UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA

Tjaša REŠETIČ

ANALIZA TRANSKRIPTOMA PLODU OLJKE (*Olea europaea* L.) S PIROSEKVENCIRANJEM

DOKTORSKA DISERTACIJA

Ljubljana, 2013

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA

Tjaša REŠETIČ

ANALIZA TRANSKRIPTOMA PLODU OLJKE (*Olea europaea* L.) S PIROSEKVENCIRANJEM

DOKTORSKA DISERTACIJA

ANALYSIS OF OLIVE FRUIT TRANSCRIPTOME (*Olea europaea* L.) BY PYROSEQUENCING

DOCTORAL DISSERTATION

Ljubljana, 2013

Doktorska disertacija je zaključek enovitega doktorskega podiplomskega študija bioloških in biotehniških znanosti. Raziskava je bila v celoti opravljena na Katedri za genetiko, biotehnologijo, žlahtnjenje rastlin in statistiko Oddelka za Agronomijo Biotehniške fakultete Univerze v Ljubljani. Del analiz je bil opravljen na Inštitutu za sredozemsko kmetijstvo in oljkarstvo, Znanstveno-raziskovalnega središča Univerze na Primorskem..

Na podlagi Statuta Univerze v Ljubljani ter po sklepu senata Biotehniške fakultete in sklepa senata Univerze z dnem 27. September 2010 je bilo potrjeno, da kandidatka izpolnjuje pogoje za neposreden prehod na doktorski študij bioloških in biotehniških znanosti ter opravljanje doktorata znanosti s področja biotehnologije. Za mentorja doktorske disertacije z naslovom » Analiza transkriptoma plodu oljke (*Olea europaea* L.) s pirosekvenciranjem« je bil imenovan doc. dr. Jernej Jakše.

Komisija za oceno in zagovor:

Predsednik: prof. dr. Peter DOVČ
Univerza v Ljubljani, Biotehniška fakulteta, oddelek za zootehniko

Član: doc.dr. Jernej JAKŠE
Univerza v Ljubljani, Biotehniška fakulteta, oddelek za agronomijo

Član: doc. dr. Dunja BANDELJ
Univerza na Primorskem, Znanstveno-raziskovalno središče Koper, Inštitut za sredozemsko kmetijstvo in oljkarstvo
Univerza na Primorskem, Fakulteta za matematiko, naravoslovje in informacijske tehnologije

Datum zagovora: 20.12.2013

Doktorat je rezultat lastnega raziskovalnega dela. Podpisana se strinjam z objavo svojega dela v polnem tekstu na spletni strani Digitalne knjižnice Biotehniške fakultete. Izjavljam, da je delo, ki sem ga oddala v elektronski obliki, identično tiskani verziji.

Tjaša REŠETIČ

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD Dd

DK UDK 634.63:606:577.2(043)

KG Oljka/'Istrska belica'/cDNA knjižnica/pirosekvenciranje /biotehnologija

KK AGRIS F30

AV REŠETIČ, Tjaša, dipl. inž. agr.

SA JAKŠE, Jernej (mentor)

KZ SI-1000 Ljubljana, Jamnikarjeva 101

ZA Univerza v Ljubljani, Biotehniška fakulteta, Podiplomski študij bioloških in biotehniških znanosti, področje biotehnologije

LI 2013

IN ANALIZA TRANSKRIPTOMA PLODU OLJKE (*Olea europaea* L.) S PIROSEKVENCIRANJEM

TD Doktorska disertacija

OP XI, 111 str., 8 pregl., 23 sl., 140 vir.

IJ sl

JI sl/en

AI Oznake izraženih nukleotidnih zaporedij (EST) in ostala cDNA zaporedja so ena izmed primarnih orodij, ki nam hitro in poceni zagotovijo neposredne informacije o izraženih delih genoma. Za namen izdelave oljčnih EST smo vzorčili razvijajoče plodove oljk preko celotne razvojne faze. Izolirali smo RNA vzorce in preverili njeno integriteto. Normalizirano cDNA knjižnico smo izdelali s pomočjo dupleks-specifične nukleaze (DSN). Nukleotidna zaporedja smo določili s pomočjo pirosekvenatorja Roche 454 FLX, ki spada v kategorijo naslednjih generacij določevanja nukleotidnih zaporedij. Pridobili smo 560.578 zaporedij, katerih povprečna dolžina je bila 286 bp, skupna dolžina pa 160 Mb. Konkatemerna zaporedja smo ločili, ter jim odstranili dele zaporedij, uporabljenih pri manipulaciji cDNA knjižnice, poli A regije in vsa prekratka zaporedja. Na koncu smo pridobili 577.025 očiščenih zaporedij oljčne cDNA s povprečno dolžino 242 bp in N50 vrednostjo 294 bp. Zaporedja smo na osnovi podobnosti in dolžine prekrivanja združili v domnevna konsenzus zaporedja z uporabo različnih programov za združevanje zaporedij: TGICL2.1, Mira 3.2, iAssembler 1.3, PAVE 2.5, Newbler 2.3, Newbler 2.6 in komercialni programski paket CLC Genomics Workbench. Naši rezultati so pokazali ustreznost pridobljenih NGS podatkov in razpoložlivih tehnologij za združevanje zaporedij transkriptoma nemodelnih organizmov. Z Blast2go orodjem smo uspešno anotirali 51 % od vseh zaporedij, ter jim pripisali vloge na ravni bioloških procesov, celičnih komponent in molekularnih funkcij. Cilj nadaljne raziskave je bil iz pridobljenih nukleotidnih podatkov določiti primerne referenčne gene (RGs) za analize razvijajočih se plodov oljke s pomočjo PCR v realnem času (qPCR). Določili smo 29 kandidatnih RGs in 12 točk vzorčenja plodov, ki so zajemala pet glavnih faz razvoja oljčnih plodov. Glede na rezultate geNorm algoritma, sta se za najboljša RGs izkazala TIP41 sorodni protein (TIP41) in TATA vezavni protein (TBP). Z uporabo teh dveh RGs smo analizirali štiri gene, ki sodelujejo v metabolizmu maščobnih kislin in dokazali različne vzorce izražanja, povezane z razvojem mezokarpa in zorenjem oljčnih plodov.

KEY WORDS DOCUMENTATION

ND Dd

DC UDC 634.63:606:577.2(043)

CX Olive/'Istrska belica'/cDNA library/pirosequencing /biotehnology

KK AGRIS F30

AV REŠETIČ, Tjaša, dipl. inž. agr.

SA JAKŠE, Jernej (supervisor)

PP SI-1000 Ljubljana, Jamnikarjeva 101

PB University of Ljubljana, Biotechnical Faculty, Postgraduate Study of Biological and Biotechnical Sciences, Field: Biotechnology

PY 2013

TY ANALYSIS OF OLIVE FRUIT TRANSCRIPTOME (*Olea europaea* L.) BY PYROSEQUENCING

DT Doctoral dissertation

NO XI, 111 p., 8 tab., 23 fig., 2 ann., 140 ref.

LA sl

Al sl/en

AB Sequencing of ESTs is one of the primary tool for gene discovery, which provides fast and economical information on the parts of the expressed genome. The research has been conducted for generating new EST sequences from developing olive fruit. Olive fruits were sampled during the whole development phase. RNA was isolated and checked for the integrity. The normalized cDNA library was constructed using Kamchatka crab duplex-specific nuclease. 560,578 sequences of average length of 286 bp (160 Mb in total) were generated using 454 Titanium FLX sequencing technology. Obtained concatemer sequences were separated. We have removed parts of the sequences used in the manipulation of a cDNA library, poly A region, and all too short sequence. So we finally gained 577,025 purified olive cDNA sequences with an average length of 241 bp. Performance of seven different assemblers or assembler wrappers (TGICL2.1, Mira 3.2, iAssembler 1.3, PAVE 2.5, Newbler 2.3, Newbler 2.6 and commercial CLC Genomics Workbench) was compared. Our results demonstrated suitability of NGS data and available assembly methodologies for transcriptome assembly of a non model organism. The Blast2Go tool successfully revealed an annotation for 51% of all sequences that describe gene products in terms of their associated biological processes, cellular components and molecular functions. The aim of the further investigation was to develop suitable reference genes (RGs) for RT-qPCR studies of developing olive fruit from 29 RG candidates. We used 12 sampling points to cover the five stages of olive fruit development. According to the results of the geNorm algorithm, the two best RGs were TIP41-like family protein (*TIP41*) and TATA binding protein (*TBP*). Using the two new RGs, four genes involved in the metabolism of fatty acids were studied and showed distinct expression patterns associated with mesocarp development and ripening stages.

KAZALO VSEBINE

KAZALO PREGLEDNIC

KAZALO SLIK

OKRAJŠAVE IN SIMBOLI

ang.	angleško
ATP	Adenozin-5'-trifosfat
BAC	umetni bakterijski kromosomi (ang. *bacterial artificial chromosome)*
bp	bazni par
$CaCl_2$	kalcijev klorid
cDNA	komplementarna DNA (ang. Complementary DNA)
CTAB	cetil trimetil amonijev bromid
DNA	deoksiribonukleinska kislina
dNTP	deoksi nukleotid trifosfat
DTT	ditiotreitol
EDTA	etilendiamintetraocetna kislina- dinatrijeva sol
EtBr	etidijev bromid
FAS	sinteza maščobnih kislin
GMO	gensko spremenjeni organizmi (ang. geneticaly modified organism)
HDL	holesterol HDL. dobri holesterol
HPLC	tekočinska kromatografija visoke ločljivosti (ang. high-pressure liquid chromatography)
LDL	holesterol LDL. slabi holesterol
LOX	lipoksigenaza
TAG	triacilglicerol
Taq	*Thermus aquaticus*
TBE	tris-boratni-EDTA elektroforetski pufer

SLOVARČEK

BLAST	lokalni algoritem poravnave BLAST (ang. Basic Local Alignment Search Tool)
BLAT	lokalni algoritem poravnave BLAT (ang. Blast Like Alignment Tool)
DBG	de Brujin graf metoda (ang. de Brujin Graph)
EST	izraženo nukleotidno zaporedje, tudi v množini izražena nukleotidna zaporedja (ang. Expressed Sequence Tag/Tags)
FRET	fluorescentne resonančne energije (ang. Flourescent Resonance Energy Transfer)
NGS	Naslednje generacije določevanja nukleotidnih zaporedij (ang. Next Generation Sequencing)
OLC	(ang. Overlap/Layout/Consensus)
PCR	verižna reakcija s polimerazo (ang. Polymerase Chain Reaction)
RFLP	polimorfizem dolžine restrikcijskih fragmentov (ang. Restriction Fragment Length Polymorphisem)
SFF	(ang. standard flowgram format)
SSAHA2	(ang. Sequence Search and Alignment by Hashing Algorithm)
qPCR	verižna reakcija s polimerazo v realnem času (ang. Quantitative Polymerase Chain Reaction)
QTL	kvantitativni lokusi (ang. Quantitative Trait Loci)
WGS	določanje nukleotidnega zaporedja po postopku WGS (ang. Whole Genome Shotgun)

1 UVOD

V slovenskih oljčnih nasadih je 'Istrska belica' najbolj zastopana sorta. K nagli širitvi v istrske oljčnike po pozebi leta 1956 so pripomogle številne pozitivne lastnosti, med drugimi odpornost na nizke temperature, samooplodnost, ter dobra in redna rodnost (Bandelj Mavsar in sod., 2005). 'Istrska belica' je poznana tudi po visoki vsebnosti skupnih biofenolov, ki je višja v primerjavi z mnogimi drugimi italijanskimi sortami (Uccella, 2000).

Razvoj plodov je genetsko reguliran proces na katerega vplivajo tudi okoljski dejavniki. Oljčni plodovi v razvoju se spreminjajo v velikosti, sestavi, barvi, teksturi, okusu in odpornosti na okoljske dejavnike. Za opredelitev in določitev genov, ki sodelujejo v teh procesih v plodovih, so razvili različna genomska orodja (npr. izražena nukleotidna zaporedja, mikromreže, itd.). Izražena nukleotidna zaporedja (ESTs, angl. Expressed Sequence Tags) in cDNA zaporedja so ena izmed primarnih orodij, ki nam zagotovijo neposredne informacije o transkriptih, ki kodirajo dele genoma in so eden od pomembnejših virov za raziskovanje transkriptoma (Nagaraj in sod., 2006). Baze podatkov EST so zelo koristno orodje za odkrivanje genov in markerjev, gensko kartiranje in funkcijske študije pri preučevanih organizmih (Ozgenturk in sod., 2010).

Pri določevanju zaporedij cDNA prihaja pogosto do problemov zaradi homopolimernih adeninskih repov, zato se poskuša ta problem rešiti z njihovim odstranjevanjem oz. krajšanjem. Eden izmed načinov je tudi vnašanje mesta za restrikcijsko endonukleazo tipa II *Gsu*I, ki cepi 16|14 bp stran od prepoznavnega mesta. Na tak način pripravljena knjižnica cDNA naj bi bila primernejša za direktno ali transkripcijsko sekvenciranje, saj zaradi krajših poli A mest ne prihaja do zdrsov polimeraze ali prekomernega signala pri pirosekvenciranju (Shibata in sod., 2001).

Pridobljene sekvenčne informacije o transkriptih gredo skozi različne faze obdelave podatkov, kot so sestavljanje zaporedij za opredelitev domnevnih transkriptov, anotacija sestavljenih podatkov in njihova uporaba. Celotno urejanje informacij transkriptoma ni enostavno, saj posamezna zaporedja lahko vsebujejo napake in polimorfizme, ki onemogočajo njihovo optimalno obdelavo (Kumar in Blaxter, 2010). Slednje je še posebej izrazito pri visoko heterozigotnih organizmih, kar je tudi oljka.

Znani so fiziološki in biokemični podatki o rasti, razvoju in zorenju oljčnih plodov, vendar pa v glavnih genskih podatkovnih bazah še vedno ni veliko podatkov o zaporedjih genov in genskih produktih oljke (Galla in sod., 2009). Z izboljšanjem našega znanja o sestavi genov in njihovem izražanju, bomo pridobili nove informacije o procesu razvoja, fiziologiji dozorevanja, primarnem metabolizmu in sintezi zdravilnih substanc biofenolov

v oljčnem plodu. Te informacije nam lahko pomagajo pri izboljšanju kvalitativnih in kvantitativnih lastnosti oljčnih produktov.

Osnovni namen študije je bil pridobiti kakovostno normalizirano knjižnico cDNA, ki predstavlja vse razvojne stadije oljčnega plodu sorte 'Istrska belica'. S pomočjo genomskega pristopa smo želeli določiti izražanje genov v različnih stopnjah razvoja oljčnih plodov.

Normalizirano cDNA smo uporabili za določevanje nukleotidnih zaporedij s pomočjo novih tehnologij sekvenciranja (Roche 454). V naši raziskavi smo uporabili večje število programov za obdelavo in združevanje zaporedij, t.i. zbirnikov (angl. assembler), kot so TGICL (Partea in sod., 2003), MIRA (Chevreux in sod., 2000), iAssembler (Zheng in sod., 2011), gsAssembler 2.3. in gsAssembler 2.5, Pave 2.5 (Soderlund in sod., 2009), CLC Genomics Workbench in tako poskušali pridobili podatke o optimalni programski opremi.

S pomočjo nadaljne analize zaporedij EST iz razvijajočega plodu oljke smo pridobili informacije o kandidatnih genih, ki imajo pomembno funkcijo v primarnem in sekundarnem metabolizmu oljčnih plodov, s PCR analizo genske ekspresije s PCR v realnem času (qPCR) pa smo preverili tkivno in časovno specifičnost izražanja določenih genskih transkriptov oljke, ki smo jih pridobili z določevanjem zaporedij knjižnice. Pričakujemo, da bodo rezultati doprinesli k novim spoznanjem o biokemijskih karakteristikah primarnega (sinteza maščobnih kislin) in sekundarnega (sinteza biofenolov) metabolizma oljčnega plodu. Prav tako bodo rezultati imeli doprinos k boljšemu bazičnemu znanju bioloških procesov pri rastlinah. Sekundarni metabolizem biofenolov je zelo pomemben in zanimiv zaradi njihovega doprinosa k stabilnosti in trajnosti oljčnega olja, učinkovine pa so tudi zanimive s farmacevtskega stališča.

Predlagana študija razvoja EST zaporedij oljčnega plodu bo poleg tega uporabna za raziskovalne skupine, ki delajo na raziskavah oljke. Preko orodij primerjalne genomike lahko rezultate apliciramo tudi na raziskave ostalih ekonomsko pomembnih rastlin.

V raziskavi bomo:

- pridobili dovolj kakovostne vzorce RNA iz razvijajočih plodov oljk, ki bodo primerni za razvoj normalizirane cDNA knjižnice;
- z odstranitvijo poli A regij pripomogli k boljši izvedbi določevanja nukleotidnih zaporedij;
- določili večjo količino nukleotidnih podatkov za razvijajoči plod oljke;
- optimalno združili pridobljena zaporedja s pomočjo izbranega zbirnika;
- določili transkripte, ki so povezani s primarnim in sekundarnim metabolizmom oljčnega plodu;

- potrdili tkivno specifično izražanje za nekatere ključne transkripte s pomočjo PCR v realnem času (qPCR).

Naloga zajema naslednje delovne sklope:

- vzorčenje plodov oljk (sorte 'Istrska belica') skozi celotno obdobje razvoja od začetka junija, ko se zaključi faza cvetenja, do sredine novembra, ko se prične obiranje plodov oz. nastopi fiziološka zrelost;
- izolacijo RNA, kjer smo uporabili Spectrum Plant Total RNA Extraction Kit-a in izdelavo normalizirane cDNA knjižnico s pomočjo dupleksno specifične nukleaze (Bogdanova is sod., 2008);
- optimizacijo cDNA knjižnice in določevanje nukleotidnih zaporedij s pomočjo Roche 454 tehnologije naslednje generacije;
- bioinformacijsko obdelavo podatkov in združitev prečiščenih zaporedij na osnovi podobnosti v domnevna konsenzna zaporedja z uporabo zbirnikov TGICL (Partea in sod., 2003), MIRA (Chevreux in sod., 2000), iAssembler (Zheng in sod., 2011), gsAssembler 2.3. in gsAssembler 2.5, Pave 2.5 (CITAT), CLC Genomics Workbench;
- primerjava različnih združevanj s pomočjo programa BLAT (Kent, 2002), s katerim pridobimo število unikatnih zaporedij posameznega združevanja v primerjavi z ostalimi programi;
- uporaba BLASTN algoritma in TBLASTX algoritma (Altschul in sod., 1990), ki nam omogoča iskanje homologij na podlagi domnevnega prevedenega aminokislinskega zaporedja EST-ja (v vseh 6 možnih bralnih okvirjih) in je natančnejši za iskanje homologij med bolj oddaljenimi vrstami (Grant in sod., 2000) ter izbor najboljšega seta združenih podatkov;
- funkcijska analiza oljčnih zaporedij z uporabo programskega paketa Blast2go, ki izražena nukleotidna zaporedja razdelijo v tri glavne Gene Ontology sklope (celični prostor, molekularne funkcije, biološki procesi) (Götz in sod., 2008);
- določitev primernih referenčnih genov za qPCR analizo razvijajočih se oljčnih plodov ter tkivna in časovna specifičnost izražanja določenih genskih transkriptov, vpletenih v metabolizem maščobnih kislin oljke s pomočjo qPCR.

2 PREGLED OBJAV

2.1 OLJKA

2.1.1 Botanična klasifikacija oljke

Znanstveno ime oljke je *Olea europaea* L., katerega botanični izraz Olea naj bi izhajal iz latinskega izraza »oleum« in grškega izraza »elaia«, ki pomeni rastlino, ki proizvaja olje. Ime »*europaea*« pa je predlagal Linnaeus leta 1764, kar kaže na to, da je oljka evropskega izvora in tipična rastlina mediteranskega območja (Ganino in sod., 2006).

Botanično oljka pripada družini Oleaceae, ki jo večina klasifikacij deli na dve glavni podružini, in sicer Jasminoideae in Oleideae (Ganino in sod., 2006). Vendar pa najnovejša klasifikacija opušča to delitev in predstavlja novo delitev družine Oleaceae na 5 plemen, ki vsebujejo 25 rodov s 600 različnimi vrstami. Pod omenjenih pet plemen štejemo Myxopyreae (vključuje rodove *Myxopyrum, Nyctanthes* in *Dimetra*), Fontanesieae (vključuje rod *Fontanesia*), Forsythieae (vključuje rodova *Forsythia* in *Abeliophyllum*), Jasmineae (vključuje rodova *Jasminum* in *Menodora*) in Oleae, slednji pa vsebuje podplemena Ligustrinae (rodova *Syringa* in *Ligustrum*), Schreberinae (rodova *Schrebera* in *Comoranthus*), Fraxininae (rod *Fraxinus*) in Oleinae (s preostalimi 12 rodovi) (Wallander in Albert, 2001). Za pleme Oleae je značilno, da je monofiletska skupina, katere predstavniki imajo osnovno število kromosomov 23 in jih združujejo številne anatomske, morfološke in kemijske podobnosti (Wallander in Albert, 2001). Med rodovi družine *Oleaceae* najdemo predvsem rodove, ki se gojijo kot okrasne rasline, z izjemo rodov *Fraxinus* L. in *Olea* L., ki sta zelo pomembna tudi s kmetijskega in ekonomskega stališča.

Oljko uvrščamo v rod *Olea*, ki vsebuje 17 vrst in podvrst. Te vrste in podvrste pripadajo trem linijam, ki se razprostirajo preko Afrike, južne Evrope, Azije in Oceanije (Besnard in sod., 2009). Znotraj rodu *Olea* je tudi tako imenovani oljčni kompleks oziroma kompleks *O. europaea*, ki ga sestavlja 6 podvrst značilnih za določeno geografsko območje (Kernerman in sod., 1992). *O. europaea* podvrsta *europaea* je poznana kot sredozemska oljka, *O. e.* podvrsta *laperrinei* (Bratt. & Trab.) se nahaja v Saharskem gorovju, podvrsto *O. e. cuspidata* (Wall.) najdemo od južne Afrike in Egipta do Arabije in Kitajske, *O. e.* podvrsta *guanchina* se nahaja na Kanarskih otokih, *O. e.* podvrsta *maroccana* (Greut. & Burd.) je značilna za južni Maroko, ter *O. e.* podvrsta *cerasiformis* (Webb & Berth.), ki je endemična na otoku Madeira. V mediteranskem območju razlikujemo dve varieteti sredozemske oljke, in sicer gojeno sredozemsko oljko (*Olea europaea* L. subsp. *europaea* var. *europaea*) in divjo oljko (*Olea europaea* L. subsp. *europaea* var. *sylvestris* = var. *oleaster*). Divje oljke ali oleastre delimo glede na to ali uspevajo brez kultiviranja v sredozemskih gozdovih ali pa nastanejo s spontanim križanjem med divjimi rastlinami in sortami v opuščenih nasadih (Besnard in sod., 2002b).

2.1.2 Biološke značilnosti oljke

Oljka je verjetno eno izmed prvih gojenih sadnih dreves, ki jih danes naštejemo kar 800 milijonov v svetovnem merilu, največ v sredozemskem območju (Kaniewski in sod., 2012). Sredozemska oljka je počasi rastoča zimzelena vrsta, tolerantna na sušne razmere. Drevo lahko zraste tudi do 15 m višine, vendar pa so bolj pogoste grmičaste oblike razrasti. Oljka razvije močan, razraščen in širok, vendar dokaj plitek koreninski sistem, ki se v ugodnih talnih razmerah redkokdaj razvije globlje od 60 do 70 cm. Deblo je predvsem pri starejših rastlinah masivno, sivkaste barve, luknjasto in grčavo. Ker je za oljko značilno, da se deblo ne debeli enakomerno so letnice ob njegovem prerezu slabo opazne, deblo pa zaradi neenakomerne debelitve postane nagubano. Listi so podolgovati, ozki, suličasti do linearni s celimi robovi, v srebrno-zeleni barvi, na spodnji strani prekriti s sivobelimi dlačicami. Cvet oljke je dvospolen, hermafroditen, kar pomeni, da so v istem cvetu moški in ženski (rodni) organi. Majhni kremasto-beli cvetovi cvetijo v maju ali začetku junija in so združeni v grozdasta socvetja 10 do 40 cvetov (Slika 1), ki pa večinoma odmrejo tekom razvoja in ob obiranju dajo le nekaj plodov na socvetje (Sancin, 1990). V sredozemski regiji imajo glavni agronomski pomen oljčni plodovi, saj so vir oljčnega olja, uporabljajo pa se tudi kot vloženi plodovi (Vossen, 2007).

Slika 1: Cvetoča oljka sorte ‘Istrska belica’ (foto: D. Bandelj)
Figure 1: Flourishing olive tree variety ‘Istrska belica’ (photo: D. Bandelj)

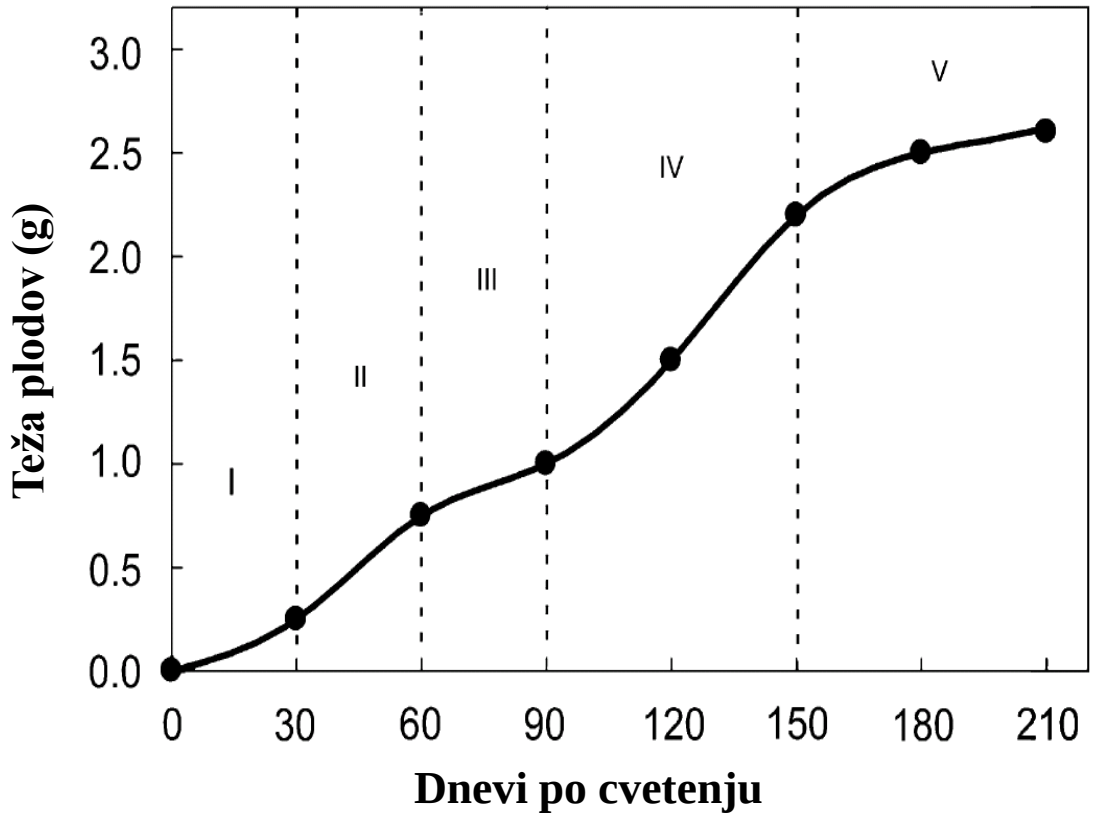
Oljčni plod sestavljajo eksokarp ali kožica, mezokarp ali meso in endokarp ali koščica. Slednja je sestavljena iz lesnatega ovoja, ki prekriva eno ali redkeje dva embrija. Celotna teža plodu zajema 70 % - 90 % mezokarpa, 9 % - 27 % endokarpa in 2 % - 3 % embrija.

Ob obiranju plodov naj bi mezokarp vseboval 60 % vode, 30 % olja, 4 % sladkorjev, 3 % proteinov in suho snov. Endokarp naj bi ob obiranju vseboval 10 % vode, 30 % celuloze, 40 % ostalih ogljikovih hidratov in le 1 % olja, medtem ko naj bi embrio vseboval 30 % vode, 27 % olja, 27 % ogljikovih hidratov in 10 % proteinov (Conde in sod., 2008). V primerjavi z oleastri ima večina gojenih oljk večje in bolj mesnate plodove, z višjim deležem olja (Kaniewski in sod., 2012).

Slika 2: Razvojne faze plodov sorte 'Istrska belica' (foto: T. Rešetič)
Figure 2: Fruit developmental stages of variety 'Istrska belica' (photo: T. Rešetič)

Med razvojem plodu prihaja do sprememb v njegovi velikosti, sestavi, barvi, teksturi in aromi. Razvoj in zorenje olj̆nih plodov je kombinacija biokemičnih in fizioloških pojavov, ki so posledica genske kontrole in vpliva okolja. Razvoj plodu oljke traja 4-5 mesecev in ga delimo na 5 glavnih faz razvoja (Slika 2). Prva faza vključuje oploditev in zasnovo plodu, zanjo je značilna hitra delitev celic in pospešena rast embria. Druga faza zajema razvoj semena, zanjo je značilno obdobje hitre rasti plodu (predvsem endokarpa), ki je posledica tako delitve celic, kot tudi rasti celic. V tretji fazi pride do otrditve koščice, rast plodu se upočasni, saj se celice endokarpa nehajo deliti in otrdijo. Četrta faza zajema razvoj mezokarpa, pri katerem pride do razširitve celic mesa plodu in intenzivne akumulacije olja. V peti fazi poteka zorenje plodu. V tem času se spremeni trdota in svetlo zelena barva plodu, ki prehaja preko rumene v škrlatno rdečo, vijolično in na koncu do

skoraj črne barve (Conde in sod., 2007). Teh pet faz razvoja oljčnega plodu je predstavljeno na Sliki 3.



Slika 3: Faze rasti in razvoja oljčnega plodu: (i) oplodnja in nastavek plodov, ta faza traja do 30 dni po cvetenju, značilne zanjo so hitre celične delitve za rast embria, (ii) razvoj semena, faza hitre rasti plodu zaradi intenzivnih celičnih delitev in povečanja, ki vključuje v glavnem rast in razvoj endokarpa (koščica), razvoj mesa (mezokarpa) je neznaten, (iii) otrditev koščice, v tej fazi se rast plodu umiri, ker se prenehajo debeliti celice endokarpa, seme otrdi, (iv) razvoj mezokarpa predstavlja drugo periode rasti plodu na račun povečanja celic in intenzivne akumulacije olja in (v) zorenje, ko se plod obarva iz temno zelene v svetlo zeleno / vijolično bravo (Conde in sod., 2009)

Figure 3: Olive fruit growth and stages of fruit development: (i) fertilization and fruit set, from flowering to approximately 30 d afterwards, characterized by rapid early cell division promoting embryo's growth, (ii) seed development, a period of rapid fruit growth due to both intense cell division and enlargement involving mainly growth and development of the endocarp (seed/pit), with little flesh (mesocarp) development, (iii) seed/pit hardening, during which fruit growth slows down as the endocarp cells stop dividing and become sclerified, (iv) mesocarp development, representing the second major period of fruit growth, due to the mesocarp development mainly by the expansion of preexisting flesh cells, and intense oil accumulation, and (v) ripening, when the fruit changes from dark lime green to lighter green/purple (Conde et al., 2009).

2.1.3 Izvor oljke in domestikacija

Oljčno drevo je rastlina antičnega sveta in tvori skupino najstarejših sadnih dreves v Sredozemskem območju. Je ena izmed prvih gojenih sadnih vrst in je že v času bronaste dobe predstavljala gospodarsko blaginjo za mnoge sredozemske družbe. Gojenje oljk sega v leto okrog 3500 pred našim štetjem, saj so se prvi znaki gojenja oljk (dobro ohranjene zoglenele oljčne koščice in zogleneli kosi oljčnega lesa) pojavil v bakreni Palestini okoli 3700-3500 pred našim štetjem (Zohary in Spiegelroy, 1975). Nekateri avtorji trdijo, da je bila osrčje domestikacije dolina Jordan med Galilejskim jezerom in Mrtvim morjem. Vendar so nedavna odkritja ta predel razširila na območje, kjer se raztezajo sodobni Izrael, Jordanija, Libanon in Sirija, v jugovzhodni del Turčije, ob reki Tigris in Evfrat, ter v Irak in zahodni del Irana (Kaniewski in sod., 2012). Gojena oljka izhaja iz divjih oljk in naj bi se pojavila v času neolitika, ko se je začelo širiti kmetovanje s semeni (znano kot neolitska revolucija), kar potrjuje nenadno povečanje semen v fosilnih ostankih v času bakrene dobe. Zgodnjo domestikacijo oljke je omogočilo tudi enostavno vegetativno razmnoževanje dragocenejših dreves (npr. tistih z večjimi plodovi) in vzpostavitev oljčnih nasadov (Zohary in Spiegel-Roy, 1975). Kljub splošnemu prepričanju, da gojenje oljk izvira iz bakrene Palestine, pa vzorci genetskih variacij ne podpirajo popolnoma te hipoteze (Besnard in sod., 2007).

Genetske študije predvidevajo, da je bila domestikacija oljke dolgotrajen in neprekinjen proces, ter da kultivarji izhajajo iz različnih virov populacij. Kljub temu pa analize genoma dedovanega po materini strani (analize kloroplastne in mitohondrijske DNA) kažejo na to, da večina sodobnih sort vsebuje materno linijo genov, ki je razširjena po celotnem sredozemskem območju (Besnard in sod., 2002a). Obstaja hipoteza, da je bila ta linija prvotno porazdeljena po vzhodnem Sredozemlju, nato pa se je s človeško dejavnostjo, ki je vključevala razširjanje kultiviranih oljk, razporedila po preostalem delu Sredozemlja. Nedavne raziskave, ki temeljijo na pregledu celotnega kloroplastnega genoma, podpirajo to hipotezo in navajajo, da je območje Levanta (obmorsko področje med Anatolijo in Egiptom) najverjetnejši izvorni center treh E1 haplotipov, ki so pogosti (>85 %) v kultivarjih in bi zato ta regija lahko predstavljala začetno območje domestikacije oljke (Besnard in sod., 2011). Po preostalih delih Sredozemlja so nato sledile tako imenovane sekundarne domestikacije oljk, ki so nastale s križanjem na novo uvedenih kultivarjev in takratnimi lokalnimi oblikami oljk (Kaniewski in sod., 2012). Dandanes tako poznamo več kot 2000 kultivarjev na Sredozemskem območju, njihova porazdelitev pa sovpada z divjimi sorodniki (Zohary in Spiegel-Roy, 1975). Kljub temu, da je oljka občutljiv temperaturni bioindikator Sredozemske regije, pa je morala tekom kultivacije preseči svoje naravne bioklimatske meje, saj jo danes gojimo na višjih nadmorskih višinah in geografskih širinah z distribucijo, ki daleč presega meje uspevanja divje oljke (Carrion in sod., 2010).

2.1.4 Kultivar ‘Istrska belica’

V slovenskih oljčnih nasadih je ‘Istrska belica’ najbolj zastopana sorta, k njeni nagli širitvi v istrske oljčnike po pozebi leta 1956 pa so pripomogle njene številne pozitivne agronomske lastnosti. Po navedbah je ‘Istrska belica’ avtohtona sorta in izvira iz območja Doline in Boljunca pri Trstu, vendar o njeni avtohtonosti ne obstajajo nobeni dokazi, zato jo uvrščamo med udomačene sorte. Znana je pod številnimi sinonimi, kot so: ‘Belica’, ‘Cepljena Belica’, ‘Žlahtna Belica’, ‘Bijelica’, ‘Istarska Bjelica’, ‘Bianchera’, ‘Bianca Istriana’, ‘Biancara’ (Bandelj Mavsar in sod., 2005) (Slika 4).

Slika 4: Travniški nasad oljk sorte ‘Istrska belica’ (foto: D. Bandelj)
Figure 4: Grassland olive grove of variety ‘Istrska belica’ (photo: D. Bandelj)

Za 'Istrsko belico' je značilna bujna, pokončna in metlasta rast, zaradi katere je oblikovanje krošnje zahtevno. Listi so suličasti, srednje veliki, socvetja pa srednje velika z 12 do 15 cvetovi. Cveti od konca maja do začetka junija, odvisno od vremenskih razmer. Je samooplodna sorta, ki v skrbno obdelanih nasadih dobro in redno rodi. Plodovi dozorevajo pozno od sredine novembra do sredine decembra in so ob začetku obiranja zelene barve, ki kasneje prehaja preko škrlatne v skoraj črno. ‘Istrska belica’ je zelo odporna proti nizkim

temperaturam, zato je priporočljiva za naše pridelovalno območje. Prav tako je odporna na razne bolezni in škodljivce, vendar je občutljiva na napad oljčne muhe in pavjega očesa (Sancin, 1990). Oljevitost sorte je visoka (do 20%) in daje olje dobre kakovosti, ki je znano po svoji bogati aromi, saj je zanj značilna visoka vsebnost biofenolov, ki dajejo olju grenkobo in pikantnost, ter daljšo dobo ohranjanja (Bešter in sod., 2008). Visoka vsebnost skupnih biofenolov pri sorti 'Istrska belica' je lahko tudi do dvakrat večja v primerjavi s sorto ‘Leccino’ (Butinar in sod., 2006), višja pa je tudi v primerjavi z nekaterimi drugimi italijanskimi sortami (Uccella, 2000).

2.1.5 Oljčno olje

Visoko biološko vrednost olja so intuitivno spoznali že v preteklih zgodovinskih dobah, saj je oljka ali njeno olje pogostoma prisotna v takratni mitologiji in bogoslužnih obredih (Sancin, 1990). Kljub temu, pa je bilo oljčno olje precej časa redko uporabljeno kot živilo. Kot del prehrane so go uporabljali le na obmorskih predelih Sredozemlja, in sicer samega ali kot dodatek k omakam in juham. Kakovost oljčnega olja so prvotno določali z okušanjem, Rimljani pa so bili prvi, ki so ga delili v tri različne razrede glede na kakovost (Bartolini in Petruccelli, 2002). Oljčno olje je bilo nekoč tesno povezano predvsem s kmečko in trgovsko civilizacijo mediteranskih narodov (Slika 4), vendar dandanes, ko stremimo h kakovostnejšemu življenju in prehrani, zopet odkrivamo stare in pozabljene vrednote tega hranila, kar potrjujejo tudi številna odkritja na področju človekove prehrane (Sancin, 1990). Svetovna poraba oljk in oljčnih izdelkov se je občutno povečala predvsem v visoko razvitih deželah, kot so Amerika, Evropa, Japonska, Kanada in Avstralija (Ryan in Robards, 1998). K temu je prispevala predvsem tradicionalna Mediteranska prehrana, ki temelji na rednem uživanju kakovostnega oljčnega olja in velja za eno izmed najbolj učinkovitih diet, saj dokazano varuje človeški organizem pred nekaterimi boleznimi sodobnega časa, kot so kardiovaskularne bolezni in določene oblike raka (Ryan in Robards, 1998; Soler-Rivas in sod., 2000). Oljčno olje naj bi ugodno vplivalo na človekovo zdravje predvsem zaradi visoke vsebnosti nenasičenih maščobnih kislin z eno dvojno vezjo, ter drugih snovi z različnim biološkim delovanjem, kot so tokoferoli, karotenoidi, fosfolipidi in fenoli. Te komponente prispevajo tudi k unikatnemu okusu in aromi oljčnega olja (Covas in sod., 2009).

Oljke so znane, kot eno izmed najbolj pogosto gojenih sadnih dreves na svetu. Kar 98 % pridelovalnih površin, 99 % gojenih dreves in 99 % proizvodnje oljk pripada državam Mediteranskega območja in območja Bližnjega Vzhoda. Po ocenah organizacije FAO (Food and Agriculture Organization) naj bi oljčni nasadi leta 2009 zavzemali kar 9.9 milijonov hektarov pridelovalnih površin. Največji proizvajalec je Španija z 2,500,000 ha, sledita ji Italija in Grčija. V obdobju 2008/2009 je bilo na svetu proizvedenih 2,9 milijonov ton oljčnega olja, od tega je 1/3 proizvodnje pokrivala Španija, 1/4 Italija in 1/5 Grčija. Na

podlagi trgovske vrednosti oljčno olje prispeva 15% delež svetovnega trgovanja z olji (Ghanbari in sod., 2012).

Slika 5: Pridobivanje oljčnega olja; muzej v Španiji (foto: D. Bandelj)
Figure 5: Olive oil extraction; museum in Spain (photo: D. Bandelj)

Med oljčnimi olji so najkakovostnejša in najbolj cenjena ekstra deviška oljčna olja. Ekstra deviška oljčna olja so olja, ki jih pridobimo iz plodov oljk z mehanskimi ali drugimi fizičnimi procesi ekstrakcije pod blagimi toplotnimi pogoji. Ti procesi ne smejo povzročiti v olju nobenih sprememb, dovoljeno je le pranje, dekantacija, centrifugacija in filtracija. Med ta olja ne uvrščamo olja, ki jih izločimo s topili ali procesi esterifikacije in vse mešanice z olji drugačne narave (Ghanbari in sod., 2012).

2.1.6 Biokemija oljčnega plodu in olja

V oljčnih plodovih je olje koncentrirano predvsem v perikarpu (96 % - 98 %). Podobno kot ostala rastlinska olja ga sestavljajo maščobe, ki so sposobne saponifikacije (umiljenja) in maščobe, ki saponifikacije niso sposobne. Prve sestavlja zmes različnih triacilgliceridov (98 %-99 %), ki so organske kemične snovi, estri alkohola glicerola z višjimi maščobnimi

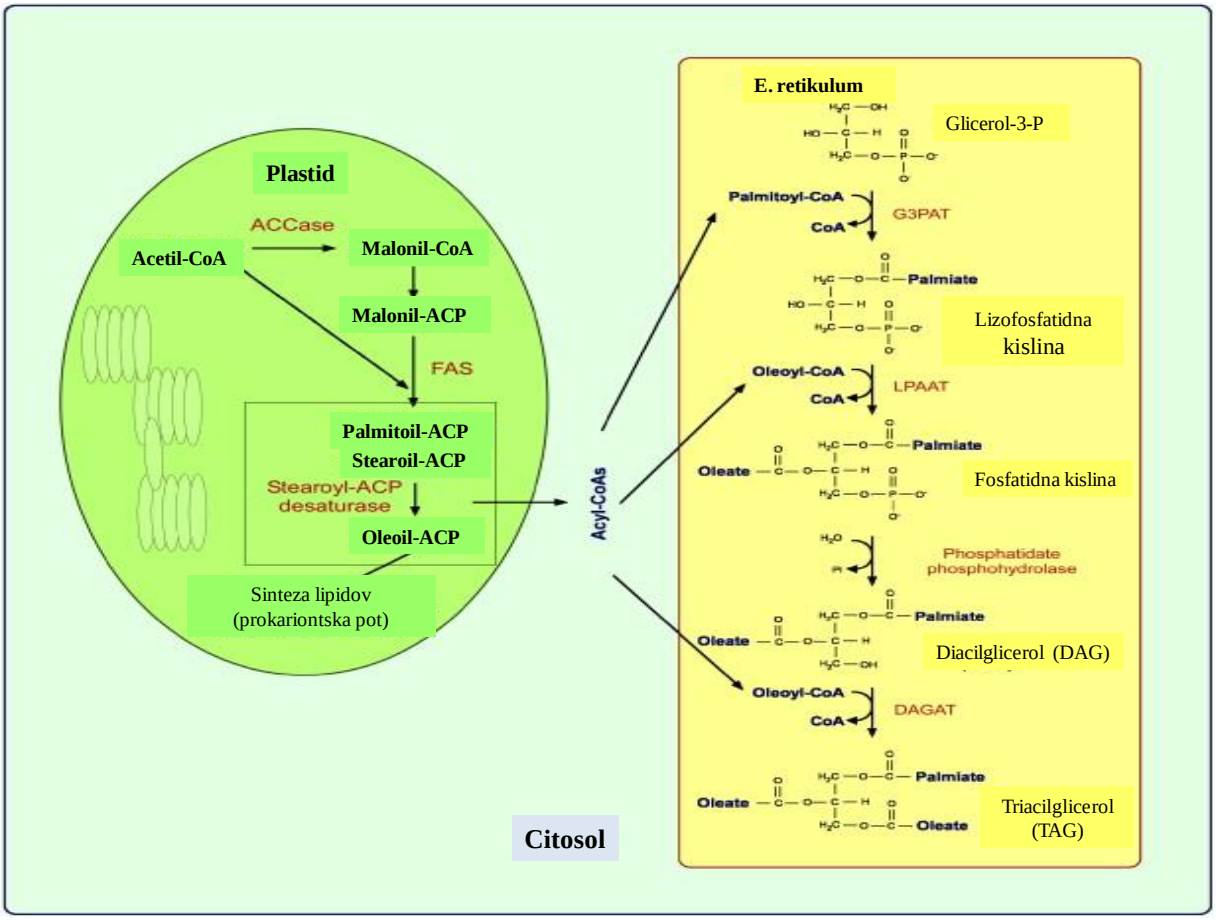
kislinami. Druge pa sestavljajo predvsem aromatične snovi (0,5 %-2 %), kot so ogljikovodiki, voski, steroli, alkoholi, vitamini, tokoferoli, biofenoli, itd.. Triacilgliceridi oljčnega olja vsebujejo predvsem enkratnenasičene maščobne kisline, manjšo količino nasičenih maščobnih kislin in večjo količino večkratnenasičenih maščobnih kislin (Aparicio in Aparicio-Ruiz, 2000). Mnoge zdravstvene študije so pokazale, da tradicionalna mediteranska prehrana, ki vključuje oljčno olje, kot osnovno prehrambno sestavino, dokazano zmanjšuje možnost nastanka srčnih obolenj in rakavih obolenj. Ugodne vplive oljčnega olja na človekovo zdravje pripisujejo predvsem visoki vsebnosti enkratnenasičenih maščobnih kislin v oljčnem olju (Gertz in Kochhar, 2001).

