

Univerzita Karlova v Praze
Přírodovědecká fakulta
Katedra antropologie a genetiky člověka

Diplomová práce

**Diverzita mitochondriální DNA
jemenských populací: příspěvek
regionálního vzorkování k
interpretaci genetické struktury jihu
Arabského poloostrova**

autor: Bc. Jakub Rídl

školitel: Mgr. Viktor Černý, PhD.

Praha, 2006

BIBLIOGRAFICKÁ IDENTIFIKACE

NÁZEV DIPLOMOVÉ PRÁCE	Diverzita mitochondriální DNA jemenských populací: příspěvek regionálního vzorkování k interpretaci genetické struktury jihu Arabského poloostrova
AUTOR DIPLOMOVÉ PRÁCE	Bc. Jakub Rídl
VYSOKÁ ŠKOLA, FAKULTA	Univerzita Karlova v Praze, Přírodovědecká fakulta
OBOR STUDIA	Biologie
SPECIALIZACE	Antropologie a genetika člověka
ROK OBHAJOBY	2006
VEDOUcí DIPLOMOVÉ PRÁCE	Mgr. Viktor Černý, PhD.
KLÍČOVÁ SLOVA	diverzita mtDNA – HVS-I – Jemen – regionální vzorkování – populační srovnání

Svoluji k zapůjčení této práce ke studijním účelům, zároveň prosím o evidenci zájemců a přesnost případného citování převzatých údajů:

Jméno a příjmení:

Datum vypůjčení:

PROHLÁŠENÍ

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Veškeré literární prameny a informace, které jsem při přípravě využil jsou uvedeny v seznamu použité literatury.

Praha, srpen 2006



.....

PODĚKOVÁNÍ

Rád bych na tomto místě poděkoval svému školiteli dr. Viktorovi Černému, který vymyslel a zrealizoval celý projekt výzkumu diverzity mitochondriální DNA jemenských populací, na kterém jsem se mohl podílet. Děkuji mu také za motivující přístup k mé práci a veškeré rady a připomínky. Můj vděk patří rovněž všem členům archeogenetické laboratoře Archeologického ústavu AVČR. Převážně Mgr. Martině Žaloudkové, která vedla mé první, ale také všechny následující kroky v laboratoři. Dále děkuji členům Oddělení biologické antropologie Katedry antropologie FF ZČU v Plzni, kteří mne vždy motivovali k vědecké činnosti.

ABSTRAKT

Archeogenetika se zabývá aplikací molekulárně-genetických metod na otázky evoluce člověka a šíření jednotlivých lidských populací. Díky svým vlastnostem je ve středu zájmu převážně mitochondriální DNA (mtDNA). Přes důležitou polohu Jemenu – spojnice mezi Afrikou, Blízkým východem a Asií – nebyla diverzita mtDNA v této oblasti systematictěji studována. Tato diplomová práce prezentuje analýzu 150 nových sekvencí HVS-I mtDNA ze tří oblastí Jemenu – Al-Achkúmu, Tihamy a Hadramatu. Díky regionálnímu postupu při sběru dat jsme byli schopni odhalit rozdíly v diverzitě mtDNA v různých oblastech Jemenu včetně jejich vztahů k dalším populacím Afriky a Blízkého východu. Soubory Al-Achkúm a Tihama se v populačně genetickém srovnání s okolními populacemi blíží populacím Blízkého východu, zatímco populace z Hadramautu spadá blíže populacím východní Afriky. Kromě vlastních závěrů nabízí diplomová práce postup pro budoucí výzkumy genetické diverzity populací jižní Arábie.

OBSAH

1	ÚVOD	8
1.1	Vlastnosti mitochondriální DNA	9
1.1.1	Organizace lidského mitochondriálního genomu	10
1.1.2	Počet kopií v buňce	10
1.1.3	Dědičnost	11
1.1.4	Rekombinace	12
1.1.5	Mutační rychlost	14
1.1.6	Selekční neutralita	15
1.2	Studium mtDNA	16
1.2.1	Přístupy ke studiu evoluční a populační historie	17
1.2.1.1	Fylogenetický přístup	17
1.2.1.2	Populační přístup	19
1.3	Přehled výzkumů mitochondriální DNA	20
1.4	Archeologický a historický kontext	22
1.5	Cíle práce	24
2	MATERIÁL A METODY	24
2.1	Soubory	24
2.2	Laboratorní postupy	28
2.3	Analýza dat	29
2.3.1	Klasifikace haplotypů	30
2.3.2	Statistické metody	30
3	VÝSLEDKY	31
3.1	Intrapopulační úroveň	36

3.2 Populační srovnání (F_{ST})	41
3.3 Rozdělení genetické diverzity (AMOVA)	43
3.4 Haplotypy sdílené mezi populacemi	45
4 DISKUZE	46
5 ZÁVĚR.....	51
6 PŘÍLOHY	53
7 LITERATURA.....	55
8 RESUMÉ.....	69

1 ÚVOD

Archeogenetika se zabývá aplikací molekulárně-genetických metod při studiu lidské minulosti. Zahrnuje dvě cesty: 1) výzkumy genetické diverzity dnešních populací a 2) analýzu tzv. aDNA přímo z ostatků minulých jedinců (Černý et al., 2003; Černý et al., 1997). Zájem o DNA přitom vychází z představy molekuly jako dokumentu minulosti. Britský genetik Brian Sykes používá metaforu matrice, kterou si každý nese v buňkách a do které se jednou za čas zapíše nový záznam prostřednictvím mutací (Sykes, 2004).

Studium genetických rozdílů mezi jedinci však předchází samotnému objevení struktury DNA. Zřejmý doklad byl podán v roce 1901, kdy Karl Landsteiner popsal rozdílné krevní skupiny (Landsteiner, 1901; Landsteiner, 1931) a následně jejich využití pro forenzní praxi (Landsteiner & Richter, 1903). První metodologické nástroje pro studium genetické diverzity populací poskytla populační genetika, jejíž konstituování v první polovině 20. století je spjato především se jmény Fisher, Haldane a Wright. V 50. a 60. letech začaly systematické výzkumy genetické diverzity lidských populací zacílené na polymorfismy v řetězcích proteinů. K přesunu od proteinů ke přímému studiu genů došlo v 80. letech (souhrnně např. Avise, 2000; Renfrew, 1992).

Oproti studiu klasických (chromosomálních) markerů a hodnocení rozdílů mezi populacemi na základě frekvence alel přináší analýza uniparentálně děděných lokusů nové možnosti – sledovat rekombinacemi nepřerušené linie, jejich vztahy znázornit fylogenetickým schématem a zasadit do časového a geografického rámce. Díky specifickým vlastnostem je mitochondriální DNA (mtDNA) v posledních desetiletích cílem intenzivního zájmu v rámci výzkumů lidské populační historie, evoluce či šíření současných lidí (souhrnně Forster, 2004; Pakendorf & Stoneking, 2005; Richards & Macaulay, 2001). Později následoval také zájem o Y chromosom (souhrnně Hammer & Zegura, 2002; Underhill, 2004). V poslední době můžeme zaznamenat posun od výzkumů jednoho

systemu markerů ke kombinaci více lokusů s rozdílným způsobem dědičnosti (Dupanloup et al., 2004; Jorde et al., 2000; Knight et al., 2004).

Prezentovaná práce se zabývá genetickou diverzitou jemenských populací na základě sekvenace prvního hypervariabilního segmentu (HVS-I) kontrolního regionu mitochondriální DNA. Zacílena je především na populačně genetické srovnání s publikovanými daty z okolních regionů. Úvodní části textu jsou věnovány přehledu vlastností a studia mtDNA, který zároveň otevírá některé sporné otázky. Dále je přehled zaměřen na dosavadní výzkumy mtDNA s ohledem na roli jižní části Arabského poloostrova. Přehled archeologických a historických pramenů z oblasti Arabského poloostrova zahrnuje následující sekce. Úvodní kapitola je zakončena prezentací cílů práce.

1.1 Vlastnosti mitochondriální DNA

Mitochondrie jsou organely v cytoplazmě eukaryotických buněk zodpovědné (mimo jiné) za aerobní metabolismus – produkují a ukládají energii v podobě ATP a teplo pro udržování tělesné teploty. Z hlediska struktury je důležité zmínit přítomnost dvou membrán – vnější a vnitřní, jejíž povrch je znásoben zvlněním a vybíhá dovnitř v útvary zvané kristy. ATP slouží jako přenašeč energie, kterou je buňka schopna dále využít. Vzniká na vnitřní membráně jako produkt procesu oxidační fosforylace (OXPHOS) s využitím potravních kalorií a kyslíku (souhrnně např. Scheffler, 2000).

Většina genetického materiálu jedince (více než 99%) je obsažena v jádře buněk – jde o jadernou, nukleární DNA (nDNA). Mitochondrie však obsahují vlastní mitochondriální DNA (mtDNA). Mitochondriální proteiny jsou převážně kódované v jaderném genomu a do mitochondrií jsou dopravovány. Část proteinů je ale kódována geny mitochondriálního genomu translatovanými na mitochondriálních ribosomech. Mitochondrie jsou tak pod duální genetickou kontrolou (Taanman, 1999).

1.1.1 Organizace lidského mitochondriálního genomu

Lidský mitochondriální genom je kruhová molekula o délce 16568 bp (MITOMAP: A Human Mitochondrial Genome Database. <http://www.mitomap.org>, 2006) tvořená dvěma řetězci, které se liší obsahem G a T bází. Rozlišuje se H (heavy) a L (light) řetězec. Část molekuly je tvořena trojvláknovou strukturou, kde je H-řetězec nahrazen krátkým úsekem komplementárním k L-řetězci, který formuje kličku, podle níž je tato část nazývána jako D-loop (displacement loop). D-loop tvořící přibližně 7% molekuly mtDNA je jediným nekódujícím úsekem. Obsahuje začátek replikace H-řetězce včetně promotorů pro transkripci a představuje tak kontrolní region. Kódující oblast neobsahuje introny. Sestává z 37 genů pro 13 polypeptidů (uplatňujících se v OXPHOS systému), 22 molekul tRNA (díky specifickému genetickému kódu mitochondrií jde o kompletní počet potřebný pro translaci všech 13 proteinů) a 2 molekuly rRNA (Anderson et al., 1981; Lightowlers et al., 1997; Taanman, 1999). Většina genů je obsažena na H-řetězci, přesto se pořadí bází sekvencí konvenčně uvádí podle L-řetězce (Anderson et al., 1981).

V průběhu evoluce došlo k transportům některých úseků mtDNA do jaderného genomu, kde ztratily svou původní funkci (tzv. Numts) (Ricchetti et al., 2004; Tourmen et al., 2002; Woischnik & Moraes, 2002; Zhang & Hewitt, 1996).

1.1.2 Počet kopií v buňce

Buňky obsahují v jádře pouze dvě kopie nDNA (od matky a otce), zatímco v cytoplasmě různé množství mitochondrií (v závislosti na typu a energetických nárocích tkáně), které nesou více kopií mitochondriálního genomu. Odhad celkového počtu molekul mtDNA činí 10^3 - 10^4 kopií na buňku (Legros et al., 2004; Lightowlers et al., 1997). Díky tomu lze snadno získat mtDNA pro analýzu, zvláště pak při izolaci aDNA z biologických pozůstatků (např. Stone & Stoneking, 1999).

Stav, kdy jedinec nese všechny molekuly mtDNA stejné, se nazývá homoplasmie. Celková homogenita mtDNA jedince je udržována mechanismem zúženého hrdla lahve (bottleneck) v průběhu oogeneze. Z celkového množství se na konečném počtu kopií mtDNA ve zralém oocytu uplatňuje náhodně generovaná část několika desítek molekul (Brown et al., 2001). Je však běžné, že je člověk nositelem více typů mtDNA důsledkem somatických mutací během života či přenosu mutací vzniklých v zárodečných buňkách. Tento stav se označuje jako heteroplasmie. Tully et al. (2000) zaznameli frekvenci výskytu heteroplasmie přibližně 14% u souboru 253 náhodně vybraných jedinců. Existují tedy různé úrovně výskytu molekul mtDNA – mitochondrie, buňka, tkáň, jedinec a populace.

1.1.3 Dědičnost

Ve zralém oocytu je přibližně 100000 – 400000 mitochondrií, každá pravděpodobně s jednou kopií mtDNA (Cummins, 2001). V průběhu oogeneze jsou buňky aerobně metabolicky neaktivní, což přispívá k ochraně proti vzniku nových mutací. Spolu s efektem zúženého hrdla lahve jde o další mechanismus, který přispívá ke stavu homoplasmie u budoucího jedince (souhrnně Cummins, 2001). Menší počet mitochondrií obsahuje také spermie (St John et al., 2000), odhady počtu kopií mtDNA se však velmi různí (od průměrně 1,4 kopie až po počet 1000-1200) (srovnej Diez-Sanchez et al., 2003; May-Panloup et al., 2003). Mitochondrie jsou soustředěny do středního segmentu bičíku a poskytují energii pro jeho pohyb (jsou tedy energeticky aktivní) (Alberts et al., 1994). Během fertilizace přináší spermie do oocytu nejen genetickou informaci jaderného genomu, ale také mitochondrie s mtDNA. Tyto otcovské mitochondrie jsou však krátce po oplodnění terčem destruktivních mechanismů - jsou označeny ubiquitinací a následně eliminovány (Sutovsky et al., 2000; Thompson et al., 2003).

Striktně mateřská dědičnost mtDNA byla považována za pravidlo (Giles et al., 1980) a jde o jedno z východisek pro studium mtDNA. V roce 2002 byl však publikován případ pacienta, jehož mtDNA ve svalech byla převážně otcovského původu (Schwartz & Vissing, 2002). Mitochondrie od otce tedy mohou uniknout destruktivním mechanismům, což narušuje představu o striktně mateřské dědičnosti mtDNA. Zbývá však zodpovědět otázku, jak často k takovým případům dochází a jaký je eventuální dopad na fylogenetické a populačně genetické analýzy. Zdá se, že prezentovaný případ představuje vzácný jev. Další výzkumy zatím nedokládají příklady otcovského původu (např. Filosto et al., 2003; Giles et al., 1980; Howell et al., 2003; Taylor et al., 2003) a mtDNA je stále nahlížena jako lokus s maternálním typem dědičnosti (Schwartz & Vissing, 2003).

1.1.4 Rekombinace

S maternální dědičností souvisí další důležitý předpoklad – v mitochondriálním genomu nedochází k rekombinacím (Birky, 1995). Následkem toho odráží celá molekula mtDNA jednotnou historii (na rozdíl od rekombinantních úseků nDNA, které jsou složeny z haplotypových bloků s různou historií – Pääbo, 2003). Některými autory byla však navržena možnost, že mtDNA může procházet rekombinacemi. Tyto závěry původně vycházely z nepřímých dokladů založených (1) na zjištění, že různé linie mtDNA vykazují mutace na stejných místech (Eyre-Walker et al., 1999) a (2) na sledování vazebné nerovnováhy (linkage disequilibrium, LD) (Awadalla et al., 1999).

Ve fylogenetických schématech se objevují homoplasie – stejné mutace u oddělených linií. Jedno z možných vysvětlení rozšíření shodných mutací je důsledek rekombinace mezi mtDNA liniemi (Eyre-Walker et al., 1999). V rámci molekuly mtDNA však existují místa, ve kterých dochází častěji k mutacím – mutační hot-spots. Homoplasie byly jinými autory připsány právě vzniku paralelních mutací v těchto místech (Howell & Smejkal, 2000; Stoneking, 2000).

Geny, které se přenášejí do další generace společně, jsou ve vazbě; úplně vázané geny jsou nerekombinovatelné. U rekombinantní DNA závisí pravděpodobnost přenosu dvou genů společně na jejich vzdálenosti – se zvětšující se vzdáleností mezi geny klesá vazebná rovnováha. Podle Awadalla et al. (1999) vykazuje mtDNA – stejně jako rekombinantní DNA – negativní závislost mezi vazebnou nerovnováhou a vzdáleností mezi úseky.

Interpretace nepřímých dokladů rekombinace mtDNA je předmětem diskuze a některé statistické metody a fylogenetická východiska uváděných studií se stala terčem kritiky (srovnej Awadalla et al., 2000; Hagelberg, 2003; Jorde & Bamshad, 2000; Kivisild & Villems, 2000; Kumar et al., 2000; Macaulay et al., 1999b; Parsons & Irwin, 2000). Další studie neprokázaly důkazy o rekombinaci mtDNA (Elson et al., 2001; Ingman et al., 2000; Piganeau & Eyre-Walker, 2004).

Splňují mitochondrie a mtDNA předpoklady pro rekombinaci? Zjednodušeně řečeno, aby mohlo docházet k rekombinaci mtDNA, musí (1) existovat schopnost rekombinace a (2) možnost mtDNA molekul si vyměňovat části. Bylo prokázáno, že savčí mitochondrie vykazují potřebný aparát – obsahují aktivní rekombinázu, která se zřejmě podílí na opravných mechanismech (Thyagarajan et al., 1996). Pro splnění druhé podmínky je nezbytný kontakt dvou molekul z rozdílných zdrojů. Je známo, že mitochondrie jsou schopny se spojovat, vytvářet retikulace a také se rozdělit (Scheffler, 2000). Zdá se také, že může docházet k fúzi a výměně obsahu mitochondrií (Legros et al., 2002; Legros et al., 2004; Ono et al., 2001). Autoři studie Enríquez et al. (2000) však docházejí k závěru, že jde spíše o výjimečné případy (Enríquez et al., 2000).

Dalším předpokladem vlivu rekombinací je interakce mtDNA molekul rozdílného původu. Důležité zjištění přineslo sledování dosud jediného známého případu jedince s otcovskými mitochondriemi (viz výše, Schwartz & Vissing, 2002). Ukázalo se, že molekuly mtDNA od matky a otce jsou schopny si vyměňovat úseky – jde tedy o přímé potvrzení

schopnosti rekombinace (Kraytsberg et al., 2004; Ladoukakis & Eyre-Walker, 2004). Pokud ovšem považujeme paternální původ mtDNA za vzácný jev, stejně tak vzácná je i role rekombinací mtDNA. Opět tedy zůstává otevřená otázka, nakolik frekventovaný by mohl být výskyt rekombinace lidské mtDNA v populaci a zda a jak by došlo k ovlivnění závěrů populačních a fylogenetických studií (Ladoukakis & Eyre-Walker, 2004).

