

UNIVERZA V LJUBLJANI  
BIOTEHNIŠKA FAKULTETA  
ŠTUDIJ BIOTEHNOLOGIJE

Lidija BERKE

**BIOINFORMACIJSKA ANALIZA OZNAK  
IZRAŽENIH ZAPOREDIJ (EST) HMELJA (*Humulus  
lupulus L.*) IN NJIHOVA APLIKACIJA PRI RAZVOJU  
MOLEKULARNIH MARKERJEV**

DIPLOMSKO DELO

Univerzitetni študij

Ljubljana, 2008

UNIVERZA V LJUBLJANI  
BIOTEHNIŠKA FAKULTETA  
ŠTUDIJ BIOTEHNOLOGIJE

Lidija BERKE

**BIOINFORMACIJSKA ANALIZA OZNAK IZRAŽENIH ZAPOREDIJ  
(EST) HMELJA (*Humulus lupulus L.*) IN NJIHOVA APLIKACIJA PRI  
RAZVOJU MOLEKULARNIH MARKERJEV**

DIPLOMSKO DELO  
Univerzitetni študij

**BIOINFORMATIC ANALYSIS OF EXPRESSED SEQUENCE TAGS  
(ESTs) IN HOP (*Humulus lupulus L.*) AND THEIR APPLICATION IN  
MOLECULAR MARKER DEVELOPMENT**

GRADUATION THESIS  
University studies

Ljubljana, 2008

Diplomsko delo je zaključek Univerzitetnega študija biotehnologije. Opravljeno je bilo na Katedri za genetiko, biotehnologijo in žlahtnjenje rastlin Oddelka za agronomijo Biotehniške fakultete Univerze v Ljubljani.

Po sklepu Komisije za dodiplomski študij oddelka za biotehnologijo z dne 2. 7. 2008 je bil za mentorja diplomskega dela imenovan doc. dr. Jernej Jakše.

Komisija za oceno in zagovor:

Predsednik:                   prof. dr. Branka Javornik  
                                       Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za agronomijo

Član:                           doc. dr. Jernej Jakše  
                                       Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za agronomijo

Član:                           izr. prof. dr. Gregor Anderluh  
                                       Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo

Datum zagovora:

Delo je rezultat lastnega raziskovalnega dela. Podpisana se strinjam z objavo svoje naloge v polnem tekstu na spletni strani Digitalne knjižnice Biotehniške fakultete. Izjavljam, da je naloga, ki sem jo oddala v elektronski obliki, identična tiskani verziji.

Lidija Berke

## KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD	Dn
DK	UDK 633.791: 577.2 (0.43.2)
KG	genetika rastlin / gensko kartiranje / hmelj / <i>Humulus lupulus</i> / alfa kisline / molekularni markerji / EST / SNP / SSR / halkon sintaza / valerofenon sintaza
KK	AGRIS F30
AV	BERKE, Lidija
SA	JAKŠE, Jernej (mentor)
KZ	SI-1000 Ljubljana, Jamnikarjeva 101
ZA	Univerza v Ljubljani, Biotehniška fakulteta, Medoddelčni študij biotehnologije
LI	2008
IN	BIOINFORMACIJSKA ANALIZA OZNAK IZRAŽENIH ZAPOREDIJ (EST) HTELJA ( <i>Humulus lupulus</i> L.) IN NJIHOVA APLIKACIJA PRI RAZVOJU MOLEKULARNIH MARKERJEV
TD	Diplomsko delo (univerzitetni študij)
OP	XII, 28 [5] str., 5 pregl., 10 sl., 2 pril., 54 vir.
IJ	sl
JI	sl/en
AI	14944 hmeljnih EST-jem, ki so večinoma izhajali iz trihomov, smo odstranili vektorska zaporedja, odstanili prekratka zaporedja in zaporedja nizke kakovosti ter preostanek zbrali v gruče. Dobili smo 1318 gruč, ostalo pa je 76 posameznih zaporedij. Pregledali smo vsebnost mikrosatelitov v nastalih soseskah in nekaj največjih sosesk primerjali z zaporedji v podatkovni zbirki GenBank. Sosesko, ki je ustrezala genu za valerofenon sintazo (VPS), smo primerjali z zaporedjem za gen, da smo locirali intron. Izdelali smo začetne oligonukleotide, katerih produkti so skupaj pokrivali celoten gen. Za izdelavo začetnih oligonukleotidov za halkon sintazo (CHS) smo uporabili zaporedja, ki so se nahajala v GenBanku. Reakcije PCR za oba gena smo optimizirali, izbranim produktom določili nukleotidno zaporedje in poiskali domnevne SNP-e. Na podlagi najdenih SNP-ov in njihove segregacije pri hmeljni družini Magnum × 2/1 smo gena kartirali na skupino povezanosti Magnum-8, kar pomeni, da ne gen za VPS ne gen za CHS ne ležita na QTL-u za vsebnost α-kislin. Domnevne SNP-e, katere smo našli z določanjem nukleotidnega zaporedja PCR produktov gena za VPS, smo primerjali z EST-ji v gruči za VPS in ugotovili, da je več kot polovico SNP-ov mogoče določiti že iz poravnav EST-jev v gruče, kar pospeši in poceni kartiranje.