2.1.6.1 Maščobne kisline

Biosinteza maščobnih kislin (Slika 7) se odvija znotraj plastidov in je dobro poznana. Sinteza se prične s karboksilacijo acetil-CoA v malonil-CoA (Sanchez in Harwood, 2002). Reakcijo, ki poteka v dveh korakih katalizira acetil-CoA karboksilaza (heteromerni kompleks 4 podenot), ki vsebuje biotin prostetično skupino. Omenjen encim je v večji meri odgovoren za celotno biosintezo maščobnih kislin. Malonil skupina se prenese na acil transportni protein (ACP). Maščobna kislina se podaljša iz malonil-ACP in acetil-CoA preko reakcij, ki jih katalizira kompleks posameznih encimov (sintetaze maščobnih kislin - FAS). Pri oljki je bila aktivnost kompleksa FAS proučena na vodotopni frakciji pulpe iz razvijajočih se plodov z uporabo radioaktivno označenega malonil-CoA kot prekurzorja (Sanchez in Harwood, 1992). Re-esterificirani oleati in palmitati se transportirajo v citosol kot acil-CoAs (‘‘pot pri evkariontih’’) in v Kennedijevi poti služijo kot acetiltransferaze na endoplazmatskemu retikulumu z vlogo kopičenja triacilglicerolov (TAG). Sinteza TAG-ov iz glicerola-3-fosfata in tvorjenje maščobnih kislin v plastidih poteka po 4 serijskih reakcijah, ki jih katalizirajo 3 acetiltransferaze in specifičen encim fosfohidrolaza (Sanchez in Harwood, 2002). Informacije o genskih regulacijah metabolne poti maščobnih kislin pri oljkah so še vedno zelo omejene.

Določenih je nekaj genov, ki sodelujejo v metabolnih poteh maščobnih kislin. Poznan je gen, ki določa sintezo oleinske kisline pri oljki (stearoil-ACP Δ9-desaturaza), poznana pa je tudi celotna dolžina njegovega cDNA zaporedja. Ekspresijska analiza tega gena je pokazala, da je le ta reguliran z razvojem oljčnega plodu (Haralampidis in sod., 1998). Poghosyan in sod. (1999) so iz knjižnice oljčnih plodov izolirali cDNA, ki kodira plastidno ω-3 desaturazo, odgovorno za sintezo linolenskih maščobnih kislin. Poravnava z drugimi desaturaznimi zaporedij, je pokazala močno homologijo s plastidno ω-3 desaturazo *fad7*. Rezultati ekspresije domnevnega *fad7* gena, so bili v skladu z njegovimi znanimi vlogami, kot so tvorba tilakoidnih membran in zagotavljanju α-linolenskih signalnih molekul, ki so še posebej pomembne v rastlinskih tkivih, ki sodelujejo pri transportu in razmnoževanju. Prav tako so iz cDNA knjižnice oljčnih plodov izolirali enoil-ACP reduktazo, ki ima pomembno vlogo pri odstranitvi trans-nesaturiranih dvojnih vezi, za izgradnjo saturirane

acil-ACP (Poghosyan in sod., 2005). Banilas in sod. (2007) so iz oljke izolirali ER-tip ω-3 *FAD* gena. Ekspersijski vzorec *Oe*FAD3 v tkivih semen in mezokarpu oljčnih plodov iz različnih razvojnih faz, je pokazal minimalen prispevek tega gena v biosintezi in modofikacijah oljčnega olja, dokazana pa je bila njegova vloga pri razvoju ženskih gametofit v oljki. Banilas in sod. (2007) so iz oljke izolirali tudi cDNA dveh ω desaturaz, ki imajo ključno vlogo pri pretvorbi oleinske maščobne kisline v linolejno maščobno kislino. Southeren analiza je pokazala, da je bil *OeFAD2* gen prisoten v dveh kopijah, medtem ko je bil *OeFAD6* gen prisoten v eni kopiji. Ekspresijska analiza z metodo RT-PCR je pokazala, da se oba gena izražata v vseh tkivih oljke, vendar so višje stopnje akumulacije mRNA odkrili v reproduktivnih organih in celicah, ki se hitro razmnožujejo ali shranjujejo lipide. Rezultati ekspresije obeh genov so bili prav tako v skladu z njunimi predvidenimi vlogami pri tvorbi membran ob delitvi celic, pri tvorbi tilakoid, shranjevanju lipidov in proizvodnji signalnih molekule, ki vplivajo na razvoj rastlin ali obrambo rastlin (Banilas in sod., 2005). Diacilglicerol aciltransferaze (DGAT) katalizirajo zadnji korak pri sintezi triacilglicerolov (TAG). Banilas in sod. so iz cDNA knjižnice oljčnih plodov pridobili tip-2 diacilglicerol aciltransferaze (DGAT2). S primerjalno transkripcijsko analizo so primerjali izražanje DGAT2 in DGAT1 med razvojem oljčnih plodov. Rezultati so pokazali, da med razvojem mezokarpa sovpadajo izražanja različnih oljčnih diacilglicerol aciltransferaze, vidnejša pa je bila vpletenost DGAT2 v razvoj cvetnih brstov in zorenje plodov (Banilas in sod., 2011).



Slika 6: Poenostavljena biosintezna pot lipidov pri oljki prikazuje nastanek maščobnih kislin v plastid in nastanek triacilglicerola v endoplazmatskem retikulumu. Acil-ACP se proizvede v plastidu s pomočjo FAS compleksa (sintetaza) in se uporabi za plastidno produkcijo lipidov ali pa se prenese v citosol kot acil-CoA. Tu se vgradi preko Kenedijeve poti endoplazmatskega retikuluma v triacilglicerole (Conde in sod., 2009).
Figure 6: Simplified olive lipid biosynthesis pathway showing fatty acid formation within the plastid and triacylglycerol assembly within the endoplasmatic reticulum. The Acyl-ACPs produced in the plastid by the FAS complex are either used directly for plastidic lipid production or are exported to the cytosol as acyl-CoAs. The latter are used by the acyltransferases of the Kennedy pathway on the endoplasmatic reticulum for the synthesis of TGAs (Conde et al., 2009).

Glavne maščobne kisline, ki jih najdemo v oljčnem olju so palmitinska kislina (C16:0), palmitoleinska kislina (C16:1), stearinska kislina (C18:0), oleinska kislina (C18:1), linolna kislina (C18:2) in linolenska kislina (C18:3). Oljčna olja večine kultivarjev vsebujejo palmitinsko, stearinsko, oleinsko in linolno kislino, ostale maščobne kisline pa so prisotne v manjših deležih. Glavna komponenta oljčnih olj je vedno oleinska kislina, ki predstavlja kar 55-75% maščobnih kislin v oljčnem olju. Ta mononenasičena maščobna kislina zvišuje lipoproteine visoke gostote (HDL-dobri holesterol) in apoproteine A, ter zmanjšuje lipoproteine nizke gostote (LDL-slabi holestrol) in apoproteine B. S tem zmanjšuje možnost nastanka srčnih in žilnih obolenj, ki so najbolj razširjena bolezen razvitega sveta

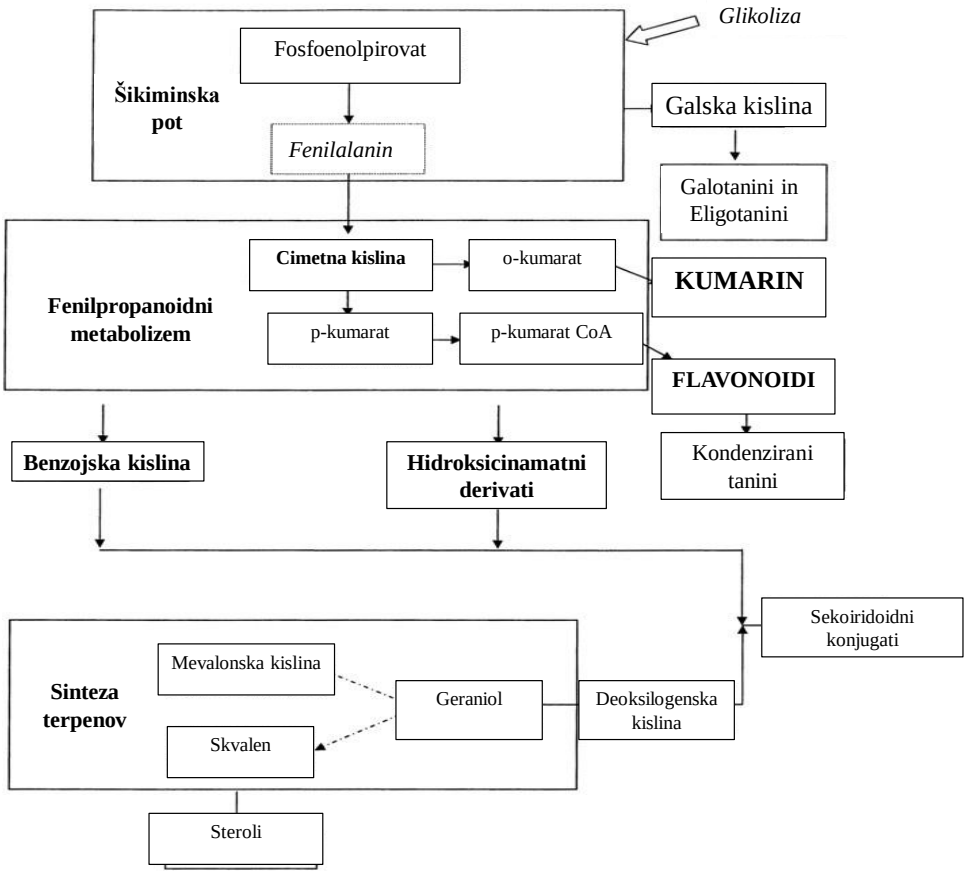
(Grundy, 1997). Linolna, ter druge večkratnenasičene maščobne kisline, so esencialne kisline. To pomeni, da jih organizem ni zmožen sam sintetizirati, zaradi česar je nujno, da jih uživamo s hrano. Pomen esencialnih nenasičenih maščobnih kislin je predvsem ta, da v človeškem organizmu uravnavajo pomembne biokemične reakcije (Viola in Viola, 2009). Relativno visoka vsebnost enkratnenasičenih maščobnih kislin, nižja vrednost nasičenih maščobnih kislin in precejšnja vsebnost večkratnenasičenih maščobnih kislin dajejo oljčnemu olju visok prehranski status.

2.1.6.2 Biofenoli

Biofenoli so sekundarni metaboliti, ki jih oljka tvori v času svoje rasti in dozorevanja plodov. Imajo potencialni antioksidativni vpliv, ki igra pomembno vlogo v kemični, organoleptični in prehrambni vlogi oljčnega olja. Biofenoli so lahko enostavne substituirane spojine z majhno molekulsko maso, ki imajo na aromatskem obroču vezano eno ali več hidroksilnih skupin, lahko pa so tudi kompleksnejše strukture vezane na monoterpene. Biofenole oljčnih olj (Slika 9) tvorijo fenolne kisline, fenolni alkoholi, hidroksi-izokromani, flavonoidi, lignani ter sekoiridoidi. Fenolne kisline so bile prva fenolna spojina, ki so jo našli v deviškem oljčnem olju in je poleg fenil-alkohola, hidroksi-izokrome in flavonoidov prisotna v oljčnem olju v manjših količinah. Prevladujejo pa fenolne spojine, kot so sekoiridoidi in lignani (Bendini in sod., 2007). Tako lignani kot flavonoidi so pogosti tudi v drugih živilih, medtem ko so sekoiridoidi značilnost oljčnih biofenolov. Sekoiridoidi so tesno povezani z iridoidi, ki so podskupina monoterpenov. Za sekoiridoide je značilna eksociklična dvojna vez na položaju 8,9- oziroma oleozid, značilen za rasline iz družine *Oleaceae* (Ryan in sod., 2002). Značilna oleozida sta olevropein, ki je ester elenolne kisline in 2-(3,4-dihidroksifenil) etanola (3,4-DPHEA) in ligstrozid, ki je ester elenolne kisline in 2-(4-hidroksifenil) etanola (*p*-HPEA). Lignani oljčnih olj so pinorezinol, acetoksipinorezinol in hidroksipinorezinol, flavonoida pa sta luteolin in apigenin. Pretvorbene oblike dveh glavnih sekoiridoidnih glukozidov oljčih plodov – ligstrozida in olevropeina dajejo oljčnim oljem značilno aromo in okus, kjer pri sorti ‘Istrska belica’ še posebej prevladujeta izrazita sadežnost in pikantnost (Bandelj Mavsar in sod., 2005). Ker sekoiridoidi niso topni v olju, ostane v oljčnem olju po mehanični ekstrakciji le majhen delež teh komponent (Servili in Montedoro, 2002). Kljub temu so sekoridoidi pomembna sestavina oljčnega olja, saj vplivajo na njegove zdravstvene in senzorične lastnosti. Sekoiridoidi v oljčnih plodovih imajo pomembno vlogo pri preprečevanju ateroskleroze in pri inhibiciji LDL peroksidacije, številne študije pa so pokazale, da so te spojine tudi dobra preventiva proti rakavim obolenjem in osteoporozi. Zlasti olevropein, hidroksitirosol in oleokantal so pokazali pozitivne učinke na zdravje ljudi.

Fenolne spojine se tvorijo preko šikiminske poti in fenilpropanoidnega metabolizma. V rastlinah je šikiminska pot odgovorna za tvorbo dveh aromatičnih kislin, in sicer

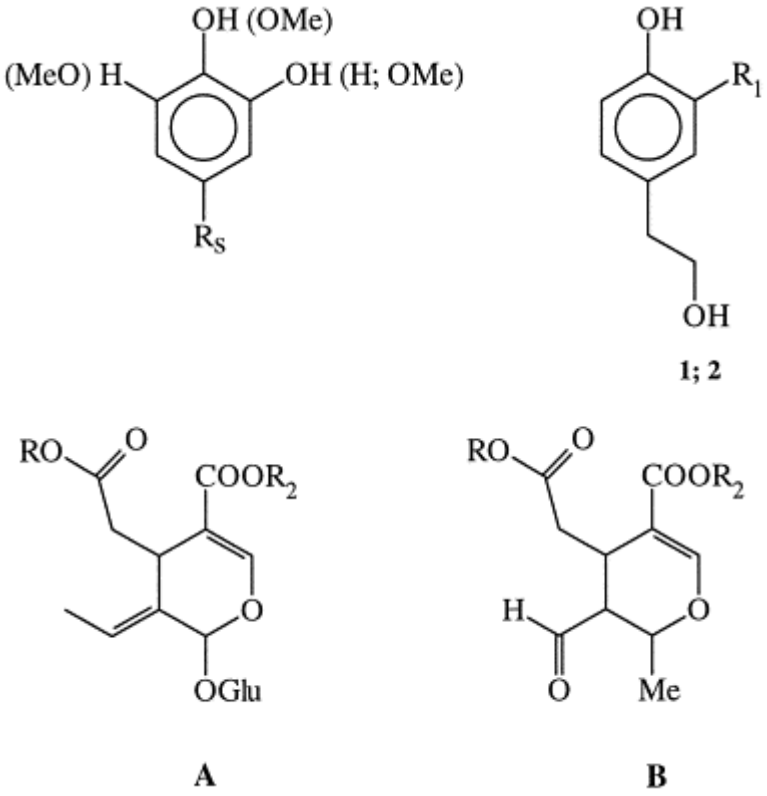
fenilalanina in tirozina. Ogljikovi hidrati so splošni vir ogljikovih atomov v metabolizmu organizma in zagotavljajo prekurzorje potrebne za sintezo sekundarnih metabolitov, kot so acetati, alifatske amino kisline in šikiminske kisline. Neoksidativna glikoliza glukoze proizvaja fosfoenolpiruvat in eritroza-4-fosfat, ki predstavljata začetne reagente za tvorbo šikiminskih kislin ali šikiminsko pot (Ryan in Robards, 1998). Do danes metabolizem skoiridoidov še ni povsem pojasnjen, vendar so za nekatere vrste *Oleaceae* že predlagali možne matabolne poti teh spojin. Sekoiridoidi so kumarinu podobne spojine, ki izhajajo iz iridoidov z odprtjem ciklopentanskega obroča. Iridoidi nastajajo znotraj sekundarnega metabolizma monoterpenov in imajo značilno ogrodje, v katerem je šestčlenski heterociklični obroč, pripojen k ciklopentanskem obroču. Odprtje tega obroča vodi do formacije sekoksiloganina, ki predstavlja matično spojino sekoiridoidov. V vrstah *Oleaceae* konjugati sekoiridoidov, kot je olevropein, vsebujejo fenolni del kot posledico esterefikacije, ki naj bi nastala z razvejanjem v poti mevalonske kisline v kateri se združita sinteza terpenov (oleozidni del) in metabolizem fenilpropanoida (fenolni del) (El Riachy in sod., 2011). Ta pot je shematsko prikazana na Sliki 8.



Slika 7: Shematski prikaz poti, ki predstavlja povezavo med fenilpropanoidnim metabolizmom in potjo mevalonske kisline (El Riachy in sod., 2011).
Figure 7: The coupling of phenylpropanoid and mevalonic acid pathways for the formation of secoiridoid conjucagates (El Riachy et al., 2011).

Fenolne spojine se skozi razvojne faze plodu, ter z mehansko obdelavo plodov spreminjajo tako kvalitativno kot tudi kvantitativno. Ligstrozid in olevropein, ki ju vsebujejo sveži plodovi, lahko zaradi poškodb ali pa pri sami predelavi vstopita v tri možne pretvorbeno-reakcijske poti. Prvi dve sta encimska ali kemijska pretvorba do aldehidne (zaprte) oblike olevropein (ligstrozid) aglikona, oziroma do hidroksi oblike olevropein (ligstrozid) aglikona. Tretja pot pa je že kar sam antioksidativen razpad, ko sekoiridoida že praktično ščitita oljke, oljčno drozgo ali predelano olje pred škodljivimi avtooksidativnimi spremembami. Pretvorba od aglikonov do nadaljnih kemijskih oblik – dialdehidne oblike olevropein aglikona in prevladujoče dialdehidne oblike dekarboksimetil olevropein aglikona, je postopna in spremlja pravilno predelano in primerno skladiščeno olje ves njegov življenjski vek. Identične pretvorbe veljajo za ligstrozid. Vse dokler sekoiridoidi ne zreagirajo do svojih končnih oblik – aromatiskih alkoholov tirosola in hidroksitirosola, so olja lahko senzorično bogata in skladna. Ko se pretvorba bliža koncu je vsebnost skupnih biofenolov še vedno relativno visoka, vendar je olje že pusto in ponavadi tudi antioksidativno šibko, saj v njem prevladuje tirosol, ki nima antioksidativnih značilnosti (Bandelj Mavsar in sod., 2005).

Vsebnost biofenolov je genetsko pogojena, vendar pa nanjo lahko vplivajo pedoklimatske razmere, agronomske tehnologije in zorjenje plodov. Več študij je pokazalo različno vsebnost biofenolov v oljčnih plodovih različnih italijanskih (Esti in sod., 1998) in španskih sort (Brenes in sod., 1999). Najbolj razširjena sorta v slovenskih oljčnikih 'Istrska belica' je poznana po visoki vsebnosti skupnih biofenolov, ki je lahko tudi za dvakrat večja v primerjavi s sorto Leccino (383 mg/kg vs 156 mg/kg, sončnično olje je v isti študiji imelo le 14 mg/kg skupnih polifenolov) (Butinar in sod., 1999). Skupni biofenoli pri sorti 'Istrska belica' so bili višji tudi pri primerjavi s štirimi drugimi italijanskimi sortami (Uccella, 2000). Za deviška oljčna olja pridelana iz sorte ''Istrska belica'' poročajo o vrednosti skupnih biofenolov tudi do 600 mg/kg (Bučar-Miklavčič in sod., 2006). Spektofotometrične meritve, predvsem meritve s HPLC, opravljene v zadnjih 5-7ih letih, so v vzorcih olj sorte Istrka belica pokazale značilno prevlado (visok relativni delež) različnih pretvorbenih oblik sekoiridoidov. Skupna vsebnost predvsem sekoiridoidnih biofenolov se lahko od letine do letine zelo spreminja. Vendar pa je značilnost dobrih olj, kot so olja iz sorte 'Istrska belica', da kjub morebitni nižji vsebnosti skupnih biofenolov delež glavnih sekoiridoidov ostaja v glavnem nespremenjen (Bandelj Mavsar in sod., 2005).



Slika 8: Glavne kemijske strukture oljčnih biofenolov: 1 – tirozol R1=H; 2 – hidroksitirozol R1=OH; A – olevropein R=1 R_2=Me, ligstrozid R=2 R_2=Me, oleozid R=H R_2=H; B – elenojska kislina R=H R_2=Me, tirozilelenolat R=2 R_2=Me, hidroksitirozilelenolat R=1 R_2=Me (Uccella, 2000)
Figure 8: Major chemical structures of olive biophenolics: 1- tyrosol R_1=H; 2 - hydroxytyrosol R_1=OH; A - oleuropein R=1 R_2=Me, ligstroside R=2 R_2=Me, oleoside R=H R_2=H; B – elenolic acid R=H R_2=Me, tirosilelenolate R=2 R_2=Me, hydroxytyrosilelenolate R=1 R_2=Me (Uccella, 2000)

2.1.6.3 Aromatične spojine

Proizvodnja aromatičnih spojin deviškega oljčnega olja poteka v rastlinskih organih z oksidacijo maščobnih kislin preko znotrajceličnih biogenetskih procesov. Določene aromatične spojine so prisotne že v nepoškodovanih plodovih, druge pa nastanejo ob pridobivanju oljčnega olja, ko se s stiskanjem plodov poruši struktura celic in pride do encimskih reakcij zaradi prisotnosti kisika. Na splošno velja, da endogeni rastlinski encimi, ki so del poti lipoksigenaze, vplivajo na pozitivno aromo oljčnega olja, medtem ko oksidacija in eksogeni encimi, ki ponavadi nastanejo zaradi mikrobnega delovanja, vplivajo na poslabšanje arome oljčnega olja (Angerosa in Basti, 2001; Kalua in sod., 2007).

2.2 DOLOČEVANJE NUKLEOTIDNEGA ZAPOREDJA DNA

2.2.1 Zgodovinski pregled

Genski kod je osnova biologije življenja, zato je poznavanje DNA zaporedij genoma, ter zapisa genov znotraj genoma v organizmu, postal temeljni vir bioloških informacij. Določevanje nukleotidnega zaporedja DNA je metoda pomnoževanja DNA, s katero določimo nukleotidno zaporedje našega DNA vzorca. Za to potrebujemo vzorec DNA, začetne oligonukleotide, DNA-polimerazo in označene nukleotide, s katerimi lahko določimo na novo nastalo zaporedje DNA. Začetki določevanja nukleotidnega zaporedja segajo v konec sedemdesetih let prejšnjega stoletja, ko sta se pojavili dve metodi določevanja zaporedja; Maxam-Gilbertova in Sangerjeva metoda. Maxam-Gilbertova metoda je temeljila na kemični modifikaciji DNA in rezanju DNA na specifično označenih mestih. Metoda je bila prvotno zelo priljubljena, kasneje pa je z izboljšavami povezanimi z možnostjo avtomatizacije postopka, pobudo prevzela Sangerjeva metoda, ki temelji na encimatskem vgrajevanju nukleotidov in njihovih analogov.

Frederick Sanger je v sedemdesetih letih dvajsetega stoletja razvil metodo, ki je bila v izpopolnjeni obliki v uporabi pri določitvi človeškega genoma, kot tudi pri določitvi ostalih živalskih in rastlinskih genomov (Hamilton in Buell, 2012). Pri določanju nukleotidnega zaporedja genoma s klasično Sangerjevo metodo genomsko DNA najprej razrežejo na krajše fragmente, ki jih nato vstavijo v plazmid in namnožijo v bakteriji *Escherichia coli*. Namnoženo DNA nato uporabijo v verižni reakciji pomnoževanja, ki se naključno zaustavi z vključitvijo fluorescentno označenega dideoksinukleotida. Končni produkt je mešanica različno dolgih verig DNA, ki so označene z ustreznim fluoroforom glede na končni nukleotid. Visoko-ločljivostno ločevanje označenih verig s kapilarno elektroforezo in določitev fluoroforov omogoča določitev nukleotidnega zaporedja.

Sangerjeva metoda in njene izboljšave so 30 let prevladovale na področju določevanja zaporedja DNA. Prvi prostoživeči organizem, ki so mu določili nukleotidno zaporedje genoma s pomočjo Sangerjeve metode, je bila bakterija *Haemophilus influenzae* (Fleischmann in sod., 1995). S tem so raziskovalci stopili korak naprej, saj so dokazali, da je določevanje zaporedja celotnega genoma »na enkrat« (WGS) izvedljivo. Čeprav je bila WGS (ang. whole genome shotgun) metoda izvedljiva pri manjših mikrobih, katerih velikost genoma zajema le nekaj megabaz (Mb), pa aplikacija te metode za evkariontske organizme ni bila možna, zaradi težav povezanih z nadaljnim zlaganjem zaporedij. V devetdesetih so prvi WGS projekti na evkariontih temeljili na uporabi knjižnic velikih insertov, kjer so dele genoma vključevali v umetne bakterijske kromosome (BAC) ali umetne kromosome kvasovk (YAC). Te so nato *in vivo* pomnožili in jim določili nukleotidno zaporedje vstavljenega DNA zaporedja, ki so ga nato zložili (Hamilton in Buell, 2012).

Arabidopsis thaliana (navadni repnjakovec) je bil zaradi majhnega genoma in njegove ustreznosti kot rastlinski modelni organizem, prva rastlina, ki so jo uporabili za *de novo* določitev nukleotidnega zaporedja genoma. Projekt se je začel izvajati leta 1996 na pobudo Arabidopsis genom iniciative, z uporabo pristopa od BAC-a do BAC-a (BAC-by-BAC) in z Sangerjevo metodo določevanjem nukleotidnega zaporedja Projekt se je zaključil z objavo v letu 2000 (Kaul in sod., 2000). Kmalu za tem (leta 2005) so z istim pristopom določili tudi genom riža (*Oryza sativa*) (Matsumoto in sod., 2005).

Izkupiček Sangerjeve platforme se je močno povečal z razvojem kapilarnih naprav za določevanje zaporedij, ki so omogočale hkratno določevanje zaporedja 96-ih reakcij. Z večjim izkupičkom zaporedij, izboljšanimi algoritmi za sestavo zaporedij in povečano računalniško močjo, je WGS sekvenciranje in sestavljanje celotnih evkariontskih genomov postalo izvedljivo. Ta metoda je odpravila drag in zamuden korak identifikacije in sestavljanja kozmidnih/BAC/YAC klonov v urejeno celoto (Hamilton and Buell, 2012).

Učinkovitost Sangerjeve WGS metode so leta 2000 potrdili pri vinski mušici (*Drosiphila melanogaster*), metoda pa je bila sprejeta tudi s strani rastlinske skupnosti (Adams in sod., 2000). Njena uporabnost je bila prikazana tudi pri človeku z določitvijo WGS zaporedja J. Craig Venter-ja (Levy in sod., 2007). V sledečih letih je bilo z omenjeno metodo določeno veliko število genomov, vendar noben od njih ni bil povsem dokončan, saj so bile zaključne faze pregledovanja zaporedij, določevanja neskladij in zlaganja sosesk zamudno, zahtevno in drago opravilo (Goff in sod., 2002; Ming in sod., 2008; Paterson in sod., 2009; Schmutz in sod., 2010; Tuskan in sod., 2006; Yu in sod., 2002). Do sedaj sta genoma *A. thaliana* in riža edina dokončana genoma dvokaličnic in enokaličnic.

2.2.2 Naslednje generacije določevanja nukleotidnih zaporedij

V zadnjih 25 letih se je z določevanjem zaporedij DNA popolnoma spremenil naš pogled na rastlinsko biologijo. Postala je mogoča določitev nukleotidnega zaporedja številnih genov in posledično tudi zaporedja ustreznih proteinov in njihovih funkcij. Informacije o genskih polimorfizmih so olajšale genetsko kartiranje, kloniranje genov in razumevanje evolucijsko pogojenih sorodstvenih vezi, ter omogočile začetek študija biotske raznolikosti (Delseny in sod., 2010).

Sangerjeva metoda je bila kar trideset let najbolj uporabljena metoda za določevanje nukleotidnih zaporedij. Nenehne potrebe po daljših odčitkih in hitrejšem določevanju zaporedja, so jo pripeljale do skoraj popolne optimizacije, vendar je kljub temu ostala razmeroma počasna in cenovno nedostopna za manjše raziskovalne skupine. Z napredki na področju mikrofluidov, nanotehnologije in informatike so se začele pojavljati nove, alternative metode sekvenciranja, ki so obetale še hitrejše in cenejše možnosti določevanja

zaporedja. Leta 2003 je J. Craig Venter Science Foundation obljubila 500.000 dolarjev nagrade tistemu, ki bo prvi določil zaporedja celotnega človeškega genoma za 1000 dolarjev, kar je spodbudilo tekmovanje med podjetji, za razvoj tehnike določevanja nukleotidnega zaporedja, ki bo to omogočala. Vse tehnike določevanja nukleotidnega zaporedja, ki so tako začele nastajati in imajo potencial za dosego tega cilja, poimenujemo s skupnim izrazom naslednje generacije sekvenciranja (NGS) (Delseny in sod., 2010).

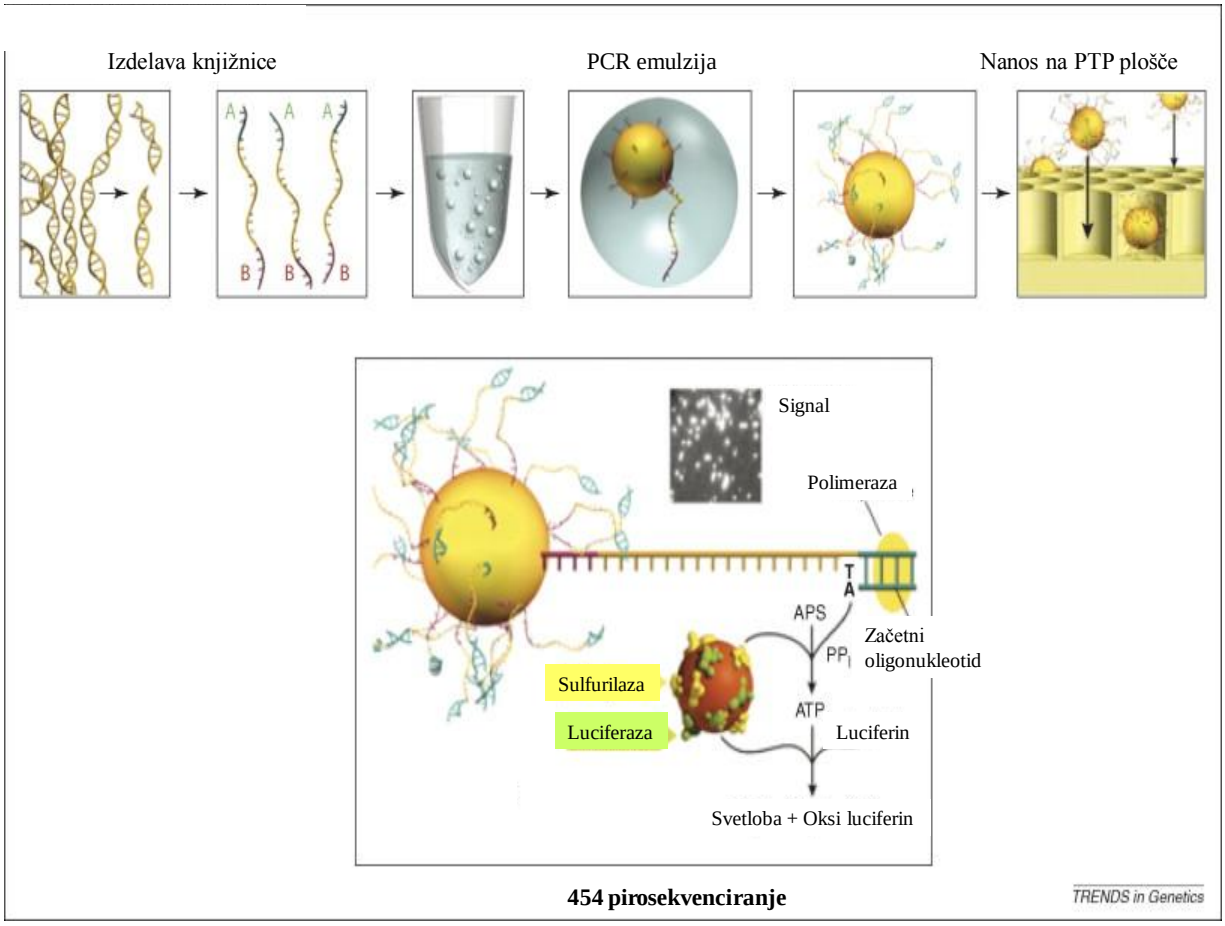
Komercialno dostopne in trenutno najbolj uporabljene med novimi tehnikami so Rochev 454 pirosekvenator, Illumina tehnologija sekvenciranja s sintezo in SOLiD platforma podjetja Aplied Biosystems, ki uporablja sekvenciranje z ligacijo. Vse zgoraj naštete tehnologije imenujemo tudi tehnologije druge generacije sekvenciranja, saj predstavljajo obdobje, ki je sledilo po določevanju zaporedij z Sangerjevo metodo (Hamilton and Buell, 2012). V zadnjih letih pa so se pričele razvijati platforme naslednje generacije, ki jih uvrščamo v tretjo generacijo določevanja zaporedij. Te metode omogočajo določevanje zaporedja ene same molekule DNA, brez predhodnega pomnoževanja (angl. Single molecule sequencing). Med aparature tretje generacije, ki so tudi komercialno dostopne, uvrščamo Pacific Biosciences PacBio RS tehnologijo. Medtem ko vse tehnologije do sedaj uporabljajo za določevanje zaporedja emitacijo svetlobe, pa se platforma Ion Torrent odlikuje po tem, da je to prva naprava, ki ne meri svetlobe (angl. post-light sequencer), ampak meri spremembo pH (Rothberg in sod., 2011).

Naslednje generacije sekvenciranja imajo več različnih potencialnih aplikacij na področju genomike, kot je razvoj genetskih markerjev, QTL (ang. *quantitative trait loci)* kartiranje, analiza ekspresije genov, populacijska in asociacijska genetika in ne nazadnje določevanje nukleotidnega zaporedja celotnih genomov. Vsaka metoda ima na določenem področju svoje prednosti in pa tudi pomanjkljivosti, zato se raziskovalci poslužujejo tehnik glede na način in zahteve svojega dela. Za določevanja zaporedij modelnih organizmov, pri katerih je dostopna velika količina genomskih informacij, sta najpogosteje uporabljeni Illumina in SOLiD tehnologiji sekvenciranja, saj imata kljub krajšim dolžinam (35-150 baz) večji izkupiček sekvenc, ki pa jih pri modelnih organizmih lahko kartiramo na referenčni genom ali transkriptom. Vendar pa je večina objavljenih projektov na nemodelnih organizmih za določevanje zaporedij uporabila Roche 454 platformo, saj omogoča pridobitev daljših zaporedij (danes že do 1000 bp s platformo FLX+), ki jih je lažje sestaviti in bolje anotirati (Hamilton in Buell, 2012).

2.2.3 454 pirosekvenciranje

Leta 2005 so Marguiles in sod. poročali o novi, alternativni metodi določevanja nukleotidnega zaporedja, poznani kot pirosekvenciranje (Margulies in sod., 2005). Platforma, ki je bila kreirana s strani družbe 454 Life Sciences in je danes last podjetja Roche, uporablja emulzijsko PCR reakcijo za namnoževanje fragmentov (Gupta, 2008).

Knjižnice DNA fragmentov se lahko konstruirajo s katerokoli metodo, ki producira kratke enoverižne fragmente, na katere so na koncu vezani adapterji. PCR (verižna reakcija s polimerazo) reakcija poteka v oljni raztopini s pomočjo nosilnih mikro kroglic (angl. beads) na katere se vežejo DNA fragmenti knjižnice, ki jo želimo pomnožiti. Ti fragmenti imajo na koncih vezane specifične adapterje, ki so komplementarni adapterjem vezanim na nosilne mikro kroglice. Ob vezavi fragmenta na kroglico in hkratni tvorbi emulzijske reakcije med oljem in vodo se omogoči tvorba vodnih kapljic v katerih so ujete kroglice s pripetimi DNA fragmenti. V teh nastalih mikro reaktorjih se prične pomnoževanje fragmenta ob dodatku začetnih oligonukleotidov, ki imajo zaporedja komplementarna adapterjem na koncu fragmentov. Nastala dvoverižna DNA se zatem denaturira in novonastala veriga se s pomočjo adapterja pripne na krogljici na novo adaptersko zaporedje in postopek se ponovi. Ob koncu amplifikacije vsaka krogljica nosi nekaj milijonov kopij unikatnega DNA fragmenta. Sledi prekinitev emulzijske reakcije, denaturacija DNA in nanos kroglic na pikotiterske (PTP) plošče. Plošče vsebujejo milijone majhnih luknjic, v katere se lahko usede le po ena krogljica. Te luknjice predstavljajo individualne reaktorje za določevanje nukleotidnega zaporedja. 454 tehnologija uporablja za določevanje nukleotidnega zaporedja tako imenovano pirosekvenciranje, ki temelji na kemolumniscenci. Ta nastane s pomočjo encimske kaskade po sprostitvi pirofosfata ob vezavi nukleotidov. Ko se nukleotid veže na nastajajočo DNA molekulo s pomočjo DNA-polimeraze, se sprosti molekula pirofosfata. Pirofosfat se nato s pomočjo encima sulfurilaza pretvori v molekulo ATP, ta pa se porabi za oksidacijo luciferina z luciferazo pri čemer nastane kemiluminiscentni signal. Na PTP ploščo se nanašajo posamezni nukleotidi, opazuje pa se svetlobni signal, ki nastane ob vključevanju nukleotidov. Sorazmerno s številom nukleotidov vključenih v zaporedje se veča moč signala. Po vključitvi nukleotida, aparat slika PTP ploščo, spere odvečne nukleotide in doda naslednjo bazo (Slika 9). Prve Roche 454 naprave so lahko določile zaporedja dolga do 110 bp, danes pa najnovejše nadgradnje naprave omogočajo branje dolžin tudi do 1000 bp (Diaz-Sanchez in sod., 2013). Glavna omejitev 454 sistema so homopolimerne regije DNA (zaporedja enakih baz, npr. $(C)_n$, $(A)_n$), saj lahko pride do napak pri odčitavanju intenzitete signala nukleotidov. Prednosti 454 sistema pa so njegovi relativno dolgi odčitki in hitrost postopka.



Slika 9: Shematski prikaz poteka 454 pirosekvenciranja (Bajang in sod., 2011).
Figure 9: Schematic representation of 454 pyrosequencing (Bajabng et al., 2011).

Z uporabo platform naslednjih generacij sekvenciranja lahko raziskujemo transkriptome organizmov globlje in širše. Nižji stroški sekvenciranja, učinkovitejša sestava podatkov in možnost določitve številčnosti transkripta so omogočili vpogled v različne rastlinske vrste. Kar zadeva rastline je bila Roche 454 tehnologija pogosto uporabljena pri določevanju zaporedij prepisov, saj daljša zaporedja dajejo več informacij in omogočajo enostavnejšo sestavo in obdelavo podatkov (Bajgain in sod., 2011; Troncoso-Ponce in sod., 2011). Uporabili so jo pri določevanju novih transkriptov dobro poznanega genoma *A. thaliana* (Weber in sod., 2007) kot tudi pri določevanju visoko izraženih transkriptov genoma kumare (*Cucumis sativus*) (Ando and Grumet, 2010). Pri pšenici so bile analize genoma dolgo časa otežene, zaradi njegove velikosti in poliploidnosti. 454 Roche tehnologija je omogočila sekvenciranje 17 giga bp velikega, heksaploidnega genoma pšenice (*Triticum Aestivum*) (Brenchley in sod, 2012). *Aegilops sharonensis* je diploidni divji sorodnik pšenice in predstavlja nekakšen rezervar genetske raznovrstnosti, ter bi lahko imel pomemben agronomski pomen predvsem kot vir novih odpornosti na bolezni. Da bi pridobili genetske podatke so Bouyioukos in sod. (2013) s pomočjo Roche 454 tehnologije posekvencirali cDNA knjižnico pridobljeno iz tkiva listov dveh geografsko ločenih akcesij.

Določevanje zaporedij mnogih rastlinskih vrst je izboljšalo tudi razumevanje vloge podvojenih regij in transpozonov znotraj genoma. Pri dveh divjih sorodnikih pšenice, *Aegilops cylindrica* in *A. geniculata*, so Senerchia in sod. (2013) s pomočjo Roche 454 tehnologije določili zaporedja od katerih je več 70% predstavljalo zapise za znane traspozone (transposable elements – TE), predvsem LTR transpozone. Roche 454 tehnologijo so uporabili tudi Alagna in sod. (2009), ko so s primerjalnim sekvenciranjem štirih različnih cDNA zbirk dveh olječnih genotipov pridobili informacije o spreminjanju genske ekspresije med razvojem olječnih plodov in med dvema genotipoma. Munoz-Merida in sod. (2013) pa so s kombinacijo določevanja zaporedij s Sangerjem in 454 pirosekvenciranjem določili najobsežnejše število zaporedij do sedaj, v želji da bi določili ESTs iz različnih tkiv v različnih razvojnih stopnjah oljke. Iz 2 M podatkov so pridobili 81020 posameznih genov.

2.2.4 Oznake izraženih nukleotidnih zaporedij

Čeprav hitrost določevanja zaporedij genomov v rastlinski biologiji zaostaja za tistimi pri sesalci in mikrobih, se je uporaba genomike razmahnila med pod-področja rastlinske znanosti kot so kmetijstvo, gozdarstvo, biokemija, genetika, vrtnarstvo, patologija in sistematska biologija. Večina genskih raziskav na kulturnih rastlinah je osredotočenih na razumevanje genskih mehanizmov in s tem na izboljšanje kakovosti in količine proizvodov. Z razvojem tehnologij sekvenciranja raziskovalci enostavneje pridobijo večje količine podatkov, ki nastanejo iz izraženih genov v določenih stadijih ali tkivih organizma (Ozgenturk in sod., 2010). Oznake izraženih nukleotidnih zaporedij (ang. Expressed sequence tags – ESTs), tako predstavljajo enega izmed načinov kako pridobiti informacije o zaporedjih preučevanih organizmov. ESTs so kratki, 200 do 800 baz dolgi, naključno izbrani prepisi sekvenc pridobljenih iz cDNA knjižnic določenih z eno sekvenčno reakcijo. Zaporedja informacijske RNA v celici predstavljajo kopije genov, ki se izražajo in nosijo pomembne informacije. Vendar molekul RNA ne moremo uporabiti za direktno kloniranje, zato jih je potrebno prepisati v dvoverižno cDNA z uporabo encima reverzna transkriptaza. Pridobljeno cDNA lahko kloniramo in tako zgradimo knjižnico genov, ki so izraženi v določeni celici ali tkivu organizma. Kasneje so cDNA kloni naključno prepisani iz 5' in 3' smeri, da dobimo kratka izražena nukleotidna zaporedja (ESTs) (Bonaldo in sod., 1996).
Uporabnost in aplikacija ESTs je bila prvič prikazana pri človeku, kjer so uspešno identificirali nove gene, ki se izražajo v možganih (Adams in sod., 1992). Delno določanje nukleotidnega zaporedja molekul cDNA (običajno enoverižnih molekul dolžine okoli 500 bp) ima več prednosti v analizi genoma, kot so npr. uporaba izraženih nukleotidnih zaporedij (ESTs) za kartiranje markerjev RFLP (ang. Restriction Fragment Length Polymorphism) (Inoue in sod., 1994), služijo kot sonde za pregled knjižnic YAC ali BAC pri konstrukciji fizičnih kart (Umehara in sod., 1995) ali pa služijo kot primer izraženih genov v točno določenem razvojnem obdobju, okolju ali pa po specifičnem tretiranju. Visoko kakovostne knjižnice cDNA so dobro orodje za odkrivanje pomembnejših genov

preko primerjave že objavljenih in okarakteriziranih nukleotidnih zaporedij DNA v sekvenčnih podatkovnih bazah (Newman in sod., 1994). ESTs služijo še kot vstopna točka za analizo izražanja genov, zaradi česar je ta sistem dodatno atraktiven. Ko ESTs dobimo iz različnih virov oz. reprezentativnih knjižnic cDNA, so le ti zanesljivi za oceno izražanja genov. Določanje zaporedij transkriptoma za nemodelne organizme je bolj priljubljeno, saj je cenovno ugodnejše in računalniško vodljivejše, kot sekvenciranje celotnega genoma. Vseeno pa še vedno zagotovi dovolj informacij za izpolnjevanje zahtev številnih raziskovalnih skupin.

