

**Jihočeská univerzita v Českých Budějovicích  
Přírodovědecká fakulta**

**Populační struktura chobotnice *Eledone Moschata*  
ve Středozemním moři na základě RAD seq dat**

Bakalářská práce

**Daniela Kotalová**

Školitelka: Mgr. Marie Drábková

Školitelka – specialistka, konzultantka: Mgr. Pavlína Kočová

České Budějovice 2019

## **Bibliografické údaje**

Kotalová D., 2019: Populační struktura chobotnice *Eledone Moschata* ve Středozezemním moři na základě RAD seq dat. [Population structure of octopus *Eledone moschata* in the Mediterranean sea on the basis of RAD seq data. Bc. Thesis, in Czech.] - 50 p., Faculty of Science. University of South Bohemia, České Budějovice, Czech Republic.

## **Anotace**

By analysis of mitochondrial DNA and the genomic method by reduced representation genomic methods (ddRAD and high-throughput sequencing), this study explores the population structure of *Eledone moschata* in the Mediterranean, especially in Tyrrhenian and Adriatic sea.

## **Prohlášení**

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

Prohlašuji, že v souladu s §47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdání textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem odhalování plagiátů.

V Českých Budějovicích dne 17. 4. 2019

## **Poděkování**

Tímto bych ráda poděkovala své školitelce Marii Drábkové za odborné vedení, za připomínky a rady při řešení dané problematiky, ale také za vstřícnost, trpělivost a za čas, který mi věnovala. Také bych ráda poděkovala Janu Štefkovi, Pavlíně Kočové a Jakubovi Vlčkovi za cenné rady. Velké díky také patří holkám na bytě i ostatním přátelům, kteří vždy našli správná slova pro povzbuzení a rodičům za podporu kterou mi po celou dobu poskytovali.

# Obsah

1. Úvod.....	1
2. Literární rešerše.....	2
2.1. Next-gen sekvenování.....	2
2.1.1. Illumina.....	3
2.2. Metody redukující genom.....	4
2.2.1. CRoPS, (van Orsouw et al., 2007).....	4
2.2.2. RRLs (van Tassell et al., 2008).....	4
2.2.3. GBS (Elshire <i>et al.</i> , 2011).....	5
2.2.4. RAD seq.....	5
2.3. Zpracování RAD seq.....	8
2.4. <i>Eledone moschata</i> .....	9
2.4.1. Základní informace.....	9
2.4.2. Distribuce.....	11
3. Cíle práce.....	12
4. Metody.....	13
4.1. Sběr vzorků.....	13
4.2. Extrakce DNA.....	15
4.3. Jedno genová analýza.....	15
4.3.1. PCR.....	15
4.3.2. Gelová elektroforéza.....	16
4.3.3. Přechištění.....	16

4.3.4.	Zpracování sekvencí .....	17
4.4.	Rad seq.....	18
4.4.1.	Příprava knihoven ddRAD.....	18
4.4.2.	Illumina .....	19
4.4.3.	Zpracování dat .....	21
4.4.3.1.	Stacks .....	21
4.4.3.2.	Finerad structure .....	27
5.	Výsledky.....	28
5.1.	Jednogenová analýza .....	28
5.1.1.	Fylogenetický strom .....	28
5.1.2.	Populační síť .....	30
5.2.	Celogenomová analýza .....	31
5.2.1.	Stacks.....	31
5.2.2.	FineRAD structure.....	36
6.	Diskuze.....	37
7.	Závěr.....	40
8.	Reference.....	41
9.	Přílohy .....	50

# 1. Úvod

Pro studium populační struktury je výhodné mít velké množství genetických dat, aby byla zajištěná vysoká informativní hodnota pro odlišení jemných populačních rozdílů. Populační studie se mohou zkoumat pomocí různých metod, například pomocí analýzy mitochondriálních genů anebo genomových metod. Pro výzkum populačních, ekologických, fylogenetických i fylogeografických studií se donedávna používaly hlavně mitochondriální geny, které však neposkytují dostatečné pokrytí genetické informace na zjištění jemné populační struktury organismů. Jedním z mitochondriálních genů používaných pro populační studie je cytochromoxidáza I (dále COI). Výhodou tohoto genu je jeho vysoká rychlost změny a variabilita uvnitř druhu (Cox a Hebert, 2001). V některých případech je možné použít COI i pro odlišení blízce příbuzných druhů (Hebert et al., 2003). Nyní se na populační studie začínají používat metody, které dokáží hustěji pokrýt genom a produkovat velké množství genomických dat. Tyto metody využívají SNP (jednonukleotidový polymorfismus, single-nucleotide polymorphism) jako genetické markery. SNP je definován jako variance nukleotidu na jedné pozici v genomu v rámci jedinců stejného druhu. Jejich výhodami jsou například distribuce po celém jaderném genomu, jejich asociace s neutrální nebo adaptivní variací a jejich relativně jednoduchá genotypizace. Díky SNP můžeme odhalit i velmi jemné populační struktury.

Přestože hlavonožci hrají klíčovou úlohu v mořských trofických sítích a ekosystémech (Clarke, 1996; Piatkowski et al., 2001), doposud bylo provedeno velmi málo výzkumů genetické struktury hlavonožců (viz například Cheng, 2015, De Luca et al. 2016, Morse et al., 2018). Tato práce se zabývá studií populační struktury *Eledone moschata* (Lamarck, 1798) v Tyrhénském a Jaderském moři. Studie je zajímavá pro svou unikátnost, jelikož populační struktura chobotnic pomocí RAD seq nebyla zatím studována. První část práce se věnuje studiu populace pomocí genu COI. V druhé části je populace studovaná pomocí genetických markerů získaných z celého genomu metodou RAD seq. Následně jsou vypočteny statistiky pro populace v každém moři a porovnány výsledky metod populačních studií na jednom markeru (COI) a napříč genomem (RAD seq).

## 2. Literární rešerše

Pro objevování, validaci a hodnocení genetických markerů v populacích jsou významné metody NGS (Next generation sequencing). NGS (masivně paralelní sekvenování) jsou metody, které umožňují rychlé a levné sekvenování celých genomů a produkují obrovské množství výstupních dat. V případě populačních studií, NGS využívá techniky redukující genom, jako jsou CROPS (complexity reduction of polymorphic sequences), RAD seq (Restriction site associated DNA sequencing), GBS (genotyping-by-sequencing) a ddRAD seq (double digest RAD seq), které účinně štěpí genom v homologních místech a umožňují tak identifikaci SNP a také jejich typizaci pro desítky nebo stovky tisíc markerů u desítek jedinců. Pomocí metod využívajících NGS ve spojení s technikami redukujícími genom je možné objevit tyto markery v každém genomu z jakéhokoliv organismu, dokonce i v populacích, ve kterých je k dispozici málo nebo žádné genetické informace.

### 2.1. Next-gen sekvenování

Genomický výzkum dnes vyžaduje hloubku informací, které klasické Sangerovo sekvenování nemůže poskytnout. Během posledních let tak došlo ve studiu populací k zásadnímu posunu směrem ke genomové analýze pomocí NGS, které produkuje velké množství dat a umožňuje sekvenování a porovnávání celých genomů. Před tímto zlomem Sangerovo sekvenování bylo nejpoužívanější metodou téměř dvě desetiletí a vedlo k řadě monumentálních úspěchů, včetně dokončení nejkompletnější sekvence lidského genomu (Collins et al., 2004).

Komerčně dostupné techniky z rodiny NGS poskytují firmy, kterými jsou například Roche/454, Illumina/Solexa, Life/APG a Helicos BioSciences (Metzker, 2009). Největší výhodou, kterou NGS může nabídnout, je schopnost produkovat relativně levně enormní množství dat. Schopnost sekvenovat celý genom mnoha organismů umožnila provádět rozsáhlé srovnávací a evoluční studie, které byly před několika lety nepředstavitelné. Touto metodou se například odhalila rychlost a vlastnosti nových spontánních mutací u *Drosophila melanogaster* (Keightley et al. 2009) nebo u *Caenorhabditis elegans* (Denver et al., 2009).

Je třeba zdůraznit, že úspěšný NGS projekt vyžaduje také odborné znalosti po bioinformatické stránce, aby byla zaručena vysoce kvalitní data a výsledky (Buermans

a Dunnen, 2014). V souvislosti s přívalem nových dat, která NGS produkuje, jsou kladeny vyšší požadavky na programování a používání infromatických zdrojů, jelikož jednoduché manuální zpracování kvůli množství dat se stává prakticky nemožné.

### **2.1.1. Illumina**

Mezi nejpoužívanější metody NGS patří Illumina (the Illumina Genome Analyzer; Bentley et al., 2008). Illumina, představená v roce 2006, je založena na konceptu SBS ('sequencing by synthesis') pro rychlé a relativně přesné sekvenování. Tento proces identifikuje báze (nukleotidy), které jsou přidávány do řetězce DNA. Každý přidávaný nukleotid má fluorescenční značku s barvou specifickou pro každou bázi. Po přidání do řetězce je tento fluorescenční signál vyzářen a zaznamenán citlivou kamerou. Můžeme tak určit pořadí DNA sekvence. Desítky milionů DNA fragmentů jsou amplifikovány a čteny současně. Díky tomuto sekvenování jsme schopni detekovat polymorfismy, včetně jednotlivých nukleotidových mutací, inzercí a delecí, změn počtu kopií a chromozomálních aberací, v genomové DNA.

Amplifikace a sekvenace se odehrávají na „flow cell“. Každá flow cell je rozdělena do drážek. Podle toho, jaká platforma Illuminy se využívá, je flow cell rozdělena do 1 (miSeq), 2 (HiSeq2500) nebo 8 (HiSeq2000, případně HiSeq2500 v rychlém módu) drážek. Na vnitřním povrchu drážek jsou kovalentně připevněny dva typy oligonukleotidů komplementární ke specifickým adaptérům, které jsou ligovány na fragmenty z genomových knihoven. Standardní způsob sekvenování trvá 3 - 6 dní, rychlý mód trvá 1 - 3 dny. Tato metoda produkuje 100-150 millionů paralelně sekvenovaných fragmentů a výstup obsahuje 7,5-35 Gb. Délka readů, které Illumina produkuje, může být 50, 100, 150 nebo 250 bp (pár bazí, base pair). Platforma Illumina MiSeq produkuje ready o velikosti až 300 bp. Velikost readů je dána přístrojem, na kterém amplifikace probíhá a jeho nastavením. Frekvence chybovosti Illumina MiSeq i HiSeq2000 je 0,1% a nejfrekventovanější chybou je substituce jednoho nukleotidu (Fox et al., 2014).

Pro získání přesnější/kompletnější infromace se využívá metoda paired-end. Tímto způsobem je možné dosáhnout vysoce kvalitních sekvenčních dat díky produkci dvojnásobného počtu readů. Tato metoda totiž sekvenuje oba konce fragmentů. Paired-end usnadňuje detekci genomického uspořádání a repetitivních elementů.



## **2.2. Metody redukující genom**

### **2.2.1. CRoPS, (van Orsouw et al., 2007)**

Metoda CRoPS (complexity reduction of polymorphic sequences) byla první metodou pro NSG na identifikaci polyformismů v jedincích. K sekvenování se používá malý soubor oblastí vybraných napříč genomem bez toho, aby se musel sekvenovat celý genom. Tato metoda začleňuje do ligovaných adaptérů identifikátory v podobě krátkých unikátních sekvencí DNA připojených ke každému fragmentu z dané knihovny/vzorku (barcodes), které jsou známy pod pojmem multiplex identifier sequences (MIDs) na Roche Genome Sequencer platformě (Margulies et al., 2005). Pro každý vzorek DNA se používá adaptér s rozdílnými barcody. Barcody se mohou používat na bioinformatické rozdělení readů a jejich přiřazení k jednotlivým vzorkům, což je výhodné u populačních studií pro sekvenování více jedinců v jedné drážce na NGS sequencing platformě.

### **2.2.2. RRLs (van Tassell et al., 2008)**

Metoda RRLs (reduced-representation libraries) byla nejprve použita pro generování SNP mapy lidského genomu (Altshuler et al., 2000) za použití kapilárního sekvenování. Posléze byla adaptovaná na NGS a byla používána ke generování desítek milionů SNPs na NGS platformě.

Pomocí této metody je genomová DNA z více jedinců často štěpena restrikcí enzymem. Fragменты jsou selektovány na velikost 300–700bp, následně smíchány a poté sekvenovány. V tomto případě jsou sekvenovány pouze konce fragmentů, ale protokol může být modifikován pro sekvenci celých fragmentů (Kerstens et al., 2009). Získá se částečné, ale celogenomové pokrytí za minimum nákladů. Je-li k dispozici vysoce kvalitní referenční genom mohou být údaje ze sekvenování se sníženou reprezentací mapovány na referenční genom, nebo může být provedena reference guided assembly (složení sekvencí vedené referenčním genomem).

