

**Univerzita Karlova v Praze**

**Přírodovědecká fakulta**

Studijní program: Biologie

Studijní obor: Botanika



**Bc. Lucie Hyková**

Systematika a fylogeografie diploidních zástupců druhového komplexu  
*Arabidopsis arenosa*

Systematics and phylogeography of diploid lineages of *Arabidopsis arenosa* complex

Diplomová práce

Vedoucí práce: Mgr. Eliška Záveská, Ph.D.

Praha, 2016

## Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 15. 8. 2016

Podpis

## Poděkování

Na tomto místě bych v první řadě chtěla poděkovat své školitelce Elišce Záveské za celkové vedení práce, rady a podněty a také za připomínky ke vznikajícímu textu této práce. Dále bych chtěla poděkovat Gábině Fuxové za praktické rady a tipy, které mi jako zkušenější kolegyně předávala při laboratorní práci, a také Filipu Kolářovi za užitečné připomínky, zejména v závěru práce.

Ráda bych poděkovala také své rodině a přátelům, kteří mě během práce povzbuzovali a podporovali.

## Abstrakt

Druhový komplex *Arabidopsis arenosa* je blízce příbuzný modelovému druhu *Arabidopsis thaliana*. Areál rozšíření diploidní linie pokrývá větší část střední a jihovýchodní Evropy s pozoruhodnou diverzitou ekologických nik od mořské hladiny po vysokohorské vrcholky. Genetické vztahy mezi diploidními populacemi *A. arenosa* stále nejsou příliš známé. Pomocí 14 jaderných mikrosatelitových markerů byla prozkoumána evoluční historie diploidního cytotypu napříč celým areálem rozšíření. V rámci komplexu *Arabidopsis arenosa* bylo objeveno 5 linií, které se vyskytují v biogeograficky odlišných oblastech. Tyto linie byly označeny jako Západní Karpaty, Východní Karpaty, Panonská nížina, Dinárské hory a Pobaltí (podle toho, ve kterém geografickém regionu se tyto linie nachází).

## Klíčová slova

*Arabidopsis*, druhový komplex *Arabidopsis arenosa*, fylogeografie, mikrosatelity

## Abstract

The *Arabidopsis arenosa* species complex is closely related to the model species *Arabidopsis thaliana*. Distribution range of the diploid lineage covers most parts of Central and South-eastern Europe with a remarkable diversity of ecological niches from sea level to alpine peaks. There is still little known about genetic relationships among the diploid populations of *A. arenosa*. The evolutionary history of the diploid cytotype across its entire range was explored by using 14 nuclear microsatellite loci. Five lineages occupying biogeographically distinct regions were identified in *Arabidopsis arenosa* complex. These lineages were labelled as Western Carpathians, Eastern Carpathians, Pannonian Lowland, Dinaric Alps and Baltic Shore (according to geographical regions where these lineages occur).

## Keywords

*Arabidopsis*, *Arabidopsis arenosa* complex, phylogeography, microsatellites

## Obsah

1. Úvod	7
1.1. Cíle a otázky	8
2. Literární rešerše	11
2.1. Fylogeneze a fylogeografie	11
2.2. Evoluční procesy (hybridizace, polyploidizace)	13
2.2.1. Hybridizace	13
2.2.2. Polyploidizace	13
2.3. Rod <i>Arabidopsis</i> (DC.) Heynh.	15
2.3.1. Druhový komplex <i>Arabidopsis arenosa</i>	15
2.3.1.1. Problematická taxonomie	17
2.4. Použité metody	23
2.4.1. Molekulární markery	20
2.4.1.1. Mikrosatelity	20
2.4.1.2. RAD sekvenování	22
2.4.2. Metody analýzy molekulárních dat	23
3. Metodika	25
3.1. Sběr vzorků	25
3.2. Laboratorní práce	30
3.2.1. Izolace DNA	30
3.2.2. Analýza mikrosatelitů a optimalizace multiplexu primerů	30
3.2.3. Sekvenování mikrosatelitů	34
3.2.4. Kontrola vzájemné kompatibility primerů	34
3.2.5. PCR a ředění produktů	34
3.2.6. Elektroforéza	35
3.3. Fragmentační analýza a vyhodnocování primárních dat	36
3.4. Analýza mikrosatelitových dat	36
4. Výsledky	38
4.1. Optimalizace multiplexu primerů	38
4.1.1. Kontrola vzájemné kompatibility primerů v Multiplexu 2	39
4.2. Molekulární analýzy	41
4.2.1. Charakteristiky populací <i>Arabidopsis arenosa</i>	42

4.2.2. Analýza hlavních koordinát	49
4.2.3. Bayesovské shlukování podle programu Structure	51
4.2.4. Analýza molekulární variance (AMOVA)	60
5. Diskuze	64
6. Závěr	69
7. Literatura	70

## Přílohy

A. Mapy lokalit sbíraných populací	78
B. Tabulky souhrnných informací o multiplexech primerů	79
C. Laboratorní protokoly	80
C.1. Izolace DNA	80
C.2. Rozpisy pro PCR	81
C.2.1. Reakční mix pro MyTaq polymerázu	81
C.2.2. Reakční mix pro REDTaq polymerázu	81
C.2.3. Reakční mix pro QIAGEN Multiplex PCR Kit	82
C.2.4. Koncentrace primerů v primer-mixu M2	82
C.3. Programy v termocykleru	82
C.3.1. Program pro MyTaq polymerázu	82
C.3.2. Program pro REDTaq polymerázu	83
C.3.3. Program pro QIAGEN Multiplex PCR Kit	83
C.4. Přечиštění octanem	83
C.5. Příprava na fragmentační analýzu	84
C.6. Příprava na sekvenování mikrosatelitů	84
D. Tvorba dimerů v rámci reverse primeru F19G10 pěti různými způsoby	85

## 1. Úvod

*Arabidopsis arenosa* (L.) Lawalrée je druhový komplex z čeledi Brassicaceae, který zahrnuje taxony o různých ploidních úrovních (2x a 4x), které také často hybridizují. Velké množství druhů z čeledi Brassicaceae hybridizuje nebo polyploidizuje, a proto je tato čeleď mimo jiné studována z hlediska evolučních procesů, jako jsou právě hybridizace a polyploidizace (Marhold & Lihová 2006). Příkladem mohou být studie rodu *Cardamine* (Lihová et al. 2007), *Rorippa* (Bleeker 2003), *Arabis* (Karl et al. 2012) a *Draba* (Jordon-Thaden et al. 2010). Jedním z nejstudovanějších druhů z čeledi Brassicaceae je známý modelový druh *Arabidopsis thaliana* (L.) Heynh. (Innan et al. 1997; Bennetzen 1999; Van der Zwan et al. 2000; Mitchell-Olds 2001), který je zároveň blízce příbuzný modelovému druhu *Arabidopsis thaliana* (Al-Shehbaz et al. 1999; Clauss & Koch 2006; Koch & Matschinger 2007). Navzdory skutečnosti, že modelovému druhu byla věnována velká pozornost ze strany vědců zkoumajících jeho genom, informace o jeho nejbližších příbuzných (zejména *Arabidopsis arenosa*) jsou nejednoznačné a chybí zde přehledné shrnutí (Clauss & Koch 2006; Schmickl et al. 2012).

Tito příbuzní druhu *Arabidopsis thaliana* se v posledních dvou dekáдах také přesouvají do popředí zájmu evolučních biologů, protože mohou těžit z velkého množství poznatků, které jsou o modelovém druhu k dispozici, a zároveň značně rozšiřují rozpětí genetické variability a přírodních fenotypů vhodných pro další experimenty (např.: Hunter & Bomblies 2010; Koenig & Weigel 2015). Konkrétní divoce rostoucí druhy rodu *Arabidopsis* byly v nedávné době ustaveny modelovými druhy ke zkoumání evoluce self-compatibility rostlin, adaptace rostlin na lokální podmínky a počínající speciace (např.: *A. lyrata*; Leinonen et al. 2011; Haudry et al. 2012), tolerance vůči těžkým kovům a regulace doby kvetení (např.: *A. halleri*; Aikawa et al. 2010; Meyer et al. 2015).

Zatímco *Arabidopsis arenosa*, což je geneticky nejvariabilnější druh z tohoto rodu (Hohmann et al. 2014), u kterého se navíc v určitých oblastech přirozeně vyskytují společně diploidní a autotetraploidní populace (Schmickl et al. 2012), byl zatím využit pouze pro rozluštění molekulární podstaty adaptace na duplikaci genomu (Yant et al. 2013). Na rozdíl od ostatních druhů *Arabidopsis* je totiž známo velmi málo o jeho rozsáhlém „pattern“ genetické variability a evoluční historie, zvláště u diploidního



cytotypu, což následně omezuje jeho využití pro řešení obecných otázek botaniky, například polyploidizace. Jediná genetická studie zabývající se *A. arenosa* ve velkém ale nikoli celém rozsahu variability, vychází z AFLP (Amplified Fragment Length Polymorphism) a plastidových markerů. Díky nim byly identifikovány oblasti s vyšší genetickou diverzitou, ale nepodařilo se odlišit hlavní evoluční linie uvnitř skupiny (Schmickl et al. 2012). Nedávná úspěšná aplikace RAD sekvenování (Restriction Associated DNA sequencing) pomohla odhalit společný původ jinak široce rozšířených tetraploidních populací (Arnold et al. 2015). Tato studie však byla zaměřena převážně na tetraploidy, takže poskytla pouze omezené informace o genetické struktuře diploidního cytotypu.

Problémem *Arabidopsis arenosa* je neustálené taxonomické pojetí, které se různí od jednoho monofyletického druhu (to používá většina studií, jako např.: Yant et al. 2013; Arnold et al. 2015) až po rozpoznávání deseti různých taxonů (diploidních a tetraploidních druhů nebo poddruhů, často neformálně označených), což je obvykle případ systematicky a lokálně zaměřených studií (Měsíček & Goliášová 2002; Schmickl et al. 2012).

## 1.1. Cíle a otázky

Cílem této diplomové práce je zhodnotit použití mikrosatelitových markerů pro účely fylogeografie druhového komplexu *Arabidopsis arenosa* a porovnat výsledky této klasické metody s moderní metodou RAD sekvenování, která byla použita jako hlavní zdroj informací ve studii Kolář et al. (2016), kde byl analyzován tentýž dataset, který je součástí mé diplomové práce. Dalšími cíli je prozkoumat diverzitu a evoluční vztahy mezi diploidními liniemi komplexu *Arabidopsis arenosa*, analyzovat jeho současné rozšíření a navrhnout taxonomické řešení pro tento komplex. Tato práce je zaměřena pouze na studium diploidů, protože vztahy na diploidní úrovni jsou základem pro pochopení vztahů mezi diploidy a tetraploidy. Zde navržený systém pak může dopomoci k řešení složitějších vztahů i mezi taxony na různých ploidních úrovních.

K zodpovězení evolučních a taxonomických otázek byla využita již zmíněná analýza mikrosatelitů (také tzv. SSRs – Simple Sequence Repeats), tj. úseků DNA bohatých na tandemové repetice, která se využívá zejména pro studium genetických

vztahů v rámci blízce příbuzných druhů, a pro odhad genetických charakteristik na populační úrovni. I přes bouřlivý vývoj nových molekulárních technik využívajících sekvenování nové generace (angl. zkratka NGS – Next Generation Sequencing) v poslední dekádě je klasická metoda mikrosatelitů stále hojně využívána (Säll et al. 2004; López-Vinyallonga et al. 2015). A to zejména díky kombinaci nízké ceny a množství informací, které metoda poskytuje, zejména je-li použito větší množství těchto markerů za použití tzv. multiplexování. Výhodou metody mikrosatelitů jsou zejména kodominantní dědičnost a multialelická povaha dat, která získáváme – tj. že analýzou těchto dat je možné zobrazit všechny přítomné alely jedince (a následně všechny alely v populaci) a také vztahy mezi jedinci a populacemi. Další jeho výhodou je vysoká variabilita na úrovni poddruhů vyplývající z vysoké mutační rychlosti (Tautz 1989). Díky tomu jsou mikrosatelity široce využívány ke studiu genetické variability, struktury a dynamiky na druhové i populační úrovni (Provan et al. 1999; Fér & Hroudová 2009; Harbaugh et al. 2009; Lepais et al. 2013).

Cílem mé diplomové práce je odpovědět na základní evoluční a fylogeografické otázky.

- a) Kolik hlavních evolučních linií je možné rozlišit na diploidní úrovni v rámci komplexu *A. arenosa* na základě studia mikrosatelitů?
- b) Jaká je struktura genetické diverzity těchto evolučních linií v rámci komplexu *A. arenosa*?
- c) Do jaké míry probíhá mezi jednotlivými liniemi genový tok?
- d) Jak se jednotlivé evoluční linie šířily v prostoru a čase, a jak je možné vysvětlit jejich současné rozšíření?
- e) Jak se liší poznatky o genetické struktuře zjištěné na základě studia mikrosatelitů a na základě RAD-sekvenačních dat?



Obrázek 1: Ilustrace *Arabidopsis arenosa*.

Převzato z [https://en.wikipedia.org/wiki/Arabidopsis\\_arenosa](https://en.wikipedia.org/wiki/Arabidopsis_arenosa)

## 2. Literární rešerše

### 2.1. Fylogeneze a fylogeografie

Tato práce se zabývá fylogenezí a fylogeografií druhového komplexu *Arabidopsis arenosa*, jejíž nedílnou součástí je snaha o rekonstrukci a pochopení vztahů mezi populacemi tohoto druhu. Populace *A. arenosa* jsou vysoce geneticky diverzifikované, což vedlo v minulosti k popisu několika samostatných taxonů v rámci druhu, obvykle na úrovni poddruhů (jako např.: *Arabidopsis borbasii* subsp. *carpatica*, *Arabidopsis petrogena* subsp. *petrogena* a *Arabidopsis nitida*), které byly navíc často popsány jen formálně, a nebyly publikovány. Tato klasifikace *Arabidopsis arenosa* do (více či méně formálních) subspecií však není jednoznačná a navíc ji komplikuje existence přechodných morfotypů u některých populací, což může být způsobeno hybridizací nebo polyploidizací. Míra hybridizace a polyploidizace však nebyla u *Arabidopsis arenosa* nikdy cele popsána – právě naopak, většina studií je zaměřena velice konkrétně například na hybridizaci mezi diploidním *Arabidopsis thaliana* a tetraploidním *A. arenosa* za vzniku *A. suecica* (Kamm et al. 1995; Bushell et al. 2003) nebo na adaptace související s duplikací celého genomu (Hollister et al. 2012; Yant et al. 2013). Recentní studie (Kolář et al. 2016), ve které jsou zahrnuty i výsledky této diplomové práce, se zaměřuje právě na rekonstrukci vztahů mezi diploidními zástupci v rámci komplexu *Arabidopsis arenosa*. Cílem této studie bylo zkoumání genetické struktury diploidního cytotypu *Arabidopsis arenosa* napříč jeho celým areálem rozšíření a testování korelace mezi genetickou strukturou a ekologií druhu. K tomu byly využity následující molekulární markery: celogenomové SNPs (Single Nucleotide Polymorphisms) a jaderné mikrosatelity. Z takto získaných dat byly zrekonstruovány fylogenetické vztahy, které odhalily rozsáhlý „pattern“ genetické diverzity a naznačily evoluční historii diploidního cytotypu (Kolář et al. 2016).

Jedním z nejdůležitějších průlomů ve výzkumu biogeografie byl rozvoj molekulárních technik a specifických markerů nezkreslených vlivem vnějšího prostředí na studované jedince (Schaal et al. 1998). Zjištění, že geografické rozmístění koresponduje nejen s rozložením genetické diverzity ale také s genetickými

vzdálenostmi, umožnilo vznik nové oblasti na poli vědy: fylogeografie (Avice et al. 1987).

Fylogeografie se zabývá popisem a studiem diferenciací populací na základě jejich geografického rozšíření a jejich genetické příbuznosti v rámci druhu (příp. druhového komplexu) a zkoumá šíření taxonů v prostoru a čase, jakož i míru genového toku mezi populacemi i mezi vyššími taxonomickými jednotkami, mezi kterými se neutvořila kompletní reprodukční bariéra (Avice et al. 1987). Šíření taxonu může být ovlivněno strukturou krajiny, složením půdy, kompeticí a klimatem. Podle Milankovičovy teorie se během pleistocénu střídala teplá období s dobami ledovými jako důsledek pravidelného kolísání oběžné dráhy Země kolem Slunce (Imbrie 1985). Poslední doba ledová v severní části Evropy a Severní Ameriky, jejíž maximum (LGM – Last Glacial Maximum) nastalo v období před 18000 až 24000 let, silně ovlivnila současnou distribuci druhů (Hewitt 2000).

Všeobecně se tedy předpokládá, že současné vzorce geografického rozložení genetické variability na severní polokouli byly významně ovlivněny klimatickými změnami, ke kterým docházelo v průběhu pleistocénu. Díky migracím v dobách ledových se měnily původní areály rozšíření rostlin (například latitudinálními i altitudinálními posuny); obvykle byly omezeny na menší plochu nebo roztrženy na malé dílčí areály, což vedlo obecně k diverzifikaci populací. Během rekolonizací (v době interglaciálů) pak navíc často docházelo k sekundárnímu kontaktu mezi těmito roztrženými a původními populacemi (Koch & Kiefer 2006). Tyto změny v distribuci organismů (posuny, expanze, fragmentace i extinkce) sehrály významnou roli při formování současných genetických „patterns“ mezi druhy i v rámci jednotlivých druhů (Taberlet et al. 1998; Hewitt 2000), takže dnešní rozmístění druhů nereflektuje pouze jejich preference určitých habitatů ale také historii rekolonizace území.

## 2.2. Evoluční procesy (hybridizace, polyploidizace)

### 2.2.1. Hybridizace

Hybridizace je proces, při kterém dochází ke křížení dvou geneticky více či méně odlišných druhů. Mezidruhová hybridizace často přispívá ke zvýšení genetické variability a může vést ke vzniku nových taxonů (Marhold & Lihová 2006). Jedná se o tzv. sympatrickou speciaci, což je způsob vzniku nového druhu ve stejném čase a prostoru, ve kterém se vyskytuje i druh původní (Dieckmann & Doebeli 1999) – v tomto případě oba rodičovské druhy vzniklého hybrida.

K hybridizaci běžně dochází mezi taxony o stejné ploidní úrovni (typicky se stejným základním chromosomovým číslem, ale i s odlišnými chromosomovými čísly), a potom se jedná o tzv. homoploidní hybridizaci. Pokud dochází k hybridizaci mezi taxony na různé ploidní úrovni, mluvíme o tzv. heteroploidní hybridizaci nebo také allopolyploidizaci. Z toho vyplývá, že hybridizace a polyploidizace jsou jevy, které spolu velmi úzce souvisejí (Hegarty & Hiscock 2005).

V rodě *Arabidopsis* je hybridizace poměrně častým jevem (Schmickl & Koch 2011). V této studii Schmickl & Koch (2011) byl odhalen přirozeně se vyskytující hybrid mezi tetraploidními zástupci *Arabidopsis arenosa* a *A. lyrata* (*A. arenosa* – donor pylu, *A. lyrata* – akceptor pylu), který vykazuje přechodné morfologické znaky mezi oběma rodičovskými druhy. Dobře doloženým případem hybridizace mezi již vymezenými druhy v rámci rodu *Arabidopsis*, kdy navíc dochází ke speciaci, je druh *Arabidopsis suecica* (Fries) Norrlin, Meddel., což je produkt hybridizace mezi diploidním *Arabidopsis thaliana* a tetraploidním *A. arenosa*, kde *A. arenosa* je opět donorem pylu a *A. thaliana* jeho akceptorem (Mummenhoff & Hurka 1994; Jakobsson et al. 2006).

### 2.2.2. Polyploidizace

Polyploidizace je odborný termín označující proces pomnožení chromosomových sad organismů. Takové organismy se pak nazývají polyploidní jedinci nebo také jednoduše „polyploidi“. Polyploidizace je dalším z mechanismů sympatrické speciace, tj. vzniku biologických druhů vyskytujících se ve stejném prostoru, a zvláště u rostlin je poměrně běžná (Otto & Whitton 2000). Oproti jiným speciálním

mechanizmům dochází při polyploidizaci k okamžité změně, tj. nově vzniklý polyploid je od počátku izolovaný od rodičovského druhu a při zpětném křížení vznikají nefertilní jedinci (Jakobsson et al. 2006).

U polyploidizace rozlišujeme dva základní typy – autopolyploidizace a allopolyploidizace – které se liší mechanismem vzniku polyploidů (Marhold & Lihová 2006). Autopolyploid vzniká zmnožením vlastního genomu, zatímco allopolyploid křížením mezi odlišnými druhy, tj. hybridizací různých genomů následovanou polyploidizací (Ramsey & Schemske 2002). Autopolyploidie je v rostlinné říši běžnější, než se všeobecně předpokládá, ale je jí věnována menší pozornost ze strany vědců než allopolyploidii (Parisod et al. 2010). Obvykle se studie autopolyploidů zaměřují na různé adaptace spojené s duplikací genomu (angl. Whole Genome Duplication – WGD), jako například Yant et al. (2013) nebo Hollister et al. (2012). Dnes se předpokládá, že opakované duplikace celého genomu sehrály důležitou roli v evoluci rostlin (Yant et al. 2013).

V čeledi Brassicaceae zcela běžně dochází k mezidruhovému genovému toku (hybridizaci) a poměrně často se zde vyskytují polyploidní zástupci, což naznačuje, že procesy polyploidizace a hybridizace sehrály zásadní roli při speciaci v této čeledi (Marhold & Lihová 2006). Předpokládaný počet polyploidních druhů činí odhadem 37 % (Warwick & Al-Shehbaz 2006). Tento odhadovaný podíl může být vyšší, vezmeme-li v úvahu i druhy, u nichž došlo k duplikaci genomu před dlouhou dobou, tzv. paleopolyploidy. Potom je odhadovaný podíl polyploidních druhů až 50 % (Koch et al. 2003). Příkladem mohou být paleohexaploidní druhy z rodu *Brassica* (Marhold & Lihová 2006). Co se týče rodu *Arabidopsis*, ten se oddělil od tribu *Brassicaceae* ještě předtím, než došlo u *Brassicaceae* k hexaploidizaci. K rozdělení došlo někdy v období před 7,9 až 14,6 miliony let (Lysak et al. 2005). Navzdory faktu, že polyploidizace je v rodě *Arabidopsis* poměrně běžným jevem, bylo na ni doposud zaměřeno jen málo pozornosti. U rodu *Arabidopsis* se setkáváme s několika v přírodě se vyskytujícími tetraploidními druhy. Jedná se o stabilní allopolyploidní druhy *A. suecica* (Fries) Norrlin, Meddel. a *A. kamchatica* (Fisch. ex DC.) O’Kane & Al-Shehbaz (Schmickl & Koch 2011). Dalším případem přirozeně se vyskytujících tetraploidních populací je druhový komplex *Arabidopsis arenosa*, který zahrnuje diploidní i tetraploidní taxony (Měsíček 1970; Schmickl et al. 2012). U těchto taxonů není dosud známo, jestli se jedná o autopolyploidy nebo allopolyploidy, ani jestli vznikli jednou v historii nebo opakovaně (Claus & Koch 2006).

### 2.3. Rod *Arabidopsis* (DC.) Heynh.

Jméno *Arabidopsis* bylo původně navrženo de Candollem v roce 1821 jako sekce rodu *Sisymbrium* L. Roku 1842 pozvedl Heynhold taxon *Arabidopsis* na úroveň biologického rodu a za jeho jediný druh označil *Arabidopsis thaliana*, bez ohledu na fakt, že tento druh vůbec nebyl zahrnut v původní de Candollově pojetí sekce *Arabidopsis*. To by ve výsledku znamenalo, že jméno *Arabidopsis* by bylo pouhým synonymem dříve publikovaných názvů. Avšak vzhledem k významu *Arabidopsis thaliana* pro rostlinnou biologii v dnešní době se taxonomové dohodli na definitivním řešení – tj. zachování tohoto jména v nomenklatuře navzdory jeho komplikované historii. K tomuto „zakonzervování“ došlo v roce 1988 v Berlíně (Al-Shehbaz & O’Kane 2002).