Zgodnja devetdeseta so zaznamovala začetek razvoja genomike, saj so se pojavile prve avtomatizirane metode določevanja zaporedij, ki so temeljile na uporabi flourescentno označenih dideoksinukleotidov (Sangerjeva metoda). Ta tehnologija je omogočila prvo obsežnejšo določitev genov s pomočjo izraženih nukleotidnih zaporedij (ESTs). Mark Adams je prvi uporabil izraz EST v povezavi z določitvijo genov in projektom človeški genom leta 1991(Adams in sod., 1993). Enostransko določevanje zaporedij cDNA klonov je omogočilo odkritje genov iz točno določenih delov tkiva organizmov. Ta pristop je sprejel rastlinski biolog in ga prvi uporabil pri določevanju zaporedij EST-jev pri rastlinskem modelnem organizmu *Arabidopsis thaliana* (Newman in sod., 1994). Tradicionalno so projekti transkriptomov temeljili na Sangerjevi dideoksi metodi sekvenciranja, vendar so jo pričele izpodrivati metode nove generacije sekvenciranja, ki imajo znatno višjo zmogljivost in nižjo ceno na določitev baze. Po zaključku projektov na genomih različnih vrst organizmov, se je hitro povečalo število EST-jev in razvile so se podatkovne baze, ki so na voljo uporabnikom. Danes lahko najdemo v dbEST podatkovni bazi NCBI (National Center for Biotechnology) 2.3390.105 rastlinskih EST-jev, ki predstavljajo 33% celotne dbEST baze podatkov in zajemajo 733 rastlinskih vrst (http://www.ncbi.nlm.nih.gov/dbEST/). Z uporabo teh podatkov lahko sklepamo o funkcijah mnogih genov, glede na homologijo z znanimi geni (Ozgenturk in sod., 2010). Na voljo so številne objave, ki poročajo o razvoju knjižnic EST pri rastlinah. Pri jablani so Newcomb in sod. (2006) analizirali 150.000 EST iz 43 knjižnic z namenom, da bi ugotovili prisotnost potencialni molekulskih markerjev uporabnih za gensko kartiranje in ugotovili družine reprezentativnih proteinov in genov potencialno pomembnih za kakovost - (Newcomb in sod., 2006). Pri hmelju (*Humulus lupulus* L.) sta dve skupini razjasnili pomembno biokemijsko pot sekundarnih metabolitov z določitvijo nukleotidnega zaporedja 20.000 EST-jev. Tako je bilo identificiranih več encimov metiltransferaze ksantohumolne biosinteze (Nagel in sod., 2008) in več pogledov na razumevanje molekulske osnove za akumulacijo terpenov (Wang in sod., 2008; Nagel in sod., 2008; Wang in sod., 2008). Pri aktinidiji je bila razvita večja skupina EST-jev (132.577) za potrebe razvoja novih kultivarjev z novimi lastnostmi kot so okus, izgled, zdravilne substance. Na ta način so ugotovili in razjasnili delovanje genskih družin genov, ki določajo okus in vonj, ki so jih primerjali s kemijskimi analizami spojin. Določeni so bili tudi EST-ji povezani z zdravilnim učinkom in sicer sintezo askorbinske kisline in kininske

kisline. Identificirani so bili tudi geni odgovorni za mehčanje plodov v različnih fazah (Crowhurst in sod., 2008). Čeprav so bili pri oljkah razviti mnogi molekulski markerji, ni bilo narejenih veliko EST študij. Zadnji pregled opravljen v aprilu 2013 kaže, da je v dbEST podatkovni bazi pri NCBI shranjenih 10.379 EST-jev oljke (*Olea europaea*). Kar 3.734 EST-jev so objavili Ozgenturk in sod. (2010), ter prispevali najširšo EST kolekcijo oljke do sedaj. Predstavili so dve cDNA knjižnici iz mladih listov in nezrelih plodov oljke z željo, da bi odkrili nove gene in njihovo funkcijo v oljki. Prvo večjo kolekcijo ESTs pri oljki, ki pa je del Sequence Read Archive-a (SRA), so objavili Alagna in sod. (2009), kjer so predstavili 4 knjižnice z 261.485 zaporedji. S primerjalnim sekvenciranjem štirih različnih cDNA zbirk dveh olječnih genotipov so pridobili informacije o spreminjanju genske ekspresije med razvojem olječnih plodov in med dvema genotipoma, ki imata kontrastno akumulacijo fenolov v plodovih (Galla in sod., 2009). Zadnjo objavo na temo ESTs pri oljki so prispevali Bazakos in sod. (2012). Opazovali so tolerantnost oljke na stres, po izpostavitvi pogojem slanosti. Tako so primerjali odzive na ravni transkriptoma pri sorti tolerantni na sol in pri sorti, kjer se po izpostavitvi slanim pogojem stres izrazi. Objavili so 1,956 ESTjev s povprečno dolžino 381 bp. Secchi in sod. (2007) so naredili molekularno študijo vodnega transporta v oljki sorte 'Leccino', tako da so določili cDNA zaporedja sorodna z družino akvaporin (AQP) genov. Zaporedja cDNA oljke so uporabili tudi Kaya in sod. (2013) za določitev SNPs s pomočjo Illumina tehnologije. Prav tako so knjižnico cDNA uporabili Dundar in sod. (2013) za proučevanje alternativne rodnosti pri oljki sorte 'Ayvalik'.

2.3 OBDELAVA PODATKOV

2.3.1 Sestava nukleotidnih zaporedij

Nove tehnologije sekvenciranja so poenostavile tako strategije določevanja nukleotidnih zaporedij, znižale število napak, izredno povečale hitrost določevanja zaporedij genoma, kot tudi močno zmanjšale stroške sekvenciranja in s tem omogočile analizo zaporedij mnogih še neraziskanih organizmov (Delseny in sod., 2010). Kljub prednostim, ki nakazujejo na širšo uporabo NGS tehnologij v prihodnosti, pa so tu tudi pomanjkljivosti in omejitve, ki predstavljajo velik izziv izdelovalcem platform kot tudi izdelovalcem programske opreme za obdelavo surovih podatkov, pridobljenih z NGS. Zaradi kratkih dolžin zaporedij, ki jih pridobimo s tehnologijami NGS, ostaja še vedno največja težava sestava teh zaporedij v daljša zaporedja, ki predstavljajo soglasje vseh zloženih zaporedij (angl. consenzus). Mnoga podjetja zato poskušajo razviti nove programske opreme z izboljšanimi algoritmi, da bi dosegli učinkovitejšo obdelavo podatkov. Ti programi so prilagojeni posameznim NGS tehnologijam in dajejo pri različnih tehnologijah različno učinkovite rezultate. Mnogi novejši programi se dandanes že uspešno bojujejo s težavami, ki jih povzroča sestava in obdelava kratkih zaporedij, pridobljenimi s tehnologijami NGS. Vendar sestava krajših odčitkov zahteva večjo pokritost za uspešno določitev prekrivanja

zaporedij, kar pa odpira mnoga računalniška vprašanj v zvezi z obvladovanjem večjih zbirk podatkov (Delseny in sod., 2010; Imelfort in Edwards, 2009).
Pri sestavljanju nukleotidnih zaporedij se ustvarja hierarhična struktura podatkov iz katerih lahko nato pridobimo informacije, ki jih nosijo sama zaporedja. Nukleotidna zaporedja se združujejo v tako imenovane soseske (ang. contig), soseske pa naprej v tako imenovane supersoseske (supercontig). Pri nastajanju sosesk se najbolj podobna zaporedja poravnajo med seboj in tvorijo soglasje zaporedja, preostala zaporedja pa se glede na podobnost v naslednjih krogih poravnavajo s soglasjem zaporedja in tako ustvarjajo vedno nova soglasja zaporedja, dokler ne dobimo poravnave vseh vključenih zaporedij. Supersoseske določajo vrstni red in usmeritev sosesk, ter velikost vrzeli med soseskami. Najbolj pogosto uporabljen podatkovni format za sestavo zaporedij je FASTA format, kjer je zaporedje soseske predstavljeno v obliki niza znakov A, C, G, T, ter morebitnih drugih znakov s posebnim pomenom. Uspešnost združevanja nukleotidov merimo glede na velikost in pravilnost sestave kontigov in superkontigov. Velikost združkov je običajno podana s statistiko, ki vključuje maksimalno dolžino kontiga, povprečno dolžino kontiga, skupno dolžino kontigov in N50 vrednost (predstavlja točko polovice masne distribucije). Uspešnost združevanja zaporedij lahko primerjamo tudi z referenčnimi zaporedji, kadar obstajajo zaupanja vredne reference (Miller in sod., 2010).

Postopki združevanja zaporedij niso enostavni, saj lahko posamezni prepisi vsebujejo napake in polimorfizme, ti pa otežujejo prepoznavanje prekrivajočih se delov zaporedij. Pri ne-normaliziranih podatkih je lahko tudi številčnost posameznih prepisov različna, kar prav tako vpliva na uspešnost prekrivanja. Zato se v zadnjih letih, pri različnih NGS tehnologijah, znanstveniki pogosto poslužujejo sistematične primerjave različnih programov za združevanje zaporedij, z željo po določitvi optimalnih rezultatov (Delseny in sod., 2010).

2.3.2 Programi za združevanje nukleotidnih zaporedij

Prvi algoritmi, ki so se uporabljali pri programih za združevanje zaporedij, so bili razviti proti koncu sedemdesetih (Peltola in sod., 1984; Staden, 1979). Začetna prizadevanja so bila usmerjena predvsem v večkratno prileganje zaporedij, da bi pridobili končno postavitev združenih zaporedij, ter soglasje zaporedja iz katerega bi lahko sklepali na sestavo molekule DNA. To prvotno metodo so poimenovali »Overlap/Layout/Consensus« (OLC) metoda (Huang in Madan, 1999). V začetku devetdesetih pa so pri združevanju zaporedij začeli poudarjati predvsem formalizacijo, primerjavo in razvrščanje zaporedij. Sledila je objava dveh člankov, v katerih so zaporedja prepisov oz. dele prepisov vključili neposredno v graf (Idury in sod., 1995; Kececioglu in Myers, 1995). V grafu vsak vozel prestavlja zaporedje, dva vozla pa sta povezana med seboj z robom, kadar je med njima prisotno prekrivanje levega dela prvega zaporedja z desnim delom drugega zaporedja.

Končno zaporedje DNA so tako določili z pregledom robov grafa. Ta pristop je postal temelj za razvoj novih generacij za združevanja zaporedij DNA.

Programe za združevanje zaporedij danes delimo na različne kategorije, in sicer glede na obliko grafov, ki jih le ti uporabljajo. Metoda OLC se zanaša na uporabo prekrivajočih grafov (overlap graph), metoda De Bruijn Graph (DBG) pa temelji na uporabi k-mernih grafov. Graf je izraz, ki se pogosto uporablja v računalništvu. Predstavlja niz tako imenovanih vozlišč in robov, ki ta vozlišča povezujejo. Če robovi vodijo le v eni smeri, graf imenujemo usmerjeni graf. Pomembno je, da vsak usmerjen rob predstavlja povezavo iz enega tako imenovanega vira vozlišča do drugega tako imenovanega ponora vozlišča. Robovi se združujejo v poti, ki v razpršeni mreži vodijo do vozlišč, in sicer tako da ponor določenega vozlišča tvori izvorno vozlišče za vsa nadaljnja vozlišča. Posebna vrsta poti, ki se imenuje preprosta pot, pa je tista, ki vsebuje samo različna vozlišča (vsako vozlišče je obiskano največ enkrat) (Miller in sod., 2010).

OLC in DGB sta robustni metodi za združevanje zaporedij, ki se do neke mere zanašata na uspešnost prekrivanja med zaporedij. Ujemanje zaporedij je pri obeh metodah predstavljeno v obliki usmerjenega grafa, ki sta si pri obeh metodah podobna, če ne enakovredna (Myers, 2005). Metoda OLC neposredno združuje prekrivanja različno dolgih zaporedij, kot je tudi pričakovano pri podatkih, ki vsebujejo dolga zaporedja z nizko pokritostjo. De Brujin graf metoda pa je najbolj pogosto uporabljena pri sestavi večjega števila kratkih zaporedji, ki jih pridobimo s Solexa in SOLiD platformami. Temelji na uporabi k-mernih grafov, ki so omejeni predvsem na kratka prekrivanja (skupni k-mer) enotne velikosti. K-merni grafi ne potrebujejo vsak z vsakim poravnave parnih zaporedij za določitev prekrivajočih mest, ne shranjujejo posameznih zapisov in njihovih prekrivajočih mest, ter stisnejo redundantna zaporedja. OLC metoda vsebuje dejanska zaporedja in lahko porabi razpoložljivost pomnilnika na velikih genomih (Miller in sod., 2010).

Obe metodi se morata spopadati z motečimi podatki. Napake pri sekvenciranju lahko povzročijo lažno pozitivno ali lažno negativno ujemanje. K-merni grafi so bolj občutljivi za te napake, saj vsaka napačno določena baza vodi naprej do napačnega vozlišča.

Metoda OLC se je razvila ob uporabi dolgih zapisov, medtem ko se je uporaba metode DBG razširila ob uvedbi kratkih zapisov. Metodo OLC se tako uporablja predvsem pri sestavi daljših zapisov dolgih od 100 do 800 bp, medtem ko se metodo DBG uporablja pri sestavi krajših zapisov dolžine od 25 do 100 bp. V prihodnosti naj bi bili v uporabi predvsem srednje dolgi prepisi, za katere naj bi bila preferenčna uporaba OLC metode (Miller in sod., 2010).

Ne glede na to katero metodo uporabljajo programi za združevanje zaporedij, so vsem skupne določene osnovne funkcije:

- odkrivanje in popravljanje napak, ki temeljijo na sestavi zaporedja prepisa,
- oblikovanje grafa za predstavitev prepisov in njihovih ujemajočih se regij,
- zmanjšanje števila enostavnih poti v grafu,
- odprava poti, ki so nastale kot posledica napake,
- razpad kompleksov, ki so nastali kot posledica polimorfizma,
- poenostavitev pentelj z uporabo podatkov zunaj grafa,
- pretvorba pridobljenih poti v soseske in super soseske,
- redukcija poravnav s soglasjem zaporedja.

Čeprav nove generacije določevanja zaporedij prinašajo velik napredek, je tukaj še vedno veliko težav, ki jih je potrebno rešiti. Cena določevanja zaporedij je močno upadla, vendar sestava zaporedij ter anotacija zaporedij še vedno ostajata zapleteni, dolgotrajni in tudi dragi. Glavna želja razvijalcev programov za združevanje zaporedij je povečati kakovost samih sekvenc, kot tudi združevanja. Za to bodo potrebni boljši in novejši algoritmi, pridobitev daljših prepisov sekvenc in posledično nove strategije sekvenciranja. Hkrati si razvijalci programov želijo pridobiti več učinkovitih informacij iz posameznih sekvenc, ter tako omogočiti nižjo pokritost na genomu in s tem višji donos (Delseny in sod., 2010).

2.3.3 Anotacija zaporedij

Ko določimo nukleotidno zaporedje, sledi korak anotacije in interpretacije sekvence, pri kateri dodamo biološke informacije določenemu nukleotidnemu zaporedju. V prvi fazi anotacije, ki jo imenujemo tudi fizična anotacija, določimo število anotiranih genov, njihovo strukturo in natančne meje. Pomembna je tudi določitev ponavljajočih se zaporedij in različnih tipov transpozicijskih elementov. Uspešnost anotacije je v veliki meri odvisna tudi od kakovosti sekvence (Delseny in sod., 2010).

Prvi avtomatiziran korak anotacije se imenuje *ab initio* anotacija, pri katerem različni računalniški programi predvidijo gene, ki se nahajajo vzdolž sekvenc. Programi se morajo biti sposobni prilagoditi značilnostim danega genoma. Korak *ab initio* anotacije je zaključen s prikazom ujemanj v genomskih brskalnikih (angl. genom browser platforms). Ti brskalniki vključujejo in grafično prikazujejo različne vire informacij, ki jih pridobijo iz različnih podatkovnih baz, kot so ESTs, FL-cDNAs in Gene Othology podatkovne baze, ter iz informacij, ki jih nosijo transpozoni, proteinska zaporedja, strukturni motivi ali pa homologija z drugimi vrstami (Brent, 2008). Prvi računalniški programi, ki so jih razvili za anotacijo zaporedij, so prinesli precej zmede, saj so združevali sosednje gene, razdruževali samostojne gene, določevanje eksonov/intronov ni bilo natančno, različna transpozonska zaporedja pa se bila prepoznana kot geni (Delseny in sod., 2010). Dandanes imamo

postavljene različne podatkovne baze, ki vsebujejo ponavljajoča zaporedja in transpozicijske elemente rastlin, ter tako omogočajo bolj natančno anotacijo zaporedij (Chaparro in sod., 2007; Ouyang and Buell, 2004). Dostopna so tudi različna orodja, ki omogočajo anotacijo majnih RNA genov (Meyers in sod., 2008). Z razširitvijo določevanja nukleotidnega zaporedja genomov različnih organizmov, si bo pri anotaciji v prihodnosti mogoče pomagati tudi s primerjalnimi pristopi (Proost in sod., 2009).

Drugo fazo anotacije imenujemo funkcijska anotacija, pri kateri se posameznim genom določi njihove funkcije. Ta proces je prav tako mogoče avtomatizirati s programi za poravnavo sekvenc, ki določijo homologijo z že poznanimi geni iste ali druge vrste. Vendar funkcijska anotacija lahko le predvideva o funkcijah določenega gena in potrebuje za potrditev laboratorijski preizkus, ki pa pogosto ni izveden zaradi dolgotrajnega procesa (Delseny in sod., 2010).

Blast2go programski paket je eno izmed bioinformatskih orodij za funkcijsko anotacijo, ter analizo genov in proteinskih sekvenc. To orodje je bilo razvito z namenom, da ustvarijo uporabnikom prijazen vmesnik na osnovi Gene Otology anotacije. Z razvojem orodja je prišlo do izboljšav pri funkcionalnosti anotacije, orodje pa podpira tudi baze kot so Enzyme code (EC), KEGG Maps (Likic, 2006) in InterPro baza (Costanzo in sod., 2011). Blast2go uporablja lokalne ali oddaljene BLAST iskalnike, da poišče sekvence, ki so podobne naloženim sekvencam, ter jim pripiše vlogo na ravni bioloških procesov, celičnih komponent in molekularnih funkcij (Gotz in sod., 2011).

2.3.4 Verižna reakcija s polimerazo v realnem času (qPCR)

Verižna reakcija s polimerazo v realnem času (qPCR) temelji na metodi verižne reakcije s polimerazo (PCR), ki jo je razvil Kary Mullis v osemdesetih letih prejšnjega stoletja (Valasek and Repa, 2005). Pri tej metodi specifičen del DNA pomnožujemo z DNA polimerazo in specifičnimi začetnimi oligonukleotidi, ter tako lahko naredimo več kot milijardo kopij. PCR v realnem času prav tako kot navadno PCR reakcijo sestavljajo trije glavni koraki, ki vključujejo denaturacijo pri kateri pride do razklenitve dvoverižne DNA, prileganje pri katerem pride do vezave začetnega oligonukleotida na enoverižno DNA matrico, ter podaljševanje pri katerem DNA polimeraza sintetizira novo verigo vzdolž DNA matrice in nastane nova dvoverižna DNA. V vsakem ciklu naj bi tako imeli dvakrat več produkta kot v predhodnem ciklu, vendar se to dejansko ne zgodi, ker se po določenem številu ciklov med reakcijo reagenti porabijo in reakcija doseže plato. DNA se učinkovito podvaja samo do platoja, zato z metodo PCR v realnem času merimo produkt v eksponentni fazi, ko je pomnoževanje DNA še učinkovito. Meritve produkta so sorazmerne začetni količini DNA, zato PCR v realnem času omogoča kvantifikacijo. Metoda PCR v realnem času tako združuje podvojevanje DNA in detekcijo pomnoženih

produktov, zato za detekcijo produkta gelska elektroforeza ni več potrebna (Valasek in Repa, 2005).

V praksi metoda deluje tako, da kamera zazna svetlobo, ki je sproščena iz fluorokroma vezanega v novo sintetizirani PCR produkt. Interkalirajoča barvila so najbolj pogosta metoda za detekcijo pomnožene DNA. Absorbirajo svetlobo nižje valovne dolžine in oddajajo svetlobo višje valovne dolžine. Njihova flourescenca se močno poveča ob vezavi na dvovijačno DNA, zato količina dsDNA vpliva na intenziteto signala. Prvotno se je kot barvilo uporabljal etidijev bromid, kasneje pa ga je nadomestilo barvilo SYBR green, ki ima večjo afiniteto do dvoverižne DNA. Slabost fluorescentnih barvil je ta, da se vežejo nespecifično, torej se barvilo veže na dsDNA neodvisno od nukleotidnega zaporedja produkta, torej tudi na npr. dimere začetnih oligonukleotidov (Gachon in sod., 2004). Vendar specifičnost vezave lahko preverimo z analizo disociacijske krivulje pomnoženega produkta, določitvijo tališča produkta, ali z gelsko elektroforezo. Pri specifičnih metodah zanavanja pa imamo poleg dveh začetnih oligonukleotidov v reakciji tudi sondo, ki se komplementarno veže s tarčnim zaporedjem med oba začetna oligonukleotida. Sonde so specifične za določeno zaporedje in imajo 5' in 3' konce označene s flourescentnima barviloma. Na enem koncu ima sonda poročevalec (ang. reporter) in na drugem dušilec (ang. quencher). Ob vezavi sonde na tarčno zaporedje sta flourescentni barvili blizu skupaj in dušilec lahko absorbira poročevalski signal. Pojav imenujemo prenos fluorescentne resonančne energije (FRET, flourescent resonance energy transfer) (Valasek in Repa, 2005).

Začetni oligonukleotidi so pomemben parameter, ki lahko močno vpliva na uspešnost metode PCR v realnem času, še posebej ko uporabljamo interkalirajoča barvila kot metodo zaznavanja pomnoženih produktov. Prihaja lahko do napak, kot so komplementarna hibridizacija na 3' koncu znotraj samega zaporedja ali do hibridizacije med začetnimi oligonukleotidi. Na trgu je veliko plačljivih in neplačljivih računalniških programov, ki po svojih algoritmih izračunajo pomembne lastnosti začetnih oligonukleotidov in na ta način ocenijo njihovo ustreznost.

Fluorescenca reakcijske zmesi, ki se meri med samo reakcijo, nam omogoča prikaz namnoževanja produkta v realnem času. Graf pomnoževanja dobimo tako, da izrišemo krivuljo odvisnosti fluorescence od števila ciklov reakcije. Delimo ga na tri faze: bazna linija, eksponentna faza in plato. Za pridobivanje podatkov je primerna le eksponentna faza, v kateri flourescenca linearno narašča z namnoževanjem produkta. V analizah rezultatov, pridobljenih s PCR v realnem času, pozitivno reakcijo detektiramo s kopičenjem fluorescenčnega signala. Prag je količina signala, ki kaže na statistično značilno povečanje signala glede na signal bazne linije. Običajno ga inštrumenti samodejno nastavijo na vrednost, ki je desetkrat večja od standardne deviacije vrednosti flourescence v bazni liniji. Lahko pa je nastavljena ročno na poljubno vrednost. Pražni

cikel (ang. treshold cycle, Ct) je cikel, pri katerem fluorescenca preseže nastavljeni prag. Vrednost Ct je obratno sorazmerna začetni količini DNA (Caraguel in sod., 2011; Huggett in sod., 2005).

Pri metodi PCR v realnem času poznamo absolutni in relativni način kvantifikacije. Z absolutno kvantifikacijo določimo količino tarčne molekule v neznanem vzorcu s pomočjo umeritvene krivulje. Pri relativni kvantifikaciji pa določimo razliko v izražanju tarčnega gena v neznanem vzorcu, v primerjavi z izražanjem tarčnega gena v referenčnem vzorcu. Pri relativni kvantifikaciji ne potrebujemo umeritvene krivulje, vendar obstajajo različni matematični modeli, na podlagi katerih določimo količino tarčne molekule (Gachon in sod., 2004).

Pri rastlinskih študijah lahko metodo PCR v realnem času aplikativno uporabimo na dva načina: za detekcijo in kvantifikacijo tuje DNA (patogenih in simbiotskih mikroorganizmov rastlin, transgenih organizmov, GMO), ter pri študijah genske ekspresije (Gachon in sod., 2004). Pri oljki sorte 'Leccino' so Galla in sod. (2009) s pomočjo metode real time PCR preučevali ekspresijo 86 genov v treh različnih stadijih razvoja oljčnega plodu. Prav tako so Muzzalupo in sod. (2012) z metodo real time PCR preučevali ekspresijo LOX gena pri plodovih oljk vzorčenih v različnih razvojnih fazah, ter vpliv tega gena na prisotnost aromatičnih spojin v oljčnem olju. Alagna in sod. (2012) so v plodovih oljk merili koncentracije glavnih fenolnih spojin, kot so olevropein, demetilolevropein, 3–4 DHPEA-EDA, ligstrozid, tirozol, hidroksitrozol, verbaskozid in lignani. Prisotnost njihove mRNA so preučevali pri plodovih vzorčenih na dvanajstih sortah oljk v treh različnih fazah razvoja.

3 MATERIALI IN METODE

3.1 ZBIRANJE IN PRIPRAVA RAZISKOVALNEGA MATERIALA

3.1.1 Vzorčenje razvijajočih plodov oljk

Plodove oljk sorte 'Istrska belica' smo vzorčili skozi celotno obdobje razvoja plodov. Vzorčenje je potekalo od začetka junija do konca septembra. Na koncu smo imeli 22 časovnih točk vzorčenja skozi celotno obdobje razvoja plodu (Preglednica 1).

Preglednica 1: 22 časovnih točk vzorčenja plodov sorte 'Istrska belica', predstavljene so tudi koncentracije izoliranih vzorcev RNA in razmerja A260/A280
Table 1: Twenty-two sampling points of fruits variety 'Istrska belica', it also presents the concentration of isolated RNA samples and the A260/A280 ratio

Vzorec	Datum obiranja	Koncentracija (ng/µl)	A260/A280
1	09.06.2009	642	1,75
2	16.06.2009	581	1,81
3	24.06.2009	555	1,74
4	29.06.2009	680	1,74
5	06.07.2009	322	1,81
6	14.07.2009	965	1,82
7	21.07.2009	681	1,75
8	29.07.2009	639	1,8
9	05.08.2009	616	1,75
10	11.08.2009	485	1,74
11	18.08.2009	604	1,81
12	25.08.2009	358	1,72
13	31.08.2009	1031	1,86
14	14.09.2009	151	2,10
15	23.09.2009	183	2,08
16	31.09.2009	264	2,10
17	06.10.2009	194	2,10
18	14.10.2009	201	2,09
19	21.10.2009	165	2,08
20	29.10.2009	220	2,10
21	04.11.2009	158	2,07
22	23.11.2009	188	2,08

Takoj po obiranju smo plodove zmrznili v tekočem dušiku in jih shranili do uporabe pri -80 °C. Celokupno RNA smo izolirali iz vsakega vzorca posebej.

3.1.2 Izolacija RNA

Za izolacijo RNA smo prvotno uporabili metodo, ki uporablja TRIZOL ali TRI reagent, ki je bil originalno opisan v delu Chomzynski (1993). Vendar z uporabo te metode nismopridobili optimalnih rezultatov (degradirana RNA), zato smo za izolacijo RNA naknadno uporabili Spectrum Total Plant RNA Extraction Kit proizvajalca Sigma-Aldrich, ki se je izkazal za primernega.

Postopek izolacije je bil naslednji:

1. Vzorce smo s pomočjo tekočega dušika strli v terilnicah, ter 150 mg vzorca prenesli v 2-ml mikrocentrifugirko, ki je bila predhodno ohlajena na ledu.
2. V nadaljevanju smo uporabili raztopino za lizijo (ang. lysis solution), kateri smo pred uporabo dodali 2-merkaptoetanol (2-ME). Na 1 ml raztopine za lizijo smo dodali 10 µl 2-ME in dobro premešali. Nato smo 500 µl mešanice dodali 150 mg zdrobljenega vzorca in vse skupaj vrtinčili 30 sekund. Vzorce smo nato inkubirali v vodni kopeli na 56 °C 5 minut. Vzorce smo nato centrifugirali pri maksimalni hitrosti 3 minute, da so se celični ostanki usedli na dno.
3. Filtrirne kolone (ang. filtration column) smo vstavili v 2-ml zbirne tubice (ang. collection tube), ter odpipetirali supernatant po liziji v filtrirno kolono. Pri pipetiranju smo pazili, da nismo v kolono prenesli tudi delov celične usedline. Filtrirne kolone smo zaprli in jih centrifugirali 5 minut pri maksimalni hitrost, da smo se znebili preostalih celičnih ostankov. Lizat, ki je šel preko kolon, smo shranili v zbirnih tubicah.
4. Nato smo vzorcu dodali 750 µl raztopine za vezavo (ang. binding solution) in vse skupaj takoj premešali s pomočjo pipete. Kolone za vezavo (ang. binding tube) smo vstavili v 2-ml zbirne tubice, ter 700 µl vzorca nanesli na kolono. Kolone smo zaprli in jih centrifugirali 1 minuto, da se je RNA vezala na silicijevo membrano. Tekočino, ki je prišla čez kolone v zbirne tubice, smo odlili, ter postopek ponovili spreostankom vzorca.
5. Na kolone smo nato dodali 500 µl raztopine za izpiranje 1 (ang. wash solution 1) in jih centrifugirali pri maksimalni hitrosti 1 minuto. Tekočino, ki je prišla skozi kolone smo odstranili, kolone pa ponovno vrnili v zbirne tubice.
6. Nato smo na kolone dodali 500 µl raztopine za izpiranje 2 (ang. wash solution 2), ki smo ji predhodno dodali 100 % etanol, kakor so zahtevala navodila za pripravo kita. Kolone smo zaprli in jih zopet centrifugirali pri maksimalni hitrosti 30 sekund. Tekočino, ki je prišla čez kolone smo odstranili in še enkrat ponovili spiranje z raztopino za izpiranje 2.
7. Tekočino, ki je prišla skozi kolone smo zopet odstranili. Kolone smo ponovno centrifugirali 1 minuto, da smo jih popolnoma osušili. Nato smo osušene kolone prenesli v nove, čiste 2-ml zbirne tubice. Točno na sredino kolone smo odpipetirali 50 µl elucijske raztopine (ang. elution solution) zaprli kolone in počakali 1 minuto.

Nato smo kolone centrifugirali pri maksimalni hitrosti 1 minuto, da se je RNA izprala iz membrane in prešla skozi kolono v zbirno tubico.

8. Očiščeno RNA smo tako pridobili v 50 µl elucijske raztopine in jo shranili pri -80 °C do nadaljne uporabe.

3.1.3 Agarozna elektroforeza

Kvaliteto izoliranih RNA vzorcev smo redno pregledovali s pomočjo gelske elektroforeze na agaroznem gelu. Uporabljali smo SeaKem agarozne gele (BMA Products, ZDA), pripravljene v 1 x TBE pufru (44,5 mM Tris HCl, 44,5 mM borove kisline in 1mM EDTA, pH 8,0). Za preverjanje kakovosti izolirane RNA smo uporabljali 1,2-odstotni gel, ki smo ga za nadaljno vizualizacijo in analizo obarvali z 0,5 µg/ml (iz založne raztopine koncentracije 10 mg/ml) koncentracijo etidijevega bromida.

Elektroforetsko ločevanje vzorcev je potekalo v horizontalni elektroforetski napravi SubCell Model 192 (Bio-Rad, ZDA) pri konstantni napetosti 130 V proti anodi v 0,5 x TBE elektroforetskem pufru. V vzorce analizirane DNA smo pred nanašanjem v agarozni gel dodali nanašalno barvilo v razmerju 1:5 (12,5-odstoten (W/v) Ficoll tip 400, 0,2-odstoten (w/v) brom fenol modro). Za oceno dolžine fragmentov smo nanesenim vzorcem na gelu dodali še 100 ng DNA markerja (GeneRuler^{TM} 1 kb DNA Ladder, Fermentas, Litva). Po končani elektroforezi smo gel prenesli na transluminator TFM-30 (UVP Inc., Anglija), opremljen z virom dolgovalovne UV svetlobe (312 nm) in dobljene rezultate fotografirali z digitalnim fotoaparatom (Nikon CoolPix, Japonska).

3.1.4 Merjenje koncentracije RNA vzorcev

Koncentracijo vseh RNA vzorcev smo izmerili s pomočjo Nano drop UV spektrofotometra (Preglednica 1).

3.1.4 Merjenje koncentracije RNA vzorcev

Koncentracijo vseh RNA vzorčkov smo izmerili s pomočjo Nano drop UV spektrofotometra (Preglednica 1). Za dosego doslednih rezultatov smo pred začetkom meritve koncentracije RNA, izmerili slepi vzorec. Za slepi vzorec smo uporabili elucijsko raztopino (3.1.2 Izolacija RNA), v kateri je bila raztopljena izolirana RNA.

Postopek:

1. V Nano drop programu smo izbrali opcijo Nukleinske kisline (ang. Nucleic acid), tip RNA, ter pritisnili opcijo Dodaj v poročilo (ang. Add to report);

2. Izbrali smo opcijo Slepa (ang. Blank) , vstavili 2 µl elucijske raztopine na spodnji podstavek za merjenje, spustili ročico v »down« položaj in pritisnili gumb Merjenje (ang. Mesure), ki se je aktiviral ob izbiri opcije Slepa;
3. Po merjenju slepega vzorca, mora biti rezultat blizu 0, krivulja pa mora biti poravnana z bazno linijo;
4. Ko je bilo merjenje zaključeno smo dvignili ročico in obrisali merilno površino;
5. Sledila je analiza vzorcev (1,5 µl) po istem postopku, le da smo izpustili začetno opcijo Sepa in pritisnili le gumb Merjenje;
6. Po končanem merjenju smo izbrali opcijo Pričetek poročila/Snemanje (ang. Start report/Recording) in nato pritisnili gumb Shrani (ang. Save) in Natisni (ang. Print).

Na koncu smo zmešali ekvimolarne količine vseh vzorcev (3 µg RNA na vzorec), da smo dobili združen, reprezentativen vzorec vseh RNA izraženih v celotnem razvojnem obdobju olčnega plodu. Združenemu vzorcu smo določili koncentracijo DNA spektrofotometrično in preverili njegovo integriteto na agaroznem gelu.

3.1.5 Izdelava, obdelava in karakterizacija normalizirane cDNA knjižnice

Reprezentativen združen vzorec RNA razvijajočih plodov oljke smo uporabili za izdelavo normalizirane cDNA knjižnice. S postopkom normalizacije izravnamo nivoje transkriptov v posameznem vzorcu (Adams in sod., 2000; Liu, 2006). Ker je postopek zelo zahteven, smo se poslužili storitve izdelave normalizirane knjižnice v komercialnem servisu (Evrogen Lab, Rusija). Celokupna RNA je bila uporabljena za izdelavo ds cDNA s pomočjo SMART tehnologije (Shagin in sod., 2002; Zhu in sod., 2001; Zhulidov in sod., 2004). Za sintezo primarne cDNA verige so uporabili SMART Oligo II začetni oligonukleotid (5'-AAGCAGTGGTATCAACGCAGAGTACGCrGrGrG-3'), ter CDS-GSU začetni oligonukleotid (5'-AAGCAGTGGTATCAACGCAGAGTACCTGGAG-d(T)20-VN-3'), ki ima poleg poly-A repa vključeno prepoznavno zaporedje za *Gsu*I restrikcijski encim (prepoznavno zaporedje je podčrtano, cepi pa 16/14 nukleotidov desno od njega). Mešanica začetnih oligonukleotidov (5 µL) je bila segreta na 72 °C in nato ohlajena na ledu. Sintezo primarne verige cDNA so izvedli z dodatkom reverzne transkriptaze v mešanico začetnih oligonukleotidov in RNA v končnem volumnu 10 µL. Reakcija je bila inkubirana pri 42 °C in nato ohlajena na ledu. Za izdelavo sekundarne verige so uporabili SMART PCR primer (5'-AAGCAGTGGTATCAACGCAGAGT-3'), za amplifikacijo pa so uporabili Long-Distance PCR (Barnes, 1994).

50 µl PCR reakcije je vsebovalo:

- 1µl enoverižne cDNA
- 1 x Advantage reakcijski pufer (Clontech)
- 200µM dNTP
- 0,3 µM SMART PCR začetni oligonukleotid

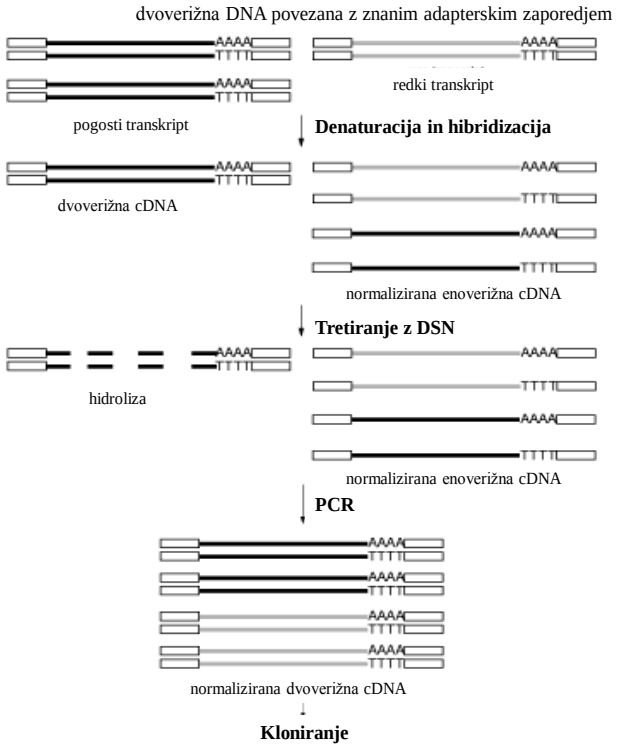
- 1 x Advantage polimerizacijska mešanica (Clontech)

18 PCR ciklov je bilo izvedenih na MJ Research PTC-2000 DNA Thermal Cycler napravi. Vsak cikel je imel naslednji temperaturni in časovni profil: 7 sek pri 95 °C ; 20 sek pri 65 °C; 3 min pri 72 °C.

Amplificirano cDNA so prečistili s PCR Purification Kompletom (QIAGEN, CA) po priporočilih proizvajalca. Očiščen produkt so koncentrirali s precipitacijo z absolutnim (96 %) etanolom. DNA pelet je bil raztopljen v miliQ vodi na končno cDNA koncentracijo 50 ng/µl. Tako pripravljeno, amplificirano cDNA so nato normalizirali z uporabo DSN normalizacijske metode (Zhulidov in sod., 2004). Normalizacija je vključevala cDNA hibridizacijo, tretiranje z dupleks-specifično nukleazo (DSN), ki jo pridobijo iz Kamčatka rakovic, (Shagin in sod., 2002) ter PCR pomnoževanje.

Postopek normalizacije je sledeč (Slika 10):

1. Hibridizacijska reakcija je vsebovala: 3 µl (približna 150 ng) raztopljene dvoverižne cDNA, 1µl 4 x hibridizacijskega pufra (200 mM HEPES-HCl, pH 8.0; 2 M NaCl). Reakcijska mešanica je bila prekrita s kapljico mineralnega olja in inkubirana pri 98 °C 3 min in pri 68 °C 5 ur.
2. Tretiranje z DSN je potekalo tako, da so hibridizacijski reakciji pri 68 °C dodali 3,5 µl miliQ vode, 1µl 5x DNAze pufra (500 mM Tris-HCl, pH 8.0; 50 mM $MgCl_2$, 10mM DTT) in 0,5 µl DSN encima. Inkubacija se je nadaljevala pri 67 °C za 20 min. Za zaključek DSN tretiranja so DSN encim inaktivirali tako, da so reakcijo segreli na 97 °C za 5 min.
3. cDNA vzorec so raztopili v 30 µl miliQ vode. Sledila je PCR reakcija (50 µl), ki je vsebovala 1 µl raztopljene normalizirane cDNA, 1 x Advantage 2 reakcijski pufer (Clontech), 200 µM dNTP, 0,3 µM SMART PCR začetni oligonukleotid, 1x Advantage 2 polimerizacijsko mešanico (Clontech). PCR reakcija je potekala v MJ Research PTC-200 DNA thermal Cycler napravi. Poteklo je 18 PCR ciklov, vsak cikel je imel naslednji temperaturni in časovni profil: 95 °C za 7 s; 65 °C za 20s; 72 °C za 3 min. 5 µg normalizirane cDNA smo pridobili po hitri pošti v obliki etanolnega precipitata.



Slika 10: Shematski prikaz poteka normalizacije cDNA knjižnice (Shagin in sod., 2002)
Figure 10: Shematic outline of DSN-normalization (Shagin et al., 2002)

Pridobljeno normalizirano cDNA knjižnico smo nato centrifugirali, odlili etanol in raztopili v TdE pufru [10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0], ji določili koncentracijo in jo nadalje okarakterizirali.

Najprej smo amplificirali del normalizirane knjižnice.

100 ul PCR reakcija je vsebovala:

- 35 ng normalizirane cDNA
- 10 x PCR pufer (10 µl)
- dNTP mešanico (10 µM)
- SMART začetni oligonukleotid (izdelani pri IDT-DNA, Belgija, 5'-AAGCAGTGGTATCAACGCAGAGT-3') (10 µM)
- Encim Dream Taq (3 U) polimeraza (Fermentas)
- Voda

PCR amplifikacijo pa smo izvedli z 8-imi ciklih pomnoževanja (94 °C 2min 1 cikel; 95 °C 7s, 65 °C 20s, 72 °C 3 min, skupaj 8 ciklov).

3.1.6 Kloniranje cDNA knjižnice

V želji, da bi knjižnico ohranili za daljše časovno obdobje, v primeru, da jo bomo kasneje še potrebovali, smo del knjižnice ligirali v pGEM-T Easy Vector (Promega, Velika Britanija), ki je uvaljavljen sisatem za uspešno kloniranje PCR produktov, za katere je znano, da pri pomnoževanju s *Taq* DNA polimerazo na 3´- koncu vsebujejo dodaten A. Vektor pripravijo na tak način, da lineariziranemu vektorju dodajo 3´- terminalni timin na oba konca.

Postopek:

1. V mikrocentrifugirke smo odpipetirali 3 µl pomnožene normalizirane cDNA.
2. PCR produktu smo dodali ligacijsko mešanico, ki je vsebovala 5 µl 2x hitrega ligacijskega pufra, 1 µl vektorja pGEM-T easy in 1 µl T4 DNA ligaze.
3. Vzorce smo inkubirali čez noč na 4 °C.

3.1.7 Transformacija kompetentnih celic

V kompetentne bakterijske celice *E. coli* genotipa XL-10 Gold smo s toplotnim šokom vnesli plazmidno DNA, čemur je sledila belo-modra selekcija uspešno transformiranih kolonij na trdnem LB gojišču v prisotnosti antibiotika karbenicilina.

Postopek:

1. Pripravili smo petrijeve posodice s trdnim LB gojiščem (1-odstotni (w/v) NaCl, 1-odstotni (w/v) tripton, 0,5-odstotni (w/v) kvasni ekstrakt, 2-odstotni (w/v) Difco agar (Becton Dickinson), ki smo mu dodali selekcijski antibiotik karbenicilin (150 mg/l), induktor promotorja gena lacZ IPTG (Duchefa) (0,2 mM iz založne raztopine s koncentracijo 0,1 M) ter X-gal (Duchefa) (40 µg/ml, iz založne raztopine s koncentracijo 20 mg/ml v dimetil formamidu) za modro-belo selekcijo.
2. Mikrocentrifugirke smo ohladili na ledu ter v vsako odpipetirali 100 µl kompetentnih celic bakterije *E. coli*.
3. V vsako mikrocentrifugirko s kompetentnimi celicami smo dodali 2 µl ligacijske reakcije z normalizirano cDNA, rahlo premešali in 20 min inkubirali na ledu.
4. Vzorčke smo izpostavili kratkemu (45-50 sek) temperaturnemu šoku (45 °C) ter jih takoj za 3 min prenesli na led.
5. Ligacijsko transformirani mešanici smo dodali 800 µl tekočega LB medija brez selekcijskega antibiotika (kot zgoraj, brez agaraja), ter 1,5 ure inkubirali pri 37 °C z mešanjem, da se je gen za odpornost na ampicilin, kodiran na plazmidu, uspešno izrazil v bakterijah, ki so sprejele plazmid.
6. Transformirano kulturo smo v štirih redčitvah (10, 25, 50, 100 µl) nanesli na LB gojišča ter jih preko noči inkubirali pri 37 °C.

7. Na osnovi belo-modre selekcije in potrjene uspešne transformacije smo celotno ligiranocDNA knjižnico razmazali po trdnem LB gojišču. Naslednji dan smo na plošče dodali 5 ml tekoćčega LB medija, rahlo stresali 30 min in inokulatprenesli v 0,5 l tekočega LB gojišča z dodatkom slekcijskega antibiotika Tekoče gojišče smo 8 ur inkubirali pri 37 °C in 130 obr/min.
8. Delu bakterijske kulture smo dodali glicerol do koncentracije 20% in zamrznili pri -80 °C. Preostanek tekočega gojišča z bakterijami smo centrifugirali in shranili usedlino bakterij za izolacijo plazmidov.

3.1.8 Izolacija plazmidne DNA

Plazmidno DNA smo izolirali z uporabo Gen*Catch*^{TM} Plasmid DNA Miniprep kompleta (Epoch Biolabs, ZDA), ki zagotavlja hitro in preprosto metodo čiščenja plazmidne DNA po principu vezave DNA na membrane na osnovi silicija.

Postopek:

1. V usedlino bakterij smo dodali 17,2 ml MX1 pufra za resuspendacijo usedline. Raztopino smo razdelili na tubice po 200 µl.
2. Z dodatkom 250 µl MX2 pufra, rahlemu mešanju in 5 min inkubaciji na sobni temperaturi smo izvedli lizijo celic.
3. Sledila je nevtralizacija z dodajanjem 350 µl MX3 pufra ter nežnim mešanjem. V tej fazi se je tvorila oborina.
4. Po 10 min centrifugiranju pri maksimalnem številu obratov smo supernatant previdno prenesli v pripravljene kolone ter s 30 sek centrifugiranja pri 5.000 onr./min supernatant precedili skozi membrano, ki je vezala DNA.
5. Kolono smo spirali s 500 µl WF pufra in 30 sek centrifugirali pri 5000 obr./min.
6. Sledilo je ponovno spiranje kolone s 750 µl WS pufra in ponovno centrifugiranje.
7. Po zadjem centrifugiranju (1 min, 12000 obr./min) smo kolono prenesli v novo 1,5 ml mikrocentrifugirko in na sredino membrane dodali 75 µl elucijskega pufra, centrifugirali ter ulovljeno plazmidno DNA shranili na -20 °C.

