



Press Release

Seeking the origin of indigenous languages in South America

University of Tübingen computational linguist investigates kinships of the Tupí-Guaraní language family using methods from molecular biology

Dr. Karl Guido Rijkhoek
Director

Janna Eberhardt
Research reporter

Phone +49 7071 29-76788
+49 7071 29-77853

Fax +49 7071 29-5566
karl.rijkhoek[at]uni-tuebingen.de
janna.eberhardt[at]uni-tuebingen.de

www.uni-tuebingen.de/aktuell

Tübingen, 15.06.2023

A new study indicates that one of the largest of the indigenous language families in Latin America originated in the sixth century BCE in the basin of the Rio Tapajós and Rio Xingu, near the present-day city of Santarém in the Brazilian state of Pará. There are around fifty languages in the Tupí-Guaraní language family, which gave us words like “jaguar” and “piranha.” Now, Dr. Fabrício Ferraz Gerardi from the University of Tübingen’s Institute of Linguistics and a team of international researchers have used methods developed in the field of molecular biology to compare and investigate the Tupí-Guaraní languages. This has shed light on how the languages are related to each other, as well as on their geographical and chronological evolution. The new study has been published in the latest edition of *PLOS ONE*.

Little is known about the history of the Tupí-Guaraní language family. It includes about 40 languages still spoken today and at least another nine that have died out. The number of speakers per language ranges from less than one hundred, as in Amondawa and Juma, to six million, as in Paraguayan Guaraní. Only a few of the Tupí-Guaraní languages have been written down. “It is mainly the extinct languages that we know from phonetic transcriptions noted down by researchers in past centuries,” Fabrício Gerardi says.

Comparison of basic vocabulary

For the relationship analysis of the various Tupí-Guaraní languages, the research team used comparative lists of basic vocabulary. They asked, for example: Are the words for “leg,” “sing,” or “bat” the same or similar in the languages studied? Or do they not share a common root? “In molecular biological relationship analysis, for example of different animal or plant species, the respective gene sequences are used. They indicate which areas are the same or similar. The general random rate of gene

changes – mutations – can also be used to estimate how long ago two related species split off from a common ancestor," Gerardi explains. The mutations in the genes of biological species correspond to phonetic shifts or substitutions in related languages. Thus, in Tupinambá, one of the Tupí-Guaraní languages, the tapir is called "tapi?ir"; in Awetí, a language that split off from these languages, it is called "tapi?it."

Large-scale analyses of the vocabulary and grammatical structures of the Tupí-Guaraní languages using algorithms from molecular biology can be used to create a family tree. "We wanted to know what the tree looked like, how strongly related individual languages were to each other, how old each language was, and when it split into new languages," Gerardi says.

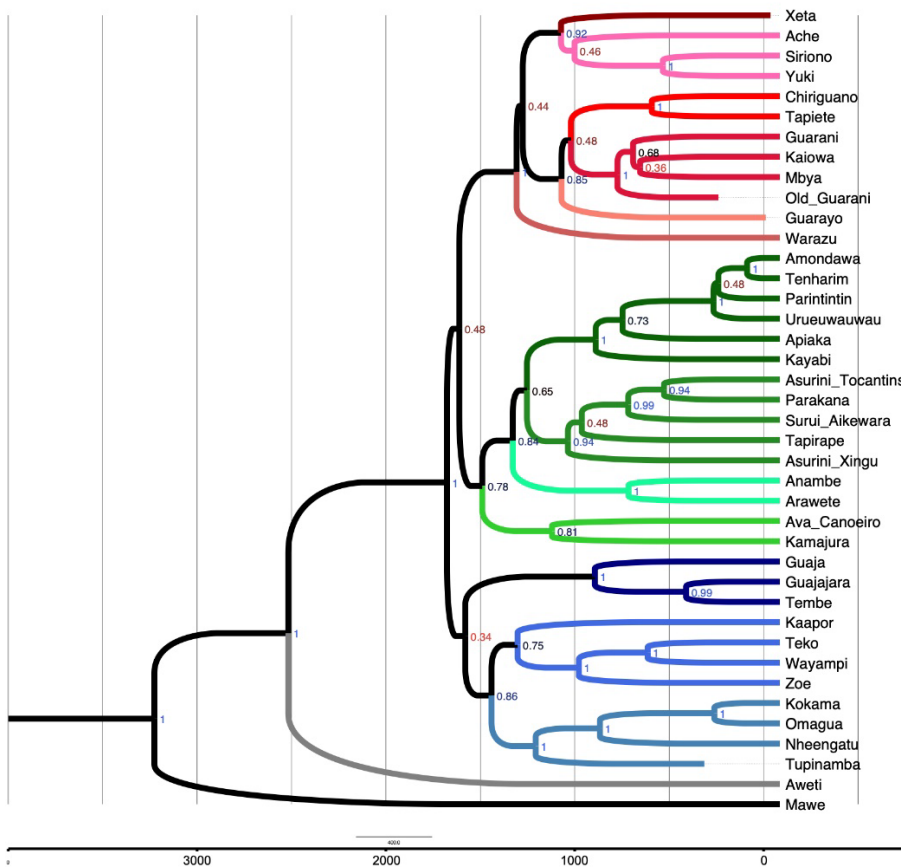
Timeframe calibrated via archeological finds

