

**Jihočeská univerzita v Českých Budějovicích**  
**Přírodovědecká fakulta**

**Koevoluce genů a jazyků – kritická rešerše**

Bakalářská práce

**Klára Vodičková**

Školitel: RNDr. Pavel Duda, Ph.D.

Konzultant: prof. RNDr. Jan Zrzavý, CSc.

České Budějovice 2023

Vodičková, K. 2023: Koevoluce genů a jazyků – kritická rešerše. [Coevolution of genes and languages – a critical review. Bc. Thesis. in Czech] – 46 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

### **Anotace**

In this bachelor's thesis, I review studies that utilize phylogenetic or population-genetic methods to study coevolution or correlation between genes and languages in different regions of the world and at different spatial scales. First, I describe the theoretical foundations of the study of gene-language coevolution and the reasons for expecting correspondence between population history and language history or lack thereof. Next, I review published studies, the data and methods they use, and their results. Finally, I discuss the limitations of the data and methods used to uncover coevolution or correlation between genes and languages.

### **Prohlášení**

Prohlašuji, že svoji bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

V Českých Budějovicích, 7. 12. 2023

.....

Klára Vodičková

## **Poděkování**

Ráda bych poděkovala svému vedoucímu bakalářské práce RNDr. Pavlu Dudovi, Ph.D. za odborné vedení, cenné rady a trpělivost při zpracování této práce.

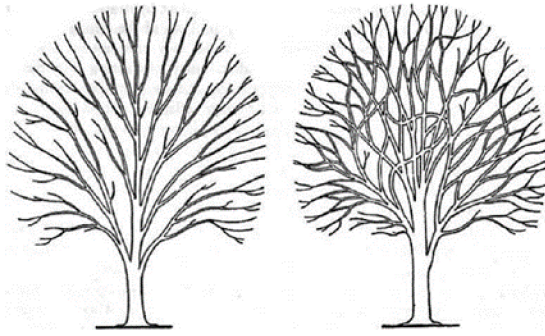
## **OBSAH**

1. ÚVOD .....	1
1.1. Argumenty pro a proti koevoluci genů a jazyků .....	3
1.2. Koevoluce a její předpoklady .....	5
1.3. Genetická data a genetické fylogeneze .....	6
1.4. Lingvistická data, jazyková klasifikace a jazyková fylogeneze .....	8
2. VÝSLEDKY .....	12
2.1. Evropa.....	14
2.2. Kavkaz .....	18
2.3. Asie .....	19
2.4. Indonésie a Melanésie .....	21
2.5. Afrika.....	24
2.6. Svět .....	28
3. DISKUSE A ZÁVĚR .....	32
4. SEZNAM LITERATURY .....	38
5. SEZNAM LITERATURY – TABULKA.....	43

## 1. ÚVOD

Myšlenka společné evoluce genů a jazyků je předmětem trvalého zájmu antropologů, lingvistů a genetiků. Jako první se jí zabýval již Charles Darwin, který v knize *O vzniku druhů* (Darwin, 1859) navrhl, že dokonalý rodokmen lidstva, který by věrně odrážel příbuzenské vztahy lidských etnik, by zároveň umožňoval přirozenou klasifikaci jazyků. Koevoluce genů a jazyků se však začala skutečně zkoumat až mnohem později, v 80. letech 20. století. Na počátku tohoto výzkumu stál italský populační genetik Luigi Luca Cavalli-Sforza, který jako první porovnal fylogenetický strom lidských populací, sestavený na základě genetických dat se stromem, založeným na klasifikaci jazyků (Cavalli-Sforza et. al., 1988). Jeho práce, ač kontroverzní, vzbudila zájem o aplikaci fylogenetických a molekulárně-genetických metod na lingvistická data a jejich potenciál pro rekonstrukci populační historie lidstva (Atkinson & Gray, 2005). V návaznosti na tuto práci byla na konci 80. a počátku 90. let publikována řada studií, které se snažili posoudit shody a neshody mezi distribucí genetických variant a jazyků nebo jazykových rodin (Barbujani & Sokal, 1990; Cavalli-Sforza et al., 1992; Harding & Sokal, 1988; Sokal 1989). Cavalli-Sforza byl přesvědčen, že znalost molekulární fylogeneze lidských populací umožní predikovat vztahy mezi jazyky a naopak (Cavalli-Sforza, 1991).

Antropologové a lingvisté však byli k myšlence propojení populační historie a historie jazyků vždy skeptičtí (Boas, 1911; Sapir, 1931). Důsledně odmítali představu, že kulturní změny jsou jakkoli navázané na biologické změny a zároveň zdůrazňovali rozdíly mezi genetickou dědičností a kulturním přenosem. Geny se dědí „vertikálně“, z generace a na generaci, zatímco jazyk má schopnost se šířit „horizontálně“, mezi nepříbuznými jedinci a populacemi. Zatímco evoluce druhů a populací je charakterizována postupným vývojem v rámci příbuzenských linií (anageneze) a jejich štěpením (kladogeneze), v evoluci jazyků dochází velmi často k horizontálnímu přenosu – splývání dříve vzniklých linií (syngeneze). Rodokmen jazyků, o kterém psal Darwin, tak vůbec není stromovitý, ale odpovídá spíše hustému keři se vzájemně propletenými větvemi (Kroeber, 1923; Obr. 1.). Tuto představu, která se nazývá *difusionalismus*, zastávali nejen mnozí lingvisté, ale také někteří evoluční biologové (např. Gould, 1991).



**Obr. 1.** Strom života a strom lidské kultury (Kroeber, 1923)

V posledních dekádách si biologové uvědomili, že biologická evoluce nemusí být pouze stromovitá. K horizontálnímu přenosu dochází na všech úrovních stromu života. Například mitochondrie a chloroplasty jsou původem bakteriální symbionti, přičemž k horizontálnímu přenosu jejich genů do jádra buňky dochází podnes. Retrotranspozony, původně parazitické úseky DNA, se často přenášejí mezi druhy. Rozsáhlý horizontální přenos DNA z nepříbuzných organismů (včetně rostlin a hub) byl popsán u vířníků pijavenek (*Bdelloidea*). Mezdruhová hybridizace je evidentně velmi běžná. K druhům hybridního původu patří třeba jelen milu (*Panolia eldi*), vlk rudohnědý (*Canis rufus*), paviánek kipunji (*Rungwecebus kipunji*), levhart obláčkový (*Pantera uncia*), nebo koneckonců moderní člověk (*Homo sapiens*), jehož mnohé populace vznikly hybridizací raně moderních lidí s neandrtálci (Duda, 2017; Pääbo, 2015). Existence horizontálního přenosu neznamená, že by biologická evoluce nebyla víceméně stromovitá, ostatně jenom díky existenci fylogeneze můžeme horizontální přenos odhalit.

Dnešní kritici myšlenky koevoluce genů a jazyků však netvrdí, že ke koevoluci nedochází, spíše poukazují na to, že se nejedná o čistě paralelní vývoj, ale že za vývojem jazyků stojí složitější procesy, které je třeba při studiu koevoluce zohlednit.

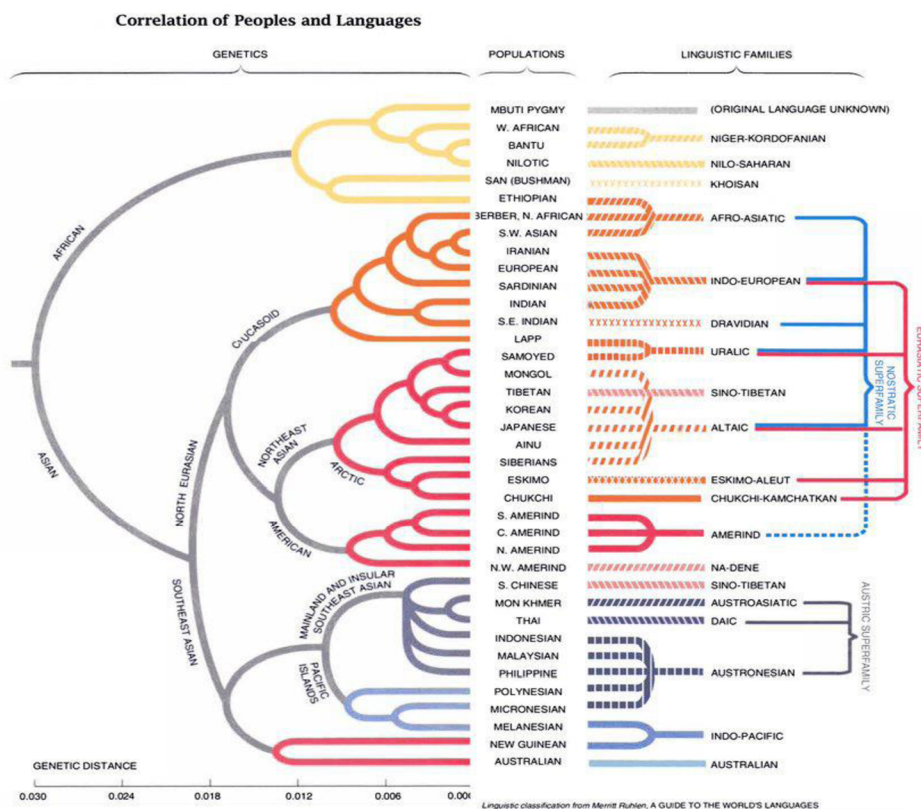
Od publikování prvních studií na toto téma uplynulo více než třicet let této problematice bylo věnováno i několik rešerší (např. Campbell, 2015; Pakendorf, 2014, 2015). Dosud ale nevznikla práce, která by publikované studie shromáždila, porovnála a diskutovala v nich použitá data a metody a jejich výsledky. Cílem této bakalářské práce je:

- 1) Popsat teoretická východiska studia koevoluce genů a jazyků.
- 2) Shromáždít případové studie, které posuzují a kvantifikují koevoluci či korelaci mezi geny a jazyky formálními fylogenetickými nebo populačně-genetickými metodami.

- 3) Popsat jejich výsledky a posoudit v nich použitá genetická a lingvistická data a statistické metody a jejich vhodnost pro studium koevoluce.
- 4) Shrnout současný stav našich znalostí koevoluce genů a jazyků, popsat praktické problémy spojené se studiem těchto evolučních systémů a jejich možná řešení.

### 1.1. Argumenty pro a proti koevoluci genů a jazyků

Důvody, proč očekávat, že evoluční historie a výsledná geografická distribuce genů a jazyků si budou vzájemně odpovídat, jsou obecně srozumitelné. Jazyk se učíme od rodičů, takže se předává spolu s geny. Populační expanze nebo migrace a následná geografická izolace, které vedou ke genetické diferenciaci, pravděpodobně povedou také k jazykové diverzifikaci. Jazyky fungují jako bariéry pro tok genů, podobně jako bariéry geografické (moře, pohoří apod.). V důsledku lze očekávat korespondenci mezi distribucí genetických variant a lingvistických uskupení (Cavalli-Sforza et. al., 1988; Cavalli-Sforza, 1991; Obr. 2).



**Obr. 2.** Srovnání genetické fylogeneze, založené na souboru klasických polymorfismů (vlevo) a jazykové klasifikace na úrovni jazykových rodin a velkorodin (vpravo) (Cavalli-Sforza, 1991)

Hlavní argumenty proti univerzální koevoluci genů a jazyků jsou následující: Zatímco geny se šíří vertikálně, z rodičů na potomky, jazyk se může šířit také horizontálně, bez ohledu na příbuzenství. Neexistuje žádný deterministický vztah mezi tím, jaké geny v sobě nosíme a jakým jazykem hovoříme. Zatímco člověk má jen jeden genom a populace jeden genofond, lidé a jejich populace mohou hovořit i více jazyky. Jazyky se rychle mění, běžně vymírají a vymření jazyka zpravidla neznamena vymření populace mluvčích, ale přechod na jiný jazyk, což narušuje očekávanou korespondenci mezi geny a jazyky (Campbell, 2015; McMahon, 2004).

V některých oblastech světa je praktikována tzv. jazyková exogamie, kdy je pravidlem, že partneři mluví odlišným jazykem. Buďto hovoří každý jiným jazykem, ale navzájem si rozumí, nebo jeden z partnerů přijme jazyk toho druhého. V oblastech, kde je tato forma exogamie praktikována (např. Vaupés nebo Chorote-Nivaclé-Wichí v Jižní Americe) nelze korespondenci mezi geny a jazyky očekávat. Lingvisté často poukazovali na existenci jazykových svazů (*Sprachbunde*), kde dochází k intenzivním výpůjčkám slov i gramatiky mezi vzdáleněji příbuznými jazyky. Jazykové svazy tvoří např. jazyky Balkánského poloostrova, jazyky v Pobaltí, Jižní Africe, na Etiopské vysočině, v Indii a Pákistánu, na západním pobřeží USA a Kanady nebo ve Střední Americe (Campbell, 2015).

Lidé často migrují jak na krátké, tak na dlouhé vzdálenosti. Dálkové migrace mají různé příčiny, např. klimatické změny nebo válečná tažení a s nimi spojené územní expanze. Lidé často migrují z ekonomických, náboženských nebo politických důvodů, někdy spontánně, jindy nuceně (Bhugra et al., 2011). Migranti, nesoucí si své geny i jazyk, přicházejí na své cestě a v místě, kde se usadí, do kontaktu s jinými populacemi. Často mezi nimi následně dochází k toku genů, v místě usazení časem i k výpůjčkám slov a ovlivňování jazyků, které jsou spolu dlouhodobě v kontaktu. Často uváděným příkladem dálkové migrace a jejich důsledků pro genetické a jazykové složení populace jsou Maďaři. Maďaři jsou po genetické stránce příbuzní středoevropským etnikům, hovořícími slovanskými a románskými jazyky, zatímco maďarština je příbuzná ugrofinským (uralským) jazykům Khanty, Mansi a Mari, kterými se mluví v povodí Obu. V 9. století přišli na území dnešního Maďarska lidé, kteří migrovali z oblasti Uralu. Díky jejich vojenské a ekonomické převaze místní obyvatelé postupně přešli na ugrofinský jazyk, zatímco méně početná skupina migrantů geneticky splýnula s místní populací (Longobardi et al., 2015; Tömöry et al., 2007).



Korelaci mezi geny a jazyky narušuje také vznik kreolských jazyků a jazyků širší komunikace. Kreolské jazyky vznikají smísením a zjednodušením různých jazyků. Tyto jazyky jsou založené převážně na evropských jazycích, které se šířily v době kolonialismu, např. na angličtině nebo francouzštině, vznikaly však i z mimoevropských jazyků, např. z arabštiny. Mají jednodušší gramatiku a stabilní slovní zásobu, která pochází z více jazyků (Romaine, 2017). Na podobném principu jsou založeny i jazyky širší komunikace, které vznikají mísením a zjednodušováním místních dialektů. Prvním takovým jazykem byla zjednodušená forma řečtiny, nazývaná koiné, kterou se hovořilo na území, kam zasahoval řecký kulturní vliv ve 4. stol. př. n. l. Tento jazyk mimo jiné obsahoval řadu slov, převzatých z jiných jazyků. Termín koiné je dodnes používán jako obecné označení pro jazyky širší komunikace (Colvin, 2011). Podobné důsledky má kodifikace jazyka, kdy dochází ke stanovení pravidel pravopisu, výslovnosti i slovníku (Daneš, 2009). Ke kodifikaci jazyka nedochází spontánně, ale nařízením shora, a vede k homogenizaci jazyka v rozsáhlé oblasti, která nemusí být homogenní geneticky.

