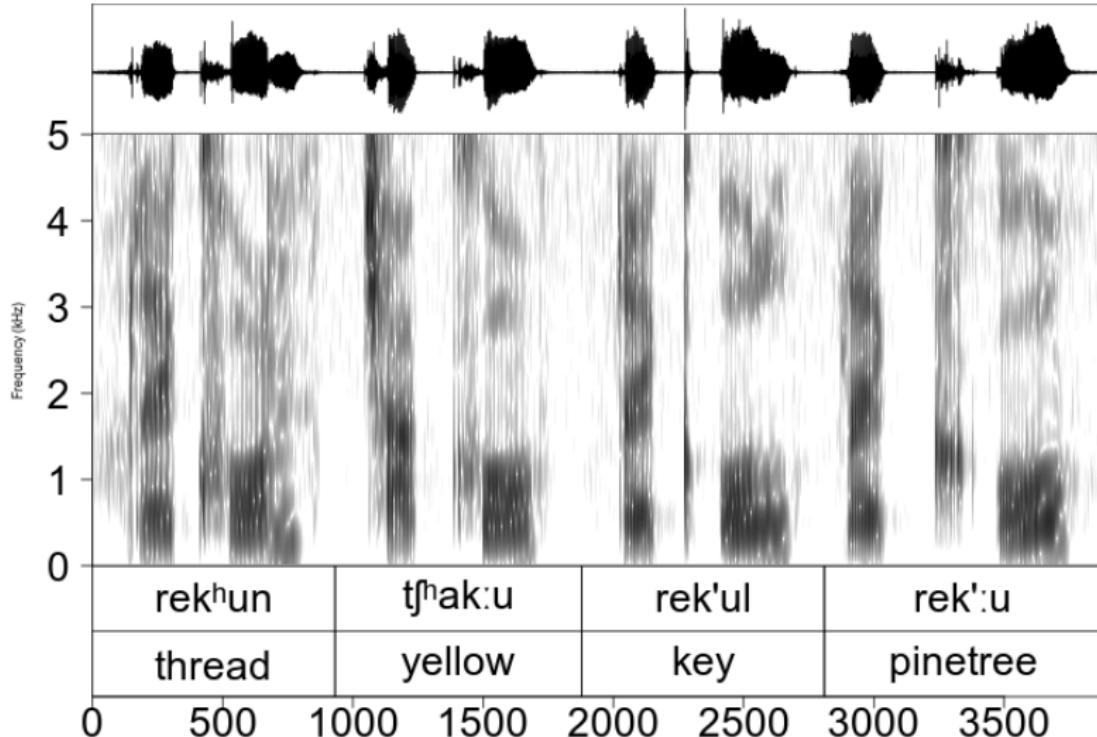
- Akhwakh (Andic) – 91
- Tindi (Andic) – 84
- Chamalal (Andic) – 73
- Archi (Lezgic) – 78
- ...
- Ingush (Nakh) – 34
- Bezhta (Tsezic) – 34
- Inkhoqwari (Tsezic) – 33
- ejectives
- gemenation (fortis)
- labialization

# Consonants: Akwakh vs. Inkhoqwari

b	d			dʒ				g					
	d <sup>w</sup>			dʒ <sup>w</sup>				g <sup>w</sup>					
p <sup>h</sup>	t <sup>h</sup>	ts	ts:	tʃ	tʃ:	tɬ	tɬ:	k	k:	q	q:		?
	t <sup>hw</sup>	ts <sup>w</sup>	ts: <sup>w</sup>	tʃ <sup>w</sup>	tʃ: <sup>w</sup>	tɬ <sup>w</sup>	tɬ: <sup>w</sup>	k <sup>w</sup>	k: <sup>w</sup>	q <sup>w</sup>	q: <sup>w</sup>		? <sup>w</sup>
p'	t'	ts'	ts: <sup>'</sup>	tʃ'	tʃ: <sup>'</sup>	tɬ'	tɬ: <sup>'</sup>	k'	k: <sup>'</sup>	q'	q: <sup>'</sup>		
	t <sup>w</sup>	ts <sup>w</sup>	ts: <sup>w</sup>	tʃ <sup>w</sup>	tʃ: <sup>w</sup>	tɬ <sup>w</sup>	tɬ: <sup>w</sup>	k <sup>w</sup>	k: <sup>w</sup>	q <sup>w</sup>	q: <sup>w</sup>		
		z		ʒ						β			ɸ
		z <sup>w</sup>		ʒ <sup>w</sup>						β <sup>w</sup>			
		s	s:	ʃ	ʃ:	ɸ	ɸ:	x	x:	χ	χ:	h	h
		s <sup>w</sup>	s: <sup>w</sup>	ʃ <sup>w</sup>	ʃ: <sup>w</sup>	ɸ <sup>w</sup>	ɸ: <sup>w</sup>	x <sup>w</sup>	x: <sup>w</sup>	χ <sup>w</sup>	χ: <sup>w</sup>		h <sup>w</sup>
m	n												
w	r					l		j					

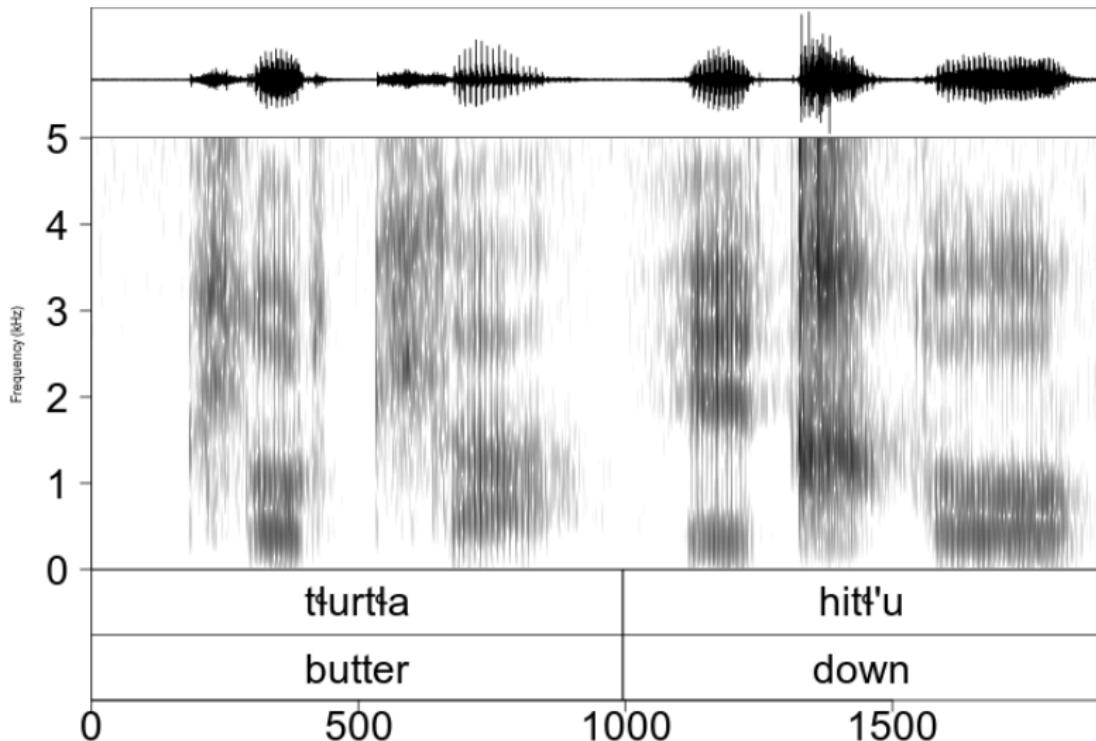
b	d		dʒ		g			
p <sup>h</sup>	t <sup>h</sup>	ts	tʃ	tɬ	k	q		?
p'	t'	ts'	tʃ'	tɬ'	k'	q'		
		z	ʒ		β	ɸ		
		s	ʃ	ɸ	χ	h		
m	n							
w	r			l	j			

