Oznaka poročila: ARRS-RPROJ-ZP-2013/117

ZAKLJUČNO POROČILO RAZISKOVALNEGA PROJEKTA

A. PODATKI O RAZISKOVALNEM PROJEKTU

1.Osnovni podatki o raziskovalnem projektu

Šifra projekta	J4-2296
Naslov projekta	Transkriptom plodu oljke ? razvoj oznak izraženih zaporedij (EST) oljčnega plodu za študije tkivno specifičnih metabolnih poti
Vodja projekta	16379 Jernej Jakše
Tip projekta	J Temeljni projekt
Obseg raziskovalnih ur	4171
Cenovni razred	C
Trajanje projekta	05.2009 - 04.2012
Nosilna raziskovalna organizacija	481 Univerza v Ljubljani, Biotehniška fakulteta
Raziskovalne organizacije - soizvajalke	1510 Univerza na Primorskem, Znanstveno-raziskovalno središče Universita del Litorale Centro di ricerche scientifiche
Raziskovalno področje po šifrantu ARRS	4 BIOTEHNIKA 4.03 Rastlinska produkcija in predelava 4.03.01 Kmetijske rastline
Družbeno-ekonomski cilj	08. Kmetijstvo

2.Raziskovalno področje po šifrantu FOS[1]

Šifra	4.04	
- Veda	4	Kmetijske vede
- Področje	4.04	Kmetijska biotehnologija

B. REZULTATI IN DOSEŽKI RAZISKOVALNEGA PROJEKTA

3.Povzetek raziskovalnega projekta[2]

SLO

Začetna faza projekta je zajemala izdelavo podatkovne baze transkriptov (EST zaporedij), ki naj bi obsegala zaporedja pridobljena iz razvijajočega plodu oljke. Za ta namen smo

skozi celotno dobo razvoja oljčnega plodu vzorčili plodove oljk na tedenski osnovi. Za vzorčenje smo izbrali slovenski kultivar 'Istrska belica', ki izstopa po značilnem profilu biofenolnih substanc. V drugi fazi tega projekta je bila iz plodov oljk, ki smo jih vzorčili skozi celotno obdobje zorenja, izdelana normalizirana cDNA knjižnica. Pred uporabo knjižnice je bilo opravljenih več testov njene kakovosti. Pridobljeno normalizirano knjižnico smo pomnožili in del knjižnice transformirali v bakterije ter shranili za nadaljno uporabo pri -80°C. Ostanek knjižnice smo uporabili za nadalnje sekvenciranje s pomočjo novih tehnologij sekvenciranja (Roche 454). Pred samim sekvenciranjem z Roche 454 tehnologijo smo cDNA obdelali z *Gsu*I encimom in se tako poskusili izognili težavam, ki jih povzroča polyA konec pri določanju nukleotidnega zaporedja. Raziskovalno delo je bilo nato usmerjeno v nadaljno obdelavo sekvenčnih podatkov, pridobljenih iz mRNA razvijajočih plodov. Po bioinformatski obdelavi podatkov sekvenciranja z različnimi bioinformatskimi orodji, smo prečiščena zaporedja na osnovi podobnosti združil v domnevna konsenzus zaporedja z uporabo različnih zbirnikov. Uporabili smo naslednje programe za združevanje zaporedij: TGICL, MIRA, iAssembler, Newbler 2.3. in 2.6., pave 2.5 in CLC Genomic Workbench 4.5. 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. Ti kriteriji so vključevali osnovne podatke meritev različnih programov za združevanje zaporedij, BLAT primerjavo zaporedij, ki nam poda število unikatnih zaporedij posameznega združevanja v primerjavi z ostalimi podatki, ter podatke ujemanja z referenčno bazo podatkov. Rezultate programa, ki se je izkazal za najboljšega, smo uporabil pri nadaljni funkcijski analizi podatkov s programom Blast2go, s pomočjo katerega smo posameznim sekvencam pripisal vlogo na ravni bioloških procesov, celičnih komponent in molekularnih funkcij. Z Blast2go anotacijo smo pridobil tudi podatke o skupinah genov, ki so vključeni v biosintetične procese maščobnih kislin, v procese razvoja plodov in procese sekundarnih metabolitov. Zaključil smo tudi RT-qPCR analizo določanja seta transkriptov za referenčne gene, ki bodo imeli pomembno vlogo pri nadaljnem določanju tkivno specifičnega izražanja nekaterih ključnih transkriptov in z analizo izražanja izbranih genov vključenih v metabolizem maščobnih kislin.

ANG

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', which is outstanding for its high proportion of biophenols, 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. Several quality tests were conducted before the library was used. The acquired normalized cDNA library was amplified and partly transformed into bacteria to be stored for further use at -80° C. The remainder of the library was used for sequencing, using next generation sequencing technologies (Roche 454). Before sequencing with Roche 454 pyrosequencing technology, the cDNA was treated with *Gsu*I enzyme to avoid problems that can be caused by polyA tails. The research then focused on further bioinformatics processing of the sequence data obtained from the mRNA of developing olive fruits. After processing the sequencing data with various bioinformatics tools, we wanted to reconstruct the obtained cDNA sequences in the longest possible transcripts. In order to achieve this, we compared the performance of various assembly programs that are available, in order to identify the most suitable program for the analysis of olive 454

transcriptome. The following assembly programs were tested: TGICL, MIRA, iAssembler, gsAssembler 2.3 and 2.6, pave 2.5 and CLC genomic Workbench 4.5. We took several criteria into consideration when selecting the best *de novo* assembly, including assembly statistics, ratio of novel sequences and alignments to reference database sequences. The results of the best assembling program, iAssembler, were further used for functional annotation with Blast2GO. 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. Annotation defined groups of genes involved in fatty acid biosynthetic processes, fruit development and secondary metabolic processes. We also completed RT-qPCR analysis of transcripts to determine a set of reference genes, which will play an important role in further determination of tissue-specific gene expression and expression analysis of selected genes involved in the metabolism of fatty acids.

4.Poročilo o realizaciji predloženega programa dela na raziskovalnem projektu[3]

V nadaljevanju opisujemo aktivnosti, ki so potekale v sklopu časovnega poteka projekta.

1) VZORČENJE OLJK IN IZOLACIJA RNA

Plodove oljk sorte 'Istrska belica' smo vzorčili skozi celotno obdobje razvoja v dveh letih. Vzorčenje je potekalo od 1. 6. 2008 do 24. 11. 2008 (leto 2008) ter od 9. 6. 2009 do 11. 11. 2009 (leto 2009). Na koncu smo imeli 22 časovnih točk vzorčenja skozi celotno obdobje razvoja oljčnega plodu. Za izolacijo RNA smo prvotno uporabili metodo, ki uporablja TRIZOL reagent, ki je bil originalno opisan v delu Chomzynski (1993). Z uporabo te metode nismo pridobili optimalnih rezultatov, zato smo za izolacijo RNA naknadno uporabili Spectrum Total Plant RNA Extraction Kit, ki se je izkazal za primernega (Priloga 1). Na koncu smo zmešali ekvimolarne količine vseh vzorcev obdobja razvoja, da smo dobili združen, reprezentativen vzorec vseh RNA izraženih v celotnem razvojnem obdobju oljčnega plodu.

2) IZDELAVA, OBDELAVA IN KARAKTERIZACIJA NORMALIZIRANE cDNA KNJIŽNICE

Reprezentativen vzorec združene RNA smo poslali v servis za izdelavo normalizirane cDNA knjižnice (Evrogen Lab, Russia). Celokupna RNA je bila uporabljena za izdelavo ds cDNA, s pomočjo SMART tehnologije. Tako pripravljeno, amplificirano cDNA so nato normalizirali z uporabo DSN normalizacijske metode. Normalizacija je vključevala cDNA denaturacijo, tretiranje z duplex-specifično nukleazo in PCR pomnoževanje.

Pridobljeno normalizirano cDNA knjižnico smo okarakterizirali in ji določili koncentracijo. Del knjižnice smo ligirali v pGEM-T-easy plazmid. Na ta način bomo knjižnico ohranili za daljše časovno obdobje. Preostanek knjižnice pa smo uporabili za nadalnje določevanje nukleotidnega zaporedja s pomočjo novih tehnologij sekvenciranja (Roche 454). Ker pa prisotnost homopolimernih odsekov, kot so poly A/T konci, povzroča težave pri določevanju nukleotidnih zaporedij v cDNA knjižnicah, 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 (ki smo ga vnesli z adapterjem). Po odstranitvi smo cDNA dodali adapterje in knjižnico pomnožili v PCR z minimalnim številom ciklov. Na ta način smo se poizkusili izognili težavam, ki jih povzroča poly-A regija pri določanju nukleotidnega zaporedja.

Sledilo je primerjalno določevanje nukleotidnega zaporedja za preizkus uspešnosti izreza poly-A regije. Primerjali smo rezultate določevanja nukleotidnega zaporedja klonom

cDNA in *Gsu*I-cDNA. Iz vsake knjižnice smo 192-im klonom v PCR pomnožili cDNA insert. Za prvo cDNA knjižnico, ki ni bila tretirana z *Gsu*I encimom, smo naredili obojestransko sekvenčno rekacijo (384 reakcij), za drugo cDNA knjižnico, ki je bila tretirana z *Gsu*I encimom, pa smo naredili enostransko sekvenčno reakcijo (192 reakcij). Pri prvi knjižnici je polovica sekvenčnih reakcij bila nezadovoljive kakovostni, prvenstveno zaradi prisotnih poly-A regij. Od skupno 384 zaporedij jih je bilo le 158 (41 %) uporabnih za nadaljno analizo. 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. Rezultat pa vseeno potrjuje dobro opravljen proces normalizacije, saj ni presežnih zaporedij v knjižnici. Iz druge knjižnice pa smo določili nukleotidno zaporedje 192-im vzorcem samo iz ene strani, uspešnost izvedbe večine sekvenčnih reakcij bi potrdila smiselnost odstranjevanja poly-A regije. Od 192 zaporedij je bilo kar 158 (81 %) primernih za nadaljno analizo, kar pomeni, da smo se pomočjo *Gsu*I restrikcijskega encima uspešno odstranili poly-A regije. Samo manjši delež zaporedij je še vseboval homopolimerne regije. Zaporedja v tej knjižnici so bila v povprečju dolga 484 bp, skupno smo določili 76.485 bp dolžine DNA. Na koncu smo vsa zaporedja združili skupaj in preverili redundantnost, skupno smo določili 112.134 bp DNA in 287 enkratnih zaporedij v povprečni dolžini 390 bp. Na podalgi teh rezulatatov smo se odločili, d aje smiselno sekvenciranje *Gsu*I tretirane knjižnice.

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. Nukleotidno zaporedje smo določili polovici regije pikotiterske plošče, kjer dobimo do 500.000 zaporedij (točk) (Priloga 2). Ker smo sekvencirali konkatemere cDNA (združene cDNA oz. parna zaporedja) smo morali hibridna zaporedja razdružiti glede na prisotnost SNX linkerja, ki obdaja cDNA. Zato smo uporabili skripto sff_extract skupaj z rutinami programa SSAHA2. Zaporedja, ki smo jih pridobili, smo vključili tudi v proces čiščenja zaporedij. S skripto seqclean, ki uporablja megablast algoritem, smo odstranili dele zaporedij z ostanki linkerjev in prekratka zaporedja (pod 75 bp). Tako smo na koncu pridobili 577.025 zaporedij v skupni dolžini 139.419.844 bp. Povprečna dolžina zaporedij je bila 241 bp, medtem ko je N50 vrednost bila 294 bp (192.189 vseh zaporedij). Vsebnost GC je bila 40,90 %. Podatki so navoljo na spletu preko NCBI SRA arhiva (http://www.ncbi.nlm.nih.gov/sra/SRX215662).

4) BIOINFORMATSKA OBDELAVA

Teh 577-tisoč zaporedij predstavlja končna cDNA zaporedja oljke, ki smo jih analizirali v naslednjem koraku združevanja, kjer smo želeli pravilno rekonstruirati (zložiti) zaporedja cDNA in pridobiti čim daljšo možno dolžino. Ta korak je bil tudi najbolj delovno zahteven. V tem koraku smo se odločili za podrobnejšo analizo našega seta podatkov z različnimi programi, ki so na voljo. Uporabili smo naslednje programe za združevanje zaporedij: TGICL (ki je skripta, ki uporablja program CAP3), MIRA, iAssembler (ki je skripta, ki istočasno uporablja CAP3 in MIRA), gsAssembler 2.3 (originalni Rochev program za združevanje), pave 2.5 in CLC Genomic Workbench 4.5 (program, ki uporablja novejšo metodologijo združevanja, ki temelji na metodi uporabe de brujin grafa) (Priloga 3). Kjer je bilo možno, smo kot merilo združevanja uporabili 96 % identičnost in minimalno prekrivanje 40 bp. Omenimo naj, da so vsi programi razen CLC-ja odprtokodni in brezplačni. Po koncu analize smo rezultate razdelili v dve skupini, in sicer v združena zaporedja in v preostala zaporedja (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 rezultate združevanja in njihovo hitrost združevanja. Rezultate združevanja smo ocenili tudi s kartiranjem združenih sekvenc na lokalno izdelane podatkovne baze proteinov, te pa smo primerjali z BLASTX in BLAT algoritmoma (Priloga 4).
Rezultati združevanja so se zelo razlikovali med programi po številu združenih kontigov in po količini sekvenčne informacije, ki so jo bili sposobni uporabiti in po številu zaporedij, ki so ostala nezdružena. Najoptimalnejši zbirnik bo združil največ sekvenc v najdaljša zaporedja in ostalo bo minimalno število preostalih zaporedij. V tej kategoriji ocenjevanja se je najslabše izkazal zbirnik Newbler 2.3, saj je lahko združil 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, MIRA in iAssembler (50.000). Zbirnik iAssmbler je združil največ kontigov daljših od 1.000 bp (2.363) in tudi največ kontigov daljših od 500 bp (21.879). Najdaljše zaporedje 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 pri ocenjevanju združenih podatkov. V tej kategoriji se je najslabše izkazal CLC program, medtem ko so ostali bili primerljivi z dolžinami okrog 500 bp. Po hitrosti pa 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 (Priloga 4).
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. Za primerjavo smo uporabili program BLAT, ki izvede globalne primerjave zaporedij med sabo. Analiza je pokazala, da zbirnik iAssembler zavzame zaporedja tudi ostalih zbirnikov. Najslabše sta se odrezali obe verziji programa Newbler, s preko 10.000 zaporedji, ki jih nimata zastopanih. Ostali štirje zbirniki so bili med sabo primerljivi (Priloga 4).
Na koncu smo izvedli še 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. 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 (Priloga 4).
Z upoštevanjem vseh parametrov analize zbirnikov smo določili, da program iAssembler zajame največjo skupino transkriptov oljke.
Rezultate programa, ki se je izkazal za najboljšega smo uporabil pri nadaljni funkcijski analizi podatkov s programom Blast2go, s pomočjo katerega smo 51% sekvenc uspešno pripisal vlogo na ravni bioloških procesov, celičnih komponent in molekularnih funkcij. Z Blast2go anotacijo smo pridobil tudi podatke o skupinah genov, ki so vključeni v biosintetične procese maščobnih kislin (264 sekvenc, GO: 0006633), v procese razvoja plodov (24 sekvenc, GO: 0010154) in procese sekundarnih metabolitov (950 sekvenc, GO: 0019748). S pomočjo baze KEGG pathway pa smo pridobili sheme procesnih poti v katere so vključene anotirane sekvence.
Za bioinformatski del analize imamo v pripravi objavo.

5) RT-qPCR analiza
Izvedli smo tudi sklop analiz kvantitativnega PCR oljčnih genov. V prvi fazi smo morali izbrati set referenčnih genov, ki bi bil uporaben za kvantitativne analize. Plodove oljk smo vzorčili v petih glavnih fazah razvoja, ter na koncu dobili 12 vzorčnih točk. S pomočjo literature smo izbrali 29 potencialnih referenčnih genov, med katerimi so bili tradicionalno uporabljeni geni in tudi novejši referenčni geni. S programom BLASTX smo v naših zaporedjih odkrili najbolj identična zaporedja oljke, ki smo jih uporabili v nadaljnem poskusu uporabnosti teh genov za interne kontrole pri genski ekspresiji oljke. Stabilnost potencialnih referenčnih genov, je bila ocenjena z uporabo geNorm programa. V naši raziskavi sta bila TIP41-like family protein (*TIP41*) in TATA binding protein (*TBP*) identificirana kot najbolj stabilna gena. Prišli smo do ugotovitve, da kombinacija dveh referenčnih genov (*TIP41* in *TBP*) zadostuje za normalizacijo, saj vključitev tretjega gena ni povečala variacije. Oba referenčna gena smo nato uporabili pri normalizaciji nivoja ekspresije štirih genov, ki so potencialno vključeni v metabolizem maščobnih kislin (fatty acyl-ACP thioesterase A, *FatA*; poplar; stearoyl-ACP desaturase, *SAD1*; acyl-CoA thioesterase family protein, *Acot*; lipoxygenase 1, *LOX1*) in so pokazali različne vzorce izražanja, povezane z razvojem mezokarpa in zorenjem plodov.
Za delo qPCR analize imamo objavo v recenziji v reviji Molecular Breeding (Priloga 5).

5.Ocena stopnje realizacije programa dela na raziskovalnem projektu in zastavljenih raziskovalnih ciljev[4]

Projektna skupina ocenjuje stopnjo realizacije programa dela in zastavljenih raziskovalnih ciljev kot uspešno. Vse zastavljene cilje smo uspešno realizirali:
1) Vzorčenje plodov, izolacija RNA: vzorčenje plodov oljke sorte 'Istrska belica' je potekalo v dveh rastnih sezonah (2008 in 2009) skozi celo dobo nastanka in razvijanja plodov (od julija do decembra). Za vzorčenje smo izbrali drevesa (biološke ponovitve) v redno oskrbovanem in negovanem nasadu. Plodove smo vzorčili tedensko, takoj po obiranju smo jih zamrznili v tekočem dušiku in shranili do uporabe pri -80 ^{o}C. Iz vzorcev smo izolirali RNA, ki je še vedno na voljo v laboratoriju.
2) Izdelali smo normalizirano cDNA knjižnico: normalizacija knjižnice nam omogoča bolj enakomerno zastopanost nukleotidnih zaporedij v vzorcu. Za ta namen smo uporabili storitev servisa Evrogen. Tako nenormalizirano, kot normalizirano knjižnico smo shranili za dolgotrajno shranjevanje v cDNA bakterijski knjižnici, ki je na voljo zainteresiranim uporabnikom in za naše nadaljne delo. Tukaj smo uporabili tudi inovativno odstranjevanje poly-A koncev mRNA z uporabo *GsuI* restrikcijskega encima.
3) Določanje nukleotidnega zaporedja: z izbiro NGS pristopa določevanja nukleotidnega zaporedja s 454 pirosekvenčno tehnologijo smo pridobili kar 560,578 zaporedij v skupni dolžini 160 Mbp sekvenčnih podattkov. Ti so javno dostopni preko NCBI SRA podatkovne baze na naslovu http://www.ncbi.nlm.nih.gov/sra/SRX215662
4) Bioinformatska obdelava podatkov: sekvenčni podatki za oljkin transkriptom so bili prvi NGS podatki v našem laboratoriju s katerimi smo se srečali pri obdelavi. Posebnost obdelave teh podatkov je njihova količina, kar zahteva tudi posebno strojno in programsko opremo. Z obdelavo teh podatkov smo uspešno osvojili potrebne veščine pregledovanja zaporedij, čiščenja in zlaganja. Pri zlaganju smo testirali različne programe in jih primerjali med sabo. Za funkcijsko analizo oljčnih zaporedij smo uporabili programski paket blast2go. Vsakemu oljčnemu EST zaporedju smo določili najbližji proteinski BLASTX zadetek, ki smo ga analizirali z omenjenim programom. Posebej smo izluščili skupino transkriptov, povezanih s sekundarnim metabolizmom. Bioinformatski del rezultatov je v pripravi za objavo.
4) Analiza genske ekspresije s PCR v realnem času: v sklopu tega dela smo uspešno določili oljčne gene za normalizacijo reakcije in preverili časovno specifično izražanje

genov vpletenih v primarni metabolizem maščobnih kislin. Delo je v recenziji v reviji Molecular Breeding.
5) Identifikacija EST zaporedij, uporabnih pri genotipizaciji oljk: analiza zaporedij je potrdila, da so le-ta vir mikrosatelitnih zaporedij in tudi polimorfizmov enojnih nukleotidov. Informacije o mikrosatelitih smo delili s Francosko raziskovalno skupino iz INRA Montpellier, ki so razvili delujoče mikrosatelitne lokuse (predstavitev na kongresu VIIth International symposium on olive growing, San Juan - Argentina, 2012).

6.Utemeljitev morebitnih sprememb programa raziskovalnega projekta oziroma sprememb, povečanja ali zmanjšanja sestave projektne skupine[5]

V prvem letu izvajanja projekta »J4-2296: Transkriptom plodu oljke: razvoj oznak izraženih zaporedij (EST) oljčnega plodu za študije tkivno specifičnih metabolnih poti« je prišlo do spremembe uporabe predlagane tehnologije določevanja nukleotidnih zaporedij. V projektnem predlogu smo predvideli določevanje nukleotidnega zaporedja 7.000-im klonom z uporabo Sangerjeve tehnologije. V tem času je prišlo do bistvenih sprememb na področju tehnologije določevanja DNA zaporedij z uvedbo t.i. tehnologij naslednje generacije (NGS). Zaradi tega smo se odločili za uporabo Rocheve tehnologije 454, ki je po dolžini pridobljenega zaporedja primerljiva s Sangerjevo tehnologijo, cena pridobljene baze pa je pri 454 tehnologiji 100x nižja. Na ta način smo pridobili večjo količino podatkov – 500 Mbp kot v primerjavi s prvotno predlagano Sangerjevo tehnologijo, kjer bi pridobili okrog 5 Mbp.
Ostalih sprememb povezanih s predvidenim programom raziskovalnega projekta, kot je bil zapisan v predlogu ali s povečanjem ali zmanjšanjem sestave projektne skupine ni bilo.

7.Najpomembnejši znanstveni rezultati projektne skupine[6]

	Znanstveni dosežek			
1.	COBISS ID		2176211	Vir: COBISS.SI
	Naslov	*SLO*	Analiza transkriptoma razvijajočega plodu oljke (Olea europaea L.) z uporabo naslednjih generacij določevanja nukleotidnih zaporedij	
		ANG	Transcriptome analysis of developing olive fruit (Olea europaea L.) using next generations sequencing technology	
	Opis	*SLO*	Avtor je na mednarodnem posvetu o novih raziskovalnih pristopih v oljkarstvu, ki je potekal februarja 2012 v Kopru, predstavil znanstveni prispevek z naslovom Analiza transkriptoma razvijajočega plodu oljke (Olea europaea L.) z uporabo naslednjih generacij določevanja nukleotidnih zaporedij.	
		ANG	Author presented a scientific paper entitled Transcriptome analysis of developing olive fruits (Olea europaea L.) using next generations sequencing technology at the international conference on new research approaches in olive growing, which took place in Koper in February 2012.	
	Objavljeno v		Univerza na Primorskem, Znanstveno-raziskovalno središče, Univerzitetna založba Annales; Novi raziskovalni pristopi v oljkarstvu; 2012; Str. 129-137; Avtorji / Authors: Rešetič Tjaša, Bandelj Mavsar Dunja, Javornik Branka, Jakše Jernej	
	Tipologija		1.08 Objavljeni znanstveni prispevek na konferenci	

8.Najpomembnejši družbeno-ekonomski rezultati projektne skupine[7]

	Družbeno-ekonomski dosežek			
1.	COBISS ID		7268985	Vir: COBISS.SI
	Naslov	*SLO*	Analiza izraženih nukleotidnih zaporedij tekom razvoja oljčnih plodov (Olea europaea) z uporabo 454 pirosekvenciranja	
		ANG	EST analysis of genes during fruit development in Olea europaea (olive) using 454 pyrosequencing	
	Opis	*SLO*	Avtorica je na znanstvenem kongresu Slovenskega genetskega društva s plakatom predstavila svoj znanstveni prispevek z naslovom Analiza izraženih nukleotidnih zaporedij tekom razvoja oljčnih plodov (Olea europaea) z uporabo 454 pirosekvenciranja. Kongres je potekal v Mariboru, septembra 2012.	
		ANG	Author of the poster presentation entitled EST analysis of genes during fruit development in Olea europaea (olive) using 454 pyrosequencing presented his work at the Scientific Congress of Slovenian Genetic Society. Congress was held in Maribor, Slovenia, September 2012.	
	Šifra		B.03 Referat na mednarodni znanstveni konferenci	
	Objavljeno v		Genetic Society of Slovenia; Genetika 2012; 2012; Str. 159; Avtorji / Authors: Rešetič Tjaša, Bandelj Mavsar Dunja, Javornik Branka, Jakše Jernej	
	Tipologija		1.12 Objavljeni povzetek znanstvenega prispevka na konferenci	
2.	COBISS ID		2086099	Vir: COBISS.SI
	Naslov	*SLO*	Transkriptom razvijajočega plodu oljke (Olea europaea L.) pridobljen z naslednjo generacijo določevanja nukleotidnih zaporedij	
		ANG	Transcriptome of developing olive fruit (Olea europaea) obtained by next generation sequencing technology	
	Opis	*SLO*	Na koloviju iz genetike Slovenskega genetskega društva, ki je potekal septembra 2011, je avtorica predstavila znanstveni prispevek z naslovom Transkriptom razvijajočega plodu oljke (Olea europea L.) pridobljen z naslednjo generacijo dolečevanja nukleotidnih zaporedij. V sklopu dela je kandidatka opravljala doktorat znanosti.	
		ANG	At genetic colloquium of Slovenian Genetic Society, which was held in September 2011, the author has presented a scientific paper entitled Transcriptome of developing olive fruit (Olea europaea) obtained by next generation sequencing technology. The candidate has doing his Ph.D. studies, while she was involved in research activities.	
	Šifra		D.09 Mentorstvo doktorandom	
	Objavljeno v		Slovensko genetsko društvo; 2. kolokvij iz genetike, Piran, 16. september 2011; 2011; Str. 70; Avtorji / Authors: Rešetič Tjaša, Bandelj Mavsar Dunja, Javornik Branka, Jakše Jernej	
	Tipologija		1.12 Objavljeni povzetek znanstvenega prispevka na konferenci	
3.	COBISS ID		7029625	Vir: COBISS.SI
	Naslov	*SLO*	Podatki pirosekvenciranja za sestavo transkriptoma oljčnega plodu (Olea europaea) in primerjavo uspešnosti različnih programov za združevanje zaporedij	
		ANG	Pyrosequencing data for the de novo assembly of the olive fruit (Olea europaea L.) transcriptome and performance comparison of several assemblers	
	Opis	*SLO*	V okviru 9. kongresa Slovenskega biokemijskega društva, ki je potekal v Mariboru oktobra leta 2011, je avtor predstavil znanstveni prispevek z naslovom Podatki pirosekvenciranja za sestavo transkriptoma oljčnega	

			plodu (Olea europaea) in primerjavo uspešnosti različnih programov za združevanje zaporedij.
		ANG	AT the 9th Congress of Slovenian Biochemical Society, which was held in Maribor in October of 2011, the author presented a scientific paper Pyrosequencing data for the de novo assembly of the olive fruit (Olea europaea L.) transcriptome and performance comparison of several assemblers.
	Šifra		B.03 Referat na mednarodni znanstveni konferenci
	Objavljeno v		Zavod za zdravstveno varstvo; Abstract book; 2011; Str. 109; Avtorji / Authors: Rešetič Tjaša, Bandelj Mavsar Dunja, Javornik Branka, Jakše Jernej
	Tipologija		1.12 Objavljeni povzetek znanstvenega prispevka na konferenci
4.	COBISS ID		6196601 Vir: COBISS.SI
	Naslov	*SLO*	Izdelava transkriptoma razvijajočega plodu oljke (Olea europaea)
		ANG	Towards the transcriptome of the developing olive Olea europaea L. fruit
	Opis	*SLO*	Avtor je na mednarodni konferenci Slovenskega biokemijskega in genetskega društva, ki je potekal septembra 2009, s predstavil znanstveni prispevek z naslovom Izdelava transkriptoma razvijajočega plodu oljke (Olea europaea).
		ANG	Author of the poster presentation entitled towards the transcriptom of the developing olive Olea europaea L. fruit presented his work at the joint congress of the Slovenian Biochemical Society and the Genetic Society of Slovenia with international participation.
	Šifra		B.03 Referat na mednarodni znanstveni konferenci
	Objavljeno v		Slovenian Biochemical Society;Genetic Society of Slovenia; Book of abstracts; 2009; Str. 177; Avtorji / Authors: Rešetič Tjaša, Bandelj Mavsar Dunja, Javornik Branka, Jakše Jernej
	Tipologija		1.12 Objavljeni povzetek znanstvenega prispevka na konferenci
5.	COBISS ID		7484281 Vir: COBISS.SI
	Naslov	*SLO*	Molekulska orodja za genetsko mapiranje in asociacijske študije
		ANG	Molecular tools for genetic mapping and association studies
	Opis	*SLO*	Predstavitev dela na mednarodni konferenci o oljki. Rezultati so neposredno pridobljeni iz projketa sekvenciranja in s sodelovanjem s tujo inštitucijo INRA Montpellier.
		ANG	Work was presented at international Olive conference. Results were gaind from the projecet through collaboration with international group INRA Montpellier.
	Šifra		B.03 Referat na mednarodni znanstveni konferenci
	Objavljeno v		s. n.; VIIth International symposium on olive growing, San Juan, Argentina 25-29 September 2012; 2012; Str. P-30; Avtorji / Authors: Essalouh L., Zine El Aabidine A., Contreeas S., Ben Sadok I., Santoni S., Khadari B., Jakše Jernej, Bandelj Mavsar Dunja
	Tipologija		1.12 Objavljeni povzetek znanstvenega prispevka na konferenci