1.1.5 Mutační rychlost

Původní předpoklad konzervativnosti molekuly mtDNA díky její organizaci (Awise, 2000) byl jednoznačně vyvrácen v roce 1979, kdy Brown, Georgie a Wilson prokázali její rychlou evoluci – přibližně 10krát rychlejší než v případě jaderné DNA (Brown et al., 1979). Tento jev byl následně připsán vysoké mutační rychlosti, pro což existují různá, navzájem se nevylučující vysvětlení – omezení opravných mechanismů, vliv volných radikálů (vedlejších produktů metabolických pochodů), nepřítomnost vazby mtDNA na histony. Ukázalo se také, že mutační rychlost vykazuje poziční heterogenitu - nejvíce mutací vzniká ve dvou hypervariabilních segmentech kontrolního regionu (HVS-I a HVS-II) a v rámci molekuly mtDNA existují rychle mutující místa, tzv. mutační hot spots (Meyer & von Haeseler, 2003; Meyer et al., 1999).

Tato zjištění mají důsledky pro výzkumy genetické diverzity na populační úrovni. Mitochondriální DNA dokáže za kratší časový interval nashromáždit větší množství změn, které mohou vést k diverzifikaci populací. Význam hrají také v možnosti datování relativně nedávného oddělení mitochondriálních linií, vycházející s předpokladu, že dokážeme mutační rychlost odhadnout (srovnej Heyer et al., 2001; Ho et al., 2005; Howell et al., 2003; Tamura & Nei, 1993).

1.1.6 Selekční neutralita

Považujeme-li mtDNA za selekčně neutrální lokus, který nepodléhá přírodnímu výběru na základě okolních podmínek, pak současná struktura genetické diverzity lidských populací bude výsledkem genetického driftu a migrací (pokud necháme stranou možný vliv rekombinací a otcovské dědičnosti). Odhalení této struktury lze poté zpětně použít jako indikátor evoluce člověka a disperze jednotlivých lidských populací. Přestože schopnost mutací mtDNA vyvolat určitá onemocnění a ovlivnit tak fenotyp jedince je známa (např. Wallace, 1994), povětšinou se nepředpokládá systematické ovlivnění mtDNA genofondu populace v důsledku přírodního výběru.

Ukazuje se ale komplexita vztahů mtDNA, fenotypu a prostředí, která zahrnuje somatické i zárodečné mutace a jak kódující, tak nekódující úseky mtDNA. Některé mutace schopné ovlivnit fenotyp jedince (například vzhledem k věku dožití) se zdají být ve vazbě k určitým haploskupinám, což ústí ve funkční odlišnosti mezi haploskupinami (Coskun et al., 2003). Některé vzory v rozdělení genetické diverzity mtDNA, výsledky fylogenetických analýz či poměr synonymních a nesynonymních mutací byly připsány vlivu selekce (Elson et al., 2004; Mishmar et al., 2003; Torroni et al., 2001). Diskutována je například možnost adaptace na klimatické podmínky. Procesy produkce ATP a tepla stojí v mitochondriích proti sobě – nižší efektivita tvorby ATP znamená vyšší tepelný výtěžek a naopak. Varianty z protilehlých pólů této rovnováhy mohou poskytovat nositeli odlišné predispozice pro adaptaci v rozdílných teplotních pásmech (Coskun et al., 2003; Mishmar et al., 2003). Jiné studie však tyto závěry nepotvrdily (Elson et al., 2001; Kivisild et al., 2006) a někteří vědci nabízejí jiná vysvětlení (Kivisild et al., 2006).

Vzhledem ke klíčové roli mitochondrií v energetice buňky by možnost působení selekce na mtDNA nebyla zcela překvapivá. Dopad na závěry současných výzkumů diverzity mtDNA lidských populací lze

těžko odhadovat, neboť by opět záleželo na míře intenzity působení případného přírodního výběru.

1.2 Studium mtDNA

Historie výzkumů mtDNA je ve znamení snahy o zvýšení rozlišovací schopnosti metod (Richards & Macaulay, 2001). Prvotní výzkumy byly založené na štěpení mtDNA několika restrikčními endonukleázami (Brown, 1980; Cann et al., 1987). Objevení polymerázové řetězové reakce (PCR) (Mullis & Faloona, 1987) a metod pro sekvenci umožnilo provádět restrikční analýzu PCR produktů a sekvenování částí kontrolního regionu. Pozornost byla obrácena převážně na studium sekvencí HVS-I, přesto však až systematické použití RFLP dat z kódujícího regionu doplněné o sekvence HVS-I vedlo k odhalování struktury vztahů mezi liniemi a rekonstrukci jejich historie (Macaulay et al., 1999a; Torroni et al., 1996).

Současné výzkumy vrcholí v sekvenování celého mtDNA genomu (Herrnstadt et al., 2002; Ingman & Gyllensten, 2003; Ingman et al., 2000; Kong et al., 2003; Maca-Meyer et al., 2001; Macaulay et al., 2005; Tanaka et al., 2004). Vedle postupného sekvenování celé molekuly mtDNA lze však nahlížet mutace v rámci celého mitochondriálního genomu jinými metodami: např. odhalením (a následnou sekvenací) polymorfních míst pomocí speciální gelové elektroforézy (CSGE) (Finnilä et al., 2000; Finnilä et al., 2001) nebo multiplex SNP analýzou (Brandstätter et al., 2003). Otázkou zůstává, zda sekvenování kontrolního regionu doplněné restrikční analýzou kódujícího regionu poskytuje dostatečnou míru rozlišení, nebo je skutečně potřeba tento postup nahradit nákladnou analýzou celého mtDNA genomu (srovnej Pakendorf & Stoneking, 2005; Torroni et al., 2006).

1.2.1 Přístupy ke studiu evoluční a populační historie

Molekulárně genetické výzkumy současných populací lze rozdělit do dvou hlavních přístupů (Forster et al., 2001; Pakendorf & Stoneking, 2005), které můžeme označit jako 1) fylogenetický a 2) populační přístup. Přestože řada studií využívá částečně oba přístupy (Kivisild et al., 2004; Quintana-Murci et al., 2004; Richards et al., 2002; Rosa et al., 2004; Salas et al., 2002), je důležité zmínit, že jde oddělené směry, které mají vlastní východiska a metodologii.

1.2.1.1 Fylogenetický přístup

Základní bod určující fylogenetický přístup (souhrnně Forster, 2004) je jednotka výzkumu. Tu představuje jedinec, přesněji řečeno jednotlivá sekvence mtDNA. Konkrétní uspořádání genetických markerů na jednom chromosomu jedince se označuje jako haplotyp. V případě mitochondriální DNA jde o kombinaci mutací v konkrétní sekvenci mtDNA. Soubor haplotypů odvozených od jednoho společného předka, které sdílí některé mutace (někdy také geografické rozdělení), tvoří haploskupinu (Torroni et al., 1996). V ideálním případě tak haploskupiny představují monofyletickou skupinu. Jejich rozpoznání a pojmenování však vedlo v některých případech ke sloučení linií, které jednoho společného nesdílejí a parametry monofyletické skupiny nesplňují (Pakendorf & Stoneking, 2005). Tyto problémy klasifikace vyplývají většinou z nízkého stupně molekulárního rozlišení studovaných mtDNA genomů, ale také z dosud ne zcela poznané míry poziční heterogenity v mutační rychlosti.

Za předpokladu mateřské dědičnosti a nerekombinovatelnosti mtDNA jsou všechny současné haplotypy konečnými body nepřerušovaných linií. Všechny tyto linie se navíc někde v minulosti sbíhají do jednoho společného výchozího haplotypu. Na základě rozdílů mezi odvozenými haplotypy, které byly akumulovány během vývoje linií a vedly k jejich diverzifikaci, lze sestavit fylogenetické schéma. Jeho podoba je výsledkem matematických metod založených na principu maximální parsimonie, distančních matic nebo pravděpodobnostním přístupem.

Zvýšená mutační rychlost a výskyt evolučně nestabilních míst (hot spots) může vést ke zpětným mutacím a konvergenci. Díky tomu nelze s jistotou rozlišit původní (ancestrální) a odvozené haplotypy, takže pro jeden soubor dat existuje zpravidla více možných fylogenetických stromů. Častěji se dnes proto využívají fylogenetické sítě zachovávající sporné body, které se projeví vznikem retikulací (Bandelt et al., 1999; Bandelt et al., 1995).

Na základě odhadů mutační rychlosti mtDNA lze datovat oddělování haplotypů a vznik haploskupin (Forster, 2004). Jejich geografickou distribucí se zabývá fylogeografie (Avice, 2000), která do problematiky přispívá prostorovou dimenzí. Důležitým bodem analýz je rekonstrukce a časové a geografické vymezení zakládajících, ancestrálních haplotypů (tzv. founder analysis) (Forster, 2004; Richards et al., 2000).

Tento přístup umožňuje aplikovat fylogenetickou metodologii i na vnitrodruhové úrovni, kde každý haplotyp představuje operační taxonomickou jednotku (OTU). Není třeba předem definovat populaci jakožto jednotku výzkumu, čímž odpadá často problematická typologizace jednotlivých populačních celků sestavovaných často pouze podle etnologických (jazyk/kultura) kritérií (Avice, 2000).

Výhodou fylogenetického přístupu je možnost datování, která umožňuje fylogenetické schéma haploskupin zasadit do časového perspektivy. Problém však může být na straně interpretace výsledků. Vztah mezi fylogenezí haploskupin a minulostí lidských populací nebývá přímočarý. Fylogenetické schéma odráží pouze historii haplotypů. Každá populace obsahuje více haploskupin a oddělení haploskupin neznamena oddělení populací (Barbujani et al., 1998; Carvalho-Silva et al., 2006; Cordaux & Stoneking, 2003; Pakendorf & Stoneking, 2005). Vznik haploskupiny (i více haploskupin) je ale často vázán na populační expanzi, jejíž existence je jinými metodami nepostižitelná. Není bez zajímavosti, že molekulární datování většiny dnes známých

haploskupin spadá do období, kdy zpočátku početně omezené populace lovecko-sběračských komunit demograficky narůstaly (např. Forster, 2004).

1.2.1.2 Populační přístup

Jednotkou výzkumu je populace. V populační genetice ji teoreticky můžeme vymezit jako reprodukční jednotku. Ovšem při sběru dat je třeba postupovat na základě geografického vymezení (ev. etnické příslušnosti či jiných kritérií).

Zjišťované charakteristiky spadají na intrapopulační a interpopulační úroveň. Intrapopulační úroveň je dána popisnými statistkami souboru vypovídající o variabilitě uvnitř zvolené (ovzorkované) populace. Na interpopulační úrovni jde o rozdělení genetické diverzity a populační srovnání. Analýza molekulárního rozptylu (AMOVA) je obdobou analýzy rozptylu (ANOVA), která rozděluje celkovou variabilitu na variabilitu uvnitř skupin a variabilitu mezi skupinami. Vstupními daty mohou být frekvence haploskupin zjištěné u zkoumaných populací, nebo sekvenční rozdíly v souboru vyjádřené průměrným počtem párových rozdílů. Nejpoužívanějším indikátorem pro genetické vzdálenosti mezi populacemi je koeficient inbreedingu F_{ST} původně navržený pro diploidní data jako míra odchylky pozorované od očekávané frekvence heterozygotů v rámci metapopulace a subpopulací. Vzdálenosti mezi populacemi lze vizualizovat vícerozměrnými statistickými metodami, jako je analýza hlavních komponent (Richards et al., 2002) nebo vícerozměrné škálování (Černý et al., 2004).

Výhodou populačního přístupu je možnost odvozování vztahů mezi reálnými lidskými populacemi a nikoli pouze haplotypy. Problémem při snaze rekonstrukce minulých populačních událostí je ale absence časového rámce. Stejně rozdělení genetické diverzity může být výsledkem několika procesů odehrávajících se v různých obdobích (Barbujani & Bertorelle, 2001). Samotné vícerozměrné metody jsou

explorační – umožňují odhalit strukturu v datech nikoli testování hypotéz (Richards et al., 2002).

1.3 Přehled výzkumů mitochondriální DNA

Počáteční výzkumy fylogeneze lidské mitochondriální DNA byly interpretovány jako doklady původu anatomicky moderního člověka v Africe (Cann et al., 1987) v souladu s modelem druhového nahrazení (souhrnně o modelech evoluce anatomicky moderního člověka viz Stringer, 2002). Přes kritickou odezvu (Templeton, 1993; Templeton, 1994) došly recentní studie k podobným výsledkům. Subsaharská Afrika obsahuje nejdiverzifikovanější a nejstarší haploskupiny, od kterých se haplotypy nacházené v celém ostatním světě oddělili později (Ingman et al., 2000; Quintana-Murci et al., 1999). Podle některých autorů však na základě současných dat zahrnujících rovněž jadernou DNA nelze jednoznačně zamítnout platnost dalších modelů původu moderních lidí (např. Eswaran et al., 2005; Garrigan et al., 2005; Plagnol & Wall, 2006; Templeton, 2002). Přesto lze usuzovat na významnou roli Afriky v evoluci a šíření anatomicky moderního člověka (Eswaran et al., 2005; Ray et al., 2005; Relethford & Jorde, 1999) a badatelé se snaží nalézt stopy pravěkých migrací ve struktuře genetické diverzity recentních populací (souhrnně Forster, 2004; Forster & Matsumura, 2005; Torroni et al., 2006).

Nejstarší populační expanze a následné migrace probíhající v rámci Afriky jsou spojené s haploskupinami L0 – L2 (Watson et al., 1997), které jsou specifické pro dnešní populace subsaharské Afriky. Přibližně před 80000 lety došlo k diverzifikaci haploskupiny L3, jejíž dva deriváty – haploskupiny M a N – byly migracemi vyneseny do Eurasie (Maca-Meyer et al., 2001; Quintana-Murci et al., 1999; Watson et al., 1997) a pokrývají dnes veškerou mtDNA diverzitu populací mimo Afriku. Haploskupiny odvozené od haploskupiny M jsou převážně u populací jižní a východní Asie, Austrálie a Ameriky, deriváty N haploskupiny nacházíme v západní Asii a Evropě (souhrnně Forster, 2004; Torroni et al., 2006).

Prvotní migrace lidských populací z Afriky mohly probíhat dvěma cestami: 1) severní přes Sinaj a 2) jižní přes úžinu Báb el-Mandab (Forster & Matsumura, 2005; Lahr & Foley, 1998; Stringer, 2000). Zatímco severní cesta vedoucí po souši se zdá být přirozenější, sporným bodem je osídlení bližší Evropy o několik tisíciletí později, než vzdálenější jihozápadní Asie a Austrálie (ale srovnej s Cordaux & Stoneking, 2003). Současný pohled tudíž uvažuje o poměrně rychlé prvotní migraci skrze jižní cestu – přes Báb el-Mandab a pobřežní trasou přes jih Arabského poloostrova a Indii až do Austrálie (Maca-Meyer et al., 2001; Macaulay et al., 2005; Metspalu et al., 2004).

Za tohoto scénáře hraje oblast Jemenu důležitou roli jako první bod mimo Afriku na jižní migrační trase. Problémovým bodem výzkumů DNA dnešních populací při zacílení na dávné události je však možnost smazání jejich odrazu recentnějšími vlivy. Proto jsou výzkumy genetických stop prvotních migrací převážně zaměřeny na populace určitým způsobem izolované – např. obyvatelé Andamanských ostrovů nebo členy kmene Orang Asli z Malajsie (Forster & Matsumura, 2005; Macaulay et al., 2005; Thangaraj et al., 2005).

Vedle otázky evoluce a šíření anatomicky moderního člověka se výzkumy diverzity mtDNA zaměřují v užším měřítku na jednotlivé regiony či populace. Z pohledu diverzity mitochondriální DNA je však oblast Jemenu terra intacta. Jedinou studií, která se této oblasti systematictěji věnuje je Kivisild et al. (2004). Autoři se zabývají genovým tokem přes úžinu Báb el-Mandab na základě sekvencí mtDNA z Etiopie a Jemenu, který je zastoupen vzorky z Kuvajtu od probandů, kteří podle vlastního tvrzení z Jemenu pocházejí. Studie podává podrobnou fylogeografickou analýzu zjištěných haplotypů. Ovšem vlivem nejistého původu vzorků nelze tato data jednoznačně použít pro stanovení míry a struktury diverzity u populací Jemenu.

1.4 Archeologický a historický kontext

Úžina Báb el-Mandab je široká přibližně 26km. Přesto, že v časovém úseku od evoluce moderního člověka docházelo vlivem klimatických změn ke kolísání hladiny moří, ukazuje se, že nikdy pokles hladiny nevedl k tvorbě pevninského mostu mezi Afrikou a Arabským poloostrovem (Fernandes et al., 2006). Případné migrace podél jižní trasy tak musely zahrnovat plavbu přes Báb el-Mandab (Petraglia & Alsharekh, 2003), nebo postupovat podél pobřeží (Stringer, 2000).