### 2.2.3. GBS (Elshire *et al.*, 2011)

Technika GBS (genotyping-by-sequencing) je jednoduchá a vysoce multiplexní. Při multiplexingu se sloučí několik knihoven do jednoho sekvenčního běhu. GBS se používá pro populační studie a na mapování znaků v různých organismech. Tento postup je založen na next-gen sekvenování. GBS zahrnuje štěpení genomové DNA restrikcími enzymy a sekvenování konců všech výsledných fragmentů. Adaptéry s barcody a běžné adaptéry jsou ligovány na štěpené fragmenty a vytváří tak fragmenty s kombinací adaptérů: barcode + běžný adaptér, barcode + barcode a běžný adaptér + běžný adaptér. Vzorky jsou smíchány a amplifikovány na platformu Illumina (Pandey, Nutter a Prediger, 2008). Ne všechny fragmenty s navázanými adaptéry jsou sekvenovány, jelikož spousta fragmentů neprojde bridge amplifikací: buď adaptéry neobsahují specifické barcody a nebo jsou příliš dlouhé (>1kb). Pro sekvenování jsou amplifikovány pouze krátké vzorky obsahující kombinaci barcode + běžný adaptér. Avšak i tak je sekvenován velký počet kratších fragmentů.

### 2.2.4. RAD seq

RAD seq je metoda, která vytváří krátké fragmenty DNA, kdy je genomová DNA štěpena restrikcími enzymy. Tyto fragmenty jsou následně sekvenovány a zpracovány například pomocí programu Stacks (Catchen *et al.*, 2013). Při použití této metody se díky sestřihu v restrikcích místech snižuje komplexita genomu. Genom je reprezentován ready, ve kterých mohou být lehce detekována polymorfnní místa. RAD seq je jednoduchá a poměrně levná metoda, se kterou je možné identifikovat v organismu více než 13,000 SNPs a to i v rámci velkých populací (Baird *et al.*, 2008). SNPs jsou velice hojné genetické markery rozmístěné napříč genomem a díky tomu jsou uplatňovány v populační genetice a evoluční biologii ve studii dědičnosti genomových oblastí (Berger *et al.*, 2001, Stickney *et al.*, 2002). Díky technologii RAD seq, která sekvenuje stejné genomové oblasti napříč jedinci, je možné identifikovat velké množství SNP. Jeffries *et al.* (2016) porovnávali, zda pro průzkum populační struktury je lepší zkoumat málo markerů, jako jsou například mikrosatelity, na velkém množství vzorků, nebo více markerů (SNPs) na málo jedincích. I když RAD seq (restriction site-associated genomic DNA) obsahoval pouze 17,6% vzorků, prokázalo se, že je účinnější metodou při zkoumání populační struktury a vlivu IBD (isolation by distance) než mikrosatelity (Jeffries *et al.*, 2016).

RAD seq se používá například ve studii fylogeografie (Takahashi et al., 2015) a v populační genetice na genomické úrovni (Hohenlohe et al., 2010, Ren et al., 2016). RAD seq je také vhodný pro rekonstrukci fylogenetických vztahů mladších skupin/kládů, ve kterých je napříč druhy ponechán dostatečný počet orthologních restričních míst (Rubin et al., 2012). RAD seq dokáže řešit i vztahy mezi blízce příbuznými druhy (Díaz-Arce et al., 2016). Tuto techniku můžeme využívat na nemodelové organismy (Rodríguez-Ezpeleta et al., 2016, Wang et al., 2016, Yang et al., 2016) nebo na genotypování když referenční genom není dostupný, jako tomu je například u ekologických studií (například Andrews et al., 2016). Konkrétně metoda RAD seq byla použita při detekci populační struktury ve volně žijících populacích, jako jsou například klauni (Planes et al., 2009), sledi (Lamichhaney et al., 2012) a ančovičky (Zarraonaindia et al., 2012).

Pro kvalitnější alignmenty se využívá PE RAD seq (paired-end sequencing; Willing et al., 2011). Nevýhodou RAD seq je, že vyžadují vysokou kvalitu DNA (Suchan et al., 2016) a také mohou ve výsledku být mezery v genomové coverage. Byly vyvinuty další varianty pro lepší pokrytí genomu, jako například ddRADseq (Peterson et al., 2012), 2b-RAD (Wang et al., 2012), SLAF-seq (Sun et al., 2013) a hyRAD (Suchan et al., 2016).

### **ddRADseq - Double Digest RADseq**

ddRAD seq (double digest restriction-site associated DNA sequencing) je levná metoda pro objevení De Novo SNP a genotypování na modelových i nemodelových organismech (Valencia et al., 2018). Hlavním cílem této techniky je usnadnit genotypizaci napříč velkým počtem jednotlivců (stovky nebo více) pro celou řadu markerů (stovky až stovky tisíc). Díky své flexibilitě je tento modifikovaný přístup RAD seq použitelný pro velké množství biologických studií v širokém spektru organismů (Lal et al., 2016, Recknagel et al., 2013, Zhou et al., 2014).

ddRAD seq se od ostatních odlišuje v tom, že používá dva restriční enzymy následované přesnou selekcí velikosti. Získáváme tedy knihovnu obsahující pouze fragmenty blízko cílové velikosti. Ve srovnání s existujícími přístupy RADseq umožňuje ddRADseq větší flexibilitu, robustnost a také podstatný pokles nákladů (Peterson et al., 2012).

## **2b-RAD**

2b-RAD je zjednodušený a flexibilní přístup pro genotypování RAD. Používá restriční enzymy typu 2b. Tyto enzymy (například BsaXI a Alfi) štěpí genomovou DNA a produkují fragmenty s jednotnou délkou, které jsou ideální pro sekvenování na stávajících platformách Next-gen (Wang et al., 2012). Tato metoda se využívá například pro fylogenetické studie (Seetharam a Stuart, 2013, Wang et al., 2013).

## **SLAF-seq**

SLAF-seq je efektivním řešením pro rozsáhlé genotypování a k nalezení De Novo SNP. Tato metoda získala název díky specifickému lokusu amplifikovanému sekvenčním sekvenováním (SLAF-seq). Technologie SLAF-seq poskytuje hluboké sekvenování pro zajištění genotypové přesnosti (Sun, 2013). Tato metoda se používá například pro sestavení genetických map (Liu et al., 2016, Xu et al., 2015).

## **HyRAD**

Tato jednoduchá a nákladově efektivní metoda umožňuje sekvenování ortologních lokusů i ze vzorků s vysokou degradací DNA, což otevírá nové cesty výzkumu v oblasti muzejní genomiky (Linck et al., 2017). Nezávisle na přítomnosti restričního místa zlepšuje pokrytí vzorků mezi lokusy. Vzhledem k tomu, že se nespolehá na přítomnost restričního místa, není tedy citlivá na polymorfismy lokusů mezi vzorky v restričních místech, které obvykle způsobují výpadek lokusů (Suchan et al., 2016).

## **RAD marker microarray**

RAD marker microarray usnadňuje rychlé genetické mapování mutací. Použitím těchto mikročipů Miller et al. (2007) lokalizovali dříve nezmapované mutace u *Danio rerio* do genomových oblastí jen s přesností několika centiMorgans na délku. RAD mikročipy tedy umožňují screening tisíců markerů v jediné hybridizaci (Miller, 2007).

## 2.3. Zpracování RAD seq

### Stacks

Program Stacks se používá na zkoumání populační analýzy. Umožňuje analyzovat desítky tisíc genetických markerů. Je to software pipeline pro vytváření lokusů z krátkých sekvencí, kterými jsou například data generované na platformě Illumina. Stacks byl vyvinut pro práci s daty založenými na restričních enzymech za účelem tvorby genetických map a zkoumání populační genomiky a fylogeografie. Tento program byl využit pro zpracování RAD seq dat v nejrůznějších studiích. Jeho citace se objevuje již ve více než tisíci článkách.

Stacks účinně organizuje a sestavuje velké množství krátkých sekvencí z vícera vzorků. Stacks identifikuje lokusy v souboru jedinců, buď De Novo, nebo zalignované s referenčním genomem. Stacks obsahuje statistický model pro identifikaci sekvenčních polymorfismů a rozlišuje je od sekvenčních chyb. Stacks využívá katalog k zaznamenávání všech lokusů identifikovaných v populaci a přiřazuje jednotlivce do tohoto katalogu, aby určil, které alely haplotypu jsou přítomny na každém lokusu u každého jedince.

Stacks probíhá v šesti hlavních fázích. Nejprve jsou reads demultiplexovány a vyčištěny programem *process\_radtags*. Další tři etapy zahrnují hlavní pipeline Stacks: budování lokusů (*ustacks*), vytvoření katalogu lokusů (*cstacks*) a nakonec porovnávání s katalogem (*sstacks*). V páté etapě je spuštěn program *gstacks* za účelem assemblování a sloučení paired-end sekvencí. V poslední fázi je prováděn program *populations* v závislosti na typu vstupních dat. Program *populations* může generovat data v několika dalších formátech pro další populační, evoluční, fylogeografické a fylogenetické studie (Emerson et al., 2010, Querejeta et al., 2016 Lescak et al., 2015).

## 2.4. *Eledone moschata*

### 2.4.1. Základní informace

Tato práce pracuje s organismem *Eledone moschata*, neboli chobotnicí pižmovou (musky octopus) patřící do čeledi Octopodidae, kam patří hlavně bentické chobotnice. Chobotnice čeledi Octopodidae mají 8 ramen a nemají žádnou tělesnou schránku ani ploutve. Mnoho druhů se vyskytuje na kamenitém nebo korálovém útesu s velkým množstvím úkrytů. Jiné se ukrývají zahrabáním do sedimentu. Všechny chobotnice mají dobře vyvinutou slinnou žlázu, která obsahuje paralyzující neurotoxiny sloužící k znehybnění kořisti. Samci mají modifikované hektokotylové chapadlo, typicky je to třetí pravé rameno. Při páření samec tímto chapadlem vloží spermatofor do vejcovodu samice. Samice pak může uchovávat jeho sperma do té doby, než je připravena klást vajíčka. Všechny samice bentických chobotnic pečují o své snůšky a chrání je, dokud se nevylíhnou (Norman, 2000).

*Eledone moschata* (Obr. 1) má tělo dlouhé do 15 cm, ramena jsou dlouhá až 40 cm. Vyskytuje se ve Středozezemním moři. Chobotnici pižmovou lehce identifikujeme pomocí jedné řady přísavek na spodní straně ramen. Avšak jednu řadu přísavek má i sesterský druh *Eledone cirrhosa* vyskytující se sympatricky. Ale chobotnici *Eledone moschata* snadno rozpoznáme od druhu *Eledone cirrhosa* tím, že po těle i plovacích blanách má roztroušené černé tečky. Její obecný název (chobotnice pižmová, musky octopus) vznikl díky jejímu pižmovému pachu, který tato chobotnice vypouští po vyndání z vody. Dospělí samci mají na špičce ramen protažené přísavky, které jsou rozděleny do dvou řad. Samice klade méně než 500 velkých vajíček, které jsou 10 mm dlouhé (Norman, 2000). Přestože se jedná o komerční druh, který rybáři často nachází ve svých sítích, populační studie na organismu *Eledone moschata* nebyly zatím provedeny.

*Eledone moschata* se živí širokou škálou korýšů, měkkýšů a ryb. Korýši jsou preferováni a druhy, kterými se tato chobotnice živí zahrnují *Maja squinado*, *Maja crispata*, *Macropodia rostrata*, *Macropodia longirostris*, *Pisa tetraodon*, *Dorippe lanata*, *Lisa chiragra*, *Lambrus angulifrons*, *Lambrus massena*, *Inachus dorsettensis*, *Carcinus aestuarii*, *Pachygrapsus marmoratus*, *Xantho poressa*, *Pilumnus hirtellus*, *Goneplax rhomboides*, *Pagurus prideauxi*, *Ilia nucleus* a *Squilla mantis*; měkkýši, které *E. moschata* konzumuje jsou například *Chlamys varia*, *Mytilus galloprovincialis*, *Sepia orbigniyana*, *Illex coindetii*, *Loligo subulata*. *E. moschata* rovněž konzumuje ryby *Engraulis encrasicolus*, *Sardina*

*pilchardus*, *Mullus barbatus*, *Diplodus annularis*, *Merluccius merluccius*, *Merlangius merlangus*, *Lepidotrigla cavillon*, *Trachurus trachurus* a *Trisopterus minutus* (Şen, 2007). *E. moschata* během lovu používá hmatového vjemu k průzkumu okolí a reaguje i na hmatové podněty během počátečního kontaktu s kořistí (Belcari et al., 2002).

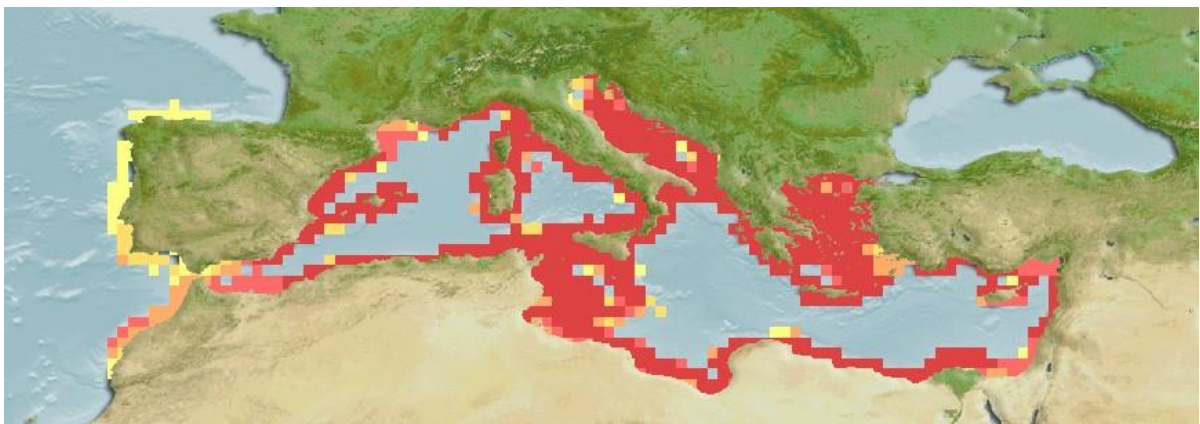


Obr. 1: Nákres *E. moschata* (Orbigny, 1845).