## Gemination and ejectives (example from Zilo Andi)



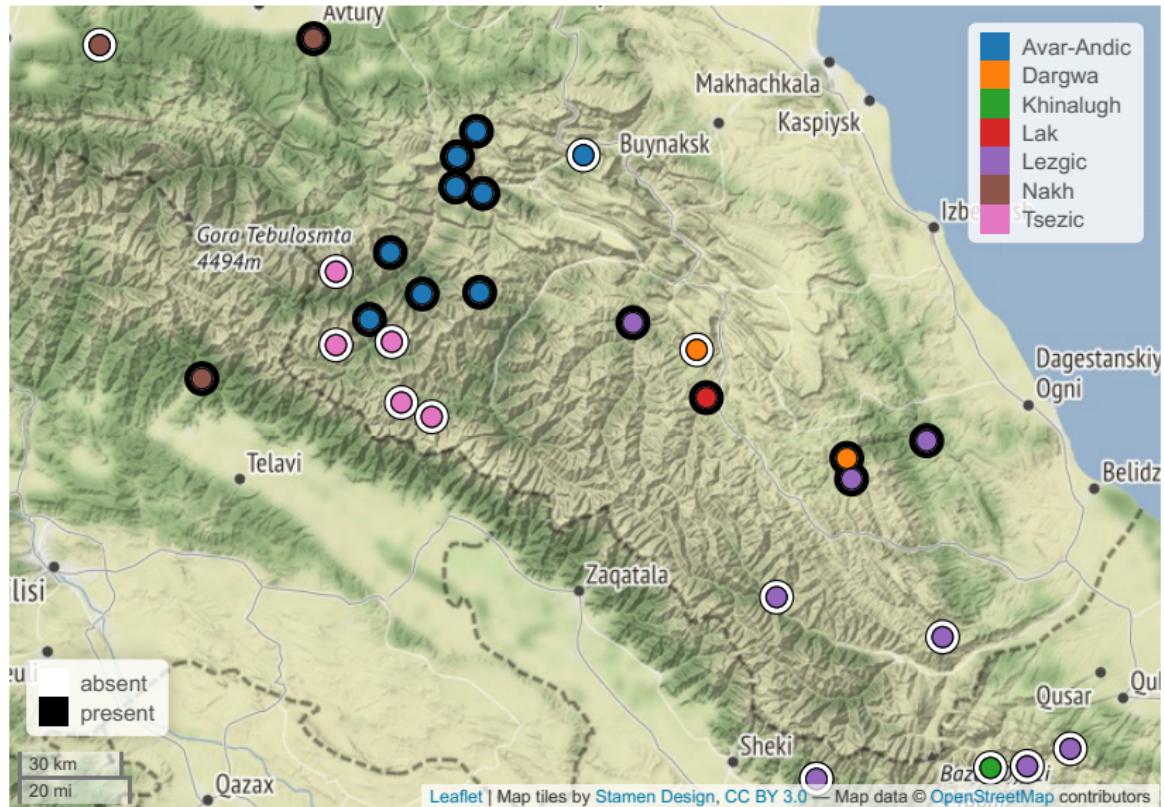
- Geminated aspirated have longer closure.
- Geminated ejectives have longer burst.

## Laterals (example from Zilo Andi)



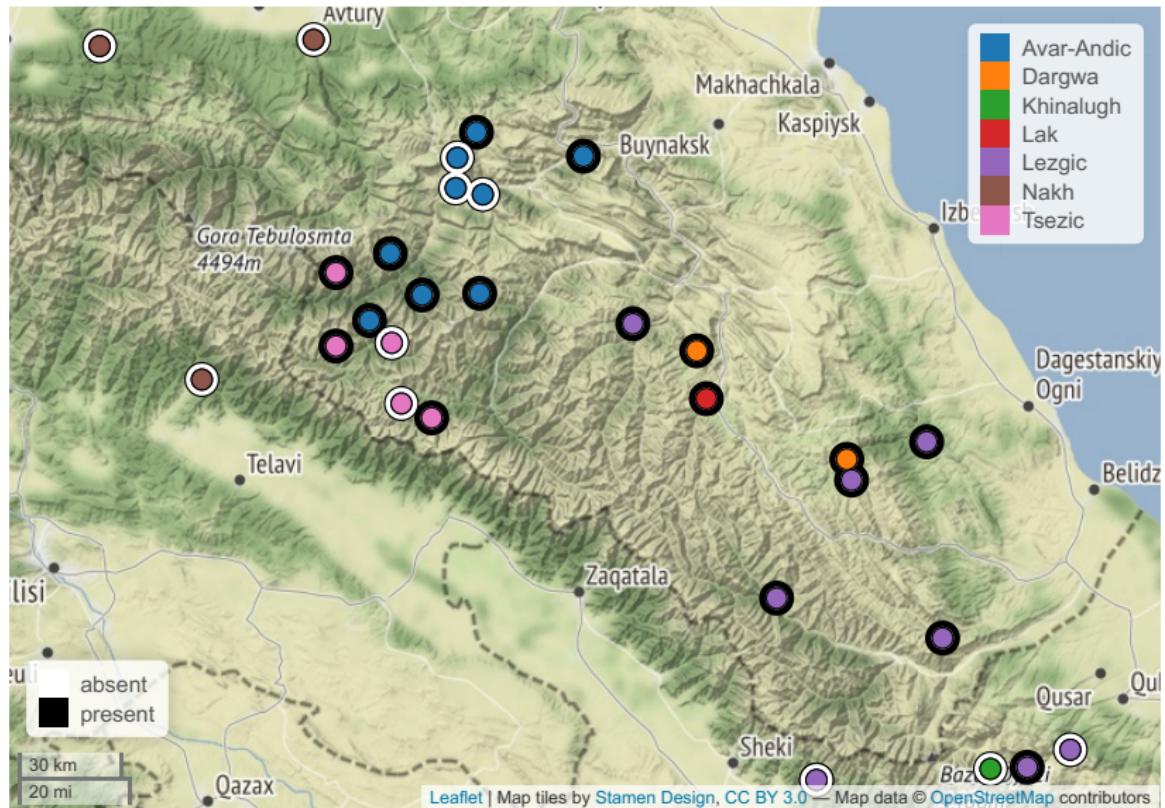
Looks like it is not t̪ but t̪: in Andic...

# Geminate consonants



Geminate consonants are absent in Tsezic and most Lezgic.

# Labialised consonants



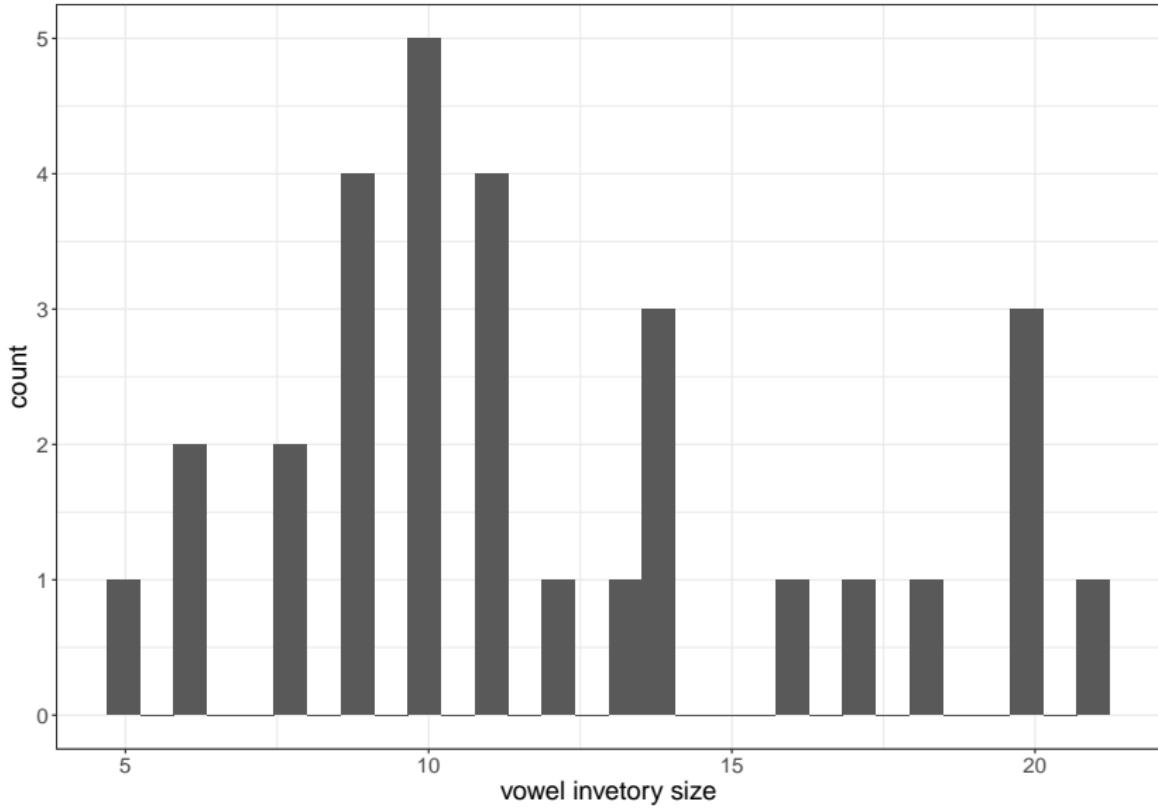
Labialised consonants are absent in Nakh, but more or less common through other branches.