Podle studie O’Kane & Al-Shehbaz (1997) je rod *Arabidopsis* rozlišován na 9 druhů, které jsou shrnuty následovně:

- *Arabidopsis arenosa* (L.) Lawalrée
- *Arabidopsis cebennensis* (DC.) O’Kane & Al-Shehbaz
- *Arabidopsis croatica* (Schott) O’Kane & Al-Shehbaz
- *Arabidopsis halleri* (L.) O’Kane & Al-Shehbaz
- *Arabidopsis lyrata* (L.) O’Kane & Al-Shehbaz
- *Arabidopsis neglecta* (Schultes) O’Kane & Al-Shehbaz
- *Arabidopsis pedemontana* (Boissier) O’Kane & Al-Shehbaz
- *Arabidopsis suecica* (Fries) Norrlin
- *Arabidopsis thaliana* (L.) Heynhold

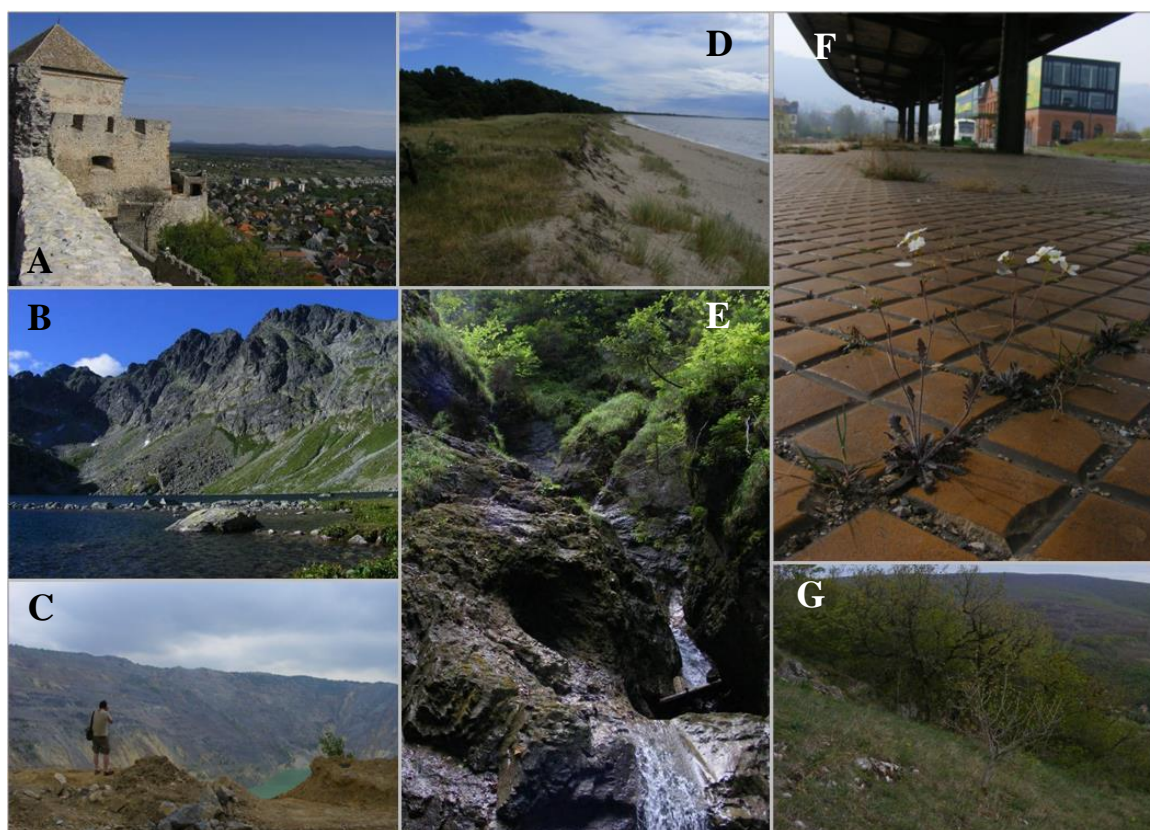
#### 2.3.1. Druhový komplex *Arabidopsis arenosa*

*Arabidopsis arenosa* agg. je nejdiverzifikovanějším druhovým komplexem v rámci rodu *Arabidopsis*. Zahrnuje jednoleté, dvouleté i krátkověké vytrvalé rostliny, diploidní i tetraploidní zástupce (Al-Shehbaz & O’Kane 2002). Diploidní jedinci mají 16 chromozomů ( $2n = 16$ ) a tetraploidní mají 32 chromozomů ( $2n = 32$ ) (Měsíček 1970). Oba cytotypy, diploidní i tetraploidní, vykazují překvapující diverzitu habitatů od mořského pobřeží po alpská stanoviště (rozpětí cca 2500 výškových metrů), a zároveň od submediteránního Balkánského poloostrova až po subarctickou severní část Skandinávie – rozpětí cca 3000 km zeměpisné šířky (Hoffmann 2005, Kolář et al. 2015).



Taková ekotypová variabilita je pozoruhodná a v celém rodě *Arabidopsis* ojedinělá (pouze s výjimkou *Arabidopsis kamchatica*), což naznačuje velký potenciál pro využití *Arabidopsis arenosa* například ke studiu lokálních adaptací.

Areál rozšíření *Arabidopsis arenosa* zahrnuje celou temperátní oblast Evropy a má tři centra genetické diverzity – jedno v Karpatech, druhé na Balkánském poloostrově a třetí ve východní až jihovýchodní části Alp (Schmickl et al. 2012). Distribuce diploidních a tetraploidních cytotypů se překrývá v oblastech východních a západních Karpat, v menší míře pak v předhůří Julských Alp na území Slovinska. Tato zjištění týkající se rozšíření komplexu *A. arenosa* a jeho cytotypů v Evropě byla založena na rozsáhlém samplingu *A. arenosa* skrze celý areál jeho rozšíření a lze je považovat za dosud nejucelenější (Kolář et al., 2015). Diploidní zástupci *A. arenosa* se vyskytují převážně ve střední, severní a východní Evropě, nerostou v podmínkách Mediteránu ani v příliš chladných severních oblastech. V severních oblastech výskytu převažují tetraploidní jedinci, což vyplynulo z terénního sběru populací po celé Evropě (Kolář et al. 2015). Diploidní linie se vyskytuje hlavně na přirozených habitatech ve vyšších nadmořských výškách (Al-Shehbaz & O’Kane 2002). Mezi primární stanoviště *Arabidopsis arenosa* patří rokliny a vlhké stinné skály (jako jsou například ve Slovenském ráji). Ve vyšších nadmořských výškách roste na horských trávnících, subalpínských karech a na březích horských jezer. Sekundárně roste na antropicky ovlivněných stanovištích, jako jsou například zříceniny hradů nebo místa podél železničních tratí a silnic. Tato stanoviště jsou preferována zejména tetraploidními zástupci (Kolář et al. 2015). Typická stanoviště *A. arenosa* jsou pro ilustraci vyobrazena na obrázku 2. Zástupce komplexu *Arabidopsis arenosa* lze nalézt ještě na jednom typu stanovišť, poněkud odlišném od výše jmenovaných, a to jsou písčné duny na mořském pobřeží – jedná se o populace z pobřeží Baltského moře (Al-Shehbaz & O’Kane 2002; Schmickl et al. 2012; Kolář et al. 2015). Z hlediska substrátu *Arabidopsis arenosa* preferuje bazické podloží, vápenaté nebo křemičité. Jeho areál výskytu nicméně zahrnuje i pohoří s kyselým podložím, ale v takovém případě roste pouze na bazických výchozech (Clauss & Koch 2006).



Obrázek 2: Primární a sekundární stanoviště obývaná diploidními zástupci *Arabidopsis arenosa*. Přírodní stanoviště: B (břeh horského jezera), D (mořské pobřeží), E (stinné a vlhké skalní štěrbin) a G (kamenité stepi). Stanoviště ovlivněná lidskou činností: A (zřícenina hradu), C (opuštěný lom) a F (chodník).

Fotky převzaty z práce Kolář et al. 2015.

### 2.3.1.1. Problematická taxonomie

Druhový komplex *Arabidopsis arenosa* v současné době zahrnuje skupinu nedokonale definovaných diploidních a tetraploidních taxonů, které jsou díky široké škále existujících morfotypů mnohdy nejednoznačně popsány, případně jsou popsány pouze neformálně (Koch & Matschinger 2007; Schmickl et al. 2012). Dále se budu zabývat pouze diploidními taxony, protože situace mezi tetraploidními taxony je velmi komplikovaná a k jejímu pochopení je třeba jasné porozumění problematice diploidů. Na problematiku diploidů je právě zaměřena tato diplomová práce a níže jsou popsány pouze známé diploidní taxony.

Jedním z možných vysvětlení vysoké morfologické a karyologické diverzity, která vedla k popsání mnoha druhů v rámci skupiny *Arabidopsis arenosa*, může být

glaciální historie daného území, jehož větší část nebyla v průběhu pleistocénu nikdy zaledněna. Díky tomu zde mohlo přežít množství izolovaných populací v refugiích, které se adaptovaly na lokální podmínky, a daly vzniknout lokálně odlišným formám (Měsíček 1970). K celkové komplexitě skupiny přispívá i vnitrodruhová hybridizace a opakovaná duplikace genomu (Marhold & Lihová 2006).

Hlavní problematika spočívá v tom, že některé klasifikační systémy uznávají určité taxonomické jednotky za odlišné druhy, zatímco jiné je považují pouze za poddruhy druhu *Arabidopsis arenosa*. V dnešní době je tedy druhový komplex *A. arenosa* členěn na dva až šest druhů nebo poddruhů, což záleží na přístupu zvoleném vědci.

Karpatské populace byly svého času velmi intenzivně studovány Měsíčkem, což vedlo k rozlišení čtyř diploidních druhů na základě jejich odlišných morfologických znaků (Měsíček 1970; Měsíček 1998):

- *Arabidopsis borbasii* subsp. *carpatica*
- *A. nitida*
- *A. petrogena* subsp. *petrogena*
- *A. neglecta*

Nicméně první tři z těchto taxonů nebyly formálně popsány a jejich názvy publikovány, a tudíž nemohou být plně uznávány vědeckou veřejností. Také nebylo provedeno žádné zhodnocení jejich vztahů k ostatním druhům v rámci komplexu pomocí moderních molekulárních a cytologických metod. Posléze definovali O’Kane & Al-Shehbaz (1997) na základě studia DNA pouze 2 poddruhy:

- *Arabidopsis arenosa* (L.) Lawalrée subsp. *arenosa*
- *A. arenosa* (L.) Lawalrée subsp. *borbasii* (Zapalowicz) O’Kane & Al-Shehbaz

Nejrecentnější studie Schmickl et al. (2012), která si kladla za cíl poprvé analyzovat druhový komplex z hlediska jeho genetické diverzity a vztahů shrnuje přehled popsaných diploidních linií takto:

- *Arabidopsis arenosa* (L.) Lawalrée subsp. *arenosa*
- *Arabidopsis carpatica*, nom. prov.

- *Arabidopsis croatica* (Schott) O’Kane & Al-Shehbaz
- *Arabidopsis neglecta* (Schultes) O’Kane & Al-Shehbaz  
subsp. *neglecta*
- *Arabidopsis nitida*, nom. prov.
- *Arabidopsis petrogena* (A. Kern) V. I. Dorof. subsp. *petrogena*

Pracovní skupina, která se zabývá studiem *Arabidopsis arenosa* agg. na naší katedře (Přírodovědecká fakulta Univerzity Karlovy, Katedra botaniky), používá pro přehlednější práci a snazší kategorizaci vzorků toto zjednodušené členění poddruhů *Arabidopsis arenosa*: *petrogena*, *carpatica*, *nitida*, *neglecta* a *arenosa* (viz tabulka 1). Některé z těchto provizorních poddruhů jsou sice ekologicky nezaměnitelné, jako například *A. arenosa* subsp. *neglecta*, což je vysokohorský typ *A. arenosa* vyskytující se pouze v Karpatech, s typicky růžovými květy (oproti běžnějším bělokvetým formám), méně dělenými listy a celkově drobnějším vzrůstem (obrázek 3). Většinou jsou ale jen těžko jednoznačně popsateľné, a dané členění neodpovídá genetické struktuře celého komplexu.

Je tedy zjevné, že pro objasnění vztahů mezi taxony v rámci komplexu *Arabidopsis arenosa*, je třeba detailně prozkoumat vzájemnou genetickou podobnost velkého množství jedinců.



Obrázek 3: Vysokohorský typ *A. arenosa* rozeznávaný jako *A. arenosa* subsp. *neglecta*.

Foto z archivu Filipa Koláře.

## 2.4. Použité metody

### 2.4.1. Molekulární markery

Molekulární markery podávají informace o organismu, získané analýzou jeho molekul DNA nebo proteinů. Markerem se rozumí určitá konkrétní část informace, která je cíleně nebo i náhodně vybrána z celkové informace. Molekulární markery tedy vypovídají o genetické podobnosti (respektive příbuznosti) studovaných jedinců, případně celých populací nebo skupin organismů.

V dnešní době existuje řada molekulárních markerů, které se mezi sebou liší různými kritérii (např.: spolehlivostí, mírou komplikovanosti metody, množstvím vzorku DNA vstupujícím do reakcí a v neposlední řadě také množstvím peněz, které daná metoda vyžaduje). Pro každou studii je vždy nutné vybrat ten nejoptimálnější, protože v současné době neexistuje marker, který by byl naprosto univerzální a ideální. Pro systematické studie zabývající se evolucí vyšších rostlin je jednou z mnoha možností metoda analýzy mikrosatelitů (Slatkin 1995).

Ke studiu fylogeografie v nedávné minulosti (tj. pozdní pleistocén a holocén) jsou vhodné kombinace sekvenčních metod (např. cpDNA; Schmickl & Koch 2011) a hypervariabilních markerů jako je AFLP (např. Schmickl et al. 2012) nebo právě mikrosatelity použité v této práci, případně moderní NGS metody jako je RAD-sekvenování. Protože paralelně s analýzou mikrosatelitů byl komplex *A. arenosa* studován metodou RAD sekvenování (Kolář et al. 2016) a výsledky těchto analýz jsou použity pro srovnání s výsledky analýzy mikrosatelitů, je dále krátce popsána i tato metoda.

#### 2.4.1.1. Mikrosatelity

Mikrosatelity jsou tandemové repetice určitých oligonukleotidových motivů, skládající se minimálně z jednoho a maximálně ze šesti párů bází – nejčastěji se však jedná o di-, tri- a tetranukleotidy. Tyto repetitivní oblasti se nacházejí napříč celým genomem rostlin (jaderným i chloroplastovým). Běžně se k jejich označení užívá anglická zkratka SSRs – Simple Sequence Repeats, nebo také starší spojení STRs – Short Tandem Repeats (Wang et al. 1994).

Sekvence mikrosatelitů vznikají *in vivo* pravděpodobně sklouzáváním (tzv. slippage) DNA polymerázy po řetězci DNA při jeho replikaci. Počet sklouznutí polymerázy v průběhu replikace je příčinou pozorovaného délkového polymorfismu mezi mikrosatelitovými alelami (Schlötterer & Tautz 1992). To znamená, že jednotlivé alely jsou charakterizovány počtem opakování daného motivu, a tedy různě dlouhé.

Metoda analýzy mikrosatelitů byla hojně využívána v 90. letech minulého století (Bell & Ecker 1994; Innan et al. 1997; Peakall et al. 1998; Provan et al. 1999). Tato metoda je tedy poněkud starší, nicméně mezi její hlavní výhody patří spolehlivost a snadná reprodukovatelnost výsledků, vysoká míra variability a kodominantní dědičnost alel. Průběh přípravy vzorků spočívá v amplifikaci vybraných mikrosatelitových lokusů pomocí PCR a následné fragmentační analýze na kapilárním sekvenátoru, kdy se nezaznamenávají celé sekvence lokusů, pouze jejich délky. Jsou-li navíc lokusy o jiném rozsahu délek alel označeny jinou značkou (typicky pomocí fluorescenčního značení primerů), je možné v jediné reakci analyzovat i více než deset různých mikrosatelitových lokusů (tzv. multiplexování). Proto je tato metoda také relativně finančně nenáročná, jakmile jsou získány druhově specifické primery.

Hlavní nevýhodou této metody je vysoká druhová specifita tzv. flanking regions, což jsou sekvence ohraničující danou repetitivní oblast. To je důvod, proč tato metoda není univerzální, a je vždy nutné pro každý studovaný druh vyvinout konkrétní primery pro amplifikaci mikrosatelitových sekvencí (Zane et al. 2002). Primery lze získat například sekvenováním genomu nebo prohledáním databází již publikovaných primerů, optimálně používaných pro nějaký blízce příbuzný druh tomu, který hodláme zkoumat.

Pro účely této diplomové práce byly použity mikrosatelitové primery ze studie Clauss et al. (2002), ve které bylo na základě sekvencí genomu *Arabidopsis thaliana* vyvinuto 42 primerů, které byly schopné amplifikace mikrosatelitů v genomu jeho blízce příbuzných druhů (*A. halleri*, *A. lyrata*, *A. petraea*) a jednoho fylogeneticky vzdálenějšího druhu (*Arabis drummondii*). Tyto mikrosatelitové markery schopné docela úspěšné mezidruhové amplifikace (79 až 100 %) jsou užitečné pro další výzkumy ekologie a evoluce v rámci rodu *Arabidopsis* a dalších blízce příbuzných zástupců čeledi Brassicaceae (Clauss et al. 2002). Proto z nich byly vybrány primery pro použití v této diplomové práci k aplikaci na genom *Arabidopsis arenosa*.



#### 2.4.1.2. RAD sekvenování

RAD (Restriction site Associated DNA) sekvenování je metoda, pomocí které lze analyzovat velké množství úseků DNA (desetitisíce i více) pocházející z různých částí jaderného i chloroplastového genomu, a které mohou mutovat buď rychle (jako např. mikrosatelity), nebo pomaleji (např. sekvence kódující geny). Pro sekvenování takového množství fragmentů se využívají metody NGS (Next Generation Sequencing), tzv. sekvenování nové generace (Davey et al. 2011). Tato metoda tak poskytuje data s velkým rozlišením vhodná pro studium fylogeografie. Navíc o studovaném organismu nemusíme mít předem žádné zásadní znalosti týkající se struktury genomu a při určité optimalizaci protokolu lze, navzdory finanční náročnosti NGS sekvenování, analyzovat velké množství jedinců a populací (stovky jedinců) za cenu srovnatelnou s klasickými metodami jako je AFLP (Záveská, osobní sdělení).

Tato metoda je založena na principu fragmentace DNA pomocí restriktivního enzymu zkombinované s PCR amplifikací určitých vybraných genomických fragmentů. Fragmenty jsou následně sekvenovány na některé z platforem NGS (nejčastěji na sekvenátorech Illumina). Výsledná čtení (tzv. reads) představují cca 100 – 250 bp dlouhé sekvence (podle typu zvolené sekvenační techniky), které jsou porovnávány mezi jedinci. Nalezené rozdíly v sekvencích mezi jedinci se označují SNP (= Single Nucleotide Polymorphism) a jsou dále používány ve fylogeografických analýzách, které se využívají i pro vyhodnocování tradičnějších dat, jako je např. analýza hlavních koordinát (PCoA), Bayesovské shlukování a podobné (Baird et al. 2008).

Metoda RAD sekvenování nachází uplatnění při zkoumání různých divoce se vyskytujících („nemodelových“) organismů, u kterých není k dispozici referenční genom (Emerson et al. 2010; Hohenlohe et al. 2010; Baxter et al. 2011; Bus et al. 2012). Na rozdíl od datasetů pro analýzy fylogenetických stromů, které obsahují relativně dlouhé sekvence menšího počtu genů, obsahují data generovaná metodou RAD sekvenování spíše kratší sekvence ale z mnohem většího množství lokusů. Tato restriktivní místa jsou konzervována napříč vzorky, takže má metoda RAD sekvenování celkově velký potenciál k detekování velkého množství SNPs v celém genomu (Eaton & Ree 2013). Z toho důvodu bývá většinou aplikována ve studiích na populační úrovni (Baird et al. 2008; Emerson et al. 2010; Baxter et al. 2011), ačkoli v poslední době je zkoušena její užitečnost i pro fylogenetické studie (např.: Rubin et al. 2012). Zvláště atraktivní je RAD

sekvenování pro systematické studie blízké příbuzných druhů, protože má potenciál například k odhalení současné nebo historické introgrese (Twyford & Ennos 2012).

Protokoly k metodě RAD sekvenování a základní informace jsou volně k dispozici na webových stránkách <https://www.wiki.ed.ac.uk/display/RADSequencing/Home>.

#### 2.4.2. Metody analýzy molekulárních dat

Primární data získaná molekulárními metodami je nutné dále analyzovat pomocí různých statistických metod, které umožní popsat charakteristiky studovaných populací či jejich vzájemné vztahy. K základním popisným parametrům populací patří očekávaná heterozygosita, pozorovaná heterozygosita, frekvence alel a celkové počty alel. Dále lze populace charakterizovat mírou inbreedingu vyjádřenou jako index inbreedingu ( $F_{IS}$ ), který je spočítán na základě frekvencí nulových alel (Chybicki & Burczyk 2009). Informace o dlouhodobé izolovanosti populací v lokalitě lze usuzovat z přítomnosti vzácných alel, protože v izolovaných populacích se hromadí mutace. Tyto vzácné alely vyjadřuje tzv. index rarity neboli DW index („frequency down-weighted marker“). Ten se zjišťuje jako počet výskytů alely v populaci ku počtu výskytů alely v celém datovém souboru (Schönswetter & Tribsch 2005).

K vyjádření prostorových vztahů mezi jedinci a populacemi (zjištění genetické struktury dat), lze využít dvou typů analýz – hierarchických a nehierarchických. Mezi nehierarchické metody patří mnohorozměrná ordinační metoda PCoA (Principal Coordinate Analysis) a Bayesovské shlukování („klastrování“) jedinců (Falush et al. 2003; Corander et al. 2013). Metoda PCoA (neboli analýza hlavních koordinát) funguje tak, že redukuje n-rozměrný prostor molekulárních znaků do 2 nebo 3 rozměrů a zobrazuje jejich vzájemné euklidovské vzdálenosti (odrážející vztahy mezi nimi) na ordinačním diagramu. Analýza hlavních koordinát je užitečná v případech, kdy máme výlučně binární znaky, kvalitativní znaky nebo smíšená data. Bayesovské shlukování zjišťuje nejpravděpodobnější model rozdělení jedinců do určitého počtu shluků. Tento nejpravděpodobnější počet (tzv. K) odpovídá počtu ancestrálních populací, ze kterých byly odvozeny populace pozorované v současnosti. Tím, že analýza vybírá toto nejpravděpodobnější K bez apriorních informací o vztazích mezi populacemi, můžeme tuto metodu považovat za explorativní. V případě programu Structure není hodnota K předem určena, toto číslo odpovídá optimálnímu počtu shluků nebo populací, do kterých jsou všichni jedinci rozděleni (Pritchard et al. 2000; Falush et al. 2003).



Ke zjištění rozdělení genetické variability uvnitř populací, mezi jednotlivými populacemi a také mezi skupinami populací se nejčastěji využívá analýza molekulární variance – AMOVA (Excoffier et al. 1992) a F-statistika (Weir & Cockerham 1984). Obě tyto metody se řadí mezi hierarchické a poskytují informace o tom, do jaké míry jsou jednotlivé populace (nebo skupiny populací) od sebe odlišné a izolované, jak moc mezi nimi dochází k výměně genetické informace a jestli jsou v některé z populací fixované unikátní alely nebo se všechny alely rovnoměrně vyskytují ve všech populacích.

### 3. Metodika

#### 3.1. Sběr vzorků

Vzorky všech populací byly sesbírány pracovní skupinou K. Marholda a F. Koláře v letech 2011 až 2013. Terénní sběr vzorků byl prováděn napříč celým areálem rozšíření *Arabidopsis arenosa*, konkrétně na území Slovenska, Polska, Chorvatska, Slovinska, Rumunska, Maďarska, Bosny a Hercegoviny, Srbska, Litvy a Německa. Sbírané populace se nacházely v nadmořských výškách od 1 metru (pobřeží Baltu) do 1823 m n. m. (populace ve Vysokých Tatrách), na primárních i sekundárních stanovištích, aby sběr reprezentoval co nejpřesněji variabilitu komplexu.

V terénu byly sbírány vzorky všech populací a následně byla v laboratoři průtokové cytometrie při Katedře botaniky Přírodovědecké fakulty UK určena jejich ploidie pomocí průtokové cytometrie (FCM), opět členy pracovní skupiny K. Marholda a F. Koláře. Tato metoda spočívá ve fluorescenčním měření celkového obsahu jaderné DNA ve vzorku (Kron et al. 2007). Ke zpracování je potřeba pouze malá část rostliny, což je výhodné vzhledem k faktu, že jedinci *Arabidopsis arenosa* nedorůstají příliš velkých rozměrů, a ještě pak bylo nutné ze vzorků izolovat DNA.

Pro analýzu mikrosatelitů bylo vybráno 64 diploidních populací s celkovým počtem 592 jedinců. Původně bylo v plánu zanalyzovat vždy 10 jedinců z každé populace, ale ukázalo se, že mnohé populace byly příliš malé a sestávaly třeba jen ze čtyř nebo pěti jedinců. Detaily o analyzovaných populacích jsou uvedeny v tabulce 1 a v mapce na obrázku 4 (nebo detailnější pohled v příloze A).

Tabulka 1: Přehled studovaných populací, jejich určení podle provizorní taxonomie, přiřazení k jedné z pěti genetických skupin (na základě výsledků ze Structure) a popis lokalit, ze kterých populace pochází.