Nejstarší nálezy, které dokládají přítomnost lidí na východním pobřeží Afriky a pravděpodobně i adaptaci na využívání mořských zdrojů, pochází z Eritreji z doby přibližně před 125 tis. lety (Walter et al., 2000). Poptávka po nálezech osídlení Arabského poloostrova z doby šíření anatomicky moderního člověka (Stringer, 2000) zatím nebyla uspokojivě naplněna. Z tohoto období nebyly nalezeny žádné kosterní nálezy ať už v důsledku nedostatečného zájmu badatelů (Petraglia & Alsharekh, 2003), nebo mohou být vlivem mořské transgrese zcela smazány. Výzkumy kamenné industrie Arabského poloostrova vykazují určitý počet nálezů z období středního paleolitu, jsou však nahlíženy jako prozatímní (souhrnně Petraglia & Alsharekh, 2003). Hlavními problémovými okruhy jsou nedostatečné detailní studie technologie výroby nástrojů, typologie a datování (Petraglia & Alsharekh, 2003; Rose, 2004). To brání v pochopení prostorových a časových charakteristik a vztahů k okolním regionům. Středně-paleolitické nálezy z Arabského poloostrova vykazují jistou podobnost ke kamenným nástrojům v severní Africe a Levantě, ale nejde o jasný ekvivalent a na základě současných dat nelze jednoznačně vyvozovat vztahy mezi populacemi (Petraglia & Alsharekh, 2003). Zdá se, že nálezy na Arabském ostrově mohou vypovídat o více migračních událostech zahrnujících obě cesty (severní i jižní); ale není vyloučen ani vlastní technologický vývoj vlivem adaptace na aktuální podmínky (souhrnně např. Petraglia & Alsharekh, 2003). Důležité nálezy pro doklad migrace přes Báb el-Mandab mohou do budoucna nabízet

lokality nejen v blízkosti úžiny samotné, ale i podél pobřeží Rudého a Arabského moře (souhrnně např. Petraglia & Alsharekh, 2003); nově také nálezy z Ománu, které navíc vykazují technologickou podobnost k východní Africe, ale zůstávají zatím nedatované (Rose, 2004).

Oblast jihu Arabského poloostrova hrála v recentnějších obdobích roli jako místo kontaktů mezi různými oblastmi. Můžeme ji zařadit do dvou obchodních okruhů. První zahrnuje povodí Nilu a Rudé moře a spojuje mimo jiné Egypt, východní Afriku a jižní část Arabského poloostrova. Druhý (afro-arabský okruh) spojuje východní Afriku s Indií (Fattovich, 2002).

Do období 7. až 4. tisíciletí př. nl. je kladen začátek obchodních cest s obsidiánem spojujících východní Súdán a pobřeží Eritreji přes Rudé moře s oblastí Tihamy v Jemenu. V polovině druhého tisíciletí př. nl. vznikl tzv. tihamský kulturní komplex zahrnující pobřeží Rudého moře na africké i arabské straně, který je charakterizován výskytem podobné keramiky. Arabové z jihu Arabského poloostrova pak ovládli obchodní cesty přes Rudé moře především po vzniku Sábského království kolem roku 800 př. nl., které ovládalo nejen západní pobřeží dnešního Jemenu, ale také oblasti Etiopie a Eritreji. Naproti tomu etiopská Aksúmská říše hrála dominantní úlohu v 1. tisíciletí nl. a v jeho polovině ovládala dnešní Jemen (souhrnně např. Fattovich, 2002). Zde byla vystřídána perskou nadvládou, která končí příchodem Islámu v 6. století.

Přestože výše nastíněné souvislosti bývají užity jako pozadí pro studie zabývající se genovým tokem mezi Afrikou a Arabským poloostrovem (Kivisild et al., 2004; Richards et al., 2003), je důležité upozornit, že navržené kulturní vztahy neznamenaají nutně také lidské migrace (Fattovich, osobní sdělení).

1.5 Cíle práce

Jemen hraje významnou úlohu pro migrace člověka jakožto důležitý bod na cestě přes úžinu Báb el-Mandab spojující východní Afriku s jihem Arabského poloostrova. Právě tudy mohla probíhat disperze anatomicky moderního člověka z Afriky do Eurasie, ale také pozdější kontakty mezi populacemi. Migrace a vztahy mezi populacemi jsou stále častěji nahlíženy prostřednictvím molekulárně genetických dat. Přesto genetická diverzita populací v oblasti Jemenu nebyla dosud systematicky studována. Jedinými příklady jsou studie zacílené na otázku genového toku mezi africkými populacemi a populacemi z jižní části Arabského poloostrova (Kivisild et al., 2004; Richards et al., 2003).

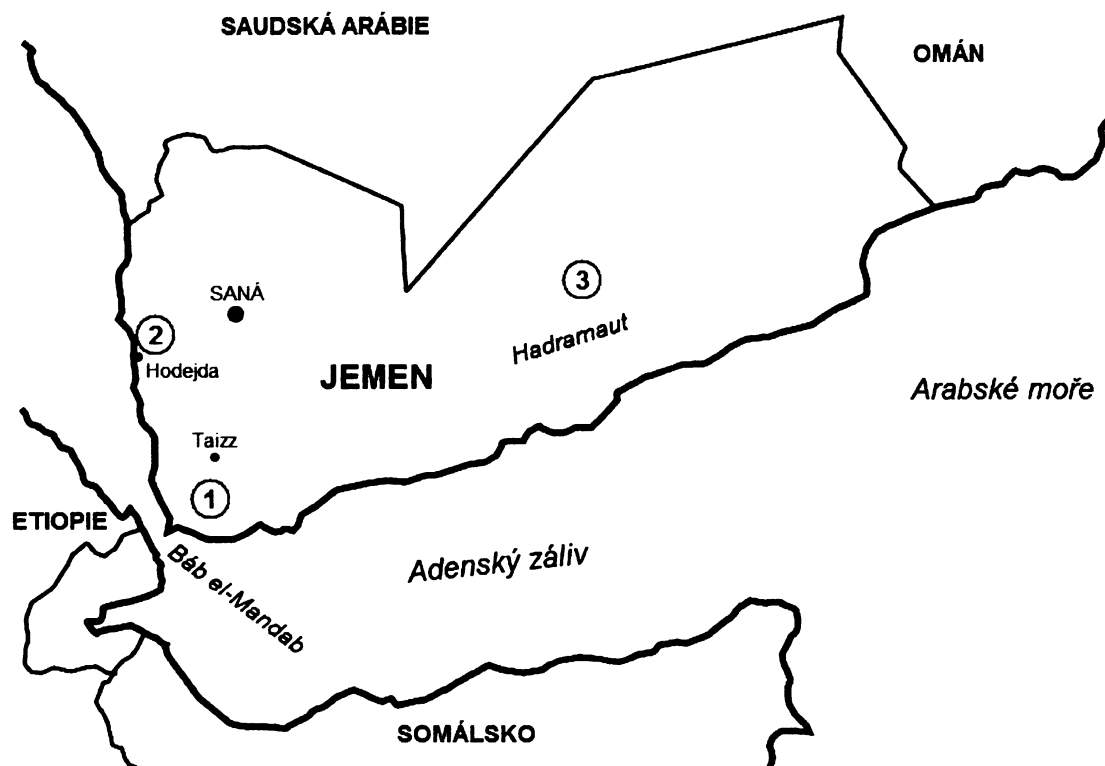
Cílem této diplomové práce je prezentovat nové sekvence HVS-I mtDNA ze tří oblastí Jemenu, na jejich základě popsat genetickou diverzitu souborů a porovnat je s populacemi, jejichž mtDNA již byla studována jinými autory. Ve středu zájmu jsou především dva okruhy. První se týká vztahů mezi populacemi na obou březích úžiny Báb el-Mandab. Jaký význam hrála úžina jakožto jedna z migračních cest mezi Afrikou a Asií? Jaké populace a regiony tyto migrace zahrnují? A jak a zda se promítají do genetické diverzity současných obyvatel? Druhý bod se zaměřuje na vztah populací podél pobřežní cesty od Etiopie až po Indický kontinent.

2 MATERIÁL A METODY

2.1 Soubory

Biologické vzorky (bukální stěry) byly získány během terénního působení Viktora Černého v oblasti Jemenu v listopadu a prosinci 2005. Při výběru probandů byly vyloučeni maternálně příbuzní jedinci a samotné vzorkování probíhalo na geografickém základě (viz Černý et al., 2004). Výsledkem jsou tři soubory ze třech lokalit: 1) JAC – 43 jedinců z oblasti jižních svahů centrálního masivu z vesnice Al-Achkúm asi 40km

jihovýchodně od města Taiz, 2) JTI – druhý vzorek 67 jedinců byl získán v několika různých vesnicích příbřežní roviny zvané Tihama poblíž Rudého moře a na západních svazích centrálního masívu v blízkosti města Hudejda, 3) JHA – 40 jedinců z Wádí Hadramaut na východě Jemenu (obr. 1). Všechny tři soubory byly vyhodnocovány samostatně.



Obrázek 1. Přibližná poloha vzorkovaných souborů v Jemenu: 1 – JAC (Al-Achkúm), 2 – JTI (Tihama), 3 – JHA (Hadramaut).

Kromě námi získaných sekvencí byla pro populační srovnání použita data z publikovaných studií. Kritéria výběru dat byla: 1) sekvence zahrnující úsek 16030 – 10370, 2) dostupné informace o geografickém původu vzorků a 3) počet nejméně 20 jedinců. Výběr sekvencí a tvorba souborů vycházela z geografické příslušnosti vzorků. U některých populací bylo zachováno etnonymum pro vymezení zeměpisné polohy. Z oblasti Arabského poloostrova byl vybrán soubor 29 Beduínů pocházejících z centrální a západní části Saudské Arábie získaných ve městě Džidda (DiRienzo & Wilson, 1991). Dále byly použity čtyři populace z Etiopie a Eritreji, které vůči Jemenu leží na protilehlém břehu úžiny Báb el-Mandab. Jde o 53 Tigrájců, 120 Amharů, 33 Oromů a 21

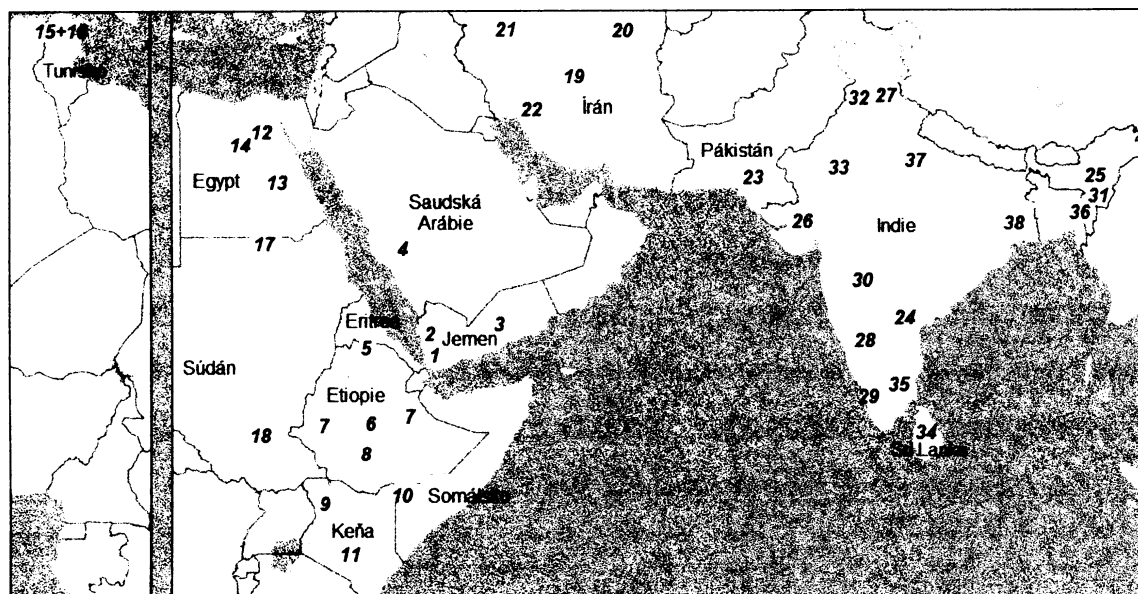
Guragů analyzovaných ve studii Kivisild et al. (2004). Z Keňi z východní Afriky pochází dále 37 Turkanů, 27 Somálců a 25 Kikuyů (Watson et al., 1997). Povodí Nilu představuje 78 Núbijců a 65 nilotů (konglomerát více etnik hovořící jazyky z nilotské větve nilosaharské rodiny, sestavený na základě geografické polohy) (Krings et al., 1999). Severovýchodní Afrika zahrnuje 59 obyvatel dolního Egypta a 35 horního Egypta (Krings et al., 1999), 35 sekvencí z oázy el-Hajez (Žaloudková & Černý, 2006) a 47 resp. 50 jedinců z vesnic Kesra a Zriba z Tuniska (Cherni et al., 2005). Přední Východ zahrnuje celkem 430 vzorků z Íránu pocházejících z centrální oblasti (36 sekvencí), severovýchodu (30 sekvencí), severozápadu (226 sekvencí) a jihozápadu země (138 sekvencí) (Metspalu et al., 2004). Indický subkontinent představuje 100 sekvencí z Karáči v Pákistánu (Quintana-Murci et al., 2004) a vzorky z Indie zahrnující různá etnika studované různými autory. Tyto vzorky byly sloučeny do souborů podle regionů: Ándhrapradéš – 243 vzorků (Bamshad et al., 1998); Ásám – 45 jedinců (Cordaux et al., 2003); Gudžarát – 53 jedinců (Metspalu et al., 2004); Himáčalpradéš – 37 sekvencí (Metspalu et al., 2004); Karnataka – 89 (Mountain et al., 1995) a 106 jedinců (Cordaux et al., 2003); Kérala – 100 vzorků (Metspalu et al., 2004); Maháráštra – 113 sekvencí (Metspalu et al., 2004); Nágáland – 43 vzorků (Cordaux et al., 2003); Paňdžáb – 65 (Kivisild et al., 1999), 40 (Cordaux et al., 2003) a 109 jedinců (Metspalu et al., 2004); Rádžasthán – 35 jedinců (Metspalu et al., 2004); Srí Lanka – 132 vzorků (Metspalu et al., 2004); Tamilnádu – 49 (Roychoudhury et al., 2001) a 99 jedinců (Cordaux et al., 2003); Tripura – 20 vzorků (Roychoudhury et al., 2001); Uttarpradéš – 103 (Kivisild et al., 1999) a 56 jedinců (Metspalu et al., 2004); Západní Bengálsko – 161 (Metspalu et al., 2004) a 34 vzorků (Roychoudhury et al., 2001).

Kompletní seznam souborů poskytuje tabulka 1, pro přibližnou geografickou polohu viz obrázek 2.

Tabulka 1. Seznam 38 souborů použitých pro genetické srovnání populací

Soubor	Stát	Region/ Populace ^a	Jazyková rodina	<i>n</i>	Reference
1 JAC	Jemen	Al-Achkúm	afroasijská	43	tato studie
2 JTI	Jemen	Tihama	afroasijská	67	tato studie
3 JHA	Jemen	Hadramaut	afroasijská	40	tato studie
4 Bd	Saudská Arábie	Beduíni	afroasijská	29	DiRienzo & Wilson (1991)
5 Tig	Etiopie/ Eritrea	Tigrajci	afroasijská	53	Kivisild et al. (2004)
6 Amh	Etiopie	Amharové	afroasijská	120	Kivisild et al. (2004)
7 Oro	Etiopie	Oromové	afroasijská	33	Kivisild et al. (2004)
8 Gur	Etiopie	Guragové	afroasijská	21	Kivisild et al. (2004)
9 Tur	Keňa	Turkanové	nilosaharská	37	Watson et al. (1997)
10 Som	Keňa	Somálci	afroasijská	27	Watson et al. (1997)
11 Kik	Keňa	Kikujové	nigerokonžská	25	Watson et al. (1997)
12 EgD	Egypt	dolní Egypt	afroasijská	59	Krings et al. (1999)
13 EgH	Egypt	horní Egypt	afroasijská	35	Krings et al. (1999)
14 ElH	Egypt	el-Hajez	afroasijská	35	Žaloudková & Černý (2006)
15 Kes	Tunisko	Kesra	afroasijská	47	Cherni et al. (2005)
16 Zri	Tunisko	Zriba	afroasijská	50	Cherni et al. (2005)
17 Nub	Súdán/ Egypt	Núbijci	nilosaharská	78	Krings et al. (1999)
18 Nil	Súdán	niloté	nilosaharská	65	Krings et al. (1999)
19 IS	Írán	střed	indoevropská	36	Metspalu et al. (2004)
20 ISV	Írán	severovýchod	indoevropská	30	Metspalu et al. (2004)
21 ISZ	Írán	severozápad	indoevropská	226	Metspalu et al. (2004)
22 IJZ	Írán	jihozápad	indoevropská	138	Metspalu et al. (2004)
23 Kar	Pákistán	Karáči	indoevropská	100	Quintana-Murci et al. (2004)
24 And	Indie	Ándhrapradéš	drávidská	243	Bamshad et al. (1998)
25 Asm	Indie	Ásám	sinotibetská	45	Cordaux et al. (2003)
26 Gud	Indie	Gudžarát	indoevropská	53	Metspalu et al. (2004)
27 Him	Indie	Himáčalpradéš	sinotibetská	37	Metspalu et al. (2004)
28 Kar	Indie	Karnataka	drávidská	195	Mountain et al. (1995); Cordaux et al. (2003)
29 Ker	Indie	Kerala	drávidská	100	Metspalu et al. (2004)
30 Mah	Indie	Maharáštra	indoevropská	113	Metspalu et al. (2004)
31 Nag	Indie	Nágáland	sinotibetská	43	Cordaux et al. (2003)
32 Pan	Indie	Paňdžáb	indoevropská	214	Kivisild et al. (1999); Cordaux et al. (2003); Metspalu et al. (2004)
33 Rad	Indie	Rádžasthán	indoevropská	35	Metspalu et al. (2004)
34 SrL	Indie	Srí Lanka	indoevropská/ drávidská	132	Metspalu et al. (2004)
35 Tam	Indie	Tamilnádu	drávidská	148	Cordaux et al. (2003); Roychoudhury et al. (2001)
36 Tri	Indie	Tripura	sinotibetská	20	Roychoudhury et al. (2001)
37 Utr	Indie	Uttarpradéš	indoevropská	159	Kivisild et al. (1999); Metspalu et al. (2004)
38 ZBe	Indie	Západní Bengálsko	indoevropská/ austroasijská	195	Metspalu et al. (2004); Roychoudhury et al. (2001)

^a Udává region, název státu, města nebo vesnice kde probíhalo vzorkování; u některých populací byl zachován etnonym pro určení přibližné polohy souboru



Obrázek 2. Přibližná poloha všech 38 souborů (čísla odpovídají tabulce 1).