# Vowels



Languages ranges from 5 to 21 vowels.

# Vowels



Most of the languages have  $10 \pm 5$  vowels.

# Vowels

- Bezhta (Tsezic) – 21
- Akhvakh (Andic) – 20
- Hunzib (Tsezic) – 20
- Bagvalal (Andic) – 20
- ...
- Aghul (Lezgic) – 6
- Tsez (Tsezic) – 6
- Avar – 5

# Vowels

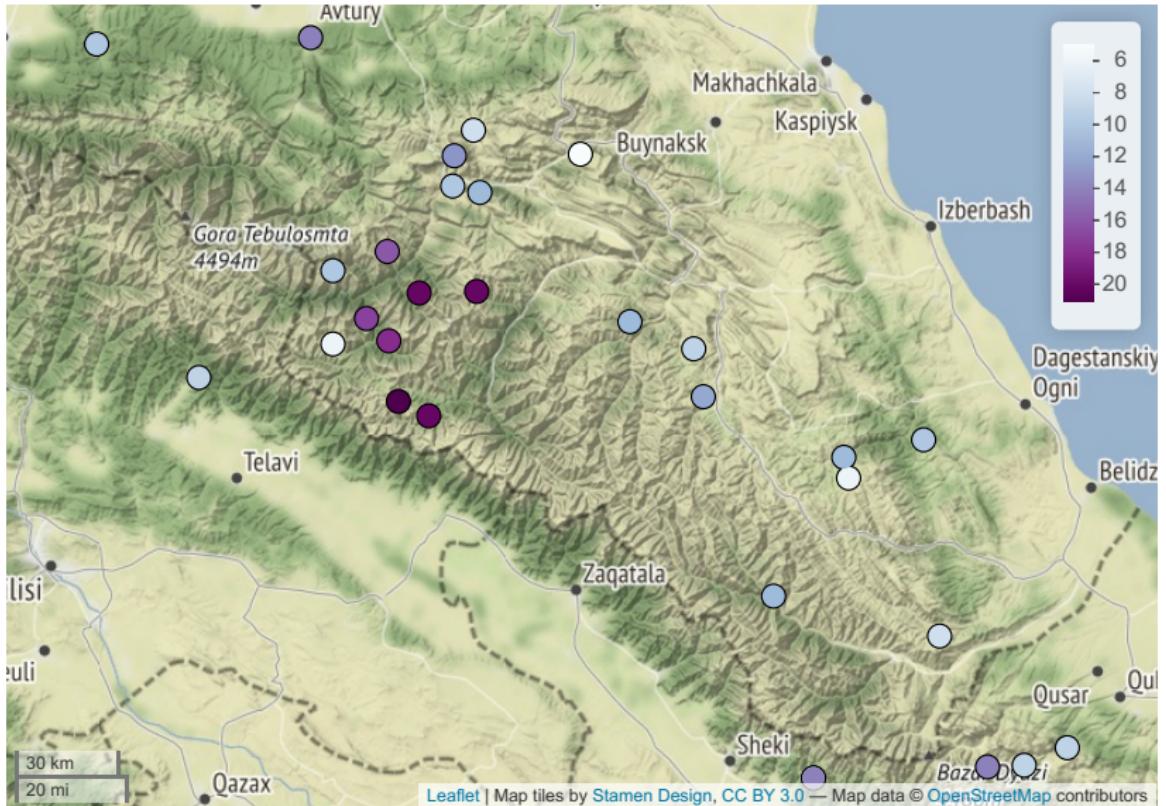
- Bezhta

i	i:	i <sup>n</sup>	y	y <sup>n</sup>				u	u:	u <sup>n</sup>
e		e <sup>n</sup>	ø					o	o:	o <sup>n</sup>
æ	æ:	æ <sup>n</sup>								
					a	a:	a <sup>n</sup>			

