



Posudek na diplomovou práci Bc. Jana Maška „Populační struktura a migrace v africkém sahelu z pohledu variability kontrolního úseku mtDNA“

Předložená diplomová práce prezentuje výsledky molekulárněgenetické analýzy kontrolní oblasti (CR) mitochondriální DNA (mtDNA) vzorku zahrnujícího 104 sekvencí pocházejících ze tří populací z Burkina Faso. Tento materiál byl doplněn o publikované údaje a celkově obsahoval 1075 sekvencí z 24 populací včetně saharských Berberů (Maurové, Tuaregové). Cílem práce bylo 1) odhadnout míru přispění 2. hypervariabilního segmentu CR (HVS-2) k přesnosti určení sekvencí do haploskupin L0–L6 a 2) odhadnout genetickou proměnlivost v rámci populačních vzorků i míru divergence mezi nimi a na základě získaných dat se pokusit odhadnout migrační historii lidských populací v oblasti západního Sahelu.

Autor dochází k závěru, že 1) HVS-2 má při určování haploskupin zpřesňující význam za předpokladu, že je referenční databáze dostatečně obsáhlá a 2) populační vzorky mtDNA jsou v rámci západního Sahelu poměrně homogenní, což vysvětluje buď vysokou mírou toku genů, nebo recentním původem těchto populací. Výsledky naznačují, že na rozložení genetické diverzity mají geografické faktory větší vliv než jazykové. Genetická blízkost populací Sahelu a oblastí jižně od něj na jedné straně a jejich odlišnost od populací saharských na straně druhé naznačuje větší význam migračního koridoru vedoucího podél sahelského pásma (pravděpodobně zejména z východu na západ) v porovnání s trans-saharskými migracemi. Sahara se tak jeví jako silnější geografická bariéra, než se dříve soudilo.

Práci jsem přečetl s chutí, i když je na ní patrné, že byla – alespoň se mi to tak jeví – psána na poslední chvíli (jak tomu pohříchu zpravidla bývá). Tomu odpovídá řada formulačních nepřesností (např. „xxx let před současností“ místo „před xxx lety“, „fylogenetická klasifikace“ (str. 8, 9, 46, 56, 62) je ve skutečnosti jen zařazování jednotlivých haplotypů do haploskupin atd.). K nim patří i ne zcela jasné vysvětlení, zda autor založil svoje analýzy pouze na sekvencích obou HVS, nebo celé CR (čemuž by napovídala délka 1424 bp) – až na str. 44 to objasňuje věta „Nejprve jsem haploskupiny určil podle úseků HVS-I a HVS-II a poté jsem prověřil celou kontrolní oblast.“

Vcelku práce působí poněkud nevyváženým dojmem, její převážnou většinu zabírá literární rešerše, naopak popis analytických metod a diskuse jsou spíše stručné. Především by práci prospělo, kdyby místo zbytečného popisu, k čemu slouží v buňce mitochondrie nebo jaký je princip elektroforézy, byly prezentovány nebo blíže specifikovány některé klíčové informace. Např. není jasné, co znamená věta „měří se proporce rozptylu alelických frekvencí v subpopulacích“ (str. 45). Celá pasáž popisující F_{ST} a Reynoldsovy distance je poněkud mlhavá; není jasné, jak autor došel zrovna ke K2P modelu (skutečnost, že v sekvencích se vyskytuje podstatně více transicí než transverzí automaticky neopravňuje k použití K2P, nehledě na to, že v těchto sekvencích existují výrazné rozdíly v zastoupení jednotlivých bází, což naznačuje nutnost použití složitějších modelů, než K2P, např. Hasegawova-Kishinova-Yanova, GTR apod.). Není jasné, co znamená „hodnota 0,4“ u gama korekce (předpokládám, že jde o parametr tvaru alfa, ale není specifikováno), taky není jasné, jak autor došel zrovna k této hodnotě, použití literárních hodnot mi nepřijde korektní, hodnota alfa by měla být odhadnuta z konkrétních dat (mimořádně z textu není zřejmé, zda byl použit kontinuální nebo diskontinuální gama model). Zcela zásadní je popis, jakým způsobem byly sekvence kombinovány do skupin pro AMOVA (str. 45), čtenář to musí dedukovat z výsledků. U Mantelova testu chybí popis, jak byly získány geografické vzdálenosti a kolik bylo použito permutačních kroků (předpokládám, že v Arlequinu byly použity permutace a ne třeba bootstrap jako v programu FSTAT).