ZKarp = Západní Karpaty, Pan = Panonská nížina, Balt = Pobaltí, Din = Dinárské hory,

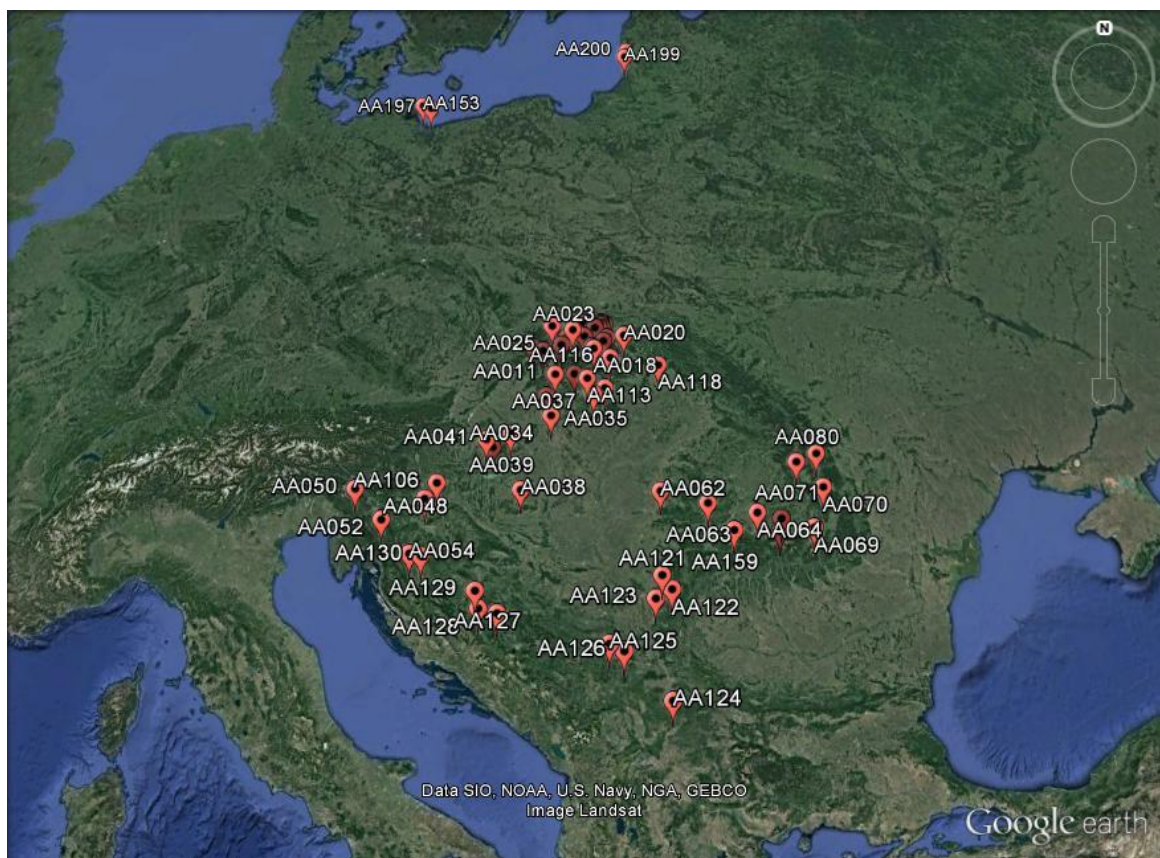
VKarp = Východní Karpaty

Kód	Provizorní taxon	Skupina	Stát	Popis lokality	Počet jedinců	Zeměpisná šířka (N)	Zeměpisná délka (E)	m n. m.
AA007	<i>petrogena</i>	ZKarp	SK	pukliny ve zdi	10	48.73897222	18.36641667	541
AA009	<i>petrogena</i>	ZKarp	SK	kamenitá step	10	48.68627778	18.63805556	598
AA011	<i>petrogena</i>	Pan	SK	příkop u silnice, v okolí les	10	48.23697222	18.99566667	247
AA012	<i>carpatica</i>	ZKarp	SK	narušená stráž nad silnicí	10	48.80625	19.13147222	683
AA013	<i>carpatica</i>	ZKarp	SK	vápencový svah v rozvolněném lese	10	48.95283333	19.26866667	637
AA016	<i>carpatica</i>	Balt	SK	vápencové kameny	10	48.96030556	20.38327778	600
AA017	<i>carpatica</i>	Balt	SK	říční břeh a přilehlé sedimenty	10	48.89569444	20.28211111	929
AA018	<i>petrogena</i>	ZKarp	SK	opuštěný kamenolom	10	48.74622222	20.02277778	467
AA020	<i>carpatica</i>	ZKarp	SK	svah nad silnicí, opadavý les	10	49.0115	20.84747222	604
AA021	<i>nitida</i>	ZKarp	SK	vápencový svah ve smíšeném lese	10	49.22891667	20.31169444	900
AA022	<i>neglecta</i>	ZKarp	SK	břeh potoka	5	49.15969444	20.24563889	1077
AA023	<i>carpatica</i>	ZKarp	SK	travertinová skalka	10	49.10725	19.43466667	574
AA025	<i>carpatica</i>	ZKarp	SK	vápencová skalka	10	49.17408333	18.8617	425
AA034	<i>petrogena</i>	Pan	HU	kamenitá step	10	47.458	18.92477778	252
AA035	<i>petrogena</i>	Pan	HU	kamenitá step	10	47.99019444	20.32744444	231
AA037	<i>petrogena</i>	Pan	HU	kamenitá step	5	47.87372222	20.03038889	898
AA038	<i>petrogena</i>	Pan	HU	kamenitá step	10	46.09916667	18.23391667	408

AA039	<i>petrogena</i>	Pan	HU	kamenitá step	10	47.09719444	17.89880556	318
AA040	<i>petrogena</i>	Pan	HU	kamenitá step	10	46.84391667	17.45138889	402
AA041	<i>petrogena</i>	Pan	HU	kameny, ruiny hradu	10	46.98227778	17.28147222	318
AA048	<i>petrogena</i>	Din	HR	svah nad silnicí, les	10	45.86908333	15.85056667	326
AA050	<i>petrogena</i>	Din	SLO	okraj silnice	4	45.96928333	14.06211667	500
AA052	<i>petrogena</i>	Din	HR	šterkové naplaveniny na říčním břehu	10	45.46061111	14.82547222	276
AA054	<i>petrogena</i>	Din	HR	kamenité údolí, podél turistické trasy	10	44.90433333	15.6111	561
AA062	<i>petrogena</i>	VKarp	RO	stinné kameny, ruiny hradu	5	46.10894444	21.72316667	261
AA063	<i>petrogena</i>	VKarp	RO	rozpadlé zdi podél cesty k ruině hradu	5	45.88969444	22.89672222	327
AA064	<i>petrogena</i>	VKarp	RO	rozvolněná kamenitá doubrava	10	45.70222222	24.11152778	570
AA066	<i>petrogena</i>	VKarp	RO	narušená stráň nad silnicí + zdi kolem	10	45.36391667	24.63755556	900
AA069	<i>petrogena</i>	VKarp	RO	stinné kameny	10	45.40558333	25.50077778	1368
AA070	<i>petrogena</i>	VKarp	RO	stinné kameny, opuštěný kamenolom	10	46.10908333	25.84636111	609
AA071	<i>petrogena</i>	VKarp	RO	stinné kameny podél potoka	10	46.57936111	25.23205556	572
AA080	<i>petrogena</i>	VKarp	RO	stinné kameny, narušený svah	5	46.71591667	25.77683333	1693
AA084	<i>neglecta</i>	ZKarp	SK	naplavený šterk, vlhké kameny	10	49.162	20.15419444	1823
AA086	<i>nitida</i>	ZKarp	SK	exponované kameny a šterk (cesta)	6	49.25925	19.72183333	1552
AA089	<i>nitida</i>	ZKarp	SK	šterk, vlhké a stinné kameny, turistická cesta	10	49.24841667	20.22761111	1207
AA090	<i>neglecta</i>	ZKarp	SK	vlhká suť	10	49.20652778	20.21505556	1625
AA091	<i>nitida</i>	ZKarp	SK	kameny a šterk (turistická cesta) + přilehlý svah	10	49.22991667	20.21902778	1751
AA106	<i>petrogena</i>	Din	HR	podél turistické cesty + zdi hradní ruiny	10	46.16166667	16.115	550

AA110	<i>petrogena</i>	Pan	SK	kamenitý svah kopce	10	47.82388333	18.77818333	161
AA112	<i>petrogena</i>	Pan	SK	kamenitý svah kopce, dubohabřina	10	48.24938333	19.51065	203
AA113	<i>petrogena</i>	Pan	SK	pastviny, suť a kameny, ruina hradu	10	48.17146667	19.8564	496
AA116	<i>petrogena</i>	ZKarp	SK	kameny podél silnice	10	48.56896667	20.47128889	400
AA118	<i>petrogena</i>	ZKarp	SK	doubrava	10	48.43313333	21.76998333	267
AA121	<i>petrogena</i>	VKarp	SRB	svah nad silnicí	10	44.6557	21.70708889	81
AA122	<i>petrogena</i>	VKarp	SRB	hlušina z dolu, okraj města	10	44.419275	21.92992778	382
AA123	<i>petrogena</i>	VKarp	SRB	stinné vápencové kameny	10	44.26528333	21.54271111	184
AA124	<i>petrogena</i>	VKarp	SRB	stinné kameny, podél silnice	10	42.59192222	21.88016111	888
AA125	<i>petrogena</i>	VKarp	SRB	kameny a štěrk podél řeky	10	43.516075	20.45257222	583
AA126	<i>petrogena</i>	VKarp	SRB	kameny a štěrk podél silnice	10	43.39950278	20.79089444	704
AA127	<i>petrogena</i>	Din	BIH	kamenitý svah nad silnicí	10	43.97502222	17.82445556	754
AA128	<i>petrogena</i>	Din	BIH	kameny a naplavený štěrk	10	44.04258333	17.37411111	772
AA129	<i>petrogena</i>	Din	BIH	kamenitý narušený svah, vedle parkoviště	5	44.33761667	17.26995556	384
AA130	<i>petrogena</i>	Din	BIH	kameny	10	44.88180833	15.89881667	217
AA137	<i>petrogena</i>	ZKarp	SK	kameny	10	48.785498	19.977113	1339
AA153	<i>arenosa</i>	Balt	D	písečné duny	4	53.98566	14.12513	22
AA156	<i>petrogena</i>	VKarp	RO	u silnice, opadavý les	10	45.58197222	24.70247222	1518
AA159	<i>petrogena</i>	VKarp	RO	u silnice	10	45.40675	23.50566667	985
AA162	<i>neglecta</i>	ZKarp	SK	vlhké kameny a štěrk, u vodopádu	10	49.15341	20.04584	1750
AA170	<i>nitida</i>	ZKarp	SK	suťový svah, štěrk	8	49.250216	20.205255	1380

AA176	<i>neglecta</i>	ZKarp	SK	břeh horského jezera	10	49.175629	20.060517	1950
AA183	<i>carpatica</i>	ZKarp	SK	narušený svah nad silnicí	10	48.9845	19.7573	735
AA197	<i>arenosa</i>	Balt	PL	písečné duny + pěšina v lese	10	53.92109	14.42157	19
AA199	<i>arenosa</i>	Balt	LT	písečné duny	10	55.378206	21.032308	1
AA200	<i>arenosa</i>	Balt	LT	písečné duny	10	55.52908	21.095556	1



Obrázek 4: Mapa lokalit zkoumaných populací – celkový pohled. Zdroj: Google Earth.

## 3.2. Laboratorní práce

Molekulární analýzy byly prováděny v DNA laboratoři Katedry botaniky Přírodovědecké fakulty Univerzity Karlovy v Praze (<http://botany.natur.cuni.cz/dna/>).

### 3.2.1. Izolace DNA

DNA pro laboratorní analýzy byla izolována ze sušené biomasy převážně listů a stonků. Byl použit CTAB protokol (Doyle & Doyle 1987) pouze s několika drobnými úpravami (viz příloha C.1.). Následně byla změřena koncentrace vyizolované DNA pomocí spektrofotometru NanoDrop (od firmy Thermo Scientific), který dokáže změřit koncentraci určité látky v objemu pouze 1  $\mu\text{l}$ . Podle zjištěných koncentrací byly vzorky DNA naředěny dvakrát destilovanou vodou ( $\text{ddH}_2\text{O}$ ) na koncentraci  $c = 10 \text{ ng}/\mu\text{l}$ . Vzorky s touto koncentrací byly používány při všech laboratorních pracích (analýza mikrosatelitů a příprava na sekvenování mikrosatelitů).

### 3.2.2. Analýza mikrosatelitů a optimalizace multiplexu primerů

Vzhledem k faktu, že mikrosatelitové primery jsou vysoce druhově specifické, je potřeba pro studovaný druh vhodné mikrosatelitové lokusy nejdříve identifikovat a navrhnout pro ně specifické primery. V případě, že jsou mikrosatelitové lokusy již známé pro některý blízkce příbuzný druh, stačí ověřit použitelnost a variabilitu primerů pro amplifikaci těchto lokusů na zkoumaném druhu. Pro zástupce rodu *Arabidopsis* (konkrétně pro druhy blízkce příbuzné komplexu *Arabidopsis arenosa*) bylo publikováno několik mikrosatelitových lokusů ve studii Clauss et al. (2002) a sedm z nich již bylo úspěšně použito i pro analýzy *A. arenosa* (Schmickl & Koch 2011; Hohmann et al. 2014). Pro účely této práce bylo tedy použito těchto sedm dříve optimalizovaných mikrosatelitových lokusů ze studie Clauss et al. (2002), které byly uspořádány do pracovního multiplexu, označovaného jako Multiplex 1 (viz tabulka 2).

Tabulka 2: Rozložení mikrosatelitových lokusů zahrnutých v Multiplexu 1 (M1). Názvy lokusů odpovídají názvům z původní studie (Clauss et al. 2002). Rozsah délek alel byl odvozen z výsledků analýz *A. arenosa* v této práci.

Fluorescenční značení	Mikrosatelitové lokusy a rozsah délek alel		
žlutá (NED)	ICE13 (210 – 282)	GC15 (116 – 168)	---
modrá (6FAM)	GC16 (129 – 174)	SLL2 (258 – 308)	NGA162 (71 – 93)
zelená (VIC)	ICE14 (200 – 245)	---	---
červená (PET)	ICE 7 (90 – 153)	---	---

Pro zvýšení informativnosti analýz bylo z publikace Clauss et al. (2002) vybráno a optimalizováno dalších dvanáct primerových párů, jejichž použitelnost a funkčnost byla testována na několika populacích z datasetu, které byly vybrány tak, aby reprezentovaly variabilitu celého komplexu *A. arenosa* a pocházely z různých provizorních poddruhů, různých ekologických stanovišť a různých částí areálu rozšíření (viz tabulka 3 a více informací v tabulce 1).

Tabulka 3: Seznam a charakteristiky populací použitých pro optimalizaci mikrosatelitových primerů a testování multiplexových reakcí.

Populace	Provizorní taxon	Geografická oblast	Stát
AA023	<i>Arabidopsis arenosa</i> subsp. <i>carpatica</i>	Západní Karpaty	Slovensko
AA040	<i>Arabidopsis arenosa</i> subsp. <i>petrogena</i>	Panonská nížina	Maďarsko
AA041	<i>Arabidopsis arenosa</i> subsp. <i>petrogena</i>	Panonská nížina	Maďarsko
AA048	<i>Arabidopsis arenosa</i> subsp. <i>petrogena</i>	Dinárské hory	Chorvatsko
AA084	<i>Arabidopsis arenosa</i> subsp. <i>neglecta</i>	Západní Karpaty	Slovensko
AA091	<i>Arabidopsis arenosa</i> subsp. <i>nitida</i>	Západní Karpaty	Slovensko
AA124	<i>Arabidopsis arenosa</i> subsp. <i>petrogena</i>	Jižní Evropa	Srbsko
AA128	<i>Arabidopsis arenosa</i> subsp. <i>petrogena</i>	Jižní Evropa	Bosna a Hercegovina
AA153	<i>Arabidopsis arenosa</i> subsp. <i>arenosa</i>	Severní Evropa	Německo
AA200	<i>Arabidopsis arenosa</i> subsp. <i>arenosa</i>	Severní Evropa	Litva

Při vlastní optimalizaci primerů byla testována funkčnost jednotlivých mikrosatelitových primerů, tj. jestli nasedají na konkrétní lokusy a dochází k jednoznačné amplifikaci v daných lokusech. V případě úspěšné amplifikace lokusu byla dále testována



jeho variabilita fragmentační analýzou. Primární data z těchto analýz pak byla vyhodnocena metodou PCoA pro zhrubý odhad informativnosti použitých markerů (zahrnuto mezi výsledky, kapitola 4). Tímto způsobem bylo ještě před započítím hlavních analýz dat postupně vyloučeno 5 primerů, které nevykazovaly dostatečnou variabilitu (tj. napříč všemi populacemi se vyskytovala pouze jedna alela), nebo ty u kterých neproběhla amplifikace během PCR. V tabulce 4 je uveden seznam nově optimalizovaných dvanácti lokusů, jakož i výsledky optimalizačních testů ohledně jejich dalšího použití.

Tabulka 4: Seznam a vlastnosti dvanácti nově vybraných a testovaných mikrosatelitových lokusů, původně publikovaných ve studii Clauss et al. (2002). Názvy lokusů odpovídají názvům z původní studie, číslované značení bylo používáno jen k usnadnění orientace při laboratorní práci.

Číslo lokusu (pracovní označení)	Název lokusu	PCR	Variabilita lokusu	Použití v Multiplexu 2
I	<b>F20D22</b>	ano	ano	ano
II	<b>F21M12</b>	ano	ano	ano
III	<b>F19G10</b>	ano	ano	ano
V	<b>F19K23-483</b>	ano	ano	ano
VI	<b>ICE12</b>	ano	ano	ano
VII	<b>ICE11</b>	ano	ano	ano
VIII	<b>MDC16</b>	ano	ano	ano
IX	<b>AthGAPAb</b>	ano	ne	ne
X	<b>ICE6</b>	ne	nelze určit	ne
XII	<b>AthCDPK9</b>	částečně	ne	ne
XIII	<b>ICE15</b>	částečně	ne	ne
XIV	<b>ICE4</b>	ano	ne	ne

Z primerů, které se pro *Arabidopsis arenosa* osvědčily, byl sestaven druhý pracovní multiplex, označený jako Multiplex 2 (viz tabulka 5). Při sestavování multiplexu (tj. při optimalizaci podmínek reakce, ve které jsou přítomny primery všech lokusů najednou) byly nejprve všechny primery testovány zvlášť. Byl zjištěn rozsah délek alel každého primeru a následně byly primery sestaveny do dvojic tak, aby se rozsah jejich

alel nepřekrýval. K rozlišení na automatickém sekvenátoru (3130xl Genetic Analyzer, Applied Biosystems® 2010), byly k dispozici čtyři fluorescenční barvy (tj. fluorescenční značení jednoho z primerů pro amplifikaci): modrá, zelená, žlutá a červená. Vnitřní standard vzorků (LIZ 600) pro analýzu na sekvenátoru byl od výrobce obarven oranžově. Tímto způsobem byly rozděleny a odlišeny primery v obou pracovních multiplexech (viz tabulky 2 a 5).

Tabulka 5: Rozložení nově optimalizovaných mikrosatelitových lokusů v Multiplexu 2 (M2). Názvy lokusů odpovídají názvům z původní studie (Clauss et al. 2002). Rozsah délek alel byl odvozen z výsledků analýz *A. arenosa* v této práci.

Fluorescenční značení	Mikrosatelitové lokusy a rozsah délek alel	
žlutá (NED)	F20D22 (159 – 213)	MDC16 (105 – 129)
modrá (6FAM)	F19G10 (169 – 185)	ICE11 (114 – 154)
zelená (VIC)	F21M12 (138 – 188)	ICE12 (215 – 251)
červená (PET)	F19K23-483 (169 – 214)	---

Při testování funkčnosti multiplexu bylo nutné vyladit optimální poměry jednotlivých primerů v primer-mixu, který je vstupní složkou PCR (QIAGEN kit), takovým způsobem, aby množství PCR-produktu jedné alely významně nepřevyšovalo PCR-produkty ostatních alel. Výsledné složení optimalizovaného primer-mixu je v příloze C.2.4. Koncentrace primeru ICE12 byla několikrát navýšena, ale i přesto se primer ICE12 nakonec nepodařilo do multiplexu pro PCR zahrnout, protože koncentrace PCR-produktů jeho alel byly vzhledem k ostatním stále příliš nízké. To znemožňovalo rozpoznání produktů na sekvenátoru a výsledky byly nezřetelné a nespolehlivé. V důsledku toho byl primer ICE12 z multiplexu pro PCR vyřazen a bylo nutné všechny PCR pro tento primer provádět zvlášť (s REDTaq polymerázou). Ze stejného důvodu byl již dříve vyřazen primer ICE7 z hromadné PCR Multiplexu 1. Laboratorní amplifikace mikrosatelitových markerů tedy spočívala v tom, provést pro každého zkoumaného jedince 4 reakce (PCR) – jednu s ICE7, jednu s ICE12 a pak dvě s multiplexy M1 a M2. Oba multiplexy byly zkompletovány teprve před odesláním vzorků na fragmentační analýzu na automatickém sekvenátoru.

Nakonec byly tedy použity oba pracovní multiplexy pro analýzu celého datasetu, takže pro všechny populace bylo analyzováno celkem čtrnáct mikrosatelitových lokusů. Díky tomu bylo získáno poměrně velké množství dat.

### 3.2.3. Sekvenování mikrosatelitů

Jednotlivé mikrosatelitové lokusy byly připraveny na sekvenování (protokol reakce je v příloze C.6.), které proběhlo v laboratoři sekvenace DNA Přírodovědecké fakulty Univerzity Karlovy. Získání sekvencí bylo důležité z toho důvodu, abych se ujistila o správnosti předpokládaných repetitivních oblastí mikrosatelitů.

Primární data ze sekvenátoru byla kontrolována pomocí programu FinchTV (<http://www.geospiza.com/ftvdlinfo.html>).

### 3.2.4. Kontrola vzájemné kompatibility primerů

Pomocí programů volně přístupných na internetu byla kontrolována struktura mikrosatelitových primerů a jejich vzájemná kompatibility nebo inkompatibilita. Konkrétně se jednalo o online analyzátoři: <http://eu.idtdna.com/calc/analyzer> a <https://www.thermofisher.com/cz/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html>.

V případě, že by primery byly kompatibilní, tvořily by páry mezi sebou, místo aby nasedaly na řetězce DNA. To by mohlo mít za následek snížení efektivity amplifikace v průběhu PCR (více v kapitole 4. Výsledky).

### 3.2.5. PCR a ředění produktů

Pro amplifikaci lokusů z multiplexů M1 a M2 byly použity odpovídající primery (Clauss et al. 2002), přičemž forward primer byl vždy značený fluorescenční značkou, jak je znázorněno v tabulkách 2 a 5. Pro přípravu polymerázových řetězových reakcí (PCR) byly použity protokoly uvedené v přílohách C.2., a v přílohách C.3. jsou rozepsané teplotní podmínky pro běh PCR v termocykleru (Mastercycler, Eppendorf).

Všechny testovací reakce a přípravné reakce na sekvenování mikrosatelitů byly prováděny s MyTaq polymerázou (Bioline). Reakce pro analýzu celého datasetu byly z finančních důvodů (za účelem úspory) prováděny s REDTaq polymerázou (Sigma-Aldrich) a na reakce multiplexů byl použit QIAGEN Multiplex PCR Kit (QIAGEN).

V průběhu práce se ukázalo, že některé PCR produkty měly příliš vysoké koncentrace fragmentů DNA, což občas způsobovalo problémy automatickému sekvenátoru při analyzování vzorků. Z toho důvodu bylo vyzkoušeno ředění PCR produktů dvakrát destilovanou vodou, které bylo již praktikováno na Multiplexu 1, a to s těmito hodnotami:

- produkty QIAGEN kitu byly ředěny 120  $\mu\text{l}$  ddH<sub>2</sub>O
- produkty REDTaq polymerázy byly ředěny 10  $\mu\text{l}$  ddH<sub>2</sub>O

U Multiplexu 2 se ředění zkušelo několikrát a nakonec se ukázalo optimální s těmito hodnotami:

- produkty QIAGEN kitu byly ředěny 120  $\mu\text{l}$  ddH<sub>2</sub>O
- produkty REDTaq polymerázy byly ředěny 90  $\mu\text{l}$  ddH<sub>2</sub>O

Také se ukázalo, že produkty REDTaq polymerázy je nutno ředit až po elektroforéze, jinak dojde k přílišnému zředění fluorescenční barvy (viz níže v kapitole Elektroforéza), a po nanesení vzorků není na gelu vidět žádný signál.

### 3.2.6. Elektroforéza

Metoda mikrosatelitů produkuje velké množství alel. Různě dlouhé alely mají zároveň různou pohyblivost v elektrickém poli, a proto je lze rozdělit a následně vizualizovat pomocí elektroforézy. Elektroforéza slouží jako vizuální kontrola toho, jestli proběhla amplifikace fragmentů DNA při PCR.

Produkty PCR byly vždy nanášeny na 1% agarózový gel, který byl připraven rozpuštěním agarózy v TAE pufru (10x tris-acetát-EDTA), a následně vložen do nádoby s tímž pufrům, připojené ke zdroji napětí. Elektroforéza probíhala při napětí 120 V po dobu přibližně 20 min.

Ke zvýraznění fragmentů DNA v gelu se používá fluorescenčních barviv, svítících v UV světle. Tato barviva jsou buď přimíchána přímo v pufru pro PCR (např. v pufru pro REDTaq polymerázu), nebo je nutné je přidat těsně před nanášením vzorků na gel. Takovým barvivem je např. ethidium bromid. K prohlížení gelů po elektroforéze byla využívána kamera Kodak Gel Logic 100, která umožňuje i pořizování snímků.

### 3.3. Fragmentační analýza a vyhodnocování primárních dat

Amplifikované fragmenty (mikrosatelitové lokusy; protokol k fragmentační analýze je v příloze C.5.) byly analyzovány na automatickém šestnáctikapilárovém sekvenátoru 3130xl Genetic Analyzer (Applied Biosystems®, 2010) v laboratoři sekvenace DNA Přírodovědecké fakulty Univerzity Karlovy (<https://www.natur.cuni.cz/biologie/servisni-laboratore/laborator-sekvenace-dna>).

Data získaná ze sekvenační laboratoře byla následně zpracovávána pomocí demo verze programu GeneMarker (SoftGenetics, LLC., State College, PA, USA, <http://www.softgenetics.com/genemarker.html>). Tento program umožňuje graficky znázornit všechny přítomné alely v každém vzorku (tj. u každého jedince). Z programu GeneMarker byla primární data manuálně převedena do tabulky (Microsoft Excel), která obsahovala veškerá data o všech jedincích a alelách ve všech lokusech. Z této datové matice pak byly vytvářeny vstupní soubory (tabulky v jiném formátu) pro další programy.