## 2.4.2. Distribuce

Vysoká míra migrace je pro hlavonožce typická z důvodu vysoké mobility. Obvykle mají velké populace s možností expanze. Z tohoto důvodu bývá jejich genetická struktura homogenní, protože migrace snižuje genetickou diferenciaci. Například u populace *Architeuthis* je úroveň nukleotidové diverzity mimořádně nízká (Winkelmann et al., 2013), což znamená, že se mezi jedinci nevyskytuje příliš velká genetická variace. *Architeuthis* postrádá rozeznatelnou strukturu pravděpodobně kvůli migraci pelagických paralarválních stádií, které jsou dispergované pomocí globální cirkulace termohalinního výměníku. Naproti tomu existují i druhy, jako například u *Sepia esculenta* (Zheng et al., 2009), s vyšší mírou genetické diverzity. Také u druhu *Octopus vulgaris* byla zjištěna diferenciaci mezi jedinci z centrálního a východního Středozemního moře (Fadhlaoui-Zid et al. 2012) a také v rámci východního Středozemního moře (Keskin a Arat 2011). Mobilita dospělých jedinců se zdá celkem limitující (Fadhlaoui-Zid et al. 2012) a nejpravděpodobnějším prostředkem rozptýlení na velké vzdálenosti je rozptyl paralarvův přes mořské proudy.

*Eledone moschata* je rozšířena po celém Středozemním moři. Její distribuce sahá i do Atlantského oceánu, kde je omezena jižním pobřežím Portugalska, Cádizským zálivem a západním pobřežím Maroka. (Guerra, 1992). *E. moschata* má pravděpodobně středomořský původ a odtud se rozptýlila do Atlantiku (Mangold a Boletzky, 1988). *E. moschata* se objevuje i v hlubinách Marmarského moře (Salman et al., 1993). Tato chobotnice se vyskytuje na písčinných dnech kontinentálního šelfu. Obvykle žije zahrabaná v sedimentu. Ve Středozemním moři se nachází hlavně od 15 do 200 metrů hloubkových s nejvyšší hustotou výskytu do 100 m, ačkoliv se v některých oblastech vyskytuje i hlouběji (Belcari a Sbrana, 1999).



Obr. 2: Výskyt *E. moschata* (www.aquamaps.org, 2016).



### 3. Cíle práce

- Zvládnutí základních metod molekulární biologie (exktrakce DNA, PCR, Sangerovo sekvenování).
- Příprava vzorků pro RAD seq (pro Illumina HiSeq)
- Bioinformatická analýza dat next-gen sekvenování.
- Vyhodnocení populační struktury na základě next-gen dat (STACKS, fineRADstructure).

## 4. Metody

### 4.1. Sběr vzorků

Bylo odebráno 24 vzorků tkáně chobotnic během několika terénních cest od roku 2015 do roku 2017 v Itálii (západní a východní pobřeží, Sardinie, Sicílie) a Chorvatsku (přehled lokalit Obr. 3 a Tab. I). Čerstvě zabité chobotnice byly zakoupeny na trhu případně přímo od rybářů a byly uchovány chlazené do zpracování. Následně byl odebrán kousek tkáně z chapadla (1 - 2 cm), který byl umístěn do mikrozkušavky s lihem (ethanol p.a.).



Obr. 3: Mapa Tyrhénského a Jaderského moře s vyznačenými lokalitami sběru vzorků a jejich četnost. Mapa (Wikimedia Commons contributors, 2017) byla upravena.

Tab. I: Lokality a GPS souřadnice sběru vzorků.

Vzorek	Země	Lokality	GPS souřadnice
OEIT 1 H	Sardínie	Oristano	39°53'15.1"N 8°31'38.9"E
OEIT 3 H	Sardínie	Oristano	39°53'15.1"N 8°31'38.9"E
OEIT 5 H	Sardínie	Oristano	39°53'15.1"N 8°31'38.9"E
OEIC 2 H	Sardínie	Cagliari	39°10'45.6"N 9°07'14.0"E
OEIC 5 H	Sardínie	Cagliari	39°10'45.6"N 9°07'14.0"E
OEIC 6 H	Sardínie	Cagliari	39°10'45.6"N 9°07'14.0"E
OECP 7 H	Chorvatsko	Pula	44°50'39.1"N 13°49'04.1"E
OECP 10 H	Chorvatsko	Pula	44°50'39.1"N 13°49'04.1"E
OEIS 1 H	Itálie	La Spezia	44°06'11.8"N 9°49'59.6"E
OEIS 4 H	Itálie	La Spezia	44°06'11.8"N 9°49'59.6"E
OEIL 6 H	Itálie	Livorno	43°31'42.9"N 10°17'47.8"E
OEIG 7 H	Itálie	Gaeta	41°12'10.8"N 13°35'03.7"E
OEIG 10 H	Itálie	Gaeta	41°12'10.8"N 13°35'03.7"E
OEIV 11 H	Itálie	Vieste	41°53'09.8"N 16°11'18.1"E
OEIV 13 H	Itálie	Vieste	41°53'09.8"N 16°11'18.1"E
OEIV 14 H	Itálie	Vieste	41°53'09.8"N 16°11'18.1"E
OEIR 17 H	Itálie	Rimini	44°04'52.8"N 12°35'27.7"E
OEIR 18 H	Itálie	Rimini	44°04'52.8"N 12°35'27.7"E
OEIM 1 H	Sicílie	Marsala	37°46'57.3"N 12°24'08.9"E
OEIM 2 H	Sicílie	Marsala	37°46'57.3"N 12°24'08.9"E
OEV 2 H	Chorvatsko	Veruda	44°49'43.2"N 13°50'17.0"E
OEV 3 H	Chorvatsko	Veruda	44°49'43.2"N 13°50'17.0"E
OeT3 H	Chorvatsko	Trogir	43°30'30.5"N 16°13'37.1"E
OeT4 H	Chorvatsko	Trogir	43°30'30.5"N 16°13'37.1"E

## 4.2. Extrakce DNA

Pro extrakci DNA byl použit malý kousek chapadla, který byl očištěn od kůže. K extrakci DNA byl použit kit DNeasy Blood & Tissue (Qiagen). Postupovalo se podle pokynů výrobce.

## 4.3. Jedno genová analýza

### 4.3.1. PCR

PCR amplifikace COI byla provedena se specifickými primery pro amplifikaci genu COI bezobratlých (sekvence a původ primerů v Tab. II). Reakční směs o objemu 20  $\mu$ l obsahovala 1  $\mu$ l templátu, 1  $\mu$ l forward primeru (5pM, F1490, Folmer et al., 1994; sekvence viz Tab. II), 1  $\mu$ l reverse primeru (5pM, H7005, Hafner et al 1994; sekvence viz Tab. II), 2  $\mu$ l pufru (PCR blue buffer, Top-bio), 0,5  $\mu$ l nukleotidů (dNTPs 10mM, ThermoFisher), 0,2  $\mu$ l Taq polymerázy (Top-bio) a 14,3  $\mu$ l H<sub>2</sub>O.

Tab. II: Přehled primerů.

Název	Sekvence primeru 5' → 3'	Forward/ Reverse	Autor
F1490	GGTCAACAAATCATAAAGATATTGG	F	Folmer et al. 1994
H7005	CCGGATCCACANCRTARTANGTRTCRTG	R	Hafner et al. 1994

Tab. III: Protokol PCR reakce.

Počáteční denaturace	Počet cyklů (Denaturace, Nasedání primerů, Syntéza)	Denaturace	Nasedání primerů	Syntéza	Finální syntéza
94°C / 5 minut	30	92°C/ 1 minuta	52°C/ 1 minuta	72°C/ 1 minuta	72°C/ 1 minuta

### 4.3.2. Gelová elektroforéza

Pro kontrolu úspěšnosti PCR reakce byla použita metoda separace pomocí gelové elektroforézy. Vzorky byly nanášeny na 1% agarózový gel. Gel byl připravován z 20 ml pufru TAE (Merck), 0,2 g agarózy (ThermoFisher) a 2  $\mu$ l GelRed (Biotium) fluorescentního barviva pro vizualizaci DNA. V elektroforetické vaničce naplněné 1% TAE bylo do jamek gelu nanášeno pomocí 6x Loading Dye (ThermoFisher) 2  $\mu$ l PCR produktu. Pro porovnání velikosti PCR produktu byl použit ladder 1 kb O'gene ruler (ThermoFisher). Elektroforéza běžela při 100 V přibližně 20 minut. Po dokončení elektroforézy byly fragmenty DNA rozdělené na gelu vizualizovány pomocí přístroje BIORAD GEL Dock EZ imager.

### 4.3.3. Přechištění

PCR produkty viditelné na gelu byly enzymaticky přechištěny metodou ExoSap-IT<sup>®</sup> (Dugan et al., 2002). Reakční směs pro enzymatické přechištění viz Tab. IV. Čištění probíhalo 15 minut při 37°C a 15 minut při 80°C. Přechištěné PCR produkty byly s primery použitými pro amplifikaci zaslány na sekvenaci do SeqMe nebo Eurofins (klasické Sangerovo sekvenování).

Tab. IV: Rozpis pro enzymatické přechištění vzorků.

Chemikálie	Množství
Exo I (BioLabs)	0,5 $\mu$ l
FastAp (ThermoFisher)	2 $\mu$ l
H <sub>2</sub> O	2,5 $\mu$ l

#### 4.3.4. Zpracování sekvencí

Výsledné sekvence byly upraveny v programu Geneious. Pomocí assemble - De Novo assembly byly výsledné chromatogramy složeny. Části se špatnou kvalitou byly odstraněny a následně byl uložen konsensus obou vláken. Byl vytvořen alignment k porovnání sekvencí. K alignmentu byly přidány referenční sekvence *Eledone moschata* (accession number KP693816), *Eledone cirrhosa* (KM517898) a *Octopus vulgaris* (AB158363). K vytvoření alignmentu byl použit algoritmus MAFFT (Katoh et al., 2002).

Pomocí programu IQ - TREE (<http://iqtree.cibiv.univie.ac.at/>, <http://www.iqtree.org/>, Trifinopoulos et al., 2016) byl vytvořen fylogenetický strom typu maximum likelihood. Nastavení bylo nastavené na výchozí hodnoty: ML, model auto selection (vybraný model TIM2+F+I; Kalyaanamoorthy et al., 2017), podpory větví jsou uvedené jako aLRT/UFBoot (Hoang et al., 2018). Strom byl vizualizován pomocí programu phylo.io (<http://phylo.io/>). Poté byl analyzován dataset pro populační síť (jen *E. moschata*, minimalizovaná chybějící data, 16 vzorků, alignment délky 770bp) v software PopArt (<http://popart.otago.ac.nz/index.shtml>, Bandelt et al.; 1999, Leigh a Bryan, 2015) algoritmem pro median joining network.

## 4.4. RAD seq

### 4.4.1. Příprava knihoven ddRAD

Pro přípravu knihoven na sekvenování byl použit protokol double-digest RAD seq (ddRAD seq). Bylo vybráno 24 vzorků chobotnic z několika lokalit Středozemního moře (Tab. I). Koncentrace DNA každého vzorku byla měřena pomocí Qubit 2.0 Broad Range. Příprava knihovny ddRAD se řídila mírně upraveným protokolem Peterson et al. (2012). Postup pro přípravu knihoven pro Illuminu metodou ddRAD byl následující:

- Nejdříve byla provedena restrikce. Bylo použito 300 ng genomické DNA z každého jedince. Následně byly přidány dva restrikční enzymy SphI + MluCI. Reakční směs o objemu 30  $\mu$ l obsahovala 3  $\mu$ l CutSmart<sup>(TM)</sup> buffer, SphI = 0,7  $\mu$ l, MluCI = 2  $\mu$ l, a individuální množství H<sub>2</sub>O. Poté byly vzorky uloženy do termobloku a inkubovány při 37°C po dobu 3h.
- Produkt štěpení byl purifikován pomocí magnetických kuliček AMPure XP beads. Ke každému vzorku o objemu 30  $\mu$ l bylo přidáno 45  $\mu$ l (1,5x objem vzorku) roztoku s kuličkami. Vzorky se nechaly odstát 15 minut, aby se navázaly fragmenty na kuličky. Zkumavky byly umístěny do magnetického separátoru, dokud nebyl supernatant čistý. Supernatant byl odstraněn aniž by se zkumavka odstranila ze stojánku. Vzorky byly dvakrát přečištěny 80% etanolem o objemu 200  $\mu$ l. Následně bylo přidáno 40  $\mu$ l elučního TE pufru, aby se pelet rozptýlil. Po uplynutí 1 minuty se zkumavky umístily znovu do magnetických stojanů a supernatant byl přenesen do čistých zkumavek.
- Následujícím krokem byla kvantifikace. Koncentrace byla změřena na přístroji Qubit 2.0 Broad Range. Postupovalo se podle pokynů výrobce.
- Poté následovala ligace. Na tuto reakci bylo použito 40  $\mu$ l vzorku. K DNA bylo přidáno 100 ng restrikčního produktu. Poté byly přidány dva adaptéry (Tab. V; Peterson et al., 2012) MID P1 a P2 o objemu 0,1  $\mu$ l a barcodes (Tab. VI), které byly individuální pro každý vzorek, o objemu 1,6  $\mu$ l (0,5 pmol/ $\mu$ l). Pak 0,1  $\mu$ l T4 DNA ligázy, 4  $\mu$ l 10x T4 pufru a individuálně byla doplněna voda do 40  $\mu$ l. Ligační reakce byla prováděna v termální cyklovači Eppendorf (ligace 23 ° C / 30 min, deaktivace ligázy 65 ° C / 10 min, pomalé chlazení 41x [64 ° C, -1 ° C / cyklus, 45s]).