Posledním procesem narušujícím očekávanou korelaci mezi geny a jazyky je vymírání jazyků, které je poměrně častým jevem (Amano et al., 2014). Podle databáze Ethnologue (Eberhard et al., 2023) je v současnosti ohroženo vyhynutím 3072, tedy přes 40 % všech jazyků. Celkem 451 popsaných jazyků je vedeno jako vyhynulých. Naprostá většina populací, které hovořily těmito jazyky, však nevymřela, ale přešla na jiný jazyk. Protože vymření jazyka není navázáno na vymření populace mluvčích, jazykové extinkce narušují korespondenci mezi jazyky a geny. Takzvané jazykové izoláty, jazyky, které nejsou zjevně příbuzné žádným jiným jazykům, jsou pravděpodobně pozůstatky větších jazykových rodin (Urban, 2021).

## **1.2. Koevoluce a její předpoklady**

Pod pojmem koevoluce se míní vztah mezi dvěma nebo více organismy, které spolu interagují a navzájem se tak ovlivňují (Thompson, 2014). Příkladem koevolučního systému je systém hostitel-parazit. Hostitel vytváří selekční tlak na parazita tím, že se snaží vyvinout obranné mechanismy proti němu. Parazit na to reaguje tím, že se snaží vyvinout mechanismy, díky kterým se obraně hostitele vyhne nebo ji překoná. Důsledkem koevoluce mezi druhy jsou podobnosti v topologii (pořadí uzlů) a chronologii (stáří uzlů) jejich fylogenezí, kdy speciální událost u jednoho druhu má za následek speciální událost u druhého druhu (Huelsenbeck et al., 1997; Page, 2003). Standardní metodou pro odhalení koevoluce je proto porovnávání fylogenetických stromů. Koevoluci tohoto typu však mezi geny a jazyky nelze očekávat, protože jazyky se vyvíjejí odlišnou rychlostí a mají částečně odlišnou dědičnost.

Klíčovým faktorem, ovlivňujícím populačně-genetickou strukturu populací je geografie. Fyzické bariéry jako jsou hory, řeky a moře, podporují paralelní vývoj genů a jazyků. Možná ještě významnější roli však hraje samotná geografická vzdálenost, která rovněž ovlivňuje genetickou i jazykovou diverzitu a její distribuci. Mezi sousedícími a blízkými populacemi dochází častěji ke genetickému mísení a k horizontálnímu přenosu, s přibývajícím vzdáleností horizontálního přenosu ubývá. Tento proces se nazývá izolace vzdáleností (*isolation-by-distance*). Pokud za vývojem genetické a jazykové diverzity stojí především tento proces, většina pozorovaných shod mezi geny a jazyky je důsledkem prostorové autokorelace (Peter et al., 2020). Klíčové otázky pro ty, kteří se koevolucí genů a jazyků zabývají, tedy jsou, jak častý je horizontální jazykový přenos a jak významný je vliv procesu izolace vzdáleností.

### 1.3. Genetická data a genetické fylogeneze

Publikované fylogeneze organismů se zakládají nejčastěji na analýze molekulárně-genetických dat: nejčastěji sekvencí nukleotidů v molekule DNA. Srovnávané sekvence jsou homologické, to znamená, že pozice v jednotlivých nukleotidech či aminokyselinách jsou zděděny od společného předka a rozdíly mezi nimi jsou důsledkem proběhnuvších změn – substitucí, inzercí či delecí. Homologické sekvence jsou k sobě přiřazeny (*alignment*) a z rozdílu mezi nimi jsou vypočítány párové vzdálenosti, z nichž se vytvoří tzv. návodný strom (*guide tree*), na jehož základě jsou pak k sobě postupně přiřazovány všechny sekvence počínaje těmi nejpodobnějšími až po ty nejvíce odlišné. (Macholán, 2014). Z těchto dat se dále určuje genetická vzdálenost taxonomických jednotek – druhů nebo populací. Tato vzdálenost je odvozena buď z počtu odlišných bází ve sledovaných sekvencích, nebo z počtu mutací, kterými se sekvence liší. Vedle fylogenezí lidských populací, založených na molekulárně-genetických datech, existují i fylogeneze založené na morfologických datech, kdy za homologii je brána podoba nebo přítomnost/nepřítomnost určitého morfologického znaku. V antropologii jsou tato data používána zřídka, nejčastěji se jedná o kraniální znaky (Macholán, 2014).

Při studiu populační historie lidstva byly využívány různé typy genetických dat. Nejstarší studie vzniklé v 80. a 90. letech minulého století používaly tzv. klasické polymorfismy – krevní skupiny, nebo jiné proteinové alomorfismy (Cavalli-Sforza et al., 1988; Harding & Sokal, 1988). Koncem 90. let se začaly využívat tzv. uniparentální markery, tedy mitochondriální DNA (mtDNA), respektive její hypervariabilní oblast a nerekombinující část chromozomu Y (NRY), které se dědí pouze po mateřské, respektive otcovské linii,

díky čemuž mohou poskytovat odlišný obrázek o historii téže populace (Bruchová et al., 2005).

Po roce 2010 se začala více využívat autozomální DNA, někdy v kombinaci s uniparentálními markery. Nejnovější studie pak používají data genomická, nejčastěji tzv. jednonukleotidové polymorfismy a krátké tandemové repetice. Jednonukleotidové polymorfismy (*Single Nucleotide Polymorphisms*, SNPs) jsou krátké sekvence, lišící se v jediném nukleotidu, které se zpravidla vyskytují v nekódujících oblastech DNA. Pokud SNPs leží na stejném chromozomu a v dostatečné blízkosti, tedy ve vazbě, dostávají se do další generace společně. Variance v SNPs jsou v populacích mapovány a tím umožňují sledování příbuznosti mezi jedinci a dávají vznik komplexnější genetické mapě takové populace. Krátké tandemové repetice (*Short Tandem Repeats*, STRs), neboli mikrosatelity, jsou opakující se krátké specifické sekvence DNA, které se skládají ze stejné jednotky dvou nebo více nukleotidů. Stejně jako SNPs se nejčastěji nacházejí v nekódujících oblastech DNA. Díky vysoké mutační rychlosti jsou mikrosatelity velmi variabilní, a to i v rámci jedné populace (McDonald, 2013). SNPs a STRs patří do kategorie tzv. genomických (*genome-wide*) dat, která jsou rozptýlena napříč celým genomem. Vedle toho se v nejnovějších studiích stále častěji objevují celogenomová data (*whole-genome*), která se zakládají na práci s celým, nebo téměř celým genomem zkoumaného organismu.

K tvorbě fylogenezí se užívají různé metody, které jsou vhodné pro zpracování různých typů dat. Mezi běžně používané metody se řadí distanční metody jako metoda neváženého párování s aritmetickým průměrem (*Unweighted Pair Group Method with Arithmetic Mean*, UPGMA), *Neighbor joining* (NJ), metoda maximální úspornosti (parsimonie), metoda maximální věrohodnosti (*maximum likelihood*) a Bayesovská analýza. Distanční metody vytvářejí příbuzenský vztah z matice párových vzdáleností, která převádí podobnosti ve studovaných znacích do číselné podoby. Tím se ztrácí informace o homologii, ale tyto metody jsou díky tomu velmi rychlé. Distanční metody UPGMA a NJ jsou shlukovací algoritmy, které postupně seskupují příbuzné druhy či populace do shluků, až dojde k vytvoření hierarchicky strukturovaného, zakořeněného či nezakořeněného stromu (Macholán, 2014). Problémem těchto metod je, že předpokládají konstantní mutační rychlost a nepracují s možností horizontálního přenosu (Macholán, 2014). Pokročilejší distanční metody, které vytváří fylogenetické sítě, např. *Neighbor-Net* (Bryant & Moulton, 2004), dokáží vizualizovat konflikt v datech, který lze interpretovat jako důsledek horizontálního přenosu.

Metoda maximální úspornosti (*Maximum parsimony*) byla po dlouhou dobu tou nejpoužívanější metodou fylogenetické rekonstrukce, díky její rychlosti a přímočarosti. Zakládá se na upřednostňování jednodušších fylogenetických hypotéz před těmi složitějšími, tedy na minimalizaci počtu evolučních kroků, nutných pro vysvětlení pozorovaných rozdílů v sekvencích či morfologii. Sdílené společné stavy znaků se vysvětlují dvěma způsoby – buď jsou zděděny od společného předka, nebo vznikly nezávisle na sobě jako důsledek homoplazie (Macholán, 2014).

Metoda maximální věrohodnosti (*Maximum Likelihood*) vysvětluje data pomocí hypotézy, při jejíž platnosti jsou pozorovaná data nejvíce věrohodná. (Macholán, 2014). Podobně funguje i Bayesovská analýza, která na základě určité hypotézy odhaduje, s jakou pravděpodobností vznikají pozorovaná data, avšak s ohledem na předpoklady o těchto datech. V tom se liší od metody maximální věrohodnosti (Macholán, 2014).

Principem molekulárních hodin je, že rychlost změn/mutací v DNA či proteinových sekvencích je v průběhu času konstantní, díky čemuž je možné stanovit čas divergence druhů a odhadnout stáří různých evolučních událostí. Tato metoda molekulárních hodin má vážné nedostatky, protože některé geny se mohou vyvíjet rychleji a snáze mutovat, na což může mít vliv i prostředí, ve kterém se organismy nacházejí (Bromham & Penny, 2003). Novější metody molekulárních hodin umožňují zohlednit odlišné mutační rychlosti v různých oblastech genomu a v průběhu času (Macholán, 2014).

#### **1.4. Lingvistická data, jazyková klasifikace a jazyková fylogeneze**

Dříve než se začaly k tomuto účelu využívat fylogenetické metody, byla jediným zdrojem příbuzenských stromů jazyků jazyková klasifikace, ve které byly jazyky seskupeny do vnitřně uspořádaných příbuzenských skupin, nazývaných jazykové rodiny. Jazyky, patřící do jedné jazykové rodiny, pochází ze společného předka neboli prajazyka. Konvenční metoda vytváření jazykové klasifikace tzv. komparativní metoda se zakládá na porovnání dvou a více potenciálně příbuzných jazyků – jejich slovníku, gramatiky, syntaxe a fonologie (Millar & Trask, 2015). Pokud je podobností dostatek a není možné je vysvětlit např. jazykovým kontaktem, je pravděpodobné, že tyto jazyky pochází ze společného předka. Postupným srovnáváním více a více jazyků lze ustanovit existenci jazykové rodiny, která zahrnuje dva jazyky, nebo více než 1000 jazyků.

Vedle toho existuje metoda tzv. masové komparace, kterou prosazoval lingvista Joseph Greenberg (Millar & Trask, 2015). Tato metoda se zakládá na srovnávání stovek až tisíců slov napříč mnoha nepříbuznými jazyky, přičemž jejich zvuková nebo významová podobnost je vnímána jako doklad možného příbuzenského vztahu i nad rámec jazykové rodiny. Takto vzniklá uskupení se nazývají jazykové velkorodiny. Většina lingvistů jazykové velkorodiny neuznává, protože podobnosti, na kterých se zakládají, mohou být víceméně náhodné. Greenberg pomocí této metody klasifikoval jazyky celého světa s různým výsledkem. Zatímco jeho klasifikace afrických jazyků do čtyř velkorodin byla ve své době dobře přijata a některé z těchto uskupení jsou uznávány dodnes (jazyky nigerokonžské nebo afroasijské), jiné velkorodiny jako např. amerindská nebo indopacifická nikdy uznávány nebyly.

Fylogeneze jazyků vychází z klasické komparativní metody, ale lingvistu v určitém kroku nahrazuje počítač. Většina fylogenezí jazyků se zakládají na lexikálních datech – seznamech slov, jejichž podobný zvuk a význam jsou důsledkem sdíleného původu (Pagel et al., 2007). Tato homologická slova se nazývají kognáty (např. číslovka *dva*, *zwei*, *two*, *due* v indoevropských jazycích). Změny kognátů je možné vystopovat po příbuzenských liniích ke společnému předkovi – prajazyku, ze kterého se vyvinula celá jazyková rodina a rekonstruovat části jeho slovníku (např. číslovka *dwóh* v praindoevropském jazyce). Počet sdílených kognátů ukazuje na to, jak moc si jsou jazyky příbuzné. Je však třeba brát v úvahu existenci tzv. falešných kognátů. Falešné kognáty jsou slova z různých jazyků, která podobně znějí a mají také podobnou strukturu, ale jejich podobnost je důsledkem nahromaděných náhodných změn (např. výraz *dog* pro psa v angličtině a v australském jazyce Mbarambaram). Jedná se o lingvistickou obdobu homoplázie či konvergence. Některé zdánlivé homologie mohou být důsledkem horizontálního přenosu (např. výraz *brother* pro bratra v maltštině není kognát českého *bratr*, ale výpůjčka z angličtiny).

Nejčastěji používaným zdrojem lexikálních dat je tzv. Swadeshův seznam slov (Haspelmath & Tadmor, 2009), který obsahuje slova tzv. základního slovníku, která najdeme téměř ve všech jazycích. Patří mezi ně označení osob, předmětů, jevů a úkonů, tedy hlavně osobní zájmena, základní číslovky a slovesa, označení členů rodiny, částí těla, přírodních jevů či některých zvířat. Existuje několik verzí Swadeshova seznamu, které se liší počtem obsažených slov. Základní verze má 100 slov, rozšířená verze jich obsahuje 207 a existují i kratší verze. Dalším, méně využívaným seznamem slov, je Leipzig-Jakartův seznam (Haspelmath & Tadmor, 2009). Ten obsahuje stejně jako Swadeshův seznam 100

univerzálních slov pro mnoho jazyků, z nichž pouze 62 slov je společných se Swadeshovým seznamem.

Stejně jako genetická data i lingvistická data lze analyzovat také pomocí distančních metod. Databáze *Automated Similarity Judgment Program* (ASJP) obsahuje 40-položkové seznamy slov pro většinu jazyků světa (Wichmann et al. 2010). Z těchto slov se počítají tzv. Levensteinovy vzdálenosti, což je minimální počet změn, nutných ke změně jednoho slova v jiné. Za změnu se počítá přidání, odebrání, nebo nahrazení jednoho symbolu jiným (tedy obdoby inzerce, delece, nebo substituce). Díky tomu, že tato metoda pracuje se změnami v sekvencích znaků (symbolů) tvořících slova, ale neuvažuje o změnách významu slov, obchází problém s identifikací homologií (kognátů) a obejde se bez lingvistické expertízy. Zároveň maximalizuje množství informace, které lze z lexikálních dat získat, a tak rekonstruovat fylogeneze, které zahrnují i vzdáleně příbuzné jazyky. Věrohodnost takových fylogenezí je obtížně věřitelná, ale data v ASJP umožňují vytvořit jazykové distance, které jsou srovnatelné s genetickými distancemi, nebo odhadnout stáří divergence jazyků (Wichmann et al. 2010).