- Avar

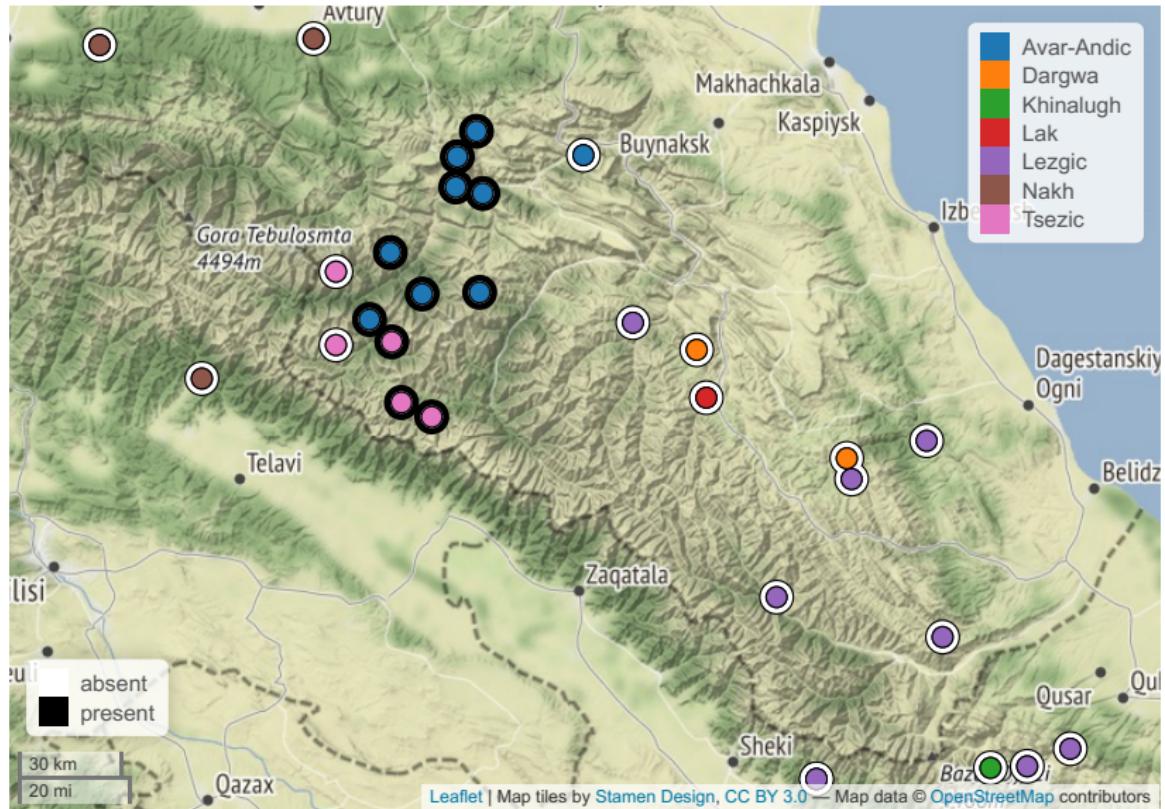
i		u
e		o
	a	

# Vowels



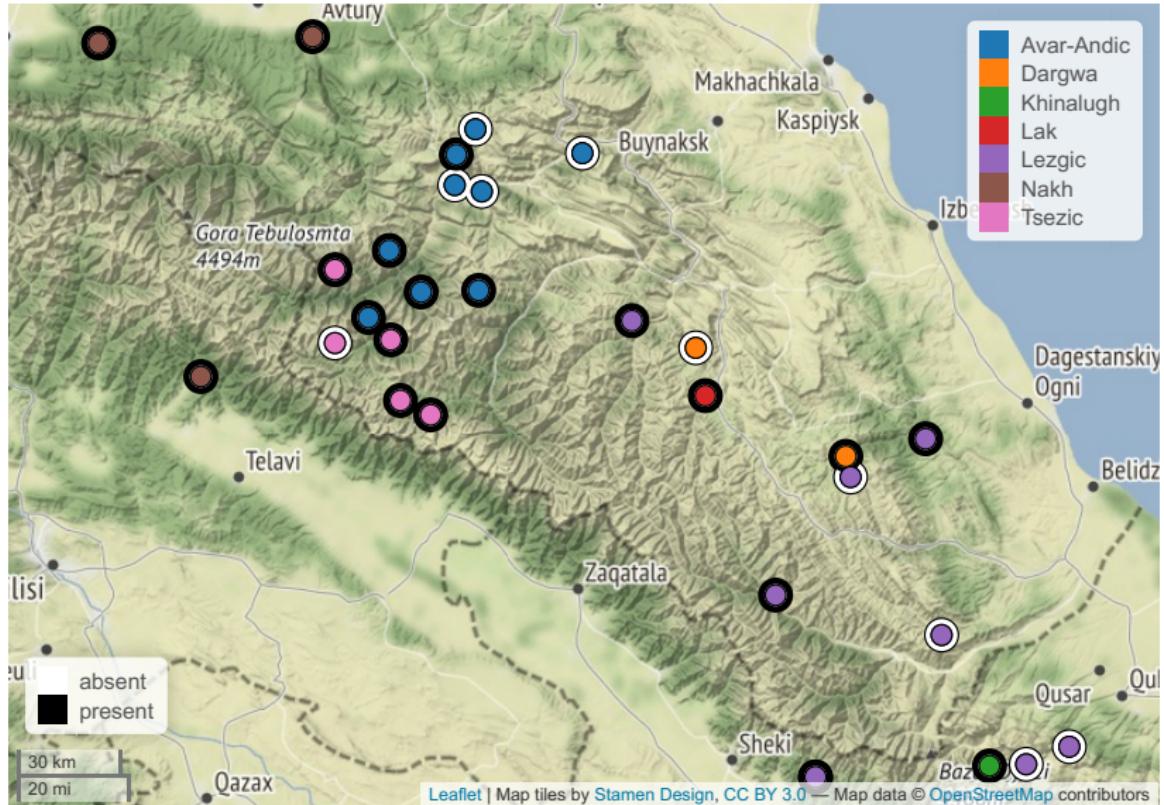
Most of the huge systems are presented in Tsezic and Andic.

# Nasal vowels



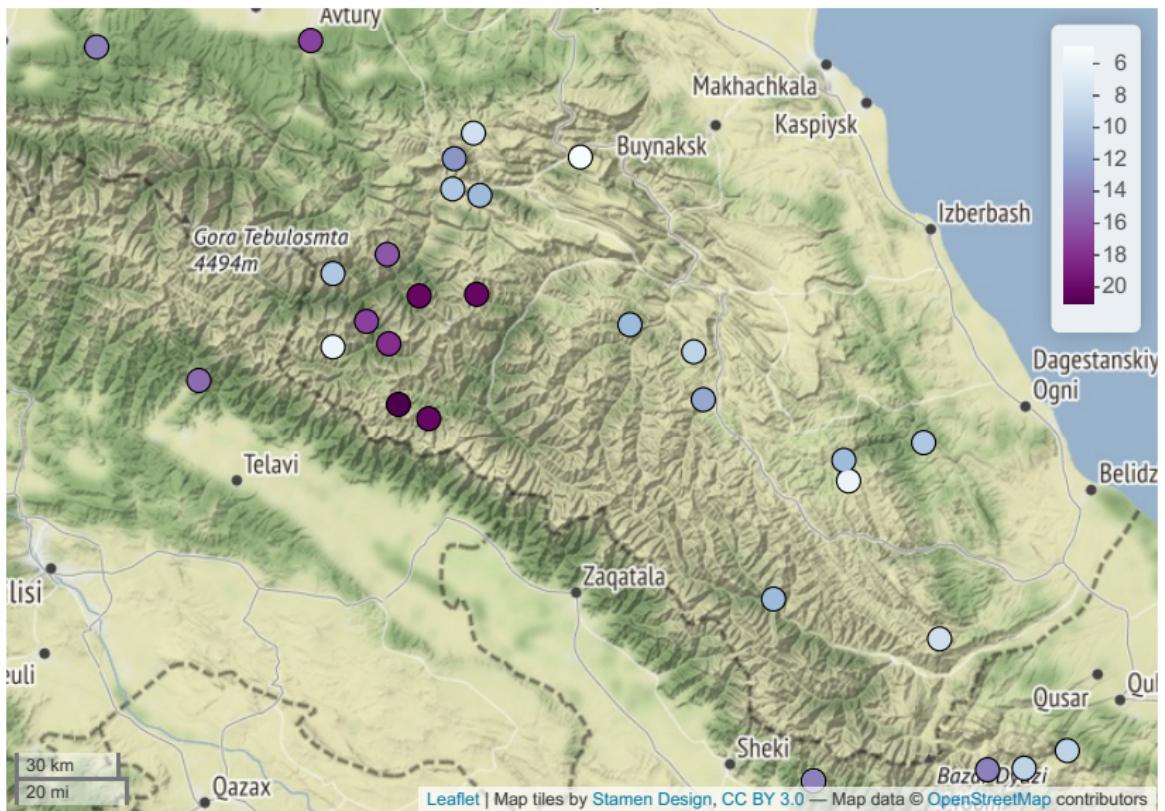
From Andic and Tsezic only Avar, Hinuq, and Tsez lack nasalized vowels.

# Long vowels

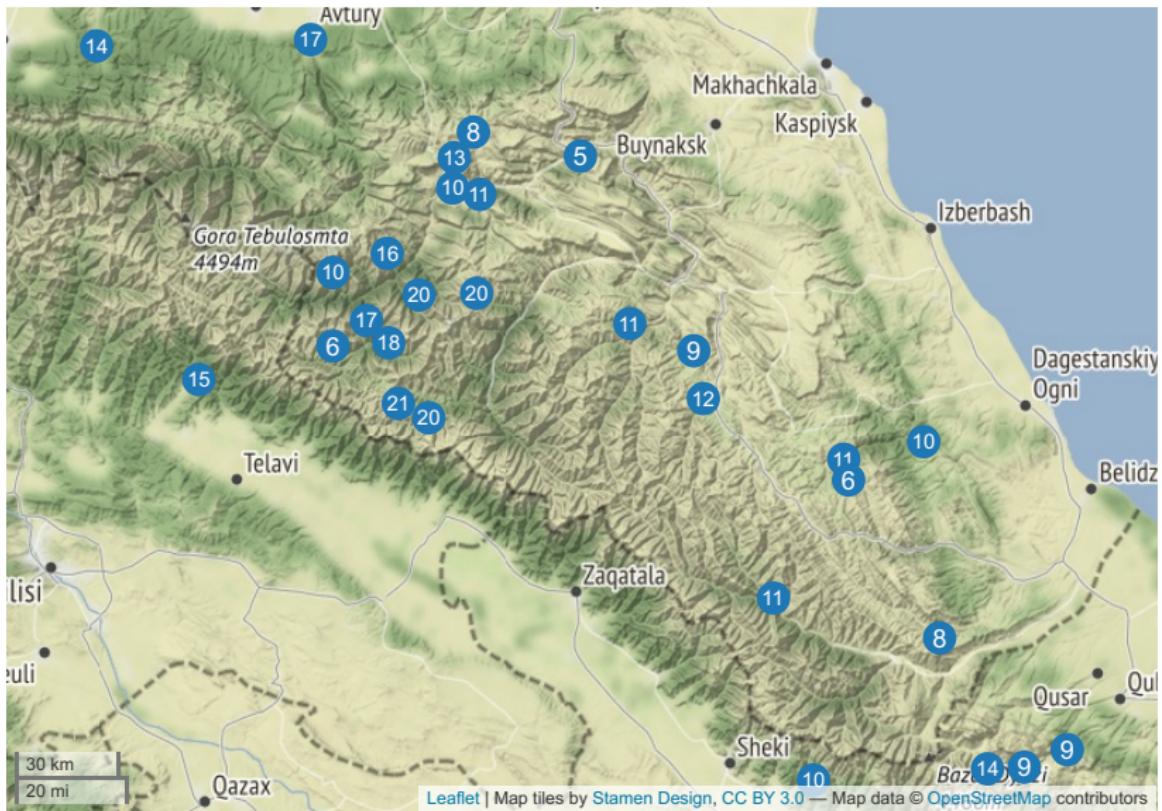


Long vowels are common to all branches.

# Vowels + diphthongs



# Vowels + diphthongs



# Outline of the talk

Introduction

Data

Methods

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

## Segments present in all languages

b	d					g	
p <sup>h</sup>	t <sup>h</sup>					k <sup>h</sup>	?
		3					
	s	f					h
m	n						
	r		l	j			

- Where are the famous ejectives?