### 3.4. Analýza mikrosatelitových dat

Jedním z praktických programů pro analýzu mikrosatelitových dat je MSA – Microsatellite Analyser ([http://i122server.vu-wien.ac.at/MSA/MSA\\_download.html](http://i122server.vu-wien.ac.at/MSA/MSA_download.html)). Program slouží k počítání základních popisných statistik (Dieringer & Schlötterer 2003), jako jsou pozorovaná ( $H_O$ ) a očekávaná heterozygosita ( $H_E$ ), index genové diverzity  $H_S$  (podle Nei 1987), průměrný počet alel a jejich frekvence výskytu. Vstupním souborem pro MSA je datová matice všech alel (pro všechny jedince, přes všechny lokusy), obsahující také informace o lokusech, jako například velikost repetitivní oblasti (udávaná v bp) a délka opakujícího se motivu (také v bp). Vstupní soubor obsahuje také informace o studovaných populacích, například jejich příslušnost ke skupinám/větším celkům nebo jsou-li populace inbrední nebo outbrední.

Dále byl spočítán DW index („frequency-downweighted marker index“), též označovaný jako index rarity (Schönswetter & Tribsch 2005), na základě matice „presence-absence“ alel použitím R skriptu AFLPdat (Ehrich 2006). Hodnota DW indexu se očekává vyšší u populací s velkým množstvím vzácných alel, tj. alel s nízkou frekvencí výskytu v celém datasetu, což obecně naznačuje dlouhodobé přežívání takové

populace v kontrastu s vlivem nedávné imigrace nebo disperze na velkou vzdálenost (Paun et al. 2008). Za účelem zjištění, zda v některých populacích nedochází k samoopylení, byl spočítán koeficient inbreedingu ( $F_{IS}$ ) v programu INEST (Chybicki & Burczyk 2009).

Zejména za účelem primární analýzy mikrosatelitových dat z optimalizačních analýz (pouze na testovacím datasetu – viz metodika), ale i pro celkový dataset, byla provedena analýza hlavních koordinát (PCoA) pomocí nejnovější verze freewarového programu PAST3.5 (<http://folk.uio.no/ohammer/past/>). PCoA je mnohorozměrná analýza, která umožňuje zobrazit shluky geneticky podobných jedinců. Vstupním souborem pro tuto analýzu byla 0-1 matice znázorňující přítomnost nebo nepřítomnost alely v lokusu.

Pomocí programu STRUCTURE v2.3.2 (Pritchard et al. 2000), který funguje na principu Bayesovského shlukování, bylo určeno optimální seskupení populací do shluků, pro odhalení genetické struktury v celém datasetu. Analýza byla nastavena pro testování modelu pro  $K = 1$  až  $K = 10$  (kde  $K$  je počet klastrů neboli shluků), přičemž pro každé  $K$  bylo provedeno 10 opakování s nastavením délky řetězce na 100.000 generací pro „burn-in“ periodu a následným 1.000.000 generací pro vlastní MCMC analýzy (Evanno et al. 2005). Optimální počet genetických skupin byl určen pomocí skriptu Structure-sum-2009.R (Ehrich et al. 2007), který vypočetl koeficienty podobnosti mezi jednotlivými běhy Structure. K určení nejlepšího „ $K$ “ bylo využito i  $\Delta K$  jako kritérium odrážející rozdíly v podobnosti jednotlivých běhů programu (Evanno et al. 2005). Ke grafickému znázornění výstupů z programu Structure byl využit program Distruct (Rosenberg 2004).

Pomocí analýzy molekulární variance (zkráceně AMOVA) bylo kvantifikováno rozdělení genetické variability v rámci a mezi populacemi celého datasetu. Stejně bylo zjištěno rozdělení genetické variability mezi hlavními skupinami (tzv. hierarchická AMOVA). Tyto skupiny byly definovány na základě výsledků Structure s přihlédnutím ke geografické distribuci populací. Analýza molekulární variance byla provedena v programu Arlequin 3.5 (Excoffier & Lischer 2010). V rámci této analýzy byly určeny i hodnoty fixačního indexu  $F_{ST}$  (Weir & Cockerham 1984).

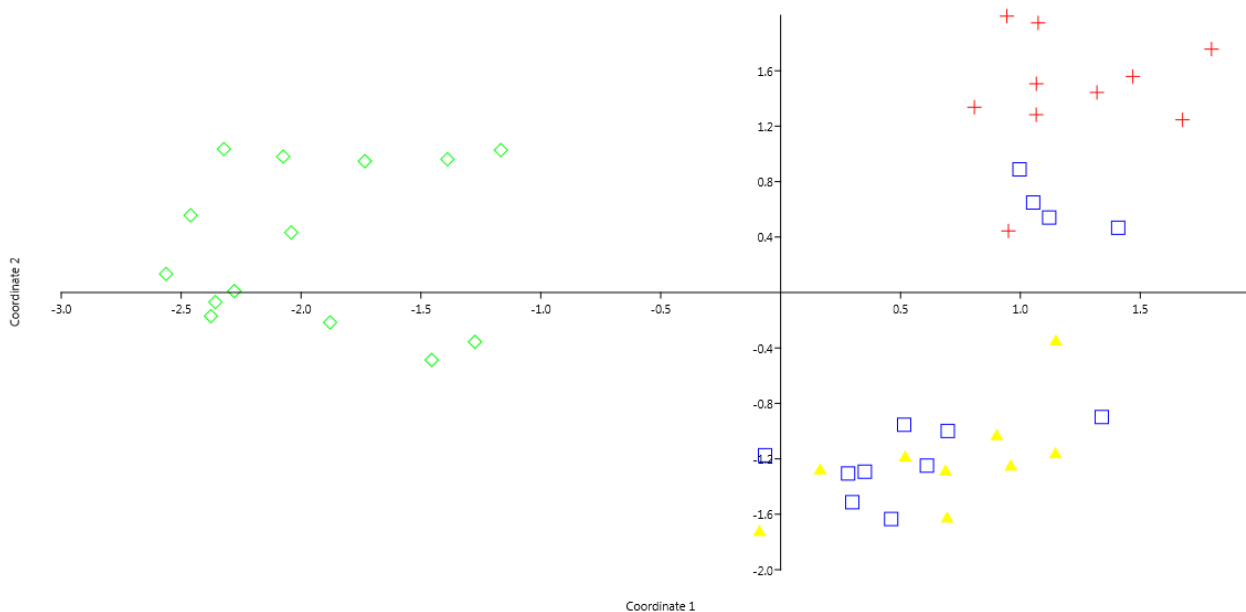
## 4. Výsledky

### 4.1. Optimalizace multiplexu primerů

Ze dvanácti testovaných mikrosatelitových lokusů bylo úspěšně amplifikováno pouze devět. Z těchto devíti lokusů byly dva lokusy (AthGAPAb a ICE4) zcela nevariabilní (viz tabulka 4). Zbývajících sedm lokusů bylo použito pro analýzu testovacího setu populací *A. arenosa* pro odhad jejich informativnosti pro fylogeografické analýzy.

Ke zjištění zda chystaná analýza těchto sedmi mikrosatelitů v kombinaci s markery z pracovního multiplexu M1 bude produkovat smysluplné výsledky, byla provedena průběžná analýza hlavních koordinát. K tomu účelu byla vytvořena datová matice, která obsahovala data ze šesti populací z testovacího souboru (viz tabulka 3, kapitola 3.2.2.) – konkrétně se jednalo o populace AA023, AA040, AA041, AA091, AA128 a AA200. Data byla získána za použití obou pracovních multiplexů M1 i M2, tj. celkem 14 mikrosatelitových lokusů. Tato datová matice byla zjednodušena tak, aby odrážela pouze přítomnost či nepřítomnost alely – binární data. Došlo tedy ke ztrátě informace o tom, jestli je daný jedinec homozygot nebo heterozygot.

Výsledkem analýzy (obrázek 5) bylo vytvoření dvou shluků zřetelně odlišených podle první osy (17,9 % variability) a třetí shluk bodů odlišený podle druhé osy (12,1 % variability). Pro přehlednost byly tyto body obarveny podle příslušnosti populací ke skupinám definovaným v rámci *Arabidopsis arenosa* na základě souhrnných výsledků z programu Structure (viz tabulka 1 nebo výsledky Structure obrázek 14).



Obrázek 5: Ordinační diagram analýzy hlavních koordinát testovacího datasetu provedené v programu PAST 3.5, jako koeficient podobnosti byl použit Jaccardův koeficient. První osa vysvětlila 17,9 % variability a druhá osa 12,1 % celkové variability. Jednotlivé body představují jedince *Arabidopsis arenosa*.

Západní Karpaty: □, Panonie: ◇, Pobaltí: ▲, Dinárské hory: +,

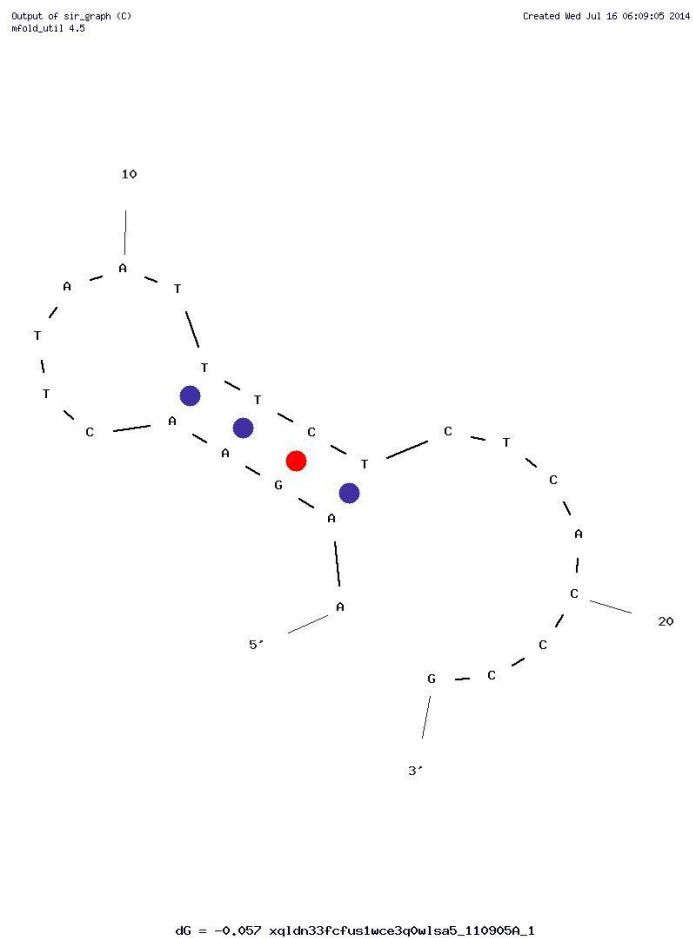
#### 4.1.1. Kontrola vzájemné kompatibility primerů v Multiplexu 2

Pomocí dvou různých online analyzátorů primerových sekvencí (viz metodika) byla zjištěna jejich vzájemná kompatibilita, a to nejen mezi jednotlivými primery nebo mezi „forward“ a „reverse“ primery, ale i v rámci jednotlivých primerových sekvencí.

Ukázalo se, že 13 různými způsoby dochází k párování bází mezi primery – tzv. Cross Primer Dimer. V pěti z těchto třinácti případů je jedním z primerů ICE12. S ohledem na tuto skutečnost (riziko snížení efektivity amplifikace v průběhu PCR) a celkovou problematičnost byl nakonec tento primer vyřazen z multiplexové PCR. Dále bylo zjištěno, že pouze u dvou primerů může dojít k tvorbě tzv. Self-Dimers, čili ke spojování určitých úseků primerových sekvencí a vzniku jakýchsi smyček. Jednalo se o primery F19G10 (reverse primer) a ICE11 (forward primer). Podle online analyzátoru



z Integrated DNA Technologies může docházet k párování bází v primeru F19G10 jedním způsobem (obrázek 6) a v primeru ICE11 pěti způsoby (příloha D).



Obrázek 6: Tvorba dimeru v rámci reverse primeru F19G10.

V praxi se ukázalo, že tvorba dimerů v rámci primerů skutečně ovlivňuje jejich funkčnost při amplifikaci mikrosatelitových sekvencí, a proto byl optimalizován primer-mix pro multiplexovou PCR. Koncentrace těchto dvou primerů byly tedy oproti ostatním zvýšeny (viz příloha C.2.4.).

## 4.2. Molekulární analýzy

Celkem bylo analyzováno 592 jedinců z komplexu *Arabidopsis arenosa*. Tito jedinci pocházeli z 64 populací v rámci celého areálu výskytu *A. arenosa*, konkrétně z území Slovenska (28 populací), Polska (1 populace), Chorvatska (4 populace), Slovinska (1 populace), Rumunska (10 populací), Maďarska (7 populací), Bosny a Hercegoviny (4 populace), Srbska (6 populací), Litvy (2 populace) a Německa (1 populace). Vzorky DNA získané z těchto jedinců byly analyzovány pomocí 14 jaderných mikrosatelitových markerů. Celkem bylo detekováno 220 alel napříč všemi lokusy, v rozmezí 7–26 alel na lokus. Shrnutí informací o jednotlivých lokusech je uvedeno v tabulce 6.

Tabulka 6: Seznam mikrosatelitových lokusů (z obou multiplexů M1 a M2) s rozsahem délek jejich alel a celkovým počtem nalezených alel v každém lokusu.

Název lokusu	Pracovní multiplex	Rozsah alel	Počet alel
ICE13	M1	210-282	23
GC15	M1	116-168	26
GC16	M1	129-174	16
SLL2	M1	258-308	7
ICE14	M1	200-245	11
nga162	M1	71-93	12
ICE7	M1	90-153	13
F20D22	M2	159-213	24
F21M12	M2	138-188	20
F19G10	M2	169-185	9
F19K23-483	M2	169-238	16
ICE12	M2	215-357	15
ICE11	M2	114-154	15
MDC16	M2	105-129	13

#### 4.2.1. Charakteristiky populací *Arabidopsis arenosa*

Pomocí mikrosatelitových markerů se podařilo detekovat 220 alel napříč všemi 14 lokusy (viz tabulka 7). Ne u všech jedinců se povedlo získat alely ze všech lokusů – v datasetu se na několika místech vyskytla chybějící data, kde určití jedinci měli alely pouze v 5-8 lokusech. Tito jedinci (jednalo se konkrétně o 12 jedinců z různých populací) byli nejprve z datasetu vyloučeni, aby příliš neovlivňovali výsledky analýz. Po otestování datasetu bez chybějících dat a kompletního datasetu se ukázalo, že vliv na výsledky je zcela zanedbatelný a dále byl používán kompletní dataset všech jedinců.

Tabulka 7: Seznam všech detekovaných alel ve všech lokusech, včetně četností jejich výskytu a procentuálního vyjádření těchto četností. Zahrnuje součet výskytů v celém datasetu.

<b>GC16</b>	<b>Frekvence</b>	<b>Procenta</b>	<b>F20D22</b>	<b>Frekvence</b>	<b>Procenta</b>
129	81	6,9	159	5	0,4
132	13	1,1	161	2	0,2
135	65	5,5	167	23	2,0
138	114	9,7	169	218	18,6
141	144	12,2	171	491	42,0
144	227	19,2	173	79	6,8
147	196	16,6	175	23	2,0
150	75	6,4	177	31	2,6
153	118	10,0	179	20	1,7
156	31	2,6	181	1	0,1
159	43	3,6	185	63	5,4
162	24	2,0	187	130	11,1
165	26	2,2	189	19	1,6
168	15	1,3	191	6	0,5
171	6	0,5	193	11	0,9
174	2	0,2	195	28	2,4
<b>SLL2</b>	<b>Frekvence</b>	<b>Procenta</b>	197	2	0,2
258	2	0,2	199	2	0,2
266	4	0,3	201	1	0,1
268	57	4,8	203	1	0,1
293	1	0,1	207	1	0,1
298	59	5,0	209	3	0,3

306	1	0,1	211	7	0,6
308	1058	89,5	213	3	0,3
<b>nga162</b>	<b>Frekvence</b>	<b>Procenta</b>	<b>F21M12</b>	<b>Frekvence</b>	<b>Procenta</b>
70	4	0,3	138	4	0,3
71	13	1,1	140	2	0,2
73	1	0,1	142	10	0,9
75	235	19,9	144	18	1,6
77	374	31,7	146	10	0,9
79	353	30,0	148	18	1,6
81	96	8,1	152	906	78,1
83	57	4,8	154	84	7,2
85	27	2,3	156	17	1,5
87	2	0,2	158	4	0,3
89	2	0,2	168	1	0,1
93	14	1,2	172	1	0,1
<b>ICE13</b>	<b>Frekvence</b>	<b>Procenta</b>	174	17	1,5
210	18	1,5	176	22	1,9
216	9	0,8	178	11	0,9
219	17	1,5	180	10	0,9
222	14	1,2	182	12	1,0
225	11	0,9	184	8	0,7
228	113	9,7	186	2	0,2
231	80	6,9	188	3	0,3
234	105	9,0	<b>F19G10</b>	<b>Frekvence</b>	<b>Procenta</b>
237	89	7,7	169	1	0,1
240	320	27,5	171	6	0,5
243	171	14,7	173	74	6,4
246	116	10,0	175	606	52,3
249	17	1,5	177	65	5,6
252	24	2,1	179	373	32,2
255	8	0,7	181	30	2,6
258	26	2,2	183	1	0,1
261	8	0,7	185	2	0,2
264	7	0,6	<b>F19K23-483</b>	<b>Frekvence</b>	<b>Procenta</b>
267	2	0,2	169	25	2,2
270	1	0,1	172	15	1,3
273	2	0,2	175	74	6,5
276	3	0,3	178	21	1,9
282	1	0,1	181	20	1,8

<b>ICE14</b>	<b>Frekvence</b>	<b>Procenta</b>	184	33	2,9
200	1	0,1	187	75	6,6
215	4	0,3	190	414	36,6
218	3	0,3	193	376	33,2
221	48	4,1	196	24	2,1
224	86	7,4	199	19	1,7
227	458	39,1	202	7	0,6
230	454	38,8	205	10	0,9
233	96	8,2	208	13	1,1
236	13	1,1	214	2	0,2
239	6	0,5	238	4	0,4
245	1	0,1	<b>ICE12</b>	<b>Frekvence</b>	<b>Procenta</b>
<b>GC15</b>	<b>Frekvence</b>	<b>Procenta</b>	215	4	0,4
116	1	0,1	221	2	0,2
120	10	0,9	223	94	8,3
124	18	1,6	225	68	6,0
126	18	1,6	227	134	11,8
128	34	2,9	229	7	0,6
130	198	17,1	233	2	0,2
132	147	12,7	237	7	0,6
134	73	6,3	243	3	0,3
136	141	12,2	245	25	2,2
138	153	13,2	247	681	59,9
140	118	10,2	249	33	2,9
142	49	4,2	251	72	6,3
144	51	4,4	299	2	0,2
146	32	2,8	357	2	0,2
148	51	4,4	<b>ICE11</b>	<b>Frekvence</b>	<b>Procenta</b>
150	13	1,1	114	3	0,3
152	4	0,3	132	11	1,0
154	8	0,7	134	11	1,0
156	11	0,9	136	7	0,6
158	5	0,4	138	431	37,3
160	15	1,3	140	2	0,2
162	4	0,3	142	358	31,0
166	1	0,1	144	5	0,4
168	1	0,1	146	137	11,9
221	1	0,1	148	1	0,1
230	1	0,1	150	160	13,8

ICE7	Frekvence	Procenta	152	21	1,8
90	9	0,8	154	7	0,6
93	25	2,1	<b>MDC16</b>	<b>Frekvence</b>	<b>Procenta</b>
96	850	72,6	105	8	0,7
99	6	0,5	107	18	1,6
105	54	4,6	109	4	0,3
120	2	0,2	111	233	20,2
123	1	0,1	113	384	33,2
132	78	6,7	115	311	26,9
135	17	1,5	117	97	8,4
138	111	9,5	119	58	5,0
141	9	0,8	121	27	2,3
147	7	0,6	123	2	0,2
153	1	0,1	125	2	0,2
			127	1	0,1
			129	11	1,0

Informace o základních populačních parametrech (jako jsou počty a frekvence alel, očekávaná a pozorovaná heterozygosita nebo genová diverzita) byly získány z analýzy v programu MSA. Tyto parametry shrnuje a srovnává tabulka 8. Počty alel v jednotlivých populacích se pohybovaly v rozmezí od 28 do 69 alel. Populace s největším počtem detekovaných alel napříč všemi lokusy pocházely převážně z Karpat.

Očekávaná heterozygosita v rámci populací se pohybovala v rozmezí od 0,31 do 0,63 a pozorovaná heterozygosita v rozmezí od 0,22 do 0,51. Nejnižší heterozygosita (očekávaná i pozorovaná) byla detekována u populace AA124, která byla zároveň charakterizována nejnižším počtem alel a nejnižší hodnotou  $H_S$  (ukazatel genové diverzity). Populace AA124 pochází z území Srbska a je to nejjihněji sebraná populace z datasetu. Naopak populace AA17 se vyznačovala nejvyšším počtem alel a nejvyššími hodnotami  $H_E$  a  $H_S$ . Tato populace byla sebrána na Slovensku. Dále se v celém datasetu se vyskytlo 8 populací s poměrně vysokým koeficientem inbreedingu  $F_{IS}$  (viz tabulka 8), u kterých lze tedy počítat se značně sníženou heterozygositou. Průměrná hodnota  $F_{IS}$  u těchto populací byla vyšší než 0,17 a zároveň spodní hranice konfidenčního intervalu byla vyšší než 0 (tyto hranice byly určeny dle vlastního uvážení). U těchto populací lze také nalézt poměrně velký rozdíl mezi očekávanou a pozorovanou heterozygositou.

Následující tabulka (tabulka 8) shrnuje kromě výše zmíněných parametrů také hodnoty DW indexu pro populace napříč všemi lokusy (viz také obrázek 7, část „b“). Tento index vyjadřuje míru vzácných alel v populaci, tj. alel s nízkou frekvencí výskytu ve zbytku datasetu. U populací s vysokým počtem vzácných alel jsou hodnoty DW indexu obvykle vyšší (například Paun et al. 2008). V rámci celého datasetu se pohybují v rozmezí od 0,12 do 1,02. Nejnižší hodnota indexu byla nalezena u populace AA124 (také zjištěn nejnižší počet detekovaných alel – viz předchozí odstavec) a nejvyšší hodnota u populace AA50. Tato populace pochází ze Slovinska a je velmi malá – k dispozici jsou pouze 4 jedinci (viz tabulka 1).

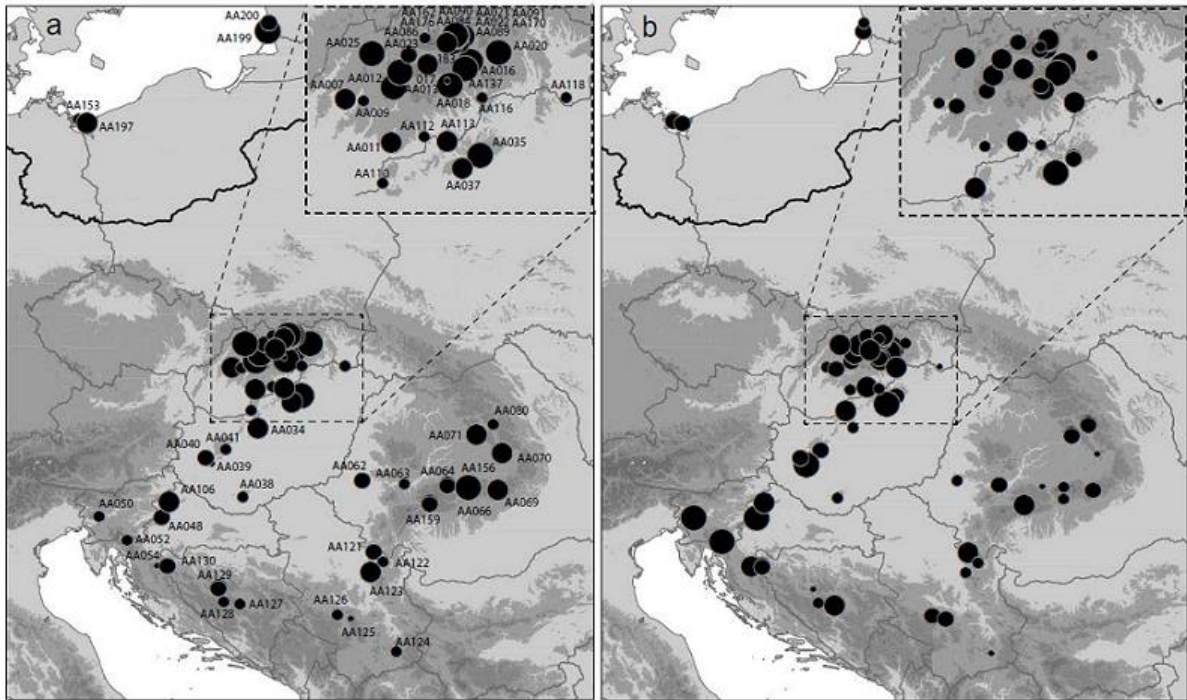
Tabulka 8: Souhrn základních populačních parametrů v přehledu pro všechny populace z datasetu. Pozorovaná ( $H_E$ ) a očekávaná ( $H_O$ ) heterozygosita a genová diverzita ( $H_S$ ) byla vypočtena programem MSA. Hodnoty indexu inbreedingu ( $F_{IS}$ ) a hranice 95% konfidenčního intervalu programem INEST. Zvýrazněné populace mají vyšší míru inbreedingu než ostatní, a tudíž u nich lze očekávat výrazněji sníženou heterozygositu. Hodnoty DW indexu spočítané pomocí skriptu AFLPdat.R vyjadřují podíl vzácných alel v populaci.