Nevýhodou lexikálních dat je, že se vyvíjejí poměrně rychle, což omezuje hloubku na nich založených jazykových fylogenezí. Uvádí se, že po asi 7000 letech již není možné rozlišit skutečné kognáty o falešných – veškeré významové a zvukové podobnosti mohou být důsledkem homoplázie. Z tohoto důvodu se vedle lexikálních dat pro rekonstrukci jazykových fylogenezí využívá také tzv. strukturních dat – syntaktických, gramatických a fonologických, o kterých se předpokládalo, že se vyvíjejí pomaleji a měla být srovnatelná napříč jazykovými rodinami (Dunn et al., 2005). Pozdější studie však ukázaly, že i když existují velmi stabilní strukturní znaky, jako např. genderové rozlišení ve třetí osobě, obecně tyto znaky nejsou v čase stabilnější než lexikální znaky. Většina z nich se vyvíjí rychleji a je náchylnější k výpůjčkám, než jsou slova základního slovníku (Greenhill et al., 2017). Na rozdíl od lexikálních dat obsahují syntaktická data méně informace, např. pořadí větných členů (podmět, předmět a sloveso) nabývá napříč všemi jazyky jen šesti možných znakových stavů. Dalším problémem je, že tyto znaky jsou funkčně integrované, takže když dojde ke změně jednoho z nich, zpravidla to způsobí změny ostatních (Greenhill et al., 2017). Vedle lexikálních a syntaktických dat se využívají i data fonologická (např. přítomnost nosových či mlaskavých hlásek).

Existuje tedy celá řada přístupů, typů dat a metod vytváření genetických a jazykových stromů, které lze v rámci studia koevoluce genů a jazyků použít. Různé genetické a jazykové markery se liší způsobem dědičnosti a rychlostí evoluce, takže na nich založené fylogeneze zachycují odlišné aspekty populační nebo kulturní historie a mají různou časovou hloubku. Lze předpokládat, že výsledky studií koevoluce budou ovlivněny tím, jaká data se pro srovnání zvolí a jaké metody srovnání se použijí.

## 2. VÝSLEDKY

Studie pro rešerši byly vyhledávány v online databázi odborné literatury Web of Science a webovém vyhledávači Google Scholar s využitím následujících klíčových slov a jejich kombinací: *gene, language, coevolution, correlation, genetic data, linguistic data, geography*. Další literatura byla vyhledávána v seznamech citovaných prací takto vyhledaných studií, mezi studii, které tyto práce citují, nebo ve studiích, které byly uvedeny jako doporučené či související na webových stránkách příslušných odborných časopisů a vydavatelství. Do rešerše byly zahrnuty studie, které zkoumají koevoluci či korelaci genů a jazyků formálními, statistickými metodami a s využitím fylogenezi, založených na analýze genetických a jazykových dat, případně kombinující genetickou fylogenezi s jazykovou klasifikací nebo s geografickými daty. Do rešerše byly zahrnuty články publikované v recenzovaných odborných časopisech s impakt faktorem, publikované po roce 1985.

Takto bylo shromážděno 27 studií, splňujících nastavená kritéria. Nejstarší z nich byla publikována v roce 1988, nejnovější v roce 2022. Více než polovina z nich byla publikována po roce 2010.

Ve čtyřech ze sedmi studií, publikovaných před rokem 2000 (Cavalli-Sforza et al., 1988, 1992; Harding, & Sokal, 1988; Poloni et al., 1997) byly použity klasické polymorfismy. Ve zbylých třech studiích (Barbujani & Sokal, 1992; Sokal, 1988, 1989) jsou genetická data zastoupena blíže nespecifikovanými alelickými sekvencemi. Ve dvou studiích (Sokal, 1988; 1989) byly vedle genetických dat pro fylogenetickou analýzu použity také morfologické (kraniální) znaky. Ve studiích, publikovaných po roce 2000 se hojně využívají uniparentální markery (mtDNA a NRY) a autozomální DNA. MtDNA a NRY byly použity samostatně nebo v kombinaci ve 12 studiích. Z těchto studií je jedna založená pouze na mtDNA (Hunley et al., 2007) a tři pouze na NRY (Balanovsky et al., 2011; Lansing et al., 2007; Rosser et al., 2000). Celkem osm studií používá autozomálními, nebo genomická data (Baker et al., 2017; Barbieri et al., 2022; González-Santos et al., 2022; Hunley et al., 2008; Longobardi et al., 2015; Matsumae et al., 2021; Tishkoff et al. 2009; Wilder, 2011). Dalších sedm studií kombinuje autozomální a genomická data s uniparentálními markery (De Filippo et al., 2012; Karafet et al., 2015; Kushniarevich et al., 2015; Mona et al., 2009; Nováčková et al., 2019; Robbeets et al., 2021; Tambets et al., 2018). Studie Dudy a Zrzavého (2016) kombinuje všechny výše uvedené typy dat.

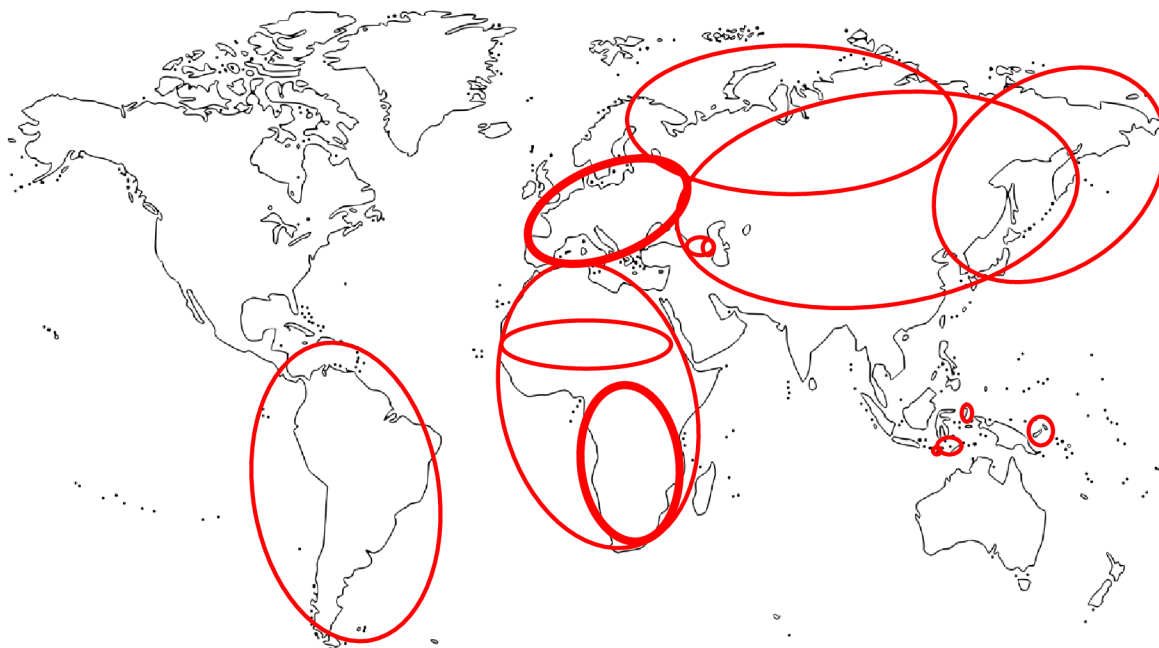


Kvalita použitých lingvistických dat se mezi studiemi značně liší. Starší studie, publikované před rokem 2000, používají pouze jazykovou klasifikaci, novější studie pracují s lexikálními daty, případně s daty syntaktickými. 12 studií, publikovaných po roce 2000, pracuje s lexikálními daty (Balanovsky et al., 2011; De Filippo et al., 2012; González; Santos et al., 2022; Karafet et al., 2015; Kushniarevich et al., 2015; Lansing et al., 2007; Longobardi et al., 2015; Matsumae et al., 2021; Nováčková et al., 2019; Robbeets et al., 2021; Rosser et al., 2000; Tambets et al., 2018), pouze tři studie pracují se syntaktickými daty (Hunley et al., 2008; Longobardi et al., 2015; Matsumae et al., 2021). Dvě z nich přitom kombinují lexikální, syntaktická a také fonologická data (Longobardi et al., 2015; Matsumae et al., 2021).

Jazykové klasifikace se buďto převádí do distančních matic (Harding & Sokal, 1988; Mona et al., 2009; Poloni et al., 1997; Sokal, 1988; Wilder et al., 2011), nebo se přímo srovnávají s genetickou fylogenezí (Cavalli-Sforza et al., 1988), nebo se formálně posuzuje, do jaké míry je lingvistická klasifikace konzistentní s fylogenezí, založenou na genetických datech (Baker et al., 2017; Cavalli-Sforza et al., 1992; Duda & Zrzavý, 2016; Hunley et al., 2007). Devět ze 14 studií, které pracují s jazykovou klasifikací, používá Ruhlenovu klasifikaci světových jazyků (Ruhlen, 1987), buď samostatně, nebo v kombinaci s jinou klasifikací. V jedné studii (Hunley et al., 2007) byly použity klasifikace Greenberga, Loukotky a Campbella. Tři studie (Baker et al. 2017; Duda & Zrzavý, 2016; Wilder, 2011) používají klasifikaci z databáze Ethnologue (Eberhard et al., 2023) a studie Barbieri et al. (2022) používá klasifikaci z databáze Glottolog (Hammarström et al., 2014).

Některé studie jsou zaměřeny na konkrétní oblasti, jiné zkoumají koevoluci genů a jazyků napříč celým světem (Obr. 3.). Ze sedmi studií, publikovaných před rokem 2000 se tři zabývají globální koevolucí genů a jazyků a čtyři se týkají Evropy. V novějších studiích, publikovaných po roce 2000 jsou zastoupeny i jiné oblasti světa. Mimo Evropu se tato problematika zkoumá na Kavkaze (Balanovsky et al., 2011; Karafet et al., 2015), v Asii (Matsumae et al., 2021; Robbeets et al., 2021; Tambets et al., 2018), v Indonésii a Melanésii (Hunley et al., 2008; Lansing et al., 2007; Mona et al., 2009; Wilder et al., 2011) a v Subsaharské Africe (De Filippo et al., 2012; González-Santos et al., 2022; Nováčková et al., 2019; Tishkoff et al., 2009). V těchto oblastech se nalézají jazyky, náležící do dvou nebo více dobře podpořených jazykových rodin. Některé z těchto oblastí mají výraznou topografii (hornatý Kavkaz, ostrovy v Indonésii a Melanésii). Nejméně zastoupenými oblastmi jsou Amerika, na kterou se zaměřuje pouze jedna studie, splňující nastavená kritéria (Hunley et al., 2007) a Austrálie. Pro tyto oblasti není k dispozici věrohodná jazyková klasifikace a je zde

dlouhodobý problém se získáváním genetického materiálu. To se týká zejména Severní Ameriky (především území USA) a Austrálie. Každou z těchto oblastí charakterizuje unikátní populační historie, která ovlivnila vztah mezi místní populačně-genetickou strukturou a distribucí jazyků. Následující část práce popisuje výsledky případových studií, zaměřených na příslušné oblasti světa: Evropa, Kavkaz, Asie, Indonésie a Melanésie, Afrika, a Svět.



**Obr. 3.** Oblasti zájmu 27 studií, zahrnutých do rešerše. Evropa je zastoupena sedmi studiemi, Subsaharská Afrika dvěma. Oblasti zájmu šesti studií, které se prezentují jako globální, nejsou na mapě vyznačeny.

Názvy jazykových rodin, skupin a izolátů použité v práci jsou převzaty z knihy *Lingvistická antropologie: jazyk, mysl a kultura* (Pokorný, 2009).

## 2.1. Evropa

Na většině území Evropy se, až na pár výjimek, mluví jazyky, spadajícími do indoevropské (IE) jazykové rodiny. Její kořeny sahají buď do doby zemědělské kulturní revoluce před 8000 lety, kdy do Evropy přišli zemědělci z Anatólie (dnešní Turecko), nebo do období před 5000 lety, kdy do Evropy migrovali pastevcí z oblasti pontsko-kaspických stepí (dnešní jihozápadní Rusko). V obou případech byla většina původních jazyků nahrazena indoevropskými, ale není jasné, jestli bylo nahrazeno také původní obyvatelstvo, nebo se indoevropské jazyky šířily horizontálně, prostřednictvím kulturní difuze. V Evropě se nachází také několik neindoevropských jazyků. Jedním z nich je baskičtina v severním Španělsku a jižní Francii. Jedná se o izolovaný jazyk, který je pozůstatkem předindoevropského osídlení.

V Evropě najdeme také zástupce uralské jazykové rodiny – finštinu, sámštinu, estonštinu a maďarštinu. Uralské jazyky ve Skandinávii jsou původem ze střední Asie a do Evropy přišli před asi 4000 lety. Původ maďarštiny sahá do 1. tisíciletí našeho letopočtu, kdy kočovní nájezdníci, původem z Uralu, dorazili do Střední Evropy a postupně asimilovali místní obyvatelstvo. Svě zástupce má v Evropě i turkická jazyková rodina. Jsou jimi gagauzština v Moldavsku a krymská tatarština na Krymu. Na Maltě se mluví afroasijským (semitským) jazykem (Comrie et al., 2003).

**Tabulka 1.** Studie koevoluce genů a jazyků v Evropě.

Autoři	Rok	Jazykové rodiny	Genetická data	Lingvistická data	Posouzení koevoluce	Koevoluce
Harding & Sokal	1988	5	Polymorfismy	Klasifikace (Ruhlen)	Mantelův test, Fenogram z genetických vzdáleností mezi jazykovými rodinami	Ano
Sokal	1988	5	97 alel + 10 kraniálních proměnných	Klasifikace (Ruhlen)	Mantelův test	Ano
Sokal et al.	1989	5	59 alel + 10 kraniálních proměnných	Klasifikace (Ruhlen)	ANOVA	Částečné
Barbujai & Sokal	1990	4	66 alelových sekvencí na 19 lokusech	Klasifikace (Ruhlen)	Posouzení lingvistické a genetické mapy	Ano
Rosser et al.	2000	6	NRY	Lexikon	Mantelův test, PCA, wombling	Částečné
Longobardi et al.	2015	3	Autozomy	Lexikon, syntax	Mantelův test, PCA	Ano
Kushniarevich et al.	2015	1	mtDNA, NRY, autozomy	Lexikon	Mantelův test, PCA, AMOVA, MDS	Ano

Evropou, případně částmi Evropy, se zabývalo sedm studií, zahrnutých do rešerše (Tabulka 1). Už ty nejstarší z nich si kladli otázku, jestli genetickou strukturu evropské populace lépe vysvětlí příbuzenské vztahy mezi jazyky, což by svědčilo o nahrazení původní populace mluvčími IE jazyků, nebo jejich prostorová distribuce a geografické bariéry, což by nahrávalo spíše modelu difuzní vlny, tedy horizontálnímu šíření IE jazyků. Robert Sokal,

jeden z hlavních průkopníků fylogenetiky a lidské populační genetiky, se tímto tématem zabýval v několika studiích (Barbujani & Sokal, 1990; Sokal, 1988; 1989). První dvě studie jsou založeny na rozsáhlém souboru klasických polymorfismů a také kraniálních znacích od tisíců jedinců z celého území Evropy. Tato genetická a morfologická data byla převedena do distanční matice, která byla porovnána s maticí založenou na jazykové klasifikaci. První studie (Sokal, 1988) k tomu využívá Mantelův korelační test, který měří závislost genetické příbuznosti na jazykové příbuznosti a geografické vzdálenosti, zatímco druhá (Sokal et al., 1989) pracuje s analýzou variance (ANOVA), která se snaží odhalit, jestli jsou za genetickou diferenciaci zodpovědné spíše jazykové bariéry, nebo geografické bariéry. Ve svém závěru se obě studie shodují – mezi jazykovými rodinami (IE, uralská, turkická, afroasijská, baskičtina) existují významné genetické rozdíly. Menší genetické rozdíly v rámci IE jazyků jsou pak vysvětlovány geografickou distribucí jejich mluvčích.