#### KEY WORDS DOCUMENTATION

DN Dn  
DD UDC 633.791: 577.2 (0.43.2)  
CX plant genetics / genetic mapping / hop / *Humulus lupulus* / alpha acids / molecular markers / EST / SNP / SSR / chalcone synthase / valerophenone synthase  
CC AGRIS F30  
AU BERKE, Lidija  
AA JAKŠE, Jernej (supervisor)  
PP SI-1000 Ljubljana, Jamnikarjeva 101  
PB University of Ljubljana, Biotechnical Faculty, Study Programme Biotechnology  
PY 2008  
TI BIOINFORMATIC ANALYSIS OF EXPRESSED SEQUENCE TAGS (ESTs)  
IN HOP (*Humulus lupulus* L.) AND THEIR APPLICATION IN MOLECULAR  
MARKER DEVELOPMENT  
DT Graduation thesis (university studies)  
NO XII, 28 [5] p., 5 tab., 10 fig., 2 ann., 54 ref.  
LA sl  
AL sl / en  
AB 14944 hop ESTs, mostly from trichomes, have been collected. The vector sequences have been clipped and short as well as low-quality sequences have been removed. Finally, the remaining sequences were clustered. 1318 clusters and 76 singletons were obtained; some of the former were consequently compared to sequences in GenBank. Also, microsatellite content of contigs was analysed. The contig that matched GenBank entry for valerophenone synthase gene (VPS) was aligned with the particular gene sequence in order to locate the intron. Primers were made whose products covered the length of the entire gene. The primers for chalcone synthase (CHS) gene were constructed using the sequences already present in GenBank. After the PCR conditions were optimized, the chosen PCR products were sequenced and checked for SNPs. Based on the discovered SNPs and their segregation in the hop family Magnum × 2/1, genes for CHS and VPS were mapped to linkage group Magnum-8. Consequently, neither gene for CHS nor gene for VPS are associated with any of the QTLs for alpha-acid content. Putative SNPs which were found by sequencing our PCR products were compared to ESTs in VPS cluster. We noted that for more than half of the SNPs that we found with sequencing, there was indication in EST cluster.