Kód populace	Počet alel	$H_E$	$H_O$	$H_S$	$F_{IS}$	Konfidenční interval - spodní (95%)	Konfidenční interval - horní (95%)	DW index
AA07	47	0,53	0,41	0,56	0,10	0	0,23	0,25
AA09	52	0,59	0,35	0,63	0,31	0,12	0,48	0,29
AA11	49	0,55	0,40	0,58	0,19	0	0,32	0,25
AA12	64	0,60	0,48	0,68	0,07	0	0,20	0,29
AA13	65	0,59	0,46	0,62	0,09	0	0,21	0,43
AA16	64	0,58	0,40	0,66	0,08	0	0,26	0,66
AA17	69	0,63	0,49	0,71	0,09	0	0,23	0,66
AA18	68	0,62	0,44	0,66	0,11	0	0,22	0,45
AA20	56	0,57	0,47	0,64	0,03	0	0,11	0,22
AA21	46	0,51	0,25	0,58	0,44	0,26	0,60	0,18
AA22	46	0,57	0,33	0,64	0,29	0,15	0,47	0,24
AA23	54	0,53	0,45	0,60	0,10	0	0,19	0,40
AA25	57	0,60	0,36	0,64	0,25	0,11	0,41	0,41
AA34	44	0,50	0,50	0,57	0,02	0	0,07	0,24
AA35	46	0,47	0,36	0,54	0,11	0	0,25	0,32
AA37	35	0,43	0,34	0,53	0,09	0	0,26	0,73
AA38	37	0,40	0,33	0,49	0,09	0	0,23	0,22
AA39	46	0,47	0,32	0,49	0,21	0	0,39	0,34
AA40	48	0,50	0,45	0,48	0,04	0	0,12	0,67

AA41	35	0,43	0,39	0,53	0,05	0	0,15	0,31
AA48	46	0,41	0,33	0,51	0,14	0	0,29	0,67
AA50	31	0,35	0,34	0,46	0,08	0	0,24	1,02
AA52	51	0,50	0,33	0,53	0,16	0	0,34	0,66
AA54	46	0,39	0,29	0,44	0,15	0	0,30	0,53
AA62	40	0,50	0,36	0,60	0,10	0	0,26	0,25
AA63	31	0,37	0,30	0,45	0,06	0	0,20	0,37
AA64	52	0,51	0,43	0,54	0,06	0	0,17	0,18
AA66	53	0,51	0,40	0,58	0,06	0	0,17	0,23
AA69	50	0,53	0,37	0,56	0,13	0	0,30	0,28
AA70	54	0,44	0,30	0,51	0,29	0,09	0,46	0,15
AA71	51	0,51	0,42	0,54	0,04	0	0,14	0,35
AA80	34	0,41	0,40	0,45	0,04	0	0,13	0,30
AA84	69	0,62	0,51	0,61	0,05	0	0,16	0,52
AA86	52	0,52	0,42	0,54	0,10	0	0,22	0,33
AA89	48	0,56	0,39	0,58	0,10	0	0,29	0,39
AA90	60	0,54	0,35	0,57	0,24	0,09	0,41	0,33
AA91	54	0,46	0,32	0,49	0,10	0	0,28	0,43
AA106	40	0,40	0,31	0,50	0,07	0	0,20	0,42
AA110	58	0,54	0,50	0,57	0,07	0	0,16	0,46
AA112	43	0,44	0,27	0,51	0,35	0,19	0,50	0,38
AA113	46	0,47	0,39	0,49	0,09	0	0,23	0,21
AA116	52	0,49	0,33	0,60	0,15	0	0,30	0,38
AA118	42	0,45	0,27	0,56	0,25	0	0,41	0,18
AA121	57	0,52	0,43	0,60	0,13	0	0,25	0,47
AA122	44	0,45	0,36	0,52	0,09	0	0,23	0,20
AA123	43	0,48	0,36	0,55	0,12	0	0,30	0,23
AA124	28	0,31	0,22	0,42	0,16	0	0,34	0,12
AA125	37	0,48	0,43	0,47	0,05	0	0,15	0,30
AA126	35	0,44	0,43	0,46	0,03	0	0,10	0,29
AA127	50	0,47	0,36	0,53	0,08	0	0,22	0,45
AA128	51	0,47	0,39	0,54	0,11	0	0,24	0,22
AA129	31	0,34	0,26	0,50	0,11	0	0,32	0,17
AA130	46	0,43	0,34	0,49	0,07	0	0,22	0,29
AA137	62	0,59	0,49	0,67	0,03	0	0,11	0,33
AA153	39	0,48	0,39	0,53	0,08	0	0,26	0,31
AA156	44	0,46	0,39	0,45	0,10	0	0,22	0,22
AA159	57	0,56	0,49	0,58	0,09	0	0,19	0,38
AA162	58	0,55	0,46	0,58	0,08	0	0,20	0,36



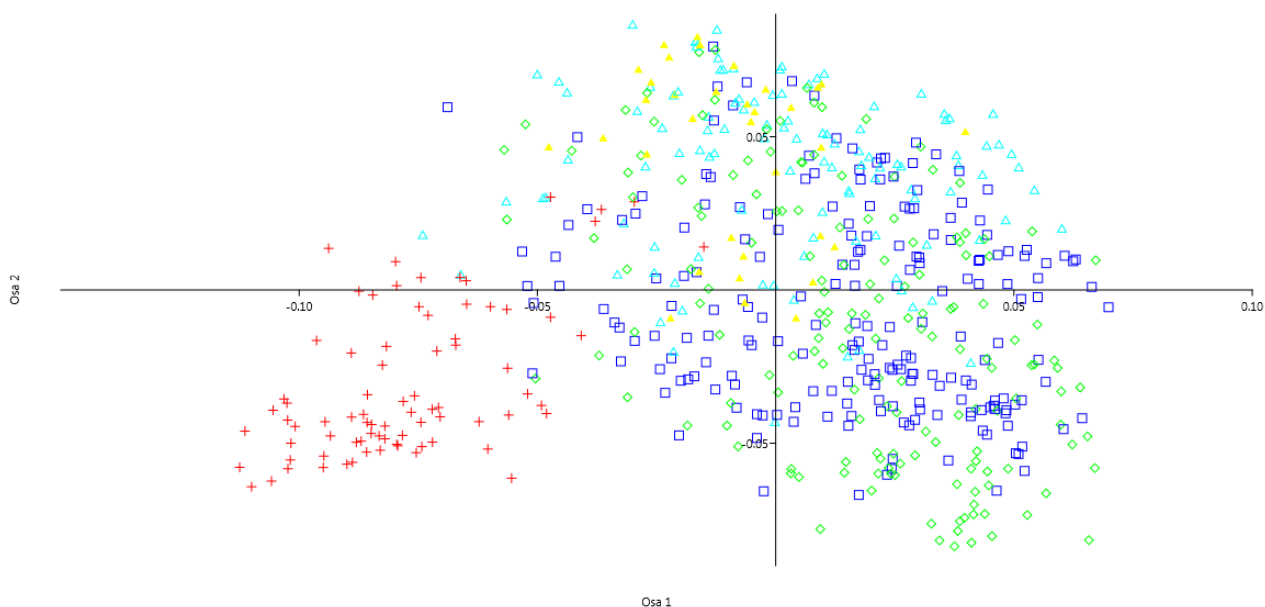
AA170	56	0,51	0,38	0,58	0,12	0	0,28	0,46
AA176	60	0,52	0,39	0,60	0,13	0	0,28	0,26
AA183	65	0,61	0,47	0,59	0,09	0	0,23	0,39
AA197	44	0,47	0,34	0,59	0,17	0	0,33	0,30
AA199	48	0,54	0,48	0,61	0,04	0	0,13	0,33
AA200	43	0,46	0,31	0,57	0,17	0,06	0,31	0,22



Obrázek 7: Rozložení genetické diverzity (část „a“) a podíl vzácných alel (část „b“), zjištěné na základě očekávané heterozygosity a DW indexu. Ve výřezu jsou detailně vidět populace ze Slovenska. Tučná linie na sever od České republiky značí maximální dosah kontinentálního ledovce během poslední doby ledové. **Obrázek převzat ze studie Kolář et al. (2016), kde byl analyzován dataset z této diplomové práce.**

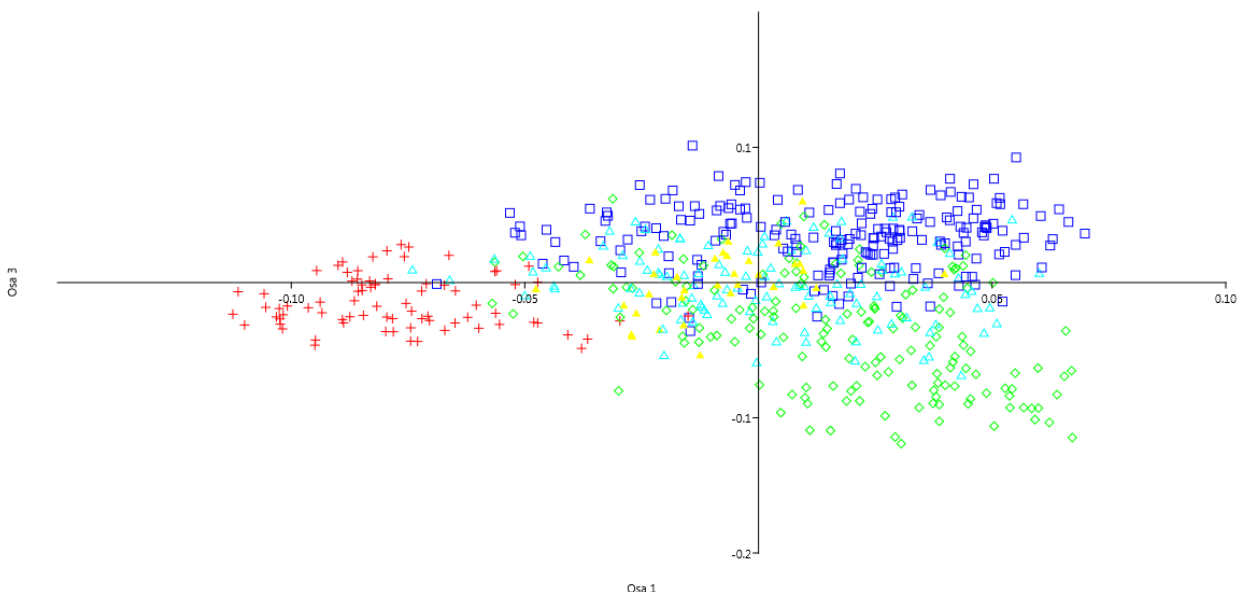
#### 4.2.2. Analýza hlavních koordinát (PCoA)

Po dokončení laboratorních prací byla znovu provedena analýza hlavních koordinát, tentokrát byl analyzován celý dataset (obrázek 8). Populace byly opět obarveny podle genetických skupin (podle výsledků Structure, detaily k zařazení populací do skupin jsou uvedeny též v tabulce 1). Díky tomu lze snadno rozeznat, že podél první osy (7,4 % variability) se mírně odlišila část populací z Panonské nížiny. Podél druhé osy (5,4 % variability) byl oddělen shluk populací z jihu Evropy. Podle třetí osy, která odhalila 4,8 % variability, se pak výrazněji odlišily populace karpatské od populací z Panonské nížiny (obrázek 9). Toto oddělení karpatských a panonských populací se potvrdilo i zobrazením vzorků z pohledu druhé a třetí hlavní koordináty (obrázek 10).



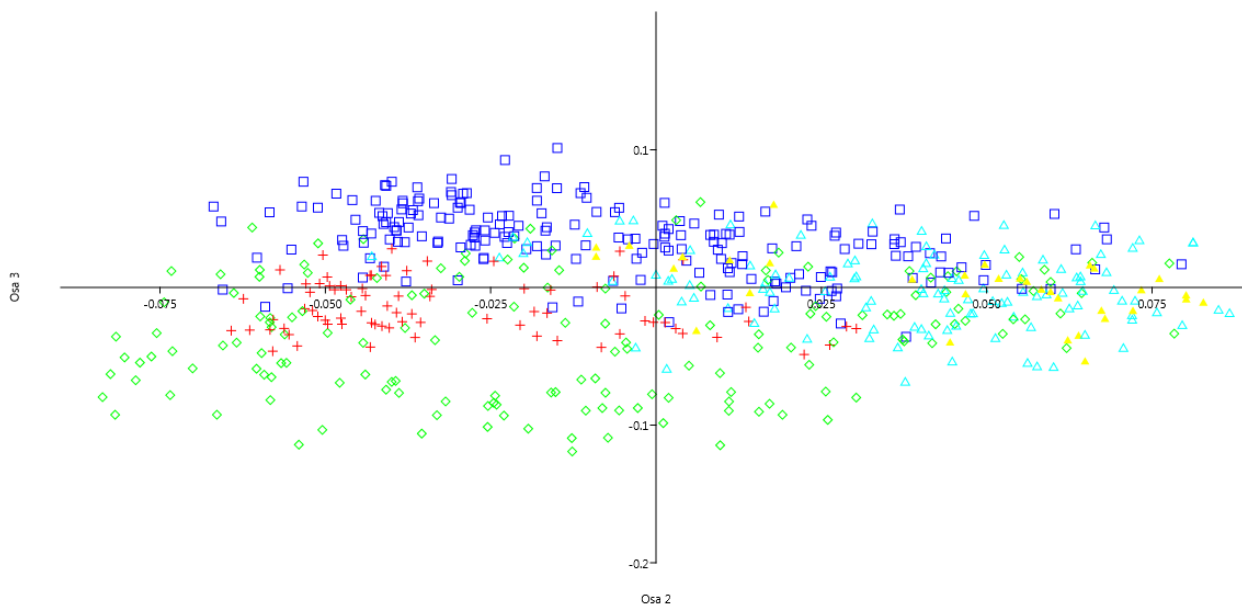
Obrázek 8: První a druhá osa ordinačního diagramu analýzy hlavních koordinát kompletního datasetu provedené v programu PAST 3.5, jako koeficient podobnosti byl použit Jaccardův koeficient. První osa vysvětlila 7,4 % variability, druhá osa vysvětlila 5,4 % variability. Jednotlivé body představují jedince.

Západní Karpaty:  $\square$ , Panonie:  $\diamond$ , Pobaltí:  $\blacktriangle$ , Dunájské hory:  $+$ , Východní Karpaty:  $\triangle$



Obrázek 9: První a třetí osa ordinačního diagramu analýzy hlavních koordinát kompletního datasetu provedené v programu PAST 3.5, jako koeficient podobnosti byl použit Jaccardův koeficient. První osa vysvětlila 7,4 % variability, třetí osa vysvětlila 4,8 % variability. Jednotlivé body představují jedince.

Západní Karpaty:  $\square$ , Panonie:  $\diamond$ , Pobaltí:  $\blacktriangle$ , Dinárské hory:  $+$ , Východní Karpaty:  $\triangleleft$

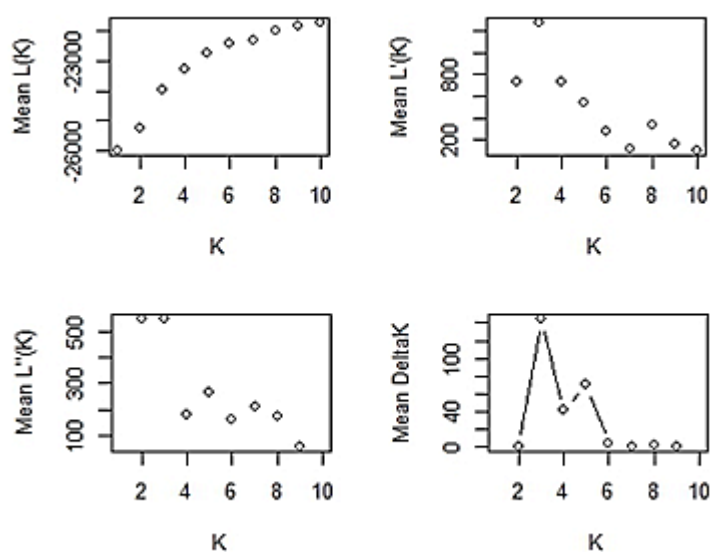


Obrázek 10: Druhá a třetí osa ordinačního diagramu analýzy hlavních koordinát kompletního datasetu provedené v programu PAST 3.5, jako koeficient podobnosti byl použit Jaccardův koeficient. Druhá osa vysvětlila 5,4 % variability, třetí osa vysvětlila 4,8 % variability. Jednotlivé body představují jedince.

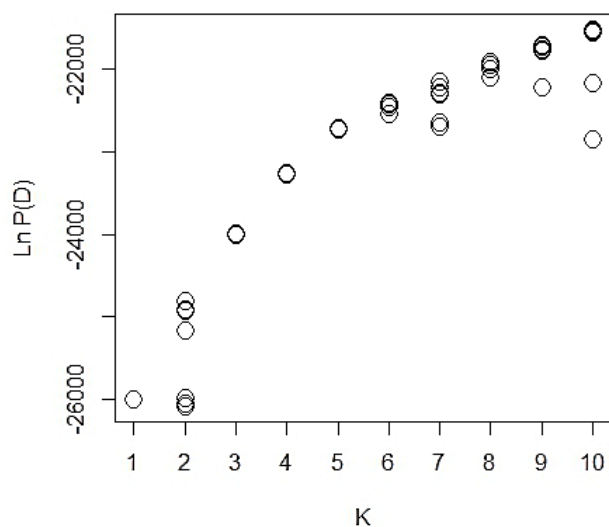
Západní Karpaty:  $\square$ , Panonie:  $\diamond$ , Pobaltí:  $\blacktriangle$ , Dinárské hory:  $+$ , Východní Karpaty:  $\triangleleft$

#### 4.2.3. Bayesovské shlukování podle programu Structure

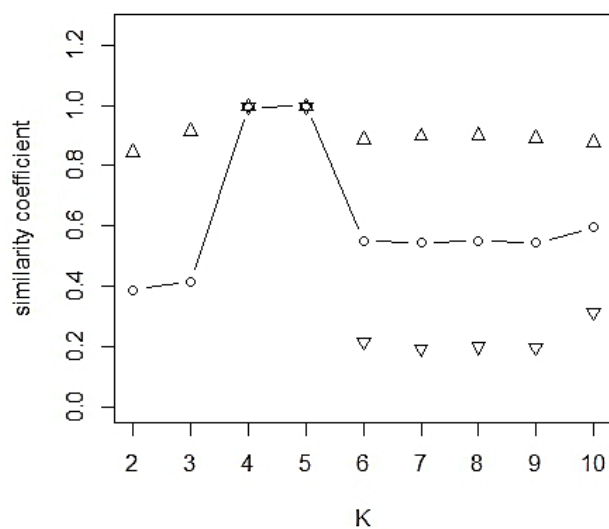
Jako nejlepší počet shluků ( $K$ ) pro datovou matici všech populací vyšlo  $K = 5$ . Tato hodnota  $K$  měla nejvyšší koeficient podobnosti opakování běhů algoritmu, vysoký logaritmus pravděpodobnosti modelu a druhé nejvyšší  $\Delta K$ . Další možností je  $K = 4$  s vysokým koeficientem podobnosti a logaritmem pravděpodobnosti, ale relativně nízkým  $\Delta K$ . Ve výsledném diagramu (obrázek 14) je zobrazena i varianta modelu pro  $K = 3$ , která měla nejvyšší hodnotu  $\Delta K$  a relativně vysokou hodnotu logaritmu pravděpodobnosti. Toto je znázorněno na grafech, které pomáhají určit optimální  $K$  (viz obrázky 11, 12, 13). Grafy ukazují pravděpodobnost modelů, které rozdělují jedince do konkrétního počtu skupin ( $K$ ) a rozdíl pravděpodobnosti mezi běhy programu pro různé počty skupin, které jsou předem určeny. Míra podobnosti mezi jednotlivými běhy programu je vyjádřena koeficientem podobnosti, a čím vyšší je podobnost výsledků, tím vyšší je i pravděpodobnost a stabilita daného rozdělení jedinců do shluků.



Obrázek 11: Výstupní grafy ze Structure-sum. Graf vlevo nahoře ukazuje na svislé ose logaritmus pravděpodobnosti modelu –  $L(K)$  – optimálního rozdělení do daného počtu skupin  $K$  zobrazeného na vodorovné ose. Graf vpravo nahoře ukazuje průměrnou míru změny pravděpodobnosti mezi výpočty s dvěma různými počty skupin  $K$  –  $L'(K)$ . Graf vlevo dole ukazuje míru změny mezi dvěma páry běhu programu, tedy mezi dvěma body předchozího grafu –  $L''(K)$ . Poslední graf ukazuje  $\Delta K$  získané jako průměr absolutních hodnot z předchozího grafu za všechny běhy programu pro dané  $K$ .



Obrázek 12: Graf znázorňující podobnost mezi jednotlivými běhy programu Structure pro předem určené hodnoty K. Na svislé ose je logaritmus pravděpodobnosti modelu a optimálního rozdělení do daného počtu skupin. Na vodorovné ose jsou dané hodnoty K. Jednoznačný výsledek pro všechna opakování algoritmu jako je vždy u  $K = 1$ , lze víceméně nalézt i u  $K = 3, 4$  a  $5$ .



Obrázek 13: Graf ukazuje koeficient podobnosti (na svislé ose) mezi jednotlivými páry běhů programu pro předem určený počet shluků K (na vodorovné ose). Na grafu je znázorněn průměrný koeficient podobnosti mezi páry běhů programu a směrodatnou odchylkou. Nejlepší vychází rozdělení do 4 nebo 5 shluků, obojí s vysokým koeficientem podobnosti a malé směrodatné odchylce.

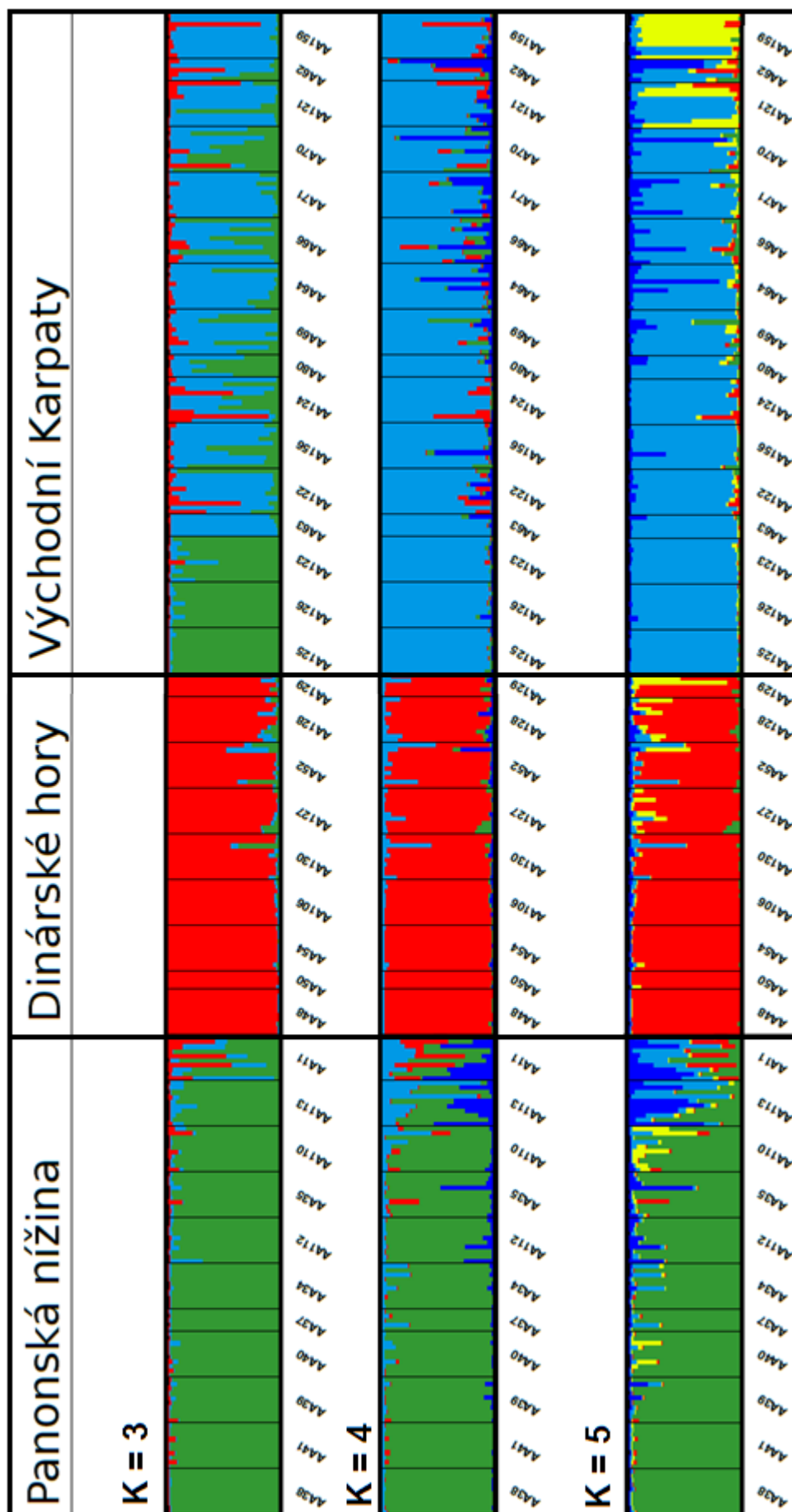
K určení nejlepšího počtu skupin bylo důležité vysoké  $\Delta K$  a také vysoký koeficient podobnosti (similarity coefficient, Falush et al. 2003) jednotlivých běhů programu vyjádřený malou směrodatnou odchylkou, což znamená, že při všech nezávislých opakování výpočtu dojde ke stejnému nebo velmi podobnému rozdělení jedinců do shluků. O rozdělení jedinců do 4 nebo 5 skupin svědčil vysoký koeficient podobnosti a logaritmus pravděpodobnosti, ale tato rozdělení neměla nejvyšší hodnotu  $\Delta K$ . Proto bylo bráno v úvahu i rozdělení dat do 3 skupin, které mělo jednoznačně nejvyšší  $\Delta K$ , ale velmi špatný koeficient podobnosti jednotlivých běhů a téměř nulovou směrodatnou odchylku. Výsledné rozdělení jedinců do 3, 4 a 5 shluků je znázorněno na obrázku 14.