Studie Barbujaniho a Sokala (1990) hledá předěly v populačně-genetické struktuře evropské populace a snaží se určit, jestli korespondují s hranicemi mezi jazyky, nebo geografickými bariérami. Vychází ze souboru alelových frekvencí a jazykové klasifikace. Používá metodu zvanou wombling, která odhaluje prudké změny v alelových frekvencích v prostoru. Většina takto odhalených předělů odpovídá hranicím rozšíření jazykových rodin (IE, uralské, afroasijské a baskičtina), jejich jazyků nebo dialektů. Z celkem 33 prudkých alelových změn jich 22 leží na fyzické bariéře, zatímco 31 genetických hranic odpovídá hranicím mezi různými jazyky. Například předěl v populačně-genetické struktuře na území Finska přibližně odpovídá hranici mezi dialekty finštiny, jejichž předci přišli do Finska z různých směrů (přes Finský záliv nebo po souši od jihovýchodu). Jiná hranice od sebe odděluje severoitalské a jihoitalské dialekty. Výsledky ukazují, že jazyky a geografie se často podílejí na výsledné populačně-genetické struktuře, často spolu s různým původem evropských populací, které přišly do kontaktu až v nedávné době.

Studie Hardingové a Sokala (1988) koreluje distanční matice, na které jsou převáděna genetická data ve formě klasických polymorfismů a jazyková klasifikace zahrnující jazyky indoevropské, uralské, turkické a afroasijské jazykové rodiny a baskičtinu. Pro jejich srovnání byl použit Mantelův korelační test, ve kterém jsou obě složky navíc srovnávány s geografickou distanční maticí. Studie dospívá k závěru, že na genetickou diverzitu má velký vliv právě geografie, kdežto jazyky k ní přispívají jen málo. Výjimkami jsou mluvčí baskičtiny, finštiny a maltštiny, kteří se odlišují jak jazykově, tak i geneticky, a to i přestože sousedí s mluvčími IE jazyků.

Studie Rosserové a kol. (2000) zkoumá Y-chromozomální diverzitu v Evropě a blízkém okolí (západní Sibiř, Kavkaz a severní Afrika). Analýza hlavních komponent (*Principal Component Analysis*, PCA) ukazuje, že distribuci Y-chromozomální variability ovlivňuje více geografická distribuce než jazyková příslušnost populací. Jako první se od ostatních oddělí africké populace, oddělené od ostatních Středozezemním mořem. Mezi Afričany a zbylými populacemi pak stojí obyvatelé Řecka a Kypru. Zatímco Maďaři mají geneticky blízko k Indoevropanům, Finové se řadí k ostatní mluvčím uralských jazyků. V souladu s tím je výsledek Mantelova korelačního testu, který ukazuje silnou korelaci mezi geny a geografii, ale slabou korelaci mezi geny a jazyky. Metoda wombling, lehce odlišná od té, kterou použil Barbujani a Sokal (1990) nachází genetické bariéry, z nichž málokterá odpovídá hranicím mezi jazyky, a spíše odpovídají geografickým bariérám.

Novější studie Longobardiho a kol. (2015) tyto závěry částečně potvrzuje. Někteří mluvčí uralských jazyků a Baskové se geneticky liší od svých indoevropsky hovořících sousedů. Tato studie pracuje s rozmanitými daty – autozomálními SNPs a lexikálními (slova) a syntaktickými (gramatika) daty. Syntaktická data umožnila autorům rekonstruovat vztahy mezi jazyky různých jazykových rodin. Použitá data – genomická, lexikální a syntaktická – byla převedena do distančních matic. Z těch pak byly metodou UPGMA vytvořeny genetické a jazykové fylogeneze. Vedle nich studie pracuje i s metodou TreeMix, umožňující detekci migrací, a dále klastrovou analýzu STRUCTURE, odhalující genetické složení populací. Porovnání stromů ukázalo, že populace hovořící příbuznými jazyky IE jazykové rodiny (germánskými, románskými a slovanskými) jsou si příbuzné také geneticky. Tři nepřibuzné jazyky – baskičtina, a uralské jazyky maďarština a finština vycházejí nepřibuzné ostatním jazykům. Baskové jsou ostatním Evropanům nepřibuzní geneticky i jazykově, Finové jsou rovněž geneticky odlišní. Maďarština je příbuzná finštině, ale mluvčí tohoto jazyka geneticky spadají do kladu tvořeného populacemi, které hovoří slovanskými a románskými (rumunština) jazyky.

Studie Kushniarevichové a kol. (2015) zkoumá genetické odlišnosti v rámci slovanské větve IE jazyků. Geneticky se rozpadá do dvou skupin – na Slovany, kteří obývají Balkánský poloostrov a ty ostatní, kteří obývají střední a západní Evropu. Mezi nimi se nachází Rakousko, Maďarsko a Rumunsko, kde se hovoří jinými jazyky a tvoří tak mezi jižními a středovýchodními Slovany fyzickou bariéru. Slované jsou také ovlivňováni svými neslovanskými sousedy. Například Češi mají ve svých genech značné germánské příměsi a celkově mají blíž k západním Evropanům. Oproti tomu Poláci, kteří také sousedí

s Německem, jsou po genetické stránce bližší východním Slovanům, přičemž si v sobě nesou malou příměs germánských genů.

Výsledky těchto studií ukazují, že populačně-genetická struktura a distribuce jazyků v Evropě si poměrně dobře odpovídají. Mezi odlišnými jazykovými rodinami v Evropě se nacházejí významné genetické rozdíly, které jsou vysvětlovány právě jazykovou příslušností. Menší genetické rozdíly v rámci převládající IE jazykové rodiny pak odpovídají spíše jejich geografické izolaci.

## 2.2. Kavkaz

Kavkaz se pyšní největší rozmanitostí jazyků v rámci regionu. Na tomto poměrně malém, hornatém území v jihozápadní Asii, mezi Černým a Kaspickým mořem, se hovoří více než 40 jazyky, které se označují jako kavkazské a dalšími jazyky, které spadají do indoevropské, turkické, semitské a mongolské jazykové rodiny. Kavkazské jazyky se dělí na jihokavkazské (kartvelské), východokavkazské (nachsko-dagestánské) a západokavkazské (abcházsko-adygejské). Na Kavkaze se se dále vyskytují indoevropské jazyky, především arménština (bazální v rámci IE jazyků), dále např. perština nebo ruština. Mimo to se zde mluví turkickými jazyky (např. azerbajdžánštinou nebo turkmenštinou), semitskými jazyky (neoaramejštinou a arabštinou) a mongolským jazykem kalmyčtinou. Jazyková situace je tedy značně odlišná od situace v Evropě. Předpokládá se, že za ohromnou rozmanitostí jazyků stojí z velké části právě hornatá krajina, která zajišťuje geografickou izolaci sousedních populací (Urban, 2020).

**Tabulka 2.** Studie koevoluce genů a jazyků na Kavkaze.

Autoři	Rok	Jazykové rodiny	Genetická data	Lingvistická data	Posouzení koevoluce	Koevoluce
Balanovsky et al.	2011	3	NRY	Lexikon	Mantelův test, AMOVA, MDS	Částečné
Karafet et al.	2015	3	mtDNA, NRY, autozomy	Lexikon	Mantelův test, PCA, AMOVA, MDS	Částečné

Dvě publikované studie (Tabulka 2) si kladou otázku, jestli izolovanost kavkazských jazyků koresponduje s mírou genetické izolovanosti jejich mluvčích. Studie Balanovského a kol. (2011) pracuje s Y-chromozomálními STR a SNPs a vedle toho s lexikálními daty pro 14 populací. Tato studie odhalila čtyři hlavní Y-chromozomální haploskupiny, z nichž se

každá vyskytuje v jiné části Kavkazu, kde se zároveň vyskytují odlišné jazykové rodiny. Pro podporu tohoto zjištění byl použit Mantelův korelační test, který ukázal na vůbec nejsilnější korelaci mezi geny a jazyky ze všech publikovaných studií. Vedle toho AMOVA ukázala, že příslušnost k jazykové rodině vysvětlí bezmála dvojnásobek genetické variability oproti příslušnosti ke skupinám, které jsou definované geograficky. Jazyková příslušnost tedy dokáže vysvětlit velkou část pozorované Y-chromozomální diverzity na Kavkaze. To podporuje koevoluci mezi geny a jazyky, která zde vzniká patrně důsledkem geografické izolace a dlouhodobého paralelního vývoje místních populací. Studie hledá vztahy mezi kavkazskými, blízkovýchodními a evropskými populacemi pomocí mnohorozměrné škálovací analýzy (*Multidimensional scaling*, MDS). MDS seskupuje kavkazské populace blíže k blízkovýchodním populacím než k evropským. Původ kavkazské Y-chromozomální linie tedy sahá na Blízký východ. Po příchodu na Kavkaz došlo ke geografické izolaci a k diferenciaci haploskupin v místech usazení jejich nositelů. Proto je k mnoha populacím a jazykům možné přiřadit konkrétní haplotypy.

Testování hypotézy koevoluce genů a jazyků v jemném geografickém měřítku se věnuje studie Karafetové a kol. (2015), která se zaměřuje na oblast Dagestánu na východním Kavkazu. Pro lepší objasnění genetické struktury této oblasti analyzuje nejen Y-chromozomální, ale také autozomální SNPs a mtDNA haplotypy. Lingvistická data jsou založena na seznamech slov. Pro objasnění vztahů mezi populacemi byla použita MDS. Autozomální a Y-chromozomální markery odlišují s částečným překrýváním populace Dagestánu, Kavkazu a Blízkého východu a mtDNA navíc vykazuje mísení s evropskými populacemi. Pro odhalení rozdělování a mísení populací byl sestaven strom pomocí TreeMix. Pro posouzení vlivu jazyků a geografie na genetickou strukturu populací byl použit Mantelův korelační test, který odhalil vysokou korelaci mezi autozomálními SNPs a jazyky u populací patřících do nachsho-dagestánské jazykové rodiny. Výsledky testů odhalily vysoce strukturované populace se silnými korelacemi mezi jazyky a geny.

Studie zaměřené na Kavkaz tedy ukazují, že geografické bariéry mohou spolu s jazykovými bariérami efektivně zamezovat toku genů a vést k dlouhodobému paralelním vývoji genů a jazyků.

### **2.3. Asie**

Tři studie zahrnuté do rešerše (Tabulka 3) se soustřeďují na oblast západní a severní Asie. Na západě se nacházejí jazyky, patřící do tří velkých jazykových rodin – indoevropské, turkické

a uralské. Uralské jazyky jsou v této oblasti původní. Mají ostrůvkovité rozšíření, sahající od Baltského moře na západě po Západosibiřskou nížinu a poloostrov Tajmyr na východě a dělí se na dvě skupiny – ugrofinské a samojedské jazyky. Jejich stáří se odhaduje na 4000–6000 let a původ sahá buď do povodí řeky Volhy, nebo do oblasti Sibiře. Turkické jazyky mají původ na jihu, na Altaji nebo v severovýchodní Číně. Indoevropské jazyky – konkrétně ruština – se do těchto oblastí rozšířily teprve nedávno. V severovýchodní Asii se vedle ruštiny a turkických jazyků, jako je jakutština, vyskytují také menší jazykové rodiny a izoláty – tunguzské a mongolské jazyky, čukotsko-kamčatské jazyky, japonština a korejština, ketština a jukagirština. Většina z nich je, spolu s turkickými jazyky, řazena do kontroverzní makro-altajské či transeurasijské jazykové velkorodiny.

**Tabulka 3.** Studie koevoluce genů a jazyků v Asii.

Autoři	Rok	Jazykové rodiny	Genetická data	Lingvistická data	Posouzení koevoluce	Koevoluce
Tambets et al.	2018	1	mtDNA, NRY, autozomy	Lexikon	Mantelův test, PCA	Částečné
Matsumae et al.	2021	11	Genome-wide	Lexikon, syntax	RDA	Ano
Robbeets et al.	2021	5	mtDNA, autozomy	Lexikon	Triangulace	Částečné

Studie Tambetsové a kol. (2018) hledá genetické doklady expanze uralských jazyků s využitím mtDNA, NRY, autozomálních a lexikálních dat. Autozomální genetická diverzita byla vyhodnocena pomocí PCA, ve které se mluvčí uralských, turkických a slovanských jazyků, zahrnutých do studie, rozmístí podél východo-západní osy v souladu s jejich geografickou distribucí a bez ohledu na jazykovou příslušnost. To podporuje také UPGMA strom, ve kterém jsou populace hovořící uralskými jazyky příbuzné nejčastěji sousedním populacím. Analýza ADMIXTURE však odhaluje populačně-genetickou komponentu, která se vyskytuje převážně u mluvčích uralských jazyků. Tato komponenta se vyskytuje u Finů a Sámů, ale postrádají ji Estonci a Maďaři, což autoři interpretují jako důsledek jazykového posunu. Mantelův korelační test odhaluje silnou korelaci mezi geny a jazyky i geografii. Genetickou variaci tedy vysvětluje kombinace obou faktorů a šíření uralských jazyků bylo alespoň z části spojené s demickou expanzí, i když v některých částech svého současného areálu (Pobaltí, Střední Evropa) se tyto jazyky šířily především horizontálně.



Studie Matsumaeho a kol. (2021) se zabývá otázkou, do jaké míry různé typy jazykových dat odrážejí historii populací severovýchodní Asie spadající do několika jazykových rodin. Studie používá genomická data (SNPs) v kombinaci s lingvistickými (syntaktickými, lexikálními a fonologickými) daty a také s hudbou. Všechna data byla nejprve převedena do matic párových distancí a analyzována metodou *Neighbor-Net*, která vytváří fylogenetické sítě, které odhalí, jak příbuzenské vztahy, tak možný horizontální přenos. Tyto sítě naznačují, že jednotlivé kulturní znaky se vyvíjejí nezávisle na populační historii. Pro další zhodnocení vztahů mezi genetickými a kulturními daty byla použita redundanční analýza (*Redundancy analysis*, RDA). Ta odhalila, že gramatické podobnosti predikují genetickou příbuznost a naopak. U ostatních kulturních znaků odhalila RDA pouze slabé korelace s populační historií. RDA analýza zohledňující vliv geografické a jazykové distance naznačuje, že místní populačně-genetická struktura se zformovala ještě před vznikem místních jazykových rodin.