9.Drugi pomembni rezultati projetne skupine[8]

Znanstveno delo v recenziji: "Validation of candidate reference genes in RT-qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids etabolism". Molecular Breeding, IF=2.852, delo je v recenziji, Priloga-5

Work in review: "Validation of candidate reference genes in RT-qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids etabolism". Molecular Breeding, IF=2.852, in review, Priloga-5

Sekvenčni podatki dostopni v NCBI SRA arhivu, to so prvi javno dostopni NGS podatki slovenske inštitucije: http://www.ncbi.nlm.nih.gov/sra/SRX215662

Raw sequencing data are available at NCBI SRA archive, these are the first publically available NGS data of Slovenian research institution: http://www.ncbi.nlm.nih.gov/sra/SRX215662

Znanstveno delo v pripravi: "Developing olive fruit transcriptome", Tree Genetics & Genomes, IF=2.335

Reasearch paper in preparation: "Developing olive fruit transcriptome", Tree Genetics & Genomes, IF=2.335

10.Pomen raziskovalnih rezultatov projektne skupine[9]

10.1.Pomen za razvoj znanosti[10]

SLO

Oljka je ena najstarejših gojenih sadnih vrst na območju sredozemskega bazena, kjer jo gojijo v glavnem za pridobivanje oljčnega olja. Pomembna je tudi z zgodovinskega, kulturnega in gospodarskega vidika in ima ključno vlogo pri ohranjanju pokrajine na določenih območjih. Globalno gledano je pridelava oljk na omenjenem območju ena najpomembnejših kmetijskih panog in oljčno olje predstavlja pomemben dejavnik v zdravi mediteranski prehrani. Študija razvoja EST zaporedij oljčnega plodu s tehnologijami naslednjih generacij sekvenciranja (NGS), ki smo jo izvedli v okviru projekta je generirala prvo večjo količino nukleotidnih podatkov oljke, ki so dostopna preko NCBI SRA arhiva. Raziskava ima odmev pri raziskovalnih skupinah, ki delajo na raziskavah oljke, saj smo s francosko raziskovalno inštitucijo INRA Montpellier skupaj razvili mikrosatelitne markerje za genotipizacijo in kartiranje olja. Projektni rezultati so omogočili karakterizacijo večjega števila genov za normalizacijo RT-qPCR študij, obenem pa smo tudi dobili vpogled v ekspresijo nekaterih genov vključenih v sintezo maščobnih kislin pri razvijajočem plodu oljke. Rezultati so tako doprinesli k novim spoznanjem o biokemijskih karakteristikah primarnega (sinteza maščobnih kislin) metabolizma oljčnega plodu. Prav tako bodo rezultati imeli doprinos širše k boljšemu bazičnemu znanju ostalih bioloških procesov pri plodu oljke. 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. Anotacija transkriptov je pokazala, daje 924 le-teh vključenih v procese sekundarnega metabolizma. Dobljeni bazični rezultati pa bodo vplivali na razvoj stroke oziroma znanosti na tem področju.

ANG

Olive is one of the oldest cultivated fruit trees and oil producing crops of the Mediterranean basin, with very high historical, cultural and economic relevance. It also plays a fundamental role in the landscape maintenance of some regions. Olive production is globally the most important agricultural branch of the aforementioned region and olive oil is the principal source of fats in the rich Mediterranean diet. This study on EST development and characterization from developing olive fruit using next generation sequencing technologies (NGS) has enabled us to generate the first larger amount of nucleotide data for olive. They are all accessible through the NCBI SRA archive. The project results are also of interest to international collaborators who are engaged in olive research. In collaboration with a French research group from INRA Montpellier, genic microsatellite markers were developed for use in genotyping and linkage analysis studies. The project results have enabled us to characterize a large number of genes used for normalization of RT-qPCR studies and to gain an insight into the expression of selected genes that are involved in the synthesis of fatty acids in developing olive fruit. The results have also contributed to new knowledge about the biochemical characteristics of the primary metabolism (synthesis of fatty acids) of olive fruit. The knowledge gained will also contribute to fundamental knowledge of other biological processes in olive fruit.
Secondary metabolism of biophenols is very important in relation to superior olive oil storage

capacity and from a pharmaceutical point of view. Annotation of transcripts yielded 924 sequences that are putatively involved in the processes of secondary metabolism. The basic results achieved will have a broad impact on the development of applications and science in this field.

10.2.Pomen za razvoj Slovenije[11]

SLO

Rezultati raziskave imajo vpliv na pridobivanje novega znanja v Sloveniji in posledično na vključevanje tega znanja v nove tehnologije v domačem prostoru ter tudi širše. Osvojili smo bioinformatske obdelave NGS podatkov, projekt je poleg drugih NGS projektov v skupini vplival na razvoj bioinformatike – odločili smo se za vzpostavitev zmogljivega računalniškega serverja s pripadajočo zmogljivo programsko opremo za obdelavo podatkov.
Rastlinska olja predstavljajo 25% kalorij, ki jih dnevno porabijo prebivalci industrijskih držav. Poleg prehranske vrednosti so rastlinska olja tudi glavna kmetijska dobrina s svetovno produkcijo preko 40 milijonov kilogramov in letno vrednostjo tržišča okrog 35 milijard evrov. Velik tržni delež in dejstvo, da vsebnost maščobnih kislin vpliva na fizikalne in prehranske vrednosti olj, sta poglavitna razloga da obstaja veliko zanimaje za možnosti vplivanja na njihovo produkcijo v rastlini in sintezo farmakološko aktivnih molekul (biofenoli in vitamini). Genomski pristopi, ki vključujejo tudi analizo EST z uporabo bioinformatike, sedaj doprinašajo nova znanja o metabolizmu oljnic in regulaciji in izražanju genov, ki vplivajo na kvaliteto in količino olj.
Projekt je pomemben tudi za izobraževanje kadrov. V sklopu projekta se je izobraževala mlada raziskovalka in osvojila veščine bioinformatske obdelave NGS podatkov. Oljka je po obsegu kmetijskih površin na 2. mestu za jablano v Sloveniji in oljčno olje z geografskim poreklom je bilo kot prvi kmetijski proizvod s kakovostno oznako v Sloveniji potrjen s strani Evropske Unije. Posredni pomen projekta za družbo se kaže tudi preko izobraževanja mladih. Izobraževanje je ključnega pomena za spodbujanje inovativnosti in prenos raziskovalnih znanj v podjetništvo. Projektni rezultati so odlična učna platforma za dodiplomske in podiplomske študente za uvajanje v delo z NGS podatki.

ANG

The research results have contributed to the acquisition of new knowledge in Slovenia and, consequently, to incorporating such knowledge into new technologies both in Slovenia and in the wider world. We have successfully gained the necessary bioinformatic skills for manipulation and analysis of NGS data. This project, together with other projects of the Biotechnical Faculty group, has made an important contribution to the development of the bioinformatics field – we have decided to develop a powerful server class computer with dedicated software for NGS data analysis.
Vegetable oils provide approximately 25% of the calories consumed by industrial nations. In addition to their dietary significance, vegetable oils are a major agricultural commodity, with worldwide production of 40 billion kilograms, worth nearly 35 billion € per year. This large market and the fact that the fatty acid composition of vegetable oils influences both their physical properties and nutritional characteristics, has encouraged considerable interest in modifying plant fatty acid production and synthesis of pharmacologically active molecules, including biophenols and vitamins. Genomic approaches, including EST sequencing, are now contributing to greater understanding of the underlying metabolism of oil plants and the regulatory networks that determine the quality and quantity of oils produced.
The project has also been important for the education of human resources. A young researcher has been trained within the context of the project tasks and she has successfully gained knowledge and the necessary bioinformatics skills. In Slovenia, olive is second to apple in terms of the area used for its cultivation. Slovenian olive oil was the first Slovene agricultural product with a quality mark certified by the European Union. This research proposal also has indirect significance for society in terms of its potential for educating the young. Education is crucial for developing innovation and for transferring research findings into the economy. The project results are also an ideal learning platform for undergraduate and graduate students for learning the skills of NGS analysis.

11.Samo za aplikativne projekte in podoktorske projekte iz gospodarstva!
Označite, katerega od navedenih ciljev ste si zastavili pri projektu, katere konkretne rezultate ste dosegli in v kakšni meri so doseženi rezultati uporabljeni

Cilj		
F.01	**Pridobitev novih praktičnih znanj, informacij in veščin**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.02	**Pridobitev novih znanstvenih spoznanj**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.03	**Večja usposobljenost raziskovalno-razvojnega osebja**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.04	**Dvig tehnološke ravni**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.05	**Sposobnost za začetek novega tehnološkega razvoja**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.06	**Razvoj novega izdelka**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.07	**Izboljšanje obstoječega izdelka**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.08	**Razvoj in izdelava prototipa**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.09	**Razvoj novega tehnološkega procesa oz. tehnologije**	
	Zastavljen cilj	◯ DA ◯ NE

	Rezultat	
	Uporaba rezultatov	
F.10	**Izboljšanje obstoječega tehnološkega procesa oz. tehnologije**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.11	**Razvoj nove storitve**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.12	**Izboljšanje obstoječe storitve**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.13	**Razvoj novih proizvodnih metod in instrumentov oz. proizvodnih procesov**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.14	**Izboljšanje obstoječih proizvodnih metod in instrumentov oz. proizvodnih procesov**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.15	**Razvoj novega informacijskega sistema/podatkovnih baz**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.16	**Izboljšanje obstoječega informacijskega sistema/podatkovnih baz**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.17	**Prenos obstoječih tehnologij, znanj, metod in postopkov v prakso**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	

F.18	**Posredovanje novih znanj neposrednim uporabnikom (seminarji, forumi, konference)**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.19	**Znanje, ki vodi k ustanovitvi novega podjetja ("spin off")**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.20	**Ustanovitev novega podjetja ("spin off")**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.21	**Razvoj novih zdravstvenih/diagnostičnih metod/postopkov**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.22	**Izboljšanje obstoječih zdravstvenih/diagnostičnih metod/postopkov**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.23	**Razvoj novih sistemskih, normativnih, programskih in metodoloških rešitev**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.24	**Izboljšanje obstoječih sistemskih, normativnih, programskih in metodoloških rešitev**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.25	**Razvoj novih organizacijskih in upravljavskih rešitev**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.26	**Izboljšanje obstoječih organizacijskih in upravljavskih rešitev**	
	Zastavljen cilj	◯ DA ◯ NE

	Rezultat	
	Uporaba rezultatov	
F.27	**Prispevek k ohranjanju/varovanje naravne in kulturne dediščine**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.28	**Priprava/organizacija razstave**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.29	**Prispevek k razvoju nacionalne kulturne identitete**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.30	**Strokovna ocena stanja**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.31	**Razvoj standardov**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.32	**Mednarodni patent**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.33	**Patent v Sloveniji**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.34	**Svetovalna dejavnost**	
	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	
F.35	**Drugo**	

	Zastavljen cilj	◯ DA ◯ NE
	Rezultat	
	Uporaba rezultatov	

Komentar

12. Samo za aplikativne projekte in podoktorske projekte iz gospodarstva!
Označite potencialne vplive oziroma učinke vaših rezultatov na navedena področja

	Vpliv	**Ni vpliva**	**Majhen vpliv**	**Srednji vpliv**	**Velik vpliv**	
G.01	**Razvoj visokošolskega izobraževanja**					
G.01.01.	Razvoj dodiplomskega izobraževanja	◯	◯	◯	◯	
G.01.02.	Razvoj podiplomskega izobraževanja	◯	◯	◯	◯	
G.01.03.	Drugo:	◯	◯	◯	◯	
G.02	**Gospodarski razvoj**					
G.02.01	Razširitev ponudbe novih izdelkov/storitev na trgu	◯	◯	◯	◯	
G.02.02.	Širitev obstoječih trgov	◯	◯	◯	◯	
G.02.03.	Znižanje stroškov proizvodnje	◯	◯	◯	◯	
G.02.04.	Zmanjšanje porabe materialov in energije	◯	◯	◯	◯	
G.02.05.	Razširitev področja dejavnosti	◯	◯	◯	◯	
G.02.06.	Večja konkurenčna sposobnost	◯	◯	◯	◯	
G.02.07.	Večji delež izvoza	◯	◯	◯	◯	
G.02.08.	Povečanje dobička	◯	◯	◯	◯	
G.02.09.	Nova delovna mesta	◯	◯	◯	◯	
G.02.10.	Dvig izobrazbene strukture zaposlenih	◯	◯	◯	◯	
G.02.11.	Nov investicijski zagon	◯	◯	◯	◯	
G.02.12.	Drugo:	◯	◯	◯	◯	
G.03	**Tehnološki razvoj**					
G.03.01.	Tehnološka razširitev/posodobitev dejavnosti	◯	◯	◯	◯	
G.03.02.	Tehnološko prestrukturiranje dejavnosti	◯	◯	◯	◯	
G.03.03.	Uvajanje novih tehnologij	◯	◯	◯	◯	
G.03.04.	Drugo:	◯	◯	◯	◯	
G.04	**Družbeni razvoj**					
G.04.01	Dvig kvalitete življenja	◯	◯	◯	◯	
G.04.02.	Izboljšanje vodenja in upravljanja	◯	◯	◯	◯	
G.04.03.	Izboljšanje delovanja administracije in javne uprave	◯	◯	◯	◯	

G.04.04.	Razvoj socialnih dejavnosti	○	○	○	○	
G.04.05.	Razvoj civilne družbe	○	○	○	○	
G.04.06.	Drugo:	○	○	○	○	
G.05.	**Ohranjanje in razvoj nacionalne naravne in kulturne dediščine in identitete**	○	○	○	○	
G.06.	**Varovanje okolja in trajnostni razvoj**	○	○	○	○	
G.07	**Razvoj družbene infrastrukture**					
G.07.01.	Informacijsko-komunikacijska infrastruktura	○	○	○	○	
G.07.02.	Prometna infrastruktura	○	○	○	○	
G.07.03.	Energetska infrastruktura	○	○	○	○	
G.07.04.	Drugo:	○	○	○	○	
G.08.	**Varovanje zdravja in razvoj zdravstvenega varstva**	○	○	○	○	
G.09.	**Drugo:**	○	○	○	○	

Komentar

13.Pomen raziskovanja za sofinancerje[12]

	Sofinancer			
1.	Naziv			
	Naslov			
	Vrednost sofinanciranja za celotno obdobje trajanja projekta je znašala:			EUR
	Odstotek od utemeljenih stroškov projekta:			%
	Najpomembnejši rezultati raziskovanja za sofinancerja			Šifra
		1.		
		2.		
		3.		
		4.		
		5.		
	Komentar			
	Ocena			

14.Izjemni dosežek v letu 2012[13]

14.1. Izjemni znanstveni dosežek

14.2. Izjemni družbeno-ekonomski dosežek

C. IZJAVE

Podpisani izjavljam/o, da:

- so vsi podatki, ki jih navajamo v poročilu, resnični in točni
- se strinjamo z obdelavo podatkov v skladu z zakonodajo o varstvu osebnih podatkov za potrebe ocenjevanja ter obdelavo teh podatkov za evidence ARRS
- so vsi podatki v obrazcu v elektronski obliki identični podatkom v obrazcu v pisni obliki
- so z vsebino zaključnega poročila seznanjeni in se strinjajo vsi soizvajalci projekta

Podpisi:

zastopnik oz. pooblaščena oseba raziskovalne organizacije:	in	*vodja raziskovalnega projekta:*
Univerza v Ljubljani, Biotehniška fakulteta		Jernej Jakše
_____		_____

ŽIG

Kraj in datum:	Ljubljana	7.3.2013

Oznaka prijave: ARRS-RPROJ-ZP-2013/117

---

[1] Opredelite raziskovalno področje po klasifikaciji FOS 2007 (Fields of Science). Prevajalna tabela med raziskovalnimi področji po klasifikaciji ARRS ter po klasifikaciji FOS 2007 (Fields of Science) s kategorijami WOS (Web of Science) kot podpodročji je dostopna na spletni strani agencije (http://www.arrs.gov.si/sl/gradivo/sifranti/preslik-vpp-fos-wos.asp). Nazaj

[2] Napišite povzetek raziskovalnega projekta (največ 3.000 znakov v slovenskem in angleškem jeziku) Nazaj

[3] Napišite kratko vsebinsko poročilo, kjer boste predstavili raziskovalno hipotezo in opis raziskovanja. Navedite ključne ugotovitve, znanstvena spoznanja, rezultate in učinke raziskovalnega projekta in njihovo uporabo ter sodelovanje s tujimi partnerji. Največ 12.000 znakov vključno s presledki (približno dve strani, velikost pisave 11). Nazaj

[4] Realizacija raziskovalne hipoteze. Največ 3.000 znakov vključno s presledki (približno pol strani, velikost pisave 11) Nazaj

[5] V primeru bistvenih odstopanj in sprememb od predvidenega programa raziskovalnega projekta, kot je bil zapisan v predlogu raziskovalnega projekta oziroma v primeru sprememb, povečanja ali zmanjšanja sestave projektne skupine v zadnjem letu izvajanja projekta, napišite obrazložitev. V primeru, da sprememb ni bilo, to navedite. Največ 6.000 znakov vključno s presledki (približno ena stran, velikost pisave 11). Nazaj

[6] Navedite znanstvene dosežke, ki so nastali v okviru tega projekta. Raziskovalni dosežek iz obdobja izvajanja projekta (do oddaje zaključnega poročila) vpišete tako, da izpolnite COBISS kodo dosežka – sistem nato sam izpolni naslov objave, naziv, IF in srednjo vrednost revije, naziv FOS področja ter podatek, ali je dosežek uvrščen v A'' ali A'. Nazaj

[7] Navedite družbeno-ekonomske dosežke, ki so nastali v okviru tega projekta. Družbeno-ekonomski rezultat iz obdobja izvajanja projekta (do oddaje zaključnega poročila) vpišete tako, da izpolnite COBISS kodo dosežka – sistem nato sam izpolni naslov objave, naziv, IF in srednjo vrednost revije, naziv FOS področja ter podatek, ali je dosežek uvrščen v A'' ali A'.

Družbeno-ekonomski dosežek je po svoji strukturi drugačen kot znanstveni dosežek. Povzetek znanstvenega dosežka je praviloma povzetek bibliografske enote (članka, knjige), v kateri je dosežek objavljen.

Povzetek družbeno-ekonomskega dosežka praviloma ni povzetek bibliografske enote, ki ta dosežek dokumentira, ker je dosežek sklop več rezultatov raziskovanja, ki je lahko dokumentiran v različnih bibliografskih enotah. COBISS ID zato ni enoznačen, izjemoma pa ga lahko tudi ni (npr. prehod mlajših sodelavcev v gospodarstvo na pomembnih raziskovalnih nalogah, ali ustanovitev podjetja kot rezultat projekta … - v obeh primerih ni COBISS ID). Nazaj

[8] Navedite rezultate raziskovalnega projekta iz obdobja izvajanja projekta (do oddaje zaključnega poročila) v primeru, da katerega od rezultatov ni mogoče navesti v točkah 7 in 8 (npr. ker se ga v sistemu COBISS ne vodi). Največ 2.000 znakov, vključno s presledki. Nazaj

[9] Pomen raziskovalnih rezultatov za razvoj znanosti in za razvoj Slovenije bo objavljen na spletni strani: http://sicris.izum.si/ za posamezen projekt, ki je predmet poročanja Nazaj

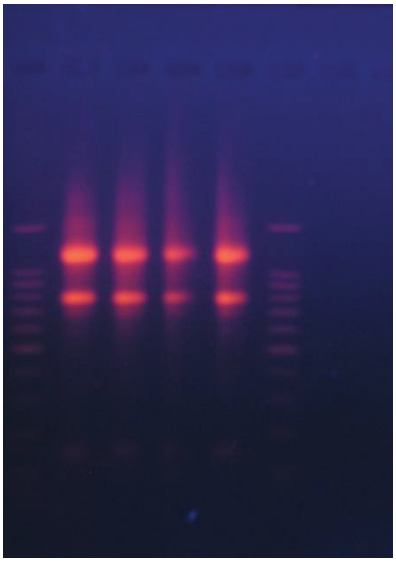
[10] Največ 4.000 znakov, vključno s presledki Nazaj

[11] Največ 4.000 znakov, vključno s presledki Nazaj

[12] Rubrike izpolnite / prepišite skladno z obrazcem "izjava sofinancerja" http://www.arrs.gov.si/sl/progproj/rproj/gradivo/, ki ga mora izpolniti sofinancer. Podpisan obrazec "Izjava sofinancerja" pridobi in hrani nosilna raziskovalna organizacija – izvajalka projekta. Nazaj

[13] Navedite en izjemni znanstveni dosežek in/ali en izjemni družbeno-ekonomski dosežek raziskovalnega projekta v letu 2012 (največ 1000 znakov, vključno s presledki). Za dosežek pripravite diapozitiv, ki vsebuje sliko ali drugo slikovno gradivo v zvezi z izjemnim dosežkom (velikost pisave najmanj 16, približno pol strani) in opis izjemnega dosežka (velikost pisave 12, približno pol strani). Diapozitiv/-a priložite kot priponko/-i k temu poročilu. Vzorec diapozitiva je objavljen na spletni strani ARRS http://www.arrs.gov.si/sl/gradivo/, predstavitve dosežkov za pretekla leta pa so objavljena na spletni strani http://www.arrs.gov.si/sl/analize/dosez/. Nazaj

Obrazec: ARRS-RPROJ-ZP/2013 v1.00
00-0D-B0-96-80-B6-2D-54-9C-D4-9A-12-E4-B8-5B-4D-4C-02-D0-F4



Priloga 1: Primeri vzorcev izolirane RNA oljčnega plodu na agaroznem gelu (izolacija komercialni komplet), kjer sta lepo vidni nerazgrajeni ribosomalni črti

M&M: Sequencing, data analysis

Normalized cDNA

½ plate GS FLX Titanium 454 sequencin

OLIVE.sff

Raw data extraction and clipping: `sff_extract.py`
Spliting concatamers: `SSAHA2`
Contaminant removal, linker trimming, polyA trimming: `SeqClean`
Results checking: `FastQC, BioPerl scripts, R`

OLIVE_CLEANED.fasta & qual

Assembly step: `CLC, Mira, iAssembler. Newbler 2.3 & 2.6, PAVE` and `tgicl: 96% identity, 40 bp min overlap`

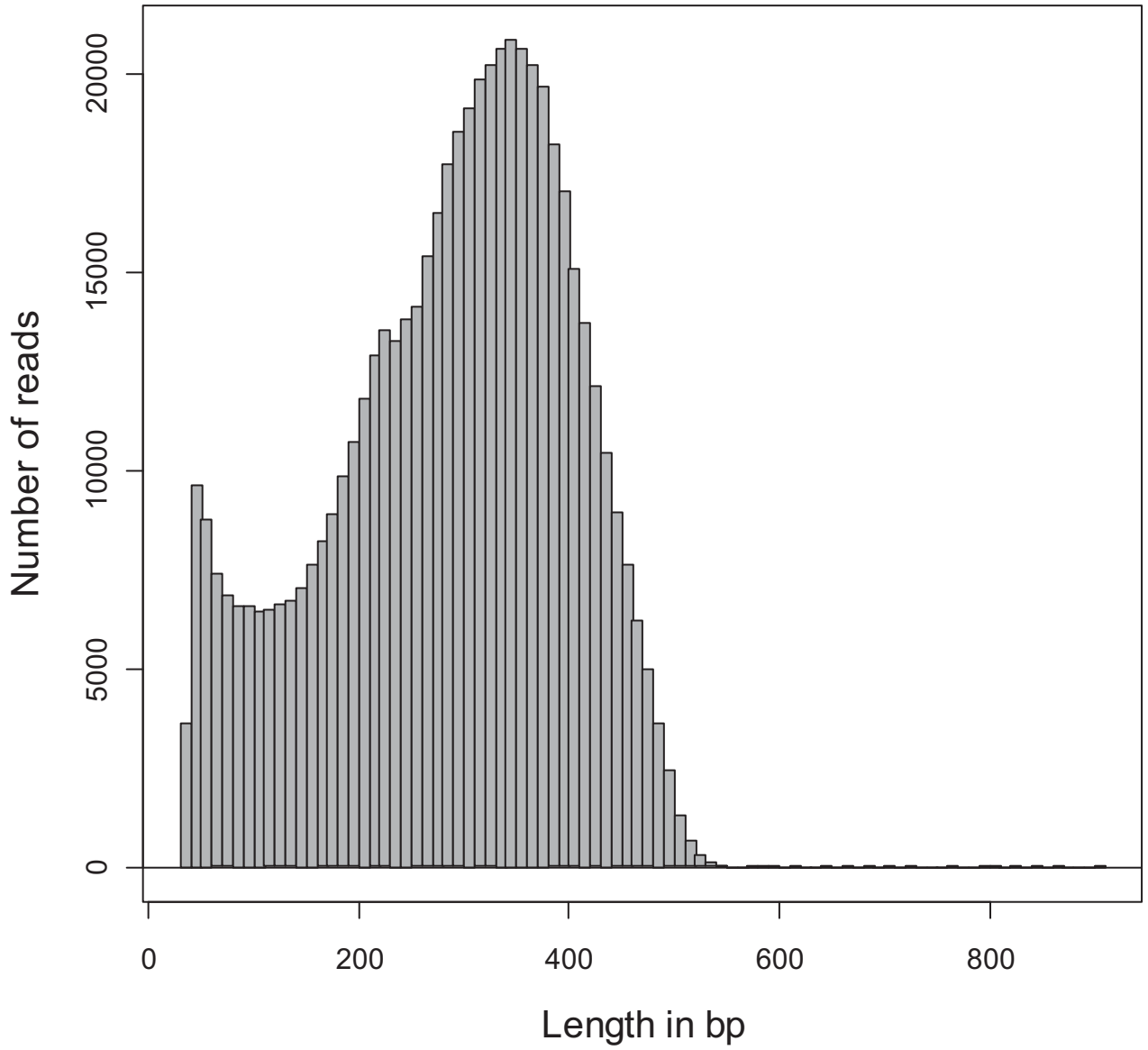
ASSEMBLED.contigs & REMAINING.singletons

Assembler performance: `BLAST, BLAT, R`

WHAT.IS.BEST.ASSEMBLY?