Je jisté, že ze získaných dat by se asi dalo vytěžit více informací, ale celkově k práci nemám (až na výše popsané výhrady) závažnější připomínky. Níže uvádím některé drobné poznámky a dotazy (diplomanta prosím o reakci pouze na ty, které jsou vyznačeny tučným písmem).

1. Název práce není zcela přesný, lépe by bylo uvést „Populační struktura a migrace v západním Sahelu“ nebo „západoafrickém Sahelu“.

2. Vyznačení oblasti Sahelu není v práci jednoznačné, např. str. 11: “ Pásmo sahelu je označení pro přechodovou zónu, jejíž severní hranici tvoří jižní okraj Sahary a na jihu je vymezena severním okrajem tropického deštného lesa.“ Naopak na str. 12 vsunuto mezi tropické pralesy a Sahel ještě pásmo savan (viz také obr. 1 na str. 15 a dále v textu).

3. str. 21: **Jak autor vysvětluje, že odhady divergencí (TMRCA) pro Y jsou většinou nižší než pro mtDNA?**

4. str. 25: **Jak sahelští pastevcí chrání svoje stáda proti mouchám tse-tse?**

5. Osobně se domnívám, že definice kontrolní oblasti (str. 33) by neměla zahrnovat geny pro t-RNA. CR a geny pro tRNA mají odlišnou funkci. Ve skutečnosti CR zahrnuje D-smyčku spolu s okolními oblastmi transkripčních promotorů.

6. Zaujala mě hodnota poměru transicí a transverzí „až 30:1“ (str. 35-36) – nemám sice zkušenosti s lidskou genetikou, ale pokud vím, pro mtDNA savců se uvádí až 10:1, z vlastní zkušenosti mohu říci, že pro sekvence CR u myší rodu *Mus* se tyto hodnoty pohybovaly kolem 2–5:1.

7. HVS-2 obsahuje 3 konzervativní segmenty (conserved sequence blocks). **Může mít tato skutečnost vliv na přesnější určení na základě HVS-2?** (Na jednu stranu CSB mohou snižovat proměnlivost sekvencí, na druhou stranu ovšem mohou snižovat nežádoucí „šum“.)

8. Obecně lze říci, že Výsledky jsou podány schematicky a popisně, výčet hodnot různých proměnných, přičemž často jsou prezentovány jako „rozdílné“, aniž se berou v úvahu intervaly spolehlivosti. Podobně u Harpendigova indexu se mluví o „vysokých“, resp. „vyšších“ a „nízkých“, resp. „nižších“ hodnotách, aniž je ukázáno, zda jsou tyto hodnoty signifikantní, tj. ve shodě s očekávaným rozdělením.

9. Autor by měl mít na paměti (a v práci by to mělo být zmíněno), že genetické ukazatele typu F_{ST} jsou validní pouze tehdy, jsou-li populace v rovnováze (např. drift – tok genů), což alespoň pro některé studované populace nemusí platit. Podobně je třeba mít na paměti, že byla studována pouze mtDNA, tzn. že výsledky neukazují na „homogenitu *populací*“, ale pouze na homogenitu *mtDNA* (tj. ženské části populace).

10. str. 55: Formulace „Mantelův test korelace matic mezi geografickými a genetickými vzdálenostmi demonstruje vliv geografie.“ není zcela přesná – statistická průkaznost naznačuje izolaci vzdáleností. (Pro vysvětlení: geografie může populace ovlivnit např. i tak, že je oddělí neprůchodnou geografickou bariérou, přičemž populace na obou stranách bariéry budou geneticky homogenní).