Rozsáhlá studie Robbeetsové a kol. (2021) hledá doklady společného původu a šíření tzv. transeurasijských jazyků. Tato studie používá autozomální SNPs, včetně staré DNA z archeologických nálezů a lexikální data, zastupující 98 jazyků. Triangulací jazykových, genetických a archeologických dat pak dospívá k závěru, že transeurasijské jazyky mají společný původ a jejich pravlast leží v severozápadní Asii, v povodí řeky Amur, odkud se začaly šířit spolu se zemědělstvím v období raného holocénu. Po úvodním rozdělení a rozšíření do oblastí kolem řeky Liao-che došlo k oddělení a šíření proto-mongolských jazyků na sever do oblasti Mongolské plošiny, proto-turkických jazyků na západ a na severovýchod, proto-tunguzské jazyky do oblastí řeky Amur, proto-korejštiny na Korejský poloostrov a proto-japonštiny na Japonské ostrovy.

#### **2.4. Indonésie a Melanésie**

Současné osídlení ostrovní jihovýchodní Asie a Oceánie je výsledkem dvou velkých, nezávislých migrací. První z nich se odehrála v Pleistocénu před více než 50 tisíci lety a vedla k osídlení Indonésie, Melanésie, Nové Guinei a Austrálie původně lovecko-sběračskou populací. Pozůstatkem této migrace jsou dnešní mluvčí papuánských a australských jazyků a izolované populace v jihovýchodní Asii (např. Andamanci nebo Semangové). Před asi 5000 lety do těchto oblastí doputovali zemědělci a mořeplavci mluvící austronéskými jazyky, kteří pocházejí z jižní Číny a Tchaj-wanu. Austronéská expanze postupovala přes Filipíny a Moluky a dále západním i východním směrem. Nakonec vedla k osídlení celé ostrovní jihovýchodní Asie (a také Madagaskaru), Melanésie, Mikronésie a Polynésie (po Havaj,

Velikonoční Ostrov a Nový Zéland). Austronéských jazyků je více než 1000 a většinou z nich mluví jen málo lidí, kteří obývají jeden ostrov, nebo jeho část. Přes vysokou diverzitu a rozsáhlý areál si tyto jazyky uchovaly velkou část společné slovní zásoby a některé gramatické prvky. V oblastech, kde se areály papuánských a austronéských jazyků překrývají, se tyto jazyky navzájem ovlivňují (Nielsen et al., 2017).

**Tabulka 4.** Studie koevoluce genů a jazyků v Indonésii a Melanésii.

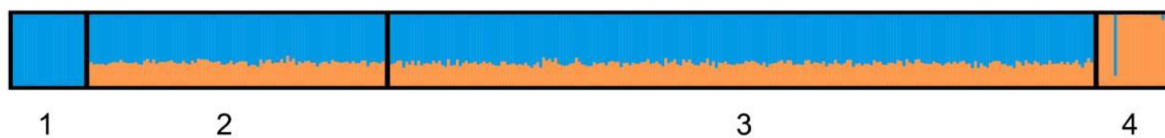
Autoři	Rok	Jazykové rodiny	Genetická data	Lingvistická data	Posouzení koevoluce	Koevoluce
Lansing et al.	2007	1	NRY	Lexikon	Mantelův test	Částečné
Hunley et al.	2008	2	Autozomy	Syntax	Test of treeness, korelační testy	Ano
Mona et al.	2009	2	mtDNA, NRY	Klasifikace (Ruhlen)	AMOVA, MDS	Ne
Wilder et al.	2011	2	Autozomy	Klasifikace (Ethnologue)	PCA, AMOVA	Ne

Dnešní populace Indonésie a Oceánie se tedy skládá ze dvou komponent – staré, papuánské a mladé, austronéské. Čtyři publikované studie (Tabulka 4) se zabývaly otázkou, jestli populačně-genetická struktura a jazyková příslušnost místních populací dodnes odráží tyto dávné migrace, nebo jestli tisíce let vzájemných interakcí těchto populací vedly ke ztrátě korespondence mezi geny a jazyky. Přitom se zaměřovali na různé části tohoto regionu. Hunley a kol. (2008) se zaměřili na Bismarckovo souostroví a ostrov Bougainville v Melanésii, kde vedle sebe žijí mluvčí papuánských a austronéských jazyků. Studie využívá autozomální STRs a gramatické a fonologické znaky od 33 místních populací. Využití strukturních dat umožnilo autorům sestavit jazykovou fylogenezi zahrnující jazyky různých jazykových rodin, podobně jako to udělala studie Longobardiho a kol. (2015). Práce testuje dva modely koevoluce mezi geny a jazyky. První model předpokládá, že korespondence je výsledkem rozdělení a následné izolace populací v počátcích expanze, tedy postupného větvení. Druhý model předpokládá průběžný tok genů a prvků jazyka mezi sousedními populacemi, tedy izolaci vzdáleností. První předpokládá shody mezi topologiemi genetických a jazykových stromů a korelaci mezi distančními maticemi, které jsou z nich vytvořené. Druhý předpokládá, že tato korelace přestane být signifikantní po zohlednění geografické distance. Porovnávání stromů a distančních matic ukázalo, že situace v západní Melanésii je složitá. K toku genů a prvků jazyka mezi blízkými populacemi dochází, a to bez ohledu na jejich

jazykovou příslušnost, což vedlo k částečnému přemazání dávené populačně-genetické struktury. Některé „papuánské“ populace jsou geneticky bližší austronéským populacím. Výměna prvků jazyka mezi papuánskými a austronéskými jazyky je však omezená. Korelace mezi geny, jazyky a geografii je v této oblasti silná, a to zejména v rozsáhlém, hornatém vnitrozemí ostrova Nová Británie.

Studie Mony a kol. (2009) se zaměřila na oblast východní Indonésie, zahrnující ostrovy Flores, Adonara, Solor, Lembata, Pantar, Alor a Východní Timor, s využitím mtDNA a NRY a jazykovou klasifikací. Genetická analýza odhalila východoasijskou (tedy austronéskou) a melanéskou („papuánskou“) komponentu v genomech místních obyvatel, ale také třetí, „indonéskou“ komponentu, která může být pozůstatkem původních lovců-sběračů, kteří do této oblasti přišli v pleistocénu. MDS grafy, založené na mtDNA a NRY shlukují mluvčí austronéských i papuánských jazyků a AMOVA neprokázala korelaci mezi genetickými a lingvistickými daty. Absence korelace mezi geny a jazyky je patrně důsledkem expanze austronéských mluvčích od západu, po které následovala expanze papuánských mluvčích od východu, přičemž obě byly spojené s intenzivním tokem genů a jazykovým posunem.

K podobnému výsledku dospěla studie Wildera a kol. (2011), která se zaměřila na souostroví Severní Moluky ve východní Indonésii. Tedy na další oblast, kde vedle sebe žijí mluvčí papuánských a austronéských jazyků. Studie analyzuje autozomální a Y-chromozomální SNPs a jazykové klasifikace. Místní populace je geneticky mimořádně homogenní. Klastrová analýza STRUCTURE odhalila intenzivní mísení mezi mluvčími austronéských a papuánských jazyků. Populace z Číny a z vnitrozemí Nové Guineje byly použity jako reference pro odhad množství asijské a papuánské příměsi v genomech místní obyvatel (Obr. 4.). Tato analýza ukázala se, že všechny populace lze charakterizovat jako hybridní, asijsko-melanéské, bez ohledu na jejich jazykovou příslušnost.



**Obr. 4.** Výsledek klastrové analýzy STRUCTURE pro dvě populačně-genetické komponenty a mísení. Obyvatelé jižní Číny (1) a vnitrozemí Papuy-Nové Guineje (4), slouží jako reference pro austronéské a papuánské předky. Mluvčí austronéských jazyků (2) a papuánských jazyků (3) na Severních Molukách jsou hybridního původu a geneticky se od sebe nijak neliší (Wilder a kol. 2011).

Studie Lansinga a kol. (2007) oproti předchozím studiím posuzuje koevoluci mezi geny a jazyky na mnohem jemnější škále. Zaměřuje se na ostrov Sumba v Indonésii, kde se mluví několika blízce příbuznými austronéskými jazyky. Studie pracuje s autozomálními SNPs a STRs a lexikálními daty (Swadeshův seznam o 200 slovech) pro 29 lokalit na ostrově. Fylogeneze, založená na těchto lexikálních dokazuje, že místní jazyky opravdu pochází z jednoho prajazyka. Podíl kognátů, sdílených s tímto prajazykem, se však mezi 29 lokalitami na ostrově liší. To stejné platí pro zastoupení Y-chromozomální haploskupiny O, pocházející z jihovýchodní Asie a charakterizující mluvčí austronéských jazyků. Podíl jazykových retencí koreluje s podílem této haploskupiny, přičemž nejvyšší je na východě ostrova, odkud podle místních legend Austronésani přišli. Studie tedy odhaluje korespondující východo-západní gradienty genetické a jazykové diverzity, které nezávisle dokládají postupnou asimilaci lovecko-sběračského obyvatelstva příchozími austronéskými zemědělci.

Celkově jsou doklady koevoluce genů a jazyků v Oceánii smíšené. Dávná populačně genetická struktura byla do značné míry přemazána pozdějším tokem genů a jazyková příslušnost odpovídá současné jazykové struktuře jen částečně. V některých oblastech však korespondenci mezi geny a jazyky najít lze.

## 2.5. Afrika

Afrika je místem, odkud moderní lidé pochází a odkud osídlili zbytek světa a je tedy pravděpodobné, že je i domovem prvních jazyků (Berwick & Chomsky, 2016). Na území Afriky vyskytují čtyři velké jazykové rodiny – afroasijská, nilosaharská, nigerokonžská a khoisanská. Afroasijská rodina je rozšířená v severní a východní Africe (berberské, čadské, omotské a kušitské jazyky) a zasahuje také do oblastí středního východu a středomoří (semitské jazyky), zatímco ostatní podskupiny se vyskytují pouze na africkém kontinentě. Nilosaharská rodina, která je rozšířená ve východní Africe a zasahuje do střední a severní Afriky, je – přinejmenším ve své inkluzivní verzi – velmi kontroverzní. Nigerokonžská jazyková rodina je s asi 1500 jazyky největší jazykovou rodinou na světě. Její nejpočetnější podskupina – bantuské jazyky – se z oblastí západní Afriky (pomezí dnešní Nigérie a Kamerunu) rozšířila do střední, východní a jižní Afriky spolu se zemědělstvím. Nigerokonžskými jazyky mluví také Pygmejové, lovci-sběrači ve střední Africe, o kterých se předpokládá, že tyto jazyky přejali dodatečně. Khoisanskými jazyky, které se vyskytují v jižní Africe, mluví lovci-sběrači Sanové a pastevcí Khoikhoiové představující nejbazálnější linie současného lidstva. Izolované jazyky, kterými hovoří lovci-sběrači Hadza a zemědělci

Sandawe ve východní Africe, bývaly řazeny ke khoisanským jazykům. Jazyková příbuznost Hadzapů a Sandawů k jihoafrickým khoisanům je však obtížně doložitelná a geneticky si také nejsou příbuzní.

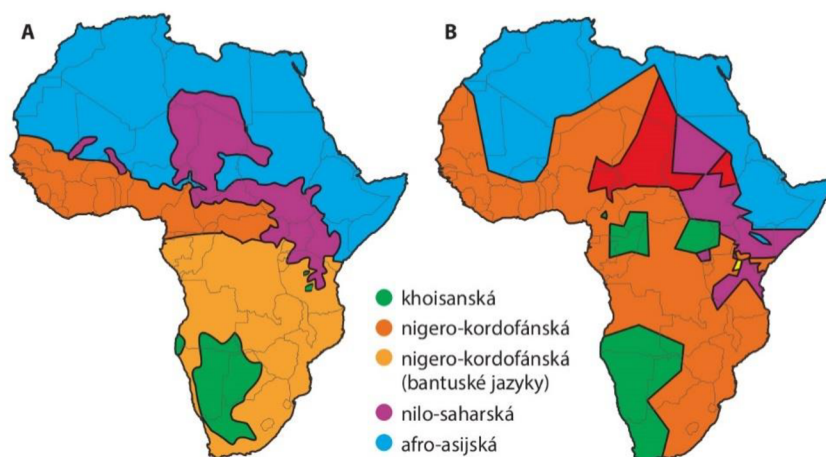
**Tabulka 5.** Studie koevoluce genů a jazyků v Africe.

Autoři	Rok	Jazykové rodiny	Genetická data	Lingvistická data	Posouzení koevoluce	Koevoluce
Tishkoff et al.	2009	4	Autozomy	Klasifikace	PCA, NJ tree, STRUCTURE	Částečné
De Filippo et al.	2012	1	mtDNA, NRY, autozomy	Lexikon	Mantelův test	Ano
Nováčková et al.	2019	3	mtDNA, NRY	Lexikon	Mantelův test, AMOVA, MDS	Částečné
González-Santos et al.	2022	1	Genome-wide	Lexikon	Mantelův test, PCA	Částečné

Vzhledem k tomu, že Africké jazyky lze rozdělit do čtyř jazykových rodin a populačně genetická struktura Afriky byla v posledních 5000 letech ovlivněna rozsáhlou demografickou expanzí, čtyři publikované studie (Tabulka 5) zkoumají, jestli distribuce genetické variability odpovídá distribuci jazykových rodin a také jaký byl vliv bantuské expanze na populačně-genetickou strukturu kontinentu.

Rozsáhlá studie Tishkoffové a kol. (2009) zkoumá populačně-genetickou strukturu obyvatel Afriky (a Afroameričanů) a jazykovou klasifikaci využívá pro interpretaci výsledků. Využívá autozomální STRs, na jejichž základě vytváří NJ stromy, PCA a klastrovou analýzu STRUCTURE. Klady tohoto stromu dobře odpovídají oblastem Afriky. NJ strom shlukuje jihoafrické khoisanské populace (Sanové a Namové), které jsou zároveň oddělené od zbytku Afričanů. Další bazální skupinou jsou Pygmejové. To ukazuje na sdílený původ těchto geograficky vzdálených populací a naznačuje, že si tyto populace kdysi mohly být příbuzné také jazykově (původní jazyk Pygmejů se nedochoval). Analýza STRUCTURE odhalila šest populačně-genetických komponent, které překvapivě dobře odpovídají jazykové klasifikaci. Tato korespondence je zjevná zejména v situaci, kdy se tyto klastry vynesou do mapy. Na druhou stranu většina afrických populací je geneticky heterogenní, přičemž převažuje nigerokonžská komponenta. Nigerokonžské populace v západní a střední Africe však zůstaly geneticky relativně homogenní, což zřejmě odráží migraci bantuských mluvčích napříč

subsaharskou Afrikou, při které se postupně mísily s původními populacemi. Intenzita tohoto mísení směrem k východu a k jihu narůstala, což je patrné i na topologii NJ stromu – bantuské populace na jihu (Zulu, Xhosa) jsou bazální, přestože představují tu nejodvozenější větev bantuských jazyků. Jde evidentně o důsledek jejich nedávného mísení s Khoisany.