3.1.9 Odstranitev poli A regij

Preostanek cDNA knjižnice smo uporabili za nadaljnje določevanje nukleotidnega zaporedja s pomočjo novih tehnologij sekvenciranja. Izbrali smo tehnologijo Roche 454 FLX, s katero lahko pridobimo do 500 bp dolge odčitke in je bila v času izbora tehnologija, ki ponujala najdaljša možna branja. Problem pa je, da prisotnost homopolimernih odsekov, kot so poli A/T repi, povzroča težave pri določevanju nukleotidnih zaporedij v cDNA knjižnicah tako pri Sangerjevi tehnologiji kot pri tehnologiji 454 (Shendure and Ji, 2008).

Zato smo pred samim sekvenciranjem z Roche 454 tehnologijo normalizirano cDNA obdelali z *Gsu*I restrikcijskim encimom, ki cepi dvoverižno cDNA 14/16 bp stran od prepoznavnega mesta, ki smo ga vnesli z adapterjem (Slika 11). Na ta način odstranimo večino homopolimernega A/T dela cDNA. cDNA knjižnico smo tretirali z *Gsu*I restrikcijskim encimom preko noči pri 30 °C.

Sestava 70 µl reakcijske mešanice je bila naslednja:
- 50 µl DNA (6750 ng)
- 3,0 µl encima *Gsu*I (15 U)
- 7,0 µL 10x reakcijskega pufra B (10x Buffer B)
- 10 µL vode

Restrikcijska reakcija je bila najprej očiščena s kloroform:izoamilalkoholom (24:1) in nato s pomočjo kita za čiščenje PCR produktov, ki odstrani začetne oligonukleotide s pomočjo silicijevih kolon (Gene JET PCR Purification Kit, Fermentas). Na ta način smo odstranili dele cDNA, ki jih je odstranil restrikcijski encim.

Po odstranitvi smo cDNA zaporedjem dodali adapterje z vključenim SMART zaporedjem (5- AAGCAGTGGTATCAACGCAGAGTCGCATT, 3-CTTCGTCACCATAGTTGCGTCTCAGCGT). Adapterje smo naročili na strani IDT (Belgija), poleg SMART zaporedja pa so imeli vključeno tudi značko ACGC na 5' koncu, za določitev konca zaporedja.

Reakcija (38 µl) je vsebovala:
- 4,0 µl 10 x ligacijski pufer
- 1,0 µl adapter
- 0,3 µl ligaza
- 32,7 µl voda

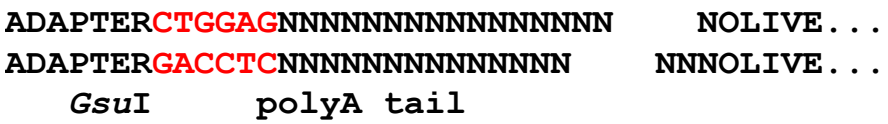
Reakcija je potekala pri 37 °C, 3 do 4 ure.

Nato smo knjižnico tretirano z *Gsu*I pomnožili s pomočjo SMART PCR začetnega oligonukleotida (5'-AAGCAGTGGTATCAACGCAGAGT-3').

Reakcija (25 µl) je vsebovala:
- 5,0 µl vzorec
- 2,5 µl 10 x PCR pufer
- 2,0 µl dNTP
- 0,75 µl SMART začetni oligonukleotid
- 0,15 µl Dream *Taq* polimeraza encim

- 14,5 µl voda

PCR amplifikacijo smo izvedli z 18-imi ciklih pomnoževanja (94 °C 2 min 1 cikel; 95 °C 7s, 65 °C 20 s, 72 °C 3 min ter jo očistili z Gene JET PCR Purification Kit, Fermentas.



Slika 11: Prepoznavno mesto za *Gsu*I restrikcijski encim.
Figure11: Recognition site for the *Gsu*I restriction enzyme.

Kvaliteto ne-normalizirane cDNA, normalizirane cDNA, cDNA po restrikciji, ter cDNA knjižnice po restrikciji in čiščenju, smo določili tudi s pomočjo naprave Agilent Bioanalyzer 2100 in uporabo čipa DNA1000. Storitev so nam izvedli na Centru za funkcijsko genomiko in bio-čipe (CFGBC, Ljubljana).

3.2 PRIMERJALNO DOLOČEVANJE NUKLEOTIDNEGA ZAPOREDJA

Sledilo je primerjalno določevanje nukleotidnega zaporedja za preizkus uspešnosti izreza poli-A regije. Del knjižnice, ki smo jo tretirali z *Gsu*I encimom in del knjižnice, ki ni bila tretirana, smo vključili v pGEM-T Easy vector (po postopku, opisanem v poglavju 3.1.6 Kloniranje PCR produktov), namnožili kolonije bakterij na LB ploščah (pripravljene po postopku, opisanem v Poglavju 3.1.7 Transformacija kompetentnih celic).

3.2.1 Direktni PCR vstavljene cDNA

Sto naključno izbranih kolonij obeh knjižnic smo pomnožili vstavljeno cDNA s postopkom direktnega pomnoževanja bakterij. Direktni PCR se uporablja, da preverimo, v katerih kolonijah se je uspešno vgradil insert, ter kakšne dolžine je, istočasno pa lahko tudi služi za neposredno določevanje nukleotidnega zaporedja. S tem se izognemo prestavljanju kolonij v tekoče gojišče in izolaciji plazmidov.

Postopek:

1. V PCR reakcijske posodice odpipetiramo 80 µl TdE pufra.
2. Z najmanjšim pipetnim nastavkom prenesemo del bakterijske kolonije v TdE pufer. Premešamo s pipeto.
3. Inkubiramo 8 min pri 96 °C, da celice počijo in izločijo plazmid.
4. Centrifugiramo, da se deli liziranih celic usedejo na dno.
5. Tako dobljeno DNA uporabimo v PCR reakciji.

15 µl PCR reakcije je vseboval:
- 5,0 µl DNA
- 2,0 µl 10x PCR pufer
- 1,6 µl dNTP
- 1,0 µl SP6 začetni oligonukleotid (10 µM)
- 1,0 µl T7 začetni oligonukleotid (10 µM)
- 0,1 µl *Taq* DNA polimeraza
- 9,3 µl voda

Po končanem PCR reakciji smo namnožene vzorce pregledali na 1 % agaroznem gelu (priprava po postopku iz poglavja 3.1.3 Agarozna elektroforeza). Preostanke PCR reakcije, pri katerih smo na gelu videli pomnoženi insert, smo očistili s pomočjo ExoSAP kompleta.

7 µl reakcijske mešanice je vsebovalo:
- 0,1 µl Exo1
- 0,5 µl FastAP
- 1,4 µl 1x PCR pufer
- 5,0 µl PCR vzorec

Reakcijsko mešanico smo inkubirali v napravi za PCR z naslednjim temperaturnim profilom: 37 °C 45 min, 80 °C 15 min, 12 °C ∞.

3.2.2 Določevanje nukleotidnega zaporedja po Sangerju

Sledilo je določevanje nukleotidnega zaporedja z uveljavljeno Sangerjevo verižno dideoksi metodo (Sanger in sod., 1977), ki temelji na cikličnem sekvenciranju s termostabilno polimerazo in fluorescentno obarvanimi dideoksi terminatorji. Uporabili smo komercialni komplet BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, ZDA).

10 µl PCR reakcije je vsebovalo:
- 2,0 µl 5x BigDye pufra
- 0,2 µl (10 µM) za vektor pGEM-Teasy specifični začetni oligonukleotid T7 ali SP6
- 0,5 µl sekvenčne mešanice BigDye 3.1
- 3,8 µl vode
- 3,5 µl vzorca DNA

Sekvenčna reakcija je potekala po sledečem temperaturnem profilu: 3 min 96 °C, nato 50 ciklov 10 sek 96 °C, 10 sek 50 °C in 4 min 60 °C. Sledila je končna inkubacija pri 72 °C za 7 min, nakar smo vzorce ohladili na 12 °C in jih shranili pri -20 °C.

3.2.3 Čiščenje produktov reakcije določevanja nukleotidnega zaporedja

Postopek:

1. Kratko smo centrifugirali vzorce.
2. Dodali smo 2,5 µl 125 mM EDTA pH 8.0 in nato 30 µl abso-lutnega etanola vsaki reakciji. Kratko smo centrifugirali, da EDTA in etanol prideta v stik z vzorcem.
3. Celotni vzorec smo odpipetirali v 96-mestno PCR ploščo.
4. Ploščo smo prekrili s folijo in premešali (5-10x obrni).
5. Inkubirali smo 15 min na sobni temperaturi, zaščiteno pred svetlobo.
6. Ploščo smo centrifugirali pri maksimalni hitrosti 55 min in 4°C, pazili smo na pravilno protiutež (rotor za mikrotiterske plošče).
7. Po centrifugiranju smo izlili etanol s hitrim gibom navzdol.
8. Ploščo smo centrifugirali obrnejno navzdol pri 190x g, na papirnati brisači, 2 min.
9. Ploščo smo inkubirali na sobni temperaturi 5 min, zaščiteno pred svetlobo.
10. Raztopili smo DNA (ne vidiš nič) v 12 ul formamida, ter prelepili s folijo. Vzorci so tako pripravljeni za nadaljno analizo na napravi ABI 3130XL. Pripravili smo tabelo z imeni v Excelu.

3.2.4 Obdelava rezultatov sekvenciranja

Sekvenciranje so izvedli na napravi ABI 3130XL na Odelku za zootehniko Biotehniške fakultete v Ljubljani. Rezultate smo prejeli v obliki kromatogramskih datotek tipa ab1.

Rezultate sekvenciranja smo obdelali z uporabo računalniškega programa CodonCode Aligner verzije 2.0.6 (CodonCode Corporation, Massachusetts, ZDA), namenjenega naprednemu urejanju in zlaganju zaporedij.

Postopek:

1. V orodni vrstici aplikacije CodonCode Alligner smo odprli meni »Datoteka« (ang. File) ter izbrali »Nov projekt« (angl. New Project). Zaporedja smo pridobili v ABI formatu, ter jih z ukazom »Uvozi podatke« (angl. Import File) prenesli v pogovorno okno projekta, kjer so bili prikazani osnovni podatki o naših vzorcih: število vzorcev, ter njihova dolžina in kvaliteta. Na osnovi teh kriterijev smo izbrali vsa nekvalitetna zaporedja (dolžina nekvalitetnega dela je predstavljala večino zaporedja, dolžina kvalitetnega zaporedja je bila krajša od 70 bp) in jih zavrgli. Zaporedja smo ročno analizirali.
2. Ostanki vektorskih zaporedij vplivajo na pravilno zlaganje zaporedij, zato smo z uporabo ukaza »Odstrani vektor« (angl. Trim Vector) znotraj orodne vrstice »Vzorci« (ang. Sample) ob primerjanju z zaporedjem uporabljenega vektorja pGEM-Teasy odstranili zaporedja plazmida z začetka in konca naših vzorcev.

3. V naslednji fazi smo nerazporejene sekvence označili ter jih z ukazom »Razvrsti« (angl. Assemble) v orodni vrstici »Gruča« (angl. Contig) na osnovi podobnosti oz. različnosti razporedili v posamezne gruče. Uporabili smo privzete kriterije, spremenili smo samo parametra podobnost=95 % in prekrivanje=65 bp. Vsa zložena zaporedja smo ročno pregledali in ročno popravili morebitne napake.

3.3 NGS 454 DOLOČEVANJE NUKLEOTIDNEGA ZAPOREDJA

cDNA knjižnico, ki smo jo tretirali z *Gsu*I encimom, smo poslali na določevanje nukleotidnega zaporedja s pomočjo Roche 454 tehnologije v podjetje GATC Biotech, Konstanz, Nemčija. Uporabili so platformo FLX s katero lahko dobimo do 500 bp dolga zaporedja. Odločili smo se za izdelavo konkatemerov cDNA (spajanje cDNA v daljše molekule), njihovo nebulizacijo in nato sekvenciranje. Na ta način lahko pridobimo tudi zaporedja daljših fragmentov in ne samo robne sekvence. Prisotnost adapterskega zaporedja sredi molekule DNA pomeni, d aje tako zaporedje himerno iz dveh zaporedij, ki ga v nadaljevanju razdružimo. Nukleotidno zaporedje smo določili polovici regije pikotiterske plošče, kjer po priporočilih proizvajalca Roche lahko pridobimo do 500.000 zaporedij (točk). Po sekvenciranju smo dobili rezultat v obliki binarne SFF datoteke (angl. standard flowgram format).

3.3.1 Bioinformatska obdelava

Večino nadaljne obdelave podatkov smo opravili na namiznem računalniku s 64-bitnim operacijskim sistemom Kubuntu verzije 10.04 Lucyd Lynx LTS, procesorjem i7 in nameščenim delovnim spominom (RAM) 12 GB. V programskem okolju smo imeli nameščena tolmača (anlg. interpreter) za programska jezika Perl in Python ter nameščene module za bioinformatske obdelave v obeh jezikih BioPerl (Stajich in sod., 2002) in BioPython (Cock in sod., 2009).

3.3.2 Pregled rezultatov sekvenciranja

V prvi fazi smo preverili količino pridobljenih podatkov. S Python skripto sff_extract, ki je del programskega paketa Mira (Chevreux in sod., 2000), smo iz binarne SFF datoteke pridobili informacije o zaporedjih in njihovih kvalitetnih vrednostih v obliki dveh ločenih datotek (FASTA in QUAL).
http://genome.cshlp.org/content/14/6/1147.abstract

Ker smo sekvencirali konkatemere cDNA (združene cDNA) smo morali hibridna zaporedja razdružiti glede na prisotnosts uporabljenih adapterskih zaporedij, ki obdajajo cDNA. Zato smo uporabili parameter skripte sff_extract, ki uporablja rutine programa

SSAHA2 (Ning in sod., 2001) za prepoznavo adapterskih zaporedij, ki jih predložimo v FASTA formatu.

Kvaliteto zaporedij smo preverili s pomočjo programa FastQC (Andrews, 2010). Cilj programa FastQC je zagotoviti preprost način za kontrolo kakovosti surovih podatkov zaporedij, pridobljenih iz visoko zmogljivih sistemov določevanja zaporedij. Zagotavlja modularno vrsto analiz, ki nam omogočajo hiter pregled nad kvaliteto naših podatkov in nas seznanijo z morebitnimi napakami, preden se lotimo nadaljnjih analiz.

Zaporedja, ki jih pridobimo v procesu sekvenciranja lahko vključujejo dele adapterskih zaporedij, dele s slabo kvaliteto, EST zaporedja pa tudu poli-A regije. Zato smo sekvence vključili tudi v proces čiščenja zaporedij. Za to smo uporabili skripto Seqclean (2010), ki uporablja nekaj kriterijev za odstranjevanje delov zaporedij. Zaporedja oljke smo pregledali na prisotnost poli-A-regij in zaporedij adapterjev, uporabljenih za izdelavo cDNA knjižnice in v procesu sekvenciranja:

- SMART OLIGO II začetni oligonukleotidi (5'-AAGCAGTGGTATCAACGCAGAGTACGCrGrGrG-3')
- CDS-GSU začetni oligonukleotidi (5'-AAGCAGTGGTATCAACGCAGAGTACCTGGAG-d(T)20-VN-3')
- smart PCR začetni oligonukleotidi (5'-AAGCAGTGGTATCAACGCAGAGT-3')
- IDT SMART začetni oligonukleotidi (5- AAGCAGTGGTATCAACGCAGAGTCGCATT, 3-CTTCGTCACCATAGTTGCGTCTCAGCGT)
- TitA (CCCATCTCATCCCTGCGTGTCTCCGACTCAG)
- TitA_rev (CTGAGTCGGAGACACGCAGGGATGAGATGG)
- TitB (CCTATCCCCTGTGTGCCTTGGCAGTCTCAG)
- TitB_rev (CTGAGACTGCCAAGGCACACAGGGGATAGG)

Zaporedja, ki so po čiščenju bila krajša od 70 bp smo zavrgli. Tako smo dobili končno datoteko očiščenih zaporedij skupaj z njihovimi kvalitetnimi vrednostmi.

3.3.3 Združevanje zaporedij

Pridobljena končna cDNA zaporedja oljke smo uporabili v naslednjem koraku združevanja, kjer smo želeli pravilno rekonstruirati (zložiti) zaporedja cDNA in pridobiti čim boljšo reprezentacijo transkriptov. Ta korak je bil tudi najbolj delovno zahteven. V tem koraku smo se odločili za podrobnejšo analizo našega seta podatkov z različnimi programi za združevanje (angl. assembler), ki so na voljo. Namen tega dela analiz je bil odkriti najboljši program oz. rutino, ki je primerna za analizo transkriptoma oljke. Uporabili smo naslednje programe za združevanje zaporedij:

- TGICL (Partea in sod., 2003);
- MIRA (Chevreux in sod, 2000);
- iAssembler (Zheng in sod., 2011);
- Newbler 2.3 (Margulies in sod., 2005);
- Newbler 2.5 (Margulies in sod., 2005);
- PAVE 2.5 (Soderlund in sod., 2009);
- CLC Genomic Workbench 4.5 (CLC, 2013)

Preglednica 2 : Lastnosti posameznih programov za združevanje zaporedij.
Table 2: Characteristics of the individual assemblers.

Zbirnik	Tip	Opis	Cena	Podpira tehnologijo
Tgicl 2.1	OLC, ESTs	skripta za CAP3	neplačljivo	Sanger
PAVE 2.5	OLC, ESTs	Skripta za CAP3, mysql integracija	neplačlivo	Sanger, 454
Mira 1.3	OLC, ESTs, genom	interativni zbirnik	neplačljivo	Sanger, 454, Illumina
iAssembler 1.2.2	OLC, ESTs	izvaja Mira in CAP3 interativne zbirnike	neplačljivo	Sanger, 454
Newbler 2.3 in 2.6	OLC, ESTs, genom	Roche program za združevanje	neplačljivo za akademike	Sanger, 454
CLC genomics Workbench 4.5	De Brujin graf, ESTs, genom	SIMD-pospeševalni algoritem zbirnikov	plačljivo	Senger in NGS podatki

Kjer je bilo možno, smo kot merilo združevanja uporabili 96 % identičnost in minimalno prekrivanje odčitkov 40 bp. Ostali parametri so se razlikovali glede na uporabljen program in so bili naslednji:

- TGICL (TIGR Gene Indices clustering tools) je za združevanje zaporedij uporabljal en procesor (-c 1), nastavitve programa pa so zajemale 30 bp za minimalno dolžino prekrivajočih se zaporedij, minimalno 96 % ujemanje prekrivajočih delov, ter 20 bp za maksimalno dolžino štrlečih delov (PID=96 OVL=30 OVHANG=20);
- Pri programu MIRA smo uporabili naslednje parametre: job=denovo, est, normal, 454 z desetimi prehodi (AS: nop=10) za sestavo 454 podatkov (mira --project=oljka --job=denovo, est, normal, 454 --notraceinfo --fasta -OUT:ort=yes:orh=yes -AS:nop=10 -SK:mnr=1 454_SETTINGS-CL:bsqc=1:cpat=1-OUT:sssip=yes 454_SETTINGS-AS:mrpc=1);

- Programa Pave in iAssembler sta uporabljala privzete parametre z minimalno 96 % identičnostjo sestavljenih regij.
- Verziji programa Roche 454 Newbler, 2.3 in 2.5 sta za sestavo zaporedij uporabila standardne parametre (40 bp prekrivanje, 96 % ujemanje), z ukazom -ace smo pridobili ACE datoteke, ki ponazarjajo sestavo sosesk in z ukazom –cpu 8 smo uporabili osem procesorjev naenkrat;
- Nastavitve programa CLC Genomic Workbench so zajemale minimalno dolžino prekrivanja 50 bp, 96 % ujemanje prekrivajočih delov ter 70 bp za minimalno dolžino kontigov.

Po koncu analize smo rezultate razdelili v dve skupini, in sicer v združena zaporedja in v preostala zaporedja, ki se niso vključila v soseske (enojna zaporedja, angl. singletons). Namen tega dela analize je bil odkriti najboljši program oz. rutino, ki je primerna za analizo transkriptoma oljke, pri čemer smo upoštevali različne kriterije pri izbiri najboljšega programa. Programe za združevanje (zbirnike) smo ocenili glede na:

- statistiko združevanja (št. sosesk, št. posameznih zaporedij, skupno št. baz, dolžina posameznih sekvenc, pokritost združevanj, št. kontigov (>=1 kbp), št. kontigov (>=500 bp), maksimalna dolžina kontigov, srednja vrednost dolžine kontigov, mediana, N50, št. kontigov v N50;
- glede na delež zaporedij, ki se niso vključila v sestavljene soseske;
- glede na primerjavo z zaporedji dveh proteinskih podatkovnih baz. Rezultate združevanja smo ocenili s primerjavo združenih zaporedij na lokalno izdelani podatkovni bazi proteinov, ki sta zajemali nepresežna (NR) proteinska zaporedja podatkovne baze NCBI (http://www.ncbi.nlm.nih.gov/protein) ter rastlinska zaporedja UNIPROT podatkovne baze (Apweiler in sod., 2010). Za primerjavo smo uporabili algoritem BLASTX (Altschul in sod., 1990) in upoštevali rezultate z E vrednostjo, ki je bila nižja od 10^{-10};
- Sledila je medsebojna analiza zastopanosti združenih zaporedij v vsaki skupini. Ideja tega načina primerjave je v tem, da odkrijemo program za združevanje zaporedij, ki vključuje največ različnih zaporedij v primerjavi z rezultati ostalih programov. Za primerjavo vseh treh zbirnikov smo uporabili program BLAT (Kent, 2002) s privzetimi parametri in izvedli lokalne parne primerjave zaporedij, pridobljenih z uporabljenimi programi za združevanje zaporedij.

Za določene analize, kot so štetje, seštevanje, primerjave smo zaradi obsežnosti in kompleksnosti uporabljali tudi programsko orodje R (Gentleman in Ihaka, 1997).

3.4 FUNKCIJSKA ANALIZA

Rezultate programa za združevanje zaporedij, ki se je izkazal za najboljšega, smo uporabil pri nadaljnji funkcijski analizi podatkov s programom Blast2go (Gotz in sod., 2011).

Postopek:

1. Nalaganje podatkov: odprli smo okno »Datoteka« (ang. »File«) in izbrali ukaz »Naloži FASTA datoteko« (ang. »Load FASTA file«), ter uvozili podatke v FASTA formatu;
2. BLAST: odprli smo okno »Blast« in izbrali ukaz »Naredi Blast« (ang. »Make Blast«), ter zagnali BLAST iskanje na NCBI ne-redundantno NR bazo podatkov;
3. Mapiranje (ang. Mapping): odprli smo okno »Mapiranje« in izbrali ukaz »Zaženi GO-Mapping korak« (ang. »Run GO-Mapping Step«);
4. Anotacija (ang. Annotation): odprli smo okno »Anotacija« in izbrali ukaz »Zaženi korak Anotacije« (ang. »Run Annotation Step«);
5. Shranitev rezultatov (ang. Save Results): odprli smo okno »Datoteka« in izbrali ukaz »Shrani B2G-projekt (»ang. B2G-project«) za shranitev pridobljenih rezultatov.

Pridobljenim genom so bili, z uporabo privzetih parametrov, dodeljeni Gene Ontology pogoji glede na pogoje, ki bili so pripisani njihovim ustreznim homologom v potakovnih bazah. Uporabili smo tudi orodje Annex, za izboljšanje rezultatov anotacije. GO (ang. Gene Onthology) zaznambe posameznih transkriptov oljke so bile nato mapirane na listo rastlinske GO Slim ontologije. Kategorizacija BLAST zadetkov in oblikovanje strukturnih krogov je bilo izvedeno na najoptimalnejših stopnjah GO-terminov z uporabo standardnih konfiguracij grafa. S KEGG pathway podatkovno bazo pa smo nato določili funkcijske poti posameznih genov.

3.4.1 Analiza s PCR v realnem času (qPCR)

3.4.1.1 Vzorci

Izvedli smo tudi analizo izbranih genov oljke s PCR v realnem času (qPCR). Vzorce plodov oljk sorte ‘Istrska belica’ smo izbrali v petih glavnih fazah razvoja plodu (Slika 2) Imeli smo 12 vzorčnih točk (vzorec 1 (14 dni po cvetenju) in vzorec 2 (29 dni po cvetenju) - fertilizacija in zasnova plodu; vzorec 3 (42 dni po cvetenju) in vzorec 4 (57 dni po cvetenju) – razvoj koščice; vzorec 5 (72 dni po cvetenju) in vzorec 6 (85 dni po cvetenju) – otrditev koščice; vzorec 7 (98 dni po cvetenju), 8 (112 dni po cvetenju), 9 (129 dni po cvetenju) in 10 (149 dni po cvetenju) – razvoj mezokarpa; vzorec 11 (158 dni po cvetenju) – zorenje plodu; vzorec 12 (182 dni po cvetenju) – prekomerno zrel plod). Vzorce plodov smo nabrali iz petih naključno izbranih dreves sorte ‘Istrska belica’, v komercialnem

olj̆nem nasadu. Takoj po obiranju smo plodove zmrznili v tekočem dušiku in jih shranili do uporabe pri -80 °C. Izolacija, kvantifikacija in preverjanje kvalitete RNA so potekali po že predhodno omenjeni postopkih opisanih v poglavju 3.1.2.

29 potencialnih referenčnih genov smo izbrali na podlagi literature in glede na razpoložljivost ortolognih sekvenc oljke v našem setu podatkov. Od tega je bilo 26 kandidatnih referenčnih genov izbranih glede na kandidatne referenčne gene pri drugih že predhodno analiziranih rastlinskih vrstah, trije kandidatni referenčni geni pa so znani kot interne reference pri določevanju gensko spremenjenih organizmov in smo jih zaradi njihovega značilno nizkega števila kopij v genomu vključili v raziskavo (Preglednica 3).

Preglednica 3: Izbor 29 kandidatnih referenčnih genov, ki so namenjeni normalizaciji ekspresije olj̆nih genov; podana so imena genov in njihove okrajšave, ki smo jih pridobili s pomočjo referenčnih vrst, ter njihove GenBank akcesijske številke in zaporedja pridobljena iz GenBank ali 454 zaporedij; referenčni geni so razvrščeni glede na geNorm razvrstitev.
Table 3: Selection of 29 candidate reference genes used for gene expression normalization experiment in olive; gene names and their abbreviations are reported with the reference species and their GenBank accession number with the olive sequence obtained either from GenBank or 454 sequences; the reference genes are ordered according to the geNorm ranking

Št.	Ime gena	Okrajšava	Določeno v vrsti	GenBank pristopno število	Reference	Vir zaporedja oljke
1	TIP41-sorodni protein	TIP41	*Solanum lycopersicum*	BT014035	Exposito-Rodriguez in sod. (2008)	454
2	TATA vezavni protein	TBP	*Solanum lycopersicum*	AK329831	Exposito-Rodriguez in sod. (2008)	454
3	Protein kinaza mRNA	Pkaba1	*Triticum aestium*	M94726	Chaouachi in sod. (2007)	454
4	Adenin fosforibozil transferaza	APRT	*Solanum tuberosum*	CK270447	Nicot in sod. (2005)	454
5	Klatrin adaptor	CLATHRIN	*Solanum lycopersicum*	SGN-U314153[2]	Exposito-Rodriguez in sod. (2008)	454
6	Domnevna sucinil-CoA ligaza	SucCoA	*Urochloa brizantha*	GE617476	Pratt in sod. 2005	454
7	14-3-3 protein	1433P	*Coffea canephora*	SGN-U627733[2]	Barsalobres-Cavallari in sod. (2009)	454

Se nadaljuje.

Nadaljevanje.

Št.	Ime gena	Okrajšava	Določeno v vrsti	GenBank pristopno število	Reference	Vir zaporedja oljke
8	Yellow leaf specifični gen 8 mRNA	YLS8	*Arabidopsis thaliana*	NM_120912	Czechowski in sod. (200)5	454
9	Ribosomalni protein L2	RPL8	*Solanum tuberosum*	CK259681	Nicot in sod. (2005)	454
10	60S ribosomalni protein	60S	*Arabidopsis thaliana*	NM_119780	Czechowski in sod. (2005)	454
11	Rotamaz ciklofilin 5	ROC5	*Arabidopsis thaliana*	NM_203166	Czechowski in sod. (2005)	454
12	Poliubikvitin 11	UBQ11	*Populus trichocarpa*	BU879229	Brunner in sod. (2004)	454
13	Ciklofilin	CYP	*Populus trichocarpa*	BU875027	Brunner in sod. (2004)	454
14	60S ribosomalni protein L7	RPL7C	*Coffea canephora*	SGN-U351477[2]	Barsalobres-Cavallari in sod. (2009)	454
15	NADH dehidrogenazna podenota F	NDHF	*Humulus lupulus*	AY289251	Maloukh in sod. (2009)	AF130163
16	Elongacijski faktor 1-α	ELNFa	*Solanum tuberosum*	AB061263	Nicot in sod. (2005)	454
17	(UDP)-glukozna pirofosforilaza	UGPase	*Solanum tuberosum*	U20345	Chaouachi in sod. (2007)	GO245620
18	Aktin 11	ACT11	*Populus trichocarpa*	CA824001	Brunner in sod. (2004)	GO243999
19	Gliceraldehide-3-fosfat dehidrogenaza	GAPDH	*Coffea canephora*	SGN-U347734[2]	Exposito-Rodriguez in sod. (2008)	454
20	Klatrin adaptor	CAC	*Solanum lycopersicum*	SGN-U314153[2]	Exposito-Rodriguez in sod. (2008)	454
21	DnaJ-podobni protein	DNAJP	*Solanum lycopersicum*	AF124139	Exposito-Rodriguez in sod. (2008)	454
22	18S ribosomalna RNA	18S	*Populus tremuloides*	AF206999	Brunner in sod. (2004)	L49289
23	Izraženo zaporedje SGNU-346908	EXP	*Solanum lycopersicum*	SGN-U346908[2]	Exposito-Rodriguez in sod. (2008)	454

Se nadljuje.

Nadaljevanje.

Št.	Ime gena	Okrajšava	Določeno v vrsti	GenBank pristopno število	Reference	Vir zaporedja oljke
24	Poliubikvitin 10	UBQ10	*Coffea canephora*	SGN-U347154[2]	Barsalobres-Cavallari in sod. (2009)	454
25	Alfa tubulin	TUA3	*Arabidopsis thaliana*	M17189	Exposito-Rodriguez in sod. (2008)	454
26	Cistein proteinaza	CYS	*Coffea canephora*	SGN-U352616[2]	Barsalobres-Cavallari in sod. (2009)	454
27	Alkohol dehidrogenaza 1	ADH1	*Zea mays*	X04050	Chaouachi in sod. (2007)	454
28	Beta-tubulin	TUBb	*Solanum tuberosum*	Z33382	Nicot in sod. (2005)	454
29	SAND sorodni protein	SAND	*Arabidopsis thaliana*	NM_128399	Czechowski in sod. (2007)	454

Za iskanje sekvenc oljke, ki so ortologne potencialnim referenčnim genom, smo uporabili dva seta zaporedij oljk. Prvi set je predstavljalo 7.080 olјčnih mRNA zaporedij iz GeneBank javno dostopne podatkovne baze, ter izražena nukleotidna zaporedja oljk, ki smo jih pridobili v sklopu 454 transkriptoma olјčnih plodov. Iz teh dveh sklopov smo izdelali lokalno podatkovno bazo s pomočjo orodja format db, ki je del BLAST paketa (Altschul in sod., 1990). Z uporabo BLASTN, BLASTX in TBLASTX algoritmov smo 29 referenčnih genov iz drugih rastlinskih vrst (Preglednica 3) primerjali z zaporedji oljk.

Prav tako smo v to primerjavo vključili štiri gene, ki sodelujejo v metabolizmu rastlinskih maščobnih kislin (maščobna acil-ACP tioesteraza A, *FatA*, XM_002303019; stearoil-ACP desaturaza, *SAD1*, AJ132636; acil-CoA tioesterazni sorodni protein, *Acot*, NM_100053; lipoksigenaza 1, *LOX1*, NM_104376).

Zaporedja oljke z najvišjo identičnostjo, z najdaljšo dolžino zadetkov in najnižjo E vrednostjo, smo uporabili v nadaljnjem poskusu uporabnosti teh genov za interne kontrole pri genski ekspresiji oljke, ter kot tarčne gene, ki so vključeni v metabolizem maščobnih kislin. Za vseh 33 zaporedij smo nato izdelali začetne oligonukleotide z uporabo programa Primer Express version 3.0 (Applied Biosystems), ki je imel naslednje nastavitve: maksimalna dolžina amplikona 110 bp, optimalna talilna temperatura 60 °C in GC vrednost med 30 % in 80 %. Vse pare začetnih oligonukleotidov je sentitiziralo podjetje Integrated DNA Technology (Leuven, Belgium).

3.4.1.2 RT-qPCR analiza in kvantifikacija ekspresije genov

Vsak vzorec RNA smo reverzno prepisali v cDNA, z uporabo High Capacity cDNA Reverse Transcription kita (Applied Biosystems, Foster City, USA) in naključno prilegajočih heksamernih primerjev.

20 µl PCR reakcije je vsebovalo:

- 10x PCR pufer 10 µl
- 25x dNTP 0,8 µl
- 10x RT začetni oligonukleotidi 2,0 µl
- Reverzna transkriptaza 1,0 µl
- Rnase inhibitor 1,0 µl
- Voda (nuclease free) 3,2 µl
- RNA 1µg

Reakcija je potekala pri 25 °C 10 min, 37 °C 60 min, 85 °C 5 sek, 4 °C ∞. Prepisane vzorce cDNA smo nato hranili pri -20 °C.

RT-qPCR je bil izveden z uporabo Fast SYBR Green tehnologije na napravi ABI PRISM 7500 Sequence Detection System (Applied Biosystems, Foster City, CA, USA). Na plošči za 96 PCR reakcij (MicroAmp Optical PCR plate, Applied Biosystems) je vsak vzorec v 20 µl PCR reakcije vseboval 10 µl FAST SYBR Green PCR Master Mix (Applied Biosystems), 2 µl cDNA in 300 nM vsakega posameznega začetnega oligonukleotida. Za pomnoževaje smo uporabili skrajšan FAST program pomnoževanja: 95 °C začetne denaturacije za 20 s, sledilo je 40 ciklov po 95 °C za 3 s in 60 °C za 30 s. Za vsak PCR vzorec smo imeli po tri tehnične ponovitve. Pri PCR vzorčkih, ki smo jih namnožili, smo nato preverili tudi talilno krivuljo in jih naložili na 2,2 % gelsko elektroforezo (kot v poglavju 3.1.3 Agarozna gelska elektroforeza), da smo preverili specifičnost namnoževanja.

Končni preizkus, v katerem smo želeli določiti izražanje referenčnih genov in tarčnih genov, je vseboval standardno krivuljo sestavljeno iz šestih zaporednih točk redčenja, ter 12 vzorcev cDNA iz različnih stopenj razvoja oljčnih plodov. Združeno cDNA vseh 12 vzorcev smo uporabili za določitev standardne krivulje, in sicer tako da smo opravili šest zaporednih štirikratnih redčitev cDNA, z začetno koncentracijo 50 ng in končno 0,05 ng (50, 12.5, 3.13, 0.78, 0.20 and 0.05 ng). Variacije med pogoni ABI naprave smo zmanjšali tako, da smo vse reakcije z enakim začetnim oligonukleotidom vključili na eno ploščo, ter na vsaki plošči naredili standardno krivuljo. RT-qPCR učinkovitost je bila določena za vsak gen s pomočjo nagiba regresijske črte v standardni krivulji, ter izračunana z ABI 7500 SDS 2.0.4 programom (Applied Biosystems).

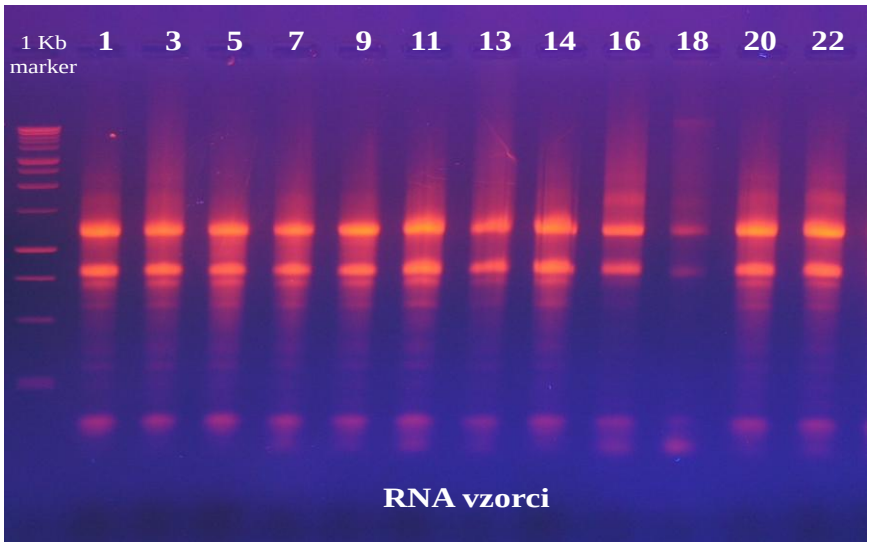
Za kvantifikacijo rezultatov smo uporabili dva pristopa. V prvem smo podatke pridobljene s serijo standardnih redčitev uporabili v skladu z metodo relativne standardne krivulje. S pomočjo standardne krivulje programska oprema SDS določi relativno količino tarčnega gena v vsakem vzorcu, tako da primerja količino vsakega vzorca s količino referenčnega vzorca. Določene relativne količine so nato prenesene v geNorm program , ta pa nato izračuna povprečno stabilnost ekspresije gena (M vrednost), ki je definirana kot povprečna parna variacija določenega gena z vsemi ostalimi kontrolnimi geni. Nizka M vrednost je pokazatelj stabilne ekspresije gena (Vandesompele in sod., 2002). Določili smo tudi parno variacijo med normalizacijskim faktorjem NF_n in normalizacijskim faktorjem NF_{n+1}, da bi dobili optimalno število referenčnih genov, ki so potrebni za normalizacijo. V drugem pristopa pa smo uporabili program RefFinder (Xie in sod., 2011), ki združuje štiri algoritme: geNorm, NormFinder, BestKeeper in primerjalno delta Cq metodo. Vsak algoritem določi svojo razvrstitev potencialnih referenčnih genov, program pa nato s pomočjo geometrijske sredine izračuna skupno razvrstitev potecialnih referenčnih genov (Xie in sod., 2011).

Dva najbolj stabilna referenčna gena smo nato uporabili pri normalizaciji nivoja ekspresije štirih genov, ki so potencialno vključeni v metabolizem maščobnih kislin.

4 REZULTATI

4.1 VZORČENJE OLJK IN IZOLACIJA RNA

Plodove oljk sorte 'Istrska belica' za potrebe razvoja transkriptoma smo vzorčili v 22 časovnih točkah skozi celotno obdobje razvoja plodov (Preglednica 1). RNA smo izolirali iz vsakega vzorca posebej z uporabo Spectrum Total Plant RNA Extraction Kita, ki se je izkazal za primernega. RNA vzorce smo pregledali s pomočjo gelske elektroforeze na agaroznem gelu. Ta analiza je pokazala fragmenta ribosomalne RNA brez vidne degradacije (Slika 12).



Slika 12: Pregled 12 RNA vzorcev (Preglednica 1) na 1,2 % gelski elektroforezi
Figure 12: Viewing 12 RNA samples (Table 1) on 1,2 % gel electrophoresis

Koncentracijo in čistoto vseh RNA vzorcev smo izmerili s pomočjo Nano drop UV spektrofotometra (GE Healthcare), ki je tudi podal razmerje absorpcijskih vrednosti izmerjenih pri 260 in 280 nm (A260/A280). Vzorci so imeli primerno razmerje z vrednostmi med 1,8 in 2,0 (Preglednica 1). Na koncu smo zmešali ekvimolarne količine vseh RNA vzorcev (3 ug RNA na vzorec), da smo dobili združen, reprezentativen vzorec vseh RNA izraženih v celotnem razvojnem obdobju olj̆nega plodu. Ta vzorec je bil uporabljen v naslednjih korakih sekvencianja transkriptoma.

4.2 NORMALIZIRANA cDNA KNJIŽNICA

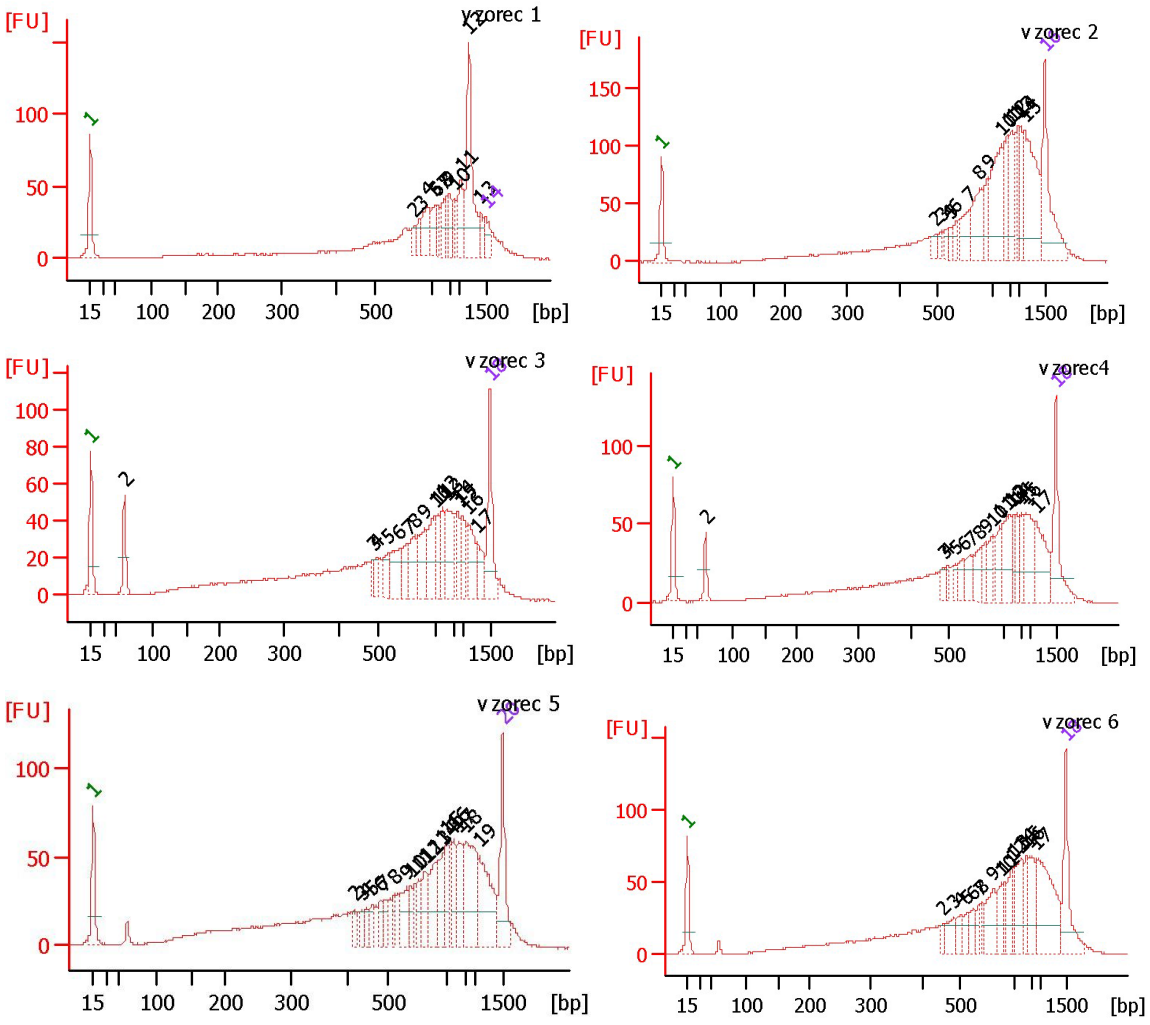
Reprezentativen vzorec RNA s koncentracijo 466 ng/µl smo poslali naprej za izdelavo normalizirane cDNA knjižnice (Evrogen Lab, Rusija) s pomočjo SMART tehnologije (Shagin in sod., 2002; Zhu in sod., 2001; Zhulidov in sod., 2004). Po opravljenem

postopku normalizacije smo nato pridobili 1200 µl normalizirane cDNA knjižnice (5.0 µg) in 450 µl ne-normalizirane cDNA knjižnice (1.0 µg). Obe knjižnici smo očistili, normalizirano cDNA knjižnico pa smo nato amplificirali in jo del ligirali v pGEM-T-easy vektor (Promega), ter jo tako ohranili za daljše časovno obdobje. Preostali del pa smo uporabili za določevanje nukleotidnega zaporedja s pomočjo novih tehnologij sekvenciranja (Roche 454).