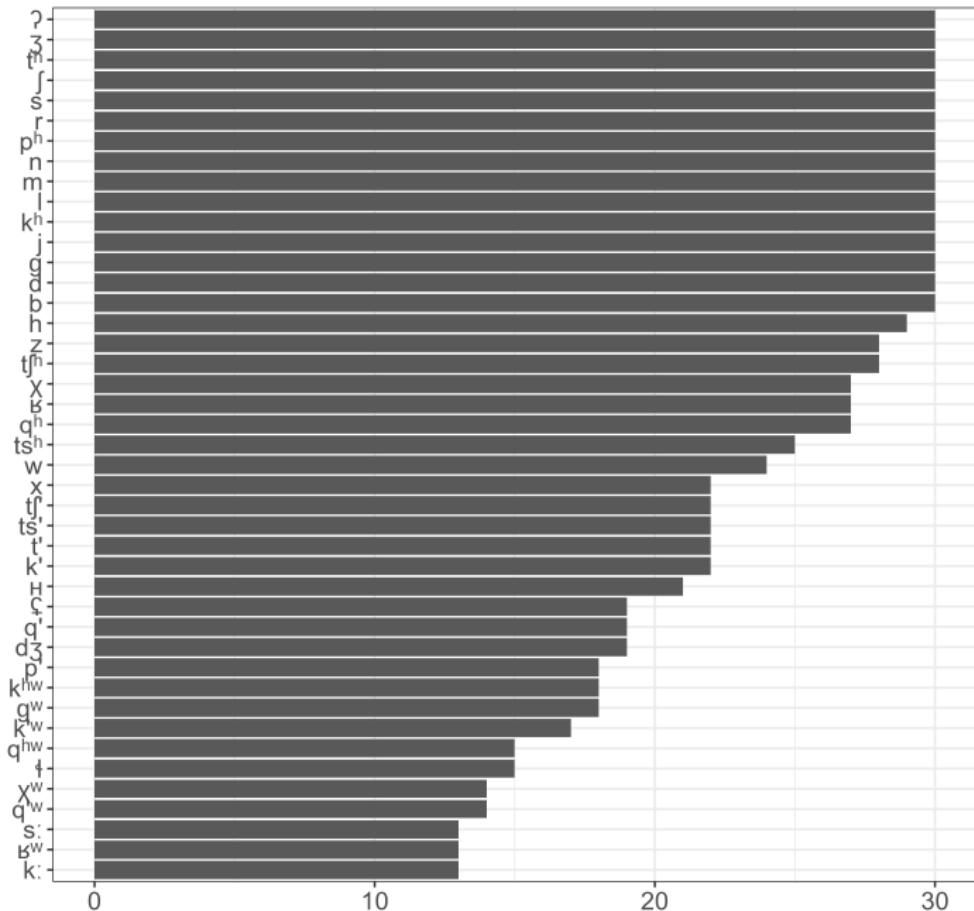
# Segments present in all languages

b	d					g	
p <sup>h</sup>	t <sup>h</sup>					k <sup>h</sup>	?
		3					
	s	f					h
m	n						
	r		l	j			

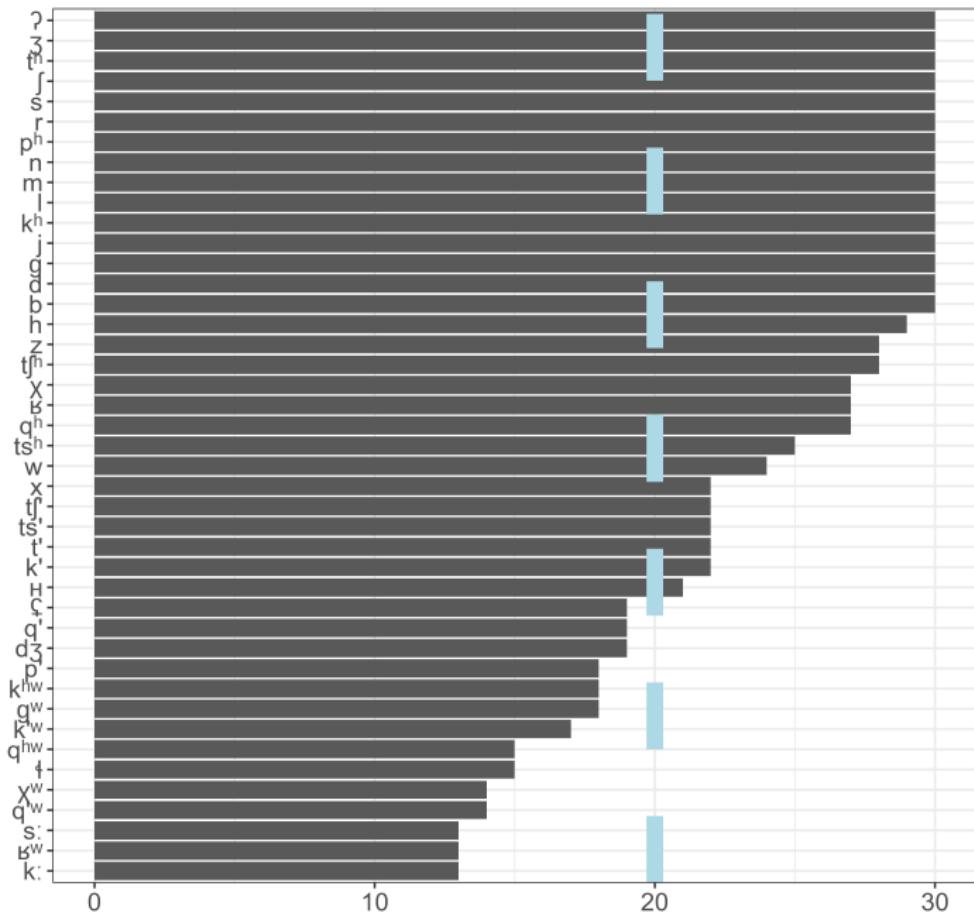
- Where are the famous ejectives? Udi:

b	d	dz	dʒ			g	
p <sup>h</sup>	t <sup>h</sup>	ts <sup>h</sup>	tʃ <sup>h</sup>			k <sup>h</sup>	q <sup>h</sup>
p	t	ts	tʃ			k	q
v		z	ʒ				β
f		s	f			χ	h
m	n						
w	r			l	j		

## Consonants: typical inventory



# Consonants: typical inventory



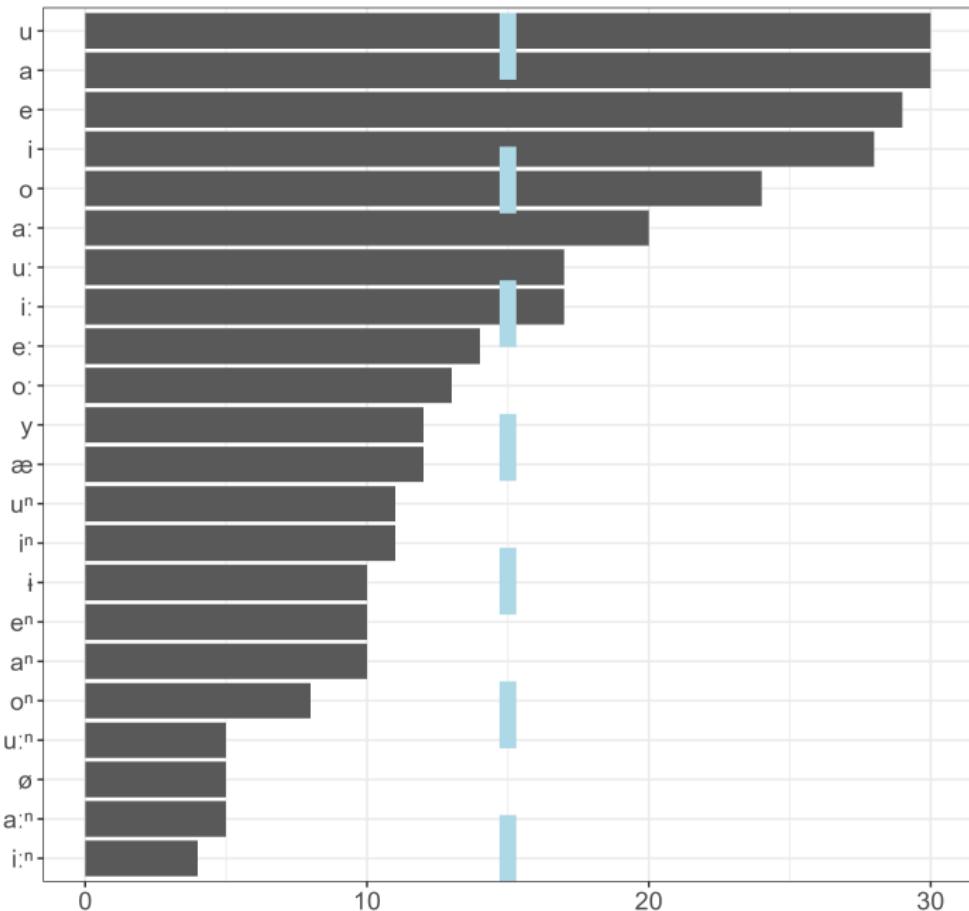
# Consonants: typical inventory

b	d					g			
p <sup>h</sup>	t <sup>h</sup>	ts	tʃ <sup>h</sup>			k <sup>h</sup>	q <sup>h</sup>		?
p'	t'	ts'	tʃ'			k'	q'		
		z	ʒ				v	f	
w		s	ʃ			x	χ	h	h
m	n								
	r			l	j				