## KAZALO VSEBINE

	str.
Ključna dokumentacijska informacija	III
Key words documentation	IV
Kazalo vsebine	V
Kazalo preglednic	VI
Kazalo slik	VII
Kazalo prilog	VIII
Okrajšave in simboli	IX
Slovarček	XI
<b>1 UVOD</b>	1
<b>2 PREGLED OBJAV</b>	2
2.1 OZNAKE IZRAŽENIH ZAPOREDIJ (EST)	2
2.2 HALKON SINTAZA (CHS) IN VALEROFENON SINTAZA (VPS)	3
2.3 MIKROSATELITI IN KARTIRANJE HMELJA	6
<b>3 MATERIAL IN METODE</b>	7
3.1 BIOINFORMACIJSKI DEL	7
<b>3.1.1 Urejanje zaporedij EST in zbiranje v gruče (angl. clustering)</b>	7
<b>3.1.2 Analiza zaporedij EST s skripto MISA</b>	7
3.2 RASTLINSKI MATERIAL	8
3.3 NAČRTOVANJE OLIGONUKLEOTIDNIH ZAČETNIKOV IN PCR VPS	8
3.4 NAČRTOVANJE OLIGONUKLEOTIDNIH ZAČETNIKOV IN PCR CHS	10
3.5 DOLOČANJE NUKLEOTIDNEGA ZAPOREDJA	11
<b>3.5.1 Izolacija produktov PCR iz agaroznega gela (CHS)</b>	11
<b>3.5.2 Neposredno čiščenje produktov PCR (VPS)</b>	11
<b>3.5.3 Reakcija za določanje nukleotidnega zaporedja</b>	11
<b>3.5.4 Čiščenje reakcije za določanje nukleotidnega zaporedja in določevanje nukleotidnega zaporedja</b>	12
<b>3.5.5 Obdelava rezultatov določanja nukleotidnega zaporedja</b>	12
3.6 KARTIRANJE GENOV ZA CHS IN VPS	12
<b>4 REZULTATI</b>	13
4.1 BIOINFORMACIJSKI DEL	13
<b>4.1.1 Rezultati urejanja in zbiranja v gruče</b>	13
<b>4.1.2 Rezultati skripte MISA</b>	14
<b>4.1.3 Primerjava najdenih SNP-ov</b>	14
4.2 PCR IN DOLOČANJE NUKLEOTIDNEGA ZAPOREDJA VPS	15
4.3 PCR IN DOLOČANJE NUKLEOTIDNEGA ZAPOREDJA CHS	15
4.4 REZULTATI KARTIRANJA	18
<b>5 RAZPRAVA IN SKLEPI</b>	21
5.1 EST	21
5.2 MIKROSATELITI	21
5.3 UPORABA SNP	21
5.4 KARTIRANJE GENOV ZA CHS IN VPS	21
<b>6 POVZETEK</b>	22

7	<b>VIRI</b> <b>ZAHVALA</b> Priloga A Priloga B	24
---	---------------------------------------------------------	----

## KAZALO PREGLEDNIC

	str.	
Preglednica 1:	Definicija mikrosatelita za iskanje po soseskah.	7
Preglednica 2:	Lastnosti sedmih parov začetnih oligonukleotidov za pomnoževanje delov gena VPS.	9
Preglednica 3:	Lastnosti treh izbranih parov začetnih oligonukleotidov za pomnoževanje gena CHS.	10
Preglednica 4:	Test $\chi^2$ za segregacijo obeh analiziranih genov pri družini Magnum $\times$ 2/1.	19
Preglednica 5:	Genetske razdalje med označevalci v skupini povezanosti 8 (Magnum-8). Ostali markerji iz te skupine izvirajo iz dela Čerenak in sod. (2006).	19

## KAZALO SLIK

	str.
Slika 1:	4
Ženski socvetji hmelja (»storžka«), ob osnovnem vretenu so lepo vidne lupulinske žleze (rumeno).	
Slika 2:	5
Dve izmed reakcij, ki jih katalizirajo rastlinske poliketid sintaze (Paniego in sod., 1999).	
Slika 3:	6
Shema strukture gena za VPS in položajev oligonukleotidnih začetnikov.	
Slika 4:	13
Vzroki odstranitve skupno 213 EST zaporedij pred nadaljnjam delom.	
Slika 5:	14
Primer rezultata zbiranja v gruče, prikazan je CL4Contig2, ki je transkript gena za VPS. Vsaka vodoravna črta ponazarja en EST v gruči.	
Slika 6:	15
Gruča EST-jev, katerih soseska predstavlja gen za VPS. Označena so tri polimorfna mesta, ki smo jih pred tem potrdili v naših produktih PCR.	
Slika 7:	16
Slika agaroznega gela s produkti PCR. Zgoraj desno: produkt PCR z začetnima oligonukleotidoma P33 in P34; spodaj levo produkt PCR z začetnima oligonukleotidoma P35 in P36, spodaj desno pa s P37 in P38.	
Slika 8:	17
Del grafičnega prikaza rezultatov določanja nukleotidnega zaporedja produkta PCR dela gena za CHS z začetnim oligonukleotidom P34. Označen je SNP – pri Magnumu (zgornja vrsta) sta vidna dva vrhova (moder in rdeč) za citozin in timin, pri 2/1 pa je viden le en rdeč vrh za timin. Desno od SNP-a je viden tudi začetek mikrosatelita GTA.	
Slika 9:	18
Primer analize potomcev polimorfnegra mesta gena <i>chs_H1</i> iz slike 8. Viden je tudi celoten trinukleotiden mikrosatelit GTA in porušenje zaporedja desno od njega pri heterozigotnih rastlinah (1, 3, 4 in 7).	
Slika 10:	20
Skupina povezanosti Magnum-8. Na levi strani je karta pred (Čerenak in sod., 2006), na desni pa po dodatku novih markerjev.	