Transcriptome dataset read statistics

RAW DATA READ LENGTH HISTOGRAM



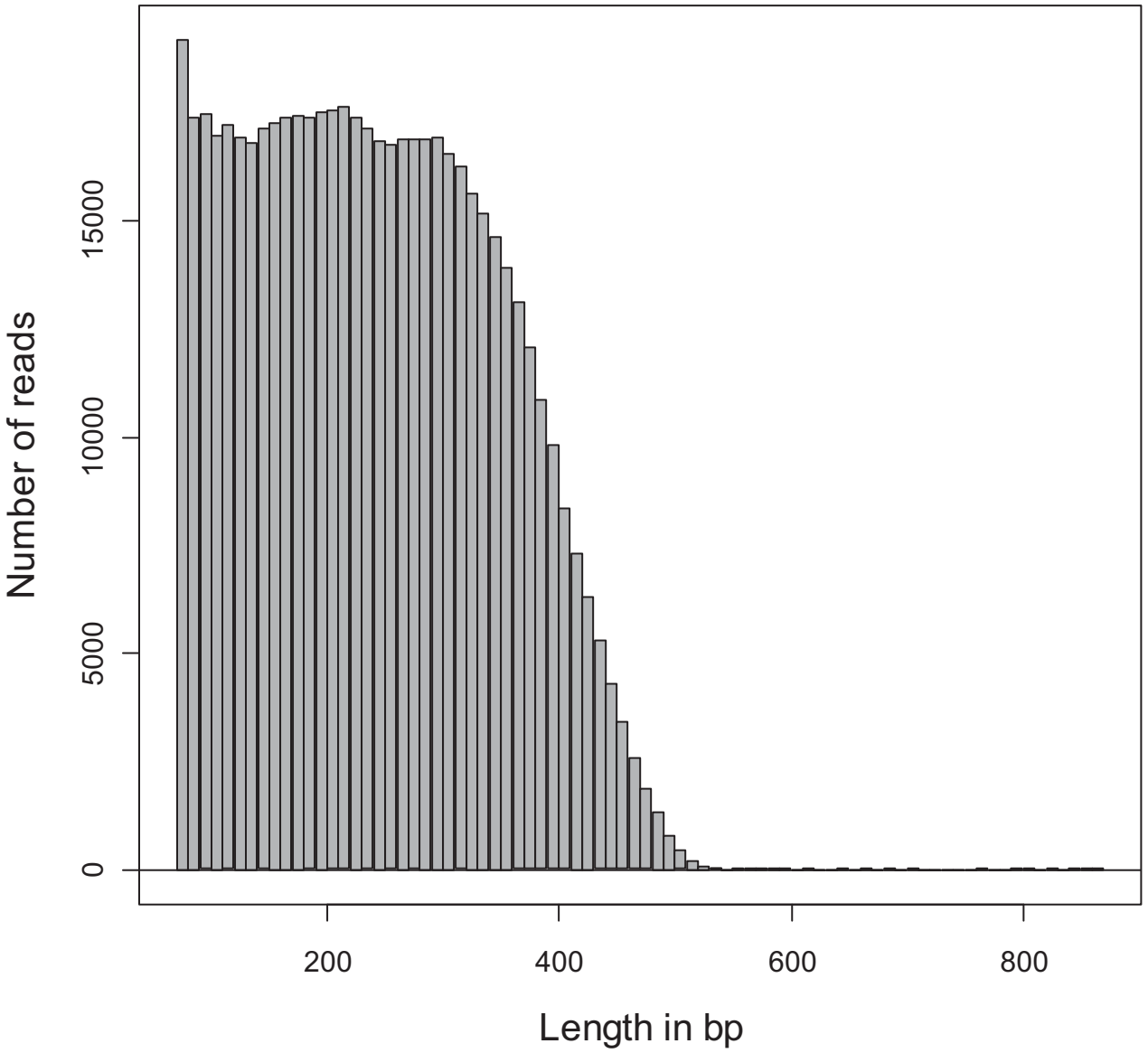
560,578 reads
160,414,301 bp

median	303 bp
average	286 bp
min	34 bp
max	904 bp
N50	343 bp

(200,290 sequences)

Transcriptome dataset read statistics

SPLIT & CLEANED DATA READ LENGTH HISTOGRAM



703,936 reads

cleaning

577,025 reads
139,419,877 bp

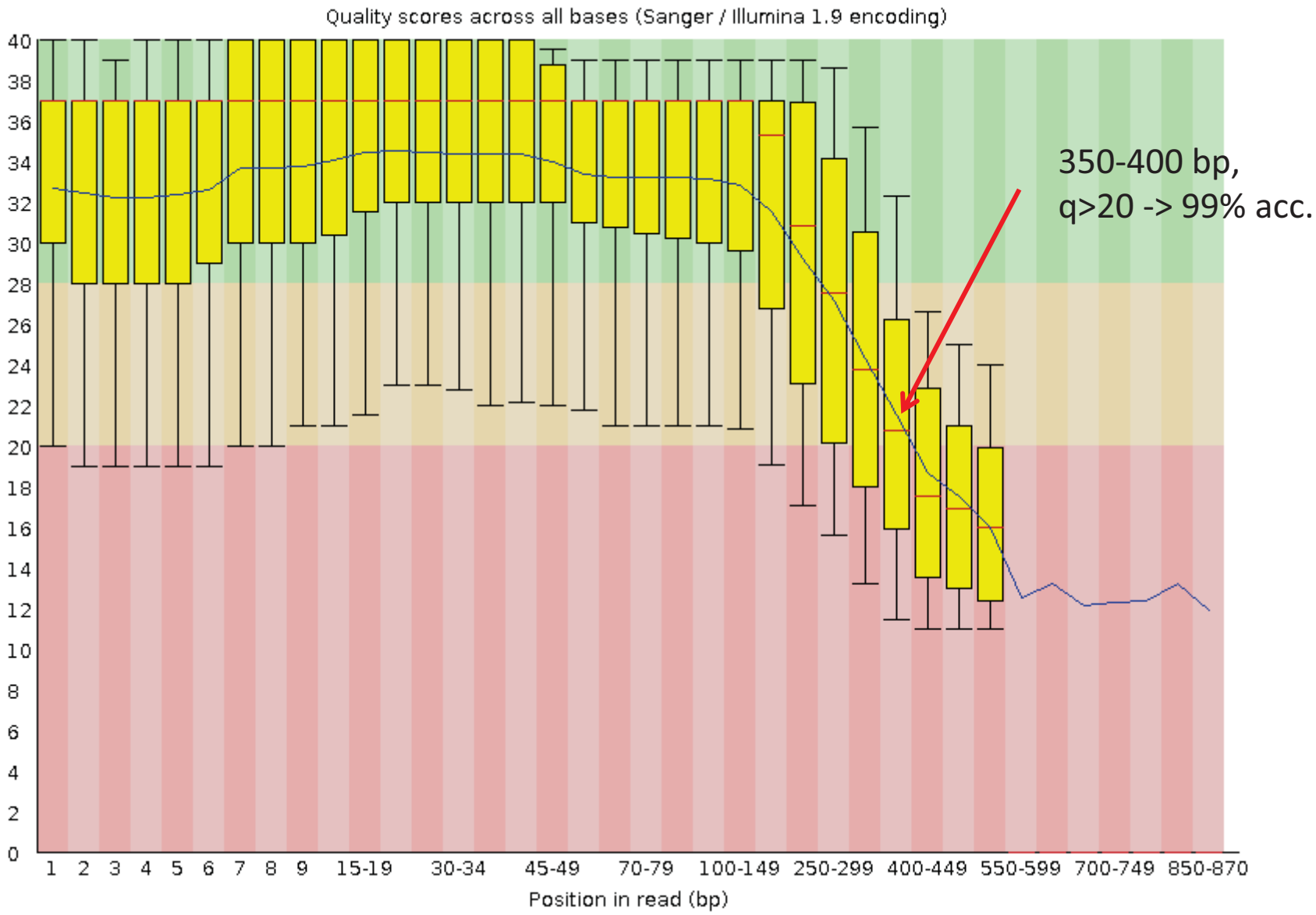
Median	236 bp
average	242 bp
min	70 bp
max	870 bp
N50	294 bp

(192,189 sequences)

Cleaning report:
-102014 adapters
-420 titA and B
-7061 polyA

Transcriptome dataset quality statistics

SPLIT & CLEANED DATA: Per base quality score



Features of assembly programmes

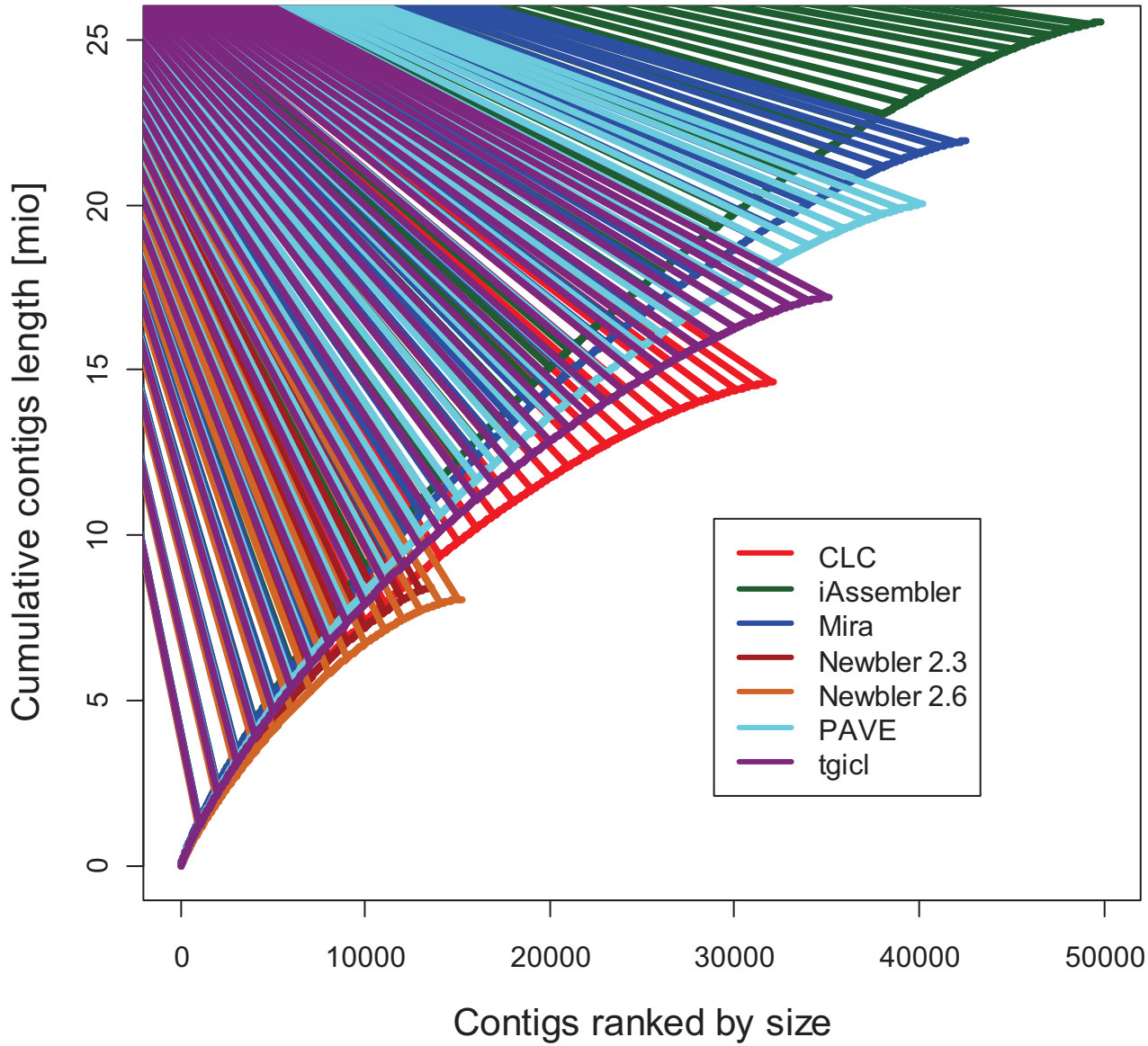
Assembler	Type	Description	Cost	Support technologies
tgicl 2.1	OLC ESTs	wrapper for CAP3	free	Sanger
PAVE 2.5	OLC ESTs	wrapper for CAP3, iterative assemblies, mysql integration	free	Sanger, 454
Mira 1.3	OLC ESTs, genomes	Can perform iterative assemblies	free	Sanger, 454, Illumina
iAssembler 1.2.2	OLC ESTs	Performing Mira and CAP3 iterative assemblies	free	Sanger, 454
Newbler 2.3 & 2.6	OLC ESTs, genomes	Software from the developer of sequencing technology	free for academics	Sanger, 454
CLC Genomics Workbench 4.5	de Brujin graph ESTs, genomes	SIMD-accelerated assembly algorithm	quote or trial	Sanger & all available NGS data

OLC=Overlap-Layout-Consensus
SIMD=Single Instruction Multiple Data
CLC – executables for other OS except Linux available as well

1) Basic assembly "length" metrics

	New 2.6	New 2.3	MIRA	iAssembler	CLC	TGICL	PAVE
Number of contigs	15,224	13,530	42,504	49,860	32,138	35,074	40,219
Total bases	8,086,878	8,439,420	21,930,174	25,529,782	14,646,256	17,215,800	20,024,716
Singletons	73,087	77,773	49,711	49,064	52,611	66,103	47,766
Singletons length	17,523,948	18,103,598	10,821,840	11,258,808	12,818,683	15,585,570	10,414,216
Assembly coverage	15.07 ×	14.37 ×	5.86 ×	5.02 ×	8.64 ×	7.19 ×	6.44 ×
No of contigs (>=1 kbp)	694	1,121	2,141	2,363	1,005	1,343	1,549
No of contigs (>= 500 bp)	8,038	9,004	18,860	21,879	11,305	14,115	16,560
Max contig length	3,456	4,336	3,738	4,473	3,142	3,032	4,619
Mean contig length	531.2	623.8	516	512	455.7	490.8	497.9
Median value	518	587	468	466	414	446	452
N50	640	687	586	585	532	559	563
No of contigs in N50	4,869	4,657	13,484	15,813	9,831	11,137	12,876
Time taken	30 min	30 min	15 hours	15 hours	5 min	41 h	12 days

Cumulative contig lengths generated by different assembly programs



2) Mapping assemblies to each other

BLAT: comparison all vs. all -> to determine the number of unique sequences in “query” assembly not present in “database” assembly

“query”

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**

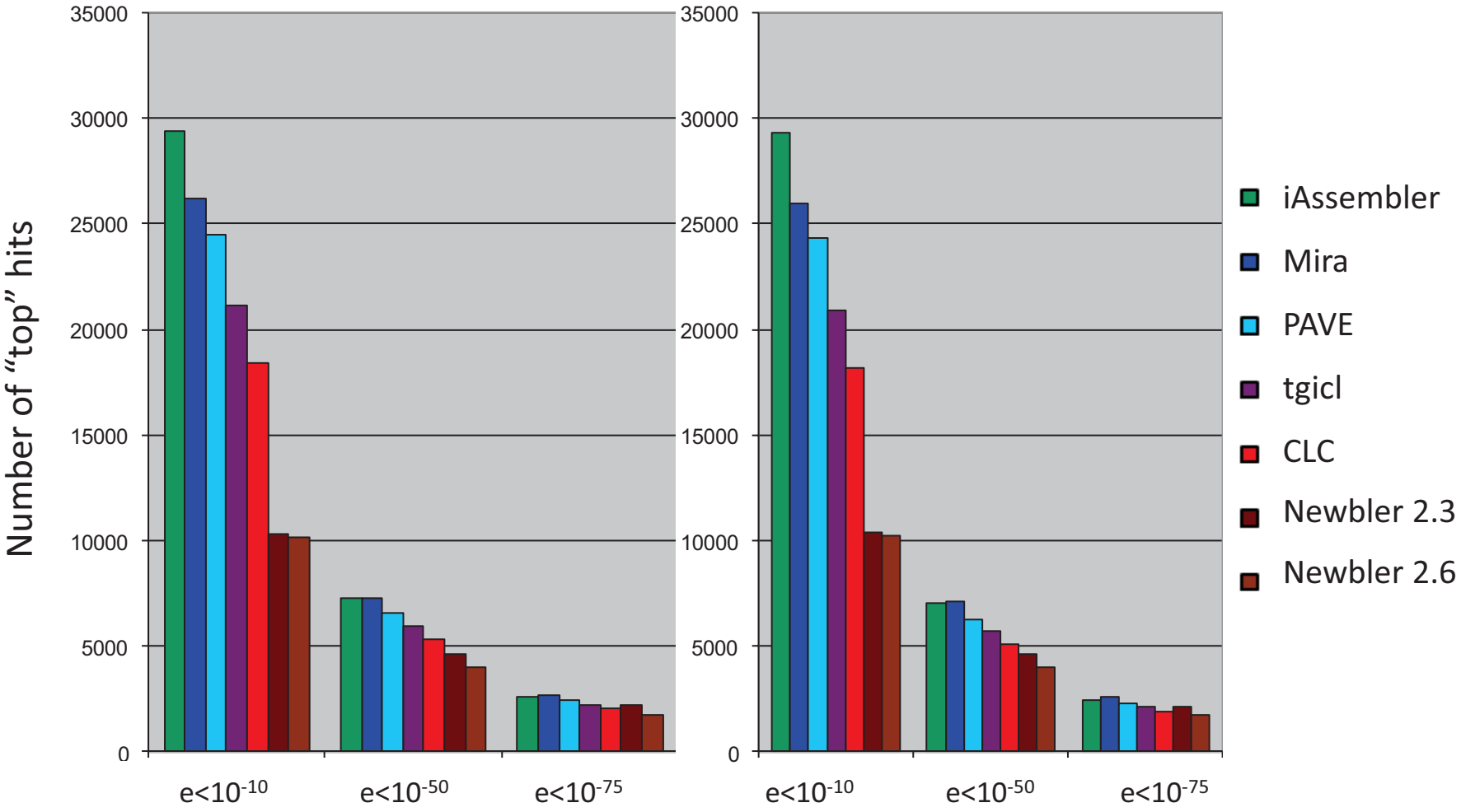
“database”

Kent WJ.Genome Res. 2002, 12(4):656-64.

3) BLASTX comparison

NR: 14,987,464 sequences; 5,132,678,026 total letters
UNIPROT PLANTS: 410,553 sequences; 143,146,364 total letters

E-value distribution



Altschul SF, et al. 1990. J Mol Biol 215 (3): 403–410.

BLASTX comparison

"Percentage of total"= length of alignment/hit length=112/212=52.8%

```
Query= oljka_rep_c2320
         (456 letters)

>gi|82698815|gb|ABB89210.1| dehydroascorbate reductase [Sesamum indicum]
          Length = 212

Score = 50.8 bits (120), Expect(2) = 1e-47
 Identities = 21/27 (77%), Positives = 23/27 (85%)
 Frame = +3

Query: 3   KTHLINFSDKPQWFLEVNPEGKVPMLK 83
           K HLIN   KPQWFLEVNPEGKVP++K
Sbjct: 38  KLHLINVDQKPQWFLEVNPEGKVPVIK 64

          Score =  163 bits (413), Expect(2) = 1e-47
 Identities = 77/86 (89%), Positives = 83/86 (96%)
 Frame = +2

Query: 80  KIDEKWITDSDVIVGIIEEKYPNPSLSPPPEISSVGSKIFPSFVKFLKSKDPSDGSEQAL 259
           K D+KWI DSDVIVG++EEKYPNPSLSPPPE+SSVGSKIFPSFVKFLKSKDP+DGSEQAL
Sbjct: 64  KFDDKWIADSDVIVGLLEEKYPNPSLSPPPEVSSVGSKIFPSFVKFLKSKDPTDGSEQAL 123

Query: 260 LNELKALDEHLKAKGPYVAGENICAV 337
           L+ELKALDEHLKAKGPYV GENICAV
Sbjct: 124 LDELKALDEHLKAKGPYVNGENICAV 149
```

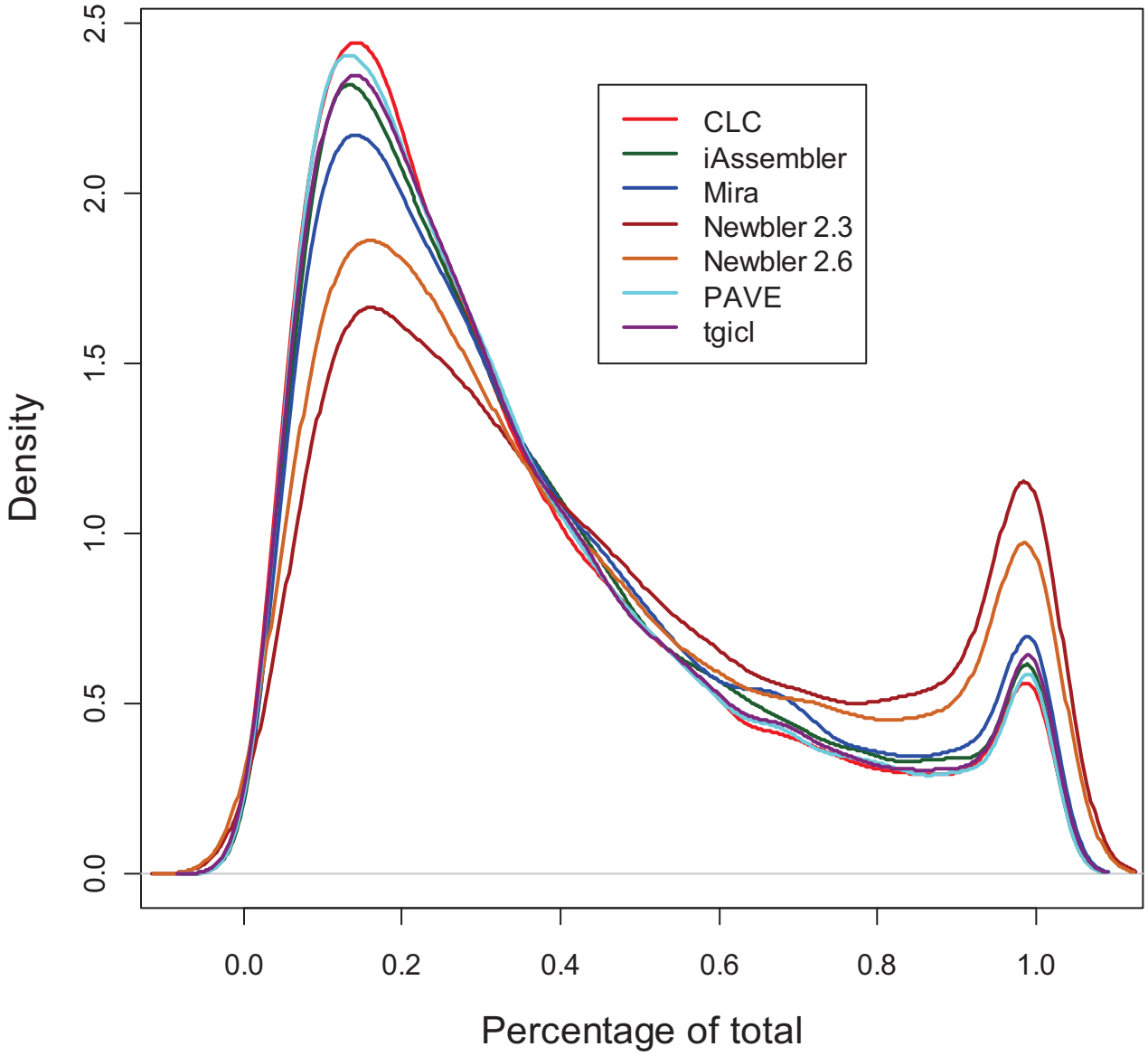
BLASTX comparison

BLASTX results: hits/no hits, unique hits, maximum occurrence of same hit, ratio of hits with alignment between 70-100% of total protein length

		CLC	iAsse	MIRA	New 2.3	New 2.6	PAVE	TGICL
NR	Contigs with hits in database	67.8%	71.2%	72.6%	83.6%	75.8%	72.6%	71.1%
Un		67.8%	71.8%	72.9%	84.2%	77.0%	72.9%	71.2%
NR	Unique hits in database	65.1%	43.6%	46.3%	57.1%	66.9%	50.6%	56.8%
Un		54.2%	33.6%	36.4%	51.8%	59.3%	40.2%	45.8%
NR	One+two hits in database	87.5%	64.2%	66.7%	83.4%	88.2%	73.2%	79.8%
Un		79.7%	54.2%	57.7%	79.7%	82.6%	63.7%	71.4%
NR	Max occurrence of same hit	10	79	300	10	14	35	14
Un		12	78	311	14	12	37	17
NR	Contigs aligned by 70-100% of db protein	6.7%	7.7%	8.4%	14.6%	11.8%	4.9%	8.2%
Un		6.6%	7.6%	8.1%	14.6%	11.7%	6.2%	8.1%

BLASTX comparison

Kernel density estimation of “percentage of total” value



Kolmogorov-Smirnov test

WHAT.IS.BEST.ASSEMBLY?