For more consonants:

- labialisation
- gemination
- laterals
- pharyngealisation

# Vowels: typical inventory



# Vowels: typical inventory

i	i:			u	u:
e				o	
		a	a:		

For more vowels:

- length
- nasalization
- pharyngealization

# Outline of the talk

Introduction

Data

Methods

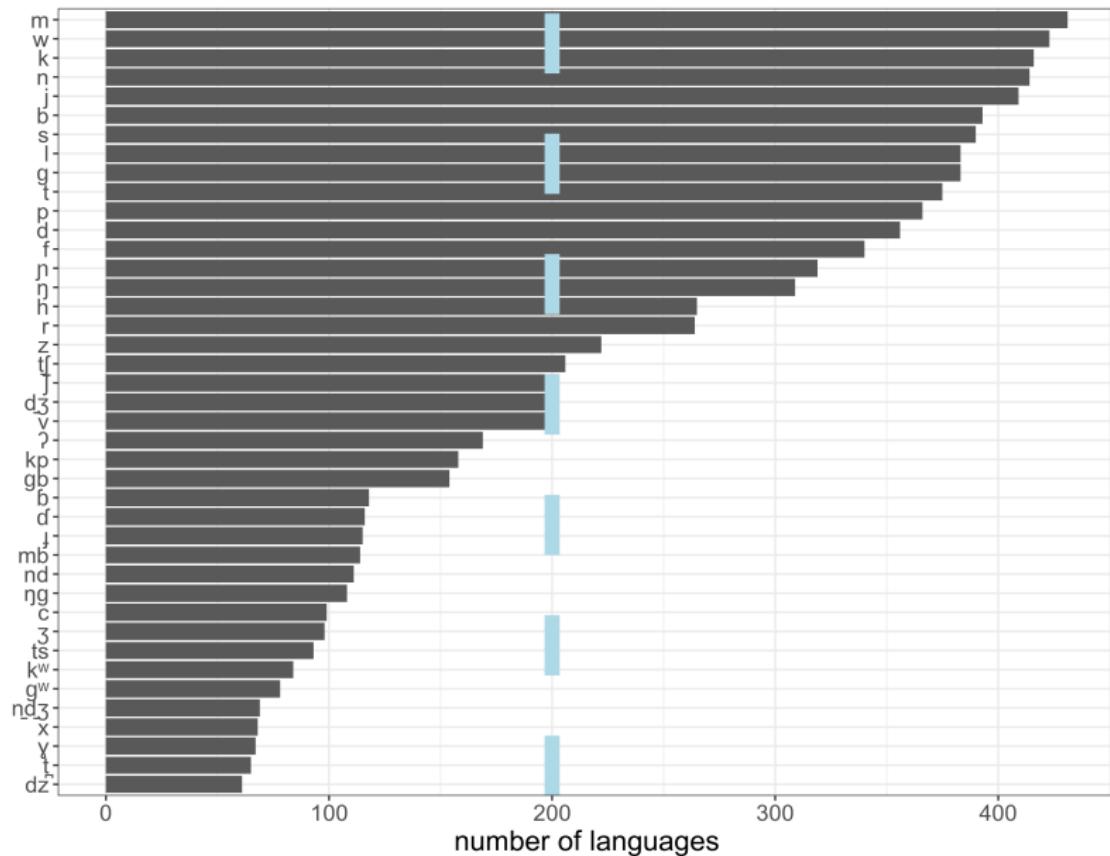
Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

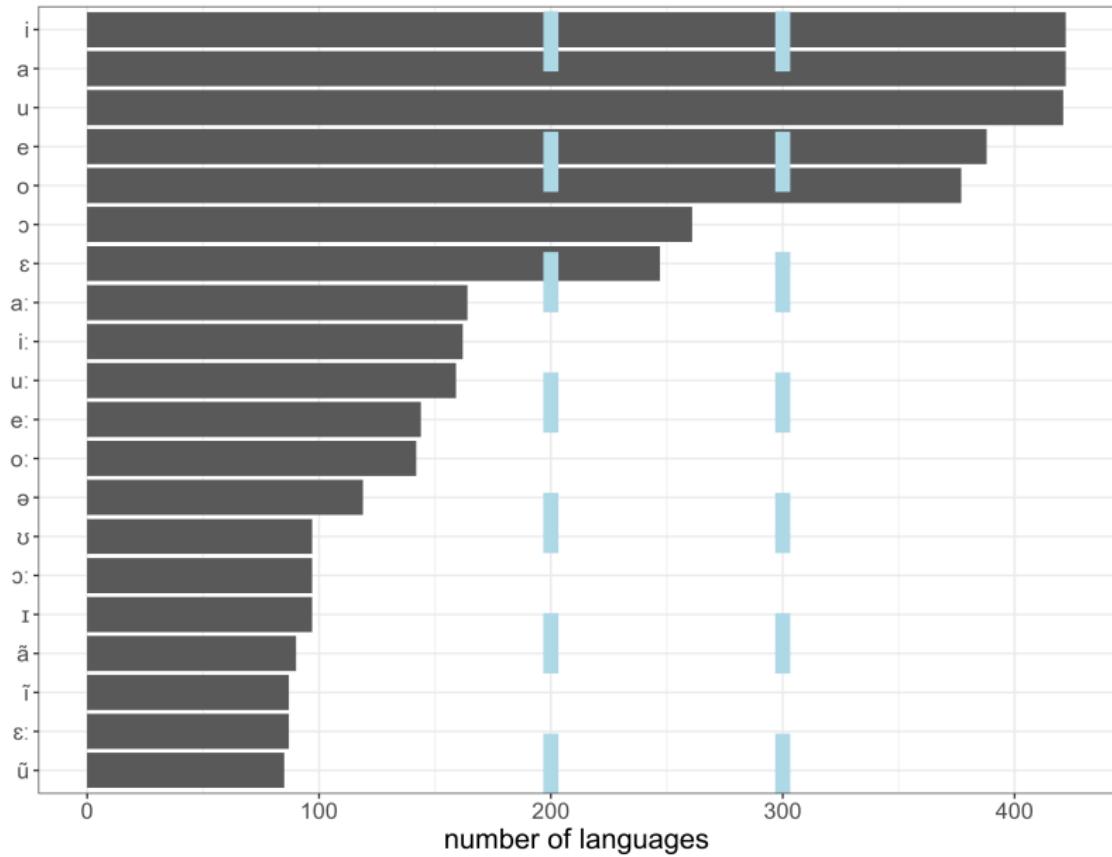
## PHOIBLE [Moran and McCloy 2019]: consonants