V modelu rozdělení populací do 3 shluků ( $K = 3$ ) se zřetelně vymeziла skupina populací pocházející z Dinárských hor (červenou barvou na diagramech i v mapě – obrázky 14 a 15). Další skupinu vytvořily populace z oblasti Panonské nížiny spolu se třemi populacemi z jihu areálu (nejjižnější část Karpat, populace AA123, AA125, AA126; zelenou barvou na diagramech i v mapě – obrázky 14 a 15). Třetí a zároveň největší shluk zahrnoval zbývající populace z celého oblouku Karpat a také populace z Pobaltí (modrou barvou na diagramech i v mapě – obrázky 14 a 15). Několik populací z oblouku Karpat vykazovalo znaky smíšeného původu, a patřily zčásti do skupiny zahrnující panonské populace (zelená, pro  $K = 3$ ) a zčásti do skupiny zahrnující karpatské populace (modrá, pro  $K = 3$ ). Jednalo se o populace AA70 a AA118. Proporcionální příslušnost jednotlivých populací k jednotlivým shlukům je vyobrazena na mapě v obrázku 15.

U modelu  $K = 4$  byla (stejně jako u  $K = 3$ ) oddělena skupina populací z Dinárských hor (červenou barvou na diagramech i v mapě – obrázky 14 a 16). Taktéž byla oddělena skupina populací z Panonské nížiny, ovšem zde již bez tří populací z jihu Karpat (zelenou barvou na diagramech i v mapě – obrázky 14 a 16). Dále se objevil rozdíl mezi shlukem populací z jihovýchodní části Karpat (světle modrou barvou na diagramech i v mapě – obrázky 14 a 16) a shlukem populací ze západních Karpat (tmavě modrou barvou na diagramech i v mapě – obrázky 14 a 16). V tomto případě se Baltské populace zobrazily jako směs obou zmíněných skupin v přibližném poměru 3:2. I v případě modelu  $K = 4$  bylo několik populací nejednoznačně zařazeno a vykazovalo znaky smíšeného původu, nejčastěji se jednalo o populace ze Západních Karpat (obrázky 14 a 16).

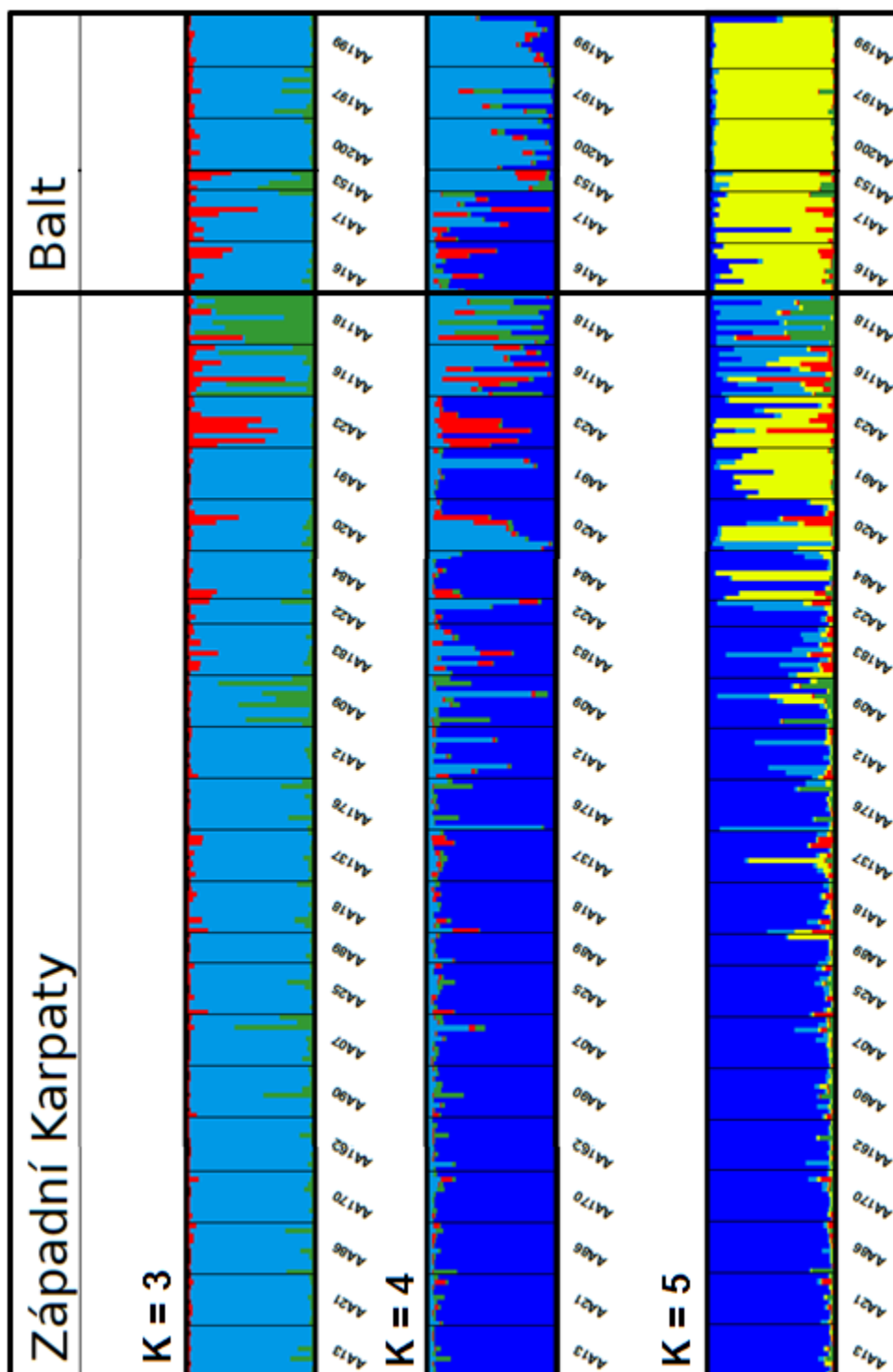
V modelu  $K = 5$  se vyzemily 4 populace z pobřeží Baltu (AA153, AA197, AA199 a AA200) do pátého shluku spolu se 2 populacemi ze Slovenska AA16 a AA17 (žlutou barvou na diagramech i v mapě – obrázky 14 a 17). Několik dalších populací z jihovýchodních a západních Karpat vykazovalo částečnou příslušnost k pátému shluku, například z území Slovenska (AA91) a Rumunska (AA159). Obě byly předchozími modely jednoznačně přiřazeny k populacím z Karpat.

Na základě výsledků těchto analýz, zejména podle modelu  $K = 5$  (obrázky 14 a 17), bylo definováno rozdělení populací do pěti hlavních skupin: Dinárské hory, Panonská nížina, Východní Karpaty, Západní Karpaty a Pobaltí.



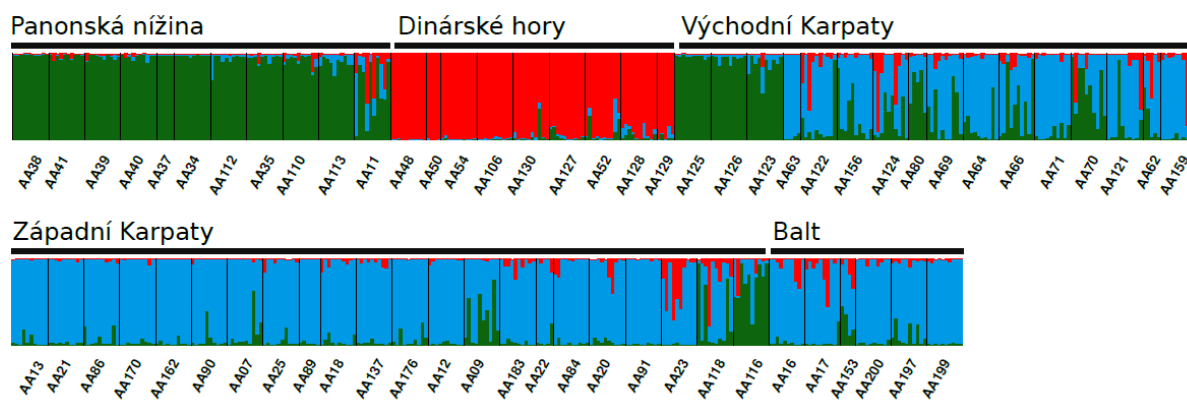
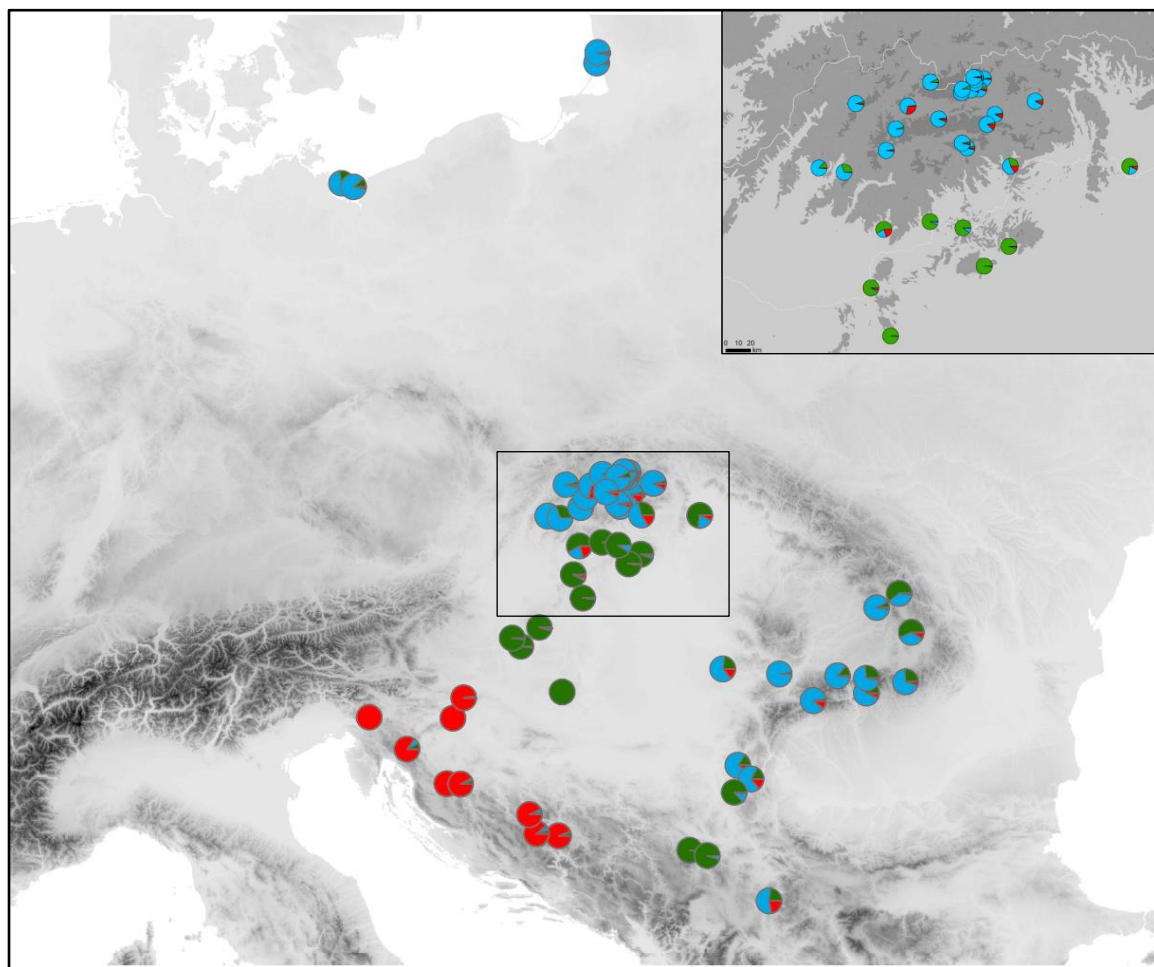
Obrázek 14 – 1. část: Shlukování jedinců do skupin (výstupy z programu Structure).



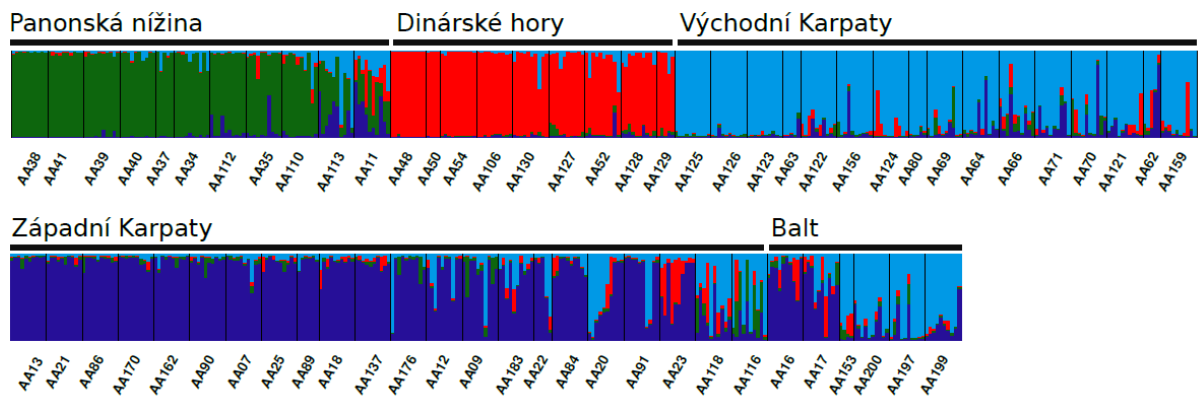
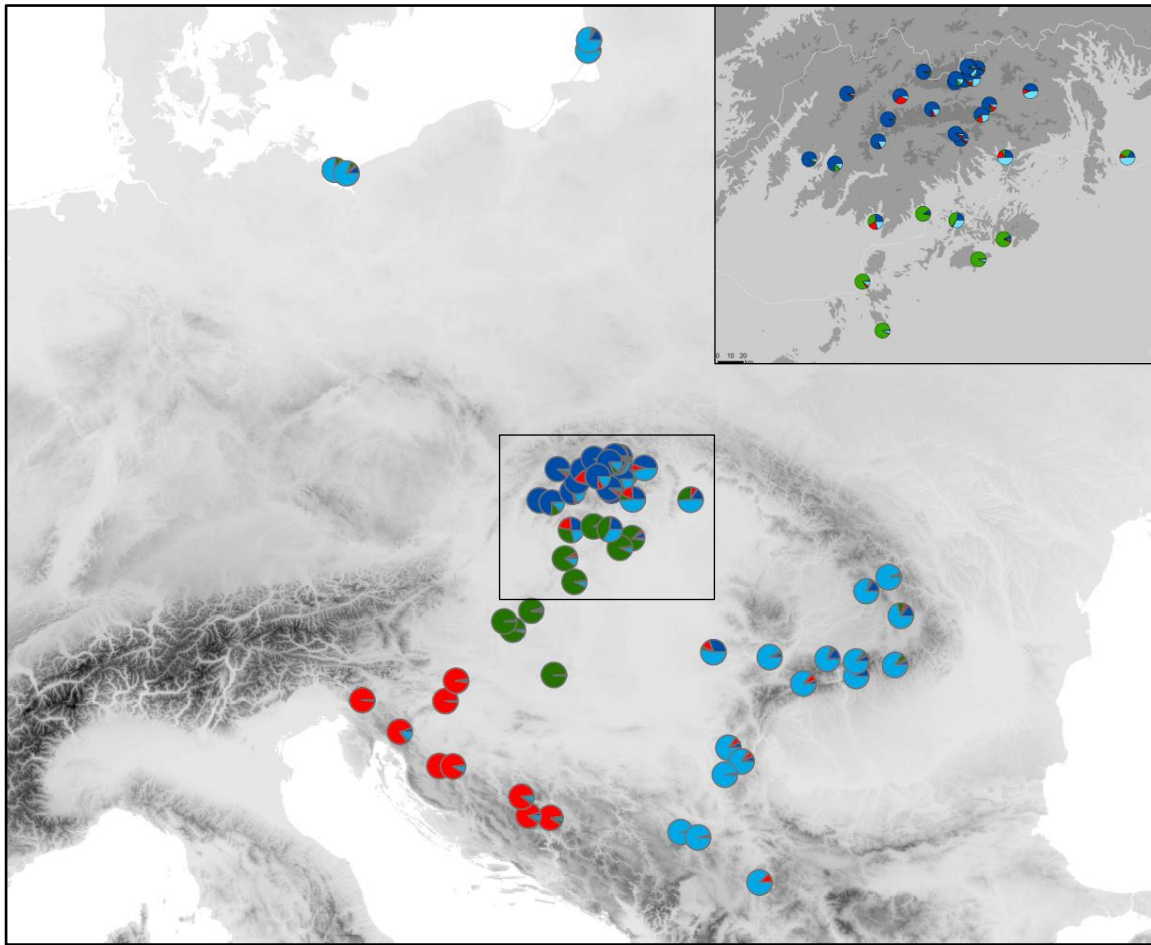


Obrázek 14 – 2. část: Shlukování jedinců do skupin podle genetické podobnosti (výstupy z programu Structure).

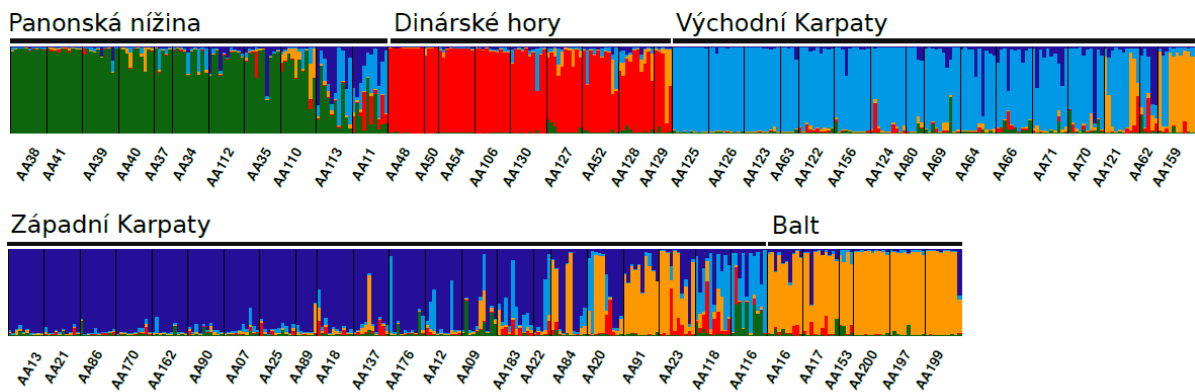
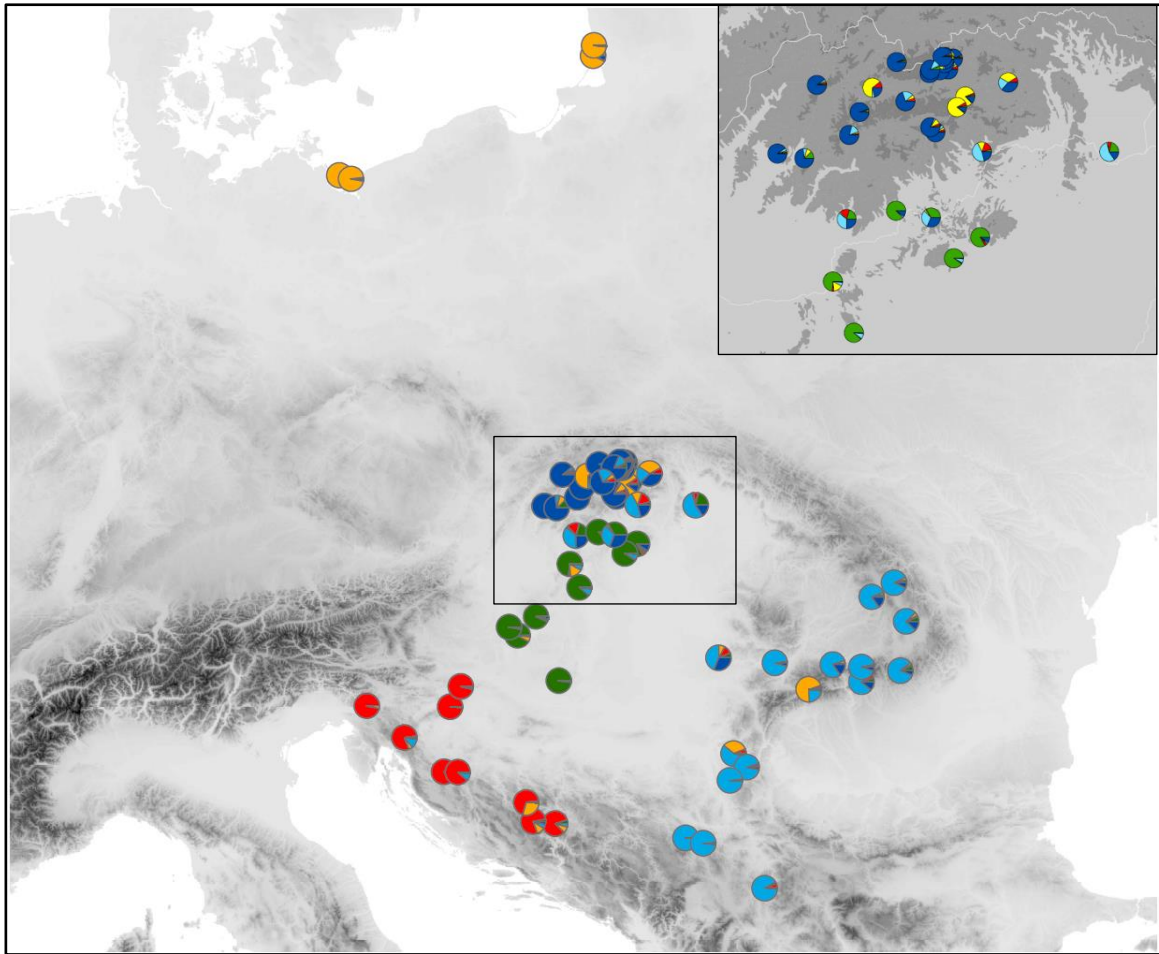
Jednotlivé sloupce jsou populace (označené kódy) a menší sloupečky uvnitř představují jedince. Sloupečky jsou obarvené podle příslušnosti k danému shluku (Panonská nížina, Dinárské hory, Východní Karpaty, Západní Karpaty a Balt). Pokud je barev více, pak poměr barev ve sloupci odpovídá pravděpodobnosti příslušnosti daného jedince do více barevně odlišených genetických shluků. Zobrazeny jsou varianty pro všechna 3 potenciálně vhodná K (3, 4 a 5).



Obrázek 15: Mapa znázorňuje geografické rozmístění zkoumaných populací *A. arenosa*, kruhové grafy ukazují jejich proporční přiřazení ke skupinám podle výsledků Structure – jedná se o model s  $K = 3$ . Rozdělení populací mezi skupiny je zobrazeno na diagramu pod mapou. Data byla zpracována v programu ArcMap v.10.



Obrázek 16: Mapa znázorňuje geografické rozmístění zkoumaných populací *A. arenosa*, kruhové grafy ukazují jejich proporční přiřazení ke skupinám podle výsledků Structure – jedná se o model s  $K = 4$ . Rozdělení populací mezi skupiny je zobrazeno na diagramu pod mapou. Data byla zpracována v programu ArcMap v.10.



Obrázek 17: Mapa znázorňuje geografické rozmístění zkoumaných populací *A. arenosa*, kruhové grafy ukazují jejich proporční přiřazení ke skupinám podle výsledků Structure – jedná se o model s  $K = 5$ . Rozdělení populací mezi skupiny je zobrazeno na diagramu pod mapou. Data byla zpracována v programu ArcMap v.10.

#### 4.2.4. Analýza molekulární variance (AMOVA)

Pomocí analýzy datasetu ze všech 14 jaderných mikrosatelitových lokusů byla zjišťována úroveň vnitropopulační a mezipopulační diferenciace u diploidních zástupců komplexu *Arabidopsis arenosa* v celém jeho areálu rozšíření. Relativně vysoké hodnoty fixačního indexu ( $F_{ST}$  až 0,28), poukázaly na výraznou genetickou diferenciaci mezi populacemi *A. arenosa*.

Analýza molekulární variance byla počítána v programu Arlequin 3.5 nejprve pro celý dataset, pak pro dataset s předem nadefinovanými genetickými skupinami (podle výsledků Structure, viz také tabulka 1) a nakonec byla zvlášť analyzována každá z pěti zmíněných genetických skupin. Všechny analýzy byly provedeny metodou  $F_{ST}$ -like (podle podílu shodných alel) a výpočty byly testovány 1000 permutacemi. Analýza celého datasetu (viz tabulka 9) poukázala na vyšší variabilitu uvnitř populací (74,4 % celkové variability) než mezi jednotlivými populacemi (25,6 % variability), avšak poměrně vysoká hodnota fixačního indexu  $F_{ST}$  (0,26) naznačuje, že populace jsou relativně dobře odlišené a vzájemně izolované.

Tabulka 9: Analýza molekulární variance (AMOVA) všech populací spočítaná v programu Arlequin 3.5 pomocí metody  $F_{ST}$ -like (podle podílu shodných alel), s počtem permutací 1000.