	CLC	iAsse	MIRA	New 2.3	New 2.6	PAVE	TGICL
Length metrics	3	7	5.5	1.5	1.5	5.5	4
Amount of assembled reads	4	6	6	1.5	1.5	6	3
BLAT mapping	3	7	5	1.5	1.5	6	4
BLASTX comparison	6.5	1.5	1.5	4.5	6.5	3	4.5
TOTAL SCORE	16.5	21.5	18	9	11	20.5	15.5
RANK	4	1	3	7	6	2	5

Molecular Breeding

Validation of candidate reference genes in RT-qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids metabolism

--Manuscript Draft--

Manuscript Number:	MOLB-D-12-02605R1
Full Title:	Validation of candidate reference genes in RT-qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids metabolism
Article Type:	Manuscript
Keywords:	RT-qPCR; olive; Olea europaea; reference gene; normalization; fruit development
Corresponding Author:	Jernej Jakse University of Ljubljana, Biotechnical faculty Ljubljana, -- Please Select -- SLOVENIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Ljubljana, Biotechnical faculty
Corresponding Author's Secondary Institution:	
First Author:	Tjasa Resetic
First Author Secondary Information:	
Order of Authors:	Tjasa Resetic
	Natasa Stajner
	Dunja Bandelj
	Branka Javornik
	Jernej Jakse
Order of Authors Secondary Information:	
Abstract:	Olive is an evergreen Mediterranean oil fruit tree with high economic, cultural and historical importance. For accurate gene expression studies of specific genes, reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) is often the method of choice, using suitable reference genes (RGs). This study identified 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), while several classical RGs proved not to be suitable. 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. In addition to identifying two RGs for future analysis of gene expression in olive fruit, our results also provide a list of potential RGs that can be easily tested in other studies of olive gene expression in different developmental stages or in biologically challenged olive tissues. The results are also valuable for future research of genes that influence the synthesis and accumulation of olive fruit metabolites.
Response to Reviewers:	This a copy paste from formated text. Please also consider to check the uploaded file responses_to_reviewers_19jan2013-FINAL and paper_corrections. Kind regards, Jernej Jakse 19/01/2013 Dear Editor, Please find the responses to reviewers and description of changes we have made to

	the manuscript »Validation of candidate reference genes in qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids metabolism« 1) Reviewer #1 Reviewer #1 did not suggest any changes to the manuscript. 2) Reviwer #2 We want to thank the anonymous reviewer #2 for the constructive review and improvement of the manuscript. We will answer all the questions from the manuscript and then questions from the cover letter. 2A) Reviewer #2 - questions in the manuscript 1) page 5, section »RNA isolation and cDNA preparation«, paragraph 1; Question »What does 'smeared tracks' mean? RNA bands with smear? Why not showing a picture?«, related to sentence: »RNA quality was regularly inspected by the integrity of ribosomal RNA bands on 1.2% formaldehyde agarose gels, which showed sharp and intense 18S and 28S ribosomal RNA bands, with a small number of smeared tracks, confirming the suitability of the isolation method.« Answer: Smeared RNA tracks in electrophoresis are a common feature from RNA degradation, when ribosomal bands show partial or complete degradation and you obtain a subset of RNA molecules of different lengths. With the isolation procedure used in our experiment, we did not observe problems related to RNA degradation or it appeared only in a few examples. If we observed any degradation, the isolation of the RNA was repeated. We decided not to show a picture of the RNA gel electrophoresis in the revised manuscript, since the procedure is quite common molecular biology practice and due to the limitation on Tables/Figures. However, we changed the sentence to: RNA quality was regularly inspected by the integrity of ribosomal RNA bands on 1.2% formaldehyde agarose gels, which showed sharp and intense 18S and 28S ribosomal RNA bands, without visible degradation, confirming the suitability of the isolation method. 2) page 6, section »Reference gene selection and primer design«, paragraph 2; Question: »(Resetic, unpublished) describe this sequence resource better.«, related to sentence: »We compared 7,080 olive mRNA or DNA sequences downloaded from GenBank and 212,795 olive 454 EST clusters of sequences against the 29 reference genes sequences (DNA and protein) from other plant species (Table 1) using BLASTN, BLASTX and TBLASTX algorithms (McGinnis and Madden 2004). Answer: The 454 sequences originate from our olive transcriptome assembly project, which is not yet published. The full set of raw sequences consists of 560,578 reads totaling 160.4 Mb. For the RT-qPCR part, the sequences were clustered using CD-HIT software into a non-redundant set of 212,795 olive sequences, which, together with NCBI sequences, were compared to the selection of reference genes. For a better description of the source we have done the following: a)We have submitted and released the entire raw transcriptome data in NCBI's SRA archive under accession number SRX215662 (http://www.ncbi.nlm.nih.gov/sra/SRX215662) b)We have changed the sentence in question and inserted an additional sentence in order to describe the source better. We have also introduced an additional reference, which is now listed in the reference section: We used 7,080 olive mRNA or DNA sequences downloaded from GenBank and 212,795 olive 454 EST clusters of sequences (Resetic, unpublished). The EST clusters originate from 454 olive transcriptome data available as raw reads in NCBI SRA

	archive (SRX215662), which were clustered using CD-HIT software (Li and Godzik, 2006) to decrease redundancy. Sequences were compared against the 29 RG… Reference was added to reference list: Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658-1659 3) page 7, section “Two-step real time PCR analysis and quantification”, paragraph 1 and 2; Question: “The final assay included a standard curve of six serial dilution points for the reference genes, genes of interest six dilutions also for the genes of interest? and twelve cDNA samples from different stages of olive fruit development with 4 dilutions? Not clear, see below.”…additional text… “The cDNA samples were used as PCR templates in a range of 4-fold series dilutions starting with 50 and ending with 0.05 ng. Please specify the 4 concentrations.” Answer: We believe that the text properly describes the standard curve experiment – we used 6 dilutions of the templates and they were prepared by a dilution series from 50 ng to 0.05 ng, which were diluted 4-fold (50, 12.5, 3.13, 0.78, 0.20 and 0.05 ng). For better understanding, we have changed the text accordingly. Before: The final assay included a standard curve of six serial dilution points for the reference genes, genes of interest and twelve cDNA samples from different stages of olive fruit development. Variation between runs was minimized by performing all of the reactions containing a single primer pair on the same plate and by including a standard curve on each plate. The cDNA samples were used as PCR templates in a range of 4-fold series dilutions starting with 50 and ending with 0.05 ng. After: The final assay included a standard curve of six serial dilution points for the RGs, genes of interest and twelve cDNA samples from different stages of olive fruit development. The cDNA samples were used as PCR templates in a range of six dilutions made in 4-fold decrements starting with 50 and ending with 0.05 ng (50, 12.5, 3.13, 0.78, 0.20 and 0.05 ng). Variation between runs was minimized by performing all of the reactions with a single primer pair on the same plate and by including a standard curve on each plate. 4) page 8, section “Two-step real time PCR analysis and quantification”, paragraph 2; Question: “How is NRQ calculated?”, related to sentence “After we had identified the two most stable reference genes, we used them for calculating the normalized relative quantities (NRQ) of four lipid metabolism target genes.” Answer: NRQ is calculated by division of the average quantity with the normalization factor. The average quantity is obtained from quantity repetitions and delivered by a Real Time program, SDS in our case. The normalization factor for the two best reference genes is calculated by the geNorm program (this is geomean of quantities of individual reference genes. Since it is a common procedure in RT-qPCR experiments, it has not been included in the manuscript. 5) page 11, section “Discussion”, paragraph 1; Questions: “We also tested two other genes selected from the GMO reference gene list, Pkaba1 and UGPase, never reported as having been used in qPCR normalization studies. UGPase proved to be a less favorable gene in our experiment, ranked as the 17th most stable gene according to which algorithm?, while Pkaba1 was among the most stable genes in our study, in 3rd place according to which algorithm?. Answer: According to the geNorm algorithm in both cases, which we added to the end of the sentence. The reviewer also suggested changes in the sentence, which we have

	incorporated: “We also tested two other genes selected from the GMO RG list, Pkaba1 and UGPase, never reported as having been used in RT-qPCR normalization studies. UGPase ranked as the 17th most stable gene, while Pkaba1 was among the most stable genes in our study, in 3rd place, in both cases according to the geNorm algorithm. 6) page 12, section “Discussion”, paragraph 2; Question: “It is interesting that one of the GAPDH primer pairs (GAPDH1) in their study was also identified as the worst reference gene, with an M value of 0.609. In our study, the GAPDH gene ranked in the second half of analyzed genes, in 19th according to which algorithm?, position with an M value of 0.68 (Fig. 2a). Answer: According to geNorm algorithm, as for 6) the changed sentence in the manuscript is: It is interesting that one of the GAPDH primer pairs (GAPDH1) in their study was identified as the worst RG, with an M value of 0.609. In our study, the GAPDH gene ranked 19th place according to geNorm, with an M value of 0.68 (Fig. 2a). 7) page 12, section “Discussion”, paragraph 3; Question: “In our case, the olive 18S reference gene had a stability value of 0.72, which is over the suggested threshold of 0.5 and was ranked in 22nd place according to which algorithm?. Answer: According to the geNorm algorithm; as for 6) the changed sentence in the manuscript is: In our case, olive 18S RG had a stability value of 0.72, which is above the suggested threshold of 0.5 and ranked 22nd according to geNorm. 8) page 12, section “Discussion”, paragraph 3; Question: “It was also shown that quantification of the expression of which gene? can be underestimated, as in the case of a potato experiment (Nicot et al. 2005).” Answer: The particular reference showed that quantification of target genes can be underestimated when 18S rRNA RG is used. We have therefore changed the sentence, also following the reviewer’s recommendations on sentence shortening: It has also been shown that expression of target genes can be underestimated (Nicot et al. 2005). 9) page 13, section “Discussion”, paragraph 1; Question: “They showed that the enzyme or the mRNA?? is already present in small drupes, embryos and endosperm.” Answer: Yes, the authors (Haralampidis et al., 1998) worked with mRNA. The mistake was corrected accordingly: “They showed that the mRNA is already present in small drupes, embryos and endosperm.” 10) page 13, section “Discussion”, paragraph 1; Question: “In a Greek study (Haralampidis et al. 1998), gene activity stayed at the maximum until the end end of what? Maturation?, without any visible drop in expression.” Answer: Gene activity stayed at the maximum until the end of the sampling period, which in their study was 28 weeks after flowering; in the Greek environment this corresponds to the end of fruit development. The sentence has been corrected accordingly, also taking into account the reviewer’s grammatical suggestion: “According to Haralampidis et al. (1998), SAD1 transcription stayed at the maximum until 28 weeks after flowering, without any visible decrease in expression.”

11) page 14, section “Discussion”, paragraph 3; Question/comment: “The correlation coefficients among standardized data obtained from values normalized with the best 9 genes and best 2 genes pair (TBP/TIP41) for all four target genes, FatA, SAD1, Acot, and LOX1 were 0.97, 0.97, 0.93, and 0.98, respectively, and these results confirm the suitability of the two reference genes selected for normalization. Low correlations between normalized (two best reference genes) and non-normalized data (0.63, 0.23, 0.04, and 0.50) reflect differences in the gene expression pattern. The presentation of non-normalized data of all four target genes showed significantly higher expression levels for two samples, 4 and 9 in all four target genes, which may be influenced by a different quantity/quality of RNA samples and by the rate of the reverse transcription step (Nolan et al. 2006). If quality of the RNA sample is not uniform, then all the study looses strength. High RNA quality is obligatory

Answer:
We think that this is the reviewer’s general comment to this paragraph of the manuscript. Isolation of high quality RNA was confirmed in our case. No action was taken here, but the whole paragraph has been rewritten, as suggested by the reviewer.

12) page 15, section “Discussion”, paragraph 2; Question/comment: “When the least stable gene ADH1 was used for normalization, low correlation coefficients compared to the two best genes normalized data were also obtained (0.34, 0.29, 0.04, and 0.25) highlighting that selection of reference genes is necessary for proper evaluation of expression studies. Similarly, a significant difference was observed in the expression pattern of two olive genes, putative polygalacturonase (PG) and farnesyl pyrophosphate synthase (FPS), when the worst internal control was used for normalization (Nonis et al. 2012). I do not see the utility of this comparison, It seems obvious to me.

Answer:
The paragraph in question is comparing our data with data from the literature – the first study of reference genes selection in olive (Nonis et al. 2012). We feel this is an important comparison and would like to keep the paragraph in the manuscript.

2B) Questions in the cover letter

1) For the amount of new information provided, the paper is way too long. The style is redundant with many unnecessary repetitions. The English form is not satisfactory, either. I recommend to shorten and simplify it as shown in the edited manuscript.

Answer:
We have substantially shortened the manuscript. The main text previously numbered 6,572 words, or 36,326 characters without spaces, and now numbers 5,793 words (11.8% reduction) or 31,666 characters without spaces (12.8% reduction). We have followed and accepted the majority of the recommendations from Reviewer#2 and removed redundancy. The paper has again been checked by a professional English proofreading service.

2) The term Reference genes(s) occurs so many times that it should be abbreviated as RG(s).

Answer:
Yes, we have inserted the suggested abbreviations in the manuscript – RG for reference gene and RGs for reference genes throughout.

3) The citation Menendez and Lupu 2007 appears not pertinent. I have no access to the full paper, but in the abstract no reference at all is made to plants, only to fatty acid synthase role in cancer: please check.

Answer:
That was our mistake, the Mendez and Lupu 2007 reference is not related to olive polyphenols. The correct reference is Cicerale et al. (2010). The manuscript and

	reference list have been corrected accordingly. Cicerale S, Lucas L, Keast R (2010) Biological activities of phenolic compounds present in virgin olive oil. Int J Mol Sci 11: 458-479 4) Materials and methods: Some clarifications are necessary (see highlighted notes). In particular, the 454 EST sequence resource should be described with some detail. Answer: Please see 2A) question 2) (same question). 5) Results: Not clear why three subsets of the data were created and analyzed separately: a justification is needed, otherwise it could seem a means of inflating the results. Answer: The systematic validation of candidate genes based on different time series resulted in different ranking of the genes based on stability values (Fig. 2). The different ranking of the candidate genes in different sample subsets emphasizes the need for detailed reference gene analysis, which is in accordance with previous studies reporting inter-tissue and experiment-dependent variation in gene stability (Gutierrez et al. 2008) and not just variation among specific plant species. Systematic validation and the use of at least two validated reference genes involved in distinct cellular functions is therefore proposed since no gene can act as a universal reference (Gutierrez et al. 2008). We have inserted a sentence in the text manuscript accordingly, p. 9: The different ranking of candidate genes in different sample subsets emphasizes the need for detailed reference gene analysis, which is in accordance with previous studies reporting inter-tissue and experiment-dependent variation in gene stability (Gutierrez et al. 2008) 6) Results: The use of the worst-performing candidate for normalization, ADH1, appears not justified, either, and the results of such comparison are obvious. Answer: Please see 2A) question 12) (same question). 7) Discussion: The data are much overdiscussed. I summarized the discussion (see edited manuscript). Answer: Thank you for the substantial improvement, we have followed the summarizing suggestions by the reviewer and substantially reduced the text (please see 2B) question 1). 8) Notes to Table 2: 454: sequence from 454 Sequence GenBank number accession provided or sequence originating from the 454 library Answer: The number 454 means sequences obtained from our 454 experiment, which is not yet published yet, although the data are now available through NCBI. Please see also 2A) question 2). 9) Supplementary Table 1: is it necessary? The primer sequences and amplicon lengths could be included in table 1. Change notes as follows: a: Amplification not achieved, b: (remove) Answer: Reviewer#3 suggested presenting efficiencies in this table, so we would like to keep it. We think that the introduction of another column of data and combining it with Table 1 would be too large for presentation. The notes have been changed as requested.

3) Reviwer #3

Reviewer #3 listed his questions and suggestions for improvement in the cover letter.

1) The performance/suitability of a reference gene is highly dependent on the experimental conditions, and even then on the experiment. Some evidence for this is also given in the paper, page 11: the study by Nonis and this study do not agree completely.
Therefore, I would like to advise to strictly stick in the conclusions and in the abstract to the evaluation of the genes in this experimental context, and to offer these genes as POTENTIALLY GOOD candidates for normalisation of gene expression data in future similar experiments. The authors suggest that they have given "the best two reference genes for use in following experiments", but that is dangerous as in other experiments the stability of the genes can be completely different. I would advise to remove such sentences that put the two genes forward as ideal reference genes for future experiments from the manuscript, and to focus on the fact that this study is valuable in offering an excellent choice of pre-evaluated candidates for other experiments.

Answer:
Yes, we agree with the suggestion. Reviewer #2 in his redrafting has already pointed out the same and changed the sentences that put the two genes forward as ideal reference genes. We have therefore these suggestions and have changed the term “the best two reference genes” throughout the paper. When we are describing the outcome of the geNorm alghoritm, we kept the term best two genes.

2) The results start with "twelve different developmental stages" whereas the M&M states "five distinct stages of fruit development". This has to be clarified. Also, where whole fruits harvested?

Answer:
Olive fruit development is described by five distinct developmental stages (from fruit set until maturity), described in a paper by Ryan et al. 1999. We covered these five stages with 12 sampling points. We have therefore changed the text in the manuscript:
Before:
“Total RNA was isolated from olive fruits from twelve different development stages of olive fruit ripening.” (p6, “Validation of reference genes”)

To:
“Total RNA was isolated from olive fruits at twelve different sampling points of olive fruit development and ripening.”

We have additionally corrected the abstract. We have replaced:
“The analyzed data points represented 12 distinct olive fruit developmental stages.”
with
“We used 12 sampling points to cover the five stages of olive fruit development.”

And in the last paragraph of Discussion, changed:
“We tested the expression stabilities of 27 selected candidate genes in olive fruits at 12 different fruit developmental stages.”
to
“We tested the expression stabilities of 27 selected candidate genes in olive fruits at 12 different sampling points of fruit development.”

Yes, we harvested the whole fruits.

3) RT-qPCR is reverse transcription quantitative PCR (real time and quantitative is the same thing)

Answer:
Thank you for pointing this out. The terms quantitative real-time polymerase chain reaction, qPCR or similar have been changed into reverse transcription quantitative polymerase chain reaction (RT-qPCR) and used throughout the paper.
The title of the paper has also been changed to “Validation of candidate reference

genes in RT-qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids metabolism". This was also suggested by Reviewer#2.

4) Was there a DNase treatment? Is there evidence of DNA presence, or can you somehow prove the absence of DNA contamination.

Answer:
Yes, we followed the on-column DNase digestion protocol as implemented in the total RNA kit by Sigma-Aldrich. However, we tested several PCR primers (ITS, microsatellites) and ran PCR with input RNA to check for amplification, which was not achieved.

5) How was [cDNA] determined?

Answer:
cDNA was not determined but we assumed 1:1 conversion from RNA. Such an approach is common in relative expression experiments, where we are not calculating the copy numbers. We have therefore changed the text in the manuscript, "Materials and methods" section:

Before:
A master mix for each PCR run was prepared using a 20 µl reaction volume containing 10 µl of FAST SYBR Green PCR Master Mix (Applied Biosystems), 10 ng of cDNA and 300 nM of each specific sense and anti-sense primer on MicroAmp Optical 96 well PCR plates (Applied Biosystems).

To:
A master mix for each PCR run was prepared using a 20 µl reaction volume containing 10 µl of FAST SYBR Green PCR Master Mix (Applied Biosystems), 2 µl of cDNA (corresponding to 10 ng of total RNA) and 300 nM of each specific sense and anti-sense primer on MicroAmp Optical 96 well PCR plates (Applied Biosystems).

6) The efficiency of the primer pairs could be listed as an extra column in suppl. table 1

Answer:
A new column with efficiencies has been introduced in Supplementary Table 1.

7) Cq (quantification cycle) should be used instead of Ct

Answer:
We have changed all Ct in manuscript, figures and tables to Cq. We have also changed the "threshold cycle" to "quantification cycle".

4) Comments for the author from editor

1) The ms will be reconsidered if authors will be willing to receive and acquire all suggestions from the reviewers, including the attached material as provided by reviewer #2.

Answer:
Please find the detailed responses to the reviewers; we have followed all the manuscript changes as suggested by reviewer#2 (pdf file named paper_corrections.pdf is attached to sow the correctons).

2) Please also consider that the total number of figures + tables should be a maximum of 5 and additional items should be placed as electronic supplementary material.

Answer:
The manuscript had 2 tables, 4 figures and 1 Supplementary table. We suggest including Table 2 as Supplementary Table 2.

Validation of candidate reference genes in RT-qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids metabolism

Tjasa Resetic[1], Natasa Stajner[2], Dunja Bandelj[1,3], Branka Javornik[2], Jernej Jakse[2]

[1]University of Primorska, Science and Research Centre of Koper, Institute for Mediterranean Agriculture and Olive Growing, Slovenia

[2]University of Ljubljana, Biotechnical Faculty, Agronomy Department, Slovenia

[3]University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Slovenia

E-mail: jernej.jakse@bf.uni-lj.si
Telephone: +386 1 3203280
Fax: +386 1 4231088

Abstract

Olive is an evergreen Mediterranean oil fruit tree with high economic, cultural and historical importance. For accurate gene expression studies of specific genes, reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) is often the method of choice, using suitable reference genes (RGs). This study identified 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*), while several classical RGs proved not to be suitable. 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. In addition to identifying two RGs for future analysis of gene expression in olive fruit, our results also provide a list of potential RGs that can be easily tested in other studies of olive gene expression in different developmental stages or in biologically challenged olive tissues. The results are also valuable for future research of genes that influence the synthesis and accumulation of olive fruit metabolites.

Keywords: RT-qPCR; olive; *Olea europaea*; reference gene; normalization; fruit development

Introduction

Reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) is one of the widely used standard methods for gene expression analysis, for its high sensitivity, high sequence specificity, no post-amplification processing (Ginzinger 2002) and high throughput potential (Heid et al. 1996). Due to its high sensitivity, any source of non-specific variation, such as sampling error, template quality and amplification efficiency may affect the final result, so normalization is necessary (Czechowski et al. 2005). Normalization of gene expression is achieved by reference to the expression of an endogenous reference gene (RG) (Kumar et al. 2011). The expression of an ideal RG should be constitutive, that is, not vary under different experimental conditions (Gachon et al. 2004). The right choice of RGs is crucial (Radonic et al. 2004).

Genes commonly used as references are usually involved in basic cellular processes and are considered to have a uniform level of expression under a range of different conditions. Recent studies have shown that commonly used RGs, often described as traditional RGs, such as *18S* rRNA, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), elongation factor-1a (*EF-1a*), polyubiquitin (*UBQ*), actin (*ACT*), alfa-tubulin and beta-tubulin (*TUA* and *TUB*), translation initiation factor 4a (*IF4a*), ubiquitin-conjugating enzyme (*UBC*) and cyclophilin (*CYP*) genes are not necessarily always the best choice for the normalization of experiments (Kumar et al. 2011). It has been shown that these traditional RGs can vary considerably in RT-qPCR experiments (Nicot et al. 2005; Remans et al. 2008). In the last few years, many studies have provided a number of novel RGs that outperform traditional RGs (Kumar et al. 2011). Validation of suitable RGs is needed before RT-qPCR studies. The use of previously non-validated RGs could greatly affect the quantification of expression levels of a target gene in an experiment (Gutierrez et al. 2008; Schmittgen and Zakrajsek 2000).

Olive (*Olea europaea* L.) is a typical evergreen tree of the Mediterranean region, grown for its fruit, which is pressed to extract highly valuable table oil or consumed pickled. In addition to containing unsaturated fatty acids, olive oil is also an important source of several biophenolic compounds specific to olive fruit, of which at least 36 have been described (Cicerale et al. 2010). Knowledge about gene regulation in metabolic pathways in olive fruits is still very limited and only a few genes involved in fatty acid metabolic pathways have been identified (Bruno et al. 2009; Conde et al. 2007; Haralampidis et al. 1998). Studies of gene expression in olive fruit employing RT-qPCR can provide insight into these biological processes. A few such studies have so far been carried out in olive, using different RGs for

normalization. Elongation factor I has been used as a RG for normalization of genes putatively involved in the main processes during olive fruit development (Galla et al. 2009). The expression of superoxide dismutase enzymes in different cell types of olive leaves has been normalized on the basis of *18S* rRNA RG (Corpas et al. 2006). The same *18S* rRNA RG was used in a study of the transcript levels of geranylgeranyl reductase gene and the content of biochemical compounds in olive pericarp (Muzzalupo et al. 2011) and in the characterization of lipoxygenase 1 (*LOX1*) transcript accumulation during different stages of olive fruit maturation (Muzzalupo et al. 2012). Hernandez et al. (2011), who studied the effects of various environmental stresses on the expression of oleate desaturase genes and fatty acid composition in olive fruit, used the ubiquitin 2 gene for normalization. Recently, a study by Nonis et al. (2012) focused on the stability of 13 candidate RGs in several stages of olive fruits and from wounded leaf tissues. Two genes, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH2*) and serine/threonine-protein phosphatase 2A (*PP2A1*), revealed by using two different algorithms, were reported to be the optimal RGs for olive fruit.

In the present study, we performed validation analysis of 29 olive genes that can be considered to be candidate RGs for studying expression in developing olive fruits. Primer pairs and amplification conditions are provided for these genes, which could also be tested for expression analysis in other olive tissue or for studies of different biological stages. Evaluation of expression stability for the tested RGs was performed by applying various scoring algorithms (Vandesompele et al. 2002; Xie et al. 2011). The RGs that were shown to be the most stable in our analyses were further used for the normalization of RT-qPCR data of four target genes (fatty acyl-ACP thioesterase A (*FatA*), stearoyl-ACP desaturase (*SAD1*), acyl-CoA thioesterase family protein (*Acot*) and lipoxygense 1 (*LOX1*)) involved in fatty acid metabolism throughout olive fruit development.

Material and methods

Plant material

Olive fruits of the variety ‘Istrska Belica’ were sampled through the periods of fertilization and fruit set, seed development, seed/pit hardening, mesocarp development and ripening (Ryan et al. 1999). Sampling was done at two-week intervals from the beginning of June until the end of November, thus yielding 12 different sampling points (Samples 1 (14 days after flowering - DAF) and 2 (29 DAF) - fertilization and fruit set; samples 3 (42 DAF) and 4 (57

DAF) - seed development; samples 5 (72 DAF) and 6 (85 DAF) - seed/pit hardening; samples 7 (98 DAF), 8 (112 DAF), 9 (129 DAF) and 10 (143 DAF) - mesocarp development; samples 11 (158 DAF) - ripening and 12 (182 DAF) – over-ripe sample). The plant material was harvested from five different olive plants from a commercial olive orchard managed by standard agricultural practice. Immediately after harvesting, the fruits were frozen in liquid nitrogen and stored at -80°C.