**Obr. 5.** (A) Výsledek klastrové analýzy STRUCTURE pro šest populačně-genetických komponent a žádné mísení, vyneseny do mapy Afriky (Tishkoff et al. 2009). (B) Geografické distribuce čtyř afrických jazykových velkorodin podle Greenberga.

Studie De Filipa a kol. (2012) studuje načasování a trasu bantuské expanze. Testuje dva modely – rané a pozdní rozdělení. Model raného rozdělení předpokládá, že předkové východních bantuských jazyků putovali před 4000 lety na sever od Konžské pánve do oblasti Velkých jezer a odtud na jih, zatímco v modelu pozdního rozdělení se východní bantuské jazyky oddělily před 2000 lety jižně od Konžské pánve a do východní Afriky přišly od jihozápadu. Také si kladla otázku, jestli za rozšířením bantuských jazyků stojí spíše démická expanze nebo kulturní difuze (jazykový posun). Studie využívala mtDNA, NRY, autozomální STR a lexikální data, která byla převedena do distancí. Mantelův test ukázal silnou korelaci mezi všemi třemi genetickými markery a jazyky. Výsledek je pak v souladu s modelem pozdního rozdělení bantuských jazyků na jižní a východní jazyky na jih od Konžského pralesa a šířením prostřednictvím expanze. To podporuje i distribuce genetických vzdáleností mezi bantuskými populacemi, která je nižší než mezi bantuskými populacemi a populacemi, které mluví nilosaharskými a afroasijskými (kušitskými) jazyky.

Tématem vztahů mezi genetickými, lingvistickými a geografickými vzdálenostmi mezi bantuskými populacemi se zabývá také studie González-Santose a kol. (2022). Využívá celogenomová SNPs bantuských i nebantuských populací spolu s lexikálními daty. První

analýzy ukázaly na genetické příměsi v bantuských populacích, vytvořené důsledkem geografické blízkosti jejich sousedů mluvících odlišnými jazyky. Stupeň genetické diference bantuských populací pravděpodobně ovlivnil postupný tok genů mezi sousedními populacemi. Výsledky korelačních testů s těmito daty neukázaly korelaci mezi geny a geografii, ale ukázaly korelace mezi jazyky s geografii i s geny. Pomocí klastrové analýzy ADMIXTURE hledá ancestrální populace a v dalších analýzách pak využívá upravená genetická data bez genetických příměsí. Výsledky ukazují na silné korelace mezi geny, geografii i jazyky. Studie dochází k závěru, že geny jsou více ovlivněny izolací a driftem, kdežto jazyky jsou více ovlivněny stářím divergence – starší skupiny vykazují vyšší jazykovou rozmanitost oproti těm, jež se diverzifikovaly později.

Práce Nováčkové a kol. (2019) se zabývala původem a vzájemnými vztahy mezi sahelskými kočovnými pastevci a usedlými zemědělci spadajícími do 3 jazykových rodin – nilosaharská, nigerokonžská a afroasijská. V této studii se vychází z mtDNA a Y-chromozomálních markerů a lexikálních dat z databáze ASJP. Lingvistická i genetická data jsou zpracována do podoby genetických a jazykových vzdáleností. Pro posouzení genetické diference mezi populacemi byla použita AMOVA a MDS a pro objasnění vztahů mezi geny, jazyky a geografii Mantelův test. Geny kočovných pastevců vykazují korelace s jazyky, ale ne s geografii. Oproti tomu genetické distance mezi usedlými zemědělci korelují s geografickou, ale ne s jazykovou distancí. Mezi zemědělci je navíc rozdíl v korelaci mezi odlišnými genetickými markery a jazyky. Jazyky s mtDNA vykazovaly téměř nulovou korelaci, zatímco Y-chromozomální STR vykazovaly sice slabší, ale statisticky významné korelace, což je důsledek patrilokality těchto populací. Míra genetické diference zemědělců je daná spíše geografickou vzdáleností, kdežto genetická diference pastevců je ovlivněna hlavně jejich jazykovou příslušností. Výsledky MDS také ukazují na hybridní původ pasteveckých populací – ty ze západního a středního Sahelu jsou odlišné od těch východosahelských. Mezi populacemi s odlišnými způsoby obživy dochází jen k minimálnímu toku genů, což potvrzuje, že tyto společnosti byly a jsou endogamní.

Populačně-genetická struktura afrického kontinentu překvapivě dobře odpovídá distribuci hlavních jazykových uskupení, i když lingvistické doklady existence některých z nich jsou sporné. Bantuská expanze, která silně ovlivnila dnešní jazykovou mapu Afriky, v některých oblastech ovlivnila také genetické složení místního obyvatelstva, zatímco jinde se tyto jazyky šířily především horizontálně.

## 2.6. Svět

Studie zabývající se koevolucí genů a jazyků na globální úrovni si nejčastěji kladou otázku, jestli populačně-genetická struktura odpovídá distribuci jazykových rodin nebo velkorodin.

**Tabulka 6.** Studie koevoluce genů a jazyků na globální škále.

Autoři	Rok	Jazykové rodiny	Genetická data	Lingvistická data	Posouzení koevoluce	Koevoluce
Cavalli-Sforza et al.	1988	21	Polymorfismy	Klasifikace (Ruhlen)	Srovnání genetického a lingvistického stromu	Částečné
Cavalli-Sforza et al.	1992	21	Polymorfismy	Klasifikace (Ruhlen)	Konzistenční index	Ano
Poloni et al.	1997	10	Polymorfismy	Klasifikace (Ruhlen)	Mantelův test, AMOVA	Částečné
Duda & Zrzavý	2016	45	Polymorfismy, mtDNA, NRY, autozomy, genome-wide	Klasifikace (Ethnologue, Greenberg-Ruhlen)	Optimalizace datových sad na základě lingvistické klasifikace na dataset superstromu.	Částečné
Baker et al.	2017	32	Genome-wide	Klasifikace (Greenberg-Ruhlen)	Pearsonův korelační koeficient	Částečné
Barbieri et al.	2022	14	Autozomy	Klasifikace (Glottolog)	1. identifikace genetických a jazykových enkláv; 2. genetická soudržnost jazykových rodin; 3. porovnání topologie a chronologie genetických a jazykových stromů	Částečné

Pionýrská studie Cavalliho-Sforzy a kol. (1988) rekonstruuje populační historii lidstva s využitím dostupných genetických dat – souboru klasických polymorfismů, převedených do matice párových genetických distancí. Z nich vytvořený fylogenetický strom pak srovnává



s Greenberg-Ruhlenovou klasifikací jazyků. Každá z jazykových rodin nebo velkorodin náleží jednomu z kladů v genetické fylogenezi (Obr. 2). Výjimkami jsou např. Pygmejové Mbuti v Konžské pánvi, Baskové ve Španělsku, nebo Laponci ve Finsku.

V navazující studii Cavalli-Sforza a kol. (1992) využívají stejný fylogenetický strom a stejnou jazykovou klasifikaci, ale srovnání topologií stromů doplňuje o formální metodu měření korespondence mezi klady v genetickém stromě a jazykovými skupinami. Měří tzv. konzistenční index, který určuje, do jaké míry je distribuce studovaného znaku v souladu s topologií fylogenetického stromu. Výsledné hodnoty jsou poměrně vysoké a podle autorů ukazují, že ke koevoluci genů a jazyků na globální škále dochází.

Poloniová a kol. (1997) analyzuje genetickou variabilitou restričního polymorfismu chromozomu Y a zkoumá, jak tuto variabilitu ovlivňují jazykové a geografické bariéry. Analýza je prezentována jako globální ale populace východní Asie, Austrálie a Ameriky jsou značně podprezentované. Lingvistická data zastupuje jazyková klasifikace. Pro posouzení populačně genetické struktury byly provedeny analýzy AMOVA a MDS. Vliv geografie a jazykové příslušnosti na genetickou strukturu byl měřen pomocí Mantelova korelačního testu. Výsledky provedených analýz odhalují tři hlavní genetická seskupení, které tvoří populace hovořící nigerokonžskými, afroasijskými a indoevropskými jazyky. Další samostatnou skupinu tvoří mluvčí khoisanských jazyků, kteří jsou však zastoupeni pouze dvěma populacemi. Mantelův test odhaluje silné korelace mezi geny, jazyky i geografii, které ukazují, že oba faktory mají vliv na genetickou strukturu, ale vliv geografie je silnější.

Práce Dudy a Zrzavého (2016) rekonstruuje globální kompozitní fylogenezi (*supertree*) lidstva s využitím genetických a jazykových dat a zkoumá konflikt mezi těmito daty a také mezi výslednou fylogenezí a jazykovou klasifikací. Metoda *supertree* umožňuje sestavit konsensus velkého množství publikovaných fylogenetických stromů různých populací, založených na různých datech – genetických i lingvistických. Metoda *matrix representation with parsimony* (MRP) převádí topologie publikovaných stromů do matic binárních znaků, které jsou následně sloučeny dohromady a analyzovány metodou maximální úspornosti a takto je vytvořen kompozitní strom (*supertree*). Práce měří konzistenční index jazykových uskupení ve dvou jazykových klasifikacích (Eberhard et al., 2023; Ruhlen, 1987). Tyto analýzy ukázaly relativně malou shodu mezi topologií stromu a jazykovou klasifikací. Vysoký konzistenční index mají např. jazyky austronéské, indoevropské afroasijské a khoisanské. V nejlepší shodě s topologií stromu jsou však kontroverzní amerindské jazyky.

Populačně-genetickou strukturací současného lidstva, migracemi a genetickým mísením a také tím, jak je tato strukturace v souladu s jazykovou klasifikací, se zabývá studie Bakerové a kol. (2017). Rozsáhlý soubor genomických dat (SNPs) podrobuje klastrové analýze ADMIXTURE, která odhalila 21 populačně-genetických komponent. Většina jedinců a populací zahrnutých do analýzy má smíšený původ. K detekci migračních událostí a s ní spojených genetických mísení byl použit program TreeMix, který odhalil tři nejdůležitější migrační události. Korelační testy ukázaly, že zatímco rozšíření některých jazyků (např. germánských nebo abcházsko-adygejských) neodpovídá genetické struktuře a historii populací, rozšíření jiných (např. berberských nebo khoianských) dobře odpovídá odhaleným populačně-genetickým komponentám a jejich geografickému rozšíření.

Shodami a neshodami mezi historií lidských populací a historií jazyků na globální úrovni se zabývá také práce Barbierové a kol. (2022). Tato studie se snaží popsat také procesy, které za odhalenými shodami a neshodami stojí. Studie pracuje s databází genomických dat (autozomální SNPs), která se nazývá *Genes and Languages Together* (GeLaTo). Pro identifikaci neshod byla použita metoda, která detekuje tzv. enklávy – populace, jejichž sousedé jsou buď geneticky odlišní (genetická enkláva), nebo mluví nepřibuzným jazykem (jazyková enkláva), přičemž mají genetické nebo jazykové příbuzné jinde. Jako jediný případ jazykové enklávy byli identifikováni Maďaři, ostatní enklávy byly genetické. Studie také zkoumala, jestli jsou si populace, které hovoří příbuznými jazyky, geneticky bližší než populace hovořící nepřibuznými jazyky. Za tímto účelem měřila, jak se mění genetická distance s přibývajícím geografickým vzdáleností pro jazykově příbuzné a jazykově nepřibuzné populace. Výsledky ukazují, že řada jazykových rodin je geneticky homogennější než populace hovořící jazyky různých rodin, které jsou od sebe stejně vzdálené v prostoru. To se týká většiny nigerokonžských, indoevropských, sinotibetských a mongolských jazyků. Naopak turkické a austronéské jazykové rodiny geneticky homogenní nejsou. Fylogeneze jazykových rodin byly datovány s využitím archeologických, historických a lingvistických dat. Genetické fylogeneze mluvčích těchto jazyků byly datovány pomocí molekulárních hodin. Srovnání datovaných fylogenezí pak ukázalo minimální shody v chronologii jazykových a genetických divergencí. Ve většině případů je genetická divergence mluvčích mnohem starší, než je stáří jednotlivých jazyků nebo celých rodin. Z výsledků vyplývá, že většina populací z databáze GeLaTo, které spadají do jedné jazykové rodiny, si je příbuzná také geneticky, přinejmenším v porovnání s mluvčími nepřibuzných jazyků v jejich okolí. Nejlepší shoda mezi geny a jazyky byla nalezena u mluvčích indoevropských jazyků. Neshody

pak byly nalezeny mezi populacemi geneticky blízkými, ale jazykově nepříbuznými, které vznikají jako důsledek jazykových posunů. Tyto neshody byly zaznamenány po celém světě, což ukazuje, že se nejedná o výjimky, ale o jevy, ke kterým v lidské populační historii dochází běžně.

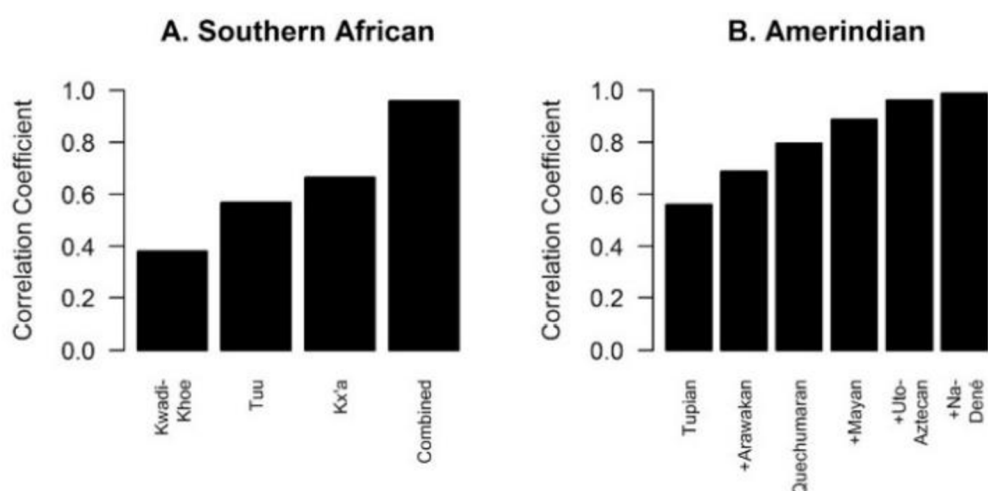
### 3. DISKUSE A ZÁVĚR

Problémy studií, které koevoluci nebo korelaci mezi geny a jazyky zkoumají, se týkají jednak použitých dat – zejména kvality lingvistických dat – a použitých metod – zejména ztráty informace obsahu při nakládání s lingvistickými daty.