Zdroje variability	Stupně volnosti	Suma čtverců	Variabilita komponent	Procento variability
Mezi populacemi	63	1597,35	1,21	25,6
Uvnitř populací	1098	3835,22	3,50	74,4
Celkem		5432,57	4,70	
$F_{ST} = 0,26$				

Účelem druhé analýzy (viz tabulka 10) bylo zjistit jaké množství variability je mezi genetickými skupinami, definovanými podle výsledků Structure (Západní Karpaty, Východní Karpaty, Panonská nížina, Dinárské hory a Pobaltí). Hodnota fixačního indexu je vysoká ( $F_{ST} = 0,28$ ), takže tato analýza molekulární variance potvrdila rozlišení datasetu na pět zmíněných skupin. Mezi populacemi v rámci genetických skupin bylo zjištěno 16 % a uvnitř populací 72,4 % celkové variability.

Tabulka 10: Analýza molekulární variance (AMOVA) všech populací s definovanými skupinami spočítaná v programu Arlequin 3.5 pomocí metody  $F_{ST}$ -like (podle podílu shodných alel), s počtem permutací 1000.

Zdroje variability	Stupně volnosti	Suma čtverců	Variabilita komponent	Procento variability
Mezi skupinami	4	565,25	0,56	11,6
Mezi populacemi v rámci skupin	59	1032,10	0,77	16,0
Uvnitř populací	1098	3835,22	3,50	72,4
Celkem		5432,57	4,85	
$F_{ST} = 0,28$				

Nakonec bylo provedeno pět jednotlivých analýz datasetů rozdělených podle genetických skupin (viz tabulky 11, 12, 13, 14 a 15) ke zjištění toho, do jaké míry jsou tyto skupiny odlišené na základě mikrosatelitových dat, jak jsou izolované a kolik je variability mezi nimi nebo v jejich rámci.

Míra mezipopulační genetické variability byla nejnižší u Baltské skupiny (konkrétně 13,4 %) a u skupiny ze Západních Karpat (13,6 %). Tyto dvě skupiny měly také nejnižší hodnoty indexu  $F_{ST}$  (pro skupinu z Pobaltí měl index hodnotu 0,13 a pro skupinu z Karpat hodnotu 0,14), což naznačuje pouze střední míru diferenciace obou skupin a naopak vysoké procento variability mezi jedinci uvnitř populací.

Naopak nejvyšší hodnota fixačního indexu ( $F_{ST} = 0,23$ ) a nejvyšší mezipopulační variabilita (22,9 %) byla zjištěna u Panonské skupiny. Obdobně vysokou míru diferenciace vykazovala i Dinárská a Východokarpatská skupina. U těchto skupin byly také zjištěny vysoké hodnoty celkové populační diferenciace ( $F_{ST}$  v rozmezí 0,21 – 0,22) a vysoká míra mezipopulační variability (v rozmezí 21,2 – 22,3).

Tabulka 11: Analýza molekulární variance populací ze skupiny Západní Karpaty, spočítaná v programu Arlequin 3.5 pomocí metody  $F_{ST}$ -like, s počtem permutací 1000.

Západní Karpaty				
Zdroje variability	Stupně volnosti	Suma čtverců	Variabilita komponent	Procento variability
Mezi populacemi	21	315,81	0,61	13,6
Mezi jedinci uvnitř populací	384	1480,34	3,86	86,4
Celkem		1796,16	4,47	
$F_{ST} = 0,14$				

Tabulka 12: Analýza molekulární variance populací ze skupiny Východní Karpaty, spočítaná v programu Arlequin 3.5 pomocí metody  $F_{ST}$ -like, s počtem permutací 1000.

Východní Karpaty				
Zdroje variability	Stupně volnosti	Suma čtverců	Variabilita komponent	Procento variability
Mezi populacemi	15	286,94	0,89	21,2
Mezi jedinci uvnitř populací	270	890,28	3,29	78,8
Celkem		1177,22	4,18	
$F_{ST} = 0,21$				

Tabulka 13: Analýza molekulární variance populací ze skupiny Panonská nížina, spočítaná v programu Arlequin 3.5 pomocí metody  $F_{ST}$ -like, s počtem permutací 1000.

Panonská nížina				
Zdroje variability	Stupně volnosti	Suma čtverců	Variabilita komponent	Procento variability
Mezi populacemi	10	218,81	0,99	22,9
Mezi jedinci uvnitř populací	197	652,41	3,31	77,1
Celkem		871,22	4,30	
$F_{ST} = 0,23$				

Tabulka 14: Analýza molekulární variance populací se skupiny Dinárské hory, spočítaná v programu Arlequin 3.5 pomocí metody  $F_{ST}$ -like, s počtem permutací 1000.

Dinárské hory				
Zdroje variability	Stupně volnosti	Suma čtverců	Variabilita komponent	Procento variability
Mezi populacemi	8	141,29	0,86	22,3
Mezi jedinci uvnitř populací	146	437,04	3,00	77,7
Celkem		578,33	3,86	
$F_{ST} = 0,22$				

Tabulka 15: Analýza molekulární variance populací ze skupiny Pobaltí, spočítaná v programu Arlequin 3.5 pomocí metody  $F_{ST}$ -like, s počtem permutací 1000.

Pobaltí				
Zdroje variability	Stupně volnosti	Suma čtverců	Variabilita komponent	Procento variability
Mezi populacemi	5	69,25	0,58	13,4
Mezi jedinci uvnitř populací	101	375,15	3,73	86,6
Celkem		444,40	4,31	
$F_{ST} = 0,13$				

Souhrnné informace o těchto skupinách jsou v tabulce 16, která shrnuje míru mezipopulační variability, hodnoty fixačního indexu  $F_{ST}$ , dále hodnoty DW indexu,  $H_E$  a  $H_S$  zprůměrované pro populace z daných skupin. U průměrných hodnot jsou udány směrodatné odchylky.

Tabulka 16: Index rarity (DW), heterozygosita ( $H_E$ ), genetická diverzita a výsledky AMOVA, sumarizované pro jednotlivé genetické skupiny. Směrodatné odchylky jsou vyznačeny „±“.

Skupiny	DW index	$H_E$	$H_S$	$F_{ST}$	Mezipopulační variabilita
Dinárské hory	0,49 ± 0,266	0,42 ± 0,055	0,50 ± 0,033	0,22	22,3 %
Pobaltí	0,41 ± 0,194	0,53 ± 0,066	0,61 ± 0,065	0,12	13,4 %
Panonská nížina	0,37 ± 0,178	0,47 ± 0,047	0,53 ± 0,036	0,23	22,9 %
Západní Karpaty	0,34 ± 0,093	0,55 ± 0,050	0,60 ± 0,045	0,14	13,6 %
Východní Karpaty	0,27 ± 0,091	0,47 ± 0,064	0,52 ± 0,060	0,21	21,2 %



## 5. Diskuze

Tato práce se zabývala zevrubnou genetickou analýzou diploidní linie komplexu *Arabidopsis arenosa* a to za použití klasické molekulární metody analýzy mikrosatelitů. Hlavním cílem práce bylo tedy prozkoumat diverzitu a evoluční vztahy mezi diploidními liniemi komplexu *A. arenosa*, analyzovat jeho současné rozšíření a diskutovat možná taxonomická řešení pro tento komplex. Dále pak porovnat výsledky analýz provedených na základě mikrosatelitových dat v této práci s výsledky obdobných analýz provedených na identickém datasetu avšak rozdílnou metodou (RAD sekvenování), publikovaných nedávno ve studii Kolář et al. (2016).

Na základě studia mikrosatelitů bylo odhaleno 5 evolučních linií na diploidní úrovni druhového komplexu *Arabidopsis arenosa*. Z analýzy dat v programu Structure vyplynulo, že nejpravděpodobnějším modelem je právě rozdělení dat (populací) na 5 skupin, které víceméně kopírují rozmístění populací do geografických oblastí. Jedná se o populace z Dinárských hor, Panonské nížiny, západní části Karpat, východní (a jižní) části Karpat a pobřeží Baltského moře. Toto rozdělení víceméně koresponduje se strukturou, která byla odhalena pomocí metody RAD sekvenování (Kolář et al. 2016).

Populace z Dinárských hor byly zjevně na svém území dlouhodobě izolované, protože mezi nimi a dalšími horskými populacemi v Karpatech se nachází Panonská nížina. Právě tyto dvě oblasti byly nejodlišnější od zbytku datasetu – vymezily se již při modelu předpokládajícím pouze 3 skupiny. Třetím shlukem pak byly populace z Karpat spolu se 4 populacemi z Pobaltí, které jsou geograficky velmi vzdálené, ale tato analýza naznačila, že geneticky jsou si naopak docela blízké. Na jejich příbuznost poukazuje i nízká složka variability mezi skupinami (pouze 7,4 %) v analýze AMOVA, v níž byly zahrnuty pouze populace z Karpat a Pobaltí (Kolář et al. 2016). Analýza hlavních koordinát (PCoA) sice neodhalila výraznou strukturu mezi populacemi *Arabidopsis arenosa*, ale potvrdila příbuznost populací z Karpat a Pobaltí, a také izolovanost Dinárské skupiny.

Jednotlivé skupiny se mezi sebou liší svou vnitřní strukturou (tj. mírou mezipopulační variability), mírou genetické diferenciaci ( $F_{ST}$ ), heterozygositou a genovou diverzitou a množstvím vzácných alel. Podle hodnot  $H_E$  a  $H_S$  bylo odhadnuto, že nejvíce genetické diverzity se nachází v Západních Karpatech a v Pobaltí. Přičemž

v Pobaltí je to celkem sporné, protože průměrné hodnoty pro celou skupinu byly zvýšeny díky hodnotám populací AA16 a AA17, které se nacházejí na Slovensku a k Pobaltí byly přiřazeny pouze na základě Bayesovského shlukování (Structure). Průměrná hodnota  $H_E$  pro baltskou skupinu bez zahrnutí populací AA16 a AA17 (tj. s přihlédnutím ke geografii) je 0,49 a s těmito populacemi 0,53 (analogicky pro  $H_S$  jsou hodnoty 0,58 a 0,61). Nejméně genetické diverzity bylo nalezeno v Dinárských horách ( $H_E = 0,42$  a  $H_S = 0,50$ ). Z hlediska vzácných alel se jako nejbohatší ukázala skupina populací z Dinárských hor a hned po ní skupina z Pobaltí, což bylo opět způsobeno zkreslením hodnot kvůli rozporu mezi výsledky Bayesovského shlukování a geografii. V případě Baltu byla hodnota DW indexu zvýšena zahrnutím populací AA16 a AA17 (0,41), bez těchto populací to bylo pouze 0,29. Naopak Dinárské hory nejsou příliš geograficky vzdálené od ostatních skupin, ale vysoký podíl vzácných alel poukazuje na dlouhodobé přežívání na daném území, možná přes období posledních glaciálů. Pokud by zde bylo glaciální refugium, pak se z něj ale nešířili jedinci k rekolonizaci okolních území, protože populace z této oblasti jsou i v současnosti geneticky izolované. Nejméně vzácných alel bylo detekováno v jihovýchodní části Karpat, což vylučuje dlouhodobou stabilitu populací na tomto území, a naznačuje, že v tomto směru (střední oblast Balkánu) se zřejmě nenacházelo významné refugium diploidních zástupců *Arabidopsis arenosa* (podrobněji o „klasickém balkánském refugiu“ níže). Poměrně vysoká míra vzácných alel byla zjištěna také v Panonské nížině, která díky nížinnému charakteru zřejmě nebyla během doby ledové příliš významně ovlivněna klimatickými změnami, a také v Západních Karpatech. Výše diskutované hodnoty parametrů genetických skupin jsou přehledně shrnuty v tabulce 16.

Je možné, že na území Slovenska mohly populace *Arabidopsis arenosa* přežívat v tzv. kryptickém severním refugiu (Stewart & Lister 2001) a to konkrétně na lokalitách na úpatí západních Karpat (Kolář et al. 2016). Tato oblast již byla dříve autory zvažována jako vážný kandidát na kryptické refugium na sever od Mediteránu (Provan & Bennett 2008; Tzedakis et al. 2013). Tyto předpoklady jsou založeny na fosilních datech (Birks & Willis 2008) a genetické struktuře některých temperátních živočichů (Kotlík et al. 2006) a rostlin (Magri et al. 2006). Dále bylo zjištěno, že západní část Karpat byla během posledního glaciálního maxima (LGM) souvisle zalesněna (Jankovská & Pokorný 2008) a přesně datované fosilie plžů zde prokázaly výskyt temperátních druhů (Juříčková et al. 2014).

Na základě mikrosatelitových dat lze těžko získat přesné informace o genovém toku mezi populacemi, ale přesto je možné usuzovat, zda k němu vůbec dochází nebo někdy v minulosti docházelo. Nejlépe je to vidět na výsledcích Bayesovského shlukování, kde vznikají směsi („admixture“) v populacích, jejichž přiřazení k jedinému shluku nebylo možné, a naznačují jejich genetickou podobnost s populacemi z jiných shluků. Velká míra „admixture“ byla zjištěna právě mezi skupinami z Karpat (všeobecně) a z pobřeží Baltu (podle modelu ze Structure, kde  $K = 5$ ). Z toho by bylo možné usuzovat na genový tok nebo sdílený ancestrální polymorfismus mezi populacemi, ovšem v případě genového toku nelze odhadnout jeho směr. Jednou možností je směr z jihu na sever předpokládající postglaciální migraci zástupců *Arabidopsis arenosa* do severních oblastí Evropy. Baltské populace by v tomto případě musely vzniknout na území Karpat a mít společné předky s těmi „dnešními Karpatskými populacemi“. Teprve v průběhu času, s tím jak se populace přesouvaly na sever, došlo k přerušení genového toku, a tyto populace se začaly odlišovat (hromadit mutace a fixovat alely). Přesto na dnešní struktuře můžeme pozorovat jedince (nebo celé populace), které se nachází třeba na Slovensku nebo v Rumunsku, ale podle molekulárních markerů jsou blízce příbuzné populacím z Litvy. Druhou možností je směr genového toku ze severu na jih, čili od Baltu směrem do Karpat. Tato varianta se mi zdá mnohem méně pravděpodobná, protože obvykle se předpokládá genový tok ve směru rekolonizace po době ledové (Taberlet et al. 1998), ale pouze na základě mých analýz mikrosatelitů nelze tuto možnost vyloučit. Z modelu, který definoval pouze 4 shluky populací, je možné vypočítat genový tok mezi oběma částmi Karpat (západní a jihovýchodní) a také to, že Baltské populace patří spíše k jihovýchodní části. Celkově to naznačuje, že tyto tři skupiny jsou si navzájem příbuznější než každá z nich samostatně se zbývajícími dvěma skupinami. Určitý genový tok nelze vyloučit ani mezi populacemi v Panonii a v Karpatech (to je zjevné ze všech variant  $K = 3, 4$  i  $5$ ), a to hlavně mezi těmi, které jsou si geograficky poměrně blízké – například populace z Tater a populace z jihu Slovenska, kde začíná nížina. Zdá se, že jedinou opravdu dlouhodobě izolovanou skupinou (bez genového toku mezi ní a ostatními skupinami) jsou populace z Dinárských hor, ovšem pouze ty z území Slovinska, Chorvatska a Bosny a Hercegoviny. Poměrně zajímavé je, že populace ze Srbska (čili geograficky velmi blízké těm z Bosny a Hercegoviny) byly jednoznačně přiřazeny ke skupině z jihovýchodních Karpat.

Řada důkazů naznačuje, že linie *Arabidopsis arenosa* z Panonské nížiny a z Karpat pravděpodobně přežily poslední velké zalednění (LGM) v tzv. kryptických severních refugiích (Stewart & Lister 2001), jak bylo již výše zmíněno. Zaprvé, populace rostoucí na Balkánském poloostrově (tj. v klasicky uváděném glaciálním refugiu, např.: Bennett et al. 1991; Taberlet et al. 1998) vykazují značně sníženou míru genetické diverzity. Navíc již zmíněné populace ze Srbska mají také nízké procento vzácných alel a geneticky jsou příbuzné s populacemi z jihovýchodních Karpat, což může znamenat poměrně nedávnou kolonizaci území středního Balkánu právě z Karpat (spojení těchto dvou regionů bylo zdokumentováno i u dalších rostlinných druhů, viz např.: Frajman & Oxelman 2007; Ronikier 2011). Také se zdá, že malý počet zřetelně odlišných populací v severní části Dinárských hor přežíval v izolaci od těchto srbských populací. Zadruhé, bazální pozice Panonské skupiny ve fylogenetickém stromě (Kolář et al. 2016) vylučuje možnost nedávného oddělení panonských populací od těch balkánských. A za třetí, dlouhodobé přežívání velkých populací *Arabidopsis arenosa* ve střední a východní Evropě je podloženo zvýšenou mírou genetické diverzity a množstvím vzácných alel v současných populacích, což bylo zjištěno díky analýze mikrosatelitových dat (viz obrázek 7, část „b“) a také studiem AFLP a plastidové DNA (Schmickl et al. 2012). Opačný možný scénář, tj. nedávná (postglaciální) expanze z Balkánu, by způsobila slabě diferenciované a geneticky méně variabilní linie ve Střední Evropě, a možná také stopy po křížení všech hlavních linií v jižních oblastech domnělého refugia.

Tato zjištění jsou částečně v rozporu s poznatky získanými ze studia chloroplastové DNA komplexu *Arabidopsis arenosa* (Schmickl et al. 2012), které identifikovaly jako pravděpodobná glaciální refugia tyto 3 oblasti: nezaledněnou část východních a jihovýchodních Alp, Karpaty a Balkán. Nicméně analýzy mikrosatelitů i SNPs (Kolář et al. 2016) jednoznačně vyloučily, že by oblast Balkánu byla refugiem pro všechny diploidy *Arabidopsis arenosa*.

Informace o vnitropopulační a mezipopulační diverzitě získané na základě dat z analýzy mikrosatelitů jsou mnohem podrobnější a přesnější než z dat získaných RAD sekvenováním. Důvodem je to, že metoda RAD sekvenování je finančně náročná, a proto její „sampling“ zahrnoval jen 2 – 4 jedince z každé populace, což není dostačující pro posouzení vnitropopulační a mezipopulační diverzity. V tomto případě je metoda mikrosatelitů výhodnější, protože umožňuje levněji analyzovat více jedinců z populací. Metoda RAD sekvenování je naopak užitečná ke zjišťování vztahů mezi populacemi –

SNPs umožnily vytvoření fylogenetických stromů pro *Arabidopsis arenosa* (viz Kolář et al. 2016), což se na základě dat z mikrosatelitových lokusů nepodařilo.

I přes rozdíly v obou metodách, vyšlo rozdělení dat do skupin podle genetické příbuznosti (v analýze programem Structure) téměř shodně, což je pozoruhodné vzhledem k faktu, že metoda RAD sekvenování vyprodukovala cca 2400 variabilních míst (SNPs) zatímco u mikrosatelitů bylo nalezeno „pouze“ 220 různých alel. V rámci těchto podobně definovaných skupin odhalila analýza mikrosatelitů větší míru „admixture“ než v případě dat z RAD sekvenování a odhalila určitou míru genového toku např. mezi oblastí Západních Karpat a Panonskou nížinou nebo Pobaltím.

## 6. Závěr

Na základě studia mikrosatelitů se mezi 64 populacemi z celého areálu rozšíření diploidních zástupců *Arabidopsis arenosa* podařilo rozlišit 5 geneticky odlišných linií, které se alopatricky vyskytují v různých částech střední a jihovýchodní Evropy (jedna z nich více prostorově izolovaná od ostatních se vyskytuje na jižní části pobřeží Baltského moře). Rozmístění těchto genetických linií z velké části odpovídá významným geografickým regionům, podle kterých byly také provizorně pojmenovány: Dinárské hory, Panonská nížina, Západní Karpaty, Východní Karpaty a Pobaltí.

Genetická struktura druhového komplexu *Arabidopsis arenosa* nastíněná v této práci je v přímém rozporu s dosavadním taxonomickým konceptem. Tři ze šesti rozeznávaných taxonů podle Schmickl et al. (2012) zapadají do geneticky definované skupiny západních Karpat, zatímco výrazně odlišené skupiny z Dinárských hor a Karpat jihovýchodních jsou v současném přístupu zcela opomíjeny. Jediným taxonem, který „pasuje“ do jedné z 5 hlavních skupin, je *Arabidopsis petrogena* (A. Kern) V. I. Dorof korespondující s populacemi v Panonské nížině (Kolář et al. 2016).

## 7. Literatura

- Aikawa, S. et al., 2010. Robust control of the seasonal expression of the Arabidopsis FLC gene in a fluctuating environment. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), pp.11632–7.
- Al-Shehbaz, I.A. & O’Kane, S.L., 2002. Taxonomy and Phylogeny of Arabidopsis (Brassicaceae). *The Arabidopsis Book*, 6(1), p.1.
- Al-Shehbaz, I.A., O’Kane, S.L. & Price, R.A., 1999. Generic placement of species excluded from Arabidopsis (Brassicaceae). *Novon*, 9(3), pp.296–307.
- Arnold, B., Kim, S.T. & Bomblies, K., 2015. Single geographic origin of a widespread autotetraploid Arabidopsis arenosa lineage followed by interploidy admixture. *Molecular Biology and Evolution*, 32(6), pp.1382–1395.
- Avise, J.C. et al., 1987. Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics*, 18(1), pp.489–522.
- Baird, N.A. et al., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3(10).
- Baxter, S.W. et al., 2011. Linkage mapping and comparative genomics using next-generation rad sequencing of a non-model organism. *PLoS ONE*, 6(4).
- Bell, C.J. & Ecker, J.R., 1994. Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. *Genomics*, 19, pp.137–144.
- Bennett, K., Tzedakis, P. & Willis, K., 1991. Quaternary Refugia of North European Trees. *Journal of Biogeography*, 18(1), pp.103–115.
- Bennetzen, J.L., 1999. Plant genomics takes root, branches out. *Trends in Genetics*, 15(3), pp.85–87.
- Birks, H.J.B. & Willis, K.J., 2008. Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity*, 1(2), pp.147–160.
- Bleeker, W., 2003. Hybridization and *Rorippa austriaca* (Brassicaceae) invasion in Germany. *Molecular Ecology*, 12(7), pp.1831–1841.
- Bus, A. et al., 2012. High-throughput polymorphism detection and genotyping in Brassica napus using next-generation RAD sequencing. *BMC Genomics*, 13(1), p.281.

- Bushell, C., Spielman, M. & Scott, R.J., 2003. The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. *The Plant cell*, 15(6), pp.1430–42.
- Clauss, M.J., Cobban, H. & Mitchell-Olds, T., 2002. Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). *Molecular Ecology*, 11(3), pp.591–601.
- Clauss, M.J. & Koch, M.A., 2006. Poorly known relatives of *Arabidopsis thaliana*. *Trends in Plant Science*, 11(9), pp.449–459.
- Corander, J. et al., 2013. BAPS : Bayesian Analysis of Population Structure. Manual v 6.0. *Bioinformatics*, pp.1–28.
- Davey, J.W. et al., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), pp.499–510.
- Dieckmann, U. & Doebeli, M.O., 1999. On the origin of species by sympatric speciation. *Nature*, 400(22), pp.354–357.
- Dieringer, D. & Schlötterer, C., 2003. MICROSATELLITE ANALYSER (MSA): A platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, 3(1), pp.167–169.
- Doyle, J.J. & Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, pp.11–15.
- Eaton, D.A.R. & Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from glowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, 62(0), pp.689–706.
- Ehrich, D., 2006. AFLPDAT: A collection of R functions for convenient handling of AFLP data. *Molecular Ecology Notes*, 6(3), pp.603–604.
- Ehrich, D. et al., 2007. Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Molecular Ecology*, 16(12), pp.2542–2559.
- Emerson, K.J., Merz, C.R. & Catchen, J.M., 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, 107(37), pp.16196–16200.
- Evanno, G., Regnaut, S. & Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), pp.2611–2620.



- Excoffier, L. & Lischer, H.E.L., 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), pp.564–567.
- Excoffier, L., Smouse, P.E. & Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 131(2), pp.479–491.
- Falush, D., Stephens, M. & Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164(4), pp.1567–1587.
- Fér, T. & Hroudová, Z., 2009. Genetic diversity and dispersal of *Phragmites australis* in a small river system. *Aquatic Botany*, 90(2), pp.165–171.
- Frajman, B. & Oxelman, B., 2007. Reticulate phylogenetics and phytogeographical structure of *Heliosperma* (Sileneae, Caryophyllaceae) inferred from chloroplast and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, 43(1), pp.140–155.
- Harbaugh, D.T. et al., 2009. Genetic structure of the polymorphic *Metrosideros* (Myrtaceae) complex in the Hawaiian Islands using nuclear microsatellite data. *PLoS ONE*, 4(3).
- Haudry, A. et al., 2012. Disentangling the effects of breakdown of self-incompatibility and transition to selfing in North American *Arabidopsis lyrata*. *Molecular Ecology*, 21(5), pp.1130–1142.
- Hegarty, M.J. & Hiscock, S.J., 2005. Hybrid speciation in plants: new insights from molecular studies. *The New phytologist*, 165(2), pp.411–423.
- Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), pp.907–913.
- Hoffmann, M., 2005. Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution*, 59(7), pp.1425–1436.
- Hohenlohe, P.A. et al., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, 6(2).
- Hohmann, N. et al., 2014. Taming the wild: resolving the gene pools of non-model *Arabidopsis* lineages. *BMC evolutionary biology*, 14(1), p.224.
- Hollister, J.D. et al., 2012. Genetic Adaptation Associated with Genome-Doubling in Autotetraploid *Arabidopsis arenosa*. *PLoS Genetics*, 8(12).