2.2 Laboratorní postupy

Laboratorní zpracování vzorků zahrnuje publikované metody (Černý et al., 2006; Černý et al., 2004). Stěrky byly uchovány v 800 μ l lyzačního roztoku (50mM Tris pH8.0, 50mM EDTA pH8.0, 50mM sacharózy, 100mM NaCl, 1%SDS), přičemž extrakce byla provedena z 200 μ l prostřednictvím kitu QIAquick PCR Purification (QIAGEN). Principem izolace je denaturování proteinových struktur chaotropní soli a následné využití kolonek se silikátovou membránou, která váže DNA, zatímco ostatní nečistoty kolonkou prochází. Adsorpce DNA membránou a opětovné uvolnění je řízeno koncentrací solí a hodnotou pH a zajišťováno PB respektive EB pufry. Kompletní protokol postupu extrakce DNA poskytuje příloha 1.

Metodou PCR byl amplifikován úsek dlouhý 440bp v rozmezí pozic 15971 a 16410 podle CRS (Anderson et al., 1981). Sekvence použitých primerů jsou: F-15971 (5'-TTA ACT CCA CCA TTA GCA CC-3') a R-16410 (5'-GAG GAT GGT GGT CAA GGG AC-3'). Celkový objem na reakci činil 25 μ l (3 μ l vzorku a 22 μ l mastermixu). PCR probíhala v teplotních cyklech 1x (94°C – 5 minut), 35x (94°C – 50 sekund; 51°C –

– 30 sekund; 72°C – 45 sekund) a 1x (72°C – 5 minut). Při každé PCR reakci byly provedeny negativní kontroly.

PCR produkty byly přečištěny pomocí kolonek QIAquick PCR Purification Kit (QIAGEN) (viz příloha 2). Pro kvantifikování PCR produktu a hodnocení negativní kontroly byly 2 μ l produktu použity do gelové elektroforézy. Hmotnost DNA byla odečtena podle hmotnostního markeru MassRuler™ DNA Ladder, Low Range (Fermentas) na 1% agarózovém gelu. Výše uvedené zpracování vzorků bylo provedeno v archeogenetické laboratoři Archeologického ústavu AVČR a kromě mne se na něm podíleli dr. Černý a Mgr. Žaloudková.

Sekvenování proběhlo komerční cestou v Centru servisních laboratoří biologické sekce UK v Praze (PřF UK – biologie, <http://www.natur.cuni.cz/~sedinova/>). Pro sekvenační reakci byl použit BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) a forward primer F-15971 pro amplifikaci L-řetězce. Reakce probíhala na ABI Prism 3100 Avant Genetic Analyzer (Applied Biosystems). Sekvence s délkovou heteroplasmii (vliv homopolymerického cytozinového úseku mezi pozicemi 16184 a 16193) byly sekvenovány i reversním primerem R-16410. Délkové heteroplasmie nebyly do dalších analýz zahrnuty.

2.3 Analýza dat

Vizuální prospekce chromatogramů byla provedena v programu Chromas Lite 2.0 (Technelysium Pty Ltd 1998-2004). Pomocí softwaru BioEdit Sequence Alignment Editor, verze 7.0.4.1 (Hall, 1999) proběhlo zarovnání sekvencí podle referenční sekvence CRS (Anderson et al., 1981). Do dalších analýz byl použit úsek dlouhý 341bp mezi pozicemi 16030 a 16370 (podle Černý et al., 2004). Sekvence byly následně exportovány do textového editoru TextPad 4.6.2 (Hélios Software Solutions 1992-2004).

2.3.1 Klasifikace haplotypů

Přiřazení sekvencí do haploskupin proběhlo na základě HVS-I mutací podle publikovaných motivů (<http://www.stats.gla.ac.uk/~vincent/founder2000/motif.html>; Kivisild et al., 2004; Salas et al., 2002). Názvy haploskupin byly aktualizovány podle recentních úprav: haploskupina L0a představuje dřívější L1a (Salas et al., 2004); haploskupiny R0 a R0a jsou dřívější preHV respektive (preHV)1 (Torroni et al., 2006).

2.3.2 Statistické metody

Pro populačně genetické vyhodnocení dat byl použit software Arlequin, verze 2.00 (Schneider et al., 2000). Popisná statistika na intrapopulační úrovni zahrnuje standardní míry diverzity. Počítána byla genová diverzita H_S , tedy pravděpodobnost náhodného vybraní dvou rozdílných haplotypů ze souboru (podle Nei, 1987). Dále nukleotidová diverzita π udávající pravděpodobnost, že dva náhodně vybrané homologní nukleotidy jsou odlišné (Nei, 1987; Tajima, 1983). Výpočet průměrného počtu párových rozdílů D_{ii} a rozptylu $V(D_{ii})$ vychází z Tajima, (1983) a Tajima (1993). Demografická expanze populace se promítá do rozložení počtu párových rozdílů (tzv. mismatch rozložení). Odchyly od normálního rozložení párových rozdílů byly testovány pomocí Harpendingova r indexu (raggedness index), který vykazuje nižší hodnoty pro expandující populace (Harpending, 1994). Jako testy neutrality byly počítány statistiky D (Tajima, 1989) a F_S (Fu, 1997). Tyto zároveň indikují populační expanzi a byly použity pro testování nulové hypotézy demograficky konstantního vývoje populace (za předpokladu selekční neutrality a rovnováhy mezi vznikem mutací a genetickým driftem).

Genetické vzdálenosti mezi námi studovanými populacemi Jemenu a publikovanými daty byly hodnoceny koeficientem F_{ST} . Metoda výpočtu vychází z párových rozdílů (Weir & Cockerham, 1984). Nulová hypotéza předpokládající žádné rozdíly mezi populacemi byla testována na základě

1000 permutací. Výsledná matice vzdáleností F_{ST} představující mnohorozměrná data byla vizualizována ve dvojrozměrném prostoru metodou vícerozměrného škálování v programu STATISTICA, verze 5.5 (StatSoft, Inc. 2000).

Analýza molekulárního rozptylu (AMOVA) (Excoffier et al., 1992) byla aplikována pro analýzu genetické struktury populací a rozložení diverzity podle geografických, lingvistických a dalších kritérií. Shlukování populací do skupin vycházelo prvotně z otázky vztahů mezi populacemi sousedícími přes úžinu Bab el-Mandab a podél jihozápadního pobřeží Asie. V dalších krocích pak podle výsledků genetických vzdáleností mezi populacemi (viz dále).

3 VÝSLEDKY

V rámci prezentované práce jsme získali 150 nových sekvencí prvního hypervariabilního segmentu mtDNA ze třech oblastí Jemenu. Celkem 15 jedinců vykazuje délkovou heteroplasmii díky homopolymerickému cytosinovému úseku mezi pozicemi 16184 a 16193. Dva jedinci jsou nositeli bodové heteroplasmie (JAC36 na pozici 16145 a JTI99 na 16355). Zaznamenána byla jedna inserce (16170iC u jedince JHA121). Tato mutace nebyla použita v dalších analýzách. Seznam mutací v rozsahu 341bp dlouhého úseku HVS-I je prezentován v tabulce 2 spolu s předběžným zařazením sekvencí do haploskupin. Mutace HVS-I neposkytují dostatečné rozlišení pro 100% klasifikaci haplotypů. Například autoři Santos et al. (2004) dokládají přibližně 50% úspěšnost klasifikace sekvencí podle HVS-I motivu. Procento haplotypů, které by bylo možné s jistotou přiřadit do haploskupin, nebylo v této práci stanoveno. Předběžně navržená klasifikace vychází z maximální shody mezi pozorovanými a publikovanými motivy. Laboratorní zpracování vzorků ve smyslu haploskupinové klasifikace (štěpení RFLP) nadále v archeogenetické laboratoři Archeologického ústavu AVČR probíhá.

Tabulka 2. Seznam mutací v 341bp dlouhém úseku HVS-I, předběžné zařazení do haploskupin a výskyt délkové heteroplasmie

Vzorek	HVS-I motiv	Haploskupina	Htp ^a
JAC1	m 069 126 145 261	J1b	0
JAC2	m 223 265T	L3e3	0
JAC3	m 066 129 182C 183C 189 234	L1c3	1
JAC4	m 069 126 145 222 261 362	J1b	0
JAC5	m 126 355 362	R0a	0
JAC6	m 183C 189 223 278	L?	1
JAC7	m 069 086 126 193 309	J1c	0
JAC8	m 126 163 186 189 294	T1	0
JAC9	m 093 126 189 294 296	?	1
JAC10	m 126 362	R0	0
JAC11	m 176 224 311	L6 nebo K	0
JAC12	m 126 294 296	T	0
JAC13	m 179 223 256G 284 311	L3h	0
JAC14	m 176 224 311	L6 nebo K	0
JAC15	m 093 129 182C 183C 189 223 240 249 331	M1a1	1
JAC16	m 126 355 362	R0a	0
JAC17	m 126 355 362	R0a	0
JAC18	m 126 163 186 189 294	T1	0
JAC19	m 093 126 355 362	R0a	0
JAC20	m 066 129 183C 189 234 358G	L1c	1
JAC21	m 126 355 362	R0a	0
JAC22	m 069 126 145 261	J1b	0
JAC23	m 189 193 189 193 278 294 309	L2a1	1
JAC24	m 069 126 193 300 309	J1c	0
JAC25	m 184 223 311	L nebo M	0
JAC26	m 069 126	J	0
JAC27	m 069 126 193 300	J1c	0
JAC28	m 126 355 362	R0a	0
JAC29	m 176 224 311	L6 nebo K	0
JAC30	m 114A 126 294	T	0
JAC31	m 126 355 362	R0a	0
JAC32	m 126 355 362	R0a	0
JAC33	m 176 224 311	L6 nebo K	0
JAC34	m 126 355 362	R0a	0
JAC35	m 071 129 145 187 189 213 223 234 265C 278 286G 294 311 360	L1c2	0
JAC36	m 067 145 278 362	HV1	0

Pokračování na další straně

Tabulka 2. Pokračování.

Vzorek	HVS-I motiv	Haploskupina	Htp ^a
JAC37	m 114A 126 294	T	0
JAC38	m 129 183C 189 223 249 311 359	M1a	1
JAC39	m 112G 126 355 362	R0a	0
JAC40	m 069 126 145 261	J1b	0
JAC41	m 114A 129 145 213 223 278 294 311	L	0
JAC42	m 223 265T	L3e3	0
JAC43	m 114A 129 145 213 223 278 311	L2b	0
JTI44	m 069 126 193 300	J1c	0
JTI45	m 223 265T	L3e3	0
JTI46	m 071	R2	0
JTI47	m 172 223 269 301 362	L4g	0
JTI48	m 223 265T	L3e3	0
JTI49	m 069 126 145 261	J1b	0
JTI50	m 093 224 311	K	0
JTI51	m 069 126 145 261	J1b	0
JTI52	m 148 172 187 188G 189 223 230 311 320	L0a2	0
JTI53	m 223 234 249 278 294 295	L2a	0
JTI54	m 126 288 292 294 296 311	T2	0
JTI55	m 071	R2	0
JTI56	m 224 311	K	0
JTI57	m 126 288 292 294 296 311	T3	0
JTI58	m 093 224 311	K	0
JTI59	m 172 223 269 301 362	L4g	0
JTI60	m 126 362	R0	0
JTI61	m 067 086 183 294 327A	?	0
JTI62	m 129 189 192 223 278 294 309	L2a1b2	0
JTI63	m 093 124 223 278 362	L3b	0
JTI64	m 182C 183C 189 223 278 290 294 309	L2a1b	1
JTI65	m 209 316	?	0
JTI66	m 093 124 223 278 362	L3b	0
JTI67	m 189 207 309 318C	L2a1b	1
JTI68	m 086 239	?	0
JTI69	m 153	?	0
JTI70	m 126 362	R0	0
JTI71	m 086 239	?	0
JTI72	m 153	?	0
JTI73	m 069 126 193 309	J1c	0

Pokračování na další straně

Tabulka 2. Pokračování.

Vzorek	HVS-I motiv	Haploskupina	Htp ^a
JTI74	m 069 126 193 310	J	0
JTI75	m 069 126 193 309	J1c	0
JTI76	m 354	?	0
JTI77	m 069 126 193 309	J1c	0
JTI78	m 069 126 193 309	J1c	0
JTI79	m 069 126	J	0
JTI80	m 069 126 193 309	J1c	0
JTI81	m 069 126	J	0
JTI82	m 172 223 311 362	L?	0
JTI83	m 148 192 223 274 311	L?	0
JTI84	m 069 126 145 222 261 311	J1b	0
JTI85	m 209 223 292 311 327	L3f1	0
JTI86	m 354	?	0
JTI87	m 126 286	?	0
JTI88	m 126 362	R0	0
JTI89	m 179 183C 189 223 294	L1c3 nebo L2a1	1
JTI90	m 209 223 224	?	0
JTI91	m 069 126 145 261	J1b	0
JTI92	m 069 126 145 261	J1b	0
JTI93	m 069 126 136 145 221 222 261	J1b	0
JTI94	m 069 126 145 261	J1b	0
JTI95	m 126 355 362	R0a	0
JTI96	m 223 292	L3e3	0
JTI97	m 069 126 193 287 300 309	J1c	0
JTI98	m 148 192 223 274 311	L?	0
JTI99	m 148 172 187 188G 189 223 230 311 320 355	L0a	0
JTI100	m 069 126	J	0
JHA101	m 182C 183C 189 223 278 290 294 309	L2a1b	1
JHA102	m 189 223 278	X	1
JHA103	m 069 126 231 319	J	0
JHA104	m 124 223 319	L3d1	0
JHA105	m 069 126 193 300 309	J1c	0
JHA106	m 172 187 189 223 265 278 294 309	L2a1b2	0
JHA107	m 126 223 290	?	0
JHA108	m 129 148 168 172 187 188G 189 223 230 278 293 311 320	L0a1	0
JHA109	m 093 209 223 292 311	L3f1	0
JHA110	m 069 126 231 319	J	0

Pokračování na další straně

Tabulka 2. Pokračování.

Vzorek	HVS-I motiv	Haploskupina	Htp ^a
JHA111	m 256 270 293	L nebo U	0
JHA112	m 067 274	HV1	0
JHA113	m 126 207 362	R0	0
JHA114	m 147G 172 223 248 355	N1a	0
JHA115	m 129 148 168 172 187 188G 189 223 230 278 293 309 311 320	L0a1	0
JHA116	m 124 223 319	L3d1	0
JHA117	m 265C	?	0
JHA118	m 124 223 300 319	L3d1	0
JHA119	m 172 183C 189 223 320	L3e2b	1
JHA120	m 148 172 187 188G 189 223 230 311 320	L0a2	0
JHA121	m 170iC 183C 189 223 274 311 319 320	L3e2b	1
JHA122	m 148 172 187 188G 189 223 230 311 320	L0a2	0
JHA123	m 148 172 187 188G 189 223 230 311 320	L0a2	0
JHA124	m 129 172 173 188A 189 223 256 278 293 294 311 360 368	L1c1	0
JHA125	m 185 209 223 327	L3e1a	0
JHA126	m 148 172 187 188G 189 209 223 230 311 320	L0a	0
JHA127	m 148 172 187 188G 189 223 230 311 320	L0a2	0
JHA128	m 124 223 278 362	L3b	0
JHA129	m 092	?	0
JHA130	m 183C 189 249 265	?	1
JHA131	m 209 223 292 311	L3f1	0
JHA132	m 148 172 187 188G 189 223 230 311 320	L0a2	0
JHA133	m 092	?	0
JHA134	m 126 153 233C 257 294 325	T5	0
JHA135	m 223 260 270 287 311	L nebo M	0
JHA136	m 185 209 223 327	L3e1a	0
JHA137	m 148 172 187 188G 189 223 230 311 320	L0a2	0
JHA138	m 129 172 173 188A 189 223 256 278 293 294 311 360 368	L1c1	0
JHA139	m 185 209 223 327	L3e1a	0
JHA140	m 126 146 189 292 294 296	T2	1
JTI141	m 129 148 172 174 187 188G 189 223 230 304 311 320	L0a	0
JTI142	m 086 290 317 343 355 368	?	0
JTI143	m 134 356	?	0
JTI144	m 168	?	0
JTI145	m 093 124 223 278 362	L3b	0
JTI146	m 069 129 145 261	J1b	0
JTI147	m 148 172 187 188G 189 223 230 311 320	L0a2	0

Pokračování na další straně

Tabulka 2. Pokračování.

Vzorek	HVS-I motiv	Haploskupina	Htp ^a
JTI148	m 069 126 145 261	J1b	0
JTI149	m 172 223 268 301T 362	L4g	0
JTI150	m 209 223 292 311	L3f1	0

^a výskyt délkové heteroplasmie (homopolymerický cytozinový řetězec mezi 16184 - 16193)

3.1 Intrapopulační úroveň

Soubor JAC z jižního Jemenu (Al-Achkúm) obsahuje celkem 43 sekvencí, z toho 27 rozdílných haplotypů definovaných na základě 44 polymorfních míst. Z celkového počtu 45 substitucí připadá 37 na tranzice a 8 na tranzverze. Celkem 67 sekvencí tvoří soubor JTI ze západního Jemnu (Tihama). Rozdílných haplotypů je 39, polymorfních míst 63. Poměr tranzicí vůči transverzím je 59/6, celkem 65 sledovaných substitucí. Třetí soubor JHA získaný ve východním Jemenu (Hadramaut) je tvořen 40 sekvencemi, které spadají do 29 haplotypů a obsahují 53 polymorfních míst. 53 zjištěných substitucí je rozděleno mezi tranzice a transverze v poměru 49/6.