Prisotnost homopolimernih odsekov v DNA zaporedju, kot so poli A/T konci, povzroča težave pri določevanju nukleotidnih zaporedij v cDNA knjižnicah (Shibata in sod., 2001). Zato smo pred samim sekvenciranjem z Roche 454 tehnologijo, cDNA obdelali z *Gsu*I restrikcijskim encimom, ki cepi dvoverižno cDNA 14/16 bp stran od prepoznavnega mesta, ter se na tak način poskusili izogniti težavam, ki jih povzroča pola A regija. Sledilo je primerjalno določevanje nukleotidnega zaporedja za preizkus uspešnosti izreza poli A regije. Primerjali smo rezultate določevanja nukleotidnega zaporedja klonom ligirane normalizirane cDNA in ligirane normalizirane *Gsu*I-cDNA. Iz vsake knjižnice smo 192-im klonom v PCR pomnožili cDNA insert. V prvem primeru (brez trtiranja z *Gsu*I encimom) smo naredili obojestransko sekvenčno rekacijo pomnoženim fragmentom (384 reakcij). Pri drugi knjižnici, ki pa je bila tretirana z *Gsu*I encimom, ligirana z adapterji, ter reamplificirana, pa smo naredili enostransko sekvenčno reakcijo (192 reakcij). Pri prvi knjižnici je bilo od skupno 384 zaporedij le 158 (41 %) uporabnih za nadaljnjo analizo. Neuporabnih sekvenc je bilo 226, od tega je bilo 158 (70 %) slabih zaporedij zaradi prisotnosti poliA regij. Po preverjanju redundantnosti v knjižnici s 95 % ujemanjem kot merilom identičnosti smo dobili 140 posameznih zaporedij in 9 združenih zaporedij. Skupno smo določili 63.655 bp DNA zaporedij, ki so bila v povprečju dolga 402 bp. Iz druge knjižnice pa smo določili nukleotidno zaporedje 192-im vzorcem samo iz ene strani. Od 192 zaporedij je bilo kar 158 (81 %) primernih za nadaljnjo analizo, le manjši delež zaporedij je še vseboval prisotne homopolimerne A regije. Takih primerov je bilo od 35 slabih zaporedij le 12 (35 %). Zaporedja v tej knjižnici so bila v povprečju dolga 484 bp, skupno smo določili 76.485 bp dolžine DNA. Na koncu smo vsa zaporedja iz obeh knjižnic združili skupaj in preverili redundantnost, skupno smo določili 112.134 bp DNA in 287 enkratnih zaporedij v povprečni dolžini 390 bp. Na podlagi teh rezultatov smo se odločili, da je smiselno sekvenciranje *Gsu*I tretirane knjižnice.

Kvaliteto ne-normalizirane cDNA, normalizirane cDNA, cDNA po restrikciji, ter cDNA knjižnice po restrikciji in čiščenju, smo preverili tudi s pomočjo naprave Agilent Bioanalyzer 2100 in uporabo čipa DNA1000. Ugotovili smo, da je pri ne-normalizirani knjižnici prisotna večja količina zaporedij z dolžino okoli 1300 bp (Slika 13, vzorec 1). Z normalizacijo (Slika 13, vzorec 2) smo število teh zaporedij uspešno zmanjšali, prisotna pa je ostala povečana količina zaporedij okoli 1500 bp. Po obdelavi knjižnice z *Gsu*I encimom se število daljših sekvenc zmanjša, pojavi pa se povečano število kratkih sekvenc dolžine okoli 60 bp (Slika 13, vzorca 3 in 4), ki so bili posledica odstranjenih delov cDNA po

tretiranju z *Gsu*I restrikcijskim encimom. Število teh zaporedij smo nato uspešno zmanjšali s čiščenjem knjižnice z uporabo silicijeve kolone (vzorec 5 in 6). Skozi vse korake po normalizaciji knjižnice pa ostaja povečano število zaporedij dolžine okoli 1500 bp (vzorec 2, 3, 4, 5 in 6).



Slika 13: Določitev kvalitete ne-normalizirane cDNA (vzorec 1), normalizirane cDNA (vzorec 2), cDNA po restrikciji (vzorec 3 in 4), ter cDNA knjižnice po restrikciji in čiščenju (vzorec 5 in 6) z napravo Agilent Bioanalyzer 2100 in uporabo čipa DNA1000.

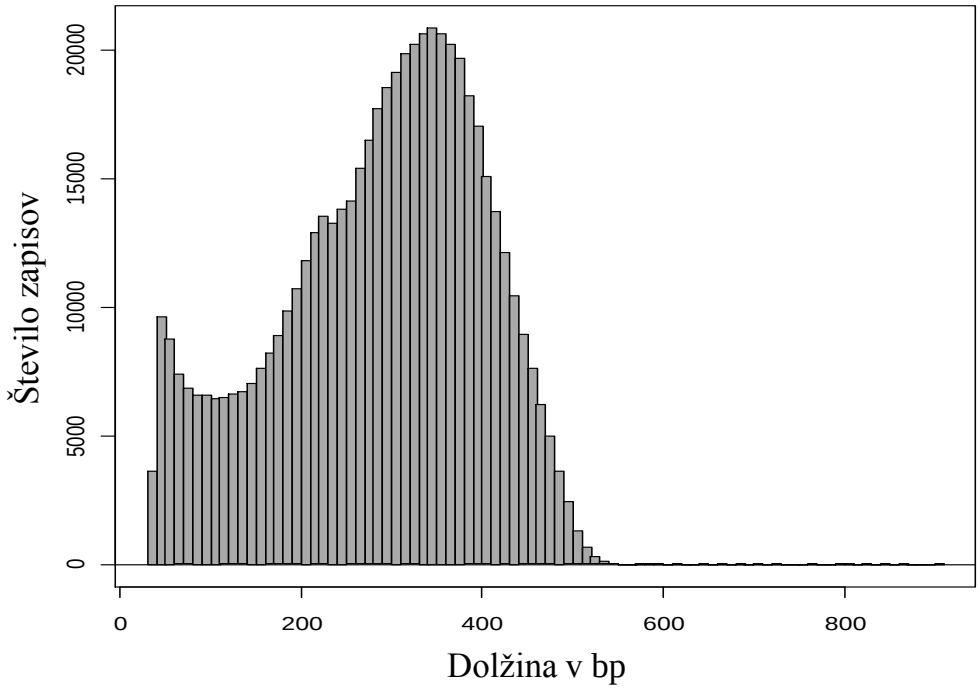
Figure 13: Quality determination of non-normalized cDNA (sample 1), normalized cDNA (sample 2), cDNA after restriction (sample 3 and 4), and cDNA library after restrictions and clining (sample 5 and 6) with a device Agilent 2100 Bioanalyzer using the chip DNA1000 .

4.3 454 PIROSEKVENCIRANJE

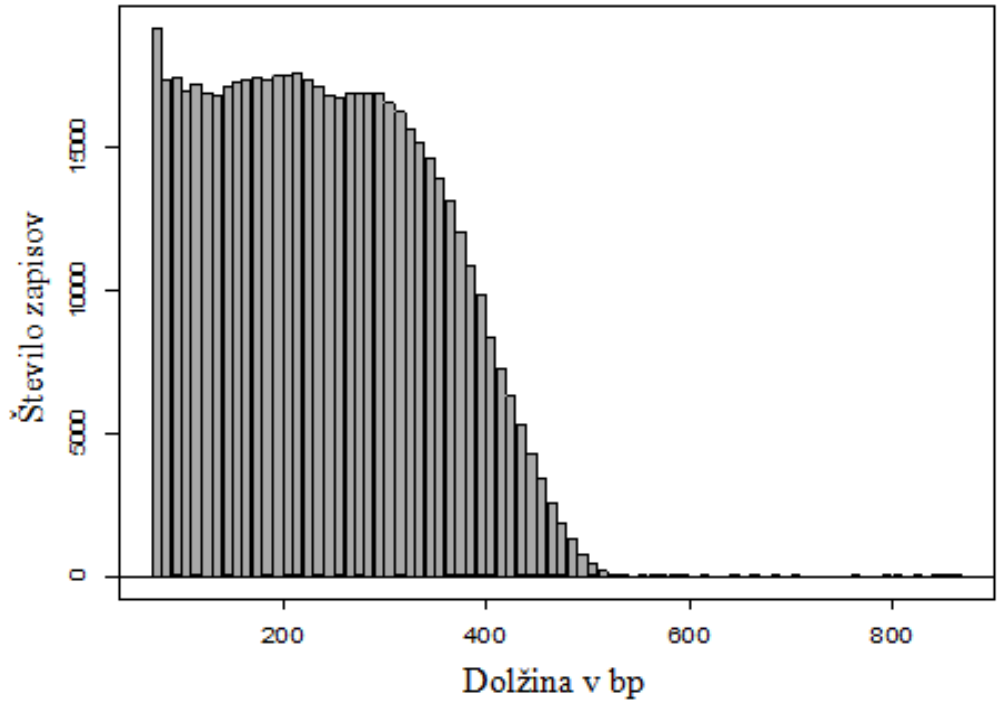
cDNA knjižnico, ki smo jo tretirali z *Gsu*I encimom, smo poslali na določevanje nukleotidnega zaporedja s pomočjo Roche 454 FLX tehnologije (GATC Biotech, Konstanz, Germany). Nukleotidno zaporedje smo določili polovici regije pikotiterske plošče, kjer dobimo do 500.000 zaporedij.

Surove podatke smo nato prejeli v obliki binarne SFF datoteke (ang. standard flowgram format). Prva analiza je pokazala, da smo pridobili 560.578 sekvenc v skupni dolžini 160.414.301 bp. Povprečna dolžina zaporedij je bila 286 bp, minimalna vrednost 34 bp, maksimalna pa 904, medtem ko je bila N50 vrednost seta 343 bp (200.290 zaporedij je imelo dolžino enako ali večjo od N50 vrednosti). Povprečna vsebnost GC je bila 41.8 % (Slika 14). Ko smo konkatemerne cDNA razdružili s pomočjo SSAHA programa, smo pridobili 703.936 zaporedij v skupni dolžini 147.278.109 bp. Povprečna dolžina sekvenc je bila 209 bp, N50 vrednost pa 295 bp (200.873 zaporedij). Povprečna vsebnost GC je bila 41.05 %. Ta zaporedja bi lahko še vedno vsebovala dele zaporedij, ki smo jih uporabili pri izdelavi cDNA ali sekvenciranju in niso del oljčne DNA. Zato smo ta zaporedja vključili tudi v proces čiščenja zaporedij, s katerim smo le-te odstranili, odstranili smo tudi morebitne poli-A regije, ter prekratka zaporedja (pod 75 bp). Na koncu smo tako pridobili 577.025 zaporedij v skupni dolžini 139.419.844 bp. Povprečna dolžina zaporedij je bila 242 bp, minimalna vrednost 70 bp, maksimalna pa 870, medtem ko je N50 vrednost bila 294 bp (192.189 vseh zaporedij). Povprečna vsebnost GC je bila 40,90 % (Slika 15).

Neobdelani sekvenčni podatki so na voljo zainteresiranim uporabnikom na spletu preko NCBI SRA arhiva (http://www.ncbi.nlm.nih.gov/sra/SRX215662).

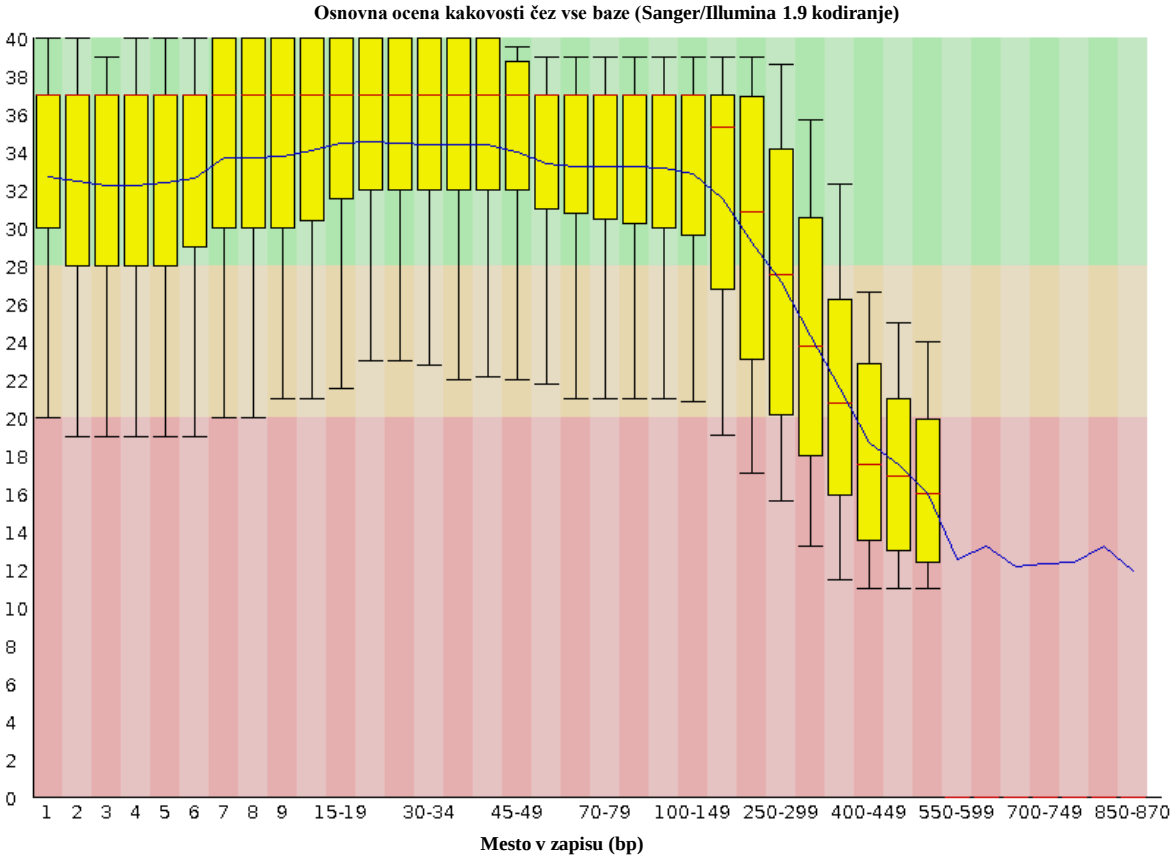


Slika 14: Histogram prikazuje dolžine surovih zaporedij
Figure 14: Histogram is showing the length of the raw sequences



Slika 15: Histogram prikazuje dolžine razdruženih in očiščenih zaporedij
Figure 15: Histogram showing the length of the sequences after splitting and clinning

Pri sekvenciranju dobimo poleg samega zaporedja tudi kvalitetno oz. Phred (Ewing in sod., 1998) vrednost posamezne baze. Zato smo določili tudi analizo kvalitetnih vrednosti transkriptoma na posamezno bazo. Kar 99 % zaporedij je imelo oceno kakovosti nad mejno vrednostjo 20, dolžina teh zaporedij pa je znašala od 350 do 400bp. S povečevanjem dolžine zaporedij se je zmanjševala njihova kvalitetna vrednost, ki je pri dolžinah nad 400 bp, padla tudi pod mejno vrednost 20 (Slika 16).



Slika 16: Osnovna ocena kakovosti razdruženih in očiščenih zaporedij. Rdeča črta pradstvalja srednjo vrednost (mediana), rumene škatle predstavljajo kvantila 0,25 in 0,75, repki predstavljajo percentila 10 in 90, medtem ko modra črta predstavlja srednjo vrednost ocene kakovosti. (Andrews, 2010)
Figure 16: Quality score of splitted and cleaned sequences. The red line represents the median value, yellow boxes are 25 and 75 percentiles, while whiskers are 10 and 90 percentiles and blue line is the mean value. (Andrews, 2010)

4.5 BIOINFORMATSKA OBDELAVA PRIDOBLJENIH ZAPOREDIJ

Odločili smo se za podrobnejšo analizo našega seta podatkov z različnimi pogrami za združevanje zaporedij: TGICL, MIRA, iAssembler, PAVE in Newbler (v2.3. in v2.6.), ki temeljijo na OLC (angl. over-layout-consensus) metodi, ter CLC Genomic Workbench 4.5, ki temelji na metodi uporabe De brujin grafa. Kjer je bilo mogoče smo kot merilo

združevanja uporabili 96 % identičnost in minimalno prekrivanje 40 bp. Po koncu analize smo rezultate razdelili v dve skupini, in sicer v združena zaporedja (kontigi-soseske) in preostala zaporedja (enojna zaporedja-singletoni). Programe za združevanje zaporedij (zbirnike) smo ocenili glede na statistiko združevanj, glede na delež unikatnih zaporedij, rezultate združevanja pa smo ocenili tudi s primerjanjem združenih zaporedij na lokalno izdelane podatkovne baze proteinov (BLASTX) (Altschul in sod., 1990) in medsebojnim primerjanjem rezultatov zbirnikov (BLAT) (Kent, 2002).

Preglednica 4: Osnovni podatki meritev različnih programov za združevanje zaporedij
Table 4: Basic data measurements of different assembly programs

Meritve	New 2.6.	New 2.3.	MIRA	iAssembler	CLC	TGICL	PAVE
Število sosesk	15.224	13.530	42.504	49.860	32.138	35.074	40.219
Skupno število baz	8.086.878	8.439.420	21.930.174	25.529.782	14.646.256	17.215.800	20.024.716
Posamezna zaporedja	73.087	77.773	74.953	49.064	52.611	66.103	47.766
Dolžina posameznih zaporedij	17.523.948	18.103.598	17.133.522	11.258.808	12.818.683	15.585.570	10.414.216
Pokritost združevanj	15.07 ×	14.37 ×	5.58 ×	5.02 ×	8.64 ×	7.19 ×	6.44 ×
Število sosesk (>=1 kbp)	694	1.121	2.141	2.363	1.005	1.343	1.549
Število sosesk (>= 500 bp)	8.038	9.004	18.860	21.879	11.305	14.115	16.560
Maksimalna dolžina sosesk	3.456	4.336	3.738	4.473	3.142	3.032	4.619
Srednja vrednost dolžine sosesk	531.2	623.8	516	512	455.7	490.8	497.9
Mediana	518	587	468	466	414	446	452
N50	640	687	586	585	532	559	563
Število sosesk v N50	4.869	4.657	13.484	15.813	9.831	11.137	12.876
Čas	30 min	30 min	15 hours	15 hours	5 min	41 h	12 days

Preglednica 5: BLAT primerjava programov za združevanje zaporedij; pridobimo število unikatnih zaporedij posameznega združevanja v primerjavi z ostalimi programi.
Table 5: BLAT comparison all vs. all to determine the number of unique sequences in "query" assembly not present in "database" assembly.

POIZVEDBA							
TGICL	1171	62	655	14587	13186	196	
PAVE	1371	255	1009	15772	13908		912
Newbler 2.6	146	76	86	1850		86	130
Newbler 2.3	30	3	10		572	3	11
MIRA	1452	342		16119	13269	663	988
iAssembler	2447		1394	17826	15049	1021	1599
CLC		481	889	14044	12720	421	963
	CLC	iAsse	MIRA	New 2.3	New 2.6	PAVE	TGICL
				BAZA			

Preglednica 6: BLASTX rezultati posameznih programov za združevanje zaporedij
Table 6: BLASTX results of individual assembly programs

		CLC	iAsse	MIRA	New 2.3	New 2.6	PAVE	TGICL
NR	Soseske z zadetki	67.8	71.2	72.6	83.6	75.8	72.6	71.1
Uniprot	v bazi (%)	67.8	71.8	72.9	84.2	77.0	72.9	71.2
NR	Enkratni zadetki	65.1	43.6	46.3	57.1	66.9	50.6	56.8
Uniprot	v bazi (%)	54.2	33.6	36.4	51.8	59.3	40.2	45.8
Nr	En + dva zadetka	87.5	64.2	66.7	83.4	88.2	73.2	79.8
Uniprot	v bazi (%)	79.7	54.2	57.7	79.7	82.6	63.7	71.4
Nr	Maksimalno št.	10	79	300	10	14	35	14
Uniprot	pojavljanja zadetka	12	78	311	14	12	37	17
NR	Sekvence z 70 % - 100%	6.7	7.7	8.4	14.6	11.8	4.9	8.2
Uniprot	pokritostjo zadetka s proteinom (%)	6.6	7.6	8.1	14.6	11.7	6.2	8.1

4.5.1 Programi za združevanje zaporedij - zbirniki

4.5.1.1 TGICL

S TGICL zbirnikom smo pridobili 35.074 združenih zaporedij v skupni dolžini 17.215.800 bp. Število kontigov daljših ali enakih 500 bp je bilo 14.115, medtem ko je bilo kontigov daljših ali enakih 1.000 bp 1.343. Maksimalna dolžina kontiga je znašala 3.032 bp, povprečna dolžina 491 bp, medtem ko je bila medijana 446 bp. N50 vrednost je znašala 559 bp. Zaporedij, ki se niso združila (singletoni) je bilo 66.103 v skupni dolžini 15.585.570 bp (Preglednica 5). Čas, ki ga je zbirnik potreboval za zaključek procesa, je

znašal 41 ur. BlastX primerjava z ne-redundantno (NR) proteinsko bazo (14.987.464 sekvenc; 5.132.678.026 znakov), rastlinsko UniProt bazo (410.553 sekvenc; 143.146.364 znakov) pri mejni e-vrednosti $<e10^{-10}$ je prinesla 21.107 zadetkov za NR bazo in 20.877 zadetkov za UniProt bazo, ter pokazala, da je 71 % kontigov, ki smo jih pridobili s TGICL zbirnikom, imelo zadetke z zapisi v podatkovnih bazah. V ne-redundantni (NR) bazi podatkov je bilo od tega kar 56.8 % edinstvenih zadetkov, medtem ko je bilo v UniProt bazi takih zadetkov 45.8 %. V NR bazi se je isti zadetek pojavil največ 14-krat, v UniProt bazi pa 17-krat. Kontigov, ki so imeli 70 % - 100 % dolžinsko ujemanje s proteini, je bilo v obeh bazah okoli 8 % (Preglednica 6).

4.5.1.2 MIRA

Z zbirnikom MIRA smo pridobili 42.504 kontigov v skupni dolžini 210.930.174 bp. Maksimalna dolžina kontigov je bila 3.738 bp, povprečna dolžina 516 bp in srednja vrednost 468 bp. Število kontigov, ki so bili daljši ali enaki dolžini 500 bp, je bilo 18.869, medtem ko je bilo 2.141 kontigov daljših ali enakih 1.000 bp. Le 49.711 posameznih sekvenc skupne dolžine 10.821.840 bp ni bilo združenih v kontige (Preglednica 5). Zbirnik MIRA je končal proces v 15 urah. Kot pri programu TGICL smo tudi tukaj opravili BlastX primerjavo in pridobili 26.235 zadetkov pri NR bazi in 25991 zadetkov pri UniProt bazi. Rezultati so pokazali, da je 72 % kontigov imelo zadetke z bazo podatkov, od tega je bilo v NR bazi 46.3 % edinstvenih zadetkov, v rastlinski UniProt bazi pa 36.4% edinstvenih zadetkov. V obeh bazah podatkov je okoli 8 % sosesk imelo 70 % - 100 % ujemanje s proteini (Preglednica 6).

4.5.1.3 iAssembler

S programom iAsseembler smo pridobili kar 49.860 kontigov v skupni dolžini 25.529.782 bp in le 49.064 singletonov s skupno dolžino 11.258.808 bp. Število kontigo, ki so bili daljši ali enaki 500 bp, je bilo 21.879, kontigov enakih ali daljših od 100 bp pa 2.363. Maksimalna dolžina kontigov je znašala 4.473 bp, povprečna dolžina 512 bp in mediana 466 bp. Celoten proces združevanja je trajal 15 ur. Maksimalna dolžina kontigov je znašala 4.473 bp, povprečna dolžina 512 bp in mediana 466 bp. Celoten process združevanja zaporedij je trajal 15 ur (Preglednica 5). Z BlastX primerjavo proti ne-redundantni (NR) proteinski bazi in rastlinski UniProt bazi z E vrednostjo <e^-10 smo pridobili skoraj 29378 zadetkov pri NR bazi in 29297 zadetkov pri UniProt bazi. 71.2 % kontigov, ki smo jih uporabili za BlastX primerjavo z NR proteinsko bazo podatkov, je imelo zadetek, od tega pa je bilo 43.6 % edinstvenih zadetkov. Kontigi, ki smo jih uporabili pri BlastX primerjavi z rastlinsko UniProt bazo podatkov, pa so imeli 71.8 % zadetkov, od tega je bilo 33.6 % edinstvenih zaporedij. Kontigov, ki so imeli 70 % - 100 % ujemanje s proteini v NR proteinski bazi podatkov je bilo 7.7 %, medtem ko je bilo takih kontigov v rastlinski UniProt bazi podatkov 7.6 % (Preglednica 6).

4.5.1.4 PAVE

Program PAVE nam je združil sekvence v 40.219 kontigov s kupno dolžino 20.024.716 bp. Dolžina najdaljšega kontiga je bila 4.619 bp, povprečna dolžina kontigov je bila 4.979 bp in mediana 452 bp. Število kontigov, ki so bili daljši ali enaki 500 bp, je bilo 16560, medtem ko je bilo število kontigov daljših ali enakih 100 bp 1.549. N50 vrednost je pri programu PAVE znašala 563 bp. Število sekvenc, ki se niso združile je bilo nizko, saj smo pridobili le 47.766 singletonov v skupni dolžini 10.414.216 bp. Združevanje zaporedij s programom PAVE je bilo zaključeno v 12 dneh (Preglednica 5). BlastX primerjava z E vrednostjo <e^-10 je prinesla 24498 zadetkov pri NR bazi in 24311 zadetkov pri UniProz bazi.. V ne-redundantni (NR) proteinski bazi podatkov je 72.6 % kontigov imelo zadetke, od tega jih je bilo 50.6% edinstvenih. V rastlinski UniProt bazi podakov pa je 72.9 % kontigov imelo zadetke, od tega je bilo 40.2 % edinstvenih zadetkov. Kontigov, ki so imeli 70 % - 100 % ujemanje je bilo v NR proteinski bazi 4.9 %, v ratlinski UniProt bazi pa 6.2 % (Preglednica 6).

4.5.1.5 Newbler (v2.3 in v2.6)

S programoma Newbler 2.3 in Newbler 2.6 smo pridobili nizko število kontigov. Newbler 2.3 je sekvence združil v 13.530 kontigov v skupni dolžini 8.439.420 bp, Newbler 2.6 pa v 15.224 kontigov v skupni dolžini 8.086.878 bp. Število kontigov, ki so bili daljši ali enaki 500 bp je bilo pri Newbler 2.3 zbirniku 9.004, pri Newbler 2.6 zbirniku pa 8.038. Medtem ko je bilo število kontigov daljših ali enkaih 1.000 bp pri Newbler 2.3 zbirniku 1.121 in pri Newbler 2.6. zbirniku 694. Dolžina najdaljšega kontiga je bila pri Newbler 2.3. programu 4.336 bp in pri Newbler 2.6 programu 3.456 bp. Število zaporedij, ki se niso združila je bilo pri obeh programih visoko. Pri programu Newbler 2.3 je bilo število posameznih zaporedij 77.773 v skupni dolžini 18.103.598 bp in pri programu Newbler 2.6 pa je bilo število posameznih zaporedij 73.087 v skupni dolžini 17.523.948 bp. Čas ki sta ga oba programa porabila za združevanje je znašal 30 min (Preglednica 4). N50 vrednost je pri Newbler 2.3 zbirniku znašala 687 bp, pri Newbler 2.6 zbirniku pa 640 bp. BlastX primerjava z ne-redundantno (NR) proteinsko bazo podatkov z E-vrednostjo <e^-10 je dala 10.346 zadetkov pri programu Newbler 2.3 in 10.168 zadetkov pri programu Newbler 2.6. BlastX primerjava z rastlinsko UniProt bazo z E-vrednostjo <e^-10 pa je dala 10.391 zadetkov pri programu Newbler 2.3 in 10.262 pri programu Newbler 2.6. BlastX rezultati kažejo, da je okoli 84% kontigov pridobljenih z Newbler 2.3 zbirnikom imelo zadetke v obeh podatkovnih bazah, od tega je bilo v NR proteinski bazi 57.1% edinstvenih zaporedij in v rastlinski UniProt bazi 51.8% edinstvenih zaporedij. S programom Newbler 2.6 je 75.8 % kontigov v NR proteinski bazi in 77 % kontigov v rastlinski UniProt bazi imelo zadetke. Od tega je bilo v NR proteinski bazi 66.9 % edinstvenih zadetkov, v rastlinski UniProt bazi pa 59.3 % edinstvenih zadetkov. Isti zadetek se je pri rezultatih pridobljenih z Newbler 2.3 pojavil največ 10 krat v NR bazi in 14 krat v UniProt bazi. Kontigov, ki so imeli 70 % -

100 % prileganje s proteini, je bilo v obeh bazah 14.6 %. Pri rezultatih pridobljenih s programom Newbler 2.6 pa se je isti zadetek pojavil največ 14 krat v NR bazi in največ 12 krat v UniProt bazi. Kontigov, ki so imeli 70 % do 100 % ujemanje s proteini je bilo v NR bazi 11.8 % in v UniProt bazi 11.7 % (Preglednica 6).

4.5.1.6 CLC

S CLC zbirnikom smo pridobili 32.138 kontigov s skupno dolžino 14.646.256 bp, maksimalno dolžino 3.142 bp, povprečno dolžino 455.7 bp in mediano 414 bp. Število kontigov, ki so bili daljši ali enaki 500 bp je bilo 11.305, število kontigov daljših ali enakih 10.00 bp pa 1.005. Zaporedij, ki se niso združila v kontige, je bilo 52.611 v skupni dolžini 12.818.683 bp. Proces združevanja je bil pri programu CLC končan v 5 minutah (Preglednica 5). Naredili smo tudi BLASTX primerjavo proti ne-redundantni (NR) proteinski bazi in rastlinski UniProt bazi z E vrednostjo <e^-10 in pridobili 18.417 zadetkov pri NR bazi in 18.143 zadetkov pri UniProt bazi. Primerjava je pokazala, da je 67.8% kontigov imelo zadetke v obeh bazah podatkov, od tega je bilo v NR bazi 65.1 % edinstvenih zaporedij, v rastlinski UniProt bazi pa 54.2 % edinstvenih zaporedij. Kontigov, ki so imeli 70 % - 100 % dolžinsko ujemanje s proteini, je bilo v NR bazi 6.7 % in v ratlinski UniProt bazi 6.6% (Preglednica 6).

4.5.2 Medsebojna primerjava rezultatov zlaganja transkriptoma

Sledila je medsebojna analiza zastopanosti združenih zaporedij v vsaki skupini. Ideja tega načina primerjave je v tem, da odkrijemo zbirnik, ki najde največ zaporedij v primerjavi z ostalimi zbirniki. Za primerjavo smu uporabili program BLAT, ki izvede globalne primerjave zaporedij med sabo. Analiza je pokazala, da zbirnik iAssembler zavzame zaporedja tudi preostalih zbirnikov, saj je število nazastopanih zaporedij manjše 500 v primerjavi z vsemi ostalimi zbirniki. Najslabše sta se odrezali obe verziji programa Newbler, s preko 10000 zaporedij, ki jih nimata zastopanih v primerjavi z vsemi ostalimi zbirniki. Število nezastopanih zaporedij je manjše od 10000 le pri primerjavi obeh Newbler zbirnikov med seboj. Ostali štirje zbirniki (TGICL, MIRA, PAVE in CLC) so bili po številu nezastopanih zaporedij med seboj primerljivi (Preglednica 5).

4.5.3 Izbira najprimernejšega zbirnika

Upoštevali smo različne kriterije pri izbiri najboljšega programa za združevanje zaporedij, kot so meritve dolžin zaporedij in število združenih zaporedij pri posameznih programih za združevanje, BLAT primerjava posameznih zbirnikov, ter BLASTX primerjava posameznih zbirnikov. Glede na dolžino združenih zaporedij se je najbolje izkazal program iAssembler, saj je sestavil najdaljše kontige in dosegel 7 točk. Sledila sta mu programa PAVE in MIRA , ki sta dosegla 5,5 točk, najkraše dolžine združenih zaporedij pa sta imela

oba programa Newbler in sta zato dosegla le 1,5 točke. Po številu združenih zaporedij so bili programi iAssembler, MIRA in PAVE enakovredni in so vsi trije dosegli po 6 točk. Najslabše sta bila zopet ocenjena programa Newbler, saj sta združila najmanjše število zaporedij. Pri BLAT mapiranju je program iAssembler dosegel vseh 7 točk, saj je zajel tudi zaporedja preostalih zbirnikov. Tudi v tem kriteriju sta bila najslabša programa Newbler, ki sta zopet dosegla le 1,5 točke. Glede na zadnji kriterij sta se najbolje izkazala programa CLC in Newbler 2.6 z doseženimi 6,5 točkami. Pri BLASTX primerjavi sta bila najslabša programa iAssembler in MIRA (Preglednica 7).

Z upoštevanjem vseh parametrov analize zbirnikov smo določili, da program iAssembler zajame največjo skupino transkriptov oljke.

Preglednica 7: Določitev najboljšega programa za združevanje zaporedij z upoštevanjem vseh kriterijev ocenjevanja.
Table 7: Determining the best program for sequence assembly with regard to all criteria of evaluation.

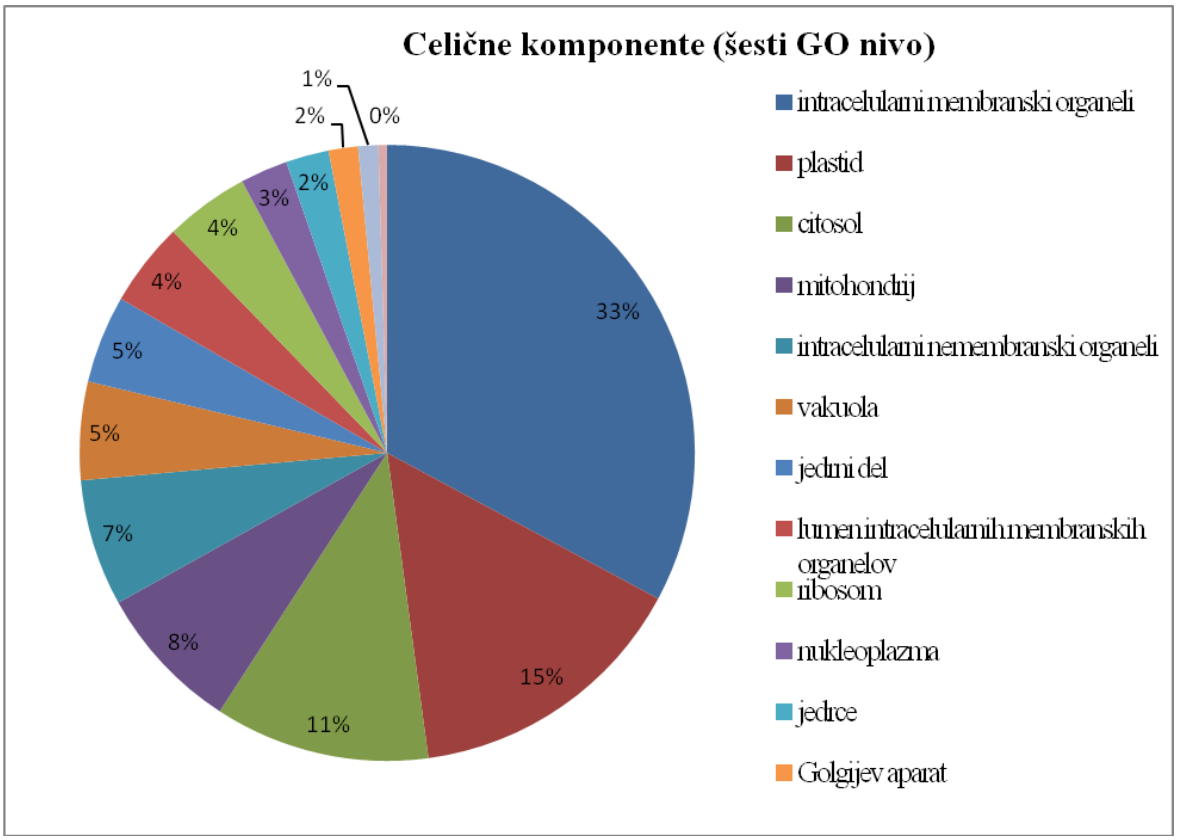
	CLC	iAsse	MIRA	New 2.3	New 2.6	PAVE	TGICL
Meritve dolžin	3,0	7,0	5,5	1,5	1,5	5,5	4,0
Št. združenih zaporedij	4,0	6,0	6,0	1,5	1,5	6,0	3,0
BLAT mapiranje	3,0	7,0	5,0	1,5	1,5	6,0	4,0
BLASTX primerjava	6,5	1,5	1,5	4,5	6,5	3,0	4,5
SKUPNO ŠT. TOČK	16,5	21,5	18	9,0	11	20,5	15,5
ZASEDENO MESTO	4,0	1,0	3,0	7,0	6,0	2,0	5,0

4.6 FUNKCIJSKA ANALIZA

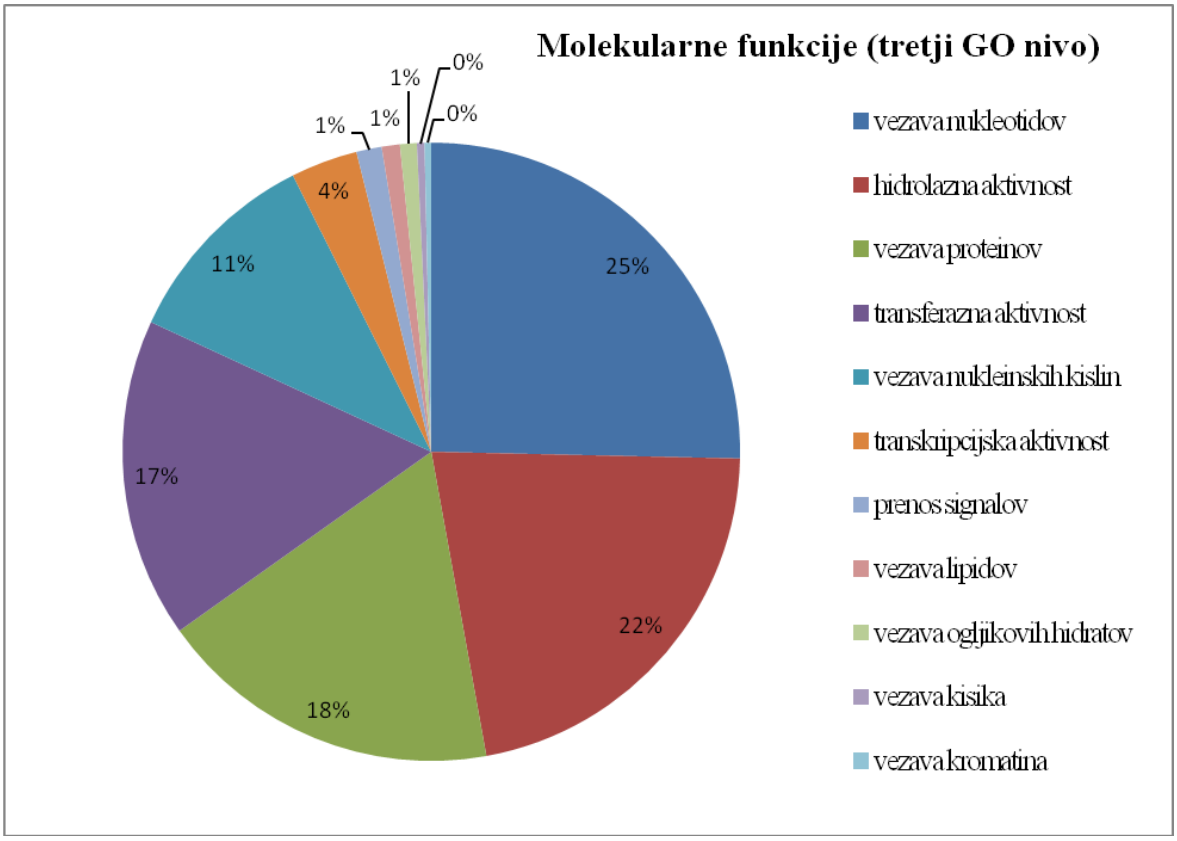
Rezultate programa iAssembler za združevanje zaporedij, ki smo jih izbrali kot najboljše, smo uporabil pri nadaljnji funkcijski analizi podatkov s programom Blast2go (Gotz in sod., 2011), s pomočjo katerega smo 25.451 zaporedjem (51 %) uspešno pripisali vlogo na ravni bioloških procesov, celičnih komponent in molekularnih funkcij. Pri celičnih komponentah je bilo največ anotiranih zaporedij na šestem GO nivoju, od tega je več kot 50% vlog pripisanih intracelularnim membranskim organelom, plastidom in citosolu, v nižjih deležih pa sledijo vloge pripisane mitohondrijem, intracelularni organelom, ki niso povezani z membrano, vakuolam, ter ostalim celičnim komponentam. Pri molekularnih funkcijah je bilo sekvencam največ vlog pripisanih na tretjem GO nivoju, od tega je bilo največ vlog pripisanih vezavi nukleotidov (25 %), hidrolazni aktivnost (22 %), vezavi proteinov (18 %), transferazni aktivnosti (17 %), vezavi nukleinskih kislin (11 %), ter v manjših deležih

transkripcijskim faktorjem, vezavi lipidov, ogljikovih hidratov in kisika. Na ravni bioloških procesov je bilo prav tako sekvencam pripisanih največ vlog na tretjem GO nivoju, in sicer kar 30. Najpogostejše vloge so bile povezane s primarnim metabolnim procesom (14 %), celičnim metabolnim procesom (7 %), makromulekularnim metabolnim procesom (7 %), biosintetičnim procesom (11 %), odgovorom na stres (8 %), katabolnim procesom (8 %), ter v manjših deležih z ostalimi biološkimi procesi. Kar 11.312 zaporedij je imelo BLAST ujemanje s podatkovno bazo trte (*Vitis vinifera*), oljka (*Olea europaea*) je bila šele na 11 mestu s 372 zaporedji.

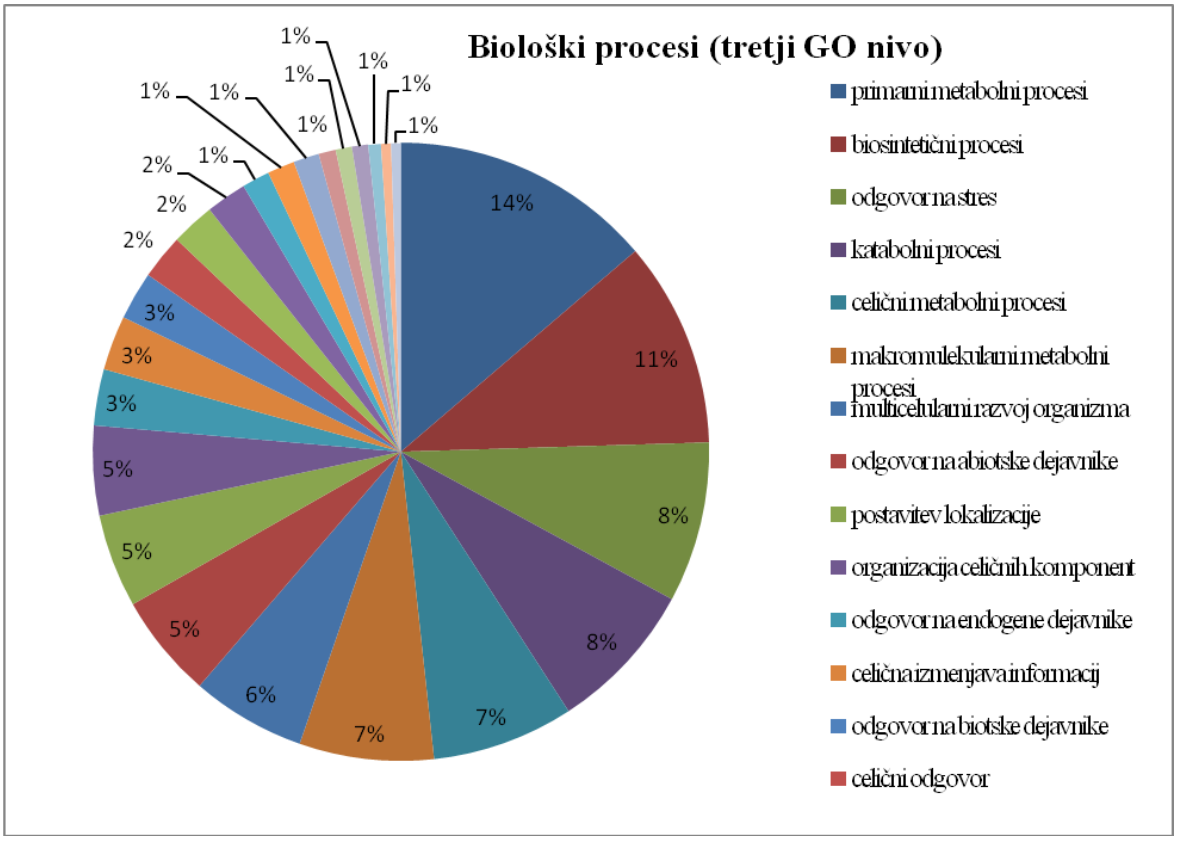
Z Blast2go anotacijo smo pridobil tudi podatke o skupinah genov, ki so vključeni v procese sekundarnega metabolizma (950 zaporedij, GO: 0019748), metabolne procese maščobnih kislin (47 zaporedij, GO: 0006631), v biosintezne procese maščobnih kislin (305 zaporedij, GO: 0006633), v procese nesaturiranih maščobnih kislin (22 zaporedij, GO: 0006636), ter metabolne (99 zaporedij, GO: 0006629) in biosintezne (30 sekvenc, GO: 0008610) procese lipidov. KEGG enciklopedijo (Kyoto Enzyclopaedia of Genes and Genomes) pa smo uporabili za iskanje zaporedij, ki kodirajo encime, ki so vkjučeni v sheme metabolnih in biosinteznih poti, ki so povezane s maščobnimi kislinami in sekundarnim metabolizmom.



Slika 17: Funkcijska analiza podatkov s programom Blast2go (Gotz in sod., 2011) na ravni celičnih komponent.
Figure 17: Functional data analysis program Blast2go (Gotz et al., 2011) on the level of cellular components.