- Hunter, B. & Bomblies, K., 2010. Progress and Promise in using *Arabidopsis* to Study Adaptation, Divergence, and Speciation. *The Arabidopsis book / American Society of Plant Biologists*, 8(8), p.e0138.
- Chybicki, I.J. & Burczyk, J., 2009. Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity*, 100(1), pp.106–113.
- Imbrie, J., 1985. A theoretical framework for the Pleistocene ice ages: William Smith Lecture. *Journal of the Geological Society*, 142(3), pp.417–432.
- Innan, H., Terauchi, R. & Miyashita, N., 1997. Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. *Genetics*, 146(4), pp.1441–1452.
- Jakobsson, M. et al., 2006. A unique recent origin of the allotetraploid species *Arabidopsis suecica*: Evidence from nuclear DNA markers. *Molecular Biology and Evolution*, 23(6), pp.1217–1231.
- Jankovská, V. & Pokorný, P., 2008. Forest vegetation of the last full-glacial period in the Western Carpathians (Slovakia and Czech Republic). *Preslia*, 80(3), pp.307–324.
- Jordon-Thaden, I. et al., 2010. Molecular phylogeny and systematics of the genus *Draba* (Brassicaceae) and identification of its most closely related genera. *Molecular Phylogenetics and Evolution*, 55(2), pp.524–540.
- Juříčková, L., Horáčková, J. & Ložek, V., 2014. Direct evidence of central European forest refugia during the last glacial period based on mollusc fossils. *Quaternary Research (United States)*, 82(1), pp.222–228.
- Kamm, A. et al., 1995. Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationships between *Arabidopsis* species. *Plant Molecular Biology*, 27(5), pp.853–862.
- Karl, R. et al., 2012. Systematics and evolution of Arctic-Alpine *Arabis alpina* (Brassicaceae) and its closest relatives in the eastern Mediterranean. *American Journal of Botany*, 99(4), pp.778–794.
- Koenig, D. & Weigel, D., 2015. Beyond the thale: comparative genomics and genetics of *Arabidopsis* relatives. *Nature reviews. Genetics*, 16(5), pp.285–298.
- Koch, M. a & Matschinger, M., 2007. Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15), pp.6272–6277.
- Koch, M. a. & Kiefer, C., 2006. Molecules and migration: Biogeographical studies in cruciferous plants. *Plant Systematics and Evolution*, 259(2-4), pp.121–142.

- Koch, M., Al-shehbaz, I.A. & Mummenhoff, K., 2003. Molecular Systematics, Evolution, and Population Biology in the Mustard Family (Brassicaceae). *Annals of the Missouri Botanical Garden*, 90(2), pp.151–171.
- Kolář, F. et al., 2015. Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae). *Biological Journal of the Linnean Society*.
- Kolář, F. et al., 2016. Genome-wide genetic differentiation driven by altitudinal niche divergence in the emerging plant model *Arabidopsis arenosa*. *Molecular biology and evolution*, (accepted).
- Kotlík, P. et al., 2006. A northern glacial refugium for bank voles (*Clethrionomys glareolus*). *Proceedings of the National Academy of Sciences of the United States of America*, 103(40), pp.14860–14864.
- Kron, P., Suda, J. & Husband, B.C., 2007. Applications of Flow Cytometry to Evolutionary and Population Biology. *Annual Review of Ecology, Evolution, and Systematics*, 38(1), pp.847–876.
- Leinonen, P.H., Remington, D.L. & Savolainen, O., 2011. Local adaptation, phenotypic differentiation, and hybrid fitness in diverged natural populations of *Arabidopsis lyrata*. *Evolution*, 65(1), pp.90–107.
- Lepais, O. et al., 2013. High Genetic Diversity and Distinctiveness of Rear-Edge Climate Relicts Maintained by Ancient Tetraploidisation for *Alnus glutinosa*. *PLoS ONE*, 8(9).
- Lihová, J. et al., 2007. Hybridization between two polyploid Cardamine (Brassicaceae) species in north-western Spain: Discordance between morphological and genetic variation patterns. *Annals of Botany*, 99(6), pp.1083–1096.
- López-Vinyallonga, S. et al., 2015. The polyploid series of the *Achillea millefolium* aggregate in the Iberian Peninsula investigated using microsatellites. *PLoS ONE*, 10(6).
- Lysak, M.A. et al., 2005. Chromosome triplication found across the tribe Brassicaceae. *Genome Research*, 15(4), pp.516–525.
- Magri, D. et al., 2006. A new scenario for the Quaternary history of European beech populations: Palaeobotanical evidence and genetic consequences. *New Phytologist*, 171(1), pp.199–221.

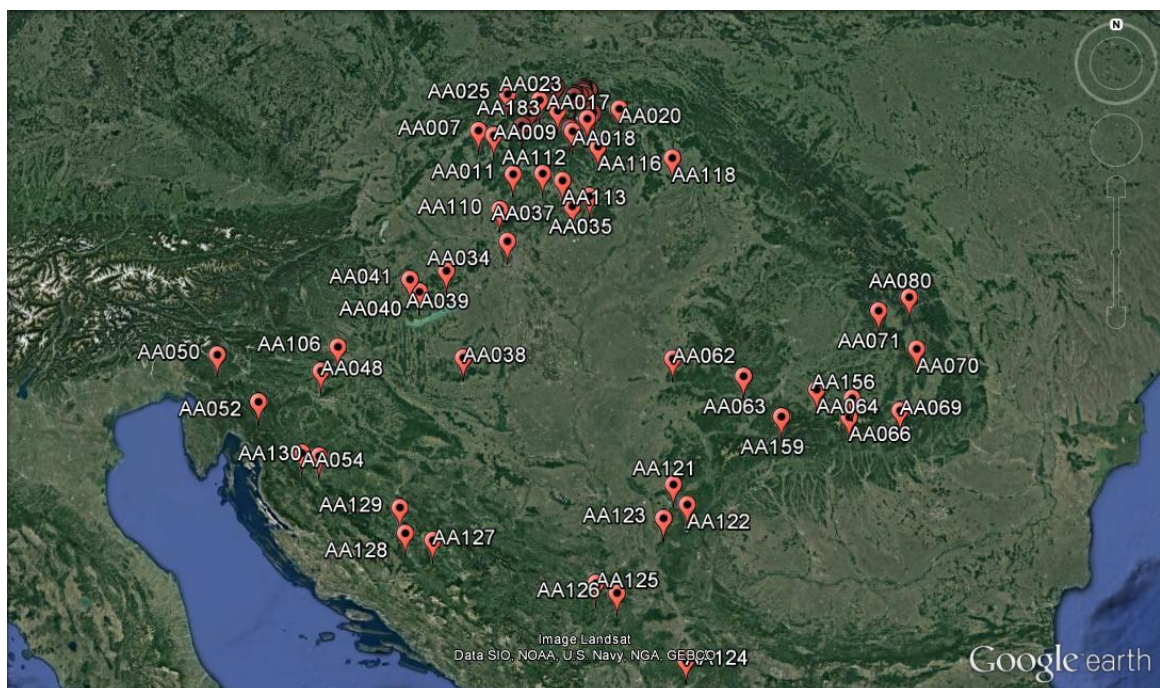
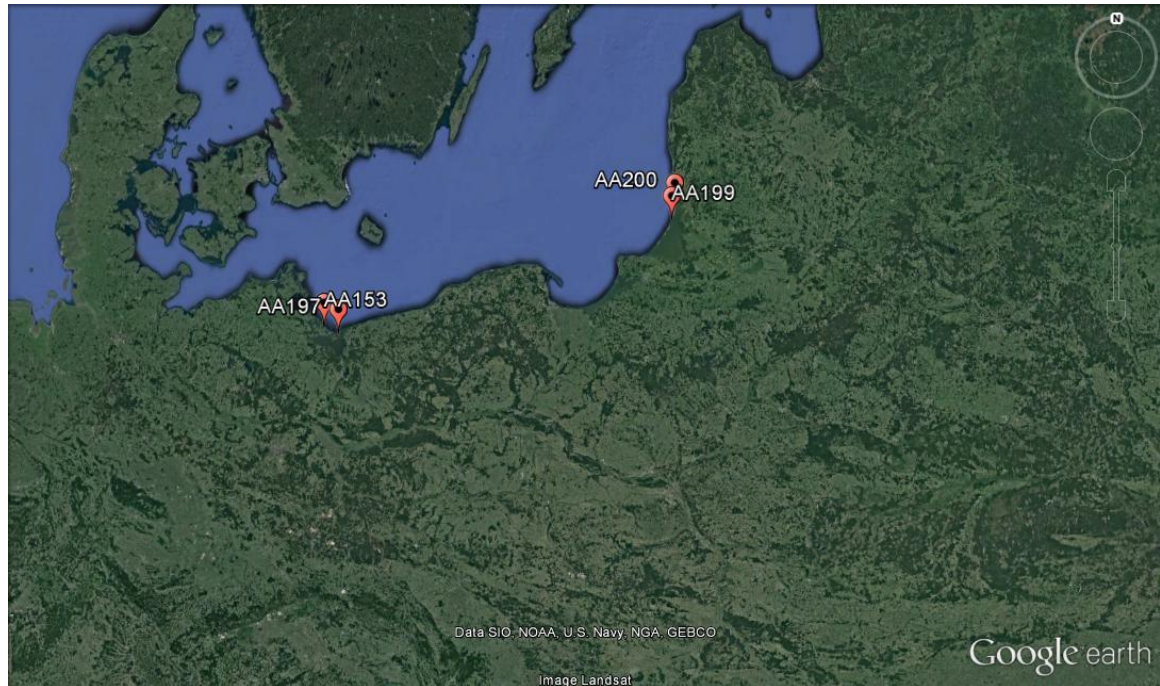
- Marhold, K. & Lihová, J., 2006. Polyploidy, hybridization and reticulate evolution: Lessons from the Brassicaceae. *Plant Systematics and Evolution*, 259(2-4), pp.143–174.
- Měsíček, J., 1970. Chromosome counts in *Cardaminopsis arenosa* agg. (Cruciferae). *Preslia*, 42, pp.225–248.
- Měsíček, J., 1998. *Cardaminopsis*. In: Marhold K, Hindák F, eds. Zoznam nižších a vyšších rastlin Slovenska – Checklist of non-vascular and vascular plants of Slovakia. Bratislava: *Veda*, pp. 395–396
- Měsíček, J. & Goliášová, K., 2002. *Cardaminopsis* (C. A. Mey.) Hayek. *Flóra Slovenska*, 4, pp.388–415.
- Meyer, C.L. et al., 2015. Intraspecific variability of cadmium tolerance and accumulation, and cadmium-induced cell wall modifications in the metal hyperaccumulator *Arabidopsis halleri*. *Journal of Experimental Botany*, 66(11), pp.3215–3227.
- Mitchell-Olds, T., 2001. *Arabidopsis thaliana* and its wild relatives: A model system for ecology and evolution. *Trends in Ecology and Evolution*, 16(12), pp.693–700.
- Mummenhoff, K. & Hurka, H., 1994. Subunit polypeptide composition of rubisco and the origin of allopolyploid *Arabidopsis suecica* (Brassicaceae). *Biochemical Systematics and Ecology*, 22(8), pp.807–811.
- Nei, M., 1987. *Molecular Evolutionary Genetics*, Columbia university press, 1987.
- O’Kane, S.L.J. & Al-Shehbaz, I.A., 1997. A Synopsis of *Arabidopsis* (Brassicaceae). *Novon*, 7(3), pp.323–327.
- Otto, S.P. & Whitton, J., 2000. Polyploid Incidence and Evolution. *Annual Review of Genetics*, 34(1), pp.401–437.
- Parisod, C., Holderegger, R. & Brochmann, C., 2010. Evolutionary consequences of autopolyploidy. *New Phytologist*, 186(1), pp.5–17.
- Paun, O. et al., 2008. Historical divergence vs. contemporary gene flow: Evolutionary history of the calcicole *Ranunculus alpestris* group (Ranunculaceae) in the European Alps and the Carpathians. *Molecular Ecology*, 17(19), pp.4263–4275.
- Peakall, R. et al., 1998. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Molecular biology and evolution*, 15(10), pp.1275–1287.
- Pritchard, J.K., Stephens, M. & Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp.945–59.

- Provan, J. et al., 1999. Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. *Molecular Ecology*, 8(3), pp.505–511.
- Provan, J. & Bennett, K.D., 2008. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, 23(10), pp.564–571.
- Ramsey, J. & Schemske, D.W., 2002. Neopolyploidy in Flowering Plants. *Annual Review of Ecology and Systematics*, 33(1), pp.589–639.
- Ronikier, M., 2011. Biogeography of high-mountain plants in the Carpathians: An emerging phylogeographical perspective. *Taxon*, 60(2), pp.373–389.
- Rosenberg, N.A., 2004. DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), pp.137–138.
- Rubin, B.E.R., Ree, R.H. & Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. *PLoS ONE*, 7(4).
- Säll, T. et al., 2004. Mode of reproduction in *Arabidopsis suecica*. *Hereditas*, 141, pp.313–317.
- Schaal, B.A. et al., 1998. Phylogeographic studies in plants: Problems and prospects. *Molecular Ecology*, 7(4), pp.465–474.
- Schlötterer, C. & Tautz, D., 1992. Slippage synthesis of simple sequence DNA. *Nucleic acids research*, 20(2), pp.211–215.
- Schmickl, R. et al., 2012. The evolutionary history of the *Arabidopsis arenosa* complex: Diverse tetraploids mask the Western Carpathian center of species and genetic diversity. *PLoS ONE*, 7(8), p.e42691.
- Schmickl, R. & Koch, M. A., 2011. *Arabidopsis* hybrid speciation processes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34), pp.14192–14197.
- Schönswetter, P. & Tribsch, A., 2005. Vicariance and dispersal in the alpine perennial *Bupleurum stellatum* L. (Apiaceae). *Taxon*, 54(3), pp.725–732.
- Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139(1), pp.457–62.
- Stewart, J.R. & Lister, A.M., 2001. Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, 16(11), pp.608–613.
- Taberlet, P. et al., 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7, pp.453–464.

- Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic acids research*, 17(16), pp.6463–71.
- Twyford, a D. & Ennos, R. a, 2012. Next-generation hybridization and introgression. *Heredity*, 108(3), pp.179–189.
- Tzedakis, P.C., Emerson, B.C. & Hewitt, G.M., 2013. Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology and Evolution*, 28(12), pp.696–704.
- Wang, Z. et al., 1994. Survey of plant short tandem DNA repeats. *Theoretical and Applied Genetics*, 88(1), pp.1–6.
- Warwick, S.I. & Al-Shehbaz, I. a., 2006. Brassicaceae: Chromosome number index and database on CD-Rom. *Plant Systematics and Evolution*, 259(2-4), pp.237–248.
- Weir, B.S. & Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), pp.1358–1370.
- Yant, L. et al., 2013. Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Current biology : CB*, 23(21), pp.2151–6.
- Zane, L., Bargelloni, L. & Patarnello, T., 2002. Strategies for microsatellite isolation: A review. *Molecular Ecology*, 11, pp.1–16.
- Zwan van der, C., Brodie, S.A. & Campanella, J.J., 2000. The intraspecific phylogenetics of *Arabidopsis thaliana* in worldwide populations. *Systematic Botany*, 25(1), pp.47–59.

## Přílohy

A. Mapy lokalit sbíraných populací – bližší pohledy na pobřeží Baltského moře (horní obrázek) a na střední a jižní část areálu rozšíření diploidní linie *Arabidopsis arenosa*





B. Tabulky souhrnných informací o multiplexech primerů

barva	NED	VIC	6-FAM	PET	VIC	6-FAM	NED
lokus	F20D22	F21M12	F19G10	F19K23-483	ICE12	ICE11	MDC16
opakovaná jednotka	GTTT	GAAA	GT	TTC	CT	GA	GA
forward primer	CCCAAGTGAGTC TGGTTC	GGCTTCTCGAAAT CTGTCC	AGTTGGTCCTCGA GCTCTCC	GGCTAAATTGCCG TTGTTGC	CTCATGGCAAAAAG AGGGAAA	TTTCAAGTTGAGA AGTGGAGTG	GAGTGGCCTCGTG TAGAGAAAAG
reverse primer	AACAAAATGAGTT TCTCTGCATG	TTACTTTTGCCTC TTGTCATTG	AAGAACTTAATTTT TCTCACCCG	GAATTCTGTAAACAT CCCAATTTCC	GCTCTCTCACCTCG AAGGTC	AAGAATTAGGCAA GAGTTTAGTGG	TGTCACCTCTTTTCC TCTGGTTTG
velikost alel	159-213	138-188	169-185	169-238	215-357	114-154	105-129
multiplex	M2	M2	M2	M2	M2	M2	M2

barva	NED	NED	6-FAM	6-FAM	VIC	6-FAM	PET
lokus	ICE13	AthZFPG (GC15)	ATTS0392 (GC16)	SLL2	ICE14	nga162	ICE7
opakovaná jednotka	ATC	CT	AAG	CA	GAT	GA	GAA
forward primer	GATCCTTCACCGG GTCTTG	TTGCGTTCCACAT TTGTTT	TTTGGAGTTAGAC ACGGATCTG	CATGTACTGGGAT TCAGTGTC	TCGAGGTGCTTTCT GAGGTT	CATGCAATTTGCGAT CTGAGG	TTCAAGGGCAGGA TCAAAAC
reverse primer	GTGGTGGAGACTC TTCGAGC	TGGGTCAATTAC ATGTAGAGA	GTTGATGCGAGCT TGATAAGC	CGTCCCTTTGTGTGG TTACACG	TACCTCACCCCTTT GACCCA	CTCTGTCACTCTTT TCCTCTGG	GTCTCACTGCTATC GTCACAGG
velikost alel	210-282	116-168	129-174	258-308	200-245	71-93	90-153
multiplex	M1	M1	M1	M1	M1	M1	M1



## C. Laboratorní protokoly

### C.1. Izolace DNA

- ze sušené biomasy (skladované v silikagelu) odebereme cca 0,5 g a vložíme do plastových zkumavek (Eppendorf, 2 ml) spolu se dvěma wolfram-karbidovými kuličkami
- zkumavky rovnoměrně umístíme do mlýnku Retsch MM200 a drtíme při frekvenci 30 ot./sek. po dobu 3 minut (v případě nutnosti ještě další 2 minuty)
- po rozdrčení materiálu pracujeme v digestoři a přidáme ke každému vzorku 900 µl roztoku CTAB (smíšeného v poměru 100:2 s merkaptoethanolem), 5 µl RNasy A a špetku PVP (polyvinylpyrolidon)
- uzavřené zkumavky promícháme na vortexu (Scientific Industries Inc., Vortex-Genie 2) a inkubujeme na termomixeru (Eppendorf, Thermomixer comfort) 30 min při teplotě 60 °C a 1400 rpm
- zkumavky dáme na 6 min do centrifugy (Sigma, 4-16K Centrifuge) a nastavíme maximální otáčky (13200 rpm), poté přepipetujeme supernatant do čistých zkumavek a zbylé usazeniny zlikvidujeme
- ke vzorkům přidáme 500 µl směsi chloroformu a isoamylalkoholu (v poměru 24:1), uzavřené zkumavky 3x převrátíme a necháme 5 minut odstát
- zkumavky umístíme do centrifugy na 6 min při 13200 rpm a pak přepipetujeme supernatant do čistých zkumavek
- pak zopakujeme poslední dva kroky (přečištění chloroformem a isoamylalkoholem)
- poté přidáme 500 µl vychlazeného isopropanolu, uzavřené zkumavky 2x převrátíme a necháme minimálně 30 min v mrazáku při -20 °C
- zkumavky umístíme do centrifugy na 3 min při 13200 rpm, supernatant opatrně vylijeme a zkumavky (s DNA ve formě pelletu usazeného na dně) necháme chvíli schnout dnem vzhůru
- ke vzorkům přidáme 400 µl vychlazeného 96% ethanolu a inkubujeme je 15 min na termomixeru při 37 °C a 1200 rpm
- uzavřené zkumavky umístíme do centrifugy na 6 min při 13200 rpm a pak opět opatrně vylijeme supernatant

- ke vzorkům přidáme 200  $\mu\text{l}$  vychlazeného 70% ethanolu a necháme cca 5 min odstát
- uzavřené zkumavky umístíme do centrifugy na 5 min při 13200 rpm a pak naposled opatrně vylijeme supernatant
- otevřené zkumavky necháme 10-15 min vysychat a pak je umístíme na termomixer, kde je sušíme při teplotě 65 °C dalších minimálně 15 minut (vysušený pellet DNA lze opatrným poklepáním oddělit od stěny zkumavky)
- vysušenou DNA rozpustíme ve 200  $\mu\text{l}$  TE pufru při inkubaci na termomixeru (10 min, 60 °C a 600 rpm)
- DNA lze skladovat v lednici nebo dlouhodobě v mrazáku (až -80 °C)

## C.2. Rozpisy pro PCR

### C.2.1. Reakční mix pro MyTaq polymerázu

ddH <sub>2</sub> O	13,4 $\mu\text{l}$
MyTaq pufr	4,0 $\mu\text{l}$
Forward („F“) primer	0,13 $\mu\text{l}$
F primer – fluorescenčně značený FAM	0,5 $\mu\text{l}$
Reverse („R“) primer	0,5 $\mu\text{l}$
MyTaq DNAPolymeráza	0,2 $\mu\text{l}$

- k 18,73  $\mu\text{l}$  mixu byl přidán 1  $\mu\text{l}$  vzorku DNA

### C.2.2. Reakční mix pro REDTaq polymerázu

ddH <sub>2</sub> O	6,2 $\mu\text{l}$
hořečnaté ionty (Mg <sup>2+</sup> )	0,5 $\mu\text{l}$
Red Taq pufr	1,0 $\mu\text{l}$
dNTP	0,2 $\mu\text{l}$
směs primerů (F+R)	0,6 $\mu\text{l}$
REDTaq DNAPolymeráza	0,5 $\mu\text{l}$

- k 9  $\mu\text{l}$  mixu byl přidán 1  $\mu\text{l}$  vzorku DNA

### C.2.3. Reakční mix pro QIAGEN Multiplex PCR Kit

ddH <sub>2</sub> O	1,0 µl
2x QIAGEN Multiplex PCR Master Mix	2,5 µl
10x primer-mix	0,5 µl

- ke 4 µl mixu byl přidán 1 µl vzorku DNA

### C.2.4. Koncentrace primerů v primer-mixu M2

- F20D22 0,75 pmol/µl
- F21M12 0,75 pmol/µl
- F19G10 0,75 pmol/µl
- F19K23-483 1 pmol/µl
- ICE11 2 pmol/µl
- MDC16 0,75 pmol/µl

## C.3. Programy v termocykleru

### C.3.1. Program pro MyTaq polymerázu

- 95 °C 1 min
  - 95 °C 30 s
  - 55 °C 30 s
  - 72 °C 45 s
  - 95 °C 30 s
  - 53 °C 30 s
  - 72 °C 45 s
  - 72 °C 10 min
  - 10 °C teplota udržovaná až do otevření víka termocykleru
- 27x
- 8x

### C.3.2. Program pro REDTaq polymerázu

- 94 °C 1 min
  - 94 °C 30 s
  - 50 °C 30 s
  - 72 °C 40 s
  - 72 °C 20 min
  - 10 °C teplota udržovaná až do otevření víka termocykleru
- } 35x

### C.3.3. Program pro QIAGEN Multiplex PCR Kit

- 95 °C 15 min
  - 94 °C 30 s
  - 60 °C 90 s
  - 72 °C 1 min
  - 60 °C 30 min
  - 10 °C teplota udržovaná až do otevření víka termocykleru
- } 35x

### C.4. Přečištění octanem

- ke vzorkům (1  $\mu$ l PCR produktů) přidáme 1  $\mu$ l octanu sodného a 10  $\mu$ l 96% ethanolu
- krátce promícháme na vortexu a stočíme v centrifuze, pak necháme vzorky 20 min v klidu stát v mrazáku
- vzorky vložíme do předem vychlazené centrifugy (Sigma, 4-16K Centrifuge) a necháme točit při teplotě 4 °C po dobu 30 min a při maximálních otáčkách (3700 rpm)
- poté zkumavky opatrně obrátíme dnem vzhůru na vrstvu buničiny, vložíme zpět do chlazené centrifugy a pustíme na cca 0,5 min při 580 rpm – necháme vytéct supernatant
- ke vzorkům přidáme 100  $\mu$ l 70% ethanolu a centrifugujeme 5 min při otáčkách 3700 rpm

- zopakujeme předchozí krok (tj. necháme v centrifuze vytéct supernatant na buničinu) a pak vzorky sušíme v otevřených zkumavkách na termomixeru 6 min při teplotě 60 °C
- takto přečištěné vzorky lze pak skladovat v mrazáku nebo rovnou připravit na fragmentační analýzu

#### C.5. Příprava na fragmentační analýzu

- ke každému vzorku přidáme 10  $\mu$ l směsi Formamidu (iontové rozpouštědlo) a vnitřního žebříčku (GeneScan™ 600 LIZ® Size Standard) dle následujícího rozpisu:
  - Formamid 10  $\mu$ l
  - LIZ 600 0,25  $\mu$ l
- DNA ve vzorcích zdenaturujeme při teplotě 95 °C za 3 min
- lze skladovat krátkodobě v lednici

#### C.6. Příprava na sekvenování mikrosatelitů

- přečištěné vzorky (PCR produkty) naředíme 5  $\mu$ l dvakrát destilovanou vodou (ddH<sub>2</sub>O)
- roztoky Reverse primerů naředíme na koncentraci  $c = 3,2$  pmol/ $\mu$ l
- směs pro automatický sekvenátor vytvoříme podle následujícího rozpisu:
  - R-primer 1  $\mu$ l
  - vzorek 1  $\mu$ l
  - ddH<sub>2</sub>O 6  $\mu$ l

## D. Tvorba dimerů v rámci reverse primeru F19G10 pěti různými způsoby