Základní statistiky hodnotící míru diverzity mtDNA ve všech 38 populacích shrnuje tabulka 3. Genová diverzita populací z Jemenu je nejnižší u populace Al-Achkúm ($H_s = 0,955$), ostatní dva soubory z Tihamy a Hadramautu dosahují srovnatelných hodnot (0,974 resp. 0,972). Nejnižší genová diverzita je u populací z Tuniska, která byla popisována již dříve (Cherni et al., 2005). Nízkou genovou diverzitu vykazují Indické soubory Nágáland a Karnataka a populace z oázy el-Hajez (od 0,940 do 0,953), kterým se přibližuje svou hodnotou populace Al-Achkúm. Populace z Tihamy a Hadramautu se v porovnání s ostatními soubory umísťují uprostřed. Nejvyšší hodnoty vykazují některé populace jak Afriky, tak Indie i Íránu. Populace východoafrických Guragů má nejvyšší genou diverzitu (1,000).

Tabulka 3. Základní statistiky udávající míru diverzity HVS-I mtDNA 38 populací.

	n	k	S	TTR	$H_s \pm SE$	$\pi \pm SE$	D_{ii}	$V(D_{ii})$	r	$P(r)$	D	$P(D)$	F_s	$P(F_s)$
Al-Achkúm	43	27	44	37/8	0,955±0,020	0,020±0,011	6,97	11,67	0,014	0,418	-1,101	0,114	-11,387	0,000
Tihama	67	39	63	59/6	0,974±0,008	0,020±0,011	6,89	10,60	0,012	0,261	-1,611	0,036	-22,108	0,000
Hadramaut	40	29	53	49/6	0,972±0,016	0,026±0,014	8,98	15,02	0,013	0,204	-0,999	0,149	-12,490	0,001
Bedulini	29	27	52	49/3	0,995±0,011	0,022±0,012	7,63	8,97	0,013	0,487	-1,593	0,031	-20,423	0,000
Tigrjci	53	46	67	64/8	0,994±0,005	0,024±0,013	8,24	9,30	0,009	0,343	-1,536	0,041	-24,843	0,000
Amharové	120	88	95	90/11	0,994±0,002	0,026±0,013	8,76	10,56	0,007	0,439	-1,634	0,014	-24,563	0,000
Oromové	33	30	57	55/4	0,992±0,010	0,025±0,013	8,64	8,53	0,029	0,020	-1,420	0,071	-21,188	0,000
Guragové	21	21	43	40/4	1,000±0,015	0,024±0,013	8,34	9,20	0,017	0,410	-1,196	0,108	-15,273	0,000
Turkana	37	34	61	60/3	0,994±0,009	0,030±0,016	10,20	13,23	0,007	0,714	-1,098	0,123	-23,478	0,000
Somálcí	27	24	43	39/4	0,992±0,013	0,021±0,011	7,05	7,61	0,015	0,432	-1,387	0,070	-15,986	0,000
Kikujové	25	23	48	44/4	0,993±0,013	0,025±0,013	8,45	11,73	0,009	0,734	-1,257	0,080	-13,984	0,000
dolní Egypt	59	53	68	64/6	0,994±0,005	0,022±0,012	7,63	11,91	0,008	0,667	-1,643	0,025	-24,931	0,000
horní Egypt	35	32	52	47/6	0,995±0,008	0,021±0,011	7,16	9,18	0,007	0,887	-1,578	0,036	-24,968	0,000
el-Hajez	35	19	35	33/2	0,953±0,017	0,022±0,012	7,41	17,35	0,019	0,270	-0,478	0,370	-2,853	0,174
Kesra	47	20	43	42/1	0,932±0,021	0,018±0,010	6,27	12,96	0,013	0,722	-1,230	0,111	-3,675	0,109
Zriba	50	16	31	30/1	0,904±0,022	0,012±0,007	3,95	7,27	0,015	0,870	-1,433	0,061	-3,219	0,131
Núbijci	78	48	70	67/7	0,972±0,009	0,024±0,012	8,21	14,31	0,005	0,780	-1,406	0,046	-24,779	0,000
niloté	65	55	67	64/7	0,994±0,004	0,024±0,013	8,28	12,02	0,004	0,833	-1,403	0,043	-24,810	0,000
Írán Střed	36	34	49	49/2	0,997±0,008	0,015±0,008	5,25	4,58	0,017	0,458	-2,009	0,009	-25,464	0,000
Írán SV	30	27	45	43/3	0,986±0,016	0,018±0,010	6,31	6,60	0,012	0,614	-1,649	0,030	-21,817	0,000
Írán SZ	226	154	108	106/7	0,987±0,003	0,016±0,009	5,50	5,84	0,010	0,637	-2,136	0,001	-24,970	0,000
Írán JZ	138	106	95	94/6	0,991±0,004	0,017±0,009	5,73	7,00	0,009	0,742	-2,130	0,001	-25,132	0,000
Karáci	100	77	88	82/8	0,992±0,003	0,016±0,009	5,52	5,84	0,010	0,713	-2,212	0,001	-25,308	0,000
Ándhrapradés	243	163	124	109/21	0,993±0,002	0,017±0,009	5,93	7,15	0,007	0,756	-2,184	0,000	-24,807	0,000
Ásám	45	36	49	48/1	0,984±0,010	0,017±0,009	5,80	5,18	0,014	0,358	-1,688	0,023	-25,323	0,000
Gudžarát	53	49	64	60/5	0,997±0,004	0,018±0,010	6,26	5,52	0,011	0,537	-1,929	0,008	-25,211	0,000
Himáčalpradés	37	31	50	48/3	0,988±0,010	0,018±0,010	5,99	5,14	0,017	0,302	-1,802	0,016	-24,719	0,000
Karnataka	195	59	61	57/5	0,941±0,009	0,014±0,007	4,69	5,94	0,025	0,107	-1,664	0,013	-25,317	0,000
Kérala	100	52	60	58/3	0,971±0,008	0,016±0,008	5,34	6,13	0,023	0,039	-1,732	0,011	-25,362	0,000

Pokračování na další straně

Tabulka 3. Pokračování.

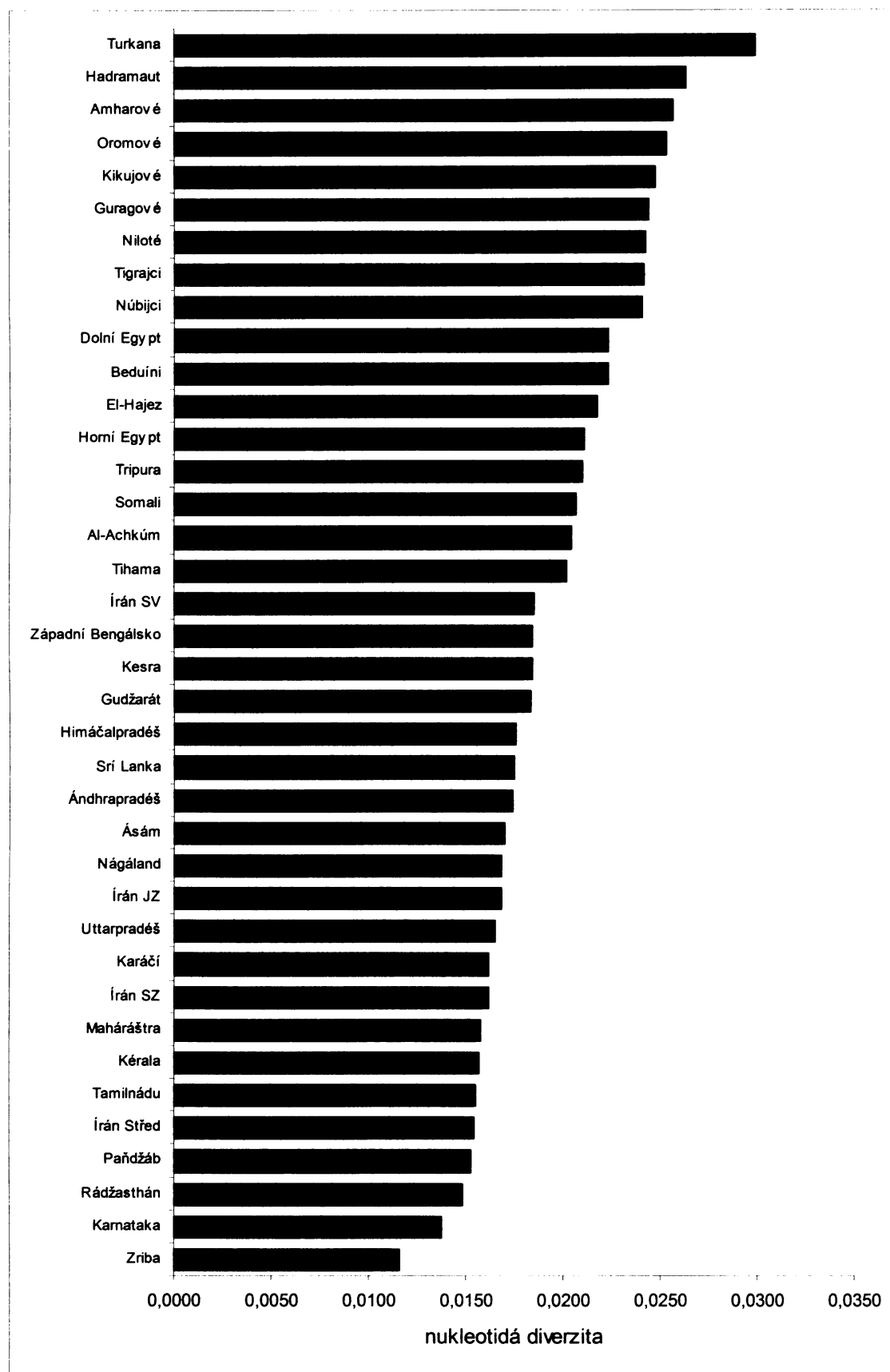
	<i>n</i>	<i>k</i>	<i>S</i>	TTR	$H_s \pm SE$	$\pi \pm SE$	<i>D_{ij}</i>	<i>V(D_{ij})</i>	<i>r</i>	<i>P(r)</i>	<i>D</i>	<i>P(D)</i>	<i>F_s</i>	<i>P(F_s)</i>
Maharáštra	113	71	82	76/9	0,983±0,005	0,016±0,008	5,37	6,32	0,011	0,624	-2,108	0,003	-25,321	0,000
Nágáland	43	21	36	34/2	0,940±0,019	0,017±0,009	5,75	8,63	0,023	0,171	-1,065	0,138	-5,889	0,022
Paňdžáb	214	123	91	87/8	0,988±0,003	0,015±0,008	5,20	5,33	0,010	0,678	-2,026	0,002	-25,093	0,000
Rádžasthán	35	28	47	44/3	0,985±0,011	0,015±0,008	5,07	4,41	0,016	0,537	-2,015	0,004	-22,703	0,000
Srí Lanka	132	93	85	76/13	0,990±0,003	0,018±0,009	5,98	6,81	0,007	0,808	-1,964	0,003	-25,081	0,000
Tamilnádu	148	89	89	80/15	0,983±0,004	0,015±0,008	5,28	5,91	0,009	0,771	-2,115	0,001	-25,236	0,000
Tripura	20	19	42	36/6	0,995±0,018	0,021±0,011	7,17	4,16	0,028	0,187	-1,574	0,046	-12,203	0,000
Uttarpradés	159	118	100	96/11	0,992±0,003	0,017±0,009	5,63	6,40	0,008	0,724	-2,153	0,003	-25,095	0,001
Západni														
Bengálsko	195	98	94	92/10	0,969±0,006	0,018±0,010	6,29	7,58	0,010	0,448	-1,883	0,006	-24,823	0,000

n - velikost souboru; *k* - počet odlišných haplotypů; *S* - počet polymorfních míst; TTR - poměr tranzice/transverze; $H_s \pm SE$ - genová diverzita \pm standardní chyba; $\pi \pm SE$ - nukleotidová diverzita \pm standardní chyba; D_{ij} - průměrný počet párových rozdílů; $V(D_{ij})$ - rozptyl sledovaných párových rozdílů; *r* - Harpendingův (raggedness) index; *P(r)* - pravděpodobnost Harpendingova indexu; *D* - Tajimův test neutrality; *P(D)* - pravděpodobnost statistiky *D*; F_s - test neutrality podle Fu (1997); $P(F_s)$ - pravděpodobnost statistiky F_s (statisticky významné hodnoty na 5% hladině významnosti pro $P(F_s)$ jsou pod 0,02).

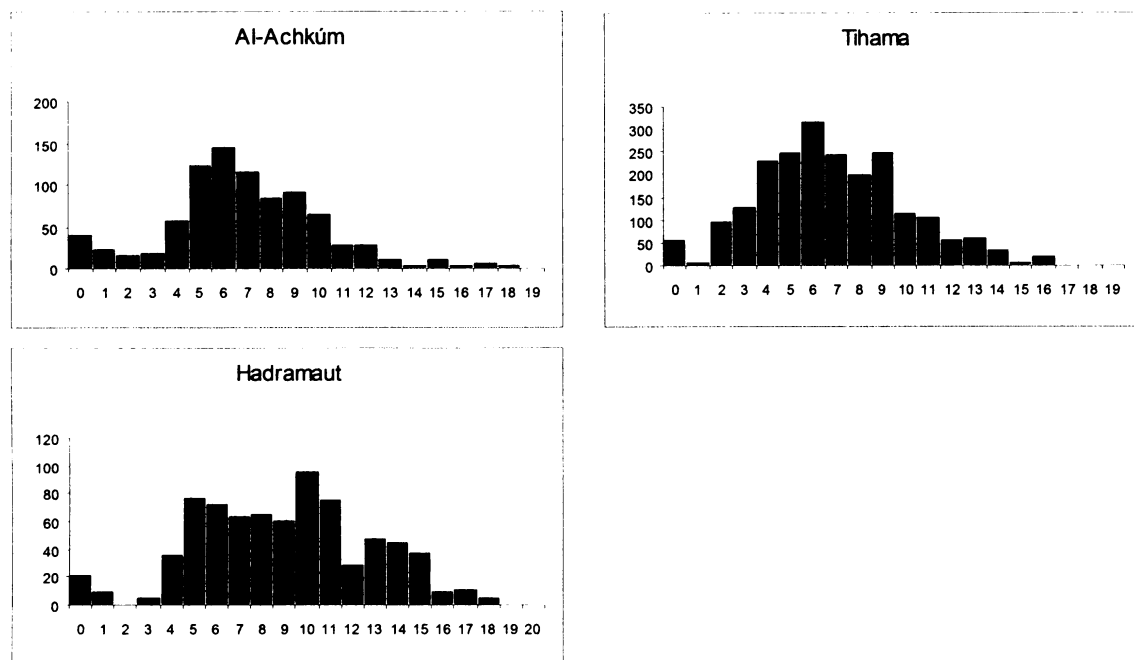
Hodnoty nukleotidové diverzity v rámci Jemenu nejvíce odlišují Hadramaut ($\pi = 0,026$), zatímco ostatní dvě populace mají hodnotu nižší (0,020). V porovnání se všemi populacemi dosahuje Hadramaut druhé nejvyšší nukleotidové diverzity a spadá mezi populace východní Afriky. Turkanové vykazují nejvyšší hodnotu (0,03), ostatní východoafrické populace se pohybují v rozmezí od 0,024 do 0,026 vyjma Somálců (0,021). Míra nukleotidové diverzity souborů Al-Achkúm a Tihama odpovídá populacím Egypta, vesnici Kesra a Beduínům. V celkovém pohledu se projevuje pokles nukleotidové diverzity směrem od východní Afriky přes severovýchodní Afriku a Arabský poloostrov k Indii a Íránu (viz graf 1). Výjimku představuje vesnice Zriba s nejnižší hodnotou (0,012). Naproti tomu vyšší míru nukleotidové diverzity vykazuje východoindický soubor Tripura (0,021).

Průměrný počet párových rozdílů kopíruje rozdělení nukleotidové diverzity. Nejvyšší počet má Turkana (10,20), Hadramaut vykazuje druhou nejvyšší hodnotu (8,98). Al-Achkúm a Tihama s obdobným počtem (6,97 a 6,89) se nacházejí mezi africkými a asijskými populacemi.

Rozložení počtu párových rozdílů jemenských populací (graf 2) nevykazuje statisticky signifikantní odchylky od předpokládaného normálního rozložení (Harpendingův r index). Ze všech populací mají hodnotu r statisticky významně vyšší pouze Oromové. V porovnání tří populací z Jemenu dosahuje Tihama statisticky významně zápornou hodnotu indexu D , svědčící o populační expanzi. Na druhou stranu F_s test, který je pro indikování vývoje velikosti populace citlivější (Pereira et al., 2001), je významný pro všechny populace. Z celkového souboru populací mají statisticky nevýznamnou hodnotu F_s pouze čtyři – el-Hajez, Kesra, Zriba a Nágáland.



Graf 1. Srovnání hodnot nukleotidové diverzity HVS-I mtDNA 38 populací.



Graf 2. Rozložení počtu párových rozdílů HVS-I mtDNA u 3 jemenských populací.

3.2 Populační srovnání (F_{ST})

Hodnoty F_{ST} a statistické významnosti těchto testů jsou uvedeny v tabulce 4. Výsledky ukazují, že všechny soubory z Jemenu se od sebe statisticky odlišují na 5% hladině významnosti. Pouze na 1% hladině významnosti je nesignifikantní rozdíl mezi populacemi z Tihamy a Al-Achkúmu. Obě tyto populace se také významně neodlišují od Beduínů ze Saudské Arábie. Al-Achkúm dále vykazuje afinitu k populaci horního Egypta, pouze při snížení hladiny významnosti na 1% je možné to samé říci o populaci z Tihamy. Tihama se dále statisticky neliší na 5% hladině významnosti od severovýchodního Íránu a na 1% hladině významnosti od centrálního Íránu. Hadramaut se naproti tomu neodlišuje od etiopské populace Oromo a od Kikujů z Keni. Na 1% hladině významnosti můžeme sledovat afinitu Hadramautu k dalším východoafrickým populacím Oromů a Somálců a také ke vzorků z egyptské oázy el-Hajez.