RNA isolation and cDNA preparation

RNA was isolated from each sample using a Spectrum Plant Total RNA Extraction Kit (Sigma-Aldrich) according to the manufacturer's protocols. RNA quality was regularly inspected by the integrity of ribosomal RNA bands on 1.2% formaldehyde agarose gels, which showed sharp and intense 18S and 28S ribosomal RNA bands, without visible degradation, confirming the suitability of the isolation method. The RNA concentrations and A260/A280 ratios were measured by a Nano Drop 2000c Spectrophotometer (Thermo Scientific) and samples were further stored at -80°C. One microgram of each RNA sample was reverse transcribed to cDNA using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, USA) employing random hexamer primers. Reversed transcribed samples were stored at -20°C.

Reference gene selection and primer design

The 29 potential RGs were selected based on literature data and on the availability of orthologous olive sequences. The selection of 26 candidate RGs was made from candidate RGs previously analyzed in other plant species. Additionally, three genes (UDP-glucose pyrophosphorylase, *UGPase*, protein kinase, *PKaba1*, and alcohol dehydrogenase 1, *ADH1*), commonly used as internal reference targets for analysis of genetically modified organisms, known to be single or low copy number genes, were included (Chaouachi et al. 2007) (Table 1).

We used 7,080 olive mRNA or DNA sequences downloaded from GenBank and 212,795 olive 454 EST clusters of sequences (Resetic, unpublished). The EST clusters originate from 454 olive transcriptome data available as raw reads in NCBI SRA archive (SRX215662), which were clustered using CD-HIT software (Li and Godzik, 2006) to decrease redundancy. Sequences were compared against the 29 RG sequences (DNA and protein) from other plant

species (Table 1) using BLASTN, BLASTX and TBLASTX algorithms (McGinnis and Madden 2004). Four genes involved in plant fatty acid metabolism (fatty acyl-ACP thioesterase A, *FatA*, XM_002303019, poplar; stearoyl-ACP desaturase, *SAD1*, AJ132636, *Gossipyum hirsutum*; acyl-CoA thioesterase family protein, *Acot*, NM_100053, *Arabidopsis*; lipoxygenase 1, *LOX1*, NM_104376, *Arabidopsis*) were included in this analysis. The longest olive hits with the highest e scores and highest identities were selected as species specific RGs or genes involved in fatty acid metabolism. For all 33 sequences, primer pairs were designed using Primer Express version 3.0 (Applied Biosystems) with the following parameter settings: max amplicon length 110 bp, optimal melting temperature 60°C, and GC content 30-80%. All primers were synthesized by Integrated DNA Technology (Leuven, Belgium). Primers and amplicon lengths are presented in Supplementary Table 1.

Two-step RT-qPCR analysis and quantification of gene expression

RT-qPCR was performed using Fast SYBR Green technology in the ABI PRISM 7500 Sequence Detection System (Applied Biosystems, Foster City, USA). A master mix for each PCR run was prepared using a 20 µl reaction volume containing 10 µl of FAST SYBR Green PCR Master Mix (Applied Biosystems), 2 µl of cDNA (corresponding to 10 ng of total RNA) and 300 nM of each specific primer on MicroAmp Optical 96 well PCR plates (Applied Biosystems). Amplification was performed using the following FAST cycling program: 95°C 20 s, 40 cycles at 95°C for 3s followed by 60°C for 30s. Three technical replicates were performed for each PCR sample. After amplification, melting curve analysis and gel electrophoresis was also performed to confirm the specificity of amplification. The final assay included a standard curve of six serial dilution points for the RGs, genes of interest and twelve cDNA samples from different stages of olive fruit development. The cDNA samples were used as PCR templates in a range of six dilutions made in 4-fold decrements starting with 50 and ending with 0.05 ng (50, 12.5, 3.13, 0.78, 0.20 and 0.05 ng). Variation between runs was minimized by performing all the reactions with a single primer pair on the same plate and by including a standard curve on each plate. The RT-qPCR efficiency was determined for each gene by using the slope of the regression line in the standard curve, calculated with ABI 7500 SDS software v2.0.4 (Applied Biosystems). All RGs and selected targets displaying efficiencies between 91% and 108% were taken into account (Supplementary Table 1).

We used two evaluation approaches for quantification. First, data from the standard dilution series was used according to the relative standard curve method. Using the standard curve, the

SDS software determines the relative quantity of target gene in each sample by comparing the quantity in each sample with the quantity in the reference sample. Relative quantities were exported to the geNorm program, which calculates the average expression stability (M value), which is defined as the average pairwise variation of a particular gene with all other control genes in a given panel of samples. A low M value is an indicator of stable gene expression (Vandesompele et al. 2002). In addition, the number of RGs required for accurate normalisation was determined by estimations of the pairwise variation of two sequential normalisation factors (V_n/V_{n+1}), which reflects the effect of including an additional gene. Vandesompele et al. (2002) suggest that this ratio should be less than 0.15 (i.e., less than 15% variation in normalization factors) to accept the proposed set as the minimum set of RGs.

In our second evaluation approach, the RefFinder program (Xie et al. 2011), which employs algorithms of four major computational programs, i.e., geNorm, Normfinder, BestKeeper and the comparative ΔC_q method, was used. Based on the classification from each program, it determines an appropriate weight for an individual gene and calculates the geometric mean of these weights for a comprehensive ranking.

After we had identified the two most stable RGs, we used them for calculating the normalized relative quantities (NRQ) of four lipid metabolism target genes.

Results

Validation of RGs

Total RNA was isolated from olive fruits at twelve different sampling points of olive fruit development and ripening. The commercial spin column RNA isolation method proved to be appropriate for olive fruit tissue. RNA samples showed intact ribosomal bands without visible degradation, with appropriate A260/280 ratios and they were amenable to cDNA synthesis. Primers for the amplification of 29 candidate olive RGs and four genes involved in fatty acid metabolism were designed using olive sequences obtained either from GenBank or our olive fruit transcriptome 454 sequences (Table 1 and Supplementary Table 1). The predicted amplicons were in a range of 90 to 100 bp. During the optimization step, the optimal primer concentration and cDNA template were defined as 300 nM and 0.5 ng/µl, respectively, since they both gave the lowest quantification cycle (C_q) vales with acceptable efficiencies. Twenty-seven primer pairs for RGs and four target genes successfully produced an amplicon

with a single dissociation curve. Two RGs did not yield amplification and were excluded from further analysis (*TUBb* and *SAND*).

For each RG, the C_q values are presented as a box-plot (Fig. 1), which shows the relative abundance of particular transcripts. The C_q values showed a range from 7.22 (*18S*) to 26.5 (*CAC*). The C_q value was particularly low for the highly abundant *18S* RG, while the C_q value range for the other 26 RGs covered a narrower range, from 21.4 (NADH dehydrogenase subunit F (*NDHF*)) to 26.5 (*CAC*). The difference represents a 34-times higher abundance of *CAC* RG compared to *NDHF*.

The average expression stability value M of the 27 candidate RGs was calculated using geNorm as the overall value for all 12 sampling points (Fig. 2a) and also for three subsets consisting of 1-4, 5-8 and 9-12, respectively (Fig. 2). The different ranking of candidate genes in different sample subsets emphasizes the need for detailed reference gene analysis, which is in accordance with previous studies reporting inter-tissue and experiment-dependent variation in gene stability (Gutierrez et al. 2008). The authors of geNorm recommend using a threshold of M=0.5 for relatively homogeneous sample panels and M=1 for heterogeneous panels, to identify genes with stable expression (Hellemans et al. 2007). As presented in Fig. 2a, *TIP41* and *TBP* (both M = 0.15) were defined as the two most stable RGs. When M was calculated for the three subsets, these two genes ranked 3rd and 4th for sampling points 1-4 (Fig. 2b, *TIP41* M=0.14, *TBP* M=0.16), 11th and 12th for sampling points 5-8 (Fig. 2c, *TBP*=0.29, *TIP41*=0.31), and 6th and 13th for sampling points 9-12 (Fig. 2d, *TBP*=0.25, *TIP*=0.35).

The geNorm program also determines the minimum number of control genes required for calculating an accurate normalization factor (NF), which is based on pairwise variation (V_n/V_{n+1}). Fig. 3a shows that the combination of the two genes (*TBP* and *TIP41*) is an adequate option for calculation of the NF in gene expression analysis of olive fruits, since the $V_{2/3}$ value in this particular case is 0.119, which is lower than the suggested cut-off value of 0.15. The overall minimum value of pairwise variation is reached with a combination of 20 candidate genes ($V_{20/21}$) and is 0.033. When the three different groups of data are considered, the pairwise value showed that the two selected genes are also sufficient for proper normalization for these particular periods ($V_{2/3}$ values 0.049, 0.053 and 0.064, Fig. 3b, c, d).

The ranking of reference genes based on all four algorithms applied in the RefFinder program (Xie et al. 2011) indicated *TBP* and *YLS8* as the two most stable genes (Supplementary Table 2). Comparing these rankings to the geNorm results, *TBP* is one of the two most stable genes in both analyses, while *TIP41* and *YLS8* ranked slightly differently but were still among the

best. The three least stable genes, *TUA3*, *CYS* and *ADH1*, are such according to all algorithms and on the comprehensive ranking list, with the exception of the BestKeeper algorithm.

Expression levels of *FatA*, *SAD1*, *Acot*, and *LOX1* genes during olive fruit development

The expression level for the three metabolic genes was quantified by a) using *TIP41* and *TBP* as RGs, b) using 9 candidate genes for which the expression stability value M was below the recommended value of 0.5 calculated for the whole set of 12 samples (Fig. 2a) and c) using *ADH1*, which had the worst stability value. The results show that *FatA*, *SAD1*, *Acot*, and *LOX1* genes were expressed in all stages of fruit development, with different relative quantification values and expression profiles (Fig. 4).

Discussion

For adequate RT-qPCR analysis, it is necessary to have a suitable RG, which allows accurate normalization of gene expression. Numerous studies have reported that no single gene expression is completely stable (Andersen et al. 2004; Pfaffl et al. 2004; Vandesompele et al. 2002). It has therefore been suggested that normalization using a single gene should be replaced by normalization based on multiple RGs, which must be experimentally identified (Vandesompele et al. 2002). The calculation of normalization factors based on more than one RG is more accurate, as confirmed by several studies (Hoerndli et al. 2004; Schmid et al. 2003). RGs with stable expression profiles should be defined for each organism. The expression stability values (M) using geNorm were less than 1 for all 27 candidate RGs, but 9 of them had an M value less than 0.5, which confirmed that these genes have acceptable expression stability (Hellemans et al. 2007). The ranking of the genes based on the expression stability values allowed us to identify two genes, *TIP41* and *TBP*; which can be used as stable references for gene expression studies in olive fruit. When another RG selection tool (RefFinder, Xie et al. 2011), which integrates four different evaluation algorithms, was used, these two genes also performed well. In a study by Reid et al. (2006) of mesocarp tissue during grapevine berry development, *TIP41* was among the top four RGs. *TIP41* was also found to be a suitable normalization reference in *Arabidopsis* (Czechowski et al. 2005). *TBP* and *TIP41* are also reported to be among the most stable RGs for studies of the development process in tomato (Exposito-Rodriguez et al. 2008). *TBP* has also been shown to be a stable reference in *Zostera marina* sea grass (Ransbotyn and Reusch 2006). The least stable genes in

our experiment were *ADH1* and *CYS*. *ADH1* is commonly used as a RG in GMO quantitative detection studies (Chaouachi et al. 2007), although it has also been previously used in a RT-qPCR experiment on coffee plant and was not considered to be a suitable RG (Barsalobres-Cavallari et al. 2009). We also tested two other genes selected from the GMO RG list, *Pkaba1* and *UGPase*, never reported as having been used in RT-qPCR normalization studies. *UGPase* ranked as the 17th most stable gene, while *Pkaba1* was among the most stable genes in our study, in 3rd place, in both cases according to the geNorm algorithm. It would therefore be reasonable also to consider other genes from the GMO list (Chaouachi et al. 2007) to be tested for RT-qPCR. Interestingly, one of our target genes, *SAD1,* is also reported to be a suitable GMO testing RG. When tested as a RG, it was the least stable according to geNorm (data not shown).

A recent study by Nonis et al. (2012) focused on the stability of candidate RGs in several stages of olive fruits and wounded leaf tissues. They evaluated 13 primer pairs, for 6 candidate RGs. All of these six genes were also considered in our study, except *PP2A*. According to their results, applying geNorm and Normfinder algorithms, *GAPDH2* and *PP2A1* were identified as the best RGs, with M values of 0.216 for *PP2A1* and 0.244 for *GAPDH2*. It is interesting that one of the *GAPDH* primer pairs (*GAPDH1*) in their study was identified as the worst RG, with an M value of 0.609. In our study, the *GAPDH* gene ranked 19th place according to geNorm, with an M value of 0.68 (Fig. 2a).

The *18S* rRNA gene is commonly used as reference in RT-qPCR analysis and has also been used in olive (Corpas et al. 2006; Muzzalupo et al. 2012). In our case, olive *18S* RG had a stability value of 0.72, which is above the suggested threshold of 0.5 and was ranked 22nd according to geNorm. The *18S* rRNA gene is usually expressed at very high levels, which is often inappropriate for normalization of weakly expressed genes (Brunner et al. 2004). It has also been shown that expression of target genes can be underestimated (Nicot et al. 2005).

On the basis of geNorm results, the expression levels of *FatA*, *SAD1*, *Acot* and *LOX1* genes were assessed in fruit samples (Fig. 4). These genes were chosen based on the expectation of a fruit developmental stage-dependent expression pattern of lipid metabolism genes.

Fatty acyl-ACP thioesterase A, *FatA*, is an intraplastidial enzyme that terminates the synthesis of fatty acids in plants and has high substrate specificity towards oleoyl-ACP (18:1-ACP). Oleic acid is the major fatty acid in olive oil and its content can be as high as 80% (Conde et al. 2008). In the first step of oleic acid formation, stearoyl-ACP is converted to oleoyl-ACP by stearoyl-ACP desaturase (*SAD1*), for which transcription profiles in different parts of olive fruit collected at different developmental stages have been studied (Haralampidis et al. 1998).

They showed that the mRNA is already present in small drupes, embryos and endosperm. Mesocarp expression began later (13 weeks after flowering) and was observed until 28 weeks after flowering. In our study, *SAD1* expression is comparable with the previously published expression levels of stearoyl-ACP desaturase obtained by Northern blot experiments (Fig. 4a) (Haralampidis et al. 1998). Expression is low at the first three sampling points (14 DAF, 29 DAF and 42 DAF), with an increase in samples 4 (57 DAF) and 5 (72 DAF), which can be attributed to embryonic gene expression. The increase in *SAD1* expression continues until sample 7 (98 DAF) and slightly declines until sample 9 (129 DAF), which can be attributed to endosperm specific developmental expression. Expression drops at sample 10 but is still present even in over-ripe sample 12. According to Haralampidis et al. (1998), *SAD1* transcription stayed at the maximum until 28 weeks after flowering, without any visible decrease in expression. In that study, Northern blot analysis was used, which in general correlates well with RT-qPCR (Dean et al. 2002); however, the olive genotype and environmental effects can explain the slightly different expression profile. The *FatA* gene expression (Fig. 4b) shows a pattern similar to *SAD1* in the first sampling points; it then starts to increase in sample 4 and reaches the highest expression in sample 9. The expression remains high up to sample 11 and then significantly drops in sample 12.

Acyl-CoA thioesterase (*Acot*) hydrolyzes fatty acyl-CoAs to free fatty acids and coenzyme A, thus providing the potential for regulation of intracellular levels of acyl-CoAs, free fatty acids and coenzyme A. This family of enzymes is suggested to have a possible role in fatty acid oxidation in animals. In plants, the first Acot gene - *ACH2* - was cloned from *Arabidopsis*. The gene was more expressed in mature tissues than in germinating seedlings, indicating that its role is probably not linked to fatty acid oxidation (Tilton et al. 2004). The *Acot* gene shows varying expression during fruit development; the highest expression is during the end stage of pit hardening and the whole stage of mesocarp development (Fig. 4c, samples 6-10). The expression declined after the mesocarp developmental stage (samples 11 and 12). The observation of expression suggests that this gene, just as the *Arabidopis Acot* gene, is not primarily involved in the beta oxidation process.

The fourth characterized expression profile for lipoxygenase 1 gene (*LOX1*) (Fig. 4d, e) showed a very large expression peak in the over ripe sample (sample 12) (Fig. 4d), although expression fluctuations can also be seen in the other 11 olive fruit samples (Fig. 4e), with an increasing trend of transcript accumulation in samples 10 and 11 (mesocarp development and ripening). The *LOX1* enzyme plays a primary role in a series of enzymatic conversion processes in the lipoxygenase (LOX) pathway, in which some of the most abundant volatiles

are formed from linoleic and linolenic acid. This pathway is particularly induced during crushing and the malaxation procedure of oil extraction (Conde et al. 2008). This is an important metabolic process for oil quality, since olive oils are characterized by their aroma, which is a complex mixture of volatile compounds (Morales et al. 1995). A recent qPCR study of *LOX1* in olive fruits showed an increasing trend of transcript accumulation towards the end of ripening in the 3 samples of two Italian cultivars (Muzzalupo et al. 2012). Our expression results of the olive *LOX1* gene support these findings, while very large *LOX1* accumulation was detected in over-ripe sample 12.

The correlation coefficients among standardized data obtained from values normalized with the best 9 genes and best two-gene pair (*TBP/TIP41*) for the genes*, FatA*, *SAD1*, *Acot*, and *LOX1* were 0.97, 0.97, 0.93, and 0.98, respectively, and these results confirm the suitability of the two RGs. Low correlations between normalized (two best RGs) and non-normalized data (0.63, 0.23, 0.04, and 0.50, respectively) appear to be mostly due to higher expression levels of all the metabolic genes in two samples, 4 and 9, which, in turn, may be influenced by a different quantity/quality of RNA samples and by the rate of the reverse transcription step (Nolan et al. 2006). When the least stable gene *ADH1* was used for normalization, low correlation coefficients compared to the two best genes normalized data were also obtained (0.34, 0.29, 0.04, and 0.25), highlighting that selection of RGs is necessary for proper evaluation of expression studies. Similarly, a significant difference was observed in the expression pattern of two olive genes, putative polygalacturonase (*PG*) and farnesyl pyrophosphate synthase (*FPS*), when the worst internal control was used for normalization (Nonis et al. 2012).

In conclusion, the presented study reports on a comprehensive analysis aimed at determining the optimal RGs for the quantification of transcript levels during olive fruit development. We tested the expression stabilities of 27 selected candidate genes in olive fruits at 12 different sampling points of fruit development. The 27 genes were further used for normalization of expression data of four target genes involved in fatty acid metabolism. On the basis of the results, we recommend two RGs, *TIP41* and *TBP*, for gene expression studies in olives.

Acknowledgment

The authors acknowledge funding by the Slovenian Research Agency, research project grant no. J4-2296 and the support of TR by grant number 1000-08-310185.

References

Andersen CL, Jensen JL, Orntoft TF (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 Res 64:5245-5250

Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG (2009) Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. BMC Mol Biol 10

Brunner A, Yakovlev I, Strauss S (2004) Validating internal controls for quantitative plant gene expression studies. BMC Plant Biol 4:14

Bruno L, Chiappetta A, Muzzalupo I, Gagliardi C, Iaria D, Bruno A, Greco M, Giannino D, Perri E, Bitonti MB (2009) Role of geranylgeranyl reductase gene in organ development and stress response in olive (*Olea europaea*) plants. Funct Plant Biol 36:370-381

Chaouachi M, Giancola S, Romaniuk M, Laval V, 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*). J Agric Food Chem 55:8003-8010

Cicerale S, Lucas L, Keast R (2010) Biological activities of phenolic compounds present in virgin olive oil. Int J Mol Sci 11: 458-479

Conde C, Agasse A, Silva P, Lemoine R, Delrot S, Tavares R, Geros H (2007) *OeMST2* encodes a monosaccharide transporter expressed throughout olive fruit maturation. Plant Cell Physiol 48:1299-1308

Conde C, Delrot S, Geros H (2008) Physiological, biochemical and molecular changes occurring during olive development and ripening. J Plant Physiol 165:1545-1562

Corpas FJ, Fernandez-Ocana A, Carreras A, Valderrama R, Luque F, Esteban FJ, Rodriguez-Serrano M, Chaki M, Pedrajas JR, Sandalio LM, del Rio LA, Barroso JB (2006) The expression of different superoxide dismutase forms is cell-type dependent in olive (*Olea europaea* L.) leaves. Plant Cell Physiol 47:984-994

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiol 139:5-17

Dean JD, Goodwin PH, 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 Mol Biol Rep 20:347-356

Exposito-Rodriguez M, Borges AA, Borges-Perez A, Perez JA (2008) Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. BMC Plant Biol 8

Gachon C, Mingam A, Charrier B (2004) Real-time PCR: what relevance to plant studies? J Exp Bot 55:1445-1454

Galla G, Barcaccia G, Ramina A, Collani S, Alagna F, Baldoni L, Cultrera NGM, Martinelli F, Sebastiani L, Tonutti P (2009) Computational annotation of genes differentially expressed along olive fruit development. BMC Plant Biol 9

Ginzinger DG (2002) Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. Exp Hematol 30:503-512

Gutierrez L, Mauriat M, Guenin S, Pelloux J, Lefebvre JF, Louvet R, Rusterucci C, Moritz T, Guerineau F, Bellini C, Van Wuytswinkel O (2008) The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnol J 6:609-618

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. J Exp Bot 49:1661-1669

Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. Genome Res 6:986-994

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 Biol 8

Hernandez ML, Padilla MN, Sicardo MD, Mancha M, Martinez-Rivas JM (2011) Effect of different environmental stresses on the expression of oleate desaturase genes and fatty acid composition in olive fruit. Phytochemistry 72:178-187

Hoerndli FJ, Toigo M, Schild A, Gotz J, Day PJ (2004) Reference genes identified in SH-SY5Y cells using custom-made gene arrays with validation by quantitative polymerase chain reaction. Anal Biochem 335:30-41

Kumar V, Sharma R, Trivedi PC, Vyas GK, Khandelwal V (2011) Traditional and novel references towards systematic normalization of qRT-PCR data in plants. Aust J Crop Sci 5:1455-1468

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658-1659

McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20-W25

Morales MT, Alonso MV, Rios JJ, Aparicio R (1995) Virgin olive oil aroma - relationship between volatile compounds and sensory attributes by chemometrics. J Agric Food Chem 43:2925-2931

Muzzalupo I, Stefanizzi F, Perri E, Chiappetta AA (2011) Transcript levels of CHL P gene, antioxidants and chlorophylls contents in olive (*Olea europaea* L.) pericarps: A comparative study on eleven olive cultivars harvested in two ripening stages. Plant Foods Hum Nutr 66:1-10

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, and technological treatments for oil extraction. Sci World J, article ID 532179, 9 pages

Nicot N, Hausman JF, Hoffmann L, Evers D (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot 56:2907-2914

Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. Nat Protoc 1:1559-1582

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. J Agric Food Chem 60:6855-6865

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. Biotechnol Lett 26:509-515

Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A (2004) Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun 313:856-862

Ransbotyn V, Reusch TBH (2006) Housekeeping gene selection for quantitative real-time PCR assays in the seagrass *Zostera marina* subjected to heat stress. Limnol Oceanograph Meth 4:367-373

Reid KE, Olsson N, Schlosser J, Peng F, Lund ST (2006) An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol 6

Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, Cuypers A (2008) Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. Planta 227:1343-1349

Ryan D, Robards K, Lavee S (1999) Changes in phenolic content of olive during maturation. Int J Food Sci Technol 34:265-274

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

Schmittgen TD, Zakrajsek BA (2000) Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods 46:69-81

Tilton GB, Shockey JM, Browse J (2004) Biochemical and molecular characterization of ACH2, an acyl-CoA thioesterase from *Arabidopsis thaliana*. J Biol Chem 279:7487-7494

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 Biol 3

Xie FL, Sun GL, Stiller JW, Zhang BH (2011) Genome-wide functional analysis of the cotton transcriptome by creating an integrated EST database. Plos One 6

Figure1
Click here to download Figure: Figure1.docx

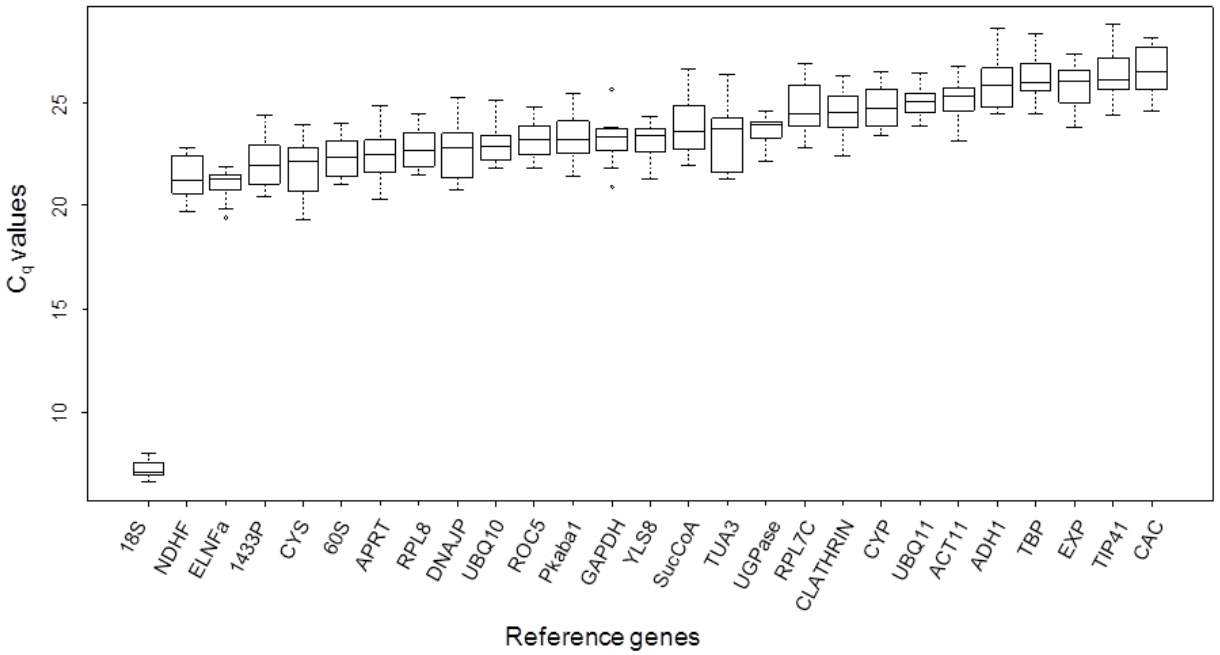


Fig. 1 Box plot presentation of quantitative cycle values (C_q) for 27 analysed reference genes. The solid line represents the median value, boxes are 25 and 75 percentiles, while whiskers are 10 and 90 percentiles and dots are outliers