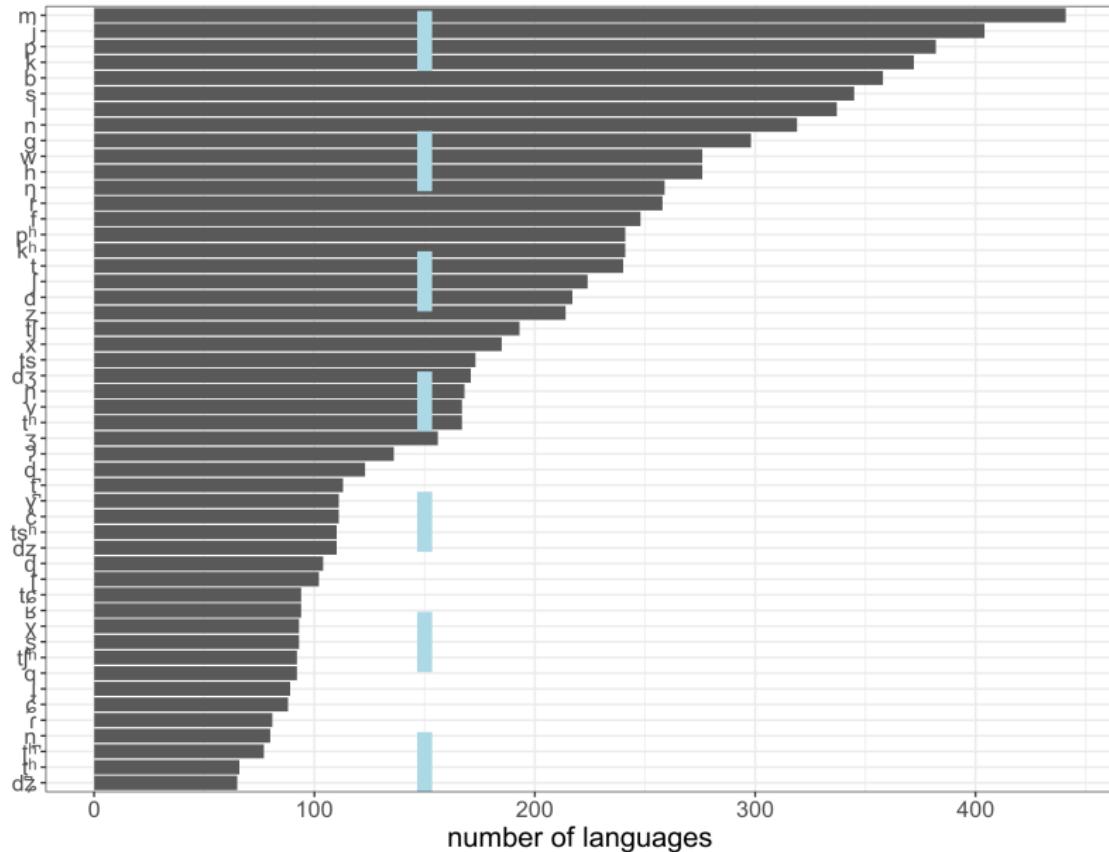
## PHOIBLE [Moran and McCloy 2019]: consonants

b	d		dʒ			g	
p	t		tʃ			k	
		z	ʃ				
f		s					h
m	n				jɪ	ɪj	
w	r			l	j		

## PHOIBLE [Moran and McCloy 2019]: vowels



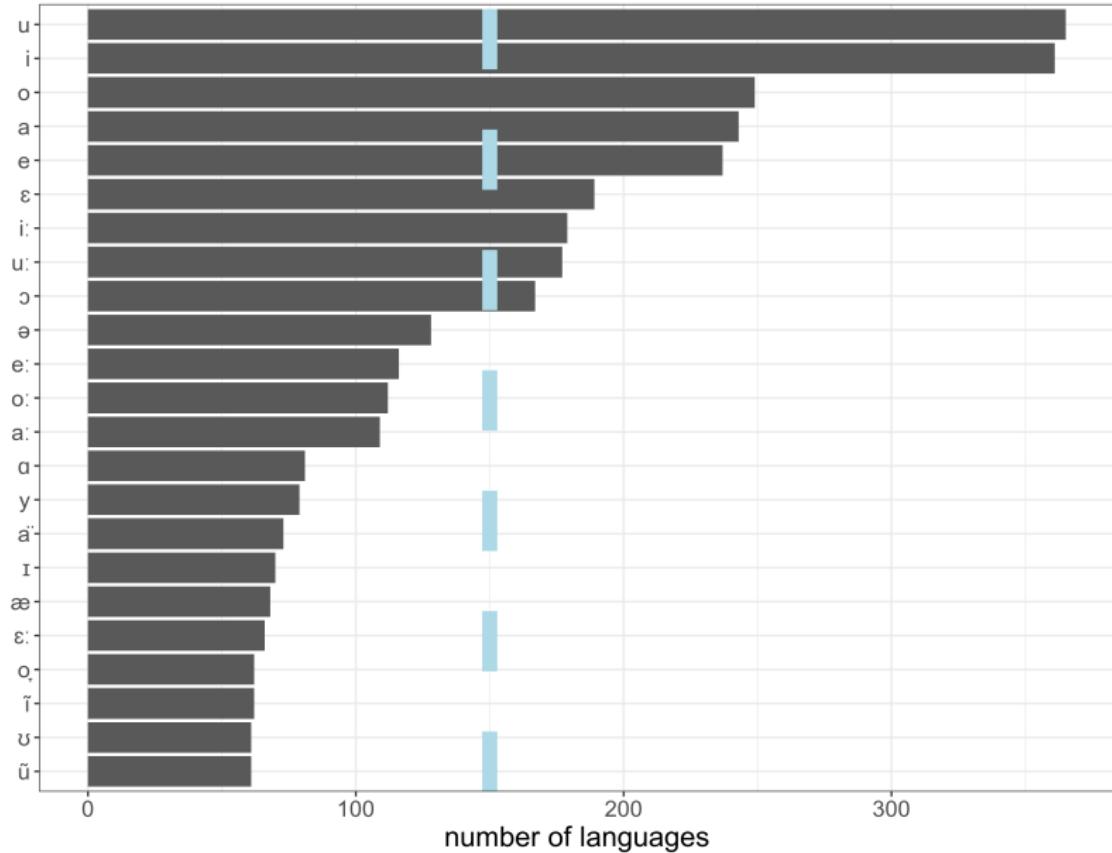
## Eurasian phonological inventories [Nikolaev et al. 2015]: consonants



## Eurasian phonological inventories [Nikolaev et al. 2015]: consonants

b	d		dʒ			g	
p	t	ts	tʃ			k	
p <sup>h</sup>	t <sup>h</sup>					k <sup>h</sup>	
		z	ʒ				
f	v	s	ʃ			x	h
m	n				jn	ŋ	
w	r			l	j		

## Eurasian phonological inventories [Nikolaev et al. 2015]: vowels



## Conclusions

- East Caucasian consonant systems are more or less typical

# Conclusions

- East Caucasian consonant systems are more or less typical
  - except uvular and laryngeal consonants
  - except ejective consonants
  - except laterals in Avar-Andic branch
  - except labialisation
  - except gemination

# Conclusions

- East Caucasian consonant systems are more or less typical
  - except uvular and laryngeal consonants
  - except ejective consonants
  - except laterals in Avar-Andic branch
  - except labialisation
  - except gemination
- East Caucasian vowel systems are typical

# Outline of the talk

Introduction

Data

Methods

Inventory size

Frequency analysis

Compare to other databases

Suprasegmental features

# Syllable structure and suprasegmental features

Syllable structure CVCC, sometimes CVL (L – sonorant or *b*).

A lot of East Caucasian languages has stress:

- some languages has more o less fixed stress (e. g. Mehweb)
- some languages has a tendency to have stress on first three syllables (Andic languages)
- a lot of languages distinguish morphological meanings via stress (e. g. plural marker in Dargic languages)

pharyngealization – mysterious feature (absent in Andic branch).

d-irʔ-an

npl-gather:ipfv-hab

‘gathers them’

d-irʔ-a<sup>f</sup>n /d-irʔ<sup>f</sup>-an/

npl-gather:ipfv-hab

‘gathers them’

ħa-d-irʔ-an

neg-npl-gather:ipfv-hab

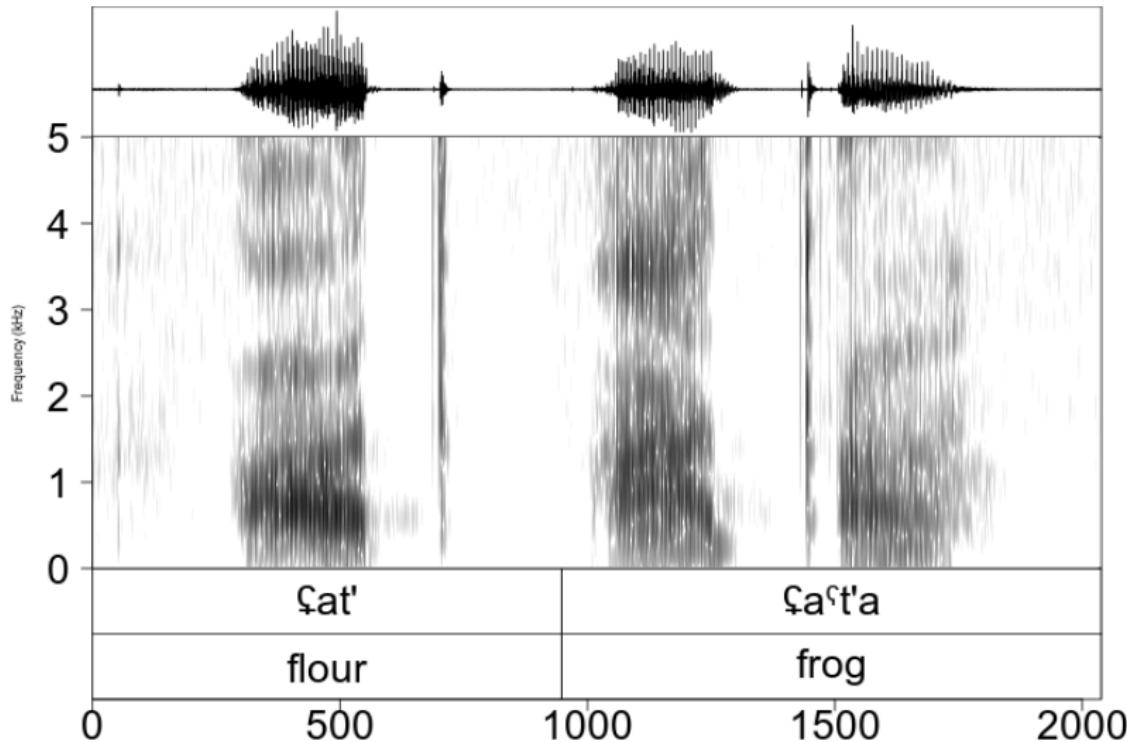
‘does not gather them’