- Dalším krokem byl poolplex. Všech 24 vzorků, lišících se v individuálním barcode adaptéru, bylo smícháno a přečištěno s AMPure UP beads (s poměrem 1,5). Finální eluce byla do 1x TE pufru o objemu 30  $\mu$ l.
- Následovala selekce podle délky fragmentů (fraction selection). Selektce probíhala na přístroji PippinPrep (Sage Science). Postupovalo se podle pokynů výrobce.
- Poté proběhla PCR amplifikace s primery obsahujícími indexy pro multiplex k použití na jedné drážce Illumina. Produkty z PippinPrep separace byly rozděleny do 3 reakcí, aby se zabránilo chybě PCR. Byl použit PCR protokol QIAGEN Multiplex PCR kit. Reakční směs obsahovala: 2x Multiplex PCR Master Mix 8  $\mu$ l, 1  $\mu$ l forward a reverse primeru, 2  $\mu$ l DNA a 4  $\mu$ l H<sub>2</sub>O. Produkty PCR amplifikace byly vizualizovány pomocí gelové elektroforézy a BIORAD GEL Dock EZ imageru (na 2% agarózovém gelu).
- Všechny PCR reakce byly smíchány zpět do jedné zkumavky a následně přečištěny s AMPure XP beads a kvantifikovány na Qubit 2.0 kit DNA Broad range.

Tab. V: Sekvence adaptéru použitých pro Illuminu.

Adaptér	Sekvence
P1	TCACTCTTTCCCTACACGACGCTCTTCCGATCT
P2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

#### 4.4.2. Illumina

Vzorky byly sekvenovány na přístroji Illumina HiSeq 2000/2500 (125 bp, paired-end, chemistry v4) na jedné lane v institutu Embl, Gencore, Heidelberg, Německo.



Tab. VI: Přehled sekvencí individuálních barcodes použitých pro rozlišení jednotlivých vzorků.

Barcodes	jedinci
CTGCG	OEIT 1 H
CTGTC	OEIT 3 H
CTTGG	OEIT 5 H
GACAC	OEIC 2 H
GAGAT	OEIC 5 H
GAGTC	OEIC 6 H
GCCGT	OCEP 7 H
GCTGA	OECP 10 H
GGATA	OEIS 1 H
GGCCA	OEIS 4 H
GGCTC	OEIL 6 H
GTAGT	OEIG 7 H
GTCCG	OEIG 10 H
GTCGA	OEIV 11 H
TACCG	OEIV 13 H
TACGT	OEIV 14 H
TAGTA	OEIR 17 H
TATAC	OEIR 18 H
TCACG	OEIM 1 H
TCAGT	OEIM 2 H
TCCGG	OEV 2 H
TCTGC	OEV 3 H
TGGAA	OeT3 H
TTACC	OeT4 H

## 4.4.3. Zpracování dat

### 4.4.3.1. Stacks

#### Čištění dat

Osekvenované knihovny z Illuminy byly zpracovány pomocí softwaru Stacks (Catchen et al. 2011). Tento program byl vytvořený za účelem zpracování dat, které byly vyprodukovány enzymatickými restrikčními metodami (RAD seq). Z těchto dat se následně mohou vytvářet genetické mapy nebo genomické a fylogenetické analýzy. Postupovalo se podle manuálu (<http://catchenlab.life.illinois.edu/stacks/manual/>).

- Nejdříve byla data přečištěna pomocí programu *process\_radtags*. Pomocí tohoto kroku byly odfiltrovány adaptéry od našich originálních vzorků, vyřazeny sekvence s nízkou kvalitou a rozříděny dle barcode na jedince. Jako vstupní soubor byly použity raw data (reads) z Illuminy.

- Poté byl spuštěn skript:

```
process_radtags -T 16 --paired -p ./raw_zip -b barcodes_Ele01 --inline_null -o
./dada/output_PR01/ -c -q -r --renz_1 sphI --renz_2 mluCI --adapter_1
TACTACTCTTTCCCTACACGACGCTCTTCCGATCT --adapter_2
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT --adapter_mm 0 --
filter_illumina,
```

kde možnost - T znamená, kolik jader bude tento program využívat,

--paired udává paired end nastavení,

-p specifikuje cestu k raw datum,

-b určuje cestu k textovému souboru s barcody (barcody v jednom sloupci, v druhém sloupci odpovídající kódy jedinců),

-inline\_null určuje podobu barcodů (toto nastavení bylo vybráno, jelikož barcode byl vložený v sekvenci),

-o determinuje cestu, kam mají být výsledné files ukládány,

-c čistí data a odstraňuje každý read, který má nejasnou bázi,

-q odstraňuje ready s nízkou kvalitou,

-r vytváří výsledný soubor s barcody a RAD seq.

Dále byly specifikovány restrikční enzymy --renz\_1 sphI --renz\_2 mluCI,

následně byly určeny sekvence a možnosti adaptérů:

```
--adapter_1 TACTACTCTTTCCCTACACGACGCTCTTCCGATCT
```

--adapter\_2 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT,  
--adapter\_mm, který určovalo počet mismatches povolených v sekvenci adaptéru.  
Nakonec bylo zadáno nastavení --filter\_illumina, které filtruje reads, které filtry  
Illuminy označily za chybné.

- Výstup programu *process\_radtags* obsahoval 4 soubory pro každý vzorek (jedince), *sample\_X.1.fq*, *sample\_X.2.fq*, *sample\_X.rem.1.fq*, *sample\_X.rem.2.fq*, kde X nahrazuje zkratku vzorku. V souborech *Y.fq* byla přečištěná data obou paired-end konců a v souborech *rem.Y.fq* jsou vyřazená data těchto paired-end vzorků kvůli nízké kvalitě. Y nahrazuje číslo, podle kterého je označené jedno z paired-end čtení.

### ***Denovo\_map.pl***

Následujícím krokem bylo zpracování genomových dat (vybraných RAD seq). Jelikož není k dispozici referenční genom, byl použit program *denovo\_map.pl*, který pro správné fungování potřebuje populační mapu. Populační mapy zajišťují informaci o zařazení vzorků do přírodních populací. Tento program využívá všechny komponenty Stacks a vytvoří linkage mapu anebo identifikuje alely v setu populací.

- Nejprve byly vytvořeny populační mapy v podobě textového souboru, kde v jednom sloupci byly sepsány kódy jedinců, a v druhém sloupci odděleným tabulátorem byli tito jedinci rozděleni do populací. Byla provedena 3 různé rozdělení populací (Tab. VII).
  - V rozdělení „**sea**” označení značí:  
POP1 populaci *Eledone cirrhosa* v Tyrhénském moři,  
POP2 *E. moschata* v Tyrhénském moři,  
POP3 *E. moschata* v Jaderském moři,  
POP4 *Octopus vulgaris*, v Tyrhénském moři,  
POP5 *Eledone cirrhosa* v Jaderském moři.
  - V jemnějším rozdělení populací „**area**” bylo rozdělení více specifikované:  
POP1 *Eledone cirrhosa* v Tyrhénském moři v oblasti Sardinie,  
POP2 *E. moschata*, Tyrhénské moře  
POP3 *E. moschata*, Jaderské moře, Chorvatsko,  
POP4 *Eledone cirrhosa*, západní pobřeží Itálie,  
POP5 *E. moschata*, Tyrhénské moře západní pobřeží Itálie,  
POP6 *E. moschata*, Jaderské moře, východní pobřeží Itálie,

POP7 *Octopus vulgaris*, Sicilie,

POP8 *Eledone cirrhosa*, Jaderské moře, Chorvatsko.

- Posledním nejjemnějším rozdělením bylo rozdělení „locality“:

POP1 *Eledone cirrhosa*, Tyrhénské moře, Sardínie,

POP2 *E. moschata*, Tyrhénské moře, Sardínie,

POP3 *E. moschata*, Jaderské moře, Istrie,

POP4 *Eledone cirrhosa*, Tyrhénské moře, LaSpezia,

POP5 *Eledone moachata*, Tyrhénské moře, La Spezia

POP6 *E. moschata*, Tyrhénské moře, Livorno,

POP7 *E. moschata*, Tyrhénské moře, Gaeta,

POP8 *E. moschata*, Jaderské moře, Vieste,

POP9 *E. Moschata*, Jaderské moře, Rimini,

POP10 *Octopus vulgaris*, Sicily,

POP11 *Eledone cirrhosa*, Jaderské moře, Trogir.

- Vstupní soubor pro program *denovo\_map.pl* byl výstupním souborem programu *process\_radtags*.
- Spouštěcí script byl následující:  

```
denovo_map.pl -T 8 -m 3 -M 4 -n 4 -o ./dada/denovo/ --samples ./dada/output_PR01/  
--popmap ./dada/popmaps/popmap --paired.
```
- Jelikož data byla paired-end, užilo se nastavení `--paired`,  
dalším nastavením bylo `-T`, které určuje, s kolika jádry bude program pracovat,  
`-m` hloubka prosekvenování lokusů,  
`-M` stanovuje počet neshod (mismatches) povolených mezi Stacks v rámci individuí  
u `ustacks` a `-n` u `cstacks`,  
`-o` určuje, kam se výstupní data budou ukládat,  
`--samples` vymezuje cestu, kde jsou uloženy výstupní data z programu  
*process\_radtags*.  
`--popmap` stanovuje cestu, kde je možné najít text file s rozdělením jedinců  
do populací.
- Tímto krokem byly vybudovány soubory shodných krátkých sekvencí, z nichž byly  
následně sestaveny lokusy. Výstupním souborem byly katalogy SNP.
- Jelikož se zjistilo, že ne všechny vzorky byly skutečně druhu *E. moschata*, celé  
Stacks se projelo znova jen se vzorky *E. moschata*.

- Populační mapy byly rozděleny také podle 3 úrovní rozdělení: hrubé rozdělení „**sea**”, střední rozdělení „**area**” a jemné rozdělení „**locality**” (Tab. VII ).
- V rozdělení „**sea**” označení POP1 značí populaci *E. moschata* v Tyrhénském moři, POP2 *E. moschata* v Jaderském moři
- V jemnějším rozdělení populací „**area**” bylo rozdělení více specifikované:
  - POP1 *Eledone moschata*, Tyrhénské moře
  - POP2 *E. moschata*, Jaderské moře, Chorvatsko,
  - POP3 *E. moschata*, Tyrhénské moře západní pobřeží Itálie,
  - POP4 *E. moschata*, Jaderské moře, východní pobřeží Itálie.
- Posledním nejjemnějším rozdělením bylo rozdělení „**locality**”:
  - POP1 *Eledone moschata*, Tyrhénské moře, Sardínie,
  - POP2 *E. moschata*, Jaderské moře, Istrie,
  - POP3 *E. moschata*, Tyrhénské moře, Livorno,
  - POP4 *E. moschata*, Tyrhénské moře, Gaeta,
  - POP5 *E. moschata*, Jaderské moře, Vieste,
  - POP6 *E. moschata*, Jaderské moře, Rimini.

## **Populations**

Program populations vypočítává populační statistiky jako například  $\pi$ ,  $F_{IS}$  a  $F_{ST}$ . Program vypsal jednotlivé nukleotidové polymorfismy (site level SNP calls) ve formátu variant call format (VCF) a také zpracoval SNPs pro analýzu v programu STRUCTURE nebo ve Phylip formátu pro fylogenetickou analýzu.

- Spouštěcí skript vypadal takto:
 

```
populations -P denovo_output/denovo_area/abc/ -M popmaps/popmap_area_abc --
genepop --fstats --hwe --smooth --structure --write_single_snp -t 16 -vcf.
```
- Nastavení -P určuje cestu k výstupním datům z programu *denovo\_map.pl*,
  - M specifikuje cestu k populačním mapám,
  - t určuje počet jader,
  - genepop poskytuje data pro program GenePop,
  - structure pro program Structure,
  - fstats porovnává populace mezi sebou a počítá  $F_{ST}$  statistiku,
  - hwe vypočítává odlišnost od Hardy-Weinbergovy rovnováhy na každý lokus,
  - smooth umožňuje kernel-smoothed  $\pi$ ,  $F_{IS}$ ,  $F_{ST}$ ,  $F_{ST}$ , a  $\Phi_{ST}$  statistiky,

--write\_single\_snp omezuje analýzu dat pouze na první SNP na lokus,  
 --vcf umožňuje výstup SNPs ve Variant Call Format (VCF).