Slika 18: Funkcijska analiza podatkov s programom Blast2go (Gotz in sod., 2011) na ravni molekularnih funkcij.
Figure 18: Functional data analysis program Blast2go (Gotz et al., 2011) on the level of molecular function.



Slika 19: Funkcijska analiza podatkov s programom Blast2go (Gotz in sod., 2011) na ravni bioloških procesov.
Figure 19: Functional data analysis program Blast2go (Gotz et al., 2011) on the level of biological process.

4.7 qPCR ANALIZA

4.7.1 Validacija referenčnih genov

Celokupna RNA je bila izolirana iz oljčnih plodov v dvanajstih točkah razvoja in zorenja plodov. Komercialni Spectrum Total Plant RNA Extraction Kit se je izkazal za primernega (intaktnost ribosomlanih fragmentov, ustrezno A260/A280 razmerje). RNA vzorce smo uporabili za sintezo cDNA. Začetne oligonukleotide za 29 potencialnih referenčnih genov in 4 tarčne gene smo izdelali z uporabo zaporedij iz GeneBank podatkovne baze in našega 454 transkriptoma oljčnih plodov (Preglednica 8). Dolžine amplikonov so bile od 90 do 100 bp (Preglednica 8).

Preglednica 8: Začetni oligonukleotidi za 29 referenčnih genov in 4 tarčne gene vključene v metabolizem maščobnih kislin s predvideno dolžino ampliciranja in predvideno efektivnostjo
Table 8: Developed primer sequences for 29 reference genes and 4 target genes involved in fatty acid metabolism with predicted amplicon lengths and efficiencies

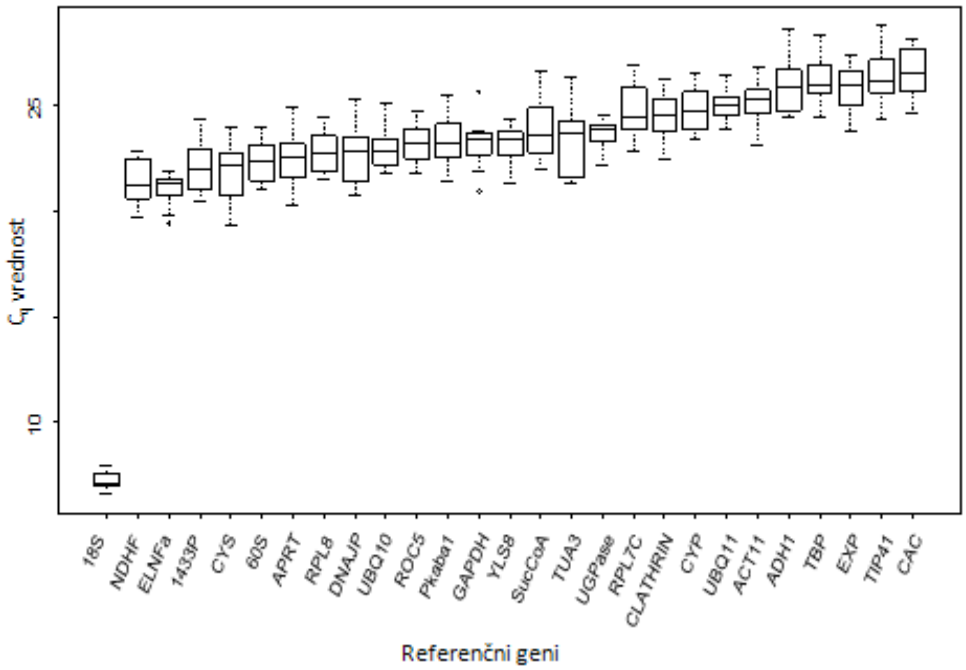
Ime gena	Okrajšava	Sekvenca začetnega oligonukeotida 5'-3'	Dolžina amplikona v bp	Efektivnost (%)
TIP41-sorodni protein	TIP41	CAACGGTGTCTCTCTTTTGACAGT TCATAAGCACTCCATCCACTCTCA	98	91
TATA vezavni protein	TBP	GAGAACAATCTTCCCTGAGACAAAA TATGAACCAGAACTATTCCCTGGAT	90	105
Protein mRNA kinaze	Pkaba1	GGAGAATACCCTTCTGGATGGA GGCTTTGAATGCAGCAGAGAT	90	101
Adenine fosforibozil transferaza	APRT	CCGATAGCCAACGCAATTG TGAGAGATACACGGGCCAAAA	90	91
Klatrin adaptor	CLATHRIN	TTTTGCCCCGAAGACACTCT GAGTAAATCTTCCATTTCGGGTACTG	97	93
Domnevni sucinil-CoA ligaza	SucCoA	TGGGAGACAAACCATCAACCA CATCCGGAGTTGATCATTAAGGT	91	93
14-3-3 protein	1433P	ACAAGTCTGCTCATGATATTGCATTA AATAGAAAACAGAGAAGTTAAGTGC AAGTC	90	93
Yellow leaf specific gene 8 mRNA	YLS8	GGTAGACCGTCTCGACGATGTC ATGATTGATCTTGGCACTGGAA	91	98
Ribosomal protein L2	RPL8	TAGCAGCAGCTTGACCACGTA GTACTGTTCGTCGGGATGCA	90	101
60S ribosomalni protein	60S	TAGCAGCAGCTTGACCACGTA GTACTGTTCGTCGGGATGCA	90	94
Rotamaz ciklofilin 5	ROC5	TTTCTCAATGGCTTTTACCACATC TGTACTGCGAAAACCGAATGG	90	103
Poliubikvitin 11	UBQ11	TCAAGGCTAAGATCCAGGACAAG CCAAAGTCCTTCCATCATCCA	90	108
Ciklofilin	CYP	ACTCTCCACCGGTGCCATT TCACTCAAAGGCTCGGCTTT	90	94
60S ribosomalni protein L7	RPL7C	ATCTGCATGGAAGATCTTGTTCAC CCCAATGGCGCCTTCA	100	101
NADH dehidrogenazna poenota F	NDHF	TTCGCCGATTTTCGCAATA TGCCCCTCAAAAGTAAGTAAATAGAT C	90	97
Elongacijski faktor 1-α	ELNFa	CTCACGTTCAGCCTTGAGCTT TGTGATTGAGAGGTTTGAGAAGGA	94	99

Se nadaljuje.

Nadaljevanje.

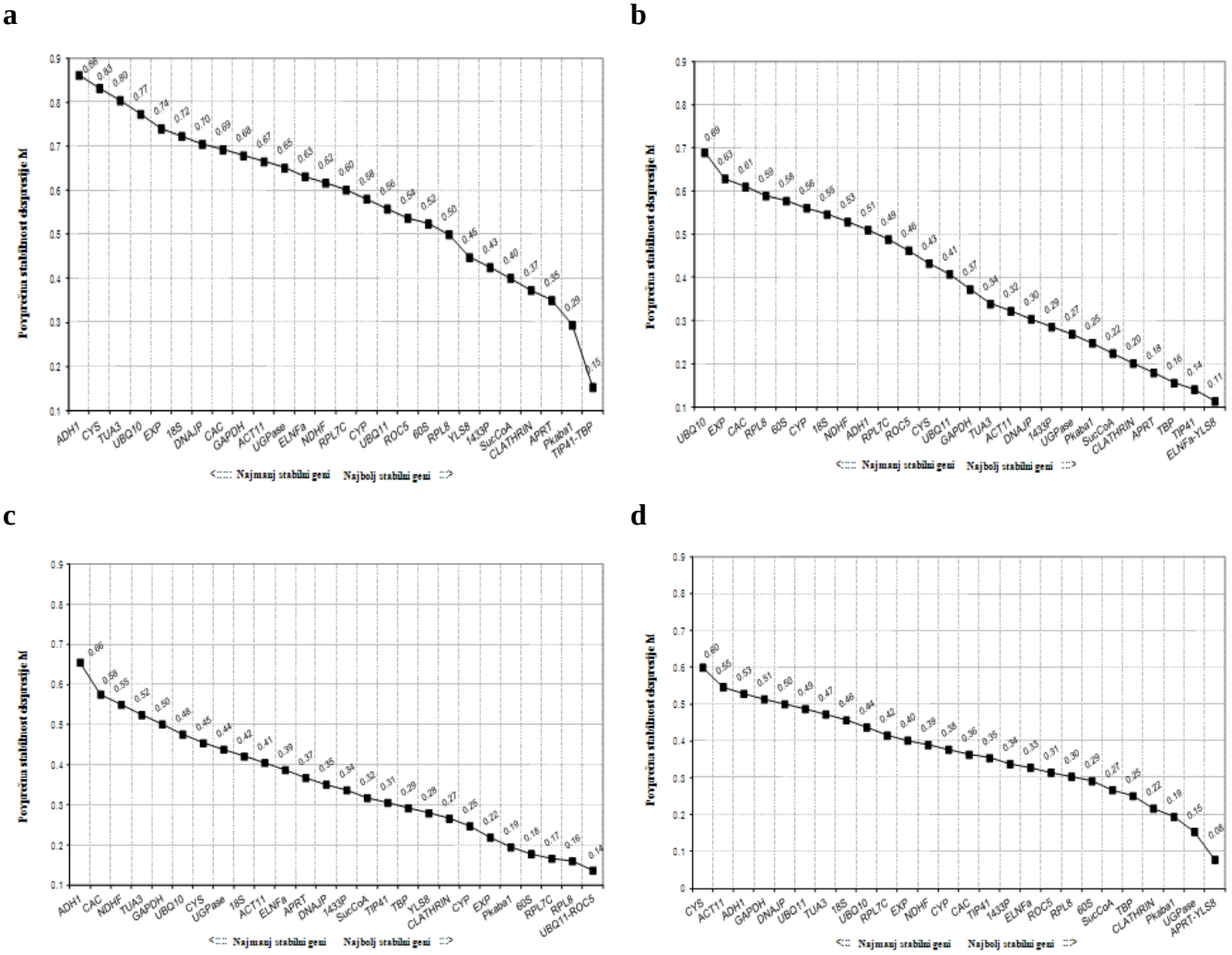
Ime gena	Okrajšava	Sekvenca začetnega oligonukeotida 5'-3'	Dolžina amplikona v bp	Efektivnost (%)
(UDP)-glukozna pirofosforilaza	UGPase	CCATGCATAAAAGATGCTGGAA TTCAAGCTTGCCACTATTCATCA	90	96
Aktin 11	ACT11	CCCAAGGCCAACAGAGAGAA GGAAAGAACGGCCTGAATAGC	90	93
Gliceraldehide-3-fosfat dehidrogenaza	GAPDH	AGCCTTGTCCTTGTCCGTAAAG TTCAGGAATCCGGAGGAGATT	90	99
Klatrin adaptor	CAC	GGCCACCTATTCAGATGGAATTT TTGTATCCACTCCTTTCCCATACC	94	102
DnaJ-podobni protein	DNAJP	CATCAGCCTCGCCAGGAA AGGTTGTGCAGGAGAAGAAGGT	90	91
18S ribosomalna RNA	18S	GGGCTCGAAGACGATCAGATAC CCGGCGGAGTCCTATAAGC	90	91
Izraženo zaporedje SGN-U346908, neznani protein	EXP	TCTCCGATGGGCAATAAACC TATGAATGTTGTATGGCCTGTTTGA	91	107
Poliubikvitin 10	UBQ10	GACAATGTCAAGGCAAAAATTCAG ACCATCCTCAAGCTGCTTACCA	90	98
Alfa tubulin	TUA3	CTGGAACTCGGTAACATCCACAT TTGAACCGGTTGATTTCTCAGA	90	101
Cistein proteinaza	CYS	ACACTGAAGAAGATTACCCCTACACA GTCTTCATAACCATCAATGGACACA	95	103
Alkohol dehidrogenaza 1	ADH1	GATGGGTCTTAAACTCTGCATCTTTA TGATCTCGGCATTTGAATGTG	90	97
Beta-tubulin	TUBb	TACGAAGAGTTCTTGTTTTGAACGTT CCTCACTGCCTCAGCCATGT	90	/
SAND sorodni protein	SAND	CCCAACCCCAAGAAAATTTCA TTTTGATCCCCTTGCTGACAA	99	/
Stearoil-ACP desaturaza	SAD1	GGGCCACTTTCATTTCTCATG CGGCAATTGTACCACATATTTGA	90	94
Maščobna acil-ACP tioesteraza A	FatA	TGAAGAGGATAATGCTAGCCTGAA TCAGCTCGTCTTGGCACAAG	90	98
Acil-CoA tioesterazni sorodni protein	Acot	GAGGCAATACAAAACGGGAATG CATTTCTGCCACCTGGTGATT	90	98
Loksigenaza 1	LOX1	CCGATGAATGGCTTGACAAA CGGATAAGGTTTTCCTGAAGACA	90	108

V postopku optimizacije smo določili, da je optimalna koncentracija primerja 300 nM in količina cDNA 0.5 ng/1µl, saj je bila ob takih koncentracijah Ct vrednost najnižja s primerno učinkovitostjo. Sedemindvajset parov začetnih oligonukleotidov potencialnih referenčnih genov in vsi štirje pari primerjev tarčnih genov so proizvedli amplikon z enojno disociacijsko krivuljo. Dva potencialne referenčna gena (TUBb in SAND) se nista pomnožila in sta bila zato izključena iz nadaljnje analize.



Slika 20: Kvantilni diagram predstavlja Cq vrednosti za 27 potencialnih referenčnih genov. Polna črta pradstvalja srednjo vrednost (mediana), škatle predstavljajo kvantila 0,25 in 0,75, repki predstavljajo percentila 10 in 90, medtem ko točke predstavljajo osamelce.
Figure 20: Box plot presentation of quantitative cycle values (C_q) for 27 analysed reference genes. The solid line represents the median value, boxes are 25 and 75 percentiles, while whiskers are 10 and 90 percentiles and dots are outliers.

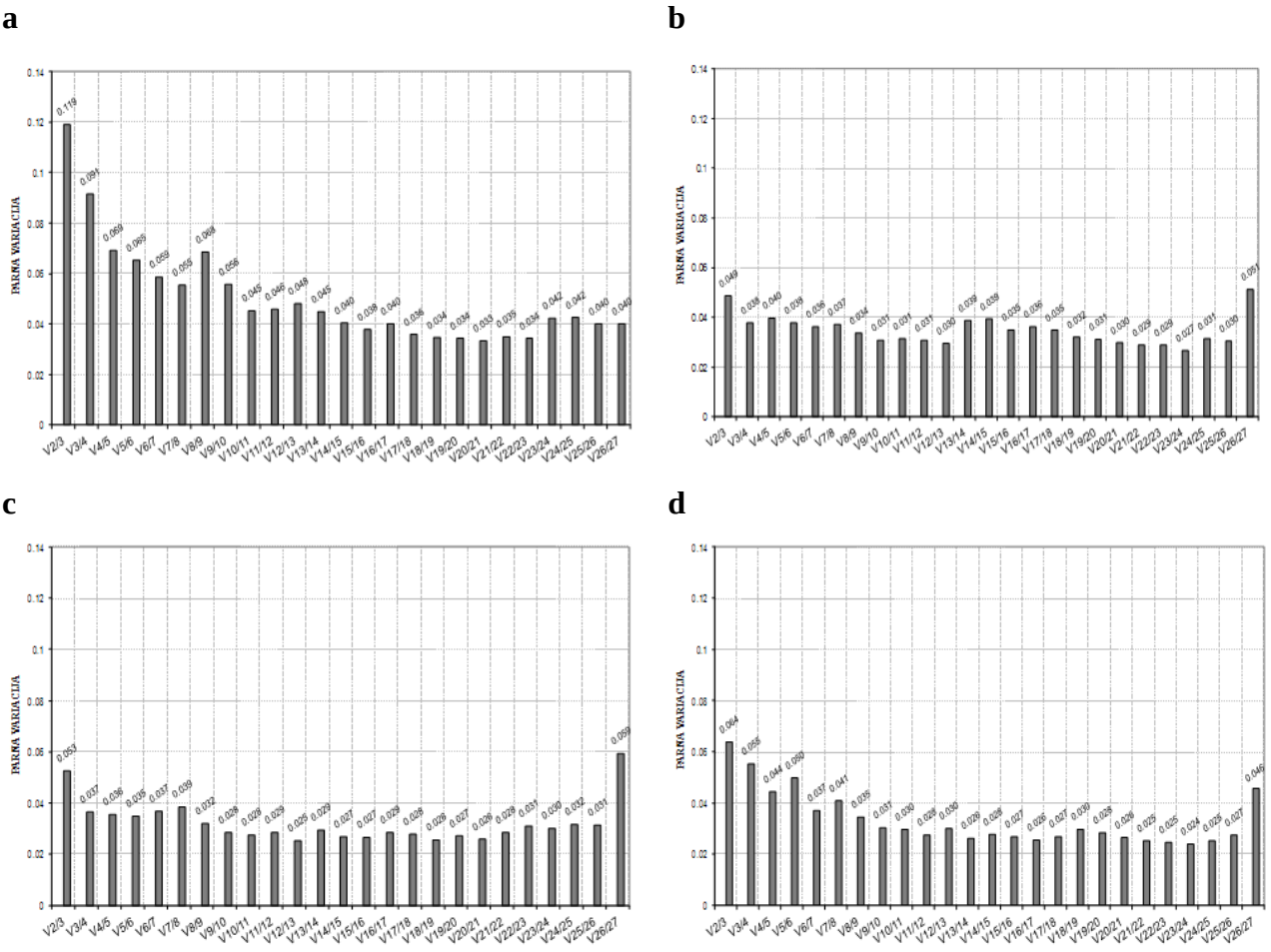
Za vsak referenčni gen, smo Cq predstavili v obliki kvantilnega diagrama (Slika 21), ki kaže relativno številčnost posameznih transkriptov. Cq vrednosti se nahajajo v razponu od 7.22 (18S) do 26.5 (CAC). Zelo nizko Cq vrednost smo dobili pri pogosto uporabljenemu 18S referenčnemu genu, medtem ko so bile pri ostalih referenčnih genih Cq vrednosti v razponu od 21.4 [NADH dehidrogenazna podenota F (NDHF)] do 26.5 (CAC). Razlika predstavlja 34-krat večjo številčnost CAC referenčnega gena nad NDHF referenčnim genom. Povprečna stabilnost ekspresije gena (M vrednost) vseh 27 kandidatnih genov je bila izračunana s programom geNorm (Vandesompele in sod., 2002) kot skupna vrednost vseh 12 vzorčnih točk, ter kot skupna vrednost točk 1-4, 5-8 in 9-12 (Slika 21).



Slika 21: Povprečna stabilnost ekspresije (M vrednost) 27 referenčnih genov oljke, izračunana z geNorm algoritmom za a) vseh 12 vzorčnih točh, b) vzorčne točke od 1 do 4, c) vzorčne točke 5 do 8, d) vzorčne točke od 9 do 12

Figure 21: Average expression stability value M of 27 evaluated olive candidate reference genes as calculated with geNorm algorithm; a) for all 12 sampling points; b) sampling points from 1 to 4; c) sampling points from 5 to 8; d) sampling points from 9 to 12

Kot lahko vidimo na Sliki 22a sta bila gena TIP41 in TBP (oba M = 0.15) definirana kot najbolj stabilna referenčna gena pri vseh 12 vzorčnih točkah. Pri vzorčnih točkah od 1 do 4 (Slika 22b, TIP41 M = 0.14, TBP M = 0.16) sta zavzela tretje in četrto mesto, pri vzorčnih točkah od 5 do 8 (Slika 22c, TBP M = 0.29, TIP41 M = 0.31) sta zavzela enajsto in dvanajsto mesto, pri vzorčnih točkah od 9 do 12 (Slika 22d, TBP M = 0.25, TIP M = 0.35) pa šesto in trinajsto mesto.



Slika 22: Parna variacija ($V_{n/Vn+1}$) med normalizacijskim faktorjem NFn in normalizacijskim faktorjem NFn+1 za določitev optimalnega števila referenčnih genov, ki so potrebni za normalizacijo. Prvi stolpec predstavlja parno variacijo med NF vrednostjo določeno za prva dva najboljša referenčna gena in NF vrednostjo določeno za prve tri najboljše referenčne gene (kot si sledijo na Sliki 18); za a) vseh 12 vzorčnih točh, b) vzorčne točke od 1 do 4, c) vzorčne točke 5 do 8, d) vzorčne točke od 9 do 12.

Figure 22: Pair-wise variations of $V_{n/n+1}$ value between the normalization factors NF_n and NF_{n+1}, used to determine the optimal number of reference genes for normalization. The first bar value represents the pair-wise variation between the NF value assessed for the two best genes and the NF value assessed for the best three genes (as ordered in Figure 18), followed by the addition of subsequent reference genes as listed in Figure 18; a) for all 12 sampling points; b) sampling points from 1 to 4; c) sampling points from 5 to 8; d) sampling points from 9 to 12.

Določili smo tudi parno variacijo (V_n/V_{n+1}) med normalizacijskim faktorjem NFn in normalizacijskim faktorjem NF_{n+1}, da bi dobili optimalno število referenčnih genov, ki so potrebni za normalizacijo. Slika 23 kaže, da je kombinacija dveh referenčnih genov (*TIP41* in *TBP*) v našem primeru zadostovala za normalizacijo, saj je bila V2/3 vrednost 0.119, kar je nižje od priporočene mejne vrednosti 0.15. Najnižja skupna vrednost parne variacije je bila dosežena s kombinacijo 20 kandidatnih genov (V20/21) in je bila 0,033. Vrednosti parne variacije so pokazale, da sta referenčna gena *TIP41* in *TBP* primerna tudi za

normalizacijo vzorcev, v razdeljenih periodah razvoja oljčnih plodov (V2/3 vrednosti 0.049, 0.053 in 0.064; Slika 23b, c, d).

Razvrstitev referenčnih genov, ki temelji na vseh štirih algoritmih, ki se uporabljajo v programu RefFinder (Xie et al. 2011), navaja TBP in YLS8 kot najbolj stabilna gena (Tabela). Če primerjamo to razvrstitev glede na rezultate geNorm-a vidimo, da je gen TBP pri obeh analizah eden od dveh najbolj stabilnih genov, medtem ko sta TIP41 in YLS8 razvrščena nekoliko drugače, vendar sta pri obeh analizah med najboljšimi. Z izjemo BestKeeper algoritma, so imeli geni TUA3, CYS in ADH1 pri vseh ostalih algoritmih in skupni razvrstitvi najslabšo stabilnost.

Preglednica 9: Preglednica referenčnih genov razporejenih s programom RefFinder (Xie in sod., 2011)
Table 9: Scoring table of reference genes using RefFinder (Xie et al., 2011)

GeNorm		**NormFinder**		**BestKeeper**		**Delta Cq**		**Comprehensive ranking**	
Ime gena	Stabilnost	Ime gena	Stabilnost	Ime gena	Stabilnost	Ime gena	Povprečje Stdev	Ime gena	Geome. sredina
TIP41	0.222	YLS8	0.292	18S	0.330	TBP	0.69	TBP	2.17
TBP	0.222	TBP	0.311	UBQ11	0.511	YLS8	0.69	YLS8	3.03
Pkaba1	0.321	Pkaba1	0.373	ELNFa	0.562	Pkaba1	0.72	Pkaba1	4.56
APRT	0.367	UBQ11	0.398	UGPase	0.568	APRT	0.74	UBQ11	5.53
CLATH	0.387	1433P	0.401	UBQ10	0.704	CLATH	0.74	TIP41	5.91
YLS8	0.415	APRT	0.413	ROC5	0.720	1433P	0.74	APRT	6.45
CAC	0.435	CLATH	0.422	YLS8	0.770	ROC5	0.74	CLAT	7.79
1433P	0.463	ROC5	0.441	GAPDH	0.784	TIP41	0.75	ROC5	7.97
SucCoA	0.482	TIP41	0.446	ACT11	0.798	UBQ11	0.76	1433P	8.32
RPL8	0.531	RPL8	0.502	CYP	0.809	RPL8	0.77	ELNFa	8.82
60S	0.557	60S	0.521	TBP	0.835	60S	0.78	RPL8	10.47
ROC5	0.572	ELNFa	0.542	RPL8	0.847	ELNFa	0.82	18S	10.5
UBQ11	0.591	CAC	0.588	EXP	0.859	CAC	0.84	60S	11.68
ELNFa	0.613	GAPDH	0.589	60S	0.878	SucCoA	0.85	UGPase	12.18
CYP	0.633	SucCoA	0.609	NDHF	0.904	GAPDH	0.86	CAC	12.84
RPL7C	0.649	DNAJP	0.658	Pkaba1	0.929	DNAJP	0.88	GAPDH	13.54
NDHF	0.664	UGPase	0.680	TIP41	0.973	CYP	0.89	CYP	14.64
UGPase	0.682	CYP	0.690	APRT	0.976	UGPase	0.90	SucCoA	14.89
ACT11	0.697	EXP	0.700	ADH1	0.979	EXP	0.93	UBQ10	16.21

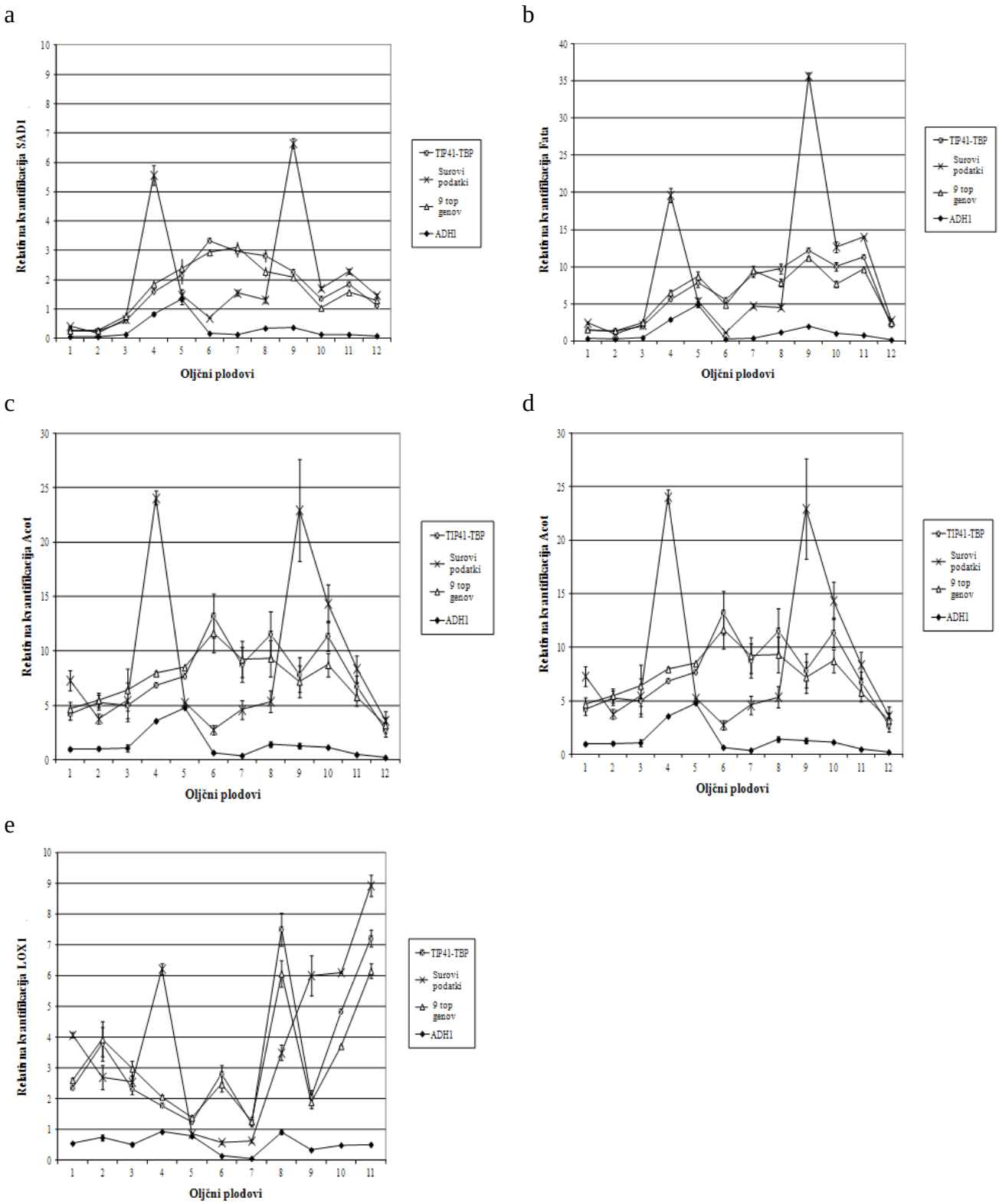
Se nadaljuje.

Nadaljevanje.

GeNorm		**NormFinder**		**BestKeeper**		**Delta Cq**		**Comprehensive ranking**	
Ime gena	Stabilnost	Ime gena	Stabilnost	Ime gena	Stabilnost	Ime gena	Povprečje Stdev	Ime gena	Geom. sredina
GAPDH	0.710	NDHF	0.744	1433P	0.986	NDHF	0.94	ACT11	16.77
DNAJP	0.725	ACT11	0.744	CHLAT	1.004	RPL7C	0.94	NDHF	17.87
EXP	0.741	RPL7C	0.764	RPL7C	1.078	ACT11	0.94	EXP	17.93
18S	0.758	18S	0.785	CAC	1.101	18S	0.99	DNAJP	18.95
UBQ10	0.789	UBQ10	0.950	DNAJP	1.122	UBQ10	1.13	RPL7C	20.08
TUA3	0.817	CYS	0.977	CYS	1.135	TUA3	1.14	ADH1	24.73
CYS	0.842	TUA3	0.992	SucCoA	1.213	CYS	1.14	CYS	25.50
ADH1	0.873	ADH1	1.099	TUA3	1.375	ADH1	1.26	TUA3	25.74

4.7.2 Nivo ekspresije genov FatA, SAD1, Acot in LOX1

Nivo ekspresije treh metabolnih genov je bil določen z uporabo (a) dveh referenčnih genov TIP41 in TBP, z uporabo (b) devetih kandidatnih genov, ki so imeli M vrednost (izračunano na podlagi vseh 12 vzorcev) pod priporočeno vrednostjo 0.5 in z uporabo (c) kandidatnega gena ADH1, ki je imel najslabšo M vrednost. Rezultati so pokazali, da so bili geni FatA, SAD1, Acot in LOX1 izraženi v vseh stopnjah razvoja plodu, z različnimi vrednostmi relativne kvantifikacije in profili izražanja (Slika 20).



Slika 23: Relativna ekspresija genov FatA, SAD1, Acot in LOX1
Figure 23: Relative expression of FatA, SAD1, Acot in LOX1 genes

5 RAZPRAVA IN SKLEPI

Oljka (*Olea europaea* L.) je šesta najpomembnejša rastlina na svetu glede produkcije olja. Izvira iz sredozemske regije in se trenutno širi tudi na druga gojitvena območja predvsem zaradi njene visoke gospodarske vrednosti. V slovenskih oljčnih nasadih je 'Istrska belica' najbolj zastopana sorta. K nagli širitvi v istrske oljčnike po pozebi leta 1956 so pripomogle številne pozitivne lastnosti, med drugimi odpornost na nizke temperature, samooplodnost, ter dobra in redna letina (Bandelj Mavsar in sod., 2005). 'Istrska belica' je poznana po visoki vsebnosti skupnih biofenolov, ki je lahko tudi za dvakrat večja v primerjavi s sorto 'Leccino' (383 mg/kg v primerjavi s 156 mg/kg, višja pa je tudi v primerjavi s štirimi drugimi italijanskimi sortami (Uccella, 2000). Za deviška oljčna olja pridelana iz sorte 'Istrska belica' poročajo o vrednostih skupnih biofenolov tudi do 600 mg/kg.

Oljčni plodovi v razvoju se spreminjajo v velikosti, sestavi, barvi, teksturi, okusu in odpornosti na okoljske dejavnike. Vsi ti procesi so genetsko regulirani, na njih pa vplivajo tudi okoljski dejavniki. Za opredelitev in označitev genov, ki sodelujejo v teh procesih v plodovih, so se razvila različna genomska orodja (izražena nukleotidna zaporedja, mikromreže, itd.) (Seymour in sod., 2008). S pomočjo genomskega pristopa bi uspeli določiti izražanje genov tudi v različnih stopnjah razvoja oljčnih plodov. Izražena nukleotidna zaporedja (ESTs) in cDNA zaporedja so ena izmed primarnih orodij, ki nam zagotovijo neposredne informacije o transkriptih, ki kodirajo dele genoma in so trenutno najpomembnejši viri za raziskovanje transkriptoma (Nagaraj in sod., 2006). EST baze podatkov so prav tako zelo koristno orodje za odkrivanje genov in markerjev, gensko kartiranje in funkcionalne študije pri preučevanih organizmih (Ozgenturk in sod., 2010).
Zaradi pozitivnih lastnosti sorte 'Istrska belica', predvsem pa dejstvo, da ima najvišje vsebnosti biofenolov, smo jo uporabili za izolacijo kodirajočih se nukleotidnih zaporedij iz razvijajočega plodu oljke. Plodove oljk sorte 'Istrska belica' smo vzorčili skozi celotno obdobje razvoja. Vzorčenje je potekalo od začetka junija, ko se zaključi faza cvetenja do sredine novembra, ko se prične obiranje plodov oz. do nastopa fiziološke zrelosti. Za vzorčenje smo izbrali drevesa v pravilno oskrbovanem in negovanem nasadu. Plodove smo vzorčili tedensko, takoj po obiranju pa jih zamrznili v tekočem dušiku in shranili do uporabe pri -80 ^{o}C. S tem smo poskrbeli, da smo dobili najbolj reprezentativen vzorec transkriptov skozi celotno obdobje razvoja plodov. Za izolacijo RNA smo uporabili Spectrum Plant Total RNA Extraction Kit, ter zmešali ekvimolarne količine vseh vzorcev, da smo dobili združen, reprezentativen vzorec vseh RNA izraženih v celotnem razvojnem obdobju oljčnega plodu. Reprezentativen vzorec RNA smo poslali naprej za izdelavo normalizirane cDNA knjižnice (Evrogen Lab, Russia).

Znotraj celic organizmov prihaja do večjih nihanj v koncentraciji različnih transkriptov. Prepis določene mRNA molekule znotraj celice je odvisen od trenutnih potreb celice oz. organizma. Zato lahko sekvenciranje celotnega evkariontskega transkriptoma zahteva

analizo tudi do 10^8 klonov iz posameznih cDNA knjižnic, da bi dobili redke sekvence, medtem ko bi bili pogosti transkripti večkrat sekvencirani. Metodo, ki zmanjša število pogostih trenskriptov in izravna koncentracije mRNA v cDNA knjižnici, imenujemo cDNA normalizacija. Normalizacija se uporablja za hitrejše odkrivanje genov znotraj cDNA knjižnic in olajša identifikacijo in analizo redkih transkriptov. Ta pristop je nujen za določevanje EST zaporedij celotnega transkriptoma, uporaben pa je tudi pri drugih aplikacijah kot so izgradnja specifičnih RNA knjižnic in funkcijsko pregledovanje (Alberts in sod., 1994).

Normaliziran cDNA vzorec olčnih plodov, ki smo jih vzorčili skozi različne faze razvoja plodu, smo nato sekvencirali s pomočjo ponudnika storitev naslednje generacije določevanja nukleotidnih zaporedij. Uporabili smo *GsuI* tretiran vzorec cDNA, kateremu smo odstranili poli-A regije iz 3' konca. Za slednji vzorec smo se odločili zato, ker se je pri poizkusu s Sangerjevim sekvenciranjem izkazal za boljšega, saj prisotnost homopolimernih A regij ni motila postopka sekvenciranja. Določevanje nuklotidnih zaporedij s pomočjo Sangerjeve metode smo opravili na ABI 3700 kapilarni napravi za določevanje nuklotidnih zaporedij. Primerjali smo rezultate določevanja nukleotidnega zaporedja klonom cDNA in *Gsu*I-cDNA. Iz vsake knjižnice smo 192-im klonom v PCR pomnožili cDNA insert. Pri normalizirani cDNA knjižnici, ki ni bila tretirana z *Gsu*I encimom smo naredili obojestransko sekvenčno reakcijo. Pri tej knjižnici je bila kar polovica sekvenčnih reakcij nezadovoljive kakovostni, prvenstveno zaradi prisotnih poli A regij. Za nadaljnjo analizo je bilo uporabnih le 41 % zaporedij s povprečno dolžino 402 bp. Vseeno pa so rezultati potrdili dobro opravljen proces normalizacije, saj ni bilo presežnih zaporedij v knjižnici. Pri drugi normalizirani cDNA knjižnici, ki je bila tretirana z GsuI encimom, ligirana z adapterji, ter reamplificirana, pa smo naredili enostransko sekvenčno reakcijo. Od 192 zaporedij je bilo kar 81 % primernih za nadaljnjo analizo, kar pomeni, da smo s pomočjo *Gsu*I restrikcijskega encima uspešno odstranili poli A regije in s tem potrdili smiselnost odstranjevanja poli A regije, saj je le manjši delež zaporedij še vseboval take regije.

Pri obeh knjižnicah smo preverili redundantnost, kot merilom identičnosti pa smo uporabili 95 % ujemanje. Število enkratnih zaporedij je bilo pri obeh knjižnicah visoko , skupne dolžine enkratnih zaporedij, ter povprečne dolžine enkratnih zaporedij pa so bile primerljive. Glavna razlika med knjižnicama pa je bila v procentu uspešnih sekvenčnih reakcij, ki jih je bilo pri *Gsu*I knjižnici kar 40 % več.

S težavami, ki jih povzročajo poli A regije se soočajo tudi drugi raziskovalci. Soderlund in sod. (2009), ki so pripravili dve normalizirani knjižnici koruze linije B73, so imeli težave z določevanjem nukleotidnih zaporedij zaradi zdrsov, ki so jih povzročali poli A regije. Yang in sod. (2005) so določili strategijo FMA (angl. failure mode analysis), s katero so želeli določiti zakaj občasno prihaja do neuspešnega branja nukleotidnih zaporedij. Napake, ki so nastale pri visoko zmogljivem sekvenciranju so sistematično pregledali, ter

določili vrsto napake in njeno pogostost pojavljanja. Napake so nato razdelili v devet kategorij. Najpogostejše napake so bile povezane s prisotnostjo poli A regij. Takasuga in sod. (2001) so iz cDNA klonov govejega osteonectin-a pridobili 36.310 izraženih nukleotidnih zaporedij (ESTs) z uporabo 10 različnih cDNA knjižnic. Določeno število teh knjižnic je imelo prisotne poli A konce, pri določenih pa so bili le ti odstranjeni z uporabo Nested Deletion Kit (Amersham Pharmacia Biotech). Nato so primerjali določevanje nukleotidnih zaporedij posameznih cDNA knjižnic in ugotovili, da odstranitev poli A regij bistveno izboljša kvaliteto določevanja sekvenc.

Shibata in sod. (2001) so prav tako uporabili metodo s katero so pri cDNA klonih skrajšali poli A regije z uporabo restrikcijskega encima tipa II, kot je encim *Gsu*I. S tem so se izognili težavam pri določevanju zaporedij pri klonih, ki so imeli poli A regije odstranjene ali krajše od sedmih adeninov. Tako so dokazali, da odstranitev poli A regij močno olajša direktno in transkripcijsko določevanje zaporedij, izognemo pa se tudi predhodni pripravi modelov, ki predstavljajo oviro pri določanju zaporedij velikih DNA molekul (Shibata in sod., 2001).

Kvaliteto ne-normalizirane cDNA, normalizirane cDNA , cDNA po restrikciji, ter cDNA knjižnice po restrikciji in čiščenju, smo preverili tudi s pomočjo naprave Agilent Bioanalyzer 2100 in uporabo čipa DNA1000 (Slika 13). Ugotovili smo, da je pri ne-normalizirani knjižnici prisotna večja količina zaporedij z dolžino okoli 1300 bp (Slika 13, vzorec 1), ki smo jih s postopkom normalizacije (Slika 13, vzorec 2) uspešno zmanjšali, vendar pa je ostala delno povečana količina zaporedij okoli 1500 bp. Po obdelavi knjižnice z *Gsu*I encimom se število daljših sekvenc zmanjša, pojavi pa se povečano število kratkih sekvenc dolžine okoli 60 bp (Slika 13, vzorca 3 in 4), kar bi lahko bila posledica večjega števila odrezanih poli A repov. Le-te smo odstranili s čiščenjem s silicijevimi kolonami (Slika 13, vzorca 5 in 6).

Določevanje nukleotidnih zaporedij transkriptoma pri nemodelnih organizmih je postalo popularno, saj je cenovno ugodnejše in računalniško vodljivejše, kakor določevanje nukleotidnega zaporedja celotnega genoma organizma, vendar še vedno prinese dovolj informacij, da izpolnjuje zahteve raziskovalnih skupin. Tradicionalno se je za določevanje izraženih nukleotidnih zaporedij (ESTs) uporabljala Sangerjeva dideoksi metoda, sedaj pa so jo pričele izpodrivati nove generacije tehnologij za določevanje zaporedij, ki imajo večji izkupiček in nižjo ceno na določitev baze. Večina projektov, ki je bilo narejenih na nemodelnih organizmih, je za določevanje nukleotidnih zaporedij uporabila Roche 454 metodo, zaradi daljših prepisov (do 400 bp) in boljših rezultatov pri sestavi in anotaciji zaporedij (Kumar in Blaxter, 2010). Tudi mi smo cDNA knjižnico, ki smo jo tretirali z *Gsu*I encimom, poslali na določevanje nukleotidnega zaporedja s pomočjo Roche 454 tehnologije (GATC Biotech, Konstanz, Germany). Nukleotidno zaporedje smo določili polovici regije pikotiterske plošče, pridobili pa smo 560.578 sekvenc v skupni dolžini

160.414.301 bp. Po koraku razdruževanja konkatemernih zaporedij in po koraku čiščenja zaporedij, smo na koncu pridobili 577.025 zaporedij v skupni dolžini 139.419.844 bp. Kar 99% zaporedij je imelo oceno kakovosti na posamezno bazo nad mejno vrednostjo, kar je kazalo na to, da so pridobljeni podatki o zaporedjih transkriptoma oljke kakovostni in primerni za nadaljno obdelavo.

Pridobljene sekvenčne informacije o transkriptih nato uporabimo v različnih fazah obdelave podatkov kot sta sestavljanje zaporedij za opredelitev domnevnih transkriptov, anotacija sestavljenih podatkov in uporaba le teh. Celotno urejanje informacij transkriptoma ni enostavno, saj posamezna zaporedja lahko vsebujejo napake in polimorfizme, ki onemogočajo njihovo optimalno obdelavo (Kumar and Blaxter, 2010). Programski zbirniki CAP3, MIRA, Newbler, Seqman NGen, CLC bio in EGassembler so zbirniki, ki se najpogosteje uporabljajo pri sestavi podatkov transkriptoma pridobljenih z Roche 454 tehnologijo. Vendar niso vsi ti programi specifično namenjeni za obdelavo podatkov transkriptoma. Za razliko od genoma, ki ga sestavljajo dolgi, neprekinjeni odseki, je transkriptom sestavljen iz mnogih prepisov različnih dolžin. Sestavo zaporedij transkriptoma otežuje tudi neenakomirno izražanje genov v organizmu, kar vpliva na neenakomirno zastopanost različnih transkriptov. Normalizirane cDNA knjižnice sicer zmanjšajo razlike v zastopanosti posameznih transkriptov, vendar ne morejo omogočiti popolnoma enakomerne razporeditve le teh (Mundry in sod., 2012). Zato pri sestavi zaporedij transkriptoma prihaja do dveh pogostih napak. Pri prvi napaki (tip I) so EST-ji pridobljeni iz alternativno združenih transkriptov ali paralogov, nepravilno združeni v en transkrip, pri drugi napaki (tip II) pa EST-ji pridobljeni iz istega transkripta niso uspešno združeni skupaj. Zheng in sod. (2011) so razvili zbirnik, ki naj bi uspešno identificiral napake, ki nastajajo pri združevanju zaporedij, ter jih avtomatsko popravljal. Program se imenuje iAssembler in je sestavljen iz sedmih modulov, ki se delijo na tri kategorije: regulator (splošni regulator), zbirnik (MIRA, CAP3 in megablast) in popravljalnik napak (popravljalnik napak tipa I in tipa II) (Zheng in sod., 2011).

Pri predhodnih analizah transkriptomov, v katerih so podatke o sestavi zaporedij pridobili s pomočjo Roche tehnologije, so pogosto uporabljali le en program za sestavo zaporedij. Samo nekaj študij pa je do sedaj sistematično primerjalo različne programe za združevanje zaporedij, v želji da bi pridobili podatke o optimalni programski opremi. Kumar in Blaxter (2010) so izpeljali sistematično primerjavo petih programov za združevanje zaporedij (CAP3, MIRA, Newbler, SeqMan and CLC), da bi določili najboljši program za združevanje zaporedij transkriptoma, z uporabo nabora podatkov iz parazitskih ogorčic *Litomosoides sigmodontis* (Kumar and Blaxter, 2010). Garg in sod. (2011) pa so vsa kakovostna zaporedja, ki so jih pridobili s 454 pirosekvenciranjem transkriptoma čičerike (*Cicer arietinum*), uporabili pri primerjavi 8 različnih programov (MIRA, Newbler 2.3. in 2.5., CAP3, TGICL, CLC, Velvet, ABySS) za združevanje zaporedij, prav tako z namenom optimizacije postopka za združevanje. Zheng in sod. (2011) pa so med seboj

primerjali delovanje zbirnikov iAssembler, MIRA, CAP3, TGICL, Phrap in Newbler, in sicer pri sestavi EST zaporedij oljke in paradižnika. Njihov glavni cilj je bil ovrednotiti delovanje zbirnika iAssembler v primerjavi z ostalimi zbirniki (Zheng in sod., 2011).

V naši študiji smo primerjali sedem različnih programov za združevanje zaporedij: TGICL, MIRA, iAssembler, PAVE in Newbler (verzija 2.3. in verzija 2.6.), ki temeljijo na OLC (over-layout-consensus) metodi, ter CLC Genomic Workbench 4.5, ki temelji na metodi uporabe De brujin grafa. Programe za združevanje zaporedij (zbirnike) smo ocenili glede na statistiko združevanj, glede na delež unikatnih zaporedij, rezultate združevanja pa smo ocenili tudi s primerjanjem združenih zaporedij na lokalno izdelane podatkovne baze proteinov in z medsebojnim primerjanjem. Rezultati združevanja so se zelo razlikovali med programi po številu združenih kontigov in po količini sekvenčne informacije, ki so jo bili sposobni vključiti v združena zaporedja in po številu zaporedij, ki so ostala nezdružena.

Najoptimalnejši zbirnik naj bi združil največ sekvenc v najdaljša zaporedja, obenem pa bi ostalo minimalno število preostalih zaporedij. V tej kategoriji ocenjevanja se je najslabše izkazal zbirnik Newbler 2.3, saj je lahko združil le 13.530 zaporedij v skupni dolžini 8,4 Mb, medtem ko je zbirnik iAssembler izdelal 49.860 zaporedij v skupni dolžini 25,5 Mb. Zbirnika MIRA in PAVE sta dosegla tudi dovolj dobre rezultate združevanja. Najmanj nezdruženih zaporedij je ostalo pri zbirnikih PAVE (8,2 %), MIRA (8,6 %) in iAssembler (8,5 %). Pri analizi, ki so jo opravili Grag in sod. (2011) je program Newbler 2.5. sestavil najmanj zaporedij, medtem ko se je za najboljšega izkazal program MIRA, saj je sestavil veliko kontigov z najdaljšo skupno dolžino. Zheng in sod. (2011) so v njihovi študiji ugotovili, da so se pri združevanju zaporedij z zbirniki MIRA, CAP3, TGICL, Phrap in Newbler pogosto pojavljale napake tipa II (Zheng in sod., 2011). Zbirnik iAssembler je uspel te napake popraviti in posledično proizvedel manj kontigov, ki pa so bili znatno daljši. Garg in sod. (2011) so po učinkovitosti med seboj primerjali tudi programa Newbler 2.3 in 2.5. Glede na skupno dolžino združenih zaporedij je bil program Newbler 2.5 kar za 38 % boljši od programa Newbler 2.3, kar pa je v nasprotju z rezultati, ki so jih v svoji študiji pridobila Kumar in Blaxter (2010), saj so le ti pokazali, da naj bi bil glede na skupno dolžino združenih zaporedij program Newbler 2.3 za 39 % boljši (Kumar and Blaxter, 2010). Ti rezultati kažejo na to, da je delovanje posameznih programov za združevanje zaporedij odvisno tudi od organizma iz katerega zaporedja izvirajo , ter da je potrebno njihovo uporabo predhodno optimatizirati. V naši študiji sta oba programa Newbler združila najmanj zaporedij, vendar je bil program Newbler 2.6 boljši kot Newbler 2.3.

Poleg velikega števila kontigov naj bi dober zbirnik omogočil tudi produkcijo kakovostnih in dolgih kontigov, ki ne bi vsebovali himernih zaporedij, ter zelo podobnih prekrivajočih se kontigov, ki nastanejo kot posledica prepisov alelov ali kot posledica nekakovostnih

podatkov. Pri naši analzi je zbirnik iAssmbler združil največ kontigov daljših od 1.000 bp (2.363) in tudi največ kontigov daljših od 500 bp (21.879). Najdaljša zaporedja pa sta združila zbirnika PAVE in Newbler 2.3 (4.619 bp in 4.336 bp). Newbler 2.3 je imel najdaljšo povprečno dolžino združenih zaporedij (623 bp), najdaljšo mediano (587 bp) in največjo N50 vrednost (687 bp), ki se uporablja kot merilo pri ocenjevanju združenih podatkov. V tej kategoriji se je najslabše izkazal CLC program, medtem ko so ostali bili primerljivi z dolžinami okrog 500 bp. Pri študijah, ki so jih opravili Grag in sod. (2011), ter Kumar in Blaxter (2010) je Newbler 2.3. prav tako imel najdaljšo povprečno dolžino združenih zaporedij in največjo N50 vrednost, vendar se je v obeh študijah slabo izkazal pri združevanju zaporedij v kontige. Tudi Zheng in sod. (2011) so pridobili najdaljše kontige z uporabo programa Newbler, vendar so te pripisali številnim napakam tipa I, ki so nepravilno združile različne transkripte v skupno zaporedje (Zheng in sod., 2011).

Prav tako bi moral dober zbirnik končati analize v sprejemljivem času, ter dobro časovno obvladovati tudi naraščajoče količine podatkov. Čeprav hitrost delovanja zbirnika ni prevladujoči kriterij, pa je kljub temu pomembna, saj omogoča obsežnejše in robustnejše analize podatkov z več pogoni z različnimi nastavitvami. Seveda pa hitrost obdelave podatkov ni povezana s kakovostjo združevanja zaporedij. Po hitrosti je izstopal CLC, ki uporablja nov algoritem poravnave, saj je delo končal v samo 5-ih minutah, medtem ko je program PAVE porabil za združevanje celih 12 dni. Ostali programi so porabili za delo 15 do 40 ur, kar je še zelo sprejemljiv čas.

Sledila je medsebojna analiza zastopanosti združenih zaporedij v vsaki skupini. Ideja tega načina primerjave je v tem, da odkrijemo zbirnik, ki najde največ različnih zaporedij v primerjavi z ostalimi zbirniki. Zbirnik, ki združi veliko zaporedij v kontige, ki se pogosto ponavljajo, je lahko slabši od zbirnika, ki združi manj zaporedij, vendar so le ta unikatna. Za primerjavo smo uporabili program BLAT, ki izvede hitre primerjave zaporedij med sabo (Preglednica 1). Analiza je pokazala, da zbirnik iAssembler najbolje povzame tudi zaporedja ostalih zbirnikov. Najslabše sta se odrezali obe verziji programa Newbler, s preko 10.000 zaporedji ostalih zbirnikov, ki jih nimata zastopanih v svojih podatkih. Ostali štirje zbirniki so bili med sabo primerljivi. V študiji, ki sta jo opravila Kumar in Blaxter (2010) je imel Newbler 2.3. prav tako najmanj unikatnih zaporedij, saj so bila skoraj vsa zastopana tudi v ostalih zbirnikih (Kumar and Blaxter, 2010). Ostali programi, ki so jih preučevali (CAP3, CLC, MIRA in Newbler 2.5.), so imeli vsi primerljivo število unikatnih zaporedij.

Pomemben kriterij, ki določa optimalno delovanje zbirnika, je tudi ta kako dobro povzema predhodno določene sekvence za ciljne vrste in kako dobro predstavlja zaporedja iz sorodnih organizmov. Na koncu smo izvedli primerjavo zbranih zaporedij vseh skupin z dvema skupinama proteinskih zaporedij (14,9 M proteinskih zaporedij NR proteinske baze in 0,5 M rastlinskih proteinov UNIPROT baze). Pri tej analizi so rezultati pokazali, da pri

zbirnikih iAssembler in MIRA pričakujemo delno fragmentacijo zaporedij, se pravi da imamo lahko zaporedje razdeljeno na dva dela (Preglednica 6). Po drugi strani, pa lahko imamo ločeni alelni obliki gena, kar bi pri močno heterozigotni rastlini kot je oljka to tudi pričakovali. Pri vseh zbirnikih smo dobili med 7-9 % zaporedij, ki kažejo ujemanje s proteini po celotni dolžini, medtem ko večina zaporedij kaže ujemanje z okrog 20 % odstotkov dolžine ujemanja (Preglednica 6).

Z upoštevanjem vseh parametrov analize zbirnikov smo določili, da program iAssembler zajame največjo skupino transkriptov oljke, sledita pa mu programa PAVE in MIRA. Najslabše sta se, z upoštevanjem vseh parametrov, izkazala obe verziji programa Newbler. Zheng in sod. (2011) so opravili obsežno oceno zbirnika iAssembler v primerjavi z ostalimi zbirniki in dokazali, da ima zbirnik iAssembler bistveno boljše rezultate in manj napak pri sestavi EST zaporedij, pridobljenimi z Roche 454 tehnologijo (Zheng in sod., 2011).

Ker se število podatkov, pridobljenih z NGS tehnologijo naglo povečuje, postajajo zbirniki, ki temeljijo na de Brujin grafu, vedno bolj popularni. Vendar pa je bila večina sedanjih zbirnikov, ki temeljijo na de Brujin grafu, zasnovana na podlagi sekvenčnih rezultatov Illumina in Solexa platform. Za 454 Roche zbirnik sta še vedno zlata standarda zbirnika MIRA in Newbler, ki temeljita na OLC metodi (Ren in sod., 2012). V naši študiji se je program MIRA dobro izkazal, saj je z upoštevanjem vseh parametrov zasedel tretje mesto. Za razliko od programa MIRA, pa se je program Newbler izkazal precej slabše, saj je zasedel zadnje mesto. V naši študiji smo preizkusili le en zbirnik, ki je temeljil na osnovi de Brujin graf metode, in sicer CLC zbirnik. Ta se je izkazal slabše kot zbirnik MIRA, vendar boljše kot zbirnik Newbler. Tudi v študiji, ki so jo opravili Ren in sod. (2012) je zbirnik MIRA dosegel boljše rezultate kot ostali de Brujin graf zbirniki, vendar pa so bili slednji enakovredni ali celo boljši od zbirnika Newbler (Ren in sod., 2012). Zato lahko sklepamo, da izbira programov za obdelavo 454 podatkov na osnovi OLC metode, ni nujno vedno najboljša izbira.

Razvoj plodov je genetsko reguliran proces, na katerega pa vplivajo tudi okoljski dejavniki. Za določitev in opredelitev genov, ki so vključeni v razvojne procese plodov različnih rastlinskih vrst, so v uporabi genomska orodja kot so: izražena nukleotidna zaporedja (ESTs), mikromreže, določanje profila transkriptoma, itd. Ta orodja so v zadnjem desetletju doprinesla k naglemu povišanju števila informacij v zvezi s transkriptomom in regulatornimi potmi, ki so vključene v fiziološke in razvojne procese v rastlinah. Na jablani je bila narejena obsežna analiza, pri kateri so uporabili vsa EST zaporeja, ki so bila na voljo v javnih podatkovnih zbirkah, da bi identificirali gene, ki so vključeni v regulatorne poti v času razvoja in rasti plodov. Drugi obsežnejši EST projekti, ki so se nanašali na razvoj plodov, so bili narejeni na breskvah (Vecchietti, 2009), melonah (Clepet, 2011) in kiviju (Crowhurst, 2008).

Znani so fiziološki in biokemični podatki o rasti, razvoju in zorenju oljčnih plodov, vendar pa v glavnih genskih podatkovnih bazah ne najdemo veliko podatkov o sekvencah genov in genskih produktih oljke (Galla in sod., 2009). Čeprav so bili pri oljkah razviti mnogi molekulski markerji, ni bilo narejenih veliko EST študij. 3734 EST-jev so objavili Ozgenturk in sod. (Ozgenturk in sod., 2010), ki so analizirali dve cDNA knjižnici iz mladih listov in nezrelih plodov oljke z željo, da bi odkrili nove gene in njihovo funkcijo, ter izvedli primerjavo izražanja genov med listi in plodovi oljke. Prvo večjo kolekcijo EST-jev pri oljki, ki pa je del Sequence Read Archive-a (SRA), so objavili Alagna in sod. (Alagna in sod., 2009), kjer so predstavili 4 knjižnice z 261,485 zaporedji pridobljeno s prvo generacijo tehnologije 454 FLX dolžina branja. S primerjalnim sekvenciranjem štirih različnih cDNA zbirk dveh oljčnih genotipov so pridobili informacije o spreminjanju genske ekspresije med razvojem oljčnih plodov in med dvema genotipoma s kontrastno akumulacijo fenolov v plodovih (Alagna in sod., 2009). Leta 2012 so nato preučevali delovanje metabolnih poti in transkriptoma med razvojem oljčnih plodov na ravni fenolnih spojin, z uporabo verižne reakcije s polimerazo v realnem času (RT-qPCR). To je privedlo do identifikacije nekaterih glavnih akterjev, ki sodelujejo pri biosintezi sekundarnih spojin v oljkah (Alagna in sod., 2012). Galla in sod. (Galla in sod., 2009) so z anotacijo genov izraženih v različnih stopnjah razvoja plodu oljke, prav tako želeli razkriti metabolne poti in poti transkriptoma v povezavi s karbohidrati, maščobnimi kislinami, sekundarnimi metaboliti, transkripcijskimi faktorji in hormoni.

Tudi mi smo v naši študiji želeli izboljšati znanje o sestavi genov in njihovem izražanju, ter s tem pridobiti nove informacije o procesu razvoja, fiziologiji dozorevanja, primarnem metabolizmu in sintezi zdravilnih substanc biofenolov v oljčnem plodu. Te informacije bi lahko pomagale tudi pri izboljšanju kvalitativnih in kvantitativnih lastnosti oljčnih produktov.

Rastlinske celice za sintezo lipidov potrebujejo energijo v obliki sladkorjev, Acetil-CoA pa predstavlja začetni substrat za sintezo oglikovih verig v maščobnih kislinah. Sinteza acetil-CoA zahteva piruvat pridobljen iz ogljikovih hidratov preko glikolize, lahko pa nastane tudi iz piruvata v mitohonrijih in je hidroliziran v acetat, ter se s pomočjo acetil-CoA sintetaze aktivira v acetil-CoA, ko mitohondriji dosežejo plastid (Conde in sod., 2008).

Biosinteza maščobnih kislin se odvija znotraj plastidov in je dobro poznana. V naših podatkih, ki smo jih pridobili s funkcijsko analizo s pomočjo programa Blast2go, smo odkrili encime, ki sodelujejo pri sintezi maščobnih kislin. Sinteza se prične s karboksilacijo acetil-CoA v malonil-CoA (Sanchez in Harwood, 2002). Reakcijo, ki poteka v dveh korakih katalizira encim acetil-CoA karboksilaza (EC:6.3.4.14), ki je bil določen tudi v naših podatkih. Acetil-CoA karboksilaza je heteromerni kompleks, ki vsebuje biotin prostetično skupino in jo sestavlja več podenot, v naši podatkih sta se pojavljali podenoti biotin karboksilaza in biotin karboksil transportni protein. Encim acetil-CoA karboksilaza

je v večji meri odgovoren za celotno biosintezo maščobnih kislin. Malonil skupina se prenese na acil transportni protein (ACP). Maščobna kislina se podaljša iz malonil-ACP in acetil-CoA preko reakcij, ki jih katalizira kompleks različnih encimov (sintetaze maščobnih kislin - FAS). Znotraj naših podatkov smo našli encima beta ketoacil sintazo II (EC:2.3.1.41) in ketoacil-ACP reduktazo (ni EC številke), ki sodelujeta v procesu podaljševanja maščobnih kislin (Conde in sod., 2008). Pri oljki je bila aktivnost kompleksa FAS že proučena na vodotopni frakciji pulpe iz razvijajočih se plodov z uporabo radioaktivno označenega malonil-CoA kot prekurzor. Produkti, ki jih je tvoril ta encimski sistem, so bile večinoma maščobne kisline, ki so vsebovale 8-18 ogljikovih atomov (Sanchez and Harwood, 1992).

Oleinska kislina je glavna maščobna kislina v oljčnem olju, njegova vsebnost pa lahko doseže do 80 % (Conde in sod., 2008). V prvem koraku nastajanja oleinske kisline se stearoil-ACP pretvori v oleil-ACP, encim stearoil-ACP desaturaza (EC:1.14.19.1; EC:1.14.19.2), ki je potreben za potek te reakcije, pa je bil odkrit tudi v naših podatkih. Transkripcijo tega encima so že preučevali v oljčnih plodovih v različnih fazah razvoja (Haralampidis in sod., 1998), kjer je bilo dokazano, da je mRNA prisotna že v majhnih plodovih, embriju in endospermu. Tudi v naši raziskavi smo preučevali ekspresijo gena stearoil-ACP desaturaze (*SAD1*) skozi različne faze razvoja oljčnih plodov sorte 'Istrska belica' (Slika 11a).

Odkrili smo tudi maščobne acil-ACP tioesteraze (EC:3.1.2.14). To so intraplastidni encimi, ki prekinejo sintezo maščobnih kislin v rastlinah. Kloroplastna stroma vsebuje dve glavni skupini acil-ACP tioesteraz. Največja je FatA skupina, ki spodbuja nastajanje oleoil-ACP (18:1-ACP), manjša pa je FatB skupina, ki spodbuja nastajanje palmitoil-ACP (Conde in sod., 2008). Ker je za nastanek glavne maščobne kisline v oljčnem olju potrebna oleoil-ACP, smo izražanje tega gena prav tako spremljali skozi različne faze razvoja oljčnih plodov sorte 'Istrska belica' (Slika 11b).

Re-esterificirani oleati in palmitati se transportirajo v citosol kot acil-CoA (''pot pri evkariontih'') in v Kennedijevi poti služijo kot acetiltransferaze na endoplazmatskemu retikulumu z vlogo kopičenja triacilglicerolov (TAG). Znotraj naših podatkov smo odkrili večje število zaporedij encima glicerol-3-fosfat aciltransferaza (G3PAT) (EC:2.3.1.15; EC:2.3.1.51) in encima diacilglicerol aciltransferaza (DAGAT) (EC:2.3.1.20). Oba encima katalizirata serijske reakcije, ki omogočata sintezo TAG-ov iz glicerol-3-fosfata in tvorjenje maščobnih kislin v plastidih. Pri oljki je bila pot sinteze TAG-ov proučena na vodotopni frakciji pulpe iz razvijajočih se plodov (Rutter in sod., 1997). Specifičnost posameznih aciltransferaz, ki so vključene v Kennedijevi poti, vpliva na TAG sestavo maščobnih kislin in s tem na kakovost posameznega rastlinskega olja. Prvi encim v tej poti, glicerol-3-fosfat aciltransferaza (G3PAT), naj bi tako sprejemal predvsem palmitoil-CoA, čeprav lahko sprejema tudi oleoil-CoA. Druga aciltransferaza, ki ni bila prisotna znotraj

naših podatkov, lizofosfatidna aciltransferaza (LPAAT), naj bi imela močno selektivnost za oleoil-CoA in bila neaktivna s saturiranimi acil-CoA. Zadnji korak v Kennedijevi poti pa katalizira diacilglicerol aciltransferaza (DAGAT), ki ima široko specifičnost in najbolj vpliva na končno sestavo nabora acil-CoA. Večji delež oleoil-CoA narekuje večjo vključitev oleinske kisline, kar je v skladu z analitičnimi podatki za sestavo maščobnih kislin oljčnega olja (Conde in sod., 2008).

V oljčnih plodovih se akumulira širok razpon sekundarnih metabolitov. Glavni delež sekundarnih metabolitov v oljki predstavljajo sekoiridoidi. To je skupina monoterpenoidov, z razcepljenim metilciklopentan skeletom, ki je značilna za družino *Oleace* in še nekatere druge družine dvokaličnic. V oljkah so sekoiridoidi prisotni v izobilju in predstavljajo na fenol vezane komponente, ki lahko vsebujejo glikozidni del. Najpomembnejši sekoiridoidi v oljčnih plodovih in deviškem oljčnem olju so olevropein, demetilolevropein, olevrozid, ligostrozid, dialdehidna oblika dekarboksimetil elenojske kisline povezane z 3,4-DHPEA ali p-HPEA (3,4-DHPEA-EDA in p-HPEA-EDA), izomer olevropein aglikona (3,4-DHPEA-EA) in ligstrozid aglikona (p-HPEA-EA) (Alagna in sod., 2012). Ker sekoiridoidi niso topni v olju, ostane v deviškem oljčnem olju po mehanični ekstrakciji le majhen delež teh komponent (Servili in Montedoro, 2002). Kljub temu so sekioridoidi pomembna sestavina oljčnega olja, saj vplivajo na njegove zdravstvene in senzorične lastnosti. Sekoiridoidi v oljčnih plodovih imajo pomembno vlogo pri preprečevanju ateroskleroze in pri inhibiciji LDL peroksidacije, številne študije pa so pokazale, da so te spojine tudi dobra preventiva proti rakavim obolenjem in osteoporozi. Zlasti oleuropein, hidroksitirozol in oleokantal so pokazali pozitivne učinke na zdravje ljudi. Sekoiridoidi prispevajo k izboljšanju kakovosti oljčnega olja in vplivajo na okus olja, saj mu dajejo njegovo grenko in pikantno noto, ter kot glavni antioksidanti vplivajo na oksidativno stabilnost oljčnega olja (Alagna in sod., 2012).

Ostali fenoli, ki jih najdemo v oljkah so fenolne kisline, fenolni alkoholi (hidroksitirozol (3,4-DHPEA) in tirozol (p-HPEA)), flavonoidi in lignani. Te spojine so prisotne v vseh delih oljčnega plodu, z najvišjo koncentracijo v pulpi. Oljčni plodovi določenih sort vsebujejo tudi visoke deleže verbaskozidov, ki so značilni za zrele plodove. Študije fenolnih spojin v mezokarpu, eksokarpu, semenu in listih oljk so pokazale, da različna tkiva vsebujejo različne spojine. Poleh fenolov najdemo v plodovih oljk tudi triterpenske kisline (maslinska kislina in oleanolinska kislina) in tokoferole (Alagna in sod., 2012).

Fenolne spojine se tvorijo preko šikiminske poti in fenilpropanoidnega metabolizma. V rastlinah je šikimiska pot odgovorna za tvorbo dveh aromatičnih kislin, in sicer fenilalanina in tirozina. Ogljikovi hidrati so splošni vir ogljikovih atomov v metabolizmu organizma in zagotavljajo prekurzorje potrebne za sintezo sekundarnih metabolitov, kot so acetati, alifatske amino kisline in šikimiske kisline. Neoksidativna glikoliza glukoze

proizvaja fosfoenolpiruvat in eritroza-4-fosfat, ki predstavljata začetne reagente za tvorbo šikimskim kislin ali šikimisko pot (Ryan in Robards, 1998).

Čeprav je glavna vloga fenilalanina proizvodnja proteinov, deluje tudi kot eden izmed glavnih prekurzorjev večine fenolnih spojin v višjih rastlinah. Prestavlja začetni korak v seriji mnogih reakcij, ki sodelujejo v splošnem fenilpropanoid metabolizmu in se nanaša predvsem na proizvodnjo 4-kumarne kisline in fenilalanina. Fenilalanin amonijeva liaza (PAL) velja za ključni encim pri biosintezi fenolov, saj sproži biosintezo številnih fenilpropanoidnih sekundarnih spojin, vključno z lignini in flavonoidnimi pigmenti. Ta encim, ki ima ključno vlogo pri nadzoru toka skupnih fenolov, je zelo občutljiv na okoljske dejavnike, kot so temperatura, poškodbe in UV svetloba (El Riachy in sod., 2011). Aktivnost encima fenilalanin amonijeva liaza (EC:2.3.1.86; EC:1.13.12.7; EC:6.2.1.12), ki je bil zelo pogost tudi v naših Blast2go podatkih, je v veliki meri odvisna tudi od stopnje zrelosti plodov. Prav tako smo v naših podatkih našli tudi encim 4-kumarat 3-hidroksilaza (EC:2.3.1.86; EC:1.13.12.7; EC:6.2.1.12), ki sodeluje pri tvorbi kavne kisline in kininske kisline, ki sta obe pomembna antioksidanta, vendar pa do sedaj še ni bil podrobno raziskan. Do danes metabolizem sekoiridoidov še ni povsem pojasnjen, vendar so za nekatere vrste *Oleaceae* že predlagali možne matabolne poti teh spojin. Sekoiridoidi so kumarinu podobne spojine, ki izhajajo iz iridoidov z odprtjem ciklopentanskega obroča. Iridoidi nastajajo znotraj sekundarnega metabolizma monoterpenov in imajo značilno ogrodje, v katerem je šestčlenski heterociklični obroč, pripojen k ciklopentanskem obroču. Odprtje tega obroča vodi do formacije sekoksiloganina, ki predstavlja matično spojino sekoiridoidov. V vrstah *Oleaceae* konjugati sekoiridoidov, kot je oleuropein, vsebujejo fenolni del kot posledico esterefikacije, ki naj bi nastala z razvejanjam v poti mevalonske kisline v kateri se združita sinteza terpenov (oleozidni del) in metabolizem fenilpropanoida (fenolni del) (El Riachy in sod., 2011). Alagna in sod. so preučevali koncentracijo glavnih fenolnih spojin, kot so oleuropein, demetiloleuropein, 3-4 DHPEA-EDA, ligstrozid, tirozol, hidroksitirozol, verbaskozid in lignani, v razvijajočih se plodovih 12 različnih sortah oljk. Petintridesetim oljčnim zaporedjem, ki so homologni genom, ki sodelujejo v poteh glavnih sekundarnih metabolitov, so določili nivo izražanja z uporabo RT-qPCR analize. Za analizo so uporabili sorte oljk, katerih plodovi imajo značilno nizko oz. visoko vsebnost fenolnih spojin. Močno korelacijo so opazili med koncentracijami fenolnih spojin in zaporedji, ki sodelujejo pri njihovi biosintezi, kar kaže na transkripcijske regulacije preučevanih poti. Med izbranimi transkripti, so bili nekateri vključeni v plastidne 2-C-metil-d-eritritol 4-fosfat (MEP) in citosolne mevalonat (MVA) poti, medtem ko so drugi kandidatni transkripti sodelovali pri sintezi sekoiridoidov (monoterpenih in fenolnih delov), drugih fenolov, terpenoidov in sterolov.

Ugotovili so, da so se ravni transkriptov, ki sodelujejo pri biosintezi sekoiridoidov (terpenski in fenolni deli) opazno zmanjševale med razvojem plodov, skladno z zmanjšanjem olevropeina v istih stopnjah razvoja.

V višjih rastlinah lahko ogljikovi gradniki terpenoidov, izopentenil difosfat (IPP) in dimetilalil difosfat (DMAPP) izvirajo iz plastidne MEP poti in citosolne MVA poti . Encimi in sorodni geni v obeh poteh so dobro poznani in okarakterizirani. Alagna in sod. (2012) so analizirali različne encime ki naj bi sodelovali v MEP poti. Encimi 1-deoksi-D-ksiluloza 5-fosfat sintaza (EC:2.2.1.7; EC:1.2.4.0), 1-deoksi-D-ksiluloza 5-fosfat reduktoizomeraza (EC:1.1.1.267) in 4-hidroksi-3-metilbut-2-enil difosfat reduktaza (EC:1.17.1.2), ki so bili določeni tudi v naših Blast2go podatkih, so imeli najmočnejšo diferencialno ekspresijo v skladu s svojo vlogo v MEP poti, ki je bila predlagana v drugih rastlinskih vrstah. Transkripti, ki kodirajo encime vključene v MVA pot, pa so pokazali popolnoma drugačen profil izražanja, kot MEP transkripti. Encimi 3-hidroksi-3-metilglutaril-koencimA reduktaza (EC:1.1.1.34), mevalonat kinaza (EC:2.7.1.36), fosfomevalonat kinaza (EC:2.7.4.2) in mevalonat difosfat dekarboksilaza (EC: 4.1.1.33), ki so vključeni v MVA pot, smo določili tudi v naših podatkih. V študiji, ki so jo opravili Alagna in sod. (2012) transkripti teh encimov niso pokazali močnih razlik v ekspresiji med različnimi fazami razvoja plodov. Za raliko od transkriptov vključenih v MVA poti je ekspresija transkriptov, ki so vključeni v MEP poti sovpadala z ekspresijo sekoiridoidov, za katero je značilno, da upada z razvojem plodov. Korelacijo med MEP potjo in transkripcijskim profilom sekoiridoidov podpira hipotezo, da ta pot bolj kot MVA pot vpliva na biosintezo terpenskega dela sekoiridoidov v oljkah. Ti rezultati so skladni s tistimi, o katerih so poročali tudi pri drugih rastlinskih vrstah (Aharoni in sod., 2005).

Biosinteza, ki vodi do tvorbe terpenskega in fenolnega dela sekoiridoidov še vedno ni popolnoma pojasnjena, zato posledično encimi, ki sodelujejo v teh poteh, ostajajo neznani. Alagna in sod. (2012) so analizirali transkripte, ki naj bi bilo vključeni v sintezo terpenskega dela sekoiridoidov. Ekspresija genov geraniol sintaze (EC:4.2.3.20; EC:4.2.3.26), geraniol 10-hidroksilaze (EC:1.14.13.71) in NADH dehidrogenaze (EC: 1.6.99.3.), ki jih najdemo tudi v naših Blast2go podatkih, je med razvojem plodov vidno upadala v skladu z upadanjem koncentracije oleuropeina. Encima geraniol sintaza (GES) in geraniol 10-hidroksilaza (GE10H) naj bi bila, glede na rezultate raziskav pri različnih rastlinskih organizmih, vključena v biosintezo iridoidnih monoterpenoidov v različnih razredih monoterpenoidnih alkaloidov. Tvorba fenolnega dela sekoiridoidov naj bi predvidoma izhajala iz tirozina in potekala preko tirozola. V oljki sta bili predlagani dve alternativni poti za tvorbo oleuropeina preko tirozola. V prvi naj bi ligstrozid predstavljal direktni prekurzor za tvorbo oleuropeina (Damtoft in sod., 1993), v drugi pa naj bi tvorba potekala preko oleuropein aglikona (Ryan in sod., 2002). V naših oljčnih Blast2go podatkih smo našli gene za arogenat dehidrogenazo (EC:1.3.1.12; EC:1.3.1.13), polifenol oksidazo (EC:1.10.3.1), tirozin dekarboksilazo (EC:4.1.1.25) in alkohol dehidrogenaza (EC:1.1.1.1), ki naj bi sodelovali v omenjenih poteh in bili v korelaciji z vsebnostjo sekoiridoidov.

Ključni geni, ki sodelujejo pri sintezi in razgradnji sekundarnih spojin v oljčnih plodovih, še niso bili določeni. Izjema je le nekaj genov vključenih v biosintezo triterpenov. Še vedno je zelo malo informacij o sestavi zaporedij oljčnega genoma, vendar se v zadnjih letih povečuje število analiz narejenih na ravni transkriptoma oljke in oljčnega metabolizma (Alagna in sod., 2009, 2012; Ozgenturk in sod., 2010; Galla in sod., 2009). S tem se odpirajo vrata k boljšemu razumevanju o nastajanju in delovanju fenolnih spojin v oljčnih plodovih, ter opredelitvi njihovih glavnih genetskih determinantov.

Za ustrezno RT-qPCR analizo, je potrebno imeti primerne referenčne gene, ki omogočajo natančno normalizacijo genske ekspresije. Številne raziskave so pokazale, da izražanje posameznih genov ni nikoli povsem stabilno (Andersen in sod., 2004; Pfaffl in sod., 2004; Vandesompele in sod., 2002), zato naj bi se normalizacija, ki temelji na uporabi enega referenčnega gena, nadomestila z normalizacijo, ki temelji na uporabi večih referenčnih genov, ki jih predhodno eksperimentalno določimo (Vandesompele in sod., 2002). Številne študije potrjujejo, da je izračun normalizacijskih faktorjev, ki temeljijo na več kot enem referenčnem genu bolj natančen (Hoerndli in sod., 2004; Schmid in sod., 2003). Za vsak organizem bi bilo potrebno določiti primerne referenčne gene s stabilnim profilom ekspresije.

Povprečna stabilnost ekspresije gena (M vrednost) je bila z uporabo programa geNorm nižja od vrednosti 1 pri vseh analiziranih kandidatnih referenčnih genih oljke, od tega jih je devet imelo M vrednost nižjo od 0.5, kar potrjuje, da imajo ti geni sprejemljivo stabilnost ekspresije (Hellemans in sod., 2007) (Slika 9a).

Razvrstitev genov glede na stabilnost izražanja (M vrednost) nam je omogočila določitev dveh genov (*TIP41* in *TBP*), ki se lahko uporabljata kot stabilna referenčna gena pri študijah izražanja genov oljčnih plodov. Ko smo za izbor najboljših referenčnih genov uporabili program ReFinder (Xie in sod., 2011), ki združuje štiri različne algoritme za vrednotenje, sta se gena *TIP41* in *TBP* prav tako izkazala za najboljša. V študiji, ki so jo Reid in sod. (2006) opravili na mezokarpnem tkivu v različnih fazah razvoja grozdnih plodov, je bil *TIP41* tudi med najbolj stabilnimi referenčnimi geni (Reid in sod., 2006). Prav tako se je *TIP41* izkazal za primernega za normalizacijo pri *Arabidopsisu* (Czechowski in sod., 2005). *TBP* in *TIP41* sta bila med najbolj stabilnimi refrenčnimi geni tudi pri študiji razvojnih procesov paradižnika (Exposito-Rodriguez in sod., 2008), TBP je bil dokazan, kot stabilna referenca pri *Zostera marina* morski travi (Ransbotyn in Reusch, 2006). Gena z najmanjšo stabilnostjo v naši študiji sta bila *ADH1* in *CYS*. ADH1 se pogosto uporablja kot referenčni gen pri kvantitativni detekciji (Chaouachi in sod., 2007), čeprav se je pri RT-qPCR študiji na rastlini kave izkazal kot neprimeren referenčni gen (Barsalobres-Cavallari in sod., 2009). Testirali smo tudi gena *Pkaba1* in *UGPase*, ki smo jih prav tako izbrali iz seznama GSO referenčnih genov. Na podlagi geNorm algoritma je bil gen *UGPase* v naši študiji uvrščen kot sedemnajsti najbolj stabilen gen, medtem ko je

bil *Pkaba1* med najbolj stabilnimi geni v naši študiji in je zasedel tretje mesto. Zato bi bilo smiselno tudi druge referenčne gene iz seznama GSO (Chaouachi in sod. 2007) predhodno testirati za uporabo RT-qPCR analiz.

V nedavni študiji, ki so jo opravili Nonis in sod. (2012), so se osredotočil na stabilnost kandidatnih referenčnih genov v različnih razvojnih fazah olječnih plodov in v ranjenih listnih tkivih (Nonis in sod., 2012). Ocenili so 13 parov začetnih oligonukleotidov za šest kandidatnih referenčnih genov. Vsi od teh šestih genov so bili obravnavani tudi v naši raziskavi, razen *PP2A*. Glede na njihove rezultate pridobljene z uporabo geNorm in Normfinder algoritmov so *GAPDH2* in *PP2A1* bili opredeljeni kot najboljši referenčni geni, z M vrednostmi 0,216 za *PP2A1* in 0,244 za *GAPDH2*. Zanimivo je, da je v njihovi študiji eden izmed *GAPDH* genov (*GAPDH1*) določen kot najslabši referenčni gen. V naši študiji je gen *GAPDH* zasedel devetnajsto mesto po geNorm-u, z M vrednostjo 0.68 (Slika 9a).

18S rRNA gen se pogosto uporablja kot referenca za RT-qPCR analize in je bil uporabljen tudi pri oljki (Corpas in sod., 2006; Muzzalupo in sod., 2012). V našem primeru je imel olječni 18S referenčni gen M vrednost 0,72, kar je nad predlaganim pragom 0,5 in je bil uvrščen na 22. mesto po geNorm-u. 18S rRNA gen je običajno močno izražen in je zato pogosto neprimeren za normalizacijo šibko izraženih genov. Prav tako je bilo dokazano, da se ekspresije ciljnih genov pri RT-qPCR analizi ne sme zanemariti (Nicot in sod., 2005).

Na podlagi rezultatov, ki smo jih pridobili s programom geNorm, smo primerjali ravni izražanja genov *FatA, SAD1, Acot* in *LOX1* v vzorcih olječnih plodov (slika 11). Ti geni so bili izbrani na podlagi znanih vzorcev izražanja genov, vključenih v metabolizem lipidov, v odvisnosti od razvojne stopnje plodov. Maščobna acil-ACP tioesteraza (*FatA*) je intraplastidni encim, ki prekine sintezo maščobnih kislin v rastlinah in spodbuja nastajanje oleoil-ACP (18:1-ACP). Oleinska kislina je glavna maščobna kislina v olječnem olju, njegova vsebnost pa lahko doseže do 80 % (Conde in sod., 2008). V prvem koraku nastajanja oleinske kisline se stearoil-ACP pretvori v oleil-ACP s pomočjo stearoil-ACP desaturaze (*SAD1*), katere transkripcijo so že preučevali v olječnih plodovih v različnih fazah razvoja (Haralampidis in sod., 1998). Dokazano je bilo, da je mRNA prisotna že v majhnih plodovih, embriju in endospermu. V mezokarpu pa se je ekspresija pričela kasneje (13 tednov po cvtenju) in je bila prisotna do 28. tedna po cvetenju. V naši študiji je bila ekspresija gena *SAD1* primerljiva z ekspresijo stearoil-ACP desaturaze, ki so Haralampidis in sod. preučevali s pomočjo eksperimenta po Northernu (Fig. 11a) (Haralampidis in sod., 1998). Ekspresija gena *SAD1* je bila nizka v prvih treh vzorčnih točkah (14 DAF, 29 DAF and 42 DAF), nato pa se je povišala v točki 4 (57 DAF) in 5 (72 DAF), kar je lahko posledica embrionalne genske ekspresije. Ekspresija gena *SAD 1* je povišana do točke 7 (98 DAF) in nato rahlo upada do točke 9 (129 DAF), kar je lahko posledica razvoja endosperma. Ekspresija nato vidno upade v točki 10, vendar je še vedno vidna tudi v

vzorčni točki 12, ki predstavlja prezreli plod. Haralampidis in sod. (1998) trdijo, da naj bi transkripcija gena *SAD1* ostala na maksimalni točki do 28. tedna po cvetenju, brez vmesnih padcev v ekspresiji. V njihovi študiji so uporabili analizo po Northernu, ki bi morala dobro sovpadati z RT-qPCR analizo (Dean in sod., 2002), vendar razlike v genotipu oljke in v okoljskih dejavnikih, lahko pojasnijo rahle razlike v ekspresiji gena *SAD1*. Gen *FatA* ima v prvih vzorčnih točkah podobeno ekspersijo kakor gen *SAD1*, v točki 4 pa začne ekspresija gena *FatA* naraščati in doseže vrh v točki 9. Ekspresija ostane visoka do točke 11 in nato vidno upade v točki 12.

Acilkoencim A tioesteraza (*Acot*) hidrolizira maščobne acilkoencime A v maščobne kisline in koencim A, s čimer zagotavlja možnost intracelularne regulacije acilkoencimov A, maščobnih kislin in koencima A. Ta družina encimov naj bi imela vlogo pri oksidaciji maščobnih kislin pri živalih. V rastlinah pa je bil prvi Acot gen –*ACH2* kloniran iz *Arabidopsisa*. Gen je bil bolj izražen v zrelih tkivih, kot pa v kalečih sadikah, kar kaže na to, da najverjetneje ni povezan z oksidacijo maščobnih kislin (Tilton in sod., 2004). *Acot* gen je med razvojem plodov kazal različno ekspresijo; ta je bila najvišja v zadnji fazi otrditve koščice in skozi celotno fazo razvoja mezokarpa (Slika 11c, vzorci 6-10), nato pa je ekspresija upadla (vzorec 11 in 12). Glede na pridobljene rezultate lahko sklepamo, da tudi naš *Acot* gen ni primarno vključen v proces beta oksidacije.

Ekspresija četrtega gena, lipoksigenaze 1 (*LOX1*), je bila izredno visoka v prezrelem plodu (vzorec 12), rahla ekspresija pa je bila prisotna tudi v ostalih 11 vzorcih oljčnih plodov, z naraščujočim trendom ekspresije v vzorcu 10 in 11 (razvoj mezokarpa in zorenje plodu). Encim *LOX1* ima pomembno vlogo v različnih pretvorbenih procesih znotraj lipoksigenazne (*LOX*) poti, pri katerih se tvorijo hlapne snovi nastale iz linolne in linolejne kisline. Ta pot se sproži med postopki mletja, malaksacije in ekstrakcije olja iz oljčnih plodov (Conde in sod., 2008). Ta metabolni proces pomembno vpliva na kvaliteto oljčnega olja, saj le to določamo tudi na podlagi arome oljčnega olja, ki je zmes različnih hlapnih spojin (Morales in sod., 1995). Nedavna qPCR študija gena *LOX1* v oljčnih plodovih je pokazala naraščujočo ekspresijo gena proti koncu zorenja plodov oljk v treh vzorcih dveh različnih Italjanskih sort (Muzzalupo in sod., 2012). Naši rezultati ekspresije *LOX1* gena podpirajo te ugotovitve, zelo visoka raven ekspresije pa je bila ugotovljena tudi v prezrelem plodu (vzorec 12).

Gene *FatA*, *SAD1*, *Acot*, in *LOX1* smo normalizirali tako z 9 najboljšimi referenčnimi gen, kot tudi z dvema najboljšima referenčnima genoma (*TBP/TIP41*). Korelacijski koeficienti med standardiziranimi podatki pridobljenimi z 9 najboljšimi referenčnimi geni in dvema najboljšima referenčnima genoma so znašali 0.97, 0.97, 0.93, in 0.98, ter potrdili primernost genov *TBP/TIP41* za normalizacijo. Korelacija med normaliziranimi (*TBP/TIP41*) in nenormaliziranimi podatki je bila nizka (0.63, 0.23, 0.04, in 0.50), kar je verjetno posledica višje ekspresije genov v točkah 4 in 9, do katere je lahko prišlo zaradi

različne kvalitete/kvantitete RNA vzorcev ali povišane stopnje reverzne transkripcije (Nolan in sod. 2006). Za normalizacijo smo uporabili tudi referenčni gen z najnižjo stopnjo stabilnosti (*ADH1*). Pridobljene podatke smo nato primerjali s podatki normaliziranimi z najboljšima referenčnima genoma in prav tako pridobili nizke koeficiente korelacije (0.34, 0.29, 0.04, in 0.25), kar kaže na to, da je izbor primernih referenčnih genov nujno potreben za pravilno vrednotenje ekspresije genov. Prav tako je bila opažena velika razlika v ekspresijskem vzorcu pri dveh oljčnih genih, domnevna poligalakturonaza (PG) in farnezil pirofosfat-sintaza (PPT), ko je bila najslabša interna kontrola uporabljena za normalizacijo (Nonis in sod., 2012).

6 POVZETEK (SUMMARY)

6.1 POVZETEK

Oljka (*Olea europaea*, družina Oleaceae) je zimzeleno drevo s plodovi, ki se uporabljajo za pridobivanje olja ali za namizne oljke. Je tipičen pokazatelj sredozemskega podnebja, razširjenost vrste pa je povezano z geografskimi in klimatskimi dejavniki ter z dolgim obdobjem njenega gojenja. Oljka soobstaja s človekom še iz časov zgodnje bronaste dobe in je bila od nekdaj tesno povezana z verovanjem, socialno-kulturnimi, zdravstvenimi in prehrambenimi potrebami človeka. Oljke so znane, kot eno izmed najbolj pogosto gojenih sadnih dreves na svetu. Kar 98 % pridelovalnih površin, 99 % gojenih dreves in 99 % proizvodnje oljk pripada državam Mediteranskega območja in območja Bližnjega Vzhoda. Po ocenah organizacije FAO (Food and Agriculture Organization) naj bi oljčni nasadi leta 2009 zavzemali kar 9.9 milijonov hektarov pridelovalnih površin. Svetovna poraba oljk in oljčnih izdelkov se je občutno povečala predvsem v visoko razvitih deželah, k temu pa je prispevala predvsem tradicionalna Mediteranska dieta, ki temelji na rednem uživanju kakovostnega oljčnega olja in velja za eno izmed najbolj učinkovitih diet, saj dokazano varuje človeški organizem pred nekaterimi boleznimi sodobnega časa (Ryan in Robards, 1998; Soler-Rivas in sod., 2000).

Oljčni plod se razvija najmanj 4 do 5 mesecev in vključuje 5 večjih faz, med katerimi prihaja do različnih sprememb v sestavi plodu, ki nastajo kot posledica razvoja plodu. Razvoj in dozorevanje oljčnega plodu sta kombinacija biokemsijkih in fizioloških dogodkov, ki so genetsko regulirani, nanje pa vplivajo tudi različni dejavniki okolja (Connor in Fereres, 2005). Znani so mnogi fiziološki in biokemični podatki o rasti, razvoju in zorenju oljčnih plodov, vendar je kljub ekonomskemu pomen in metabolnim posebnostim oljke, zelo malo znanega o zaporedjih genov in nadzoru glavnih metabolnih poti oljke.

Večina genskih raziskav na kulturnih rastlinah je osredotočenih na razumevanje genskih mehanizmov in s tem na izboljšanje kakovosti in količine proizvodov. Izražena nukleotidna zaporedja (ESTs) skupaj s cDNA zaporedji so eno izmed primarnih orodij, ki nam zagotovi neposredne informacije o transkriptih, ki kodirajo dele genoma in so trenutno najpomembnejši vir za raziskovanje transkriptoma. Določanje zaporedij transkriptoma za nemodelne organizme je bolj priljubljeno, saj je cenovno ugodnejše in računalniško vodljivejše, kot sekvenciranje celotnega genoma, vseeno pa zagotovi dovolj informacij za izpolnjevanje zahtev številnih raziskovalnih skupin. Tradicionalno so projekti transkriptomov temeljili na Sangerjevi dideoksi metodi sekvenciranja, vendar so jo pričele izpodrivati metode nove generacije sekvenciranja, ki imajo znatno višjo zmogljivost in nižjo ceno na določitev baze. Pri večini objavljenih projektov na nemodelnih organizmih so uporabili Roche-vo 454 metodo pirosekveniranja, saj od vseh

tehnik naslednje generacije sekvenciranja zagotavlja daljše prepise (okrog 400 baz) in s tem lažje sestavljanje in anotacijo zaporedij (Kumar in Blaxter, 2010).

Z namenom, da bi s pomočjo genomskega pristopa določiti izražanje genov v različnih stopnjah razvoja oljčnih plodov sorte 'Istrska belica', smo iz vzorcev oljčnih plodov, ki smo jih vzorčili skozi različne faze razvoja plodu, ustvarili normalizirano cDNA knjižnico. To smo poslali na določevanje nukleotidnih zaporedij (sekvenciranje) s pomočjo ponudnika storitev naslednje generacije določevanja nukleotidnih zaporedij (v nadaljevanju NGS). Uporabili smo *GsuI* tretiran vzorec cDNA, kateremu smo odstranili poli A regije iz 3' konca. Za slednji vzorec smo se odločili zato, ker se je pri poizkusu s Sangerjevim določevanjem zaporedij izkazal za boljšega, saj prisotnost homopolimernih A regije ni motila postopka sekvenciranja. Odločili smo se za sekvenciranje s tehnologijo Roche 454, ki temelji na pirosekvenciranju.

Skupaj smo pridobili 560.578 konkatemernih zaporedij v skupni dolžini 160.414.301 bp in povprečno dolžino 286 bp. Konkatemerna zaporedja smo nato ločili, ter jim odstranili dele linkerjev, morebitne še vedno prisotne poli A regije in odstranili vsa prekratka in neuporabna zaporedja (Seqclean perl skripta in Megablast), ki lahko dodatno otežujejo proces nadaljnje obdelave. Tako smo na koncu pridobili 577.025 zaporedij v skupni dolžini 139.419.844 bp in povprečno dolžino 241 bp. Teh 577-tisoč zaporedij predstavlja končna cDNA zaporedja oljke, ki jih analiziramo v naslednjem koraku združevanja.

V tem koraku želimo pravilno rekonstruirati (zložiti) zaporedja cDNA in pridobiti čim daljšo možno dolžino oz. rekonstruirati celoten gen. Odločili smo se za podrobnejšo analizo našega seta podatkov z različnimi programi za združevanje, ki so na voljo. Namen tega dela analiz je bil odkriti najboljši program oz. rutino, ki je primeren za analizo transkriptoma oljke. V naši raziskavi smo uporabili večje število programov za obdelavo in združevanje zaporedij (angl. assembler), kot so TGICL (Partea in sod., 2003), MIRA (Chevreux in sod., 2000), iAssembler (Zheng in sod., 2011), gsAssembler 2.3. in gsAssembler 2.5 (Margulies in sod., 2005), PAVE 2.5 (Soderlund in sod., 2009), CLC (CLC, 2013) Genomics Workbench. Najboljši program za združevanje zaporedij smo izbrali glede na primerjavo osnovnih podatkov meritev različnih programov za združevanje zaporedij (Preglednica 5), glede na BLAT primerjava (Preglednica 6) in glede na BLASTX rezultate posameznih programov za združevanje zaporedij (Preglednica 6). Z upoštevanjem vse kriterijev ocenjevanja, se je za najboljši zbirnik izkazal program iAssembler (Preglednica 8).

Sledila je funkcijska analiza izbranih oljčnih zaporedij z uporabo programskega paketa Blast2go, ki izražena nukleotidna zaporedja razdeli v tri glavne Gene Ontology sklope (Götz in sod., 2008). 25.451 zaporedjem (51 %) uspešno pripisali vlogo na ravni bioloških procesov, celičnih komponent in molekularnih funkcij. Z Blast2go anotacijo smo pridobil

podatke o skupinah genov, ki so vključeni v procese sekundarnega metabolizma (950 zaporedij, GO: 0019748), metabolne procese maščobnih kislin (47 zaporedij, GO: 0006631), v biosintezne procese maščobnih kislin (305 zaporedij, GO: 0006633), v procese nesaturiranih maščobnih kislin (22 zaporedij, GO: 0006636), ter metabolne (99 zaporedij, GO: 0006629) in biosintezne (30 sekvenc, GO: 0008610) procese lipidov.

Cilj nadaljne raziskave je bil določiti primerne referenčne gene (RGs) za analize razvijajočih se plodov oljke s pomočjo PCR v realnem času (qPCR), saj je za ustrezno RT-qPCR analizo, je potrebno imeti primerne referenčne gene, ki omogočajo natančno normalizacijo genske ekspresije. Izbrali smo 29 kandidatnih RGs (Priloga 1) in 12 točk vzorčenja, da bi zajeli pet glavnih faz razvoja olječnih plodov. Glede na rezultate geNorm algoritma, sta se za najboljša RGs izkazala TIP41 sorodni protein (TIP41) in TATA vezavni protein (TBP). Z uporabo teh dveh RGs smo določili štiri gene (maščobna acil-ACP tioesteraza A, *FatA*; stearoil-ACP desaturaza, *SAD1*; acil-CoA tioesterazni sorodni protein, *Acot*; lipoksigenaza 1, *LOX1*), ki sodelujejo v metabolizmu maščobnih kislin in dokazali različne vzorce izražanja, povezane z razvojem mezokarpa in zorenjem olječnih plodov.

V okviru doktorske naloge smo izpolnili glavne cilje, ki smo si jih zastavili. Pridobili smo dovolj kvalitetne RNA vzorce iz razvijajočih se plodov oljk, ki so bili primerni za razvoj normalizirane cDNA knjižnice. Z odstranitvijo poli A regij smo pripomogli k boljši izvedbi določevanja nukleotidnih zaporedij, določili smo večjo količino nukleotidnih podatkov za razvijajoči plod oljke ter najoptimalnejši zbirnik za sestavo teh podatkov. Določili smo tudi transkripte, ki so povezani s primarnim in sekundarnim metabolizmom olječnega plodu, ter potrdili tkivno specifično izražanje za nekatere ključne transkripte s pomočjo PCR v realnem času (qPCR). Opravljena raziskovalna študija razvoja EST zaporedij olječnega plodu bo lahko uporabna za raziskovalne skupine, ki delajo na raziskavah oljke, preko orodij primerjalne genomike pa lahko tudi za raziskave ostalih ekonomsko pomembnih rastlin.

6. 2 SUMMARY

Olive (*Olea europaea*, family Oleaceae) is an evergreen tree grown for its fruits (drupes), which yield oil and are also marketed as table or pickled olives. The olive tree is considered one of the most important indicators of the Mediterranean climate. The species' distribution is associated with geographical and bioclimatic factors, as well as being influenced by a long period of cultivation. The edible olive seems to have coexisted with humans since the early Bronze Age and has been closely associated with human religious, sociocultural, medicinal, and nutritional needs. Olives are known as one of the widely cultivated fruit crop worldover. About 98 % of the total surface area, 99 % of productive trees and 99 % of total olive production belong to the countries around the Mediterranean

basin and in the Middle-East. According to an estimate of Food and Agriculture Organization (FAO), in 2009, 9.9 million hectares (ha) were planted with olive trees. Due to rising awareness about the beneficial effects of optimal nutrition and functional foods among todays's health conscious cosmopolitan societies, the worldwide consumption of olives and olive products has increased significantly due to the traditional "Mediterranean diet", in which olive oil is the main dietary fat, is considered to be one of the healthiest because of its strong association with the reduced incidence of cardiovascular diseases and certain cancers (Ryan and Robards, 1998; Soler-Rivas et al., 2000).

Olive fruit growth and development lasts for 4–5 months and includes 5 main phases, involving cell division, cell expansion and storage of metabolites. Olive fruit development and ripening are a combination of biochemical and physiological events that occur under strict genetic control and influence of several environmental conditions (Connor and Fereres, 2005). There have been many physiological and biochemical data on the growth, development and maturation of olive fruits, but in spite of the economic importance and metabolic specificities of olives, very little is known about the sequences of genes and about control of key metabolic pathways in olives.

Most genetic research on crop plants is focused on understanding the genetic mechanisms and thereby to improve the quality and quantity of products. Expressed nucleotide sequences (ESTs), together with the cDNA sequences are one of the primary tools. They provide direct information on the transcripts that encode parts of the genome and are one of the most important resource for research of transcriptome.Transcriptome sequencing projects for non-model organisms are popular because they cost less and are more computationally achiavable than full genome sequencing project, but still yeild sufficient information to meet the requirements of many research programs. Traditionally, transcriptome projects have been based on Sanger dideoxy-sequenced expressed sequence tags (ESTs). Due to the next-generation sequencing technologies emerging, which provide much higher output than Sanger sequencing at a lower cost per base, these new technologies are now increasingly used. However, most published non-model organism projects have used the Roche 454 pyrosequencing platform, because the longer reads generated are more amenable to de novo assembly and annotation (Kumar and Blaxter, 2010).

The initial phase of the project included the production of an expressed sequence tags database (ESTs), which covered sequences derived from developing olive fruits. Olive fruits of the variety 'Istrska Belica', were sampled on a weekly basis throughout the period of fruit development. In this phase of the project, a normalized cDNA library was made from the developing olive fruits. The acquired normalized cDNA library was used for sequencing, using next generation sequencing technology (Roche 454). Before sequencing

with with Roche 454 pyrosequencing technology, the cDNA was treated with GsuI restriction enzyme to avoid problems that can be caused by polyA tails.

Together we gained 560.578 concatamer sequences in total length of 160.414.301 bp. After concatamer splitting with SSAHA2 program, sequences may still contain parts that were used in the construction cDNA and are not part of the olive DNA. Such contaminating sequences may hinder further processing, so the sequences were also included in the process of cleaning. After concatamer splitting (SSAHA2) and removal of contaminants (Seqclean), we gained 577-thousand sequences in total length of 139.419.877 bp of an average length of 241 bp.

These 577-thousand sequences represent the final cDNA sequence of olives to be examined in the next step of assembling. In this step, we want to combine cDNA sequences and obtain the longest possible length of sequences or reconstruct the full length genes. We opted for a more detailed analysis of our data using seven different assembly programs. Among these, TGICL (Partea et all., 2003), MIRA (Chevreux et all., 2000), iAssembler (Zheng et all., 2011), PAVE (Soderlund et all., 2009) and Newbler (v2.3. and v2.6) (Margulies et all., 2005) are based on OLC (over-layout-consensus) strategy, whereas CLC genomic workbench (CLC, 2013) is based on *de Bruijn* graph algorithm. We took several criteria into consideration to select the best de novo assembly, including assembly statistics (Table 5), ratio of novel sequences (Table 6) and alignments to reference database sequences (Table 7).

The results of the best assembling program iAssembler (Table 8), were further used for functional annotation with Blast2go. The Blast2go tool succeffuly revealed an annotation for 51% of all sequences (25.451sequences) that describe gene products in terms of their associated biological processes, cellular components and molecular functions (Götz et all., 2008). Annotation defined groups of genes involved in secondary metabolic processes (950 sequences, GO: 0019748), fatty acid metabolic processes (47 sequences, GO: 0006631), fatty acid biosynthetic processes (305 sequences, GO: 0006633), processes of unsaturated fatty acid (22 sequences, GO: 0006636) and metabolic (99 sequences, GO: 0006629) and biosynthetic (30 sequences, GO: 0008610) processes of lipids.

The aim of the further study was to identify RGs for RT-qPCR studies of developing olive fruit. For the appropriate RT-qPCR analysis, it is necessary to have suitable reference genes, which allow precise normalization of the gene expression. We used 29 RG candidates and 12 sampling points to cover the five stages of olive fruit development. According to the results of the geNorm algorithm, the two best RGs were TIP41-like family protein (TIP41) and TATA binding protein (TBP), while several classical RGs proved not to be suitable. Using the two new RGs, four genes (fatty acylACP thioesterase A (FatA), stearoyl-ACP desaturase (SAD1), acyl-CoA thioesterase family protein (Acot),

lipoxygenase 1 (LOX1)) involved in the metabolism of fatty acids were studied and showed distinct expression patterns associated with mesocarp development and ripening stages.

Within the framework of the doctoral thesis we fulfill the main goals and gained enough high-quality RNA samples from developing olive fruits that were suitable for the development of normalized cDNA libraries. By removing the poly A regions, we contribute to a better determination of nucleotide sequences. We have determined a larger amount of nucleotid data for the developing olive fruits and optimal program to assemble these data. We determined the transcripts that are associated with primary and secondary metabolism of the olive fruit, and confirm tissue-specific expression of some key transcripts using real-time PCR (qPCR). The research results will be useful for research groups working on olives and through the tools of comparative genomics for research on other economically important plants as well.

7 VIRI

Adams M. D., Celniker S. E., Holt R. A., Evans C. A, Gocayne J. D., Amanatides G., Scherer S. E., Li W., Hoskins R. A., Galle R. F., George R. A., Lewis S. E., Richards S., Ashburner M., Henderson S. N., Sutton G. G., Wortman J. R.. 2000. The genome sequence of Drosophila melanogaster. Science, 287: 2185-2195

Adams M. D., Soares M. B., Kerlavage A. R., Fields C. Venter J. C. 1993. Rapid cdna sequencing (expressed sequence tags) from a directionally cloned human infant brain cdna library. Nature Genetics, 4: 373-386

Aharoni A., Jongsma M. A. Bouwmeester H. J. 2005. Metabolic engineering of terpenoids in plants. Trends in Plant Science, 10: 594-602

Alagna F., D'Agostino N., Torchia L., Servili M., Rao R., Pietrella M., Giuliano G.,Chiusano M. L, Baldoni L. Perrotta G. 2009. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. Bmc Genomics, 10:399 http://dx.doi.org/ 10.1186/1471-2164-10-399 (12. maj.2013)

Alagna F., Mariotti R., Panara F., Caporali S., Urbani S., Veneziani G., Esposto S., Taticchi A., Rosati A., Rao R., Perrotta G., Servili M., Baldoni L. 2012. Olive phenolic compounds. metabolic and transcriptional profiling during fruit development. Bmc Plant Biology, 12:162 http://dx.doi.org/10.1186/1471-2229-12-162 (12. maj.2013)

Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. 1990. Basic local alignment search tool. Journal of Molecular Biology, 215: 403-410

Andersen C. L., Jensen J. L., Orntoft T. F. 2004. Normalization of real-time quantitative reverse transcription-PCR data. A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Research, 64: 5245-5250

Andrews S. FastQC. 2010. Babraham Bioinformatics. http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/ (03. nov. 2013)

Angerosa F., Basti C. 2001. Olive oil volatile compounds from the lipoxygenase pathway in relation to fruit ripeness. Italian Journal of Food Science, 13: 421-428

Aparicio R., Aparicio-Ruiz R. 2000, Authentication of vegetable oils by chromatographic techniques. Journal of Chromatography A, 881: 93-104

Apweiler R., Martin M. J., O'Donovan C., Magrane M., Alam-Faruque Y., Antunes R., Barrell D., Bely B., Bingley M., Binns D., Bower L.. 2010. The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Research, 38: D142-D148

Bajgain B., Richardson A., Price J. C., Cronn R. C., Udall J. A. 2011. Transcriptome characterization and polymorphism detection between subspecies of big sagebrush (*Artemisia tridentata*). Bmc Genomics, 12:370 http://dx.doi.org/10.1186/1471-2164-12-370 (14. apr. 2013)

Bandelj Mavsar D., Bešter E., Bučar-Miklavčič M., Butinar B., Čalija D., Kanjir Ž., Levanič T., Valenčič V., Mazi Ž. 2005. ABC o istrski belici. Koper, Univerza na Primorskem, Znanstveno raziskovalno središče: 16 str.

Banilas G., Karampelias M., Makariti I., Kourti A., Hatzopoulos P. 2011. The olive DGAT2 gene is developmentally regulated and shares overlapping but distinct expression patterns with DGAT1. Journal of Experimental Botany, 62: 521-532

Banilas G., Moressis A., Nikoloudakis N., Hatzopoulos P. 2005. Spatial and temporal expressions of two distinct oleate desaturases from olive (*Olea europaea* L.). Plant Science, 168: 547-555

Banilas G., Nikiforiadis A., Makariti I., Moressis A., Hatzopoulos P. 2007. Discrete roles of a microsomal linoleate desaturase gene in olive identified by spatiotemporal transcriptional analysis. Tree Physiology, 27: 481-490

Barnes, W. M. 1994. Pcr amplification of up to 35-kb dna with high-fidelity and high-yield from lambda-bacteriophage templates. Proceedings of the National Academy of Sciences of the United States of America, 91: 2216-2220

Barsalobres-Cavallari C. F., Severino F. E., Maluf M. P., Maia I. G. 2009. Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. Bmc Molecular Biology, 10:1 http://dx.doi.org/10.1186/1471-2199-10-1 (14. apr. 2013)

Bazakos C., Manioudaki M. E., Therios I., Voyiatzis D., Kafetzopoulos D., Awada T., Kalaitzis P. 2012. Comparative Transcriptome Analysis of Two Olive Cultivars in Response to NaCl-Stress. Plos One, 7:11 http://dx.doi.org/10.1371/journal.pone.0042931 (14. apr. 2013)

Bendini A., Cerretani L., Carrasco-Pancorbo A., Gomez-Caravaca A. M., Segura-Carretero A., Fernandez-Gutierrez A., Lercker G. 2007. Phenolic molecules in virgin olive oils. a survey of their sensory properties, health effects, antioxidant activity and analytical methods. An overview of the last decade. Molecules, 12: 1679-1719

Besnard G., de Casas R. R., Christin A., Vargas P. 2009. Phylogenetics of Olea (Oleaceae) based on plastid and nuclear ribosomal DNA sequences. Tertiary climatic shifts and lineage differentiation times. Annals of Botany, 104: 143-160

Besnard G., de Casas R. R., Vargas P. 2007 Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea*). Journal of Biogeography, 34: 736-752

Besnard G., Hernandez, Khadari B., Dorado G., Savolainen. 2011. Genomic profiling of plastid DNA variation in the Mediterranean olive tree. Bmc Plant Biology, 11:80 http://dx.doi.org/10.1186/1471-2229-11-80 (14. apr. 2013)

Besnard G., Khadari B., Berville A. 2002a. Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). Theoretical and Applied Genetics, 105: 139-144

Besnard G., Khadari B., Berville A. 2002b. *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. Theoretical and Applied Genetics, 104: 1353-1361

Bester E., Butinar B., Bucar-Miklavcic M., Golob T. 2008. Chemical changes in extra virgin olive oils from Slovenian Istra after thermal treatment. Food Chemistry, 108: 446-454

Bonaldo M,. D. F., Lennon G., Soares M. B. 1996. Normalization and subtraction. Two approaches to facilitate gene discovery. Genome Research, 6: 791-806

Bouyioukos, C., Moscou M. J., Champouret N., Hernandez-Pinzon I., Ward E. R., Wulff B. B. H. 2013. Characterisation and Analysis of the Aegilops sharonensis Transcriptome, a Wild Relative of Wheat in the Sitopsis Section. Plos One: 8 http://dx.doi.org/10.1371/journal.pone.0072782 (14. apr. 2013)

Brenchley, R., Spannagl M., Pfeifer M., Barker G. L. A., D'Amore R., Allen A. M., McKenzie N., Kramer M., Kerhornou A., Bolser D., Kay S., Waite D., Bevan M. W., Hall N. 2012. Analysis of the breadwheat genome using whole-genome shotgun sequencing. Nature, 491: 705-710

Brenes M., Garcia A., Rios J. J., Garrido A. 1999. Phenolic compounds in Spanish olive oils. Journal of Agricultural and Food Chemistry, 47: 3535-3540

Brent M. R. 2008. Steady progress and recent breakthroughs in the accuracy of automated genome annotation. Nature Reviews Genetics, 9: 62-73

Caraguel C. G. B., Stryhn H., Gagne N., Dohoo I. R., Hammell K. L. 2011. Selection of a cutoff value for real-time polymerase chain reaction results to fit a diagnostic purpose. analytical and epidemiologic approaches. Journal of Veterinary Diagnostic Investigation, 23: 2-15

Carrion Y., Ntinou M., Badal E. 2010. Olea europaea L. in the North Mediterranean Basin during the Pleniglacial and the Early-Middle Holocene. Quaternary Science Reviews, 29: 952-968

Chaouachi M., Giancola S., Romaniuk M., Laval, Bertheau Y., Brunel D. 2007. A strategy for designing multi-taxa specific reference gene systems. Example of application - ppi phosphofructokinase (ppi-PPF) used for the detection and quantification of three taxa. Maize (*Zea mays*), cotton (*Gossypium hirsutum*) and rice (*Oryza sativa*). Journal of Agricultural and Food Chemistry, 55: 8003-8010

Chaparro C., Guyot R., Zuccolo A., Piegu B., Panaud O. 2007. RetrOryza. a database of the rice LTR-retrotransposons. Nucleic Acids Research, 35: D66-D70

CLC bio. 2013. QIAGEN company.

Chevreux B., Pfisterer T., Suhai S. 2000. Automatic assembly and editing of genomic data: Genomics and Proteomics. Functional and Computational Aspects: 51-65 http://www.clcbio.com/ (4.11.2013)

Cock: J. A., Antao T., Chang J. T., Chapman B. A., Cox C. J., Dalke A., Friedberg I., Hamelryck T., Kauff F., Wilczynski B., de Hoon M. J. L. 2009. Biopython. freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25: 1422-1423

Clepet C., Joobeur T., Zheng Y., Jublot D., Huang M. Y., Truniger V., Boualem A., Hernandez-Gonzalez M. E., Dolcet-Sanjuan R., Portnoy V., Mascarell-Creus A., Cano-Delgado A. I., Katzir N. 2011. Analysis of expressed sequence tags generated from full-length enriched cDNA libraries of melon, Bmc Genomics, 12: 252 http://dx.doi.org/10.1186/1471-2164-12-252 (25. maj. 2013)

Conde C., Agasse A., Lemoine R., Delrot S., Tavares R., Geros H. 2007. OeMST2 encodes a monosaccharide transporter expressed throughout olive fruit maturation. Plant and Cell Physiology, 48: 1299-1308

Conde C., Delrot S., Geros H. 2008. Physiological, biochemical and molecular changes occurring during olive development and ripening. Journal of Plant Physiology, 165: 1545-1562

Corpas F. J., Fernandez-Ocana A., Carreras A., Valderrama R., Luque F., Esteban F. J., Rodriguez-Serrano M., Chaki M., Pedrajas J. 2006. The expression of different superoxide dismutase forms is cell-type dependent in olive (Olea europaea L.) leaves. Plant and Cell Physiology, 47: 984-994

Costanzo M. C., Park J., Balakrishnan R., Cherry J. M., Hong E. L. 2011 Using computational predictions to improve literature-based Gene Ontology annotations. A feasibility study. Database-the Journal of Biological Databases and Curation: bar004

Covas M. I., Konstantinidou M. Fito. 2009. Olive Oil and Cardiovascular Health. Journal of Cardiovascular Pharmacology, 54: 477-482

Crowhurst R. N., Gleave A. P., MacRae E. A., Ampomah-Dwamena C., Atkinson R. G., Beuning L. L., Bulley S. M., Chagne D. 2008. Analysis of expressed sequence tags from Actinidia. Applications of a cross species EST database for gene discovery in the areas of flavor, health, color and ripening. Bmc Genomics, 9:351 http://dx.doi.org/10.1186/1471-2164-9-351 (25. maj. 2013)

Czechowski T., Stitt M., Altmann T., Udvardi M. K., Scheible W. R. 2005 Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiology, 139: 5-17

Damtoft S., Franzyk H., Jensen S. R. 1993. Biosynthesis of secoiridoid glucosides in oleaceae. Phytochemistry, 34: 1291-1299

Dean J., Goodwin D. H., Hsiang T. 2002. Comparison of relative RT-PCR and northern blot analyses to measure expression of beta-1,3-glucanase in Nicotiana benthamiana infected with Colltotrichum destructivum. Plant Molecular Biology Reporter, 20: 347-356

Delseny M., Han B, Hsing Y. I. 2010. High throughput DNA sequencing. The new sequencing revolution. Plant Science, 179: 407-422

Diaz-Sanchez S., Hanning I., Pendleton S., D'Souza D. 2013. Next-generation sequencing. The future of molecular genetics in poultry production and food safety. Poultry Science, 92: 562-572

Dundar E., Suakar O., Unver T., Dagdelen A. 2013. Isolation and expression analysis of cDNAs that are associated with alternate bearing in Olea europaea L. cv. Ayvalik. Bmc Genomics, 14: 219

http://dx.doi.org/10.1186/1471-2164-14-219 (25. maj. 2013)

El Riachy M., Priego-Capote F., Leon L., Rallo L., de Castro M. D. L. 2011. Hydrophilic antioxidants of virgin olive oil. Part 1. Hydrophilic phenols. A key factor for virgin olive oil quality. European Journal of Lipid Science and Technology, 113: 678-691

Esti M., Cinquanta L., La Notte E. 1998. Phenolic compounds in different olive varieties. Journal of Agricultural and Food Chemistry, 46: 32-35

Ewing B., Hillier L., Wendl M. C., Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research, 8: 175-185

Exposito-Rodriguez M., Borges A. A., Borges-Perez A., Perez J. A. 2008. Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. Bmc Plant Biology, 8:131
http://dx.doi.org/10.1186/1471-2229-8-131 (25. maj. 2013)

Fleischmann R. D., Adams M. D., White O., Clayton R. A., Kirkness E. F., Kerlavage A. R., Bult C. J., Tomb J. F., Dougherty B. A. 1995. Whole-genome random sequencing and assembly of haemophilus-influenzae rd. Science, 269: 496-512

Gachon C., Mingam A., Charrier B. 2004. Real-time PCR. what relevance to plant studies?. Journal of Experimental Botany, 55: 1445-1454

Galla G., Barcaccia G., Ramina A., Collani S., Alagna F., Baldoni L.. 2009, Computational annotation of genes differentially expressed along olive fruit development. Bmc Plant Biology, 9: 128
http://dx.doi.org/10.1186/1471-2229-9-128 (25. maj. 2013)

Ganino T., Bartolini G., Fabbri A. 2006. The classification of olive germplasm - A review: Journal of Horticultural Science & Biotechnology, 81: 319-334

Gentleman R. in Ihaka R. 1997. The R project for statistical computing. (25. sept. 2013). http://www.r-project.org/index.html (3. nov. 2013)

Gertz C., Kochhar S. P. 2001. A new method to determine oxidative stability of vegetable fats and oils at simulated frying temperature. Ocl-Oleagineux Corps Gras Lipides, 8: 82-88

Ghanbari R., Anwar F., Alkharfy K. M., Gilani A. H., Saari N. 2012. Valuable Nutrients and Functional Bioactives in Different Parts of Olive (*Olea europaea* L.)-A Review. International Journal of Molecular Sciences, 13: 3291-3340

Goff S. A., Ricke D., Lan T. H., Presting G., Wang R. L., Dunn M., Glazebrook J. 2002, A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science, 296: 92-100.

Gotz, S., R. Arnold: Sebastian-Leon, S. Martin-Rodriguez: Tischler, M. A. Jehl, J. Dopazo, T. Rattei A. Conesa, 2011, B2G-FAR, a species-centered GO annotation repository. Bioinformatics, 27: 919-924

Grundy S. M. 1997. What is the desirable ratio of saturated, polyunsaturated monounsaturated fatty acids in the diet? American Journal of Clinical Nutrition, 66: S988-S990

Gupta K. 2008. Ultrafast and low-cost DNA sequencing methods for applied genomics research. Proceedings of the National Academy of Sciences India Section B-Biological Sciences, 78: 91-102

Hamilton J. P., Buell C. R.. 2012. Advances in plant genome sequencing. Plant Journal, 70: 177-190

Haralampidis K., Milioni D., Sanchez J., Baltrusch M., Heinz E., Hatzopoulos P. 1998. Temporal and transient expression of stearoyl-ACP carrier protein desaturase gene during olive fruit development. Journal of Experimental Botany, 49: 1661-1669

Hellemans J., Mortier G., De Paepe A., Speleman F., Vandesompele J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biology, 8: 19
http://dx.doi.org/10.1186/gb-2007-8-2-r19 (18. jun. 2013)

Hoerndli F. J., Toigo M., Schild A., Gotz J., Day P. J. 2004. Reference genes identified in SH-SY5Y cells using custom-made gene arrays with validation by quantitative polymerase chain reaction. Analytical Biochemistry, 335: 30-41

Huang X., Madan Q. A. 1999. CAP3. A DNA sequence assembly program. Genome Research, 9: 868-877

Huggett J., Dheda K., Bustin S., Zumla A. 2005. Real-time RT-PCR normalisation; strategies and considerations. Genes and Immunity, 6: 279-284

Imelfort M., Edwards D. 2009. De novo sequencing of plant genomes using second-generation technologies. Briefings in Bioinformatics, 10: 609-618

Inoue T., Zhong H. S., Miyao A., Ashikawa I., Monna L., Fukuoka S., Miyadera N., Nagamura Y., Kurata N., Sasaki T., Minobe Y. 1994. Sequence-tagged sites (stss) as standard landmarkers in the rice genome. Theoretical and Applied Genetics, 89: 728-734

Kalua C. M., Allen M. S., Bedgood D. R., Bishop A. G., Prenzler D., Robards K. 2007. Olive oil volatile compounds, flavour development and quality. A critical review. Food Chemistry, 100: 273-286

Kaniewski D., Van Campo E., Boiy T., Terral J. F., Khadari B., Besnard G. 2012. Primary domestication and early uses of the emblematic olive tree. palaeobotanical, historical and molecular evidence from the Middle East. Biological Reviews, 87: 885-899

Kaul S., Koo H. L., Jenkins J., Rizzo M., Rooney T., Tallon L. J., Feldblyum T., Nierman W., Benito M. I. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, 408: 796-815

Kaya H. B., Cetin O., Kaya H., Sahin M., Sefer F., Kahraman A., Tanyolac B. 2013. SNP Discovery by Illumina-Based Transcriptome Sequencing of the Olive and the Genetic Characterization of Turkish Olive Genotypes Revealed by AFLP, SSR and SNP Markers. Plos One, 8
http://dx.doi.org/10.1371/journal.pone.0073674 (25. maj. 2013)

Kececioglu J. D., Myers E. W. 1995. Combinatorial algorithms for dna-sequence assembly. Algorithmica, 13: 7-51

Kent W. J. 2002. BLAT - The BLAST-like alignment tool. Genome Research, 12: 656-664

Kernerman S. M., McCullough J., Green J., Ownby D. R. 1992. Evidence of cross-reactivity between olive, ash, privet, and russian olive tree pollen allergens: Annals of Allergy. 69: 493-496

Kumar S., Blaxter M. L. 2010. Comparing de novo assemblers for 454 transcriptome data. Bmc Genomics, 11: 571
http://dx.doi.org/ 10.1186/1471-2164-11-571 (18. jun. 2013)

Likic A. 2006. Databases of metabolic pathways. Biochemistry and Molecular Biology Education, 34: 408-412

Liu Z. J. 2006. Transcriptome characterization through the generation and analysis of expressed sequence tags. Factors to consider for a successful EST project. Israeli Journal of Aquaculture-Bamidgeh, 58: 328-340

Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., Berka J., Braverman M. S., Chen Y. J., Chen Z. T., Dewell S. B. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437: 376-380

Matsumoto T., Wu J. Z., Kanamori H., Katayose Y., Fujisawa M., Namiki N., Mizuno H., Yamamoto K., Antonio B. A.. 2005. The map-based sequence of the rice genome. Nature, 436: 793-800

Meyers B. C., Axtell M. J., Bartel B., Bartel D. P., Baulcombe D., Bowman J. L., Cao X., Carrington J. C., Chen X. M., Green J. 2008. Criteria for Annotation of Plant MicroRNAs. Plant Cell, 20: 3186-3190

Miller J. R., Koren S., Sutton G. 2010. Assembly algorithms for next-generation sequencing data. Genomics, 95: 315-327

Ming R., Hou S. B., Feng Y., Yu Q. Y., Dionne-Laporte A., Saw J. H., Wang W., Ly B., Lewis K. L. T., Salzberg S. L. 2008. The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature, 452: 991-996

Morales M., Alonso T., Rios M., Aparicio J. J. 1995. Virgin olive oil aroma - relationship between volatile compounds and sensory attributes by chemometrics. Journal of Agricultural and Food Chemistry, 43: 2925-2931

Mundry M., Bornberg-Bauer E., Sammeth M., Feulner P. G. D. 2012. Evaluating Characteristics of De Novo Assembly Software on 454 Transcriptome Data. A Simulation Approach. Plos One, 7:10
http://dx.doi.org/10.1371/journal.pone.0031410 (3. nov. 2013)

Munoz-Merida A., Gonzalez-Plaza J. J., Canada A., Blanco A. M., Garcia-Lopez M. D., Rodriguez J. M., Pedrola L., Sicardo M. D., Hernandez M. L., De la Rosa R. 2013. De Novo Assembly and Functional Annotation of the Olive (Olea europaea) Transcriptome. DNA Research, 20: 93-108

Muzzalupo I., Macchione B., Bucci C., Stefanizzi F., Perri E., Chiappetta A., Tagarelli A., Sindona G. 2012. LOX Gene Transcript Accumulation in Olive (*Olea europaea* L.) Fruits at Different Stages of Maturation. Relationship between Volatile Compounds, Environmental Factors Technological Treatments for Oil Extraction. Scientific World Journal, 62: 3403-3420

Myers E. W.2005. The fragment assembly string graph. Bioinformatics, 21: 79-85

Nagel J., Culley L. K., Lu Y. P., Liu E. W., Matthews D., Stevens J. F., Page J. E. 2008. EST analysis of hop glandular trichomes identifies an O-methyltransferase that catalyzes the biosynthesis of xanthohumol. Plant Cell, 20: 186-200

Newcomb R. D., Crowhurst R. N., Gleave A. P., Rikkerink E. H. A, Allan A. C. 2006. Analyses of expressed sequence tags from apple. Plant Physiology, 141: 147-166

Newman T., Debruijn F. J., Keegstra K., Kende H., McIntosh L., Ohlrogge J., Raikhel N., Somerville S., Thomashow M., Retzel E., Somerville C. 1994. Genes galore - a summary of methods for accessing results from large-scale partial sequencing of anonymous arabidopsis cdna clones. Plant Physiology, 106: 1241-1255

Nicot N., Hausman J. F., Hoffmann L., Evers D. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. Journal of Experimental Botany, 56: 2907-2914

Ning Z. M., Cox A. J., Mullikin J. C. 2001. SSAHA. A fast search method for large DNA databases. Genome Research, 11: 1725-1729

Nonis A., Vezzaro A., Ruperti B. 2012. Evaluation of RNA Extraction Methods and Identification of Putative Reference Genes for Real-Time Quantitative Polymerase Chain Reaction Expression Studies on Olive (*Olea europaea* L.) Fruits. Journal of Agricultural and Food Chemistry, 60: 6855-6865

Ouyang S., Buell C. R. 2004. The TIGR Plant Repeat Databases. a collective resource for the identification of repetitive sequences in plants. Nucleic Acids Research, 32: D360-D363

Ozgenturk N. O., Oruc F., Sezerman U., Kucukural A., Korkut S., Toksoz F., Un C. 2010. Generation and Analysis of Expressed Sequence Tags from Olea europaea L. Comparative and Functional Genomics, 12: 307-321

Paterson A. H., Bowers J. E., Bruggmann R., Dubchak I., Grimwood J., Gundlach H., Haberer G., Hellsten U., Mitros T., Poliakov A., Schmutz J. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature, 457: 551-556

Peltola H., Soderlund E., Ukkonen H. 1984. Seqaid - a dna-sequence assembling program based on a mathematical-model. Nucleic Acids Research, 12: 307-321

Pertea, G., Huang X. Q., Liang F., Antonescu V., Sultana R., Karamycheva S., Lee Y., White J., Cheung F., Parvizi B., Tsai J., Quackenbush J. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics, 19: 651-652

Pfaffl M. W., Tichopad A., Prgomet C., Neuvians T. P. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity. BestKeeper - Excel-based tool using pair-wise correlations. Biotechnology Letters, 26: 509-515

Poghosyan Z. P., Giannoulia K., Murphy D. J., Hatzopoulos P. 2005. Temporal and transient expression of olive enoyl-ACP reductase gene during flower and fruit development. Plant Physiology and Biochemistry, 43: 37-44

Poghosyan Z. P., Haralampidis K., Martsinkovskaya A. I., Murphy D. J., Hatzopoulos P. 1999. Developmental regulation and spatial expression of a plastidial fatty acid desaturase from *Olea europaea*. Plant Physiology and Biochemistry, 37: 109-119

Proost S., Van Bel M., Sterck L., Billiau K., Van Parys T., Van de Peer Y., Vandepoele K. 2009. PLAZA. A Comparative Genomics Resource to Study Gene and Genome Evolution in Plants. Plant Cell, 21: 3718-3731

Ransbotyn T., Reusch B. H. 2006. Housekeeping gene selection for quantitative real-time PCR assays in the seagrass Zostera marina subjected to heat stress. Limnology and Oceanography-Methods, 4: 367-373

Reid K. E., Olsson N., Schlosser J., Peng F., Lund S. T. 2006. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. Bmc Plant Biology, 6: 27 http://dx.doi.org/10.1186/1471-2229-6-27 (18. jun. 2013)

Ren X. W., Liu T., Dong J., Sun L. L., Yang J., Zhu Y. F., Jin Q. 2012. Evaluating de Bruijn Graph Assemblers on 454 Transcriptomic Data. Plos One, 7: 9 http://dx.doi.org/10.1371/journal.pone.0051188 (25. maj. 2013)

Rutter A. J., Sanchez J., Harwood J. L. 1997. Glycerolipid synthesis by microsomal fractions from Olea europaea fruits and tissue cultures. Phytochemistry, 46: 265-272

Ryan D., Antolovich M., Herlt T., Prenzler D., Lavee S., Robards K. 2002. Identification of phenolic compounds in tissues of the novel olive cultivar Hardy's mammoth. Journal of Agricultural and Food Chemistry, 50: 6716-6724

Ryan D., Robards K. 1998. Phenolic compounds in olives. Analyst, 123: 31R-44R.

Sancin V. 1990. Velika knjiga o oljki. Trst, Založništvo tržaškega tiska: 319

Sanchez J. J., Harwood L. 1992. Fatty-acid synthesis in soluble fractions from olive (*olea europaea*) fruits. Journal of Plant Physiology, 140: 402-408

Sanchez J. J., Harwood L. 2002. Biosynthesis of triacylglycerols and volatiles in olives. European Journal of Lipid Science and Technology, 104: 564-573

Schmid H., Cohen C. D., Henger A., Irrgang S., Schlondorff D., Kretzler M. 2003. Validation of endogenous controls for gene expression analysis in microdissected human renal biopsies. Kidney International, 64: 356-360

Schmutz J., Cannon S. B., Schlueter J., Ma J. X., Mitros T., Nelson W. 2010. Genome sequence of the palaeopolyploid soybean (vol 463, pg 178, 2010). Nature, 465: 120-120

Senerchia, N., Wicker T., Felber F., Parisod C. 2013. Evolutionary Dynamics of Retrotransposons Assessed by High-Throughput Sequencing in Wild Relatives of Wheat. Genome Biology and Evolution, 5: 1010-1020

Secchi, F., Lovisolo C., Uehlein N., Kaldenhoff R., Schubert A. 2007. Isolation and functional characterization of three aquaporins from olive (*Olea europaea* L.). Planta, 225: 381-392

Servili M., Montedoro G. 2002. Contribution of phenolic compounds to virgin olive oil quality. European Journal of Lipid Science and Technology, 104: 602-613

Sequence cleaner. 2010. Dice Holdings Inc. (4. sep. 2013). http://sourceforge.net/projects/seqclean/ (3. nov. 2013)

Shagin D. A., Rebrikov D., Kozhemyako B., Altshuler I. M., Shcheglov A. S., Zhulidov A., Bogdanova E. A., Staroverov D. B., Rasskazov A., Lukyanov S. 2002. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Research, 12: 1935-1942

Shendure J. H., Ji L. 2008. Next-generation DNA sequencing. Nature Biotechnology, 26: 1135-1145

Shibata I., Sato K., Hayatsu N., Shiraki T., Ishii Y., Arakawa T., Hara A., Ohsato N., Izawa M., Aizawa K., Itoh M., Shibata K., Shinagawa A. 2001. Removal of polyA tails from full-length cDNA libraries for high-efficiency sequencing. Biotechniques, 31: 1042-1049

Soderlund C., Johnson E., Bomhoff M., Descour A. 2009. PAVE. Program for assembling and viewing ESTs. Bmc Genomics, 10: 400 http://dx.doi.org/10.1186/1471-2164-10-400 (18. jun. 2013)

Soler-Rivas C., Espin J. C., Wichers H. J. 2000. Oleuropein and related compounds. Journal of the Science of Food and Agriculture, 80: 1013-1023

Staden R.. 1979. Strategy of dna sequencing employing computer-programs. Nucleic Acids Research, 6: 2601-2610

Stajich J. E., Block D., Boulez K., Brenner S. E., Chervitz S. A. 2002. The bioperl toolkit. Perl modules for the life sciences. Genome Research, 12: 1611-1618

Tilton G. B., Shockey J. M., Browse J. 2004. Biochemical and molecular characterization of ACH2, an acyl-CoA thioesterase from Arabidopsis thaliana. Journal of Biological Chemistry, 279: 7487-7494

Troncoso-Ponce M. A., Kilaru A., Cao X., Durrett T. P., Fan J. L., Jensen J. K., Thrower N. A., Pauly M., Wilkerson C., Ohlrogge J. B. 2011. Comparative deep transcriptional profiling of four developing oilseeds. Plant Journal, 68: 1014-1027

Tuskan G. A., DiFazio S., Jansson S., Bohlmann J., Grigoriev I., Hellsten U., Putnam N., Ralph S., Rombauts S., Salamov A. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science, 313: 1596-1604

Uccella N. 2000. Olive biophenols. novel ethnic and technological approach. Trends in Food Science & Technology, 11: 328-339

Umehara Y., Inagaki A., Tanoue H., Yasukochi Y., Nagamura Y. 1995. Construction and characterization of a rice yac library for physical mapping. Molecular Breeding, 1: 79-89

Valasek M. A., Repa J. J. 2005. The power of real-time PCR. Advances in Physiology Education, 29: 151-159

Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology, 3: 0034-003.11

Vecchietti A., Lazzari B., Ortugno C., Bianchi F., Malinverni R., Caprera A., Mignani I., Pozzi C. 2009. Comparative analysis of expressed sequence tags from tissues in ripening stages of peach (Prunus persica L. Batsch): Tree Genetics & Genomes, 5: 377-391

Viola M. 2009. Virgin olive oil as a fundamental nutritional component and skin protector. Clinics in Dermatology, 27: 159-165

Vossen. 2007. Olive oil. History, production characteristics of the world's classic oils. Hortscience, 42: 1093-1100

Wallander E., Albert A. 2001. Phylogeny and classification of Oleaceae based on rps16 and trnL-F sequence data. (vol 87, pg 1827, 2000). American Journal of Botany, 88: 390-390

Wang G. D., Tian L., Aziz N., Dai X. B., He J., King A., Zhao X., Dixon R. A. 2008. Terpene Biosynthesis in Glandular Trichomes of Hop. Plant Physiology, 148: 1254-1266

Weber A. P. M., Weber K. L., Carr K., Wilkerson C., Ohlrogge J. B. 2007. Sampling the arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiology, 144: 32-42

Xie F. L., Sun G. L., Stiller J. W., Zhang B. H. 2011. Genome-Wide Functional Analysis of the Cotton Transcriptome by Creating an Integrated EST Database. Plos One, 6: 11

Yu J., Hu S. N., Wang J., Wong G. K. S., Li S. G., Liu B., Deng Y. J., Dai L. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). Science, 296: 79-92

Zheng Y., Zhao L. J., Gao J. P., Fei Z. J. 2011. iAssembler. a package for de novo assembly of Roche-454/Sanger transcriptome sequences. Bmc Bioinformatics, 12: 453 http://dx.doi.org/10.1186/1471-2105-12-453 (18. jun. 2013)

Zhu Y. Y., Machleder E. M., Chenchik A., Li P., Siebert D. 2001. Reverse transcriptase template switching. A SMART (TM) approach for full-length cDNA library construction. Biotechniques, 30: 892-897

Zhulidov A., Bogdanova E. A., Shcheglov A. S., Vagner L. L., Khaspekov G. L. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Research, 32: 37 http://dx.doi.org/10.1093/nar/gnh031 (18. jun. 2013)

Zohary D., Spiegelroy P. 1975. Beginnings of fruit growing in old world. Science, 187: 319-327

ZAHVALA

Sprva bi se rada zahvalila mentorju doc. dr. Jerneju Jakšetu za vso ažurno pomoč in prepotrebno usmerjanje, ko sem ga potrebovala. Brez njega bi bila ta doktorska naloga veliko težje izvedljiva, končni izdelek pa gotovo ne bi bil tako dober.

Zahvalila bi se tudi sodelavcem iz Katedre za genetiko, biotehnologijo, statistiko in žlahtnjenje rastlin za čudovito izkušnjo, ki sem jo doživela v njihovi družbi in dejstvo, da so mi bili v slehernem trenutku pripravljeni pomagati pri mojem delu. Z istimi lepimi mislimi pozdravljam tudi sodelavce iz Znanstveno raziskovalnega središča na Primorskem.

Zahvaljujem se tudi Komisiji za oceno in zagovor, za konstruktivne predloge, zaradi katerih sem lahko postavila piko na i svoji doktorski nalogi in njihov hiter pregled, ki mi je omogočil, da doktoriram ravno pred najbolj praznični dnevi v letu.

V neverjetno pomoč mi je bila tudi podpora moje družine in partnerja, ki so mi stali ob strani tudi v trenutkih, ko mi je šlo najtežje in so me vseskozi iskreno podpirali pri mojem delu. Brez njihovega doprinosa ta doktorska naloga ne bi ugledala luči sveta.

Na koncu bi se zahvalila še vsem prijateljem, ki so vedno našli spodbudno besedo zame in vsem ostalim, ki jih je razveselilo dejstvo, da sem zaključila zelo pomembno poglavje v svojem življenju.

Iskrena hvala.