## KAZALO PRILOG

- Priloga A: Frekvenca tipov mikrosatelitov v analiziranih sosekah, upoštevajoč komplementarnost zaporedij.
- Priloga B: Genotipske vrednosti lokusa CHS in VPS za družino Magnum × 2/1, ki služijo kot vhodne vrednosti za program JoinMap.  
Vrednost lm pomeni, da je rastlina heterozigot, vrednost ll pa, da je homozigot.

## OKRAJŠAVE IN SIMBOLI

AFLP	polimorfizem dolžine pomnoženih restrikcijskih fragmentov (angl. amplified fragment length polymorphism)
cDNA	komplementarna DNA (angl. complementary DNA)
dH <sub>2</sub> O	deionizirana voda
dNTP	deoksi nukleotid trifosfat
DNA	deoskiribonukleinska kislina
CHS	halkon sintaza (angl. chalcone synthase)
EDTA	etilendiamintetraocetna kislina
EST	oznaka izraženega zaporedja (angl. expressed sequence tag)
ISSR	medmikrosatelitne regije (angl. inter-simple sequence repeat)
LOD	logaritem razmerja obetov (angl. logarithm of odds)
MAS	žlahtnjenje s pomočjo markerjev (angl. marker assisted selection)
mRNA	informacijska RNA (angl. messenger RNA)
MgCl <sub>2</sub>	magnezijev klorid
PCR	verižna reakcija s polimerazo (angl. polymerase chain reaction)
QTL	kvantitativni lokus (angl. quantitative trait loci)
RAPD	tehnika naključno pomnožene DNA (angl. random amplified polymorphic DNA)
RFLP	polimorfizem dolžine restrikcijskih fragmentov (angl. restriction fragment length polymorphism)
SNP	polimorfizem posameznih nukleotidov (angl. single nucleotide polymorphism)
SSR	mikrosatelit (angl. microsatellite, single sequence repeat)
STS	sekvenčno označeno mesto (angl. sequence-tagged site)
polimeraza	encim DNA polimeraza, izolirana iz organizma <i>Thermus aquaticus</i>
<i>Taq</i>	

TBE pufer	tris-boratni-EDTA elektroforetski pufer
TE pufer	tris-EDTA pufer (10 mM Tris-HCl, 1 mM EDTA; pH 8,0)
$T_m$	talilna temperatura
UTR	neprevedljiva regija (angl. untranslated region)
VPS	valerofenon sintaza (angl. valerophenone synthase)