Publikované studie často zkoumají, jestli geny – klady na fylogenetickém stromu, populačně-genetické klastry, nebo haplotypové frekvence, odpovídají jazykovým skupinám – rodinám a velkorodinám. Většina (14) studií pracuje pouze s jazykovou klasifikací, přičemž obvykle používá kontroverzní klasifikaci Merritta Ruhlena (Ruhlen, 1987), která je částečně založená na metodě masové komparace Josepha Greenberga (Greenberg, 1963). Zatímco Ruhlenova klasifikace obsahuje pouze 35 jazykových uskupení – 17 velkorodin a 18 izolátů – ostatní lingvisté řadí jazyky světa do 420 takových uskupení – rodin a izolátů (Campbell, 2015). Typickým příkladem kontroverzní jazykové velkorodiny je amerindská rodina, do které se řadí většina jazyků Severní, Střední a Jižní Ameriky, jejichž příbuznost ale není formálně doložitelná. Většina starších populačně-genetických studií, které se zaměřovaly na území Ameriky, používá právě tuto klasifikaci (Bolnick et al., 2004). Ruhlenovu klasifikaci používá devět studií (Baker et al., 2017; Cavalli-Sforza et al., 1988, 1992; Duda & Zrzavý, 2016; Harding & Sokal, 1988; Mona et al., 2009; Poloni et al., 1997; Sokal, 1988, 1989).

Uváděné korelace mezi geny a jazyky mohou být velmi vysoké (Baker et al., 2017), ale jejich informační hodnota je sporná, když jsou založené na neuznávané jazykové klasifikaci. Co víc, korelace mezi genetickými a jazykovými uskupeními je tím vyšší, čím je jazykové uskupení větší, inkluzivnější a tedy kontroverznější. Studie Bakerové a kol. (2017) na základě rozsáhlého globálního souboru genomických dat odhaluje 21 populačně-genetických klastrů a koreluje je s jazykovými skupinami. Těsnost korelace však není dána věrohodností jazykové rodiny, právě naopak. Například korelace nekontroverzních germánských jazyků a severoevropského genetického klastru je pouze 0,21, zatímco korelace kontroverzních khoisanských jazyků a jihoafrického klastru je 0,96. Oproti tomu individuální skupiny, které tvoří khoisanské jazyky vykazují nižší korelaci s genetickou strukturou (Obr. 6 A). Jednotlivé jazykové rodiny v Americe (např. tupíjské a aravacké jazyky) s genetickou strukturou korelují slabě, ale postupným přidáváním dalších rodin korelace stoupá, až dospějeme k amerindské velkorodině, která genetickou strukturu predikuje téměř dokonale ( $r = 0,96$ , Obr. 6 B), ale lingvisty není uznávána. K podobnému výsledku dopěla i studie Dudy a Zrzavého (2016), která měřila konzistenční indexy jazykových rodin na topologii stromu založeného na genetických a jazykových datech. Nejvyšší CI má právě amerindská velkorodina, což je ale

způsobeno tím, že pokrývá v podstatě celý kontinent, díky čemuž podchytí očekávanou korelaci mezi geny a geografii. Na tento problém, že čím jsou jazykové klasifikace realističtější, tím jsou doklady koevoluce slabší, upozorňuje i jediná studie týkající se Ameriky zahrnutá do rešerše – studie Hunleyho a kol. (2007). Tato studie používá tři jazykové klasifikace – Loukotkovu, Greenbergovu a Campbellovu, které se shodují na mělkých úrovních, ale na hlubších úrovních se mezi sebou liší. Práce nachází jen slabé shody mezi klasifikací a NJ stromem založeném na mtDNA. Nejlépe mu odpovídá nejstarší, Loukotkova klasifikace, pravděpodobně kvůli její shodě s geografii. Následuje Greenbergova klasifikace a minimální shoda je mezi fylogenezí a uznávanou Campbellovou klasifikací.



**Obr. 6.** Korelace mezi populačně-genetickými komponentami a jazykovými rodinami. (A) Jihoafrická populačně-genetická komponenta, Khoisanské jazyky a jejich podskupiny. (B) Americká komponenta, amerindské jazyky a jejich podskupiny (Baker et al., 2017)

Řešením tohoto problému by byla užší spolupráce mezi genetiky, antropology a lingvisty, která by vedla k upuštění od těchto kontroverzních jazykových klasifikací a využití jiných typů lingvistických dat, např. lexikálních a strukturních (Campbell, 2015).

Statistickými metodami, které se v publikovaných studiích používají nejčastěji, jsou Analýza molekulární variance (AMOVA) a Mantelův korelační test. Princip AMOVy je, že studované populace se rozdělí do skupin podle geografie a jazykové příslušnosti a analýza stanoví, jestli pozorovanou genetickou variaci lépe vysvětlí geografická, nebo jazyková struktura a jestli je rozdíl mezi nimi signifikantní. AMOVu používá sedm studií (Balanovsky et al., 2011; Karafet et al., 2015; Kushniaverich et al., 2015; Mona et al., 2009; Nováčková et al., 2019; Poloni et al., 1997; Wilder et al., 2011).

Mantelův test používá 13 studií (Balanovsky et al., 2011; De Filippo et al., 2012; González-Santos et al., 2022; Harding & Sokal, 1988; Karafet et al., 2015; Kushniarevich et al., 2015; Lansing et al., 2007; Longobardi et al., 2015; Nováčková et al., 2019; Poloni et al., 1997; Rosser et al., 2000; Sokal, 1988; Tambets et al., 2018). Tento test měří korelaci mezi dvěma maticemi párových distancí. Někdy do analýzy vstupuje třetí matice, obsahující geografické vzdálenosti, která plní funkci kontroly v tzv. částečném korelačním testu. Genetická a lingvistická data je tedy třeba převést do distancí. Zatímco pro odhad genetické distance existuje několik metod, šitých na míru různým genetickým datům, odhad jazykové vzdálenosti jsou často problém (Pakendorf; 2015). Některé studie odvozují distance od počtu sdílených kognátů (Lansing et al. 2007), počtu fonémických změn ve slovech základního slovníku (Nováčková et al., 2019), nebo počtu fonémů ve studovaných jazycích (De Filippo et al., 2012; Hunley et al., 2008). Většina studií ale volí jednodušší postup a jazykové distance odvozují z klasifikace (např. nejmenší vzdálenost mezi dialekty jednoho jazyka, největší vzdálenost mezi jazykovými velkorodinami). I podrobná klasifikace obsahuje jen několik málo úrovní, takže v porovnání s genetickými daty je málo informativní. To vnáší do analýz systematický bias, kdy se zejména vzdáleně příbuzné jazyky jeví příbuznější, než odpovídá očekávání (Hunley et al., 2008). V důsledku jsou výsledky studií, které pracují s takto vytvořenými distančními maticemi, nespolehlivé, a to i v případě, že je použita klasifikace věrohodná.

Ani AMOVA, ani Mantelův test nejsou metody pro odhalení koevoluce. Jsou to metody, které slouží pro odhalení příčin genetické diferenciace (geografických nebo jazykových) a odhalení závislosti variability sledovaných znaků (genů, jazyků) na vzdálenosti v prostoru. Dalším problémem Mantelova testu je, že má sklon poskytovat falešné pozitivní výsledky (Somers & Jackson, 2022).

Z 27 studií zahrnutých do rešerše naprostá většina z nich uvádí, že doklady pro koevoluci genů a jazyků našla. Deset z nich nachází koevoluci, 15 nachází částečnou koevoluci a pouze dvě uvádí, že žádné důkazy koevoluce nenašly. Ve skutečnosti jsou výsledky spíše smíšené.

Korelace mezi geny a jazyky vycházející z Mantelova korelačního testu dosahují hodnot od 0,02 do 0,78. Nejnižší hodnoty byly zaznamenány u mtDNA a jazyků u zemědělců obývajících oblasti Sahelu (Nováčková et al., 2019). Slabou korelaci lze vysvětlit tím, že jsou populace patrilokální – ženy opouštějí své rodné komunity, čímž se korespondence mezi mtDNA a etnickou příslušností vytrácí. Naopak nejvyšší hodnota byla naměřena pro

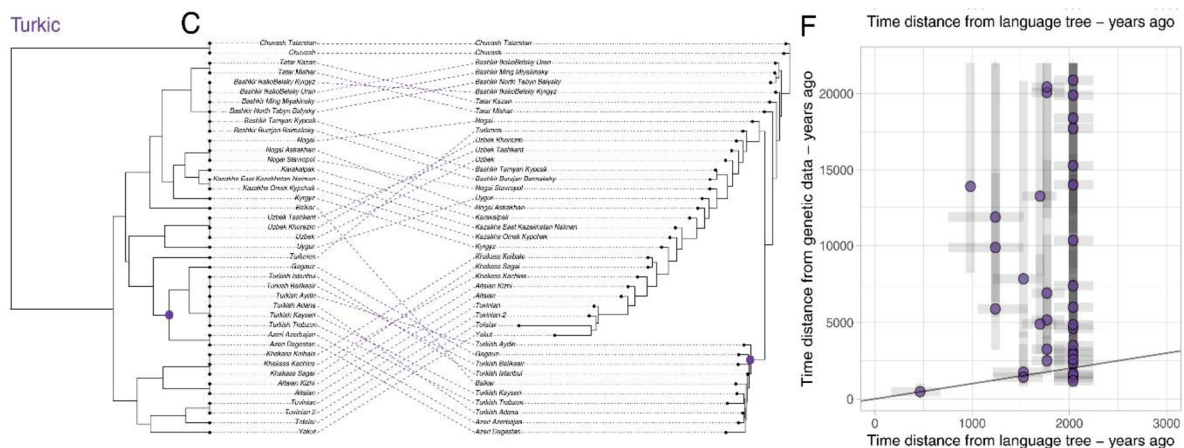
slovanskou větev indoevropských jazyků s použitím mtDNA, NRY i autozomálních SNPs (Kushiarevich et al., 2015), přičemž jsou slovansky mluvící populace geneticky poměrně homogenní a zřetelně odlišné od sousedních populací, které mluví jinými jazyky. Výjimkami jsou některé populace v kontaktních zónách, kdy např. Češi jsou blízce příbuzní svým germánským sousedům. V Evropě vychází korelace mezi geny a jazyky často poměrně silná – platí to jak pro staré studie (např. Barbujani & Sokal, 1990), tak pro ty novější a metodologicky sofistikovanější (Barbieri et al., 2022).

Klíčovým zjištěním je, že korelace mezi geny a jazyky jen málokdy přetrvávají, pokud se do analýzy zahrne také geografie. Po zohlednění geografické vzdálenosti v částečném Mantelově korelačním testu se korelace mezi geny a jazyky sníží nebo přestanou být signifikantní v 11 ze 13 případů. Pouze ve dvou studiích, z nichž jedna se zaměřuje na indonéský ostrov Sumba (Lansing et al., 2007) a druhá na východní Kavkaz (Karafet et al., 2015), korelace mezi geny a jazyky po zohlednění geografické vzdálenosti stoupne. To ukazuje, že korelace mezi geny a jazyky je dána spíše procesem izolace vzdáleností, tedy genetickým a jazykovým driftem, nikoli paralelní genetickou a jazykovou diverzifikací. Mezi sousedními a blízkými populacemi dochází k častějšímu horizontálnímu přenosu genů i jazyků, zatímco mezi vzdálenými populacemi k němu dochází méně. K vysvětlení takto vzniklých genetických a jazykových rozdílů mezi populacemi ale koevoluci nepotřebujeme.

Studie, které žádnou korelaci mezi geny a jazyky nenašly (Mona et al., 2009; Wilder et al., 2011) se obě zaměřovaly na ostrovy a souostroví ve východní Indonésii – Flores, Adonara, Solor, Lembata, Pantar, Alor, Východní Timor a Severní Moluky. Místní obyvatelstvo lze charakterizovat jako geneticky smíšené – asijsko-melanéské, což je důsledkem proběhnuvší západo-východní papuánské expanze v pleistocénu, západo-východní austronéské expanze v holocénu a pozdější papuánské expanze, která přišla z východu. Distribuce papuánských a austronéských jazyků v oblasti nijak neodráží populačně-genetickou strukturu. Je zajímavé, že studie Lansinga a kol. (2017), která se zaměřuje na nedaleký ostrov Sumba ve východní Indonésii, korelaci mezi geny a jazyky našla, ale použila k tomu odlišné metody.

Geny i jazyky jsou evoluční systémy, které v průběhu času mění svou podobu, ale genetické změny probíhají mnohem pomaleji, díky čemuž nám umožňují popsat mnohem starší populační historii. Zatímco kořen genetické fylogeneze lidstva sahá do doby před 150–200 tisíci lety, fylogeneze jazyků sahají do doby před 8–10 tisíci lety (Pagel, 2017). Geny a jazyky nás v tak informují o procesech, které probíhají na odlišných časových škálách.

Zatímco genetická data nám umožňují rekonstruovat dlouhou historii vývoje lidské populace a popsat příbuzenské vztahy mezi kontinenty a rozsáhlými oblastmi světa, jazyky nám umožňují popsat příbuzenské vztahy mezi etniky a studovat lidské migrace, které proběhly v posledních několika tisících letech (Gray et al., 2011). Odlišná časová hloubka genetických a jazykových fylogenezí je zřejmě příčinou toho, že geny a jazyky nejsou v tak silném konfliktu, jak bychom mohli očekávat (viz např. Duda & Zrzavý, 2016). Geny a jazyky jsou často komplementární, ale nejsou kongruentní. Pozorovaná shoda mezi nimi je důsledkem absence konfliktu. Studie, které se na tuto otázku zaměřily (např., Matsumae et al. 2021; Barbieri et al., 2022), ukázaly, že populačně-genetická struktura je starší, než vztahy mezi jazyky (Obr. 7).



**Obr. 7.** Srovnání mezi lingvistickým (vlevo) a genetickým (vpravo) stromem turkické jazykové rodiny (C). Korelace mezi dobou jazykové a genetické diverzity (F). (Barbieri et al., 2022)

I přes rozmanitá genetická a lingvistická data a odlišné metody jejich zpracování téměř všechny publikované studie našly alespoň nějaké doklady o shodě mezi geny a jazyky. Mnohé z těchto výsledků jsou založené na nekvalitních genetických datech a kontroverzních jazykových klasifikacích (např. Cavalli-Sforza et al., 1992; Sokal, 1988), díky čemuž nemají velkou hodnotu. Najdou se však studie, které používají kvalitní lingvistická data (např. Hunley et al., 2008; Karafet et al., 2015; Nováčková et al., 2019) nebo pracují s věrohodnou jazykovou klasifikací (Barbieri et al., 2022). Tyto studie odhalují korespondenci mezi geny a jazyky spíše v jednotlivých oblastech v rámci rozsáhlejšího regionu, nebo ve vztahu konkrétnímu genetickému markeru. Na globální úrovni spolu geny a jazyky spíše nekorelují, respektive jejich korelaci nemáme jak odhalit, protože dosud nedokážeme rekonstruovat globální fylogenezi jazyků. Pokud se však zaměříme na konkrétní oblasti a použijeme kvalitní genetická a lingvistická data, korelace nacházíme často. Demografických a socio-kulturních



procesů, které narušují korespondenci mezi geny a jazyky, je mnoho, a proto bychom neměli koevoluci genů a jazyků očekávat. Nulovou hypotézou by měla být absence koevoluce (Campbell, 2015). Na druhou stranu lingvistická data by měla být běžně součástí populačně-genetických studií, protože nám umožňují výsledky těchto studií smysluplně interpretovat, bez ohledu na to, jestli jsou geny a jazyky v souladu.