Figure2
Click here to download Figure: Figure2.docx

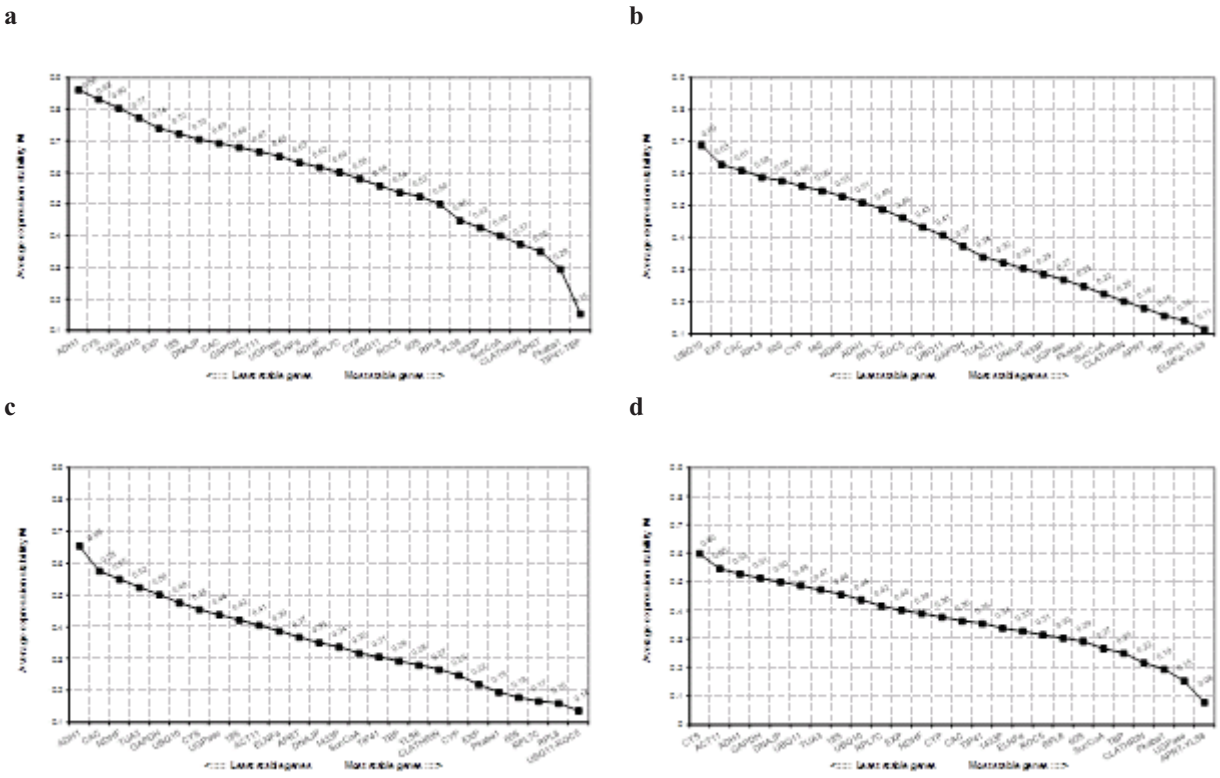


Fig. 2 Average expression stability value M of 27 evaluated olive candidate reference genes as calculated with geNorm algorithm; a) for all 12 sampling points; b) sampling points from 1 to 4; c) sampling points from 5 to 8; d) sampling points from 9 to 12

Figure3
Click here to download Figure: Figure3.docx

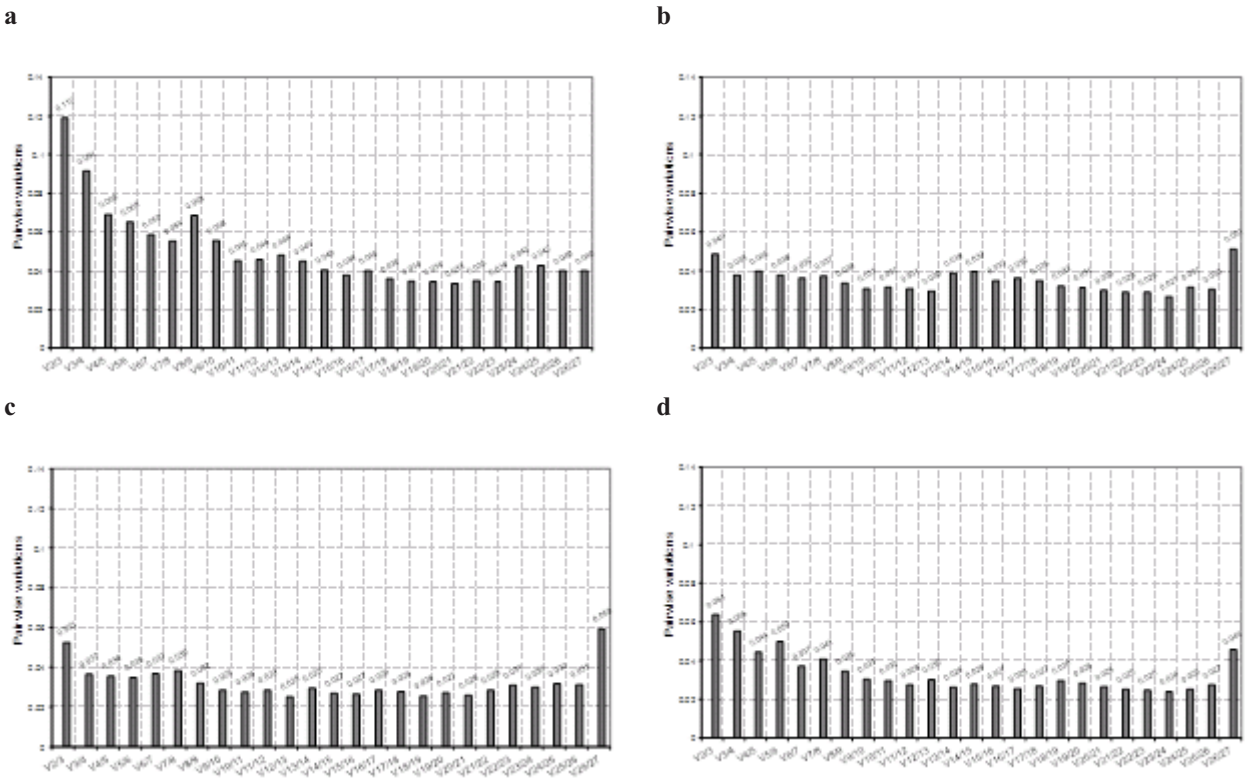


Fig 3 Pair-wise variations of $V_{n/n+1}$ value between the normalization factors NF_n and NF_{n+1}, used to determine the optimal number of reference genes for normalization. The first bar value represents the pair-wise variation between the NF value assessed for the two best genes and the NF value assessed for the best three genes (as ordered in Figure 2), followed by the addition of subsequent reference genes as listed in Figure 2; 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

Figure4
Click here to download Figure: Figure4.docx

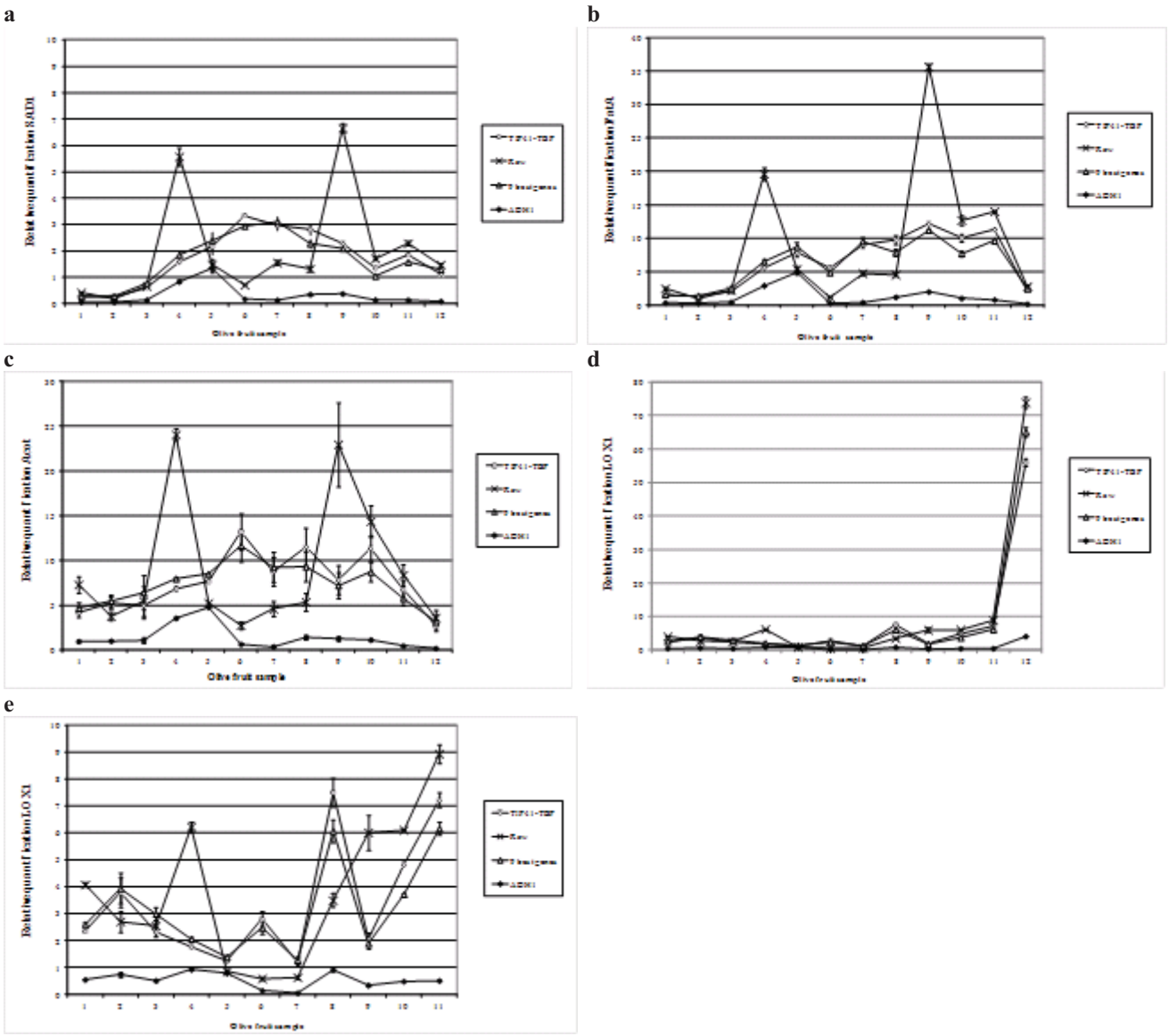


Fig. 4 Relative expression of a) *SAD1*; b) *FatA*; c) *Acot* and d) *LOX1* in all twelve analysed olive samples and e) *LOX1* in first eleven olive samples. The expression level for the four metabolic genes was quantified by using *TIP41* and *TBP* reference genes as the best two candidate genes selected by geNorm, by 9 candidate genes for which the expression stability value M was below the recommended value of 0.5 (refer to Figure 2a) and by *ADH1* reference gene, which had the highest stability value and was ranked last by geNorm among all 27 candidate genes. The expression levels are also presented as raw values.

Table1
Click here to download Table: Table1.docx

Table 1 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

No.	Gene name	Abbreviation	Reported in species	GenBank accession number	Reference	Olive sequence source[a]
1	TIP41-like family protein	TIP41	*Solanum lycopersicum*	BT014035	Exposito-Rodriguez et al. 2008	454
2	TATA binding protein	TBP	*Solanum lycopersicum*	AK329831	Exposito-Rodriguez et al. 2008	454
3	Protein kinase mRNA	Pkaba1	*Triticum aestium*	M94726	Chaouachi et al. 2007	454
4	Adenine phosphoribosyl transferase	APRT	*Solanum tuberosum*	CK270447	Nicot et al. 2005	454
5	Clathrin adaptor complex medium subunit	CLATHRIN	*Solanum lycopersicum*	SGN-U314153[b]	Exposito-Rodriguez et al. 2008	454
6	Putative succinyl-CoA ligase	SucCoA	*Urochloa brizantha*	GE617476	Pratt et al. 2005	454
7	14-3-3 protein	1433P	*Coffea canephora*	SGN-U627733[b]	Barsalobres-Cavallari et al. 2009	454
8	Yellow leaf specific gene 8 mRNA	YLS8	*Arabidopsis thaliana*	NM_120912	Czechowski et al. 2005	454
9	Ribosomal protein L2	RPL8	*Solanum tuberosum*	CK259681	Nicot et al. 2005	454
10	60S ribosomal protein	60S	*Arabidopsis thaliana*	NM_119780	Czechowski et al. 2005	454
11	Rotamase cyclophilin 5	ROC5	*Arabidopsis thaliana*	NM_203166	Czechowski et al. 2005	454
12	Polyubiquitin 11	UBQ11	*Populus trichocarpa*	BU879229	Brunner et al. 2004	454
13	Cyclophilin	CYP	*Populus trichocarpa*	BU875027	Brunner et al. 2004	454
14	60S ribosomal protein L7	RPL7C	*Coffea canephora*	SGN-U351477[b]	Barsalobres-Cavallari et al.	454

					2009	
15	NADH dehydrogenase subunit F	NDHF	*Humulus lupulus*	AY289251	Maloukh et al. 2009	AF130163
16	Elongation factor 1-α	ELNFa	*Solanum tuberosum*	AB061263	Nicot et al. 2005	454
17	(UDP)-glucose pyrophosphorylase	UGPase	*Solanum tuberosum*	U20345	Chaouachi et al. 2007	GO245620
18	Actin 11	ACT11	*Populus trichocarpa*	CA824001	Brunner et al. 2004	GO243999
19	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	*Coffea canephora*	SGN-U347734[b]	Exposito-Rodriguez et al. 2008	454
20	Clathrin adaptor complexes medium subunit	CAC	*Solanum lycopersicum*	SGN-U314153[b]	Exposito-Rodriguez et al. 2008	454
21	DnaJ-like protein	DNAJP	*Solanum lycopersicum*	AF124139	Exposito-Rodriguez et al. 2008	454
22	18S ribosomal RNA	18S	*Populus tremuloides*	AF206999	Brunner et al. 2004	L49289
23	Expressed sequence SGN-U346908, unknown protein	EXP	*Solanum lycopersicum*	SGN-U346908[b]	Exposito-Rodriguez et al. 2008	454
24	Polyubiquitin 10	UBQ10	*Coffea canephora*	SGN-U347154[b]	Barsalobres-Cavallari et al. 2009	454
25	Alpha tubulin	TUA3	*Arabidopsis thaliana*	M17189	Exposito-Rodriguez et al. 2008	454
26	Cysteine proteinase	CYS	*Coffea canephora*	SGN-U352616[b]	Barsalobres-Cavallari et al. 2009	454
27	Alcohol dehydrogenase 1	ADH1	*Zea mays*	X04050	Chaouachi et al. 2007	454
28[c]	Beta-tubulin	TUBb	*Solanum tuberosum*	Z33382	Nicot et al. 2005	454
29[c]	SAND family protein	SAND	*Arabidopsis thaliana*	NM_128399	Czechowski et al. 2007	454

[a]Sequence GenBank number accession provided or sequence originating from 454 SRA archive SRX215662

[b]Unigene accession number according to the SOL Genomics Network (http:// http://solgenomics.net/); the sequence is not present in GenBank

[c]Amplification was not achieved for two reference genes

Supplementary Table 1 Developed primer sequences for 29 reference genes and 4 target genes involved in fatty acid metabolism with predicted amplicon lengths and efficiencies

Gene name	Abbreviation	Primer sequence 5'-3'	Amplicon length in bp	Efficiency (%)
TIP41-like family protein	TIP41	CAACGGTGTCTCTCTTTTGACAGT TCATAAGCACTCCATCCACTCTCA	98	91
TATA binding protein	TBP	GAGAACAATCTTCCCTGAGACAAAA TATGAACCAGAACTATTCCCTGGAT	90	105
Protein kinase mRNA	Pkaba1	GGAGAATACCCTTCTGGATGGA GGCTTTGAATGCAGCAGAGAT	90	101
Adenine phosphoribosyl transferase	APRT	CCGATAGCCAACGCAATTG TGAGAGATACACGGGCCAAAA	90	91
Clathrin adaptor complex medium subunit	CLATHRIN	TTTTGCCCCGAAGACACTCT GAGTAAATCTTCCATTTCGGGTACTG	97	93
Putative succinyl-CoA ligase	SucCoA	TGGGAGACAAACCATCAACCA CATCCGGAGTTGATCATTAAGGT	91	93
14-3-3 protein	1433P	ACAAGTCTGCTCATGATATTGCATTA AATAGAAAACAGAGAAGTTAAGTGCAA GTC	90	93
Yellow leaf specific gene 8 mRNA	YLS8	GGTAGACCGTCTCGACGATGTC ATGATTGATCTTGGCACTGGAA	91	98
Ribosomal protein L2	RPL8	TAGCAGCAGCTTGACCACGTA GTACTGTTCGTCGGGATGCA	90	101
60S ribosomal protein	60S	TAGCAGCAGCTTGACCACGTA GTACTGTTCGTCGGGATGCA	90	94
Rotamase cyclophilin 5	ROC5	TTTCTCAATGGCTTTTACCACATC TGTACTGCGAAAACCGAATGG	90	103
Polyubiquitin 11	UBQ11	TCAAGGCTAAGATCCAGGACAAG CCAAAGTCCTTCCATCATCCA	90	108
Cyclophilin	CYP	ACTCTCCACCGGTGCCATT	90	94

		TCACTCAAAGGCTCGGCTTT		
60S ribosomal protein L7	RPL7C	ATCTGCATGGAAGATCTTGTTCAC CCCAATGGCGCCTTCA	100	101
NADH dehydrogenase subunit F	NDHF	TTCGCCGATTTTCGCAATA TGCCCCTCAAAAGTAAGTAAATAGATC	90	97
Elongation factor 1-α	ELNFa	CTCACGTTCAGCCTTGAGCTT TGTGATTGAGAGGTTTGAGAAGGA	94	99
(UDP)-glucose pyrophosphorylase	UGPase	CCATGCATAAAAGATGCTGGAA TTCAAGCTTGCCACTATTCATCA	90	96
Actin 11	ACT11	CCCAAGGCCAACAGAGAGAA GGAAAGAACGGCCTGAATAGC	90	93
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	AGCCTTGTCCTTGTCCGTAAAG TTCAGGAATCCGGAGGAGATT	90	99
Clathrin adaptor complexes medium subunit	CAC	GGCCACCTATTCAGATGGAATTT TTGTATCCACTCCTTTCCCATACC	94	102
DnaJ-like protein	DNAJP	CATCAGCCTCGCCAGGAA AGGTTGTGCAGGAGAAGAAGGT	90	91
18S ribosomal RNA	18S	GGGCTCGAAGACGATCAGATAC CCGGCGGAGTCCTATAAGC	90	91
Expressed sequence SGN-U346908, unknown protein	EXP	TCTCCGATGGGCAATAAACC TATGAATGTTGTATGGCCTGTTTGA	91	107
Polyubiquitin 10	UBQ10	GACAATGTCAAGGCAAAAATTCAG ACCATCCTCAAGCTGCTTACCA	90	98
Alpha tubulin	TUA3	CTGGAACTCGGTAACATCCACAT TTGAACCGGTTGATTTCTCAGA	90	101
Cysteine proteinase	CYS	ACACTGAAGAAGATTACCCCTACACA GTCTTCATAACCATCAATGGACACA	95	103
Alcohol	ADH1	GATGGGTCTTAAACTCTGCATCTTTA	90	97

dehydrogenase 1		TGATCTCGGCATTTGAATGTG		
Beta-tubulin[a]	TUBb	TACGAAGAGTTCTTGTTTTGAACGTT CCTCACTGCCTCAGCCATGT	90	/
SAND family protein[a]	SAND	CCCAACCCCAAGAAAATTTCA TTTTGATCCCCTTGCTGACAA	99	/
Stearoyl-ACP desaturase	SAD1	GGGCCACTTTCATTTCTCATG CGGCAATTGTACCACATATTTGA	90	94
Fatty acyl-ACP thioesterase A	FatA	TGAAGAGGATAATGCTAGCCTGAA TCAGCTCGTCTTGGCACAAG	90	98
Acyl-CoA thioesterase family protein	Acot	GAGGCAATACAAAACGGGAATG CATTTCTGCCACCTGGTGATT	90	98
Lipoxygenase 1	LOX1	CCGATGAATGGCTTGACAAA CGGATAAGGTTTTCCTGAAGACA	90	108

[a]Amplification not achieved

Supplementary Table 2
Click here to download Table: Supplementary_Table2.docx

Supplementary Table 2 Scoring table of reference genes using RefFinder (Xie et al. 2011)

GeNorm		NormFinder		BestKeeper		Delta Cq		Comprehensive ranking	
Gene name	**Stability value**	**Gene name**	**Stability value**	**Gene name**	**Stability value**	**Gene name**	**Average of Stdev**	**Gene name**	**Geomean of ranking values**
TIP41	0.222	YLS8	0.292	18S	0.33	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
CLATHRIN	0.387	1433P	0.401	UBQ10	0.704	CLATHRIN	0.74	TIP41	5.91
YLS8	0.415	APRT	0.413	ROC5	0.72	1433P	0.74	APRT	6.45
CAC	0.435	CLATHRIN	0.422	YLS8	0.77	ROC5	0.74	CLATHRIN	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	0.68	TIP41	0.973	CYP	0.89	CYP	14.64
UGPase	0.682	CYP	0.69	APRT	0.976	UGPase	0.9	SucCoA	14.89
ACT11	0.697	EXP	0.7	ADH1	0.979	EXP	0.93	UBQ10	16.21
GAPDH	0.71	NDHF	0.744	1433P	0.986	NDHF	0.94	ACT11	16.77
DNAJP	0.725	ACT11	0.744	CHLATRIN	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	0.95	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

Validation of candidate reference genes in RT-qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids metabolism

Tjasa Resetic[1], Natasa Stajner[2], Dunja Bandelj[1,3], Branka Javornik[2], Jernej Jakse[2]

[1]University of Primorska, Science and Research Centre of Koper, Institute for Mediterranean Agriculture and Olive Growing, Slovenia

[2]University of Ljubljana, Biotechnical Faculty, Agronomy Department, Slovenia

[3]University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Slovenia

E-mail: jernej.jakse@bf.uni-lj.si
Telephone: +386 1 3203280
Fax: +386 1 4231088

Abstract

Olive is an evergreen Mediterranean oil fruit treeplant with high economic, cultural and historical importance. For accurate gene expression studies of specific metabolic genes, quantitative real-time reverse transcriptase-quantitatve polymerase chain reaction (RT-qPCR or qPCR) is often the method of choice, using suitable reference genes (RGs) which, through a normalization step, ensure proper presentation of the expression data. This study identified stable reference genes RGs for RT-qPCR studies of developing olive fruit from 29 selected reference gene RG candidates., which were further used to normalize expression data of four target genes involved in fatty acid metabolism. We used 12 sampling points to cover the five stages of olive fruit developmentThe analyzed data points represented 12 distinct olive fruit developmental stages. According to the results of the geNorm algorithm, the two best reference genes RGs were TIP41-like family protein (*TIP41*) and TATA binding protein (*TBP*), while several classical reference genes RGs proved not to be suitable. Based Using on the the selection of suitable candidate two new reference genesRGs, four genes involved in the metabolism of fatty acids were normalized studied and showed distinct expression patterns associated with mesocarp development and ripening stages. In addition to giving the best identifying two reference genes RGs for future analysis of gene expression in olive fruit, these our results also provide a list of potential references RGs that can be easily applied tested in other studies of olive gene expression in different developmental stages or in biologically challenged olive tissues. The results are also valuable for future research of genes that influence the synthesis and accumulation of olive fruit metabolites.

Keywords: RT-qPCR; olive; *Olea europaea*; reference gene; normalization; fruit development

Introduction

~~Quantitative real-time~~ Reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR ~~or qPCR~~) is one of the widely used standard methods for gene expression analysis, ~~of various biological tissue states or conditions in many organisms. It has a number of advantages, such as~~ for its high sensitivity, high sequence specificity, ~~it does not require~~no post-amplification processing (Ginzinger 2002) and ~~it is amenable to~~ high ~~increased sample~~ throughput potential (Heid et al. 1996). Due to its high ~~the~~ sensitivity~~, which is significant for RT-qPCR,~~ any source of non-specific variation, such as sampling error, template quality and amplification efficiency may affect the final result, so normalization is necessary ~~for the correction of non-specific variations~~ (Czechowski et al. 2005). Normalization of gene expression is achieved by ~~measuring~~ reference to the expression of ~~various endogenous reference genes~~an endogenous reference gene (RG) (Kumar et al. 2011). The expression of ~~these~~ an ideal ~~genes~~ RG should be constitutive, that is, not vary under different experimental conditions (Gachon et al. 2004). T~~and t~~he right choice of RGs~~reference genes~~ is ~~therefore of great importance~~ crucial (Radonic et al. 2004).

Genes commonly used as references are usually involved in basic cellular processes and are considered to have a uniform level of expression under a range of different conditions. Recent studies have shown that commonly used RGs~~reference genes~~, often described as traditional RGs~~reference genes~~, such as 18S rRNA, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), elongation factor-1a (*EF-1a*), polyubiquitin (*UBQ*), actin (*ACT*), alfa-tubulin and beta-tubulin (*TUA* and *TUB*), translation initiation factor 4a (*IF4a*), ubiquitin-conjugating enzyme (*UBC*) and cyclophilin (*CYP*) genes are not necessarily always the best choice for the normalization of experiments (Kumar et al. 2011). It has been shown that these traditional RGs~~references~~ can vary considerably in RT-qPCR experiments (Nicot et al. 2005; Remans et al. 2008). In the last few years, many studies have provided a number of novel RGs~~references able to substitute for traditional reference genes in the future. Comparative studies have also shown~~ that ~~novel genes often~~ outperform traditional RGs~~reference genes~~ (Kumar et al. 2011). Validation of suitable ~~reference genes~~ RGs is ~~thus~~ needed before RT-qPCR studies. The use of previously non-validated ~~references~~ RGs could greatly affect the quantification of expression levels of a ~~the~~ target gene in an experiment (Gutierrez et al. 2008; Schmittgen and Zakrajsek 2000).

Olive (*Olea europaea* L.) is a typical evergreen tree of the Mediterranean region, grown for its fruit, which is pressed to extract~~into~~ highly valuable table oil or consumed pickled. In

addition to containing unsaturated fatty acids, olive oil is also an important source of several biophenolic compounds specific to olive fruit, of which at least 36 have been described ~~(Menendez and Lupu 2007)~~(Cicerale et al. 2010).. Knowledge about ~~the~~ gene regulation in metabolic pathways in olive fruits is still very limited and only a few genes involved in fatty acid metabolic pathways have ~~to date~~ been identified (Bruno et al. 2009; Conde et al. 2007; Haralampidis et al. 1998). Studies of gene expression in olive fruit employing RT-qPCR can provide ~~a novel~~ insight into these biological processes. A~~However, very~~ few such~~RT-qPCR~~ studies have so far been carried out in olive, using different RGs~~reference genes~~ for normalization. Elongation factor I has been used as a RG~~reference gene~~ for normalization of ~~differentially expressed~~ genes putatively involved in the main processes during olive fruit development (Galla et al. 2009). The expression of superoxide dismutase enzymes in different cell types of olive leaves has been normalized on the basis of ~~the~~ 18S rRNA RG~~reference gene~~ (Corpas et al. 2006). The same 18S rRNA RG~~reference gene~~ was used in a study of the transcript levels of geranylgeranyl reductase gene and the content of biochemical compounds in ~~the~~ olive pericarp (Muzzalupo et al. 2011) and in the ~~further work of the correlation between~~ characterization of lipoxygenase 1 (*LOX1*) transcript accumulation during different stages of olive fruit maturation (Muzzalupo et al. 2012). Hernandez et al. (~~Hernandez et al.~~ 2011), who studied the effects of various environmental stresses on the expression of oleate desaturase genes and fatty acid composition in olive fruit, used the ubiquitin 2 gene ~~sequence~~ for normalization ~~of the results~~. Recently, a study by Nonis et al. (~~Nonis et al~~. 2012) ~~was the first to focus~~focused on the stability of 13 ~~olive~~ candidate RGs~~reference genes~~ in several stages of olive fruits and from wounded leaf tissues~~, aiming at reliable quantification of transcripts~~. Two genes, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH2*) and serine/threonine-protein phosphatase 2A (*PP2A1*), revealed by using two different algorithms were reported to be the optimal ~~best~~ RGs~~reference genes~~ for olive fruit.