Tab. VII: Rozdělení všech jedinců do populací.

Jedinec	sea	area	locality
OEIT 1 H	POP1	POP1	POP1
OEIT 3 H	POP1	POP1	POP1
OEIT 5 H	POP1	POP1	POP1
OEIC 2 H	POP2	POP2	POP2
OEIC 5 H	POP2	POP2	POP2
OEIC 6 H	POP2	POP2	POP2
OECF 7 H	POP3	POP3	POP3
OECF10 H	POP3	POP3	POP3
OEIS 1 H	POP1	POP4	POP4
OEIS 4 H	POP2	POP5	POP5
OEIL 6 H	POP2	POP5	POP6
OEIG 7 H	POP2	POP5	POP7
OEIG 10 H	POP2	POP5	POP7
OEIV 11 H	POP3	POP6	POP8
OEIV 13 H	POP3	POP6	POP8
OEIV 14 H	POP3	POP6	POP8
OEIR 17 H	POP3	POP6	POP9
OEIR 18 H	POP3	POP6	POP9
OEIM 1 H	POP4	POP7	POP10
OEIM 2 H	POP4	POP7	POP10
OEV 2 H	POP3	POP3	POP3
OEV 3 H	POP3	POP3	POP3
OeT3 H	POP5	POP8	POP11
OeT4 H	POP5	POP8	POP11

Tab. VII: Rozdělení jedinců *Eledone moschata* do populací.

Jedinec	sea	area	locality
OEIC 2 H	POP1	POP1	POP1
OEIC 5 H	POP1	POP1	POP1
OEIC 6 H	POP1	POP1	POP1
OECP 7 H	POP2	POP2	POP2
OECP 10 H	POP2	POP2	POP2
OEIS 4 H	POP1	POP3	POP3
OEIL 6 H	POP1	POP3	POP3
OEIG 7 H	POP1	POP3	POP4
OEIG 10 H	POP1	POP3	POP4
OEIV 11 H	POP2	POP4	POP5
OEIV 13 H	POP2	POP4	POP5
OEIV 14 H	POP2	POP4	POP5
OEIR 17 H	POP2	POP4	POP6
OEIR 18 H	POP2	POP4	POP6
OEV 2 H	POP2	POP2	POP2
OEV 3 H	POP2	POP2	POP2

### 4.4.3.2. Finerad structure

Pomocí programu fineRADstructure (Malinsky et al., 2018) lze zkoumat populační strukturu. Celogenomová data získaná RAD sekvenováním jsou zpracovány metodou coancestry matrix. Tato metoda měří společný původ všech jedinců a hledá nejbližší haplotyp pomocí MCMC (Markov chain Monte Carlo) klastrovacího algoritmu. Analyzováni byli jen jedinci *E. moschata*. Výstupní data byla zpracována v prostředí R (R Core Team 2013), v programu RStudio (RStudio Team, 2015) pomocí upravených skriptů podle Malinský et al. (2018).

- Nejdříve byla vypočítána co-ancestry matrix pomocí skriptu:  
`/home/vacatko/Programs/fineRADstructure/RADpainter paint  
EM_S_populations.haplotypes.tsv.min2.finerad,`  
kde byl uveden jako vstupní soubor soubor z programu *populations*  
`populations.haplotypes.tsv.`
- Následně byly jedinci přiděleny do populací skriptem:  
`/home/vacatko/Programs/fineRADstructure/finestructure -x 100000 -y 100000 -z  
1000 EM_S_populations.haplotypes.tsv.min2_chunks.out  
EM_S_populations.haplotypes.tsv.min2_chunks.mcmc.xml`
  - nastavení `-x` určuje počet počátečních kroků, které nejsou v analýze zahrnuty (burn-in) v interakcích pro metodu MCMC
  - y počet iterací vzorku pro metodu MCMC (výchozí: 1000),
  - z úzký interval ve výstupním souboru pro metodu MCMC.
- Pak byl vytvořen příbuzenský strom díky skriptu:  
`/home/vacatko/Programs/fineRADstructure/finestructure -m T -x 100000  
EM_S_populations.haplotypes.tsv.min2_chunks.out  
EM_S_populations.haplotypes.tsv.min2_chunks.mcmc.xml  
EM_S_populations.haplotypes.tsv.min2_chunks.mcmcTree.xml.`
- Výsledné dokumenty byly zpracovány v R (R Core Team 2013).

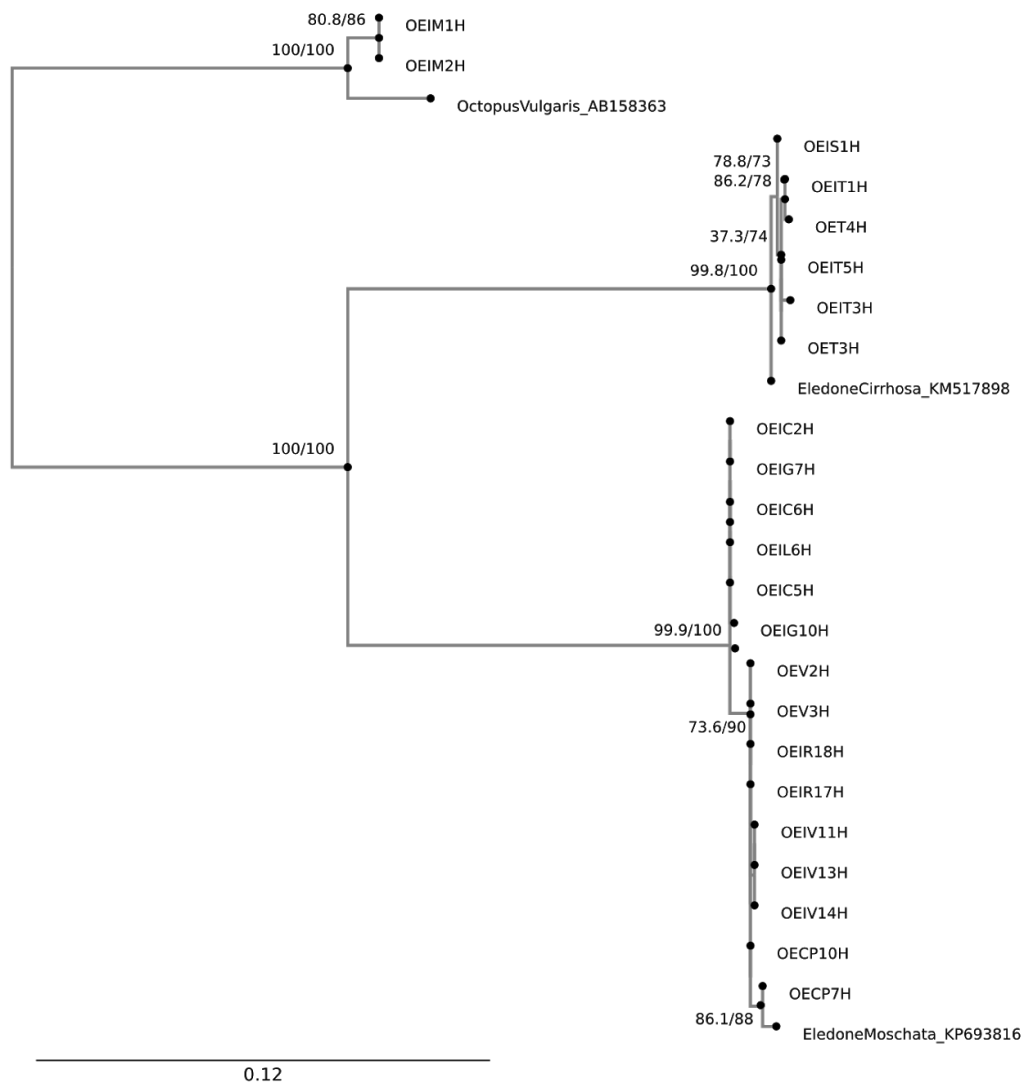


## 5. Výsledky

### 5.1. Jednogenová analýza

#### 5.1.1. Fylogenetický strom

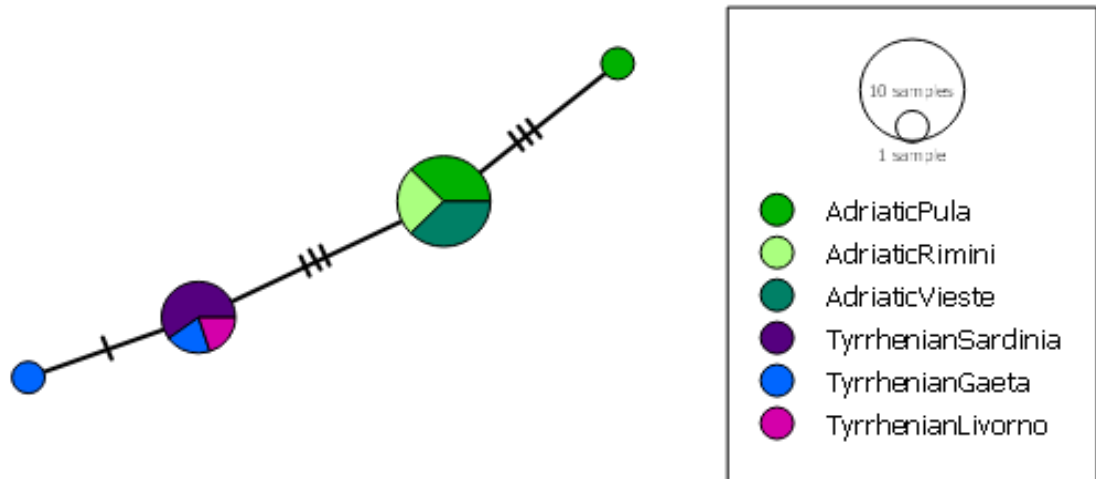
Z fylogenetického stromu (24 sekvencí a 3 referenční sekvence; Obr. 4) lze vyčíst, že některé vzorky původně označené, jako *Eledone moschata* mají blíže k *Eledone cirrhosa* (vzorky OEIS 1 H, OEIT 1 H, OEIT 3 H, OEIT 4 H, OEIT 5 H, OET 3 H) a *Octopus vulgaris* (vzorky OEIM 1 H, OEIM 2 H), což ukazuje, že během sběru vzorků došlo k záměně druhů. Tyto vzorky byly dále v populačních analýzách buď vynechány, nebo označeny jako samostatné populace. Je také vidět, že vzorky *Eledone moschata* vytváří dvě skupiny a dělí se na populaci v Tyrhénském a Jaderském moři, nicméně tyto skupiny nejsou významně oddělené. Uvnitř těchto podskupin na fylogenetickém stromu není žádná další struktura patrná.



Obr. 4: Fylogenetický strom genu COI. Pro sestavení tohoto stromu bylo použito 24 našich sekvencí a 3 sekvencí referenčních: *Eledone moschata* (accession number KP693816), *Eledone cirrhosa* (KM517898) and *Octopus vulgaris* (AB158363). Alignment byl dlouhý 1035 bp.

### 5.1.2. Populační síť

Byla vytvořena populační síť pro zkoumání vztahů mezi skupinami haplotypů v souvislosti s jejich původem (obr. 5).



Obr. 5: Síť haplotypů COI *Eledone moschata*. Bylo využito 16 sekvencí pro vytvoření populační sítě. Odstíny modré a růžové představují jedince v Jaderském moři, odstíny zelené jedince v Tyrhénském moři. Kolmé čárky na propojovacích přímkách mezi haplotypy znázorňují počet bodových mutací mezi nimi.

Z populační sítě (Obr. 5) lze vyčíst, že vzorky z Tyrhénského a Jaderského moře mají jasně oddělené populace 3 konstantními bodovými mutacemi (SNP), uvnitř těchto skupin je genetická struktura homogenní (v obou populacích pouze jediný vzorek s jiným, než centrálním haplotypem).

## 5.2. Celogenomová analýza

### 5.2.1. Stacks

#### *process\_radtags*

*Process\_radtags* odfiltroval z důvodu špatné kvality 2,9 % readu (viz tab IX).

Tab. IX: Filtrovaná data.

Celkový počet sekvencí	48775942
Neúspěšně filtrované ready	0
Ready obsahující sekvence adaptérů	66481
Nenalezené barcody	580336
Nízká kvalita	338850
Nejednoznačný RAD seq	443104
Ponechané ready	47347171

#### *denovo\_map.pl*

Po spuštění programu *denovo\_map.pl* se vytvořily katalogy SNP. Tento program poskytoval informaci i genomové coverage (viz přílohy Tab. IX). Rozpětí coverage bylo od 5,63x do 20,07x. Po této analýze zůstalo 323784 lokusů s 146074 variabilními místy.

Následně byl program spuštěn jen pro vzorky *Eledone moschata*. Genomová coverage byla stejná. Po této analýze zůstalo 137421 lokusů s 73989 variabilními místy.