## SLOVARČEK

adapter	Kratka, <i>in vitro</i> sintetizirana dvoverižna DNA, ki jo dobimo iz dveh enoverižnih komplementarnih oligonukleotidov, s katero vnesemo na konce molekul DNA znano zaporedje. Omogoča nadaljnjo manipulacijo molekul DNA.
cDNA	DNA, sintetizirana iz zrele mRNA
ekson	Regija gena, ki se prepiše in prevede v protein.
homologni geni	Skupina genov s skupnim izvorom in visoko podobnostjo zaporedij.
Indel	Mutacija, pri kateri se doda ali odstrani en ali več baznih parov.
intron	Nekodirajoči del znotraj gena, še vedno prisoten v pre-mRNA molekuli, ki se z izrezovanjem odstrani, da dobimo funkcionalno mRNA.
kapa	Posebna struktura iz 7-metilgvanozinskega ostanka, na transkript povezana s tremi fosfatnimi skupinami, ki se v jedru doda na 5' konec evkariotske mRNA. Kapa zaščiti mRNA pred razpadom in je nujna za kasnejšo translacijo v citoplazmi.
kvantitativni lokus	Področje v genomu, ki se pojavlja v statistično značilni povezavi z neko lastnostjo.
poli-A rep	Serijski 100-250 adeninskih nukleotidov, dodana na 3' konec mRNA po prepisovanju.
paralogni gen	Homologni geni znotraj vrste, ki so nastali z gensko duplikacijo po speciaciji in imajo lahko različne funkcije.
transkriptom	Skupina vseh mRNA, ki so v neki časovni točki izražene v neki celici ali populaciji celic.
trihom	Rastlinski epidermalni izrastki z različnimi funkcijami.
3' in 5' UTR	Regija transkripta RNA na 3' koncu za mestom terminacije prevajanja oz. na 5' koncu pred mestom začetka prevajanja.

## 1 UVOD

Hmelj (*Humulus lupulus* L.) je trajna ovijalka iz družine konopljevk (Cannabaceae). Rastlina je dvodomna in tujeprašnica. Storžki, ki zrastejo na ženskih rastlinah, imajo lupulinske žleze, kjer se proizvajajo pomembni sekundarni metaboliti. Storžki in njihovi ekstrakti so ključnega pomena v pivovarski industriji kot aromatičen dodatek pri varjenju piva, za hmeljne sekundarne metabolite pa se vedno bolj zanima tudi farmacevtska industrija (Miranda in sod., 1999; Gerhauser in sod., 2002).

O izvoru hmelja se še krešejo mnenja, najbolj verjetna domovina pa je Kitajska. Gotovo je dejstvo, da sta evolucijsko različni evropska in azijska skupina in da severnoameriški divji hmelj izhaja iz azijskih sorodnikov (Murakami in sod., 2006). Hmeljarstvo je v Sloveniji prisotno več kot sto let, hmelj pa je glavni izvozni artikel slovenskega kmetijstva. Po prozvodnji se naša država uvršča med deset največjih proizvajalk, kjer po površinah tradicionalno prednjačijo ZDA in Nemčija.

V pivovarstvu so posebno pomembne  $\alpha$ - in  $\beta$ -kisline (humulon, lupulon idr.), ki dajejo pivu značilno aroma. V njihovi biosintezi poti sodelujeta encima valerofenon sintaza (VPS) in halkon sintaza (CHS), homologni poliketidni sintazi. Poleg teh dveh so v hmeljnem genomu do sedaj identificirali še tri zapise za homologne gene (*chs2*, *chs3* in *chs4*). Znani so štirje domnevni kvantitativni lokusi ali QTL-i (angl. quantitative trait loci) za vsebnost  $\alpha$ -kislin in tudi podatek, da v asociaciji z njimi niso geni *chs2*, *chs3* ali *chs4* (Čerenak in sod., 2006). Zato nas zanima, če sta CHS ali VPS v asociaciji s katerim od štirih QTL-ov za vsebnost  $\alpha$ -kislin.

Pridobivanje oznak izraženih zaporedij (EST-jev) predvsem nemodelnih organizmov je danes vedno bolj pogosto, predvsem zaradi njihove cene, informativnosti in možnosti avomatizacije postopka. Njihov pomen je v tem, da predstavljajo kodogeni del genoma. V javno dostopnih podatkovnih zbirkah je na voljo več kot 10.000 zaporedij EST hmelja, ki do sedaj še niso bile opisane ali okarakterizirane.

V nalogi želimo hmeljne EST-je združiti v soseske (angl. contigs), iz le-teh pa dobiti podatke o prisotnosti in pogostosti