11. str. 59: Saharská skupina: Autor má pravdu v tom, že „izolovanost Tuaregů by mohla být vysvětlena i soudržností a opozicí vůči ostatním pastevcům zapříčiněnou soutěží o omezený přístup k pastvinám...“, vzájemná soudržnost se ale zdá být v rozporu s velkou genetickou diferenciací samotných Tuaregů, jak ukazují výsledky MDS.

12. str. 59: Nerozumím větě „Analýza korelace genetických a geografických vzdáleností pomocí Mantelova testu prokázala stav vzdálenostní izolace i tehdy, když byly geneticky odlišné skupiny Tuaregů do analýzy začleněny.“ – na str. 55 jsou prezentovány výsledky pouze jednoho testu a není jasné, zda ten Tuaregy zahrnoval, nebo ne.

13. str. 59: Tvzení, že „působení genetického driftu [u Tuaregů z Libye] naznačují i výsledky této diplomové práce.“ není přesné. 1) drift se vyskytuje ve *všech* zkoumaných populacích a 2) autor má zřejmě na mysli poměrně vysoké hodnoty párových F_{ST} vzorku TAWT – pak by bylo správnější napsat, že zmíněná populace je vůči ostatním zkoumaným populačním vzorkům izolovaná. Index F_{ST} je dán efektivní velikostí (sub)populací *a tokem genů*.

14. str. 59: Zaujala mě věta „Spojitost Tuaregů z Nigeru se subsaharskými populacemi byla detekována již dříve (Pereira et al. 2010), prozatím však pro ni není žádné vysvětlení.“ **Skutečně se o tom nic neví? Neexistují smíšené páry Tuaregů (předpokládá bych, že v tomto případě spíš mužů – bohužel nemám k dispozici frekvence jednotlivých haploskupin u saharských a sahelových Tuaregů) a žen z jiných pasteveckých etnik?** Pak by existence ne-tuarežské mtDNA u sahelských Tuaregů byla celkem logická. Z jiných částí kontinentu jsou podobné případy mezikmenových či mezietnických sňatků známy. Naopak v případě saharských Tuaregů k tomu dochází asi naprosto výjimečně. Mohl by k tomu autor říci něco víc?

15. str. 61: „Tento vztah mezi nigerskými Tuaregy z oblasti Tanut a populacemi žijícími se v sahelském pásmu může být zapříčiněn klimatickými výkyvy v průběhu holocénu, kdy byly populace Tuaregů nuceny se stáhnout ze Sahary do jižnějších oblastí a přijít tak do kontaktu se subsaharskými populacemi. Neumím ovšem vysvětlit, proč se jednalo jen o jednu skupinu Tuaregů.“ Laicky mě napadá, že v průběhu zhoršení klimatu se na jih mohli těžko stáhnout všichni Tuaregové, včetně kmenů ze severní Sahary – naopak bych čekal, že se tam stáhly (resp. už tam zůstaly) jen populace z jižních oblastí tuarežské oikumeny.

16. str. 61: „...je pravděpodobné, že jsou odvozeny z jedné ancestrální populace a/nebo že mezi nimi působil silný genový tok.“ – **co autor pokládá za pravděpodobnější vysvětlení?** Z toho, co se píše v úvodním literárním přehledu, se ukazuje, že populace v západní Africe nejsou příliš starobylé. V Diskusi se měly tyto skutečnosti konfrontovat s tím, co je uvedeno v literárním přehledu.

17. Osobně mě zaujala vysoká míra vnitropopulační divergence u většiny studovaných populací. **Rád bych se proto autora zeptal, zda je mu něco známo o úrovni genetické variability CR v lidských populacích mimo africký kontinent?** (V této souvislosti ještě poznámka: v práci se uvádí, že variabilita vykazuje geografický gradient – to by se ovšem dalo místo pouhého narativního vyjádření explicitně testovat.)

Přes některé dílčí nedostatky jsem přesvědčen, že diplomová práce Jana Maška je kvalitní, dobře zpracovaná a přináší některé zajímavé poznatky. Proto ji doporučuji k obhajobě.

V Brně dne 2. 9. 2014

Prof. RNDr. Miloš Macholán, CSc.