#### *populations*

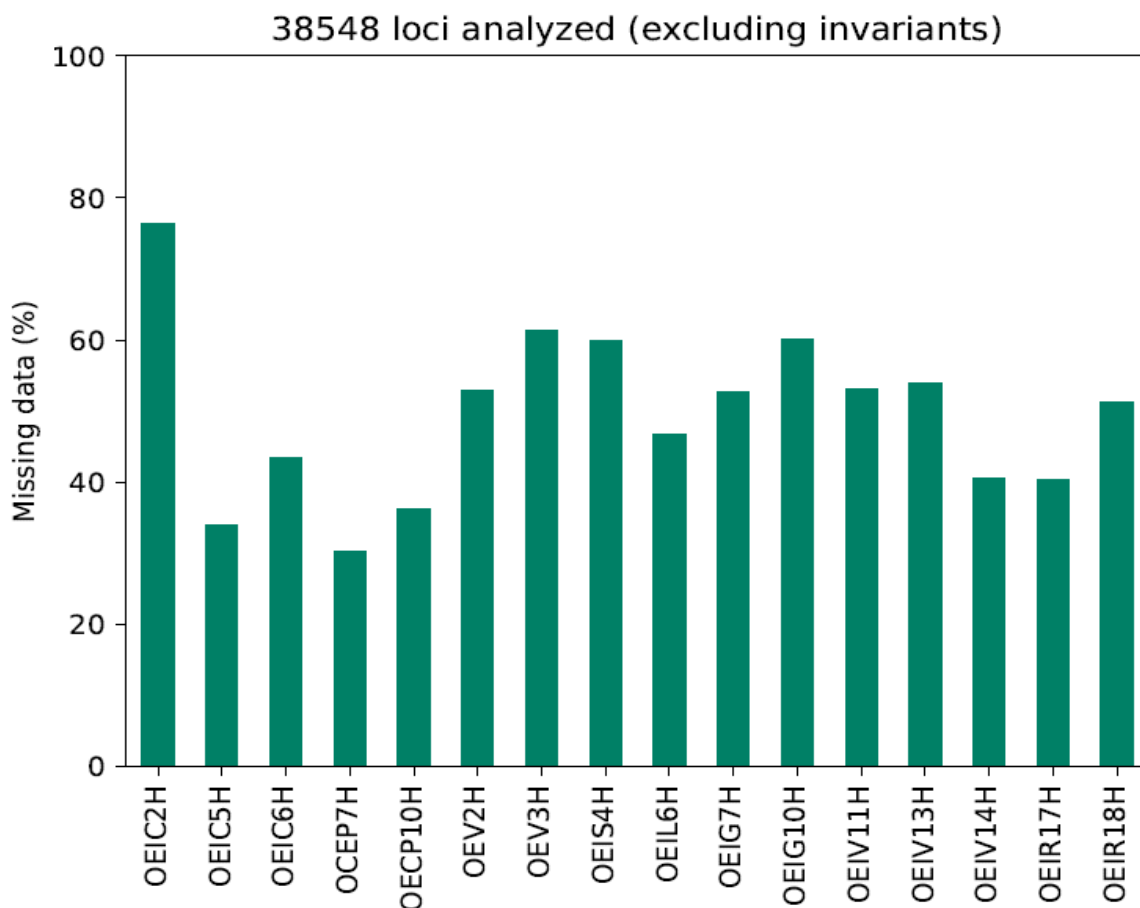
Tímto krokem byly vypočteny základní statistiky populační genetiky. Zaměřili jsme se na  $\Phi_{ST}$ , která vyjadřuje genetické vzdálenosti mezi populacemi. Statistiky (Tab. X) byly vypočítány pro dataset obsahující njen *E. moschata*. Po této analýze zůstalo 323784 lokusů s 75866 variabilními místy.

Tab. X: Genetická diverzita populací ( $\Phi_{ST}$  průměry) pro všechny vzorky s rozdělením populací „area”: POP1 *Eledone cirrhosa*, Tyrhénské moře, Sardínie, POP2 *Eledone moschata*, Tyrhénské moře, Sardínie, POP3 *Eledone moschata*, Jaderské moře, Chorvatsko, POP4 *Eledone cirrhosa*, Tyrhénské moře, západní pobřeží Itálie, POP5 *Eledone moschata*, Tyrhénské moře, západní pobřeží Itálie, POP6, *Eledone moschata*, Jaderské moře, východní pobřeží Itálie, POP7 *Octopus vulgaris*, Sicílie, POP8 *Eledone cirrhosa*, Jaderské moře, Chorvatsko.

	POP1	POP2	POP3	POP4	POP5	POP6	POP7	POP8
POP1		0,73737	0,84957	0,21604	0,61049	0,40927	1	0,14951
POP2			0,29568	0,89372	0,08547	0,26077	0,79167	0,8246
POP3				0,8212	0,24413	0,03677	0,01597	0,83022
POP4					0,90158	0,84605	0,89543	0,03492
POP5						0,21656	0,66667	0,84019
POP6							0,23768	0,80686
POP7								0,72222

Z tabulky (Tab. X) je patrné, že se většina nastavených populací od sebe významně liší (hodnoty  $\Phi_{ST}$  větší než 0,3). Nicméně některé hodnoty jsou však obtížně biologicky vysvětlitelné (POP7 - *Octopus vulgaris* v porovnání s *Eledone moschata* POP3 kdy  $\Phi_{ST} = 0,016$  a POP6 kdy  $\Phi_{ST} = 0,238$ ).

Statistiky se dále počítaly jen pro vzorky *E. moschata*. Řešila se tři různá rozdělení do populací: rozdělení hrubé „sea”, střední „area” a nejjemnější „locality”. Po této analýze zůstalo 137421 lokusů s 38548 variabilními místy. Tento program vygeneroval také graf, který ukazoval chybějící data každého vzorku (Obr. 6). Dále byly vypočteny  $\Phi_{ST}$  statistiky pro tento dataset (Tab. XI, Tab. XII, Tab. XIII).



Obr. 6: Chybějící data pro každého jedince.

Z obrázku lze vyčíst, že chybějící data jsou rozložena nerovnoměrně a jejich rozpětí je od 30% až skoro do 80 %.

V hrubém rozlišení „sea” (Tab. XI) byla vypočítaná genetická vzdálenost mezi populacemi v Tyrhenském a Jaderském moři.

Tab. XI: Genetická diverzita ( $\Phi_{ST}$  průměry) populací *E. moschata* s rozdělením populací „sea”: POP1 *E. moschata* v Tyrhenském moři, POP2 *E. moschata* v Jaderském moři.

	POP1	POP2
POP1		0,233559

Z výsledků  $\Phi_{ST}$  statistiky (Tab. XI) vyplývá, že populace *E. moschata* z Jaderského a Tyrhenského moře jsou divergované ( $\Phi_{ST} = 0,234$ ). Poměrně vysoké hodnoty značí větší genetickou vzdálenost mezi populacemi.

Jemnějším rozdělením „area” (Tab. XII) byla zkoumána jemnější genetická struktura, kdy se jedinci rozdělili do vícero populací: pobřeží Chorvatska, Sardinie a dále východní a západní pobřeží Itálie.

Tab. XII: Genetická diverzita ( $\Phi_{ST}$  průměry) populací *E. moschata* s rozdělením populací „area”. Rozdělení populací : POP1 *Eledone moschata*, Tyrhénské moře, Sardinie POP2 *E. moschata*, Jaderské moře, Chorvatsko, POP3 *E. moschata*, Tyrhénské moře západní pobřeží Itálie, POP4 *E. moschata*, Jaderské moře, východní pobřeží Itálie.

	POP1	POP2	POP3	POP4
POP1		0,299063	0,0872585	0,264215
POP2			0,247521	0,0404876
POP3				0,220334

Toto rozdělení potvrzuje hypotézu odlišnosti populací v Jaderském a Tyrhénském moři. Díky nízkým hodnotám mezi populacemi č. 1 a 3 ( $\Phi_{ST} = 0,087$ ) i 2 a 4 ( $\Phi_{ST} = 0,041$ ) je zjevné, že mají mezi sebou malou genetickou vzdálenost a lze tyto populace označit za stejné. Geneticky nejvzdálenější jsou populace od pobřeží Sardinie a Chorvatska ( $\Phi_{ST} = 0,299$ ).

V nejemnějším rozdělení

„locality” (Tab. XII) byla zkoumána jemnější genetická vzdálenost, kdy se jedinci rozdělili populací podle lokalit, kde byli sebráni.

Tab. XIII: Genetická diverzita populací *E. moschata* ( $\Phi_{ST}$  průměry) s rozdělením populací „locality”: POP1 *Eledone moschata*, Tyrhénské moře, Sardínie, POP2 *E. moschata*, Jaderské moře, Istrie, POP3 *E. moschata*, Tyrhénské moře, Livorno, POP4 *E. moschata*, Tyrhénské moře, Gaeta, POP5 *E. moschata*, Jaderské moře, Vieste, POP6 *E. moschata*, Jaderské moře, Rimini.

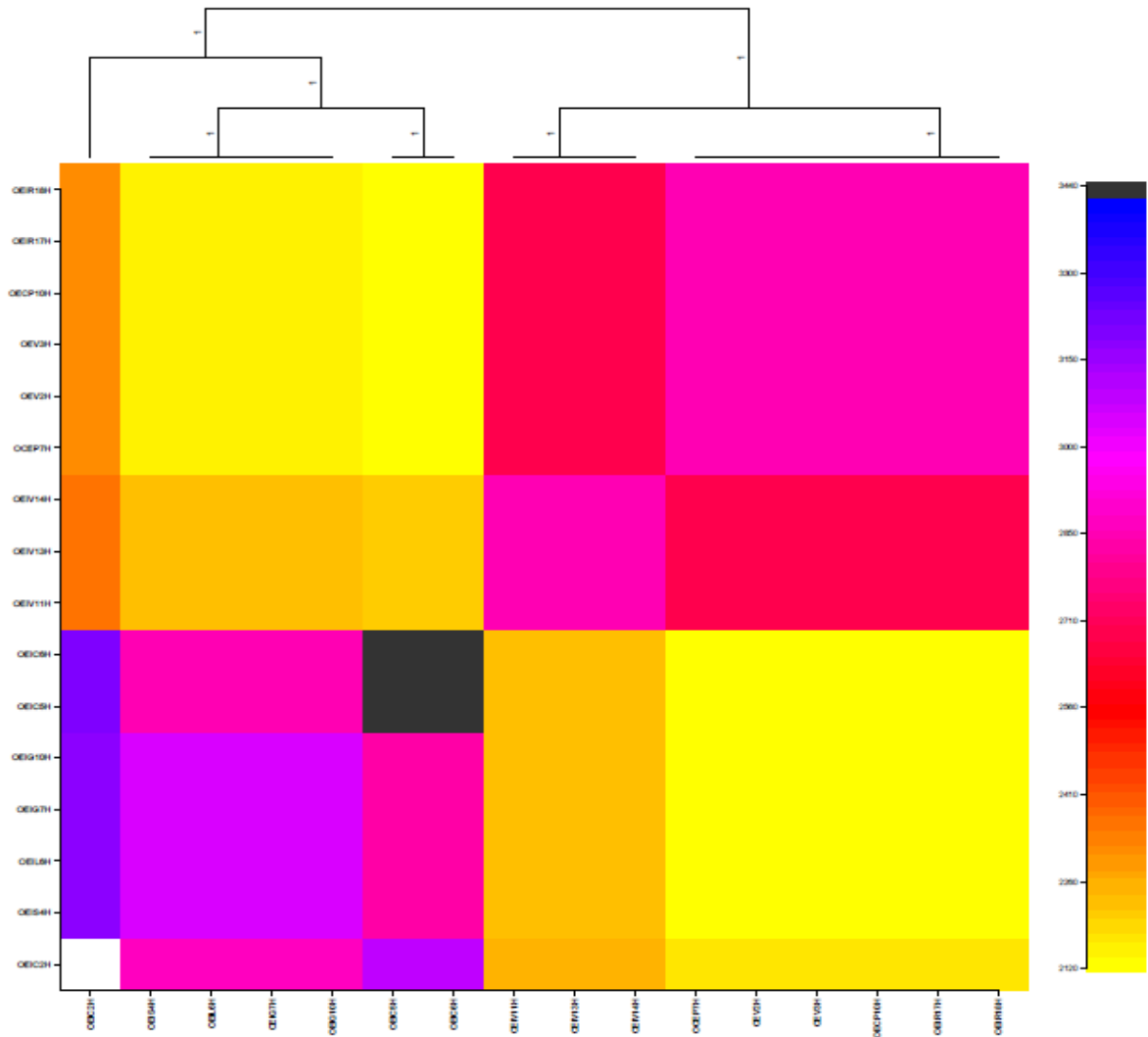
	POP1	POP2	POP3	POP4	POP5	POP6
POP1		0,297683	0,0897481	0,111533	0,268978	0,296062
POP2			0,262825	0,249078	0,0741162	0,00496087
POP3				0,00597018	0,225274	0,246277
POP4					0,214451	0,232382
POP5						0,058147

Nejjemnější rozdělení ukazuje, že i v populaci z Tyrhénského moře jsou si jedinci lehce geneticky vzdálení, jak tomu je mezi populacemi z oblasti Sardínie a Gaeta ( $\Phi_{ST} = 0,112$ ).



## 5.2.2. FineRAD structure

Metodou FineRAD structure bylo zpracováno 37509 lokusů pro 16 vzorků. Tento program vygeneroval co-ancestry matici a také strom. Výsledkem byl graf (Obr. 7), který ukazuje varianci uvnitř populací. Z grafu lze vyčíst jasná odlišnost vzorků v Tyrhénském a Jaderském moři. Z grafu je zároveň patrné jemnější strukturování uvnitř populací v každém moři.