In the present study, we performed validation analysis of 29 olive genes that can be considered to be candidate RGs ~~reference genes~~ for studying expression in developing olive fruits. ~~The list of candidates was chosen based on the literature data from other plant tissues.~~ Primer pairs and amplification conditions are provided for these genes, which could also be tested for expression analysis in other olive tissue or for studies of different biological stages. Evaluation of expression stability for the tested RGs~~reference genes~~ was performed by applying various scoring algorithms (Vandesompele et al. 2002; Xie et al. 2011). The RGs~~reference genes~~ that were shown to be the most stable in our analyses were further used for the normalization of RT-qPCR data of four target genes (fatty acyl-ACP thioesterase A

(*FatA*), stearoyl-ACP desaturase (*SAD1*), acyl-CoA thioesterase family protein (*Acot*) and lipoxygense 1 (*LOX1*)) involved in fatty acid metabolism throughout ~~the entire~~ olive fruit develop~~ment~~ment~~ing period~~.

Material and methods

Plant material

Olive fruits of the variety 'Istrska Belica' were sampled through ~~five distinct stages of fruit development, representing~~ the periods of fertilization and fruit set, seed development, seed/pit hardening, mesocarp development and ripening (Ryan et al. 1999). Sampling was done at two-week intervals from the beginning of June until the end of November, thus yielding 12 different sampling points (Samples 1 (14 days after flowering - DAF) and 2 (29 DAF) - fertilization and fruit set; samples 3 (42 DAF) and 4 (57 DAF) - seed development; samples 5 (72 DAF) and 6 (85 DAF) - seed/pit hardening; samples 7 (98 DAF), 8 (112 DAF), 9 (129 DAF) and 10 (143 DAF) - mesocarp development; samples 11 (158 DAF) - ripening and 12 (182 DAF) – over-~~-~~ripe sample). The plant material was harvested from five different olive plants from a commercial olive orchard managed by standard agricultural practice. Immediately after harvesting, the fruits were frozen in liquid nitrogen and stored ~~until use~~ at -80°C.

RNA isolation and cDNA preparation

RNA was isolated from each sample using a Spectrum Plant Total RNA Extraction Kit (Sigma-Aldrich) according to the manufacturer's protocols. RNA quality was regularly inspected by the integrity of ribosomal RNA bands on 1.2% formaldehyde agarose gels, which showed sharp and intense 18S and 28S ribosomal RNA bands, without visible degradation~~with a small number of smeared tracks~~, confirming the suitability of the isolation method. The RNA concentrations and A260/A280 ratios were measured by a Nano Drop 2000c Spectrophotometer (Thermo Scientific) and samples were further stored at -80°C. One microgram of each RNA sample was reverse transcribed to cDNA using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, USA) employing random hexamere primers. Reversed transcribed samples were stored at -20°C ~~until further use~~.

Reference gene selection and primer design

The 29 potential RGs~~reference genes~~ were selected ~~for testing~~ based on literature data and on the availability of orthologous olive sequences. The selection of 26 candidate RGs~~reference genes~~ was made from ~~previously analyzed~~ candidate RGs previously analyzed ~~reference genes~~ in other plant species~~, which are considered to be traditional or novel reference genes~~. Additionally, three~~ selected~~ genes (UDP-glucose pyrophosphorylase, ~~(~~*UGPase*~~)~~, protein kinase, ~~(~~*PKaba1*,~~)~~ and alcohol dehydrogenase 1, ~~(~~*ADH1*~~)~~), ~~are~~ commonly used as internal reference targets for analysis of genetically modified organisms, known to be ~~and are~~ single or low copy number genes, were included (Chaouachi et al. 2007) (Table 1).

We ~~compared~~ used 7,080 olive mRNA or DNA sequences downloaded from GenBank and 212,795 olive 454 EST clusters of sequences (Resetic, unpublished). The EST clusters are originating from 454 olive transcriptome data available as raw reads in NCBI SRA archive (SRX215662) which were clustered using CD-HIT software (Li and Godzik, 2006) for decreasing of the redundancy. Sequences were compared against the 29 RGs~~reference genes~~ sequences (DNA and protein) from other plant species (Table 1) using BLASTN, BLASTX and TBLASTX algorithms (McGinnis and Madden 2004). ~~Additionally,~~ F~~f~~our ~~target~~ genes involved in plant fatty acid metabolism (fatty acyl-ACP thioesterase A, *FatA*, XM_002303019, poplar; stearoyl-ACP desaturase, *SAD1*, AJ132636, *Gossipyum hirsutum*; acyl-CoA thioesterase family protein, *Acot*, NM_100053, *Arabidopsis*; lipoxygenase 1, *LOX1*, NM_104376, *Arabidopsis*) were included in this analysis. ~~chosen for further expression analysis study.~~ The longest olive hits with the highest e scores and highest identities were selected as species specific RGs~~reference genes~~ or genes involved in fatty acid metabolism ~~and further used in primer development~~. For all 33 sequences, primer pairs were designed using Primer Express version 3.0 (Applied Biosystems) with the following~~recommended~~ parameter settings: max amplicon~~fication region~~ length 110 bp, optimal melting temperature ~~at~~ 60°C, and GC content 30-80%. All primers were synthesized by Integrated DNA Technology (Leuven, Belgium). Primers and amplicons lengths are presented in Supplementary Table 1.

Two-step RT-~~real time~~ qPCR analysis and quantification of gene expression

RT-~~eal time~~q PCR was performed using Fast SYBR Green technology in the ABI PRISM 7500 Sequence Detection System (Applied Biosystems, Foster City, USA). A master mix for

each PCR run was prepared using a 20 µl reaction volume containing 10 µl of FAST SYBR Green PCR Master Mix (Applied Biosystems), 2 µl of cDNA (corresponding to 10 ng of total RNA)~~10 ng of cDNA~~ and 300 nM of each specific ~~sense and anti-sense~~ primer on MicroAmp Optical 96 well PCR plates (Applied Biosystems). Amplification was performed using the following FAST cycling program: 95°C 20 s, 40 cycles at 95°C for 3s followed by 60°C for 30s. Three technical replicates were performed for each PCR ~~reaction (~~sample~~) to increase the reliability of the results~~. After amplification, melting curve analysis and gel electrophoresis ~~analysis~~ ~~were~~was also performed~~-~~ to confirm the specificity of amplification~~as verification that the primers amplified only a single product of the expected size~~. The final assay included a standard curve of six serial dilution points for the RGs~~reference genes~~, genes of interest and twelve cDNA samples from different stages of olive fruit development. The cDNA samples were used as PCR templates in a range of six dilutions made in 4-fold decrements starting with 50 and ending with 0.05 ng (50, 12.5, 3.13, 0.78, 0.20 and 0.05 ng). Variation between runs was minimized by performing all ~~of~~ the reactions with~~containing~~ a single primer pair on the same plate and by including a standard curve on each plate. ~~The cDNA samples were used as PCR templates in a range of 4-fold series dilutions starting with 50 and ending with 0.05 ng.~~ The RT-~~real-time~~ qPCR efficiency was determined for each gene by using the slope of the regression line in the standard curve, calculated with ABI 7500 SDS software v2.0.4 (Applied Biosystems). All RGs ~~reference genes~~ and selected targets displaying efficiencies between 91~~1~~0% and 108% were taken into account (Supplementary Table 1)~~, confirming that the PCR reaction was not inhibited~~.

We used two evaluation approaches for quantification. First, data from the standard dilution series ~~to create a standard curve~~ was used according to the relative standard curve method. Using the standard curve, the SDS software determines the relative quantity of target gene in each sample by comparing the ~~target~~ quantity in each sample with~~to~~ the~~target~~ quantity in the reference sample. Relative ~~In order to determine the most stable genes, relative~~ quantities were exported to the geNorm program, which calculates the average expression stability (M value), which is defined as the average pairwise variation of a particular gene with all other control genes in a given panel of samples. A low M value is an indicator of stable gene expression ~~and this increases the suitability of a particular gene as a control gene~~ (Vandesompele et al. 2002). In addition, the number of RGs~~reference genes~~ required for accurate normalisation was determined by estimations of the pairwise variation of two sequential normalisation factors (V_n/V_{n+1}), which reflects the effect of including an additional gene. Vandesompele et al. (~~Vandesompele et al.~~ 2002) suggest that this ratio should be less

than 0.15 (i.e., less than 15% variation in normalization factors) to accept the proposed set as the minimum set of RGs~~reference genes for normalization~~.

In our second evaluation approach, the RefFinder program (Xie et al. 2011), which employs algorithms of four major computational programs, i.e. geNorm, Normfinder, BestKeeper and the comparative ΔC_{qt} method, was used ~~to compare and classify the tested candidate reference genes~~. Based on the classification from each program, it determines an appropriate weight for an individual gene and calculates the geometric mean of these weights for a comprehensive ranking.

After we had identified the two most stable ~~–~~ RGs~~reference genes~~, we used them for calculating the normalized relative quantities (NRQ) of four lipid metabolism target genes.

Results

Validation of RGs~~reference genes~~

Total RNA was isolated from olive fruits at~~from~~ twelve different sampling points~~development stages~~ of olive fruit development and ripening. The commercial spin column RNA isolation method proved to be appropriate for olive fruit tissue. RNA samples showed intact ribosomal bands without visible degradation, with appropriate A260/280 ratios and they were amenable to cDNA synthesis. Primers for the amplification of ~~twenty nine~~ 29 candidate olive RGs ~~reference genes~~ and four genes involved in fatty acid metabolism were designed using olive sequences obtained ~~from~~ either from GenBank or our ~~own~~ olive fruit transcriptome 454 sequences (Table 1 and Supplementary Table 1). The ~~designed primers amplified~~ predicted amplicons were in a range of 90 to 100 bp. During the optimization step, the optimal primer concentration and cDNA template were defined as 300 nM and 0.5 ng/µl, respectively, since they both gave the lowest quantification cycle (C_{tq}) vales with acceptable efficiencies. ~~After completion of the reactions,~~ Ttwenty-seven primer pairs for RGs~~reference genes~~ and four target genes successfully produced an amplicon with a single dissociation curve ~~from 12 olive fruit cDNA samples~~. Two RGs~~reference genes~~ did not yield amplification and were excluded from further analysis (~~beta-tubulin (~~*TUBb*~~)~~ and ~~SAND family protein (~~*SAND*~~); Table 1 and Supplementary Table 1))~~.

For each ~~–~~ RG~~reference gene~~, the ~~threshold cycle values (~~C_{qt}~~)~~ values are presented as a box-plot (Fig. 1), which shows the relative abundance of particular transcripts. The C_{qt} values showed a range from 7.22 (*18S*) to 26.5 (~~clathrin adaptor complexes medium subunit (~~*CAC*~~)~~).

The C_{qt} value was particularly low for the highly abundant *18S* RG ~~reference gene~~, while the C_{qt} value range for the other 26 RGs~~reference genes~~ covered a narrower range, from 21.4 (NADH dehydrogenase subunit F (*NDHF*)) to 26.5 (*CAC*). The difference represents ~~a 34-times~~a 34-times higher abundance of ~~the~~ *CAC* RG~~reference gene~~ compared to *NDHF*~~, while the *18S* reference is more than 600,000-times more abundant than *CAC*~~.

The average expression stability value M of the 27 ~~evaluated olive~~ candidate RGs ~~reference genes~~ was calculated using geNorm as the overall value for all 12 sampling points (Fig. 2a) and also for three subsets consisting of ~~4 consesecutive sampling points together (~~1-4~~, Fig. 2b~~, 5-8~~ Fig. 2c~~, and 9-12, respectively~~ (Fig. 2~~d~~). The different ranking of candidate genes in different sample subsets emphasizes the need for detailed reference gene analysis, which is in accordance with previous studies reporting inter-tissue and experiment-dependent variation in gene stability (Gutierrez et al. 2008). The authors of geNorm recommend using a threshold of M=0.5 for relatively homogeneous sample panels and M=1 for heterogeneous panels, to identify genes with stable expression (Hellemans et al. 2007). As presented in Fig. 2a, *TIP41* and *TBP* (both M = 0.15) were defined as the two most stable RGs~~reference genes in the overall M calculation~~. When M was calculated for the three subsets, these two genes ~~3 time series, they were both still~~ ranked ~~high: in~~ 3rd ~~-~~and~~-~~ 4th ~~place~~ for sampling points 1-4 (Fig. 2b, *TIP41* M=0.14, *TBP* M=0.16), ~~for sampling points 5-8 they were ranked~~ 11th and 12th for sampling points 5-8 ~~among all~~ (Fig. 2c, *TBP*=0.29, *TIP41*=0.31), and ~~while for sampling points 9-12 *TBP* was ranked in~~ 6th ~~place,~~ and ~~*TIP41* was ranked~~ 13th for sampling points 9-12 (Fig. 2d, *TBP*=0.25, *TIP*=0.35).

The geNorm program also determines the minimum number of control genes required for calculating an accurate normalization factor (NF), which is based on pairwise variation (V_n/V_{n+1}). ~~Our data revealed (~~Fig. 3a shows~~a)~~ that the combination of the two genes (*TBP* and *TIP41*) is an adequate option for calculation of the NF in gene expression analysis of olive fruits, since the $V_{2/3}$ value in this particular case is 0.119, which is lower than the suggested cut-off value of 0.15. The overall minimum value of pairwise variation is reached with a combination of 20 candidate genes ($V_{20/21}$) and is 0.033. When the three different groups of data ~~(sampling points 1-4, 5-8, and 9-12)~~ are considered, the pairwise value showed that the~~ selected~~ two selected genes ~~for these particular periods~~ are also sufficient for proper normalization for these particular periods ($V_{2/3}$ values 0.049, 0.053 and 0.064, Fig. 3b, c, d).

~~The 27 candidate reference genes were also analyzed for their stability by the RefFinder program (Xie et al. 2011)~~. The ~~comprehensive~~ ranking of reference genes based on all four algorithms applied in~~ the~~ the RefFinder program (Xie et al. 2011) indicated~~software revealed~~

TBP and ~~yellow leaf specific gene 8 mRNA (~~*YLS8*~~)~~ as ~~to be~~ the two most stable genes (Supplementary Table 2). Comparing these rankings to the geNorm results, *TBP* is one of the two most stable genes in both analyses, while *TIP41* and *YLS8* ranked slightly differently but were still among the best. The three least stable genes, ~~alpha tubulin (~~*TUA3*~~)~~, ~~cystein proteinase (~~*CYS*~~)~~ and *ADH1*, are such ~~identical~~ according to~~in~~ all algorithms and on the comprehensive ranking list, with the exception of the BestKeeper algorithm.

Expression levels of *FatA*, *SAD1*, *Acot*, and *LOX1* genes during olive fruit development

The expression level for the three metabolic genes was quantified by a) using *TIP41* and *TBP* as RGs~~reference genes as the two best candidate genes selected by geNorm~~, b) ~~then by~~ using 9 candidate genes for which the expression stability value M was below the recommended value of 0.5 calculated for the whole set of 12 samples (Fig. 2a) and c) using~~by~~ *ADH1* ~~reference gene~~, which had the worst~~highest~~ stability value. ~~and was ranked last by geNorm among all 27 candidate genes and accordingly assumed to be less suitable for normalization (Vandesompele et al. 2002). The results of combinations of normalized expression levels of *FatA*, *SAD1*, *Acot* and *LOX1* genes are presented in Fig. 4. The expression levels of these genes were evaluated in twelve sampled fruit developmental stages.~~ The results show that *FatA*, *SAD1*, *Acot*, and *LOX1* genes were expressed in all stages of fruit development, with different relative quantification values and expression profiles (Fig. 4).

Discussion

~~Real-time PCR (RT-qPCR) is an important technique that allows precise quantitative analysis of gene expression (Heid et al. 1996).~~ For adequate RT-qPCR analysis, it is necessary to have a suitable RG~~reference gene(s)~~, which allows accurate normalization of gene expression. ~~An ideal control gene should not be affected by the experimental conditions, including different cell types, developmental stages and/or sample treatments.~~ Numerous studies have reported that no single gene expression is completely stable ~~under the aforementioned experimental conditions~~ (Andersen et al. 2004; Pfaffl et al. 2004; Vandesompele et al. 2002). It has therefore been suggested that normalization using a single gene should be replaced by normalization based on multiple~~,~~ RGs~~best-performing reference genes~~, which must be ~~are~~ experimentally identified~~defined~~ (Vandesompele et al. 2002). The calculation of normalization factors based on more than one RG~~reference gene~~ is ~~also~~ more accurate, as ~~has~~

~~been~~ confirmed by several studies (Hoerndli et al. 2004; Schmid et al. 2003). RGs~~Internal reference genes~~ with stable expression profiles should ~~therefore~~ be defined for each organism. ~~organism before real-time PCR experiments. This particular study was conducted to detect a set of appropriate genes for expression analysis of developing fruits but it can easily be applied for different tissues or diverse biological treatments of olive plants.~~

~~The~~The expression stability values (M) using geNorm were less than 1 for all 27 candidate ~~putative~~ RGs~~olive reference genes~~, but 9 of them had an M value less than 0.5, which confirmed that these genes have acceptable expression stability (Hellemans et al. 2007). The ranking of the genes based on the expression stability values allowed us to identify ~~provided the best result for~~ two genes, *TIP41* and *TBP*; which ~~these two genes~~ can be used as stable references for ~~and are suggested for future analysis of~~ gene expression studies in olive fruit ~~studies~~. ~~The data of gene expression stability obtained by comprehensive rankings were slightly different w~~.When another RG ~~reference gene~~ selection tool (RefFinder~~),~~ ~~(~~Xie et al. 2011~~)) was used~~, which integrates four different evaluation algorithms was used, these two genes also performed well.~~.~~ ~~In comparison with other novel and traditional references, *TIP41* was identified as the most stable novel reference gene, while *TBP* belongs to the group of the traditional reference genes. It has already been shown that traditional references can be outperformed by novel reference candidate genes (Czechowski et al. 2005).~~ In a study by Reid et al. (~~Reid et al.~~ 2006) of mesocarp tissue during grapevine berry development, *TIP41* was~~ranked~~ among the top four~~-~~ RGs~~reference genes~~. ~~The~~ *TIP41* ~~gene~~ was also found to be a suitable normalization reference in *Arabidopsis* (Czechowski et al. 2005). ~~The common reference gene~~ *TBP* and *TIP41* are also reported to be among the most stable RGs ~~reference genes~~ for ~~in tomato plant~~ studies of the development process in tomato~~, in which they ranked as the most stably expressed genes among all of the novel and traditional references used in the study~~ (Exposito-Rodriguez et al. 2008). *TBP* has also been shown to be a stable reference in *Zostera marina* sea grass (Ransbotyn and Reusch 2006). ~~This clearly shows that standardization using common reference genes can be successfully improved by novel reference genes, especially when comparing samples from different tissues, organs and developmental stages.~~ The least stable genes in our experiment were *ADH1* and *CYS*. *ADH1* is ~~more~~ commonly used as aRG ~~reference gene~~ in GMO quantitative detection studies (Chaouachi et al. 2007), although it has also been previously used in a RT-qPCR experiment on coffee plant and was not considered to be a suitable RG ~~normalization reference gene~~ (Barsalobres-Cavallari et al. 2009). We also tested two other genes selected from the GMO RG ~~reference gene~~ list, *Pkaba1* and *UGPase*, never reported as having been used in RT-qPCR

normalization studies. *UGPase* ~~proved to be a less favorable gene in our experiment,~~ ranked as the 17th most stable gene, while *Pkaba1* was among the most stable genes in our study, in 3rd place, in both cases according to the geNorm algorithm. It would therefore be reasonable also to consider other genes from the GMO ~~reference gene~~ list (Chaouachi et al. 2007) to be tested ~~as possible~~ for RT-qPCR ~~reference genes~~. Interestingly, one of our target genes, *SAD1,* is also reported to be a suitable GMO testing ~~RG~~ RG~~reference gene~~. When tested ~~this particular gene (*SAD1*) was considered~~ as a ~~RG~~ RG~~reference gene~~, it was ~~ranked as~~ the least stable according to geNorm~~, replacing *ADH1* reference gene in last place~~ (data not shown).

A recent study by Nonis et al. (~~Nonis et al.~~ 2012) focused on the stability of candidate RGs ~~reference genes~~ in several stages of olive fruits and wounded leaf tissues~~, aiming at reliable quantification of transcripts~~. They ~~also selected reference genes on the basis of literature data and~~ evaluated 13 primer pairs, for ~~which represented~~ 6 candidate RGs~~reference genes~~. All of these six genes were also considered in our study except *PP2A*. According to their results, applying geNorm and Normfinder algorithms, *GAPDH2* and *PP2A1* were identified as the best RGs ~~two reference genes for olive fruit development and ripening~~, with M values of 0.216 for *PP2A1* and 0.244 for *GAPDH2*. It is interesting that one of the *GAPDH* primer pairs (*GAPDH1*) in their study was ~~also~~ identified as the worst ~~RG~~ RG~~reference gene~~, with an M value of 0.609. In our study, the *GAPDH* gene ranked ~~in the second half of analyzed genes, in~~ 19th ~~position~~ place according to geNorm, with an M value of 0.68 (Fig. 2a).

The *18S* rRNA ~~reference~~ gene is commonly used as reference in ~~quantitative~~ RT-qPCR analysi~~es~~ and has also been used in ~~previous~~ olive ~~expression experiments~~ (Corpas et al. 2006; Muzzalupo et al. 2012). In our case, ~~the~~ olive *18S* RG ~~reference gene~~ had a stability value of 0.72, which is above ~~over~~ the suggested threshold of 0.5 and was ranked ~~in~~ 22nd ~~place~~ according to geNorm. The *18S* rRNA ~~reference~~ gene is usually expressed at very high levels, which is often inappropriate for normalization of weakly expressed genes (Brunner et al. 2004). It has~~was~~ also been shown that ~~quantification of the~~ expression of target genes can be underestimated~~, as in the case of a potato experiment~~ (Nicot et al. 2005).

On the basis of geNorm results, ~~*TIP41* and *TBP* as the best two reference genes, 9 reference genes showing the lowest pairwise variation, equal to or below the recommended M value of 0.5 and *ADH1* gene ranked with the highest expression stability value, were used to normalize~~ the expression levels of *FatA*, *SAD1*, *Acot* and *LOX1* genes were assessed in fruit samples ~~of olive at different developmental stages~~ (Fig. 4). These genes were chosen based on the expectation of a fruit developmental stage~~-~~-dependent expression pattern of lipid metabolism genes.

Fatty acyl-ACP thioesterase A, *FatA*, is an intraplastidial enzyme that terminates the synthesis of fatty acids in plants and has high substrate specificity towards oleoyl-ACP (18:1-ACP). Oleic acid is the major fatty acid in olive oil and its content can be as high as 80% ~~in selected olive oils~~ (Conde et al. 2008). In the first step of oleic acid formation ~~in olive fruits~~ stearoyl-ACP is converted to oleoyl-ACP by stearoyl-ACP desaturase (*SAD1*), for which transcription profiles in different parts of olive fruit collected at different developmental stages have been ~~s~~tudied (Haralampidis et al. 1998). They showed that the ~~enzyme~~mRNA is already present in small drupes, embryos and endosperm. Mesocarp expression began~~ins~~ later (13 weeks after flowering) and was observed ~~remained in their experiment~~ until 28 weeks after flowering. In our study, ~~*SAD1* and *FatA* enzymes expression profiles were studied, since they are both included in the formation of oleic acid in two reactions that follow each other. The results of the expression study for~~ *SAD1* expression is ~~gene (Fig. 4a) are~~ comparable with ~~to~~ the previously published expression levels of stearoyl-ACP desaturase obtained by Northern blot experiments (Fig. 4a) (Haralampidis et al. 1998). Expression is ~~of the olive *SAD1* gene is present but~~ low at the first three sampling points (14 DAF, 29 DAF and 42 DAF), with an increase ~~in expression~~ in samples 4 (57 DAF) and 5 (72 DAF) which can be~~is~~ attributed to embryonic gene expression. The increase in *SAD1* expression continues until sample 7 (98 DAF) and slightly declines until sample 9 (129 DAF), which can be ~~is~~ attributed to endosperm specific developmental expression. Expression drops at sample 10, but is still present~~remains active~~ even in over ripe sample 12. According to~~In a Greek study~~ (~~(~~Haralampidis et al. (1998), *SAD1* transcription ~~gene activity~~ stayed at the maximum until 28 weeks after flowering, without any visible decrease~~drop~~ in expression. ~~They studied the expression profile by~~ In that study, Northern blot analysis was used, which in general ~~gives highly~~ correlate~~sd~~ well with ~~results by~~ RT-qPCR (Dean et al. 2002); however~~, but~~ the olive genotype and environmental effects ~~cannot be neglected since they~~ can explain~~contribute to the~~~~a~~ slightly different expression profile. The *FatA* gene expression (Fig. 4b) shows~~A~~ a pattern similar ~~distribution compared~~ to ~~the gene~~ *SAD1* ~~can be seen~~ in the first sampling points ~~of the olive *FatA* gene expression (Fig. 4b)~~; it then starts to increase in sample 4 and reaches the highest expression in~~at~~ sample 9. The expression remains high up to sample~~ing point~~ 11 and then significantly drops in ~~over ripe~~ sample 12.

Acyl-CoA thioesterase (*Acot*) hydrolyzes fatty acyl-CoAs to free fatty acids and coenzyme A, thus providing the potential for regulation of intracellular levels of acyl-CoAs, free fatty acids and coenzyme A. This family of ~~The groups of these~~ enzymes is~~are~~ suggested to have a possible role in fatty acid oxidation in animals. In plants, the first Acot gene - *ACH2* was

cloned from *Arabidopsis*. The gene was more highly expressed in mature tissues rather than in germinating seedlings, indicating that its role is probably not linked to fatty acid oxidation (Tilton et al. 2004). TExpression analysis of the *Acot* gene in olive shows varyingdifferential expression during at all stages of fruit development;, although the highest expression is during the end stage of pit hardening and the whole stage of mesocarp development (Fig. 4c, samples 6-10). The expression declined after the mesocarp developmental stage (samples 11 and 12). The observation of expression suggests that the particular olive transcript, this gene, just as with the *Arabidopis* Acot gene, is not primarily involved in the beta oxidation process.