Tabulka 4. Hodnoty F_{ST} mezi 38 populacemi (pod diagonálou), statistická významnost (nad diagonálou), kódování populací viz tabulka 1.

JAC	JTI	JHA	Bd	Tlg	Amh	Oro	Gur	Tur	Som	Kik	EpD	EhH	EiH	Kes	Zri	Nub	Nii	IS	ISV	ISZ	IJZ	Kar	And	Asm	Gud	Him	Kar	Ker	Meh	Nag	Pan	Rad	SiL	Tam	Tri	Utr	Zbe		
JAC	0.000	-	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	
JTI	0.019	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
JHA	0.112	0.075	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Bd	0.003	0.004	0.074	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Tlg	0.055	0.052	0.038	0.038	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Amh	0.080	0.072	0.025	0.059	0.004	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Oro	0.077	0.088	0.019	0.052	0.000	0.000	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Gur	0.090	0.087	0.035	0.069	0.000	0.000	0.000	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Tur	0.191	0.182	0.048	0.157	0.072	0.047	0.048	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Som	0.106	0.087	0.035	0.075	0.015	0.001	0.005	0.001	0.063	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Kik	0.163	0.132	0.007	0.120	0.048	0.015	0.017	0.021	0.000	0.023	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
EpD	0.024	0.027	0.050	0.009	0.015	0.031	0.028	0.035	0.121	0.037	0.080	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
EhH	0.012	0.020	0.053	0.000	0.016	0.038	0.029	0.039	0.130	0.043	0.090	0.000	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
EiH	0.096	0.098	0.042	0.069	0.048	0.042	0.044	0.060	0.094	0.055	0.062	0.021	0.028	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Kes	0.077	0.051	0.046	0.052	0.033	0.032	0.033	0.036	0.127	0.034	0.071	0.018	0.021	0.047	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Zri	0.091	0.068	0.128	0.072	0.082	0.100	0.113	0.134	0.235	0.140	0.190	0.048	0.047	0.121	0.087	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Nub	0.080	0.068	0.017	0.055	0.007	0.005	0.002	0.007	0.045	0.021	0.016	0.025	0.028	0.033	0.031	0.008	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Nii	0.138	0.128	0.046	0.114	0.039	0.017	0.021	0.011	0.028	0.011	0.010	0.085	0.078	0.063	0.055	0.155	0.028	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
IS	0.043	0.023	0.084	0.022	0.053	0.068	0.070	0.069	0.187	0.088	0.139	0.019	0.011	0.085	0.032	0.027	0.063	0.122	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
ISV	0.043	0.015	0.068	0.010	0.078	0.064	0.066	0.116	0.207	0.114	0.160	0.037	0.024	0.104	0.083	0.072	0.090	0.153	0.011	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
ISZ	0.043	0.025	0.132	0.021	0.083	0.110	0.110	0.128	0.258	0.128	0.193	0.034	0.028	0.115	0.080	0.028	0.103	0.173	0.008	0.020	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
IJZ	0.036	0.017	0.114	0.014	0.070	0.064	0.095	0.109	0.234	0.111	0.173	0.028	0.018	0.105	0.053	0.028	0.085	0.157	0.000	0.012	0.000	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Kar	0.084	0.048	0.068	0.058	0.060	0.070	0.083	0.064	0.195	0.076	0.132	0.053	0.052	0.128	0.050	0.083	0.071	0.127	0.024	0.046	0.045	0.036	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
And	0.094	0.058	0.087	0.068	0.066	0.068	0.077	0.087	0.192	0.059	0.118	0.055	0.055	0.112	0.043	0.073	0.074	0.117	0.029	0.053	0.061	0.050	0.007	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Asm	0.143	0.118	0.098	0.116	0.078	0.070	0.082	0.106	0.152	0.073	0.119	0.103	0.108	0.147	0.101	0.172	0.086	0.112	0.100	0.122	0.151	0.131	0.071	0.050	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Gud	0.064	0.030	0.073	0.031	0.052	0.061	0.063	0.060	0.173	0.067	0.118	0.031	0.033	0.063	0.037	0.056	0.058	0.114	0.011	0.025	0.024	0.019	0.006	0.010	0.074	0.000	++	++	++	++	++	++	++	++	++	++	++	++	++
Him	0.085	0.061	0.075	0.058	0.042	0.051	0.057	0.071	0.146	0.058	0.102	0.040	0.047	0.096	0.048	0.087	0.054	0.095	0.035	0.081	0.064	0.053	0.021	0.012	0.041	0.009	0.000	++	++	++	++	++	++	++	++	++	++	++	
Ker	0.111	0.072	0.102	0.077	0.084	0.079	0.097	0.120	0.225	0.079	0.152	0.072	0.069	0.138	0.088	0.113	0.089	0.152	0.049	0.071	0.081	0.088	0.027	0.013	0.067	0.026	0.037	0.000	++	++	++	++	++	++	++	++	++	++	
Meh	0.114	0.071	0.100	0.078	0.076	0.074	0.069	0.099	0.201	0.077	0.136	0.063	0.069	0.128	0.056	0.076	0.079	0.128	0.037	0.067	0.063	0.054	0.016	0.012	0.076	0.013	0.020	0.033	0.000	++	++	++	++	++	++	++	++	++	
Nag	0.117	0.089	0.099	0.089	0.054	0.058	0.066	0.085	0.148	0.072	0.109	0.073	0.079	0.125	0.077	0.132	0.064	0.101	0.080	0.112	0.117	0.103	0.060	0.049	0.060	0.021	0.076	0.057	0.050	0.000	++	++	++	++	++	++	++	++	
Pan	0.084	0.045	0.102	0.055	0.063	0.077	0.066	0.104	0.221	0.081	0.148	0.048	0.048	0.123	0.042	0.056	0.080	0.136	0.015	0.047	0.037	0.032	0.008	0.010	0.075	0.005	0.011	0.025	0.016	0.016	0.056	0.000	++	++	++	++	++	++	
Rad	0.097	0.054	0.075	0.066	0.053	0.063	0.066	0.077	0.159	0.054	0.105	0.052	0.056	0.119	0.052	0.105	0.081	0.105	0.041	0.061	0.089	0.054	0.006	0.000	0.059	0.013	0.009	0.017	0.007	0.002	0.045	0.011	0.000	++	++	++	++	++	
SiL	0.092	0.057	0.086	0.064	0.059	0.063	0.071	0.078	0.178	0.061	0.114	0.051	0.051	0.108	0.040	0.065	0.066	0.108	0.025	0.054	0.057	0.045	0.007	0.005	0.058	0.009	0.015	0.024	0.010	0.016	0.048	0.009	0.008	0.000	++	++	++	++	
Tam	0.107	0.068	0.079	0.073	0.069	0.060	0.073	0.068	0.187	0.058	0.114	0.059	0.058	0.105	0.048	0.087	0.070	0.114	0.035	0.081	0.072	0.060	0.021	0.006	0.045	0.018	0.016	0.013	0.018	0.018	0.054	0.018	0.011	0.012	0.000	++	++	++	
Tri	0.098	0.068	0.064	0.063	0.050	0.061	0.055	0.074	0.135	0.069	0.096	0.054	0.055	0.101	0.059	0.076	0.082	0.098	0.043	0.063	0.079	0.068	0.041	0.032	0.034	0.034	0.012	0.067	0.033	0.047	0.029	0.035	0.029	0.026	0.033	0.000	++	++	
Utr	0.090	0.056	0.066	0.065	0.066	0.071	0.066	0.064	0.200	0.070	0.134	0.055	0.055	0.124	0.055	0.076	0.074	0.128	0.032	0.057	0.057	0.047	0.003	0.008	0.069	0.012	0.024	0.022	0.010	0.009	0.058	0.009	0.001	0.006	0.019	0.041	0.000	++	
Zbe	0.111	0.090	0.068	0.086	0.075	0.064	0.081	0.087	0.172	0.053	0.110	0.070	0.069	0.107	0.058	0.111	0.074	0.107	0.060	0.084	0.085	0.084	0.035	0.021	0.049	0.033	0.023	0.036	0.036	0.030	0.046	0.038	0.023	0.026	0.023	0.058	0.031	0.000	

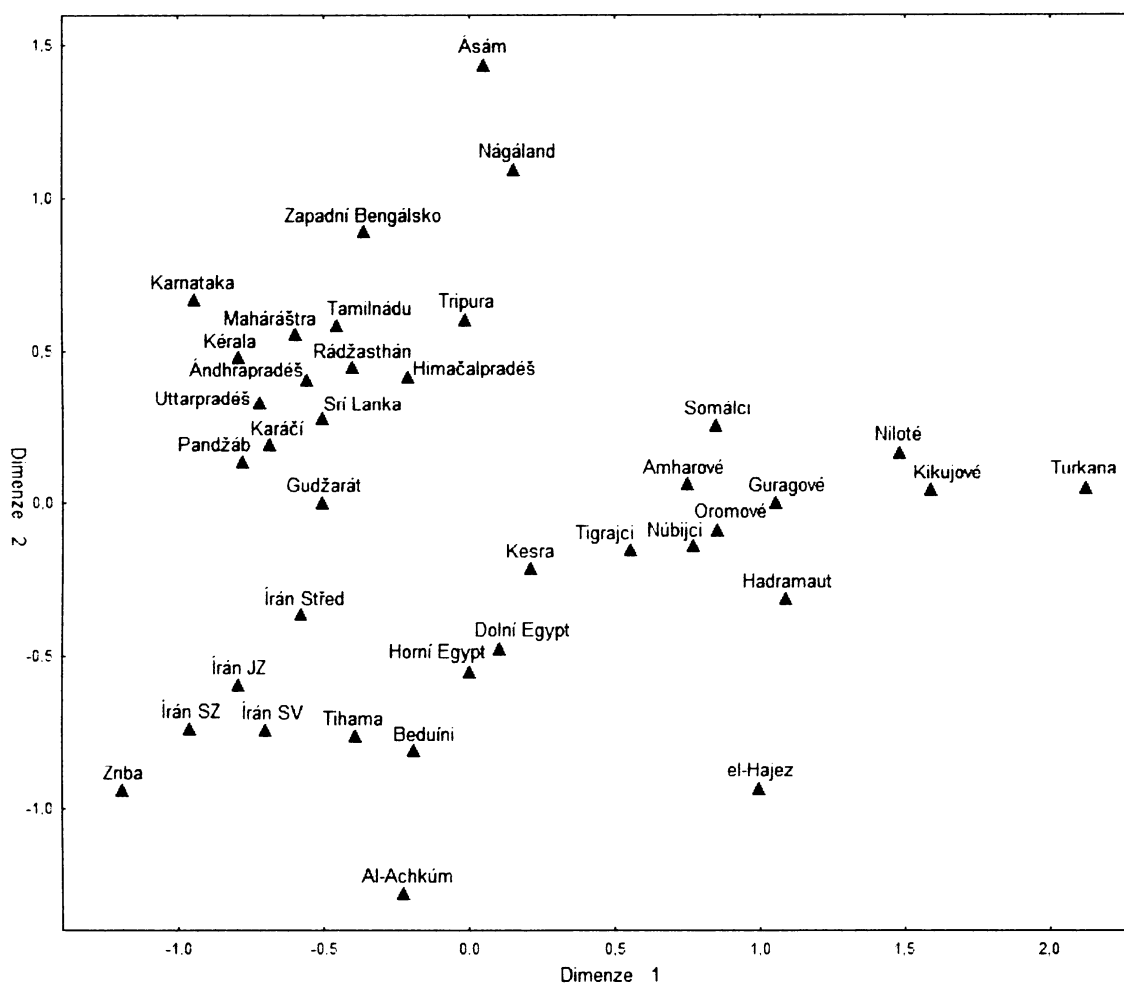
++ statisticky významné hodnoty na 5% hladině významnosti
 -- statisticky nevýznamné hodnoty na 5% hladině významnosti
 - statisticky nevýznamné

Z výsledků dalších souborů je třeba zmínit populaci Beduínů, která se statisticky neodlišuje od horního ani dolního Egypta a také od severovýchodního Íránu. Dále pak můžeme sledovat příbuznost populací severovýchodní Afriky (s výjimkou rozdílu mezi Tigrjci a Kikuji), od kterých se odlišuje pouze populace Turkanů (kromě vztahu s populací Kikujů). Zajímavým výsledkem je také odlišnost souborů z indického subkontinentu od všech ostatních populací, vyjma shody státu Gudžarát s centrálním Íránem na 1% hladině významnosti. Matice F_{ST} vzdáleností tak nenaznačuje žádné vztahy mezi populacemi podél jižní migrační cesty v úseku od Arabského poloostrova k indickému subkontinentu.

Graf vícerozměrného škálování podle matice F_{ST} vzdáleností mezi všemi 38 populacemi (graf 3) ukazuje na odlišné vazby souborů Al-Achkúm a Tihama na straně jedné a populace Hadramatu na straně druhé. Soubory ze západního Jemenu tvoří shluk spolu s Beduíny a populacemi Egypta a Íránu. Naproti tomu je třeba upozornit na afinitu Hadramautu k východoafrickým populacím, mezi něž zde patří i Núbijci a niloté. Populace z Indického poloostrova tvoří samostatnou oddělenou skupinu.

3.3 Rozdělení genetické diverzity (AMOVA)

Analýza molekulárního rozptylu poskytuje informace o genetické struktuře (tabulka 5). Nejprve byly všechny populace posuzovány dohromady jako jedna skupina. V takovém případě dosahoval podíl variability mezi populacemi uvnitř skupiny 5,81% a uvnitř populací 94,19%. Populace byly následně shlukovány podle různých kritérií. V prvním kroku jsme se zaměřili na otázku vztahů populací podél jižní migrační cesty z východní Afriky přes úžinu Báb el-Mandab a jih Arabského poloostrova až po indický subkontinent. Populace byly rozděleny do dvou skupin. První zahrnuje populace severovýchodní Afriky, Beduíny ze Saudské Arábie a soubory z Íránu. Druhá skupina populace východní Afriky spolu s povodím Nilu, námi analyzované



Graf 3. Výsledný graf vícerozměrného škálování na základě hodnot F_{ST} mezi 38 populacemi

soubory z Jemenu a indický subkontinent. Výsledky analýzy (tabulka 5, krok 1) udávají malé rozdíly mezi vytvořenými skupinami (podíl variability 2,82%), uvnitř kterých existují velké rozdíly mezi populacemi (podíl variability 4,70%). Ke zvýšení podílu variability mezi skupinami (3,32%) došlo, pokud byly populace Al-Achkúm a Tihama spolu s Beduíny přesunuty jak napovídají výsledky populačního srovnání. Ve třetím kroku byly populace sloučeny na lingvistickém základě podle jazykových rodin. Z této analýzy byly vypuštěny soubory jedinců ze Srí Lanky a Západního Bengálska, neboť náležejí k více než jedné jazykové rodině. Pouze částečný nárůst podílu variability mezi skupinami (3,67%) byl zaznamenán v porovnání s předchozí konfigurací. Nejvyšších rozdílů mezi skupinami bylo dosaženo při shlukování populací na základě výsledků F_{ST} , kdy byly vytvořeny tři skupiny podle naznačené struktury

ve výsledném grafu vícerozměrného škálování (graf 3). První slučuje populace východní Afriky, povodí Nilu a populaci z Hadramautu. Druhá skupina obsahuje severovýchodní Afriku, Írán a z Arabského poloostrova populace Beduínů, Al-Akchúm a Tihamu. Populace z Indie tvoří třetí skupinu. Za této konfigurace je podíl variability mezi skupinami 5,63%, uvnitř skupin 2,35 a uvnitř populací 92,02% (tabulka 5, krok 4).

Tabulka 5. Výsledky analýzy molekulárního rozptylu (AMOVA), podrobnosti viz text.

	Variabilita mezi skupinami	Variabilita mezi populacemi uvnitř skupin	Variabilita uvnitř populací
Všechny populace		5,81	94,19
Krok 1 (severní vs. jižní cesta)	2,82	4,70	92,48
Krok 2 (severní cesta + JAC a JTI vs. jižní cesta + JHA)	3,32	4,48	92,29
Krok 3 (podle jazykových rodin)	3,67	3,16	93,16
Krok 4 (JHA + východní Afrika + povodí Nilu vs. JAC a JTI + severovýchodní Afrika + Írán vs. Indie)	5,63	2,35	92,02

3.4 Haplotypy sdílené mezi populacemi

Pro získání dalších informací o vztazích populací Jemenu s ostatními soubory byly zjištěny haplotypy, které se shodují u populací, jež se statisticky neodlišují na 5% hladině významnosti. Počet shodných haplotypů a jejich odhadovaná haploskupinová příslušnost je uveden v tabulce 6. Soubory ze západního Jemenu sdílejí s populacemi Blízkého východu, od kterých se signifikantně neliší, haplotypy spadající do euroasijských haploskupin (derivátů makrohaploskupiny N) J1b, J1c,

T1 a haploskupiny R0 a R0a, dříve popisované jako preHV a (preHV)1 (viz např. Richards et al., 2000).

Hadramaut sdílí s populací Oromo dva haplotypy skupiny J3d1. Kikujové a Hadramautu mají společné haplotypy klasifikované jako subsaharské haploskupiny L0a1 a L0a2.

Tabulka 6. Haploskupinová příslušnost a počet sdílených haplotypů mezi vybranými populacemi

Al-Akchúm - Beduíni	J1b 3	J1c 1	R0 1	R0a 1	T1 1
Al-Akchúm - horní Egypt	R0a 8				
Tihama - Beduíni	J1b 7	R0a 3			
Tihama – severovýchodní Írán	J1b 7				
Hadramaut - Oromové	J3d1 3				
Hadramaut - Kikujové	L0a2 6	L0a1 1			

4 DISKUZE

Výsledky populačně-genetického srovnání 38 populací ukazují v celkovém pohledu na vztah mezi geografickou vzdáleností a genetickou afinitou v souladu s globálním klinálním rozdělením genetické diverzity lidských populací sledovaném na úrovni jaderné DNA (Serre & Paabo, 2004). Odpovídá tomu na jedné straně vícerozměrné škálování na základě matice F_{ST} vzdáleností, které ukazuje, že populace spadají do výsledného grafu podle zeměpisné polohy od východní Afriky přes severovýchodní Afriku a Arabský poloostrov po Blízký východ a Indický subkontinent (graf 3). Na straně druhé také pokles nukleotidové diverzity a počtu párových rozdílů kopíruje výše popsany průběh. Vysoká nukleotidová diverzita východní Afriky a negativní závislost diverzity

ostatních populací na vzdálenosti od východní Afriky byla popsána jednak na úrovni mtDNA (Černý et al., 2004; Ingman et al., 2000), tak také prostřednictvím analýz úseků jaderné DNA (Prugnolle et al., 2005; Tishkoff & Kidd, 2004).