ħa<sup>f</sup>-d-irʔ-a<sup>f</sup>n

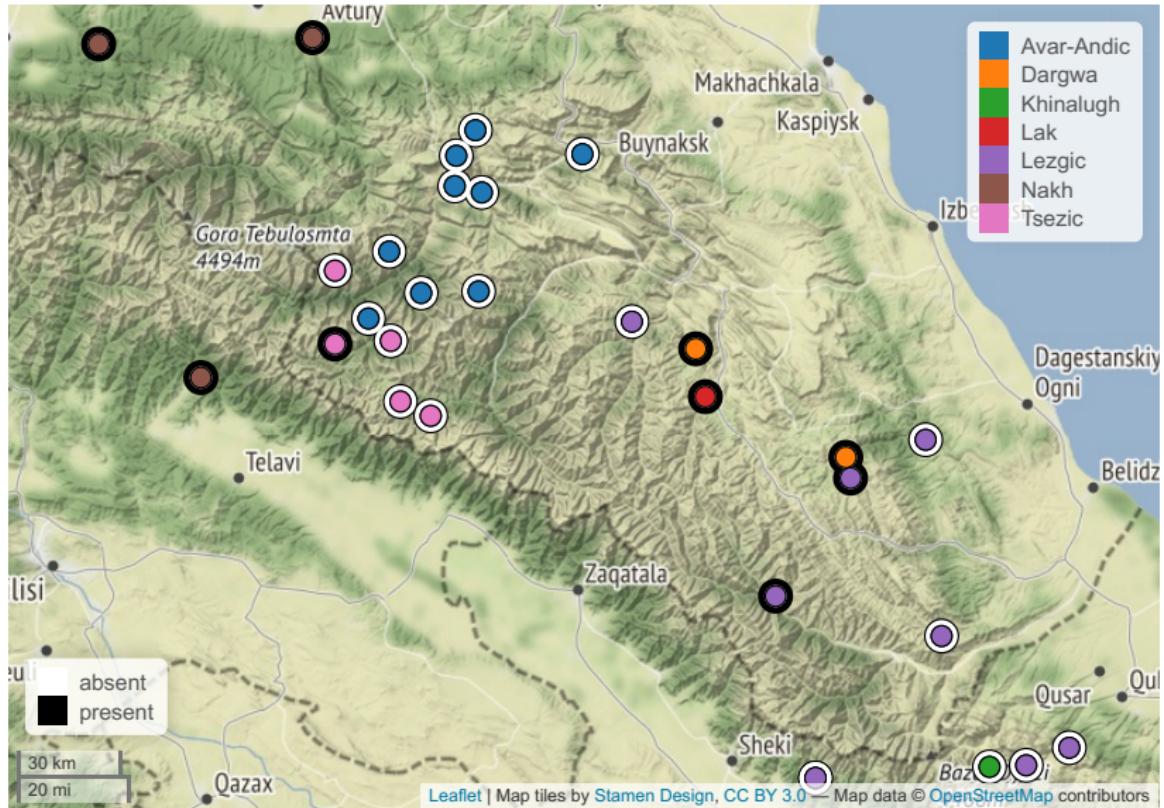
neg-npl-freeze:ipfv-hab

‘they are not freezing’

## Pharyngealization (example from Mehweb Dargwa)



# Pharyngealization



There are two enclaves: Nakh and Lak-Dargwa.

# Thanks

- to reasercher teams who collected all data
- to Viola Ivanova for helping with digitalisation of [Kibrik and Kodzasov 1990]
- to Linguistic Convergence Laboratory for organising this course

# Thanks

- to reasercher teams who collected all data
- to Viola Ivanova for helping with digitalisation of [Kibrik and Kodzasov 1990]
- to Linguistic Convergence Laboratory for organising this course
- and YOU for your attention

# Thanks

- to researcher teams who collected all data
  - to Viola Ivanova for helping with digitalisation of [Kibrik and Kodzasov 1990]
  - to Linguistic Convergence Laboratory for organising this course
  - and YOU for your attention
- 
- if you like presented maps, see [lingtypology](#) [Moroz 2017]
  - spectrograms were created with [phonfieldwork](#) [?]
  - this presentation made with [rmarkdown](#) [Xie et al. 2018] and [ggplot2](#) [wickham16]; source code and data for this presentation are available online: <https://tinyurl.com/y2oh9sg8>

## References

- Alekseev, M. E., Starostin, S. A., Klimov, G. A., and Testelets, Y. G. (2001).  
*Yazyki mira. Kavkazskiye yaziki*. Academia.
- Alexeyev, M. and Verhees, S. (202\_). Botlikh. In Koryakov, Y. and Maisak, T., editors, *Handbook of Caucasian Languages*, volume 2. De Gruyter Mouton, Berlin/New York.
- Beguš, G. (2021). Segmental phonetics and phonology. In Polinsky, M., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 1–40. Language Science Press.
- Boris, L. (2021a). Tone and intonation in languages of the caucasus. In Polinsky, M., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 1–30. Oxford University Press.

## References

- Boris, L. (2021b). Word stress in the languages of the caucasus. In Polinsky, M., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 1–30. Language Science Press.
- Grawunder, S. (2017). The caucasus. In Hickey, R., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 356–395. Cambridge University Press.
- Holisky, D. A. and Gagua, R. (1994). Tsova-Tush (Batsbi). *The indigenous languages of the Caucasus*, 4(part 2):147–212.
- Job, D. M. and Smeets, R., editors (1994). *The indigenous languages of the Caucasus*, volume 3. Caravan Books.
- Kibrik, A. E. and Kodzasov, S. V. (1990). *Sopostavitelnoye izuchenije dagestanskix yazykov: Imya. Fonetika*, volume 2. Moskovskij Gosudarstvennyj Universitet.

## References

- Kibrik, A. E., Tatevosov, S. G., Lyutikova, E. A., and Kazenin, K. I. (2001). *Bagvalinskiy yazik: Grammatika. Teksti. Slovari.* Naslediye, Moscow.
- Koryakov, Y. and Maisak, T., editors (202\_). *Handbook of Caucasian Languages*, volume 2. De Gruyter Mouton, Berlin/New York.
- Magomedbekova, Z. M. (1971). *Karatinskiy yazyk: grammaticheskiy analiz, teksty, slovar'*. Metsniereba, Tbilisi.
- Moran, S. and McCloy, D., editors (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena.
- Moroz, G. (2017). *lingtypology: easy mapping for Linguistic Typology*.
- Moroz, G. (2019). Phonology of mehweb. In Daniel, M., Dobrushina, N., and Ganenkov, D., editors, *The Mehweb language: Essays on phonology, morphology and syntax*, pages 17–39. Language Science Press.

## References

- Moroz, G., Naccarato, C., and Verhees, S. (202\_). Godoberi. In Koryakov, Y. and Maisak, T., editors, *Handbook of Caucasian Languages*, volume 2. De Gruyter Mouton, Berlin/New York.
- Nichols, J. (1994a). Chechen. *The indigenous languages of the Caucasus*, 4(part 2):3–77.
- Nichols, J. (1994b). Ingush. *The indigenous languages of the Caucasus*, 4(part 2):81–145.
- Nikolaev, D., Nikulin, A., and Kukhto, A. (2015). The database of eurasian phonological inventories.
- Smeets, R., editor (1994). *The indigenous languages of the Caucasus*, volume 4. Caravan Books.

## References

Xie, Y., Allaire, J., and Grolemund, G. (2018). *R Markdown: The Definitive Guide*. Chapman and Hall/CRC, Boca Raton, Florida. ISBN 9781138359338.