#### 4. SEZNAM LITERATURY

Amano, T., Sandel, B., Eager, H., Bulteau, E., Svenning, J. C., Dalsgaard, B., ... & Sutherland, W. J. (2014). Global distribution and drivers of language extinction risk. *Proceedings of the Royal Society B: Biological Sciences*, 281(1793), 20141574.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics*, 23(2), 147-147.

Atkinson, Q. D., & Gray, R. D. (2005). Curious parallels and curious connections - Phylogenetic thinking in biology and historical linguistics. *Syst Biol*, 54(4), 513-526. doi:10.1080/10635150590950317

Barbujani, G., & Sokal, R. R. (1990). Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences of the United States of America*, 87(5), 1816-1819. doi:10.1073/pnas.87.5.1816

Barbujani, Guido. 1997. dna variation and language affinities. *American Journal of Human Genetics* 61: 1011–1014.

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. MIT press.

Bhugra, D., Gupta, S., Bhui, K., Craig, T. O. M., Dogra, N., Ingleby, J. D., ... & Tribe, R. (2011). WPA guidance on mental health and mental health care in migrants. *World Psychiatry*, 10(1), 2-10.

Boas F (1911) *Handbook of American Indian Languages*, part 1. Washington, D.C.: Government Printing Office.

Bolnick, D. A. W., Shook, B. A. S., Campbell, L., & Goddard, I. (2004). Problematic use of Greenberg's linguistic classification of the Americas in studies of Native American genetic variation. *The American Journal of Human Genetics*, 75(3), 519-522.

Bromham, L., & Penny, D. (2003). The modern molecular clock. *Nature Reviews Genetics*, 4(3), 216-224.

Bruchová, H., Kráčmarová, A., Černý, V., & Brdička, R. (2005). Využití technologie LabMAP Luminex pro detekci SNP polymorfismů. *Klin. Biochem. Metab*, 13, 87-91.

Bryant, D., & Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution*, 21(2), 255-265.

C Comas, David, Elena Bosch, and Francesc Calafell. 2008. Human genetics and languages. *Encyclopedia of Life Sciences* (doi: 10.1002/9780470015902.a0020810).

Campbell, L. (2015). Do Languages and Genes Correlate? *Language Dynamics and Change*, 5(2), 202-226. doi:doi:http://dx.doi.org/10.1163/22105832-00502007

Cavalli-Sforza, L. L. (1991). Genes, peoples and languages. *Scientific American*, 265(5), 104-110.

Cavalli-Sforza, L. L., Minch, E., & Mountain, J. L. (1992). Coevolution of genes and languages revisited. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12), 5620-5624. doi:10.1073/pnas.89.12.5620

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., & Mountain, J. (1988). Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences of the United States of America*, 85(16), 6002-6006. doi:10.1073/pnas.85.16.6002

Colvin, S. (2011). The koine: A new language for a new world. *Creating a Hellenistic world*, 31-45.

Comrie, B., Matthews, S., & Polinsky, M. (2003). *The atlas of languages: The origin and development of languages throughout the world*.

Daneš, F. (2009). *Kultura a struktura českého jazyka*. Karolinum Press.

Darwin C. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (1st ed.). London: John Murray.

Duda, P. (2017). Kulturní fylogenetika – lidské kultury jako biologické druhy. In L. Ovčáčková (Ed.), *O původu kultury: Biologické, antropologické a historické koncepce kulturní evoluce* (pp. 230-279). Praha: Academia.

- Dunn, M., Terrill, A., Reesink, G., Foley, R. A., & Levinson, S. C. (2005). Structural phylogenetics and the reconstruction of ancient language history. *Science*, *309*(5743), 2072-2075.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the World*. Twenty-sixth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Gould SJ (1991) *Bully for Brontosaurus*. New York: Norton
- Gray, R. D., Atkinson, Q. D., & Greenhill, S. J. (2011). Language evolution and human history: what a difference a date makes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1090-1100.
- Greenberg, J. H. (1963). The languages of Africa. *International journal of American linguistics*.
- Greenhill, S. J., Wu, C. H., Hua, X., Dunn, M., Levinson, S. C., & Gray, R. D. (2017). Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*, *114*(42), E8822-E8829.
- Hammarström, H., Forkel, R., Haspelmath, M., & Nordhoff, S. (2014). *Glottolog 2.3*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online v.: <http://glottolog.org>.
- Harding, R. M., & Sokal, R. R. (1988). Classification of the European language families by genetic distance. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(23), 9370-9372. doi:10.1073/pnas.85.23.9370
- Haspelmath, M., & Tadmor, U. (2009). The loanword typology project and the world loanword database. *Loanwords in the world's languages: A comparative handbook*, *1*, 34.
- Huelsenbeck, J. P., Rannala, B., & Yang, Z. (1997). Statistical tests of host-parasite cospeciation. *Evolution*, *51*(2), 410-419.
- Kočárek, E. (2008). *Genetika: obecná genetika a cytogenetika, molekulární biologie, biotechnologie, genomika*. Scientia.

- Krause, J., & Trappe, T. (2019). *Die Reise unserer Gene: eine Geschichte über uns und unsere Vorfahren*. Ullstein Buchverlage.
- Kroeber, A. L. (1923). *Anthropology: Culture, patterns & processes* (1 ed.). New York: Harcourt, Brace & World.
- Longobardi, G., Ghirotto, S., Guardiano, C., Tassi, F., Benazzo, A., Ceolin, A., & Barbujani, G. (2015). Across language families: genome diversity mirrors linguistic variation within Europe. *American journal of physical anthropology*, *157*(4), 630-640.
- Macholán, M. (2014). *Základy fylogenetické analýzy*. Masarykova univerzita.
- McDonald, D. (2013). Lecture 8. Population Genetics VI: Introduction to Microsatellites: From Theory to Lab. Practice.
- McMahon, R. (2004). Genes and languages. *Public Health Genomics*, *7*(1), 2-13.
- Millar, R. M. C., & Trask, L. (2015). Trask's historical linguistics.
- Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, *541*(7637), 302-310.
- Pääbo, S. (2015). The diverse origins of the human gene pool. *Nature Reviews Genetics*, *16*(6), 313-314.
- Page, R. D. (Ed.). (2003). *Tangled trees: phylogeny, cospeciation, and coevolution*. University of Chicago press.
- Pagel, M. (2017). Darwinian perspectives on the evolution of human languages. *Psychonomic bulletin & review*, *24*, 151-157.
- Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*(7163), 717-720.
- Pakendorf, B. (2014). Coevolution of languages and genes. *Current opinion in genetics & development*, *29*, 39-44.
- Pakendorf, B. (2015). Historical linguistics and molecular anthropology. *The Routledge handbook of historical linguistics*, 627-642.

- Peter, B. M., Petkova, D., & Novembre, J. (2020). Genetic landscapes reveal how human genetic diversity aligns with geography. *Molecular biology and evolution*, 37(4), 943-951.
- Romaine, S. (2017). *Pidgin and creole languages*. Routledge
- Ruhlen, M. (1987). *A guide to the world's languages: Volume I, Classification*. Stanford University Press.
- Sapir E (1931) Language, race, and culture. In: Calverton VF, editor. *The Making of Man: An Outline of Anthropology*. New York: The Modern Library. pp. 142–154.
- Snustad, D. P., Simmons, M. J., Relichová, J., Doškař, J., Fajkus, J., Hořín, P., ... & Relichová, J. (2009). *Genetika*. Masarykova univerzita.
- Sokal, R. R., Oden, N. L., Legendre, P., Fortin, M. J., Kim, J., & Vaudor, A. (1989). Genetic differences among language families in Europe. *American Journal of Physical Anthropology*, 79(4), 489-502.
- Thompson, J. N. (2014). *Interaction and coevolution*. University of Chicago Press.
- Tömöry, G., Csányi, B., Bogácsi-Szabó, E., Kalmár, T., Czibula, A., Csősz, A., ... & Raskó, I. (2007). Comparison of maternal lineage and biogeographic analyses of ancient and modern Hungarian populations. *American journal of physical anthropology*, 134(3), 354-368.
- Urban, M. (2020). Mountain linguistics. *Language and Linguistics Compass*, 14(9), e12393.
- Urban, M. (2021). The geography and development of language isolates. *Royal Society Open Science*, 8(4), 202232. doi: doi:10.1098/rsos.202232
- Wichmann, S., Holman, E. W., Bakker, D., & Brown, C. H. (2010). Evaluating linguistic distance measures. *Physica A: Statistical Mechanics and its Applications*, 389(17), 3632-3639.

## 5. SEZNAM LITERATURY – TABULKA

- Baker, J. L., Rotimi, C. N., & Shriner, D. (2017). Human ancestry correlates with language and reveals that race is not an objective genomic classifier. *Scientific reports*, 7(1), 1572.
- Balanovsky, O., Dibirova, K., Dybo, A., Mudrak, O., Frolova, S., Pocheshkhova, E., ... & Genographic Consortium. (2011). Parallel evolution of genes and languages in the Caucasus region. *Molecular biology and evolution*, 28(10), 2905-2920.
- Barbieri, C., Blasi, D. E., Arango-Isaza, E., Sotiropoulos, A. G., Hammarström, H., Wichmann, S., Greenhill, S. J., Gray, R. D., Forkel, R., Bickel, B., & Shimizu, K. K. (2022). A global analysis of matches and mismatches between human genetic and linguistic histories. *Proceedings of the National Academy of Sciences*, 119(47), e2122084119.
- Barbujani, G., & Sokal, R. R. (1990). Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences*, 87(5), 1816-1819.
- Cavalli-Sforza, L. L., Minch, E., & Mountain, J. L. (1992). Coevolution of genes and languages revisited. *Proceedings of the National Academy of Sciences*, 89(12), 5620-5624.
- Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., & Mountain, J. (1988). Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences*, 85(16), 6002-6006.
- De Filippo, C., Bostoen, K., Stoneking, M., & Pakendorf, B. (2012). Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proceedings of the Royal Society B: Biological Sciences*, 279(1741), 3256-3263.
- Duda, P., & Zrzavý, J. (2016). Human population history revealed by a supertree approach. *Scientific Reports*, 6(1), 1-10.
- González-Santos, M., Montinaro, F., Grollemund, R., Marnetto, D., Atadzhanov, M., May, C. A., Mabunda, N., Maret, P., Oosthuizen, O., Oosthuizen E., & Capelli, C. (2022). Exploring the relationships between genetic, linguistic and geographic distances in Bantu-speaking populations. *American Journal of Biological Anthropology*, 179(1), 104-117.
- Harding, R. M., & Sokal, R. R. (1988). Classification of the European language families by genetic distance. *Proceedings of the National Academy of Sciences*, 85(23), 9370-9372.

- Hunley, K. L., Cabana, G. S., Merriwether, D. A., & Long, J. C. (2007). A formal test of linguistic and genetic coevolution in native Central and South America. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 132(4), 622-631.
- Hunley, K., Dunn, M., Lindström, E., Reesink, G., Terrill, A., Healy, M. E., Koki, G., Friedlaender, F. R., & Friedlaender, J. S. (2008). Genetic and linguistic coevolution in Northern Island Melanesia. *PLoS genetics*, 4(10), e1000239.
- Karafet, T. M., Bulayeva, K. B., Nichols, J., Bulayev, O. A., Gurganova, F., Omarova, J., Yepiskoposyan, L., Savina, O. V., Rodrigue, B. H., & Hammer, M. F. (2016). Coevolution of genes and languages and high levels of population structure among the highland populations of Daghestan. *Journal of human genetics*, 61(3), 181-191.
- Kushniarevich, A., Utevska, O., Chuhryaeva, M., Agdzhoyan, A., Dibirova, K., Uktveryte, I., ... & Balanovsky, O. (2015). Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data. *PloS one*, 10(9), e0135820.
- Lansing, J. S., Cox, M. P., Downey, S. S., Gabler, B. M., Hallmark, B., Karafet, T. M., Norquest, P., Schoenfelder, J. W., Sudoyo, H., Watkins, J. C., & Hammer, M. F. (2007). Coevolution of languages and genes on the island of Sumba, eastern Indonesia. *Proceedings of the National Academy of Sciences*, 104(41), 16022-16026.
- Longobardi, G., Ghirotto, S., Guardiano, C., Tassi, F., Benazzo, A., Ceolin, A., & Barbujani, G. (2015). Across language families: genome diversity mirrors linguistic variation within Europe. *American journal of physical anthropology*, 157(4), 630-640.
- Matsumae, H., Ranacher, P., Savage, P. E., Blasi, D. E., Currie, T. E., Koganebuchi, K., ... & Bickel, B. (2021). Exploring correlations in genetic and cultural variation across language families in northeast Asia. *Science advances*, 7(34), eabd9223.
- Mona, S., Grunz, K. E., Brauer, S., Pakendorf, B., Castri, L., Sudoyo, H., Marzuki, S., Barnes, R. H., Schmidtke, J., Stoneking, M., & Kayser, M. (2009). Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Molecular Biology and Evolution*, 26(8), 1865-1877.



Nováčková, J., Čížková, M., Mokhtar, M. G., Duda, P., Stenzl, V., Tříška, P., Hofmanová, Z., & Černý, V. (2020). Subsistence strategy was the main factor driving population differentiation in the bidirectional corridor of the African Sahel. *American Journal of Physical Anthropology*, 171(3), 496-508.

Poloni, E. S., Semino, O., Passarino, G., Santachiara-Benerecetti, A. S., Dupanloup, I., Langaney, A., & Excoffier, L. (1997). Human genetic affinities for Y-chromosome P49a, f/TaqI haplotypes show strong correspondence with linguistics. *The American Journal of Human Genetics*, 61(5), 1015-1035.

Robbeets, M., Bouckaert, R., Conte, M., Savelyev, A., Li, T., An, D. I., ... & Ning, C. (2021). Triangulation supports agricultural spread of the Transeurasian languages. *Nature*, 599(7886), 616-621.

Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., ... & Jobling, M. A. (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *The American Journal of Human Genetics*, 67(6), 1526-1543.

Sokal, R. R. (1988). Genetic, geographic, and linguistic distances in Europe. *Proceedings of the National Academy of Sciences*, 85(5), 1722-1726.

Sokal, R. R., Oden, N. L., Legendre, P., Fortin, M. J., Kim, J., & Vaudor, A. (1989). Genetic differences among language families in Europe. *American Journal of Physical Anthropology*, 79(4), 489-502.

Tambets, K., Yunusbayev, B., Hudjashov, G., Ilumäe, A. M., Rootsi, S., Honkola, T., ... & Metspalu, M. (2018). Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biology*, 19(1), 1-20.

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., ... & Williams, S. M. (2009). The genetic structure and history of Africans and African Americans. *science*, 324(5930), 1035-1044.

Wilder, J. A., Cox, M. P., Paquette, A. M., Alford, R., Satyagraha, A. W., Harahap, A., & Sudoyo, H. (2011). Genetic continuity across a deeply divergent linguistic contact zone in North Maluku, Indonesia. *BMC genetics*, 12(1), 1-8.