The distribution of the Tupí-Guaraní language family extends more than 4000 kilometers in both width and breadth. "In some cases, we have archaeological finds from the same area that we try to assign to the individual languages. For example, there are certain words in the languages for describing special properties of the ceramics discovered there," says Gerardi. "This allows us to establish a temporal and spatial relationship between the language and the archaeological finds. The ceramics could be dated using the radiocarbon method – so we indirectly have a temporal calibration of language development," he adds. Along the way, Gerardi and the research team were able to pinpoint the probable place of origin of the Tupí-Guaraní language family to the Tapajós-Xingu basin some 2,550 years ago. "However, to better corroborate our findings, the archaeological and linguistic evidence would need to be further explored," he says.

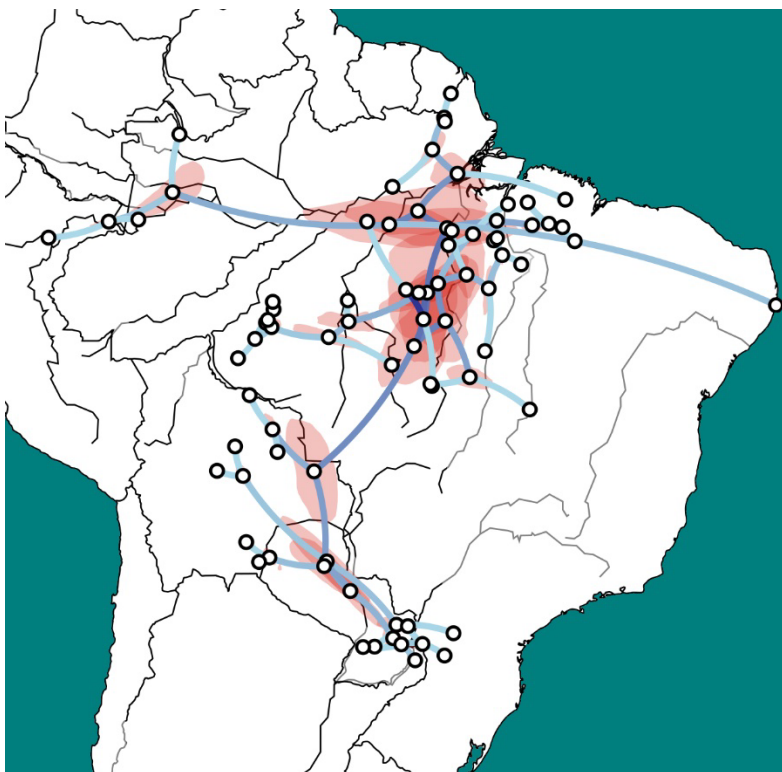
	Mawé	Awetí	Tupinambá
Leg	ʔup	ʔup	etimã
Sing	mepi	tepi	neʔenggar
Come back	aipok	ʔajpog	jeβir
Hoplias (genus)	(n)ipiuta	piutá	tare?ira
Fly (insect)	win	tin	meru
Jaguar	awiato	tawat	jawar
Anteater	arihĩ	tamajua	tamandwa
Grandfather	ase?i	amũj	amĩja
Wound	pihi	pezep	pereb
Tapir	wewato	tapi?it	tapi?ir
Sky	atipi	iwak	iβak
Genipa	wããhop	tētīpap	janipab
Bat	haki?i	tati?a	anira
Sieve (tool)	panane	kurupem	urupem
Burn (something)	wuk	api	api
Bow	moreawat	zapat	iβirapar
Star	wajkiru	tatia?it	jasitata

Tupinambá belongs to the Tupí-Guaraní language family; the Mawé and Awetí languages have split off from it and developed differently. Related words in Mawé and Awetí that do not exist in Tupinambá are highlighted in yellow. Related words in Awetí and Tupinambá, for which there is a completely different term in Mawé, are highlighted in blue. Figure: Gerardi et al., 2023, *PLOS ONE*, CC-BY 4.0

(<https://creativecommons.org/licenses/by/4.0/>)



Family tree with dating of the Tupí-Guaraní languages, from which Awetí and Mawé previously split. Figure: Gerardi et al., 2023, *PLOS ONE*, CC-BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)



Model of the geographic and kinship relationships of the Tupí-Guaraní language family: light blue lines indicate a kinship; a darker color indicates earlier migration/ separation. Geographical areas where there is an 80% probability of language separation are colored red. Figure: Gerardi et al., 2023, *PLOS ONE*, CC-BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)

Publication:

Fabrício Ferraz Gerardi, Tiago Tresoldi, Carolina Coelho Aragon, Stanislav Reichert, Jonas Gregorio de Souza, Francisco Silva Noelli: Lexical Phylogenetics of the Tupí-Guaraní Family: Language, Archaeology, and the Problem of Chronology. *PLOS ONE*,
<https://doi.org/10.1371/journal.pone.0272226>

Contact:

Fabrício Ferraz Gerardi
University of Tübingen
Humanities
Institute of Linguistics
Phone +49 7071 29-77303
fabricio.gerardi[at]uni-tuebingen.de