Genová diverzita naproti tomu vykazuje nepravidelné uspořádání napříč sledovanými oblastmi. Tento výsledek není překvapivý vzhledem k tomu, že genová diverzita odráží celkovou homogenitu či heterogenitu souboru. Na její hodnoty tak mají vliv populačně demografické události jako redukce nebo expanze populace, ale významně také způsob sběru a tvorby souborů dat. Například soubory z Indie dosahující v řadě případů vysoké genové diverzity (viz tabulka 3) jsou utvořeny na základě spolkových států. Jelikož přesnější zeměpisné informace nejsou dostupné, je možné, že takto velké regiony zahrnují více geneticky diverzifikovaných podoblastí. Jejich sloučením tak narůstá celková diverzita souboru.

Detailní prospekce našich výsledků podtrhuje význam geografického přístupu při sběru dat pro tuto studii. Předchozí práce (Kivisild et al., 2004) nabízí podrobnou analýzu haploskupin u populací Etiopie a Jemenu a sleduje genový tok mezi Afrikou a Arabským poloostrovem a vlivy z euroasijských populací. Diverzitu mtDNA Jemenu prezentují autoři jako konglomerát subsaharských a euroasijských haploskupin zastoupených v podobné míře. Vzorky byly ale získány od probandů v Kuvajtu a jejich přesný zeměpisný původ na území Jemenu nelze vymezit (Kivisild et al., 2004). Naše data ukazují, že různé oblasti Jemenu odráží vztahy s východní Afrikou a Blízkým východem rozdílně. Přes geografickou a kulturní blízkost západního pobřeží Jemenu a východní Afriky dokládánou archeologicky přinejmenším od vzniku tzv. tihamského kulturního komplexu v polovině 2. tisíciletí př. nl. (Fattovich, 2002) nevykazují vzorky z Al-Achkúmu a Tihamy předpokládanou afinitu k východoafrickým populacím. Naproti tomu spadají blíže populacím Blízkého východu – Egyptu, Saudské Arábie a Íránu. Hadramaut, který leží přibližně 800km východněji, je naopak na základě mtDNA příbuzný se

soubory z východní Afriky (viz graf 3). Tomu odpovídají rovněž výsledky analýzy molekulárního rozptylu (tabulka 5, krok 4), které rozdílnou genetickou diverzitu populací ze západu a východu Jemenu potvrzují.

Populační přístup neumožňuje zasadit sledované vztahy do časové perspektivy. Přesto relativní odhad doby populační expanze, jejíž signály vykazují bezmála všechny studované populace, poskytuje průměrný počet párových rozdílů. Delší časový úsek od populačního nárůstu poskytuje prostor pro nashromáždění většího počtu rozdílů mezi sekvencemi. Hadaramaut vykazuje větší průměrný počet párových rozdílů než ostatní dvě populace ze západního Jemenu a podobá se tak populacím východní Afriky (viz tabulka 3). Stejně tak je tomu i v případě nukleotidové diverzity. Domníváme se proto, že afinita Hadramautu a východní Afriky má časově hlubší kořeny a může reflektovat i pravěké migrace přes úžinu Báb el-Mandab na Arabský poloostrov. Naproti tomu pozice populací Tihamy a Al-Achkúmu v populačním srovnání se jeví jako následek recentních vztahů těchto oblastí s Blízkým východem. Těžko ovšem rozhodnout, zda jde o následek celkové genetické příbuznosti populací Blízkého východu navržené na základě výzkumů Y chromosomu (Shepard & Herrera, 2006), nebo zde roli hrály oddělené vlivy. Na jedné straně například nevýznamný rozdíl mezi soubory sekvencí z Al-Achkúmu a horního Egypta mohou odrážet egyptské kontakty podél Rudého moře (Fattovich, 2002) nebo pozdější genový tok z Jemenu spojený s expanzemi arabských kmenů, který naznačují výzkumy Y chromosomu (Nebel et al., 2002). Na straně druhé odlišný základ může mít genetická podobnost vzorků z Tihamy a severovýchodního Íránu. Pro zpřesnění rozdělení genetické diverzity napříč Blízkým východem by bylo potřeba implementovat další data z Arabského poloostrova a severně od něj. Například oblasti Iráku, Jordánska, Izraele či Sýrie představují v našich analýzách slepá místa (viz obrázek 2). Bohužel v současné době neprošel žádný z dostupných souborů dat kritériemi stanovenými v této studii (viz sekce 2.1). Především problémovým bodem se ukazuje být délka sekvencí a způsob sběru dat, který neumožňuje získat informace o geografickém původu

vzorků (viz např. Behar et al., 2006; Macaulay et al., 1999a; Richards et al., 2000).

Doplnit naše výsledky o časovou představu by dále umožnila fylogeografická analýza sekvencí, která se zaměřuje na rekonstrukci ancestrálních haplotypů, odhad jejich geografického původu a datování. Při řešení otázky, zda diverzita mitochondriální DNA na území Jemenu obsahuje známky prvotních migrací anatomicky moderního člověka z východní Afriky přes Báb el-Mandab, bychom se nejprve měli zaměřit na východní oblasti Jemenu. Hadramaut a jeho afinita k východoafrickým populacím odpovídá modelu disperze podél jižní cesty, zatímco recentní vlivy odhadované u západních populací mohou starší populační události překrývat.

Pravěké migrace z východní Afriky jsou spojovány především se šířením haploskupiny M (Forster & Matsumura, 2005; Macaulay et al., 2005; Quintana-Murci et al., 1999). Předběžná klasifikace HVS-I haplotypů však ukazuje, že tato haploskupina je mezi našimi sekvencemi z Hadramautu zastoupena minimálně nebo může chybět úplně (viz tabulka 2). Podle předběžných výsledků RFLP analýzy kódujícího úseku se zdá, že současná klasifikace bude v budoucnu korigována jen u přibližně 10% sekvencí (Černý, osobní sdělení). Není však jasné, zda ke vzniku haploskupiny M došlo ještě na území Afriky, nebo až v jihozápadní či jižní Asii. Zatím nejvyšší diverzita a také odhadované datování M haploskupiny bylo zaznamenáno na území Indie (Krings et al., 1999; Metspalu et al., 2004; Quintana-Murci et al., 1999), avšak vysokou diverzitu a výskyt vykazuje rovněž v Etiopii (Kivisild et al., 2004; Quintana-Murci et al., 1999). Právě Arabský poloostrov včetně přilehlých oblastí Afriky (Eritrea, Džibuti) by tak teoreticky mohl být místem vzniku haploskupiny M, což by umožnilo jednak časně rozšíření do Etiopie prostřednictvím zpětných migrací, a také expanzi dále do Indie. Současné výsledky výskytu haploskupin v námi studovaných oblastech však zatím nenaznačují, že by místem diverzifikace haploskupiny M byla oblast Jemenu.

Existuje možnost, že veškeré populační afinity, o kterých vypovídají výsledky populačního srovnání, mohou odrážet recentní události. Předchozí studie odvozuje značnou část diverzity mitochondriální DNA populace Jemenu od vlivů během posledních několika tisíců let (Kivisild et al., 2004). Pro získání další představy o podstatě námi sledovaných F_{ST} vzdáleností jsme se zaměřili na haplotypy sdílené mezi soubory z Jemenu a populacemi, jež se od nich statisticky neodlišují. V případě sekvencí z Tihamy a Al-Achkúmu a jejich protějšků u populací Blízkého východu jde o euroasijské haploskupiny (viz tabulka 6) běžné u populací Evropy, Blízkého východu a severní Afriky (Klings et al., 1999; Richards et al., 2000). Vyskytují se také ve východní Africe v Etiopii, kde jsou považovány za následek relativně časných zpětných migrací z Blízkého východu (Kivisild et al., 2004; Richards, 2003).

Soubor z Hadramautu sdílí s geneticky blízkými populacemi z východní Afriky haploskupinu J v případě Oromů a subsaharské haploskupiny L0a1 a hlavně L0a2 se souborem Kikuju (viz tabulka 6). Původ haploskupin L0a1 a L0a2 je odvozován ve východní Africe, kde se doba oddělení L0a1 odhaduje přibližně na 33350 let. Oddělení L0a2 se datuje přibližně na 8300 a předpokládá se rozšíření do centrální Afriky v posledních 10000 let (Salas et al., 2002).

Na základě těchto informací se zdá, že naše výsledky populačního srovnání odrážejí recentnější události, a to jak v případě vzorků ze západního Jemenu, tak i z oblasti Hadramautu. Subsaharské L0a haploskupiny byly předchozími studiemi přisouzeny převážně vlivu arabského obchodu otroky a jejich historie u populací Jemenu se uvažuje v časové perspektivě posledních 2500 let. Jako zdrojová oblast je zmiňována jihovýchodní Afrika díky shodě L0a haplotypů z Jemenu a Mozambiku (Kivisild et al., 2004; Richards, 2003).

Autoři Richards et al. (2003) navíc zaznamenávají výrazný rozdíl ve výskytu subsaharských haploskupin mezi arabskými populacemi a ostatními etniky na Předním Východě. Nejvyšší frekvenci dokumentují

u Arabů z oblasti Hadramautu. Tyto výsledky interpretují jako odlišný demografický a populačně-genetický dopad obchodu otroky na různá etnika. Arabské populace měli být hlavním cílovým místem pro otroky ze subsaharské Afriky. Podle srovnání s diverzitou Y chromosomu šlo hlavně o ženy (Richards et al., 2003). Naše výsledky dokládají rozdílnou afinitu vzorků z Jemenu a východní Afriky v závislosti na geografickém původu. Vystává tak otázka, zda rozdílné frekvence haploskupin přinejmenším na území Jemenu neodrážejí spíše geografické závislosti, než rozdělení podle etnických skupin. Na základě současných dat je však její zodpovězení nejisté. Kromě vzorků z Hadramautu byl v předchozí studii analyzován soubor jemenských Židů (Richards et al., 2003). Některé vzorky byly získány od jedinců mimo Jemen, např. v Izraeli (Thomas et al., 2002) a geografický původ sekvencí na území Jemenu nelze určit.

Problémem zacílení na sdílené haplotypy je redukce vztahů mezi populacemi pouze na několik sekvencí, které se mezi populacemi shodují. Přesto lze předpokládat, že právě shodné haplotypy se nejvíce podílejí na proximitě některých populací ve výsledcích F_{ST} vzdáleností. Přesnější klasifikace haplotypů doplněná o RFLP analýzu kódujících oblastí mtDNA a fylogeografická analýza je však nezbytná pro systematické zdokumentování výskytu haploskupin a jejich historie na území Jemenu.

5 ZÁVĚR

Celkem 150 nových sekvencí HVS-I mtDNA z dosud málo zdokumentované oblasti Jemenu bylo analyzováno v této studii. Výsledky jasně dokládají význam regionálního postupu při sběru dat. Díky tomu jsme byli schopni odhalit rozdíly v diverzitě mtDNA v různých oblastech jižní části Arabského poloostrova včetně jejich vztahů k dalším populacím Afriky a Blízkého východu, které se ne vždy prolínají s kulturními a lingvistickými vazbami.

Západní oblasti Jemenu – Tihama a Al-Achkúm – vykazují afinitu k populacím Blízkého východu, přestože geograficky a kulturně se blíží spíše východní Africe. Populace Hadramautu naproti tomu spadá v populačně-genetickém srovnání blíže populacím východní Afriky. Žádné vztahy nebyly sledovány mezi soubory podél pobřežní trasy od jihu Arabského poloostrova po indický subkontinent.

Na základě současných výsledků se nabízí následující postup pro budoucí výzkumy genetické diverzity populací jižní Arábie. V první řadě je to zahrnutí dalších dat z oblastí, které v geografickém rozdělení souborů použitých v této studii představují slepá místa. Jde především o sever Arabského poloostrova důležitý pro detailnější studium genového toku mezi Blízkým východem a západním pobřežím Jemenu. Další vzorky z východní a jihovýchodní Afriky na jedné straně a Ománu podél jihovýchodního pobřeží Arabského poloostrova na straně druhé mohou zase zpřesnit informace o roli úžiny Báb el-Madeb v kontaktech mezi Afrikou a Arabským poloostrovem.

Předběžné výsledky naznačují, že genetická struktura odhalená populačně-genetickým srovnáním odráží ve větší míře recentní události. Nicméně až podrobná fylogeografická analýza do budoucna umožní doplnit populační přístup (který má spíše explorační význam) o odhad geografického a časového původu genetické diverzity jemenských populací.

6 PŘÍLOHY

Příloha 1. Protokol izolace mtDNA z bukových stěrů prostřednictvím kitu QIAquick PCR Purification (QIAGEN).

1. vortexovat 1 min.
 2. odsát 200 μ l a přenést do sterilní eppendorfky (1.5 - 2 ml).
 3. přidat 1 000 μ l extrakčního pufru PB a zvortexovat.
 4. inkubovat při 60°C 1 hod s mícháním (cca 200 ot) a stočit (1 min., 12 000 g).
 5. supernatant přenést na 2x do kolonky a stočit (1,5 min, 12 000 g).
 6. přidat 200 μ l extrakčního pufru PB a stočit (1,5 min, 12 000 g).
 7. přidat 700 μ l proplachovacího pufru PE s etanolem a stočit (1,5 min, 12 000 g).
 8. vylít kolonku, osušit okraje, stočit (3 min, 12 000 g) a nechat v boxu v průvanu.
 9. vyměnit eppendorfku, přidat μ l EB pufru, inkubovat 5 min a stočit (1,5 min., 12 000g).
 10. vyhodit kolonku a eppendorfku s DNA uložit do mrazáku.
-

Příloha 2. Protokol purifikace PCR fragmentů prostřednictvím kitu QIAquick PCR Purification (QIAGEN).

1. přidat 5 objemů PB pufru k 1 objemu PCR vzorku a promíchat (na 25 μ l 125 μ l PB).
 2. přenést vzorek do kolonky a stočit (1,5 min., 12 000 g).
 3. přidat 730 μ l proplachovacího pufru PE s etanolem a stočit (1,5 min, 12 000 g).
 4. vylít kolonku, osušit okraje, stočit (3 min, 12 000 g) a nechat temperovat.
 5. vyměnit eppendorfku, přidat 30 μ l EB pufru, inkubovat 5 min a stočit (1,5 min., 12 000 g).
 6. vyhodit kolonku a eppendorfku s DNA uložit do mrazáku.
-

7 LITERATURA

- Alberts B, Bray D, Lewis J, Raff M, Roberts K 1994. *Molecular Biology of the Cell*. New York: Garland.
- Anderson S, Bankier AT, Barrell BG, Debruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG 1981. Sequence and Organization of the Human Mitochondrial Genome. *Nature* 290:457-465.
- Avise JC 2000. *Phylogeography: the history and formation of species*. Cambridge, Mass.: Harvard University Press.
- Awadalla P, Eyre-Walker A, Smith JM 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* 286:2524-2525.
- Awadalla P, Eyre-Walker A, Smith JM 2000. Questioning evidence for recombination in human mitochondrial DNA - response. *Science* 288:1931a.
- Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BVR, Rasanayagam A, Hammer MF 1998. Female gene flow stratifies Hindu castes. *Nature* 395:651-652.
- Bandelt H-J, Forster P, Röhl A 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48.
- Bandelt H-J, Foster P, Sykes BC, Richards MB 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743-753.
- Barbujani G, Bertorelle G 2001. Genetics and the population history of Europe. *Proc Natl Acad Sci USA* 98:22-25.
- Barbujani G, Bertorelle G, Chikhi L 1998. Evidence for paleolithic and neolithic gene flow in Europe. *Am J Hum Genet* 62:488-491.
- Behar DM, Metspalu E, Kivisild T, Achilli A, Hadid Y, Tzur S, Pereira L, Amorim A, Quintana-Murci L, Majamaa K, Herrnstadt C, Howell N, Balanovsky O, Kutuev I, Pshenichnov A, Gurwitz D, Bonne-Tamir B, Torroni A, Villems R, Skorecki K 2006. The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. *Am J Hum Genet* 78:487-497.
- Birky CW, Jr. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc Natl Acad Sci USA* 92:11331-11338.

- Brandstätter A, Parsons TJ, Parson W 2003. Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med* 117:291-298.
- Brown DT, Samuels DC, Michael EM, Turnbull DM, Chinnery PF 2001. Random genetic drift determines the level of mutant mtDNA in human primary oocytes. *Am J Hum Genet* 68:533-536.
- Brown WM 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc Natl Acad Sci USA* 77:3605-3609.
- Brown WM, George M, Wilson AC 1979. Rapid Evolution of Animal Mitochondrial DNA. *Proc Natl Acad Sci USA* 76:1967-1971.
- Cann RL, Stoneking M, Wilson AC 1987. Mitochondrial DNA and human evolution. *Nature* 325:31-36.
- Carvalho-Silva DR, Zerjal T, Tyler-Smith C 2006. Ancient Indian roots? *J Biosci* 2006:3-4.
- Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* 11:253-264.
- Cordaux R, Stoneking M 2003. South Asia, the Andamanese, and the genetic evidence for an "early" human dispersal out of Africa. *Am J Hum Genet* 72:1586-1590.
- Coskun PE, Ruiz-Pesini E, Wallace DC 2003. Control region mtDNA variants: Longevity, climatic adaptation, and a forensic conundrum. *Proc Natl Acad Sci USA* 100:2174-2176.
- Cummins JM 2001. Cytoplasmic inheritance and its implications for animal biotechnology. *Theriogenology* 55:1381-399.
- Černý V, Brůžek J, Brouček J, Hájek M, Brdička R 2003. Archeogenetika: nový přístup k řešení vleklých sporů antropologie? *Archeol Rozhl* 55:561-580.
- Černý V, Hájek M, Bromová M, Čmejla R, Diallo I, Brdička R 2006. mtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Hum Biol* 78:9-27.
- Černý V, Hájek M, Čmejla R, Brůžek J, Brdička R 2004. mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. *Ann Hum Biol* 31:554-569.