Obr. 7: Populační struktura jedinců *E. moschata*. Barevná stupnice vyznačuje rozsah sdílení lokusů mezi jednotlivci.

## 6. Diskuze

Fylogenetická analýza probíhala na 24 vzorcích. Sekvence mitochondriálního genu COI byly porovnány s referenční sekvencí *Eledone moschata*, *Eledone cirrhosa* a *Octopus vulgaris*. Jen 16 vzorků se skutečně zařadilo nejbližší referenční sekvenci *Eledone moschata*, dalších 6 vzorků mělo blíže k referenční sekvenci *Eledone cirrhosa* a dokonce dva vzorky byly zařazeny k referenční sekvenci *Octopus vulgaris*. Vzorky později označené jako *Octopus vulgaris* byly sebrány kolegou, který nebyl obeznámen s morfologií studovaného druhu. Pro další analýzy by bylo tedy vhodnější provádět sběr vzorků pod dohledem zkušeného pracovníka anebo poskytnout kolegům přesnější informace. Nicméně i druhy *Eledone moschata* a *Eledone cirrhosa* jsou si morfologicky velmi podobné a je snadné si je zaměnit mezi sebou, obzvláště když se jedná o mladší jedince do cca 15 cm. V případě, že si nejsme jistí správností určení druhu, je vhodné si druh ověřit pomocí amplifikace genu COI univerzálními primery pro bezobratlé. Následnou sekvenací, porovnáním sekvencí s databází GenBank a sestrojením fylogenetického stromu nebo populační struktury lze v tomto případě druh bezpečně určit, jak je vidět i v případě této práce (viz Obr. 4 a Obr. 5).

Populační struktura hlavonožců je velice málo prozkoumaná. Ve studované oblasti Středozemního moře byla zkoumána populační struktura *Octopus vulgaris* ve vícero studiích s různými genetickými markery. Maltagliati et al. (2002) zkoumali genetickou strukturu s použitím alozymů a porovnávali populace Sicílie a Kréty, které byly geneticky vzdálené od západní části Středozemního moře s nejasnou strukturou mezi ostatními populacemi v severozápadní oblasti Středozemního moře. Casu et al. (2002) pak s částí této populace pracovali a zkoumali strukturu populace pomocí jednoho mikrosatelitového lokusu a zjistili významnou populační strukturu v rámci celého Středozemního moře. Práce Keskina a Atara (2011) zkoumala genetickou strukturu *O. vulgaris* podél tureckého pobřeží s použitím mitochondriálních markerů a našla dva klastry kompatibilní s geografickou vzdáleností, jeden na východním a druhý na jižním pobřeží země. Pomocí mitochondriálního markeru cytochromoxidázy III studovali Fadhlaoui-Zid et al. (2012) genetickou strukturu *O. vulgaris* v centrálním Středozemním moři v oblasti mezi Sicílií a Tuniskem. Byla zjištěna významná genetická struktura a také vysoká genetická divergence mezi západními a východními vzorky, která by mohla být interpretována jako bod zlomu mezi západním a východním Středomoří.

Východní a západní středomořské populace se zdají být oddělené u několika mořských druhů. Tento jev se může projevat v důsledku glacialní nebo postglaciální evoluční historie moře a současných mořských proudů (např. pro mořské trávy *Posidonia oceanica* (Serra et al., 2010); pro ryby *Dicentrarchus labrax* (Bahri-Sfar et al., 2000). Populačně genomická metoda (RAD seq) zatím na populační strukturu chobotnic ani na jiné hlavonožce nebyla použita. V rámci Středozemního moře byla studována populace endemické ryby *Symphodus tinca* pomocí metody GBS. Byli studováni jedinci ze šesti míst v Jaderském a Jónském moři a získalo se 4 155 SNPs. Touto prací byla zjištěna diferenciace organismu v rámci Jaderského moře v severní a jižní oblasti.

Velice podobně jako u studie Fadhlouli-Zid et al. (2012) vyšly výsledky i v této práci s mitochondriálním markerem COI. Z fylogenetického stromu (Obr. 4) i z populační sítě COI (Obr. 5) lze vyzorovat, že struktura populace *Eledone moschata* koreluje s geografickým původem. Jedinci jsou jasně rozděleni do dvou populací a to do populace v Tyrhénském a Jaderském moři. Nejedná se o oblasti oddělené jasnou geografickou bariérou, je ale možné, že struktura populace je ovlivněna i mořskými proudy a mořskými bariérami (fronts).

Toto tvrzení potvrzují výsledky  $\Phi_{ST}$  statistiky (Tab. XI, Tab. XII, Tab. XIII). Narozdíl od základní statistiky  $F_{ST}$  je statistika  $\Phi_{ST}$  relativně neovlivněná fenoménem vysoké alelické diverzity, která snižuje odhady genetické diferenciace (Bird et al., 2011). Díky vyšším hodnotám statistiky s rozdělením „sea” (Tab. XI) se zdá, že mezi vzdálenějšími populacemi z Tyrhénského a Jaderského moře existuje diferenciace. Jemnější statistiky (Tab. XII, Tab. XIII) potvrzují toto rozdělení. Mezi vzorky z Tyrhénského i Jaderského moře jsou výrazně nižší hodnoty  $\Phi_{ST}$  oproti porovnání těchto populací mezi sebou. Bylo by však vhodné tuto hypotézu ověřit pomocí Mantelova testu (IBD). Podle tabulky (Tab. X) je zjevné, že  $\Phi_{ST}$  porovnání mezi druhy může být zkreslené, pravděpodobně kvůli málo společným lokusům.

Populační strukturu můžeme pozorovat díky koancestrální shlukovací analýze (fineRADstructure, Obr. 7). Vnitřní populační struktura je zřejmá i přes nižší počet vzorků v analýze. Geografické oblasti Tyrhénského a Jaderského moře jsou jasně oddělené. Jedinci v Tyrhénském moři jsou v populační struktuře jasně vymezení do dvou populací v oblasti kolem Sardinie a západního pobřeží Itálie. Jedinci v Jaderském moři jsou rozděleni do dvou populací na jižní a severní oblast moře. Nicméně výsledky mohou být zavádějící

kvůli vysoké hodnotě chybějících dat u některých vzorků. Například vzorek OEIC 2 H, který podle koancestrální shlukovací analýz tvoří vlastní populaci v Jaderském moři, má vysoké procento chybějících dat (skoro 80%) a výsledek proto nemusí být věrohodný.

Porovnáním fylogenetického stromu mitochondriálního genu COI (Obr. 4) a koancestrální shlukovací analýzy (Obr. 7), je na první pohled zjevné, že koancestrální shlukovací analýza vytvořila jemnější populační strukturu. Tato studie dokazuje, že metoda RAD seq je vhodnější pro populační studie, než mitochondriální gen COI, jelikož dokáže objasnit i jemnější strukturu v rámci populací.

Tato práce přináší poznatky o populační struktuře *Eledone moschata* ve Středozemním moři. Studium populační struktury tohoto organismu je zajímavé, jelikož se jedná o málo studovaný druh a také proto, že *Eledone moschata* často lovena rybáři a proto je důležité mít povědomí o tom, zda se jedná pouze o jednu populaci anebo vícero. Bylo by vhodné tuto práci rozšířit o další statistiky, jako je například Admixture, který zkoumá přítomnost DNA jedinců v rámci celé populace a zkoumá tak míru inbreedingu mezi populacemi, LD (linkage disequilibrium), který určuje nenáhodné asociace alel u různých lokusů v populaci, a nebo IBD (Isolation by distance), která porovnává fyzickou (geografickou) a genetickou vzdálenost jedinců. Bylo dobré získat více vzorků z vícera lokalit po celém Středozemním moři. Lokality, na které by se bylo vhodné zaměřit, jsou například oblasti kolem Gibraltarského průlivu, Egejského moře nebo také v okolí Tuniska. Také by bylo vhodné rozšířit dataset ze stávajících lokalit a pokračovat s těmito daty v důkladnějším výzkumu.

## 7. Závěr

Na základě genetické diferenciace populací ( $\Phi_{ST}$ ) a koancestrální shlukovací analýzy byla objasněna populační struktura *Eledone moschata*. Díky metodě ddRAD a následným next-gen sekvenováním na platformě Illumina byla zjištěna strukturovanost populace na jemnější úrovni, než pomocí analýzy mitochondriálního genu COI. Pomocí těchto metod bylo zjištěno, že jedinci jsou diverzifikováni dle geografického původu.

## 8. Reference

- Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L., and Lander, E.S. (2000).** An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407, 513–516.
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016).** Harnessing the power of RAD seq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81–92.
- Bahri-Sfar, L., Lemaire, C., Hassine, O. K. B., & Bonhomme, F. (2000).** Fragmentation of sea bass populations in the western and eastern Mediterranean as revealed by microsatellite polymorphism. *Proceedings of the Royal Society B: Biological Sciences*, 267(1446), 929–935.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008).** Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, 3(10).
- Bandelt, H., Forster, P., Röhl, A. (1999).** Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, 16(1):37–48.
- Belcari, P. and M. Sbrana. (1999).** *Eledone moschata*. In: G. Relini, J. A. Bertrand and A. Zamboni (eds.), *Synthesis of the knowledge on bottom fishery resources in Central Mediterranean (Italy and Corsica)*. *Biol. Mar. Medit.*, 6(suppl. 1): 747-752.
- Belcari, P., Tserpes, G., Gonzalez, M., Lefkaditou, E., Marceta, B., Piccinetti, Manfrin G., et al. (2002).** Distribution and abundance of *Eledone cirrhosa* (Lamarck, 1798) and *E. moschata* (Lamarck, 1798; Cephalopoda: Octopoda) in the Mediterranean Sea. *Scientia Marina*, 66(Suppl 2), 143–155
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., et al. (2008).** Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), 53–59.
- Berger, J., Suzuki, T., Senti, K.-A., Stubbs, J., Schaffner, G., & Dickson, B. J. (2001).** Genetic mapping with SNP markers in *Drosophila*. *Nature Genetics*, 29(4).
- Bird, C., Karl, S., Mouse, P., & Toonen, R. (2011).** Detecting and measuring genetic differentiation. *Phylogeography and Population Genetics in Crustacea*, 31–55.

- Buermans, H. P. J., & den Dunnen, J. T. (2014).** Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842(10), 1932–1941.
- Casu, M., Maltagliati, F., Meloni, M., Casu, D., Cossu, P., Binelli, Castelli, A. (2002).** Genetic structure of *Octopus vulgaris* (Mollusca, Cephalopoda) from the Mediterranean Sea as revealed by a microsatellite locus. *Italian Journal of Zoology*, 69(4), 295–300.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013).** Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140.
- Clarke, M. R. (1996).** Cephalopods as Prey. III. Cetaceans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 351(1343), 1053–1065.
- Collins, F.S., Lander, E.S., Rogers, J., and Waterson, R.H. (2004).** Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.
- Cox, A. J., & Hebert, P. D. N. (2001).** Colonization, extinction, and phylogeographic patterning in a freshwater crustacean. *Molecular Ecology*, 10(2), 371–386.
- De Luca, D., Catanese, G., Procaccini, G., & Fiorito, G. (2016).** *Octopus vulgaris* (Cuvier, 1797) in the Mediterranean Sea: Genetic Diversity and Population Structure. *PLoS ONE*, 11(2).
- Denver, D. R., Dolan, P. C., Wilhelm, L. J., Sung, W., Lucas-Lledo, J. I., Howe, D. K., et al. (2009).** A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proceedings of the National Academy of Sciences*, 106(38), 16310–16314.
- Díaz-Arce, N., Arrizabalaga, H., Murua, H., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016).** RAD seq derived genome-wide nuclear markers resolve the phylogeny of tunas. *Molecular Phylogenetics and Evolution*, 102, 202–207.
- Dugan, K.A., Lawrence, H.S., Hares, D.R., Fisher, C.L., and Budowle, B. (2015).** An Improved Method for Post-PCR Purification for mtDNA Sequence Analysis. *Journal of Forensic Sciences* 47.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011).** A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6.
- Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010).** Resolving postglacial phylogeography using high-

throughput sequencing. *Proceedings of the National Academy of Sciences*, 107(37), 16196–16200.

- Fadhlaoui-Zid, K., Knittweis, L., Aurelle, D., Nafkha, C., Ezzeddine, S., Fiorentino, F., Maltagliati, F. (2012).** Genetic structure of *Octopus vulgaris* (Cephalopoda, Octopodidae) in the central Mediterranean Sea inferred from the mitochondrial COIII gene. *Comptes Rendus Biologies*, 335(10-11), 625–636.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. (1994).** DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3:294–99.
- Fox, E. J., & Reid-Bayliss, K. S. (2014).** Accuracy of Next Generation Sequencing Platforms. *Journal of Next Generation Sequencing & Applications*, 01(01).
- Guerra, A. (1992).** Mollusca Cephalopoda. In: M.A. Ramos et al. (eds.), *Fauna Iberica*, vol. 1, pp. 1- 327. Museo Nacional de Ciencias Naturales - CSIC, Madrid.
- Hafner, M.S., Sudman, P.D., Villablanca, F.X., Spradling, T.A., Demastes, J.W., Nadler, S.A. (1994).** Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science*, 365:1087–1090.
- Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003).** Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, 270(Suppl\_1), S96–S99.
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S. (2018).** UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*, 35:518–522.
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010).** Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics*, 6(2).
- Cheng, S. H. (2015).** Evolution and Population Genomics of Loliginid Squids. UCLA. ProQuest ID: Cheng\_ucla\_0031D\_14167. Merritt ID: ark:/13030/m5xd459h. Retrieved from <https://escholarship.org/uc/item/0zw3h4ps>
- Jeffries, D. L., Copp, G. H., Lawson Handley, L., Olsén, K. H., Sayer, C. D., & Hänfling, B. (2016).** Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. *Molecular Ecology*, 25(13), 2997–3018.



- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S. (2017).** ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods*, 14:587-589.
- Katoh, K., Misawa, K., Kuma, K., Miyata, T. (2002).** MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059-3066.
- Keightley, P.D., Trivedi, U., Thomson, M., Oliver, F., Kumar, S., and Blaxter, M.L. (2009).** Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Research* 19, 1195–1201.
- Kerstens, H.H.D., Crooijmans, R.P.M.A., Veenendaal, A., Dibbits, B.W., Chin-A-Woeng, T.F.C., den Dunnen, J.T., and Groenen, M.A.M. (2009).** Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: Applied to Turkey. *BMC Genomics* 10, 479.
- Keskin, E., & Atar, H. H. (2011).** Genetic divergence of *Octopus vulgaris* species in the eastern Mediterranean. *Biochemical Systematics and Ecology*, 39(4-6), 277–282.
- Lal, M. M., Southgate, P. C., Jerry, D. R., & Zenger, K. R. (2016).** Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *Marine Genomics*, 25, 57–68.
- Lamarck, (1798),. *Eledone moschata*.**
- Lamichhaney, S., Barrio, A. M., Rafati, N., Sundstrom, G., Rubin, C.-J., Gilbert, E. R., et al. (2012).** Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences*, 109(47), 19345–19350.
- Leigh, J.W., Bryant, D. (2015).** PopART: Full-feature software for haplotype network construction. *Methods Ecol Evol* 6(9):1110–1116.
- Lescak, E. A., Bassham, S. L., Catchen, J., Gelmond, O., Sherbick, M. L., von Hippel, F. A., & Cresko, W. A. (2015).** Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences*, 112(52), E7204–E7212.
- Lima, F. D., Leite, T. S., Haimovici, M., Nóbrega, M. F., & Oliveira, J. E. L. (2014).** Population structure and reproductive dynamics of *Octopus insularis* (Cephalopoda: Octopodidae) in a coastal reef environment along northeastern Brazil. *Fisheries Research*, 152, 86–92.

- Linck, E. B., Hanna, Z. R., Sellas, A., & Dumbacher, J. P. (2017).** Evaluating hybridization capture with RAD probes as a tool for museum genomics with historical bird specimens. *Ecology and Evolution*, 7(13), 4755–4767.
- Liu, T., Guo, L., Pan, Y., Zhao, Q., Wang, J., & Song, Z. (2016).** Construction of the first high-density genetic linkage map of *Salvia miltiorrhiza* using specific length amplified fragment (SLAF) sequencing. *Scientific Reports*, 6(1).
- Malinsky M, Trucchi E, Lawson DJ, Falush D (2018).** RADpainter and fineRADstructure: population inference from RADseq data. *Molecular Biology and Evolution*, OpenAccess.
- Maltagliati, F., Belcari, P., Casu, D., Casu, M., Sartor, P., Vargiu, G., & Castelli, A. (2002).** Allozyme genetic variability and gene flow in *Octopus vulgaris* (Cephalopoda, Octopodidae) from the Mediterranean sea. In *Bulletin of Marine Science*, 71, 473–486
- Mangold, K., Boletzky, S.V. (, 1988).** Mediterranean Cephalopod fauna. *The Mollusca*. 12: 315-330.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005).** Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376–380.
- Metzker, M. L. (2009).** Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1), 31–46.
- Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., and Johnson, E.A. (2007).** Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17, 240–248.
- Morse, P., Kjeldsen, S. R., Meekan, M. G., McCormick, M. I., Finn, J. K., Huffard, C. L., & Zenger, K. R. (2018).** Genome-wide comparisons reveal a clinal species pattern within a holobenthic octopod—the Australian Southern blue-ringed octopus, *Hapalochlaena maculosa* (Cephalopoda: Octopodidae). *Ecology and Evolution*.
- Norman M., (2000).**, *Cephalopods: A World Guide - Octopuses, Argonauts, Cuttlefish, Squid, Nautilus*, ed. Conchbooks, 319p.
- Orbigny, A. D. d'. (1845).** *Mollusques vivants et fossiles, ou, description de toutes les espèces de coquilles et de mollusques : classées suivant leur distribution géologique et géographique.* Paris: Gide et Cie, Éditeurs, 1845.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012).** Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, 7(5)

- Piatkowski, U., Pütz, K., & Heinemann, H. (2001).** Cephalopod prey of king penguins (*Aptenodytes patagonicus*) breeding at Volunteer Beach, Falkland Islands, during austral winter 1996. *Fisheries Research*, 52(1-2), 79–90.
- Planes, S., Jones, G. P., & Thorrold, S. R. (2009).** Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences*, 106(14), 5693–5697.
- Querejeta, M., González-Esteban, J., Gómez, A., Fernández-González, A., Aymerich, P., Gosálbez, J., et al. (2016).** Genomic diversity and geographical structure of the Pyrenean desman. *Conservation Genetics*, 17(6), 1333–1344.
- R Core Team (2013).** R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Recknagel, H., Elmer, K. R., & Meyer, A. (2013).** A Hybrid Genetic Linkage Map of Two Ecologically and Morphologically Divergent Midas Cichlid Fishes (*Amphilophus* spp.) Obtained by Massively Parallel DNA Sequencing (ddRADSeq). *Genes|Genomes|Genetics*, 3(1), 65–74.
- Ren, P., Peng, W., You, W., Huang, Z., Guo, Q., Chen, N., et al. (2016).** Genetic mapping and quantitative trait loci analysis of growth-related traits in the small abalone *Haliotis diversicolor* using restriction-site-associated DNA sequencing. *Aquaculture*, 454, 163–170.
- Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U., & Irigoien, X. (2016).** Population structure of Atlantic mackerel inferred from RAD seq-derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. *Molecular Ecology Resources*, 16(4), 991–1001.
- RStudio Team (2015).** RStudio: Integrated Development for R. RStudio, Inc., Boston, MA  
URL <http://www.rstudio.com/>.
- Rubin, B. E. R., Ree, R. H., & Moreau, C. S. (2012).** Inferring Phylogenies from RAD Sequence Data. *PLoS ONE*, 7(4).
- Salman, A., Katagan, T., and Benli, H. (2002).** Cephalopod fauna of the Eastern Mediterranean. *Turkish Journal of Zoology* 26, 47–52.
- Seetharam, A. S., & Stuart, G. W. (2013).** Whole genome phylogeny for 21 *Drosophila* species using predicted 2b-RAD fragments. *PeerJ*, 1.
- Şen, H. (2007).** Food preferences of *Eledone moschata* Lamarck, 1799 (Cephalopoda: Octopodidae) in Captive Conditions. *International Journal of Natural and Engineering Sciences* 1, 29–31

- Serra, I.A., Innocenti, A.M., Di Maida, G., Calvo, S., Migliaccio, M., Zambianchi, E., Pizzigalli, C., Arnaud-Haond, S., Duarte, C.M., Serrao, E.A., et al. (2010).** Genetic structure in the Mediterranean seagrass *Posidonia oceanica*: Disentangling past vicariance events from contemporary patterns of gene flow. *Molecular Ecology* 19, 557–568.
- Stickney, H. L. (2002).** Rapid Mapping of Zebrafish Mutations With SNPs and Oligonucleotide Microarrays. *Genome Research*, 12(12), 1929–1934.
- Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N. et al. (2016).** Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens. *PLoS ONE*, 11(3).
- Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W. et al. (2013).** SLAF-seq: An Efficient Method of Large-Scale De Novo SNP Discovery and Genotyping Using High-Throughput Sequencing. *PLoS ONE*, 8(3).
- Takahashi, T., & Moreno, E. (2015).** A RAD-based phylogenetics for Orestias fishes from Lake Titicaca. *Molecular Phylogenetics and Evolution*, 93, 307–317.
- Trifinopoulos, J., Nguyen L.-T., von Haeseler, A., Minh, B.Q. (2016).** W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis.. *Nucleic Acids Res.*, 44:W232-W235.
- Valencia, L. M., Martins, A., Ortiz, E. M., & Di Fiore, A. (2018).** A RAD-sequencing approach to genome-wide marker discovery, genotyping, and phylogenetic inference in a diverse radiation of primates. *PLoS ONE*, 13(8).
- Vvan Orsouw, N.J., Hogers, R.C.J., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., Schneiders, H., van der Poel, H., van Oeveren, J., Verstegen, H., et al. (2007).** Complexity reduction of polymorphic sequences (CRoPS<sup>TM</sup>): A novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE* 2.
- Van Tassell, C.P., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C., and Sonstegard, T.S. (2008).** SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* 5, 247–252.
- Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., et al. (1998).** Large-scale identification, mapping, and genotyping of single- nucleotide polymorphisms in the human genome. *Science* 280, 1077–1082.

- Wang, J., Xue, D.-X., Zhang, B.-D., Li, Y.-L., Liu, B.-J., & Liu, J.-X. (2016).** Genome-Wide SNP Discovery, Genotyping and Their Preliminary Applications for Population Genetic Inference in Spotted Sea Bass (*Lateolabrax maculatus*). *PLoS ONE*, 11(6).
- Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012).** 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, 9(8), 808–810.
- Wang, X. Q., Zhao, L., Eaton, D. A. R., Li, D. Z., & Guo, Z. H. (2013).** Identification of SNP markers for inferring phylogeny in temperate bamboos (Poaceae: Bambusoideae) using RAD sequencing. *Molecular Ecology Resources*, 13(5), 938–945.
- Wikimedia Commons contributors, (2017).** "File:Tyrrenian Sea map-2.png," Wikimedia Commons, the free media repository,  
[https://commons.wikimedia.org/w/index.php?title=File:Tyrrenian\\_Sea\\_map-2.png&oldid=236466981](https://commons.wikimedia.org/w/index.php?title=File:Tyrrenian_Sea_map-2.png&oldid=236466981).
- Willing, E.-M., Hoffmann, M., Klein, J. D., Weigel, D., & Dreyer, C. (2011).** Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics*, 27(16), 2187–2193.
- Winkelmann, I., Campos, P. F., Strugnell, J., Cherel, Y., Smith, P. J., Kubodera, T., et al. (2013).** Mitochondrial genome diversity and population structure of the giant squid *Architeuthis*: genetics sheds new light on one of the most enigmatic marine species. *Proceedings of the Royal Society B: Biological Sciences*, 280(1759), 20130273–20130273.
- [www.aquamaps.org](http://www.aquamaps.org) (2016).** Computer generated distribution maps for *Eledone moschata* (musky octopus), with modelled year 2100 native range map based on IPCC A2 emissions scenario. [www.aquamaps.org](http://www.aquamaps.org), version of Aug. 2016. Web. Accessed 31 Mar. 2019.
- Xu, X., Xu, R., Zhu, B., Yu, T., Qu, W., Lu, L., et al. (2015).** A high-density genetic map of cucumber derived from Specific Length Amplified Fragment sequencing (SLAF-seq). *Frontiers in Plant Science*, 5.
- Yang, H., Wei, C.-L., Liu, H.-W., Wu, J.-L., Li, Z.-G., Zhang, L. et al. X.-C. (2016).** Genetic Divergence between *Camellia sinensis* and Its Wild Relatives Revealed via Genome-Wide SNPs from RAD Sequencing. *PLoS ONE*, 11(3).
- Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M. A., Manzano, C., Grant, W. S., et al. (2012).** Multiple SNP Markers Reveal Fine-Scale Population and Deep Phylogeographic Structure in European Anchovy (*Engraulis encrasicolus* L.). *PLoS ONE*, 7(7)

- Zheng, X., Ikeda, M., Kong, L., Lin, X., Li, Q., & Taniguchi, N. (2009).** Genetic diversity and population structure of the golden cuttlefish, *Sepia esculenta* (Cephalopoda: Sepiidae) indicated by microsatellite DNA variations. *Marine Ecology*, 30(4), 448–454.
- Zhou, X., Xia, Y., Ren, X., Chen, Y., Huang, L., Huang, S. et al. (2014).** Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). *BMC Genomics*, 15(1), 351.

## 9. Přílohy

Tab. IX: Hloubka pokrytí studovaných vzorků.

Jedinec	coverage
OEIS1H	14,72x
OEIT1H	6,88x
OEIT3H	7,49x
OEIT5H	5,63x
OEIC2H	7,75x
OEIC5H	17,94x
OEIC6H	17,12x
OEIG7H	16,96x
OEIG10H	13,06x
OEIL6H	16,06x
OEIS4H	12,67x
OCEP7H	18,85x
OECP10H	20,07x
OEIV11H	16,17x
OEIV13H	15,38x
OEIV14H	15,78x
OEIR17H	19,95x
OEIR18H	16,94x
OEV2H	16,39x
OEV3H	14,39x
OEIM1H	9,29x
OEIM2H	11,66x
OeT3H	18,69x
OeT4H	16,28x