The fourth characterized expression profile for lipoxygenase 1 gene (*LOX1*) (Fig. 4d, e) showed a very large expression peak in the over ripe sample (sample 12) (Fig. 4d), although expression fluctuations can also be seen in the other 11 olive fruit samples (Fig. 4e), with an increasing trend of transcript accumulation in samples 10 and 11 (mesocarp development and ripening). The *LOX1* enzyme plays a primary role in a series of enzymatic conversion processes in the lipoxygenase (LOX) pathway, in which some of the most abundant volatiles are formed from linoleic and linolenic acid. This pathway is particularly induced during crushing and the malaxation procedure of oil extraction (Conde et al. 2008). This is an important quality metabolic process for oil quality, since olive oils are characterized by their characteristic aroma, which is a complex mixture of volatile compounds (Morales et al. 1995). A recent qPCR study of *LOX1* in olive fruits showed an increasing trend of transcript accumulation towards the end of ripening in the 3 samples of two Italian cultivars (Muzzalupo et al. 2012). Our expression results of the olive *LOX1* gene support these findings, while very large *LOX1* accumulation was detected in over-ripe sample 12.

The correlation coefficients among standardized data obtained from values normalized with the best 9 genes and best two2--genes pair (*TBP/TIP41*) for the all four target genes, *FatA*, *SAD1*, *Acot*, and *LOX1* were 0.97, 0.97, 0.93, and 0.98, respectively, and these results confirm the suitability of the two RGsreference genes selected for normalization. Low correlations between normalized (two best–RGsreference genes) and non-normalized data (0.63, 0.23, 0.04, and 0.50, respectively) appear to be mostly due to reflect differences in the gene expression pattern. The presentation of non-normalized data of all four target genes showed significantly higher expression levels of all the metabolic genes in for two samples, 4 and 9 , in all four target genes, which, in turn, may be influenced by a different quantity/quality of RNA samples and by the rate of the reverse transcription step (Nolan et al. 2006). When the least stable gene *ADH1* was used for normalization, low correlation coefficients compared to the two best genes normalized data were also obtained (0.34, 0.29, 0.04, and 0.25)

highlighting that selection of RGs ~~reference genes~~ is necessary for proper evaluation of expression studies. Similarly, a significant difference was observed in the expression pattern of two olive genes, putative polygalacturonase (*PG*) and farnesyl pyrophosphate synthase (*FPS*), when the worst internal control was used for normalization (Nonis et al. 2012).

In conclusion, the presented study reports on a comprehensive analysis aimed at determining the optimal RGs~~reference genes~~ for the quantification of transcript levels during olive fruit development. We tested the expression stabilities of 27 selected candidate genes in olive fruits at 12 different sampling points of fruit ~~ripening~~ developmental ~~stages~~. The 27 genes were further used for normalization of expression data of four target genes involved in fatty acid metabolism. On the basis of the results, we recommend two RGs~~reference genes~~, *TIP41* and *TBP*, ~~which will enable more accurate and reliable normalization of qPCR results~~ for gene expression studies in olives.

Acknowledgment

The authors acknowledge funding by the Slovenian Research Agency, research project grant no. J4-2296 and the support of TR by grant number 1000-08-310185.

References

Andersen CL, Jensen JL, Orntoft TF (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 Res 64:5245-5250

Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG (2009) Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. BMC Mol Biol 10

Brunner A, Yakovlev I, Strauss S (2004) Validating internal controls for quantitative plant gene expression studies. BMC Plant Biol 4:14

Bruno L, Chiappetta A, Muzzalupo I, Gagliardi C, Iaria D, Bruno A, Greco M, Giannino D, Perri E, Bitonti MB (2009) Role of geranylgeranyl reductase gene in organ development and stress response in olive (*Olea europaea*) plants. Funct Plant Biol 36:370-381

Chaouachi M, Giancola S, Romaniuk M, Laval V, 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*). J Agric Food Chem 55:8003-8010

Cicerale S, Lucas L, Keast R (2010) Biological activities of phenolic compounds present in virgin olive oil. Int J Mol Sci 11: 458-479

Conde C, Agasse A, Silva P, Lemoine R, Delrot S, Tavares R, Geros H (2007) *OeMST2* encodes a monosaccharide transporter expressed throughout olive fruit maturation. Plant Cell Physiol 48:1299-1308

Conde C, Delrot S, Geros H (2008) Physiological, biochemical and molecular changes occurring during olive development and ripening. J Plant Physiol 165:1545-1562

Corpas FJ, Fernandez-Ocana A, Carreras A, Valderrama R, Luque F, Esteban FJ, Rodriguez-Serrano M, Chaki M, Pedrajas JR, Sandalio LM, del Rio LA, Barroso JB (2006) The expression of different superoxide dismutase forms is cell-type dependent in olive (*Olea europaea* L.) leaves. Plant Cell Physiol 47:984-994

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiol 139:5-17

Dean JD, Goodwin PH, 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 Mol Biol Rep 20:347-356

Exposito-Rodriguez M, Borges AA, Borges-Perez A, Perez JA (2008) Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. BMC Plant Biol 8

Gachon C, Mingam A, Charrier B (2004) Real-time PCR: what relevance to plant studies? J Exp Bot 55:1445-1454

Galla G, Barcaccia G, Ramina A, Collani S, Alagna F, Baldoni L, Cultrera NGM, Martinelli F, Sebastiani L, Tonutti P (2009) Computational annotation of genes differentially expressed along olive fruit development. BMC Plant Biol 9

Ginzinger DG (2002) Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. Exp Hematol 30:503-512

Gutierrez L, Mauriat M, Guenin S, Pelloux J, Lefebvre JF, Louvet R, Rusterucci C, Moritz T, Guerineau F, Bellini C, Van Wuytswinkel O (2008) The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnol J 6:609-618

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. J Exp Bot 49:1661-1669

Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. Genome Res 6:986-994

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 Biol 8

Hernandez ML, Padilla MN, Sicardo MD, Mancha M, Martinez-Rivas JM (2011) Effect of different environmental stresses on the expression of oleate desaturase genes and fatty acid composition in olive fruit. Phytochemistry 72:178-187

Hoerndli FJ, Toigo M, Schild A, Gotz J, Day PJ (2004) Reference genes identified in SH-SY5Y cells using custom-made gene arrays with validation by quantitative polymerase chain reaction. Anal Biochem 335:30-41

Kumar V, Sharma R, Trivedi PC, Vyas GK, Khandelwal V (2011) Traditional and novel references towards systematic normalization of qRT-PCR data in plants. Aust J Crop Sci 5:1455-1468

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658-1659

McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20-W25

~~Menendez JA, Lupu R (2007) Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. Nat Rev Cancer 7:763-777~~

Morales MT, Alonso MV, Rios JJ, Aparicio R (1995) Virgin olive oil aroma - relationship between volatile compounds and sensory attributes by chemometrics. J Agric Food Chem 43:2925-2931

Muzzalupo I, Stefanizzi F, Perri E, Chiappetta AA (2011) Transcript levels of CHL P gene, antioxidants and chlorophylls contents in olive (*Olea europaea* L.) pericarps: A comparative study on eleven olive cultivars harvested in two ripening stages. Plant Foods Hum Nutr 66:1-10

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, and technological treatments for oil extraction. Sci World J, article ID 532179, 9 pages

Nicot N, Hausman JF, Hoffmann L, Evers D (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot 56:2907-2914

Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. Nat Protoc 1:1559-1582

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. J Agric Food Chem 60:6855-6865

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. Biotechnol Lett 26:509-515

Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A (2004) Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun 313:856-862

Ransbotyn V, Reusch TBH (2006) Housekeeping gene selection for quantitative real-time PCR assays in the seagrass *Zostera marina* subjected to heat stress. Limnol Oceanograph Meth 4:367-373

Reid KE, Olsson N, Schlosser J, Peng F, Lund ST (2006) An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol 6

Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, Cuypers A (2008) Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. Planta 227:1343-1349

Ryan D, Robards K, Lavee S (1999) Changes in phenolic content of olive during maturation. Int J Food Sci Technol 34:265-274

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

Schmittgen TD, Zakrajsek BA (2000) Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods 46:69-81

Tilton GB, Shockey JM, Browse J (2004) Biochemical and molecular characterization of ACH2, an acyl-CoA thioesterase from *Arabidopsis thaliana*. J Biol Chem 279:7487-7494

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 Biol 3

Xie FL, Sun GL, Stiller JW, Zhang BH (2011) Genome-wide functional analysis of the cotton transcriptome by creating an integrated EST database. Plos One 6

19/01/2013

Dear Editor,

Please find the responses to reviewers and description of changes we have made to the manuscript »Validation of candidate reference genes in qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids metabolism«

1) Reviewer #1

Reviewer #1 did not suggest any changes to the manuscript.

2) Reviwer #2

We want to thank the anonymous reviewer #2 for the constructive review and improvement of the manuscript. We will answer all the questions from the manuscript and then questions from the cover letter.

2A) Reviewer #2 - questions in the manuscript

1) page 5, section »RNA isolation and cDNA preparation«, paragraph 1; Question »What does 'smeared tracks' mean? RNA bands with smear? Why not showing a picture?«, related to sentence: »RNA quality was regularly inspected by the integrity of ribosomal RNA bands on 1.2% formaldehyde agarose gels, which showed sharp and intense 18S and 28S ribosomal RNA bands, with a small number of smeared tracks, confirming the suitability of the isolation method.«

Answer:
Smeared RNA tracks in electrophoresis are a common feature from RNA degradation, when ribosomal bands show partial or complete degradation and you obtain a subset of RNA molecules of different lengths. With the isolation procedure used in our experiment, we did not observe problems related to RNA degradation or it appeared only in a few examples. If we observed any degradation, the isolation of the RNA was repeated. We decided not to show a picture of the RNA gel electrophoresis in the revised manuscript, since the procedure is quite common molecular biology practice and due to the limitation on Tables/Figures.

However, we changed the sentence to:
RNA quality was regularly inspected by the integrity of ribosomal RNA bands on 1.2% formaldehyde agarose gels, which showed sharp and intense 18S and 28S ribosomal RNA bands, without visible degradation, confirming the suitability of the isolation method.

2) page 6, section »Reference gene selection and primer design«, paragraph 2; Question: »(Resetic, unpublished) describe this sequence resource better.«, related to sentence: »We compared 7,080 olive mRNA or DNA sequences downloaded from GenBank and 212,795 olive 454 EST clusters of sequences against the 29 reference genes sequences (DNA and protein) from other plant species (Table 1) using BLASTN, BLASTX and TBLASTX algorithms (McGinnis and Madden 2004).

Answer:
The 454 sequences originate from our olive transcriptome assembly project, which is not yet published. The full set of raw sequences consists of 560,578 reads totaling 160.4 Mb. For the RT-qPCR part, the sequences were clustered using CD-HIT software into a non-redundant set of 212,795 olive sequences, which, together with NCBI sequences, were compared to the selection of reference genes.
For a better description of the source we have done the following:

a) We have submitted and released the entire raw transcriptome data in NCBI's SRA archive under accession number SRX215662 (http://www.ncbi.nlm.nih.gov/sra/SRX215662)
b) We have changed the sentence in question and inserted an additional sentence in order to describe the source better. We have also introduced an additional reference, which is now listed in the reference section:

We used 7,080 olive mRNA or DNA sequences downloaded from GenBank and 212,795 olive 454 EST clusters of sequences (Resetic, unpublished). The EST clusters originate from 454 olive transcriptome data available as raw reads in NCBI SRA archive (SRX215662), which were clustered using CD-HIT software (Li and Godzik, 2006) to decrease redundancy. Sequences were compared against the 29 RG…

Reference was added to reference list:
Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658-1659

3) page 7, section "Two-step real time PCR analysis and quantification", paragraph 1 and 2; Question: "The final assay included a standard curve of six serial dilution points for the reference genes, genes of interest six dilutions also for the genes of interest? and twelve cDNA samples from different stages of olive fruit development with 4 dilutions? Not clear, see below."…additional text… "The cDNA samples were used as PCR templates in a range of 4-fold series dilutions starting with 50 and ending with 0.05 ng. Please specify the 4 concentrations."

Answer:
We believe that the text properly describes the standard curve experiment – we used 6 dilutions of the templates and they were prepared by a dilution series from 50 ng to 0.05 ng, which were diluted 4-fold (50, 12.5, 3.13, 0.78, 0.20 and 0.05 ng).

For better understanding, we have changed the text accordingly.
Before:
The final assay included a standard curve of six serial dilution points for the reference genes, genes of interest and twelve cDNA samples from different stages of olive fruit development. Variation between runs was minimized by performing all of the reactions containing a single primer pair on the same plate and by including a standard curve on each plate. The cDNA samples were used as PCR templates in a range of 4-fold series dilutions starting with 50 and ending with 0.05 ng.

After:

The final assay included a standard curve of six serial dilution points for the RGs, genes of interest and twelve cDNA samples from different stages of olive fruit development. The cDNA samples were used as PCR templates in a range of six dilutions made in 4-fold decrements starting with 50 and ending with 0.05 ng (50, 12.5, 3.13, 0.78, 0.20 and 0.05 ng). Variation between runs was minimized by performing all of the reactions with a single primer pair on the same plate and by including a standard curve on each plate.

4) page 8, section "Two-step real time PCR analysis and quantification", paragraph 2; Question: "How is NRQ calculated?", related to sentence "After we had identified the two most stable reference genes, we used them for calculating the normalized relative quantities (NRQ) of four lipid metabolism target genes."

Answer:
NRQ is calculated by division of the average quantity with the normalization factor. The average quantity is obtained from quantity repetitions and delivered by a Real Time program, SDS in our case. The normalization factor for the two best reference genes is calculated by the geNorm program (this is geomean of quantities of individual reference genes. Since it is a common procedure in RT-qPCR experiments, it has not been included in the manuscript.

5) page 11, section "Discussion", paragraph 1; Questions: "We also tested two other genes selected from the GMO reference gene list, *Pkaba1* and *UGPase*, never reported as having been used in qPCR normalization studies. *UGPase* proved to be a less favorable gene in our experiment, ranked as the 17th most stable gene according to which algorithm?, while *Pkaba1* was among the most stable genes in our study, in 3rd place according to which algorithm?.

Answer:
According to the geNorm algorithm in both cases, which we added to the end of the sentence. The reviewer also suggested changes in the sentence, which we have incorporated:
"We also tested two other genes selected from the GMO RG list, *Pkaba1* and *UGPase*, never reported as having been used in RT-qPCR normalization studies. *UGPase* ranked as the 17th most stable gene, while *Pkaba1* was among the most stable genes in our study, in 3rd place, in both cases according to the geNorm algorithm.

6) page 12, section "Discussion", paragraph 2; Question: "It is interesting that one of the GAPDH primer pairs (GAPDH1) in their study was also identified as the worst reference gene, with an M value of 0.609. In our study, the GAPDH gene ranked in the second half of analyzed genes, in 19th according to which algorithm?, position with an M value of 0.68 (Fig. 2a).

Answer:
According to geNorm algorithm, as for 6) the changed sentence in the manuscript is:
It is interesting that one of the GAPDH primer pairs (GAPDH1) in their study was identified as the worst RG, with an M value of 0.609. In our study, the GAPDH gene ranked 19th place according to geNorm, with an M value of 0.68 (Fig. 2a).

7) page 12, section "Discussion", paragraph 3; Question: "In our case, the olive 18S reference gene had a stability value of 0.72, which is over the suggested threshold of 0.5 and was ranked in 22nd place according to which algorithm?.

Answer:
According to the geNorm algorithm; as for 6) the changed sentence in the manuscript is:
In our case, olive *18S* RG had a stability value of 0.72, which is above the suggested threshold of 0.5 and ranked 22nd according to geNorm.

8) page 12, section "Discussion", paragraph 3; Question: "It was also shown that quantification of the expression of which gene? can be underestimated, as in the case of a potato experiment (Nicot et al. 2005)."

Answer:
The particular reference showed that quantification of target genes can be underestimated when 18S rRNA RG is used. We have therefore changed the sentence, also following the reviewer's recommendations on sentence shortening:
It has also been shown that expression of target genes can be underestimated (Nicot et al. 2005).

9) page 13, section "Discussion", paragraph 1; Question: "They showed that the enzyme or the mRNA?? is already present in small drupes, embryos and endosperm."

Answer:
Yes, the authors (Haralampidis et al., 1998) worked with mRNA. The mistake was corrected accordingly:
"They showed that the mRNA is already present in small drupes, embryos and endosperm."

10) page 13, section "Discussion", paragraph 1; Question: "In a Greek study (Haralampidis et al. 1998), gene activity stayed at the maximum until the end end of what? Maturation?, without any visible drop in expression."

Answer:
Gene activity stayed at the maximum until the end of the sampling period, which in their study was 28 weeks after flowering; in the Greek environment this corresponds to the end of fruit development. The sentence has been corrected accordingly, also taking into account the reviewer's grammatical suggestion:
"According to Haralampidis et al. (1998), *SAD1* transcription stayed at the maximum until 28 weeks after flowering, without any visible decrease in expression."

11) page 14, section "Discussion", paragraph 3; Question/comment: "The correlation coefficients among standardized data obtained from values normalized with the best 9 genes and best 2 genes

pair (TBP/TIP41) for all four target genes, FatA, SAD1, Acot, and LOX1 were 0.97, 0.97, 0.93, and 0.98, respectively, and these results confirm the suitability of the two reference genes selected for normalization. Low correlations between normalized (two best reference genes) and non-normalized data (0.63, 0.23, 0.04, and 0.50) reflect differences in the gene expression pattern. The presentation of non-normalized data of all four target genes showed significantly higher expression levels for two samples, 4 and 9 in all four target genes, which may be influenced by a different quantity/quality of RNA samples and by the rate of the reverse transcription step (Nolan et al. 2006). If quality of the RNA sample is not uniform, then all the study looses strength. High RNA quality is obligatory

Answer:
We think that this is the reviewer's general comment to this paragraph of the manuscript. Isolation of high quality RNA was confirmed in our case. No action was taken here, but the whole paragraph has been rewritten, as suggested by the reviewer.

12) page 15, section "Discussion", paragraph 2; Question/comment: "When the least stable gene ADH1 was used for normalization, low correlation coefficients compared to the two best genes normalized data were also obtained (0.34, 0.29, 0.04, and 0.25) highlighting that selection of reference genes is necessary for proper evaluation of expression studies. Similarly, a significant difference was observed in the expression pattern of two olive genes, putative polygalacturonase (*PG*) and farnesyl pyrophosphate synthase (*FPS*), when the worst internal control was used for normalization (Nonis et al. 2012). I do not see the utility of this comparison, It seems obvious to me.

Answer:
The paragraph in question is comparing our data with data from the literature – the first study of reference genes selection in olive (Nonis et al. 2012). We feel this is an important comparison and would like to keep the paragraph in the manuscript.

2B) Questions in the cover letter

1) For the amount of new information provided, the paper is way too long. The style is redundant with many unnecessary repetitions. The English form is not satisfactory, either. I recommend to shorten and simplify it as shown in the edited manuscript.

Answer:
We have substantially shortened the manuscript. The main text previously numbered 6,572 words, or 36,326 characters without spaces, and now numbers 5,793 words (11.8% reduction) or 31,666 characters without spaces (12.8% reduction). We have followed and accepted the majority of the recommendations from Reviewer#2 and removed redundancy. The paper has again been checked by a professional English proofreading service.

2) The term Reference genes(s) occurs so many times that it should be abbreviated as RG(s).

Answer:
Yes, we have inserted the suggested abbreviations in the manuscript – RG for reference gene and RGs for reference genes throughout.

3) The citation Menendez and Lupu 2007 appears not pertinent. I have no access to the full paper, but in the abstract no reference at all is made to plants, only to fatty acid synthase role in cancer: please check.

Answer:
That was our mistake, the Mendez and Lupu 2007 reference is not related to olive polyphenols. The correct reference is Cicerale et al. (2010). The manuscript and reference list have been corrected accordingly.

Cicerale S, Lucas L, Keast R (2010) Biological activities of phenolic compounds present in virgin olive oil. Int J Mol Sci 11: 458-479

4) Materials and methods: Some clarifications are necessary (see highlighted notes). In particular, the 454 EST sequence resource should be described with some detail.

Answer:
Please see 2A) question 2) (same question).

5) Results: Not clear why three subsets of the data were created and analyzed separately: a justification is needed, otherwise it could seem a means of inflating the results.

Answer:
The systematic validation of candidate genes based on different time series resulted in different ranking of the genes based on stability values (Fig. 2). The different ranking of the candidate genes in different sample subsets emphasizes the need for detailed reference gene analysis, which is in accordance with previous studies reporting inter-tissue and experiment-dependent variation in gene stability (Gutierrez et al. 2008) and not just variation among specific plant species. Systematic validation and the use of at least two validated reference genes involved in distinct cellular functions is therefore proposed since no gene can act as a universal reference (Gutierrez et al. 2008).

We have inserted a sentence in the text manuscript accordingly, p. 9:
The different ranking of candidate genes in different sample subsets emphasizes the need for detailed reference gene analysis, which is in accordance with previous studies reporting inter-tissue and experiment-dependent variation in gene stability (Gutierrez et al. 2008)

6) Results: The use of the worst-performing candidate for normalization, ADH1, appears not justified, either, and the results of such comparison are obvious.

Answer:
Please see 2A) question 12) (same question).

7) Discussion: The data are much overdiscussed. I summarized the discussion (see edited manuscript).

Answer:
Thank you for the substantial improvement, we have followed the summarizing suggestions by the reviewer and substantially reduced the text (please see 2B) question 1).

8) Notes to Table 2: 454: sequence from 454 Sequence GenBank number accession provided or sequence originating from the 454 library

Answer:
The number 454 means sequences obtained from our 454 experiment, which is not yet published yet, although the data are now available through NCBI. Please see also 2A) question 2).

9) Supplementary Table 1: is it necessary? The primer sequences and amplicon lengths could be included in table 1. Change notes as follows: a: Amplification not achieved, b: (remove)

Answer:
Reviewer#3 suggested presenting efficiencies in this table, so we would like to keep it. We think that the introduction of another column of data and combining it with Table 1 would be too large for presentation.
The notes have been changed as requested.

3) Reviwer #3

Reviewer #3 listed his questions and suggestions for improvement in the cover letter.

1) The performance/suitability of a reference gene is highly dependent on the experimental conditions, and even then on the experiment. Some evidence for this is also given in the paper, page 11: the study by Nonis and this study do not agree completely.
Therefore, I would like to advise to strictly stick in the conclusions and in the abstract to the evaluation of the genes in this experimental context, and to offer these genes as POTENTIALLY GOOD candidates for normalisation of gene expression data in future similar experiments. The authors suggest that they have given "the best two reference genes for use in following experiments", but that is dangerous as in other experiments the stability of the genes can be completely different. I would advise to remove such sentences that put the two genes forward as ideal reference genes for future experiments from the manuscript, and to focus on the fact that this study is valuable in offering an excellent choice of pre-evaluated candidates for other experiments.

Answer:
Yes, we agree with the suggestion. Reviewer #2 in his redrafting has already pointed out the same and changed the sentences that put the two genes forward as ideal reference genes. We have therefore these suggestions and have changed the term “the best two reference genes” throughout

the paper. When we are describing the outcome of the geNorm alghoritm, we kept the term best two genes.

2) The results start with "twelve different developmental stages" whereas the M&M states "five distinct stages of fruit development". This has to be clarified. Also, where whole fruits harvested?

Answer:
Olive fruit development is described by five distinct developmental stages (from fruit set until maturity), described in a paper by Ryan et al. 1999. We covered these five stages with 12 sampling points. We have therefore changed the text in the manuscript:
Before:
“Total RNA was isolated from olive fruits from twelve different development stages of olive fruit ripening.” (p6, “Validation of reference genes”)

To:
“Total RNA was isolated from olive fruits at twelve different sampling points of olive fruit development and ripening.”

We have additionally corrected the abstract. We have replaced:
“The analyzed data points represented 12 distinct olive fruit developmental stages.”
with
“We used 12 sampling points to cover the five stages of olive fruit development.”

And in the last paragraph of Discussion, changed:
“We tested the expression stabilities of 27 selected candidate genes in olive fruits at 12 different fruit developmental stages.”
to
“We tested the expression stabilities of 27 selected candidate genes in olive fruits at 12 different sampling points of fruit development.”

Yes, we harvested the whole fruits.

3) RT-qPCR is reverse transcription quantitative PCR (real time and quantitative is the same thing)

Answer:
Thank you for pointing this out. The terms quantitative real-time polymerase chain reaction, qPCR or similar have been changed into reverse transcription quantitative polymerase chain reaction (RT-qPCR) and used throughout the paper.
The title of the paper has also been changed to “Validation of candidate reference genes in RT-qPCR studies of developing olive fruit and expression analysis of four genes involved in fatty acids metabolism”. This was also suggested by Reviewer#2.

4) Was there a DNase treatment? Is there evidence of DNA presence, or can you somehow prove the absence of DNA contamination.

Answer:
Yes, we followed the on-column DNase digestion protocol as implemented in the total RNA kit by Sigma-Aldrich. However, we tested several PCR primers (ITS, microsatellites) and ran PCR with input RNA to check for amplification, which was not achieved.

5) How was [cDNA] determined?

Answer:
cDNA was not determined but we assumed 1:1 conversion from RNA. Such an approach is common in relative expression experiments, where we are not calculating the copy numbers. We have therefore changed the text in the manuscript, "Materials and methods" section:

Before:
A master mix for each PCR run was prepared using a 20 µl reaction volume containing 10 µl of FAST SYBR Green PCR Master Mix (Applied Biosystems), 10 ng of cDNA and 300 nM of each specific sense and anti-sense primer on MicroAmp Optical 96 well PCR plates (Applied Biosystems).

To:
A master mix for each PCR run was prepared using a 20 µl reaction volume containing 10 µl of FAST SYBR Green PCR Master Mix (Applied Biosystems), 2 µl of cDNA (corresponding to 10 ng of total RNA) and 300 nM of each specific sense and anti-sense primer on MicroAmp Optical 96 well PCR plates (Applied Biosystems).

6) The efficiency of the primer pairs could be listed as an extra column in suppl. table 1

Answer:
A new column with efficiencies has been introduced in Supplementary Table 1.

7) C_q (quantification cycle) should be used instead of Ct

Answer:
We have changed all C_t in manuscript, figures and tables to C_q. We have also changed the "threshold cycle" to "quantification cycle".

4) Comments for the author from editor

1) The ms will be reconsidered if authors will be willing to receive and acquire all suggestions from the reviewers, including the attached material as provided by reviewer #2.

Answer:
Please find the detailed responses to the reviewers; we have followed all the manuscript changes as suggested by reviewer#2 (pdf file named paper_corrections.pdf is attached to sow the correctons).

2) Please also consider that the total number of figures + tables should be a maximum of 5 and additional items should be placed as electronic supplementary material.

Answer:
The manuscript had 2 tables, 4 figures and 1 Supplementary table. We suggest including Table 2 as Supplementary Table 2.