- Černý V, Siegllová Z, Brdička R 1997. Molekulární archeologie: aplikace molekulárně genetických metod a archeologii a jejich využití při studiu pravěkých populací. *Archeol Rozhl* 49:526-543.
- Diez-Sanchez C, Ruiz-Pesini E, Lapena AC, Montoya J, Perez-Martos A, Enriquez JA, Lopez-Perez MJ 2003. Mitochondrial DNA Content of Human Spermatozoa. *Biol Reprod* 68:180-185.
- DiRienzo A, Wilson AC 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* 88:1597-1601.
- Dupanloup I, Bertorelle G, Chikhi L, Barbujani G 2004. Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol Biol Evol* 21:1361-1372.
- Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N 2001. Analysis of European mtDNAs for recombination. *Am J Hum Genet* 68:145-153.
- Elson JL, Turnbull DM, Howell N 2004. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am J Hum Genet* 74:229-238.
- Enríquez JA, Cabezas-Herrera J, Bayona-Bafaluy MP, Attardi G 2000. Very rare complementation between mitochondria carrying different mitochondrial DNA mutations points to intrinsic genetic autonomy of the organelles in cultured human cells. *J Biol Chem* 275:11207-11215.
- Eswaran V, Harpending H, Rogers AR 2005. Genomics refutes an exclusively African origin of humans. *J Hum Evol* 49:1-18.
- Excoffier L, Smouse PE, Quattro JM 1992. Analysis of molecular variance inferred from metric distances among dna haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.
- Eyre-Walker A, Smith NH, Smith JM 1999. How clonal are human mitochondria? *Proc R Soc Lond B* 266:477-483.
- Fattovich R 2002. The development of urbanism in the northern Horn of Africa in ancient and medieval times. In: Sinclair P, editors. *The development of urbanism in Africa from a global perspective*. Uppsala: Uppsala Universiteit.

- Fernandes, Carlos A, Rohling, Eelco J, Siddall, Mark 2006. Absence of post-Miocene Red Sea land bridges: biogeographic implications. *J Biogeogr* 33:961-966.
- Filosto M, Mancuso M, Vives-Bauza C, Vilf MR, Shanske S, Hirano M, Andreu AL, DiMauro S 2003. Lack of paternal inheritance of muscle mitochondrial DNA in sporadic mitochondrial myopathies. *Ann Neurol* 54:524-526.
- Finnilä S, Hassinen IE, Ala-Kokko L, Majamaa K 2000. Phylogenetic network of the mtDNA haplogroup U in Northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. *Am J Hum Genet* 66:1017-1026.
- Finnilä S, Lehtonen MS, Majamaa K 2001. Phylogenetic network for European mtDNA. *Am J Hum Genet* 68:1475-1484.
- Forster P 2004. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Phil Trans R Soc Lond B* 359:255-264.
- Forster P, Matsumura S 2005. Did early humans go north or south? *Science* 308:965-966.
- Forster P, Torroni A, Renfrew C, Rohl A 2001. Phylogenetic Star Contraction Applied to Asian and Papuan mtDNA Evolution. *Mol Biol Evol* 18:1864-1881.
- Fu YX 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915-925.
- Garrigan D, Mobasher Z, Severson T, Wilder JA, Hammer MF 2005. Evidence for archaic Asian ancestry on the human X chromosome. *Mol Biol Evol* 22:189-192.
- Giles RE, Blanc H, Cann HM, Wallace DC 1980. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA* 77:6715-6719.
- Hagelberg E 2003. Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve. *Trends Genet* 19:84-90.
- Hall TA 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95-98.
- Hammer MF, Zegura SL 2002. The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. *Annu Rev Anthropol* 31:303-321.

- Harpending HC 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 66:591-600.
- Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 70:1152-1171.
- Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D 2001. Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet* 69:1113-1126.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* 22:1561-1568.
- Howell N, Smejkal CB 2000. Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction. *Am J Hum Genet* 66:1589-1598.
- Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am J Hum Genet* 72:659-670.
- Cherni L, Loueslati BY, Pereira L, Ennafaa H, Amorim A, El Gaaied AB 2005. Female gene pools of Berber and Arab neighboring communities in central Tunisia: Microstructure of mtDNA variation in North Africa. *Hum Biol* 77:61-70.
- Ingman M, Gyllensten U 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean Aborigines. *Genome Res* 13:1600-1606.
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-713.
- Jorde LB, Bamshad MJ 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* 288:1931a.
- Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA 2000. The distribution of human genetic diversity: a

comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 66:979-988.

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9:1331-1334.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the Gate of Tears. *Am J Hum Genet* 75:752-770.

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis KK, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman MW, Cavalli-Sforza LL, Oefner PJ 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373-387.

Kivisild T, Villems R 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* 288:1931a.

Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL 2004. African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr Biol* 13:464-473.

Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* 73:671-676.

Kraytsberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, Clayton DA, Vissing J, Khrapko K 2004. Recombination of human mitochondrial DNA. *Science* 304:981-981.

Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Paabo S, Stoneking M 1999. mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration. *Am J Hum Genet* 64:1166-1176.

Kumar S, Hedrick P, Dowling T, Stoneking M 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* 288:1931a.

- Ladoukakis ED, Eyre-Walker A 2004. Evolutionary genetics - Direct evidence of recombination in human mitochondrial DNA. *Heredity* 93:321-321.
- Lahr MM, Foley RA 1998. Towards a theory of modern human origins: Geography, demography, and diversity in recent human evolution. *Yrbk Phys Anthropol* 41:137-176.
- Landsteiner K 1901. Ueber Agglutinationserscheinungen normalen menschlichen Blutes. *Wien Klin Wochenschr* 14:1132-1134.
- Landsteiner K 1931. Individual differences in human blood. *Science* 73:403-409.
- Landsteiner K, Richter M 1903. Ueber die Verwerthbarkeit individueller Blutdifferenzen für die forensische Praxis. *Zeitschr Med Beamte* 16:85-89.
- Legros F, Lombes A, Frachon P, Rojo M 2002. Mitochondrial fusion in human cells is efficient, requires the inner membrane potential, and is mediated by mitofusins. *Mol Biol Cell* 13:4343-4354.
- Legros F, Malka F, Frachon P, Lombes A, Rojo M 2004. Organization and dynamics of human mitochondrial DNA. *J Cell Sci* 117:2653-2662.
- Lightowlers RN, Chinnery PF, Turnbull DM, Howell N 1997. Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends Genet* 13:450-455
- Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13.
- Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt H-J, Oppenheimer S, Torroni A, Richards M 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034-1036.
- Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonn -Tamir B, Sykes B, Torroni A 1999a. The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232-249.
- Macaulay V, Richards M, Sykes B 1999b. Mitochondrial DNA recombination - no need to panic. *Proc R Soc Lond B* 266:2037-2039.

- May-Panloup P, Chretien MF, Savagner F, Vasseur C, Jean M, Malthiery Y, Reynier P 2003. Increased sperm mitochondrial DNA content in male infertility. *Hum Reprod* 18:550-556.
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar D, Gilbert MT, Endicott P, Mastana S, Papiha S, Skorecki K, Torroni A, Villems R 2004. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5:26.
- Meyer S, von Haeseler A 2003. Identifying site-specific substitution rates. *Mol Biol Evol* 20:182-189.
- Meyer S, Weiss G, von Haeseler A 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152:1103-1110.
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC 2003. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100:171-176.
- Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavallisforza LL 1995. Demographic history of India and mtDNA-sequence diversity. *Am J Hum Genet* 56:979-992.
- Mullis KB, Faloona FA 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. *Methods Enzymol* 155:335-350.
- Nebel A, Landau-Tasseron E, Filon D, Oppenheim A, Faerman M 2002. Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa. *Am J Hum Genet* 70:1594-1596.
- Nei M 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Ono T, Isobe K, Nakada K, Hayashi J-I 2001. Human cells are protected from mitochondrial dysfunction by complementation of DNA products in fused mitochondria. *Nat Genet* 28:272.
- Pääbo S 2003. The mosaic that is our genome. *Nature* 421:409-412.
- Pakendorf B, Stoneking M 2005. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6:165-183.

- Parsons TJ, Irwin JA 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* 288:1931a.
- Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Biol* 65:439-458.
- Petraglia MD, Alsharekh A 2003. The Middle Palaeolithic of Arabia: Implications for modern human origins, behaviour and dispersals. *Antiquity* 77:671-684.
- Piganeau G, Eyre-Walker A 2004. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. *Heredity* 92:282-288.
- Plagnol V, Wall JD 2006. Possible ancestral structure in human populations. *PLoS Genetics* in press.
- Prugnolle F, Manica A, Balloux F 2005. Geography predicts neutral genetic diversity of human populations. *Curr Biol* 15:R159-R160.
- Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K 2004. Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet* 74:827-845.
- Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS 1999. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437-441.
- Ray N, Currat M, Berthier P, Excoffier L 2005. Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Res* 15:1161-1167.
- Relethford JH, Jorde LB 1999. Genetic evidence for larger African population size during recent human evolution. *Am J Phys Anthropol* 108:251-260.
- Renfrew C 1992. Archaeology, genetics and linguistic diversity. *Man NS* 27:445-478.
- Ricchetti M, Tekai F, Dujon B 2004. Continued colonization of the human genome by mitochondrial DNA. *PLoS Biol* 2:e273.

- Richards M 2003. The neolithic invasion of Europe. *Annu Rev Anthropol* 32:135-162.
- Richards M, Macaulay V 2001. The mitochondrial gene tree comes of age. *Am J Hum Genet* 68:1315-1320.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Norby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A, Bandelt HJ 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251-1276.
- Richards M, Macaulay V, Torroni A, Bandelt H-J 2002. In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet* 71:1168-1174.
- Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V, Torroni A 2003. Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am J Hum Genet* 72:1058-1064.
- Rosa A, Brehm A, Kivisild T, Villems R 2004. MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. *Ann Hum Biol* 68:340-352.
- Rose JI 2004. The question of Upper Pleistocene connections between East Africa and South Arabia. *Curr Anthropol* 45:551-555.
- Roychoudhury S, Roy S, Basu A, Banerjee R, Vishwanathan H, Rani MVU, Sil SK, Mitra M, Majumder PP 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum Genet* 109:339-350.
- Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A 2002. The making of the African mtDNA landscape. *Am J Hum Genet* 71:1082-1111.
- Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A 2004. The African diaspora: Mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* 74:454-465.
- Santos C, Montiel R, Angles N, Lima M, Francalacci P, Malgosa A, Abade A, Aluja MP 2004. Determination of human caucasian mitochondrial

- DNA haplogroups by means of a hierarchical approach. *Hum Biol* 76:431-453.
- Serre D, Paabo SP 2004. Evidence for gradients of human genetic diversity within and among continents. *Genome Res* 14:1679-1685.
- Shepard EM, Herrera RJ 2006. Genetic encapsulation among Near Eastern populations. *J Hum Genet* 51:467-476.
- Scheffler IE 2000. A century of mitochondrial research: achievements and perspectives. *Mitochondrion* 1:3-31.
- Schneider S, Roessli D, Excoffier L 2000. Arlequin: A software for population genetics data analysis. Ver 2. 000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.
- Schwartz M, Vissing J 2002. Paternal inheritance of mitochondrial DNA. *N Engl J Med* 347:576-580.
- Schwartz M, Vissing J 2003. New patterns of inheritance in mitochondrial disease. *Biochem Biophys Res Commun* 310:247-251.
- St. John JC, Sakkas D, Barratt CL 2000. A role for mitochondrial DNA and sperm survival. *J Androl* 21:189-199.
- Stone AC, Stoneking M 1999. Analysis of ancient DNA from a prehistoric Amerindian cemetery. *Phil Trans R Soc Lond B* 354:153-159.
- Stoneking M 2000. Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet* 67:1029-1032.
- Stringer C 2000. Coasting out of Africa. *Nature* 405:24-27.
- Stringer C 2002. Modern human origins: progress and prospects. *Phil Trans R Soc Lond B* 357:563-579.
- Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G 2000. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod* 63:582-590.
- Sykes B 2004. *Sedm dcer Eviných. Pramatky Evropanů*. Praha: Paseka.
- Taanman J-W 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1410:103-123.
- Tajima F 1983. Evolutionary relationship of DNA sequences in finite population. *Genetics* 105:437-400.

- Tajima F 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.
- Tajima F 1993. Measurement of DNA polymorphism. In: Takahata N, Clark AG, editors. *Mechanisms of Molecular Evolution. Introduction to Molecular Paleopopulation Biology*. Tokyo, Sunderland MA: Sinauer Associates, Inc., Scientific Societies Press, pp. 37–59.
- Tamura K, Nei M 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512-526.
- Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, Fuku N, Guo L-J, Hirose R, Fujita Y, Kurata M, Shinoda K-i, Umetsu K, Yamada Y, Oshida Y, Sato Y, Hattori N, Mizuno Y, Arai Y, Hirose N, Ohta S, Ogawa O, Tanaka Y, Kawamori R, Shamoto-Nagai M, Maruyama W, Shimokata H, Suzuki R, Shimodaira H 2004. Mitochondrial genome variation in eastern asia and the peopling of Japan. *Genome Res* 14:1832-1850.
- Taylor RW, McDonnell MT, Blakely EL, Chinnery PF, Taylor GA, Howell N, Zeviani M, Briem E, Carrara F, Turnbull DM 2003. Genotypes from patients indicate no paternal mitochondrial DNA contribution. *Ann Neurol* 54:521-524.
- Templeton AR 1993. The "Eve" hypotheses: A genetic critique and reanalysis. *Am Anthropol* 95:51-72.
- Templeton AR 1994. "Eve": Hypothesis compatibility versus hypothesis testing. *Am Anthropol* 96:141-147.
- Templeton AR 2002. Out of Africa again and again. *Nature* 416:45-51.
- Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L 2005. Reconstructing the origin of Andaman Islanders. *Science* 308:996.
- Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, Goldstein DB 2002. Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet* 70:1411-1420.
- Thompson WE, Ramalho-Santos J, Sutovsky P 2003. Ubiquitination of prohibitin in mammalian sperm mitochondria: possible roles in the regulation of mitochondrial inheritance and sperm quality control. *Biol Reprod* 69:254-260.

- Thyagarajan B, Padua RA, Campbell C 1996. Mammalian mitochondria possess homologous DNA recombination activity. *J Biol Chem* 271:27536-27543.
- Tishkoff SA, Kidd KK 2004. Implications of biogeography of human populations for 'race' and medicine. *Nat Genet* 36:S21-S27.
- Torrioni A, Achilli A, Macaulay V, Richards M, Bandelt H-J 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22:339-345.
- Torrioni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozari R, Obinu D, Savontaus M-L, Wallace DC 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835-1850.
- Torrioni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* 69:1348-1356.
- Tourmen Y, Baris O, Dessen P, Jacques C, Malthièry Y, Reynier P 2002. Structure and chromosomal distribution of human mitochondrial pseudogenes. *Genomics* 80:71-77.
- Tully LA, Parsons TJ, Steighner RJ, Holland MM, Marino MA, Prenger VL 2000. A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. *Am J Hum Genet* 67:432-443.
- Underhill PA 2004. Inferring human history: clues from Y-chromosome haplotypes. *Cold Spring Harb Symp Quant Biol* 68:487-493.
- Wallace DC 1994. Mitochondrial DNA sequence variation in human evolution and disease. *Proc Natl Acad Sci USA* 91:8739-8746.
- Walter RC, Buffler RT, Bruggemann JH, Guillaume MM, Berhe SM, Negassi B, Libsekal Y, Cheng H, Edwards RL, von Cosel R, Neraudeau D, Gagnon M 2000. Early human occupation of the Red Sea coast of Eritrea during the last interglacial. *Nature* 405:65-69.
- Watson E, Forster P, Richards M, Bandelt HJ 1997. Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691-704.
- Weir BS, Cockerham CC 1984. Estimating F-statistics for the analysis of population-structure. *Evolution* 38:1358-1370.

- Woischnik M, Moraes CT 2002. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res* 12:885-893.
- Zhang D-X, Hewitt G, M. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol Evol* 11:247-251.
- Žaloudková M, Černý V 2006. Původ egyptské populace z oázy el-Hajez - předběžné výsledky analýzy mtDNA. In: Šajnerová A, editors. *Sborník HMČ UK*. Praha, pp. 91-99.

8 RESUMÉ

Mitochondrial DNA variation in the Yemeni populations: the role of the regional based sampling and the genetic structure of the South Arabia.

Archaeogenetics applies the molecular-genetic methods in the study of the anatomically modern human evolution and the dispersals of the human populations. Due to its features the mitochondrial DNA (mtDNA) is in the particular interest. Yemen played an important role in the migrations of humans as it lies on the southern route between Africa and Eurasia. However, the mtDNA variation of the Yemeni populations has not yet been systematically analyzed. This diploma thesis presents the analysis of the new 150 sequences of the mtDNA HVS-I from the three regions of Yemen – Al-Akchum, Tihama and Hadramawt. Our results indicate the importance of the regional based sampling strategy. We were able to reveal the regional differences in the mtDNA variation of the three samples and detect the different affinities to surrounding populations. Populations from Al-Achkum and Tihama display the proximity with Near East populations, while the Hadramawt sample shows relationships with populations from the East Africa. In addition to our own conclusions, this paper presents the future possibilities of the research of the mtDNA variation in the South Arabia.

Keywords: mtDNA variation – HVS-I – Yemen – regional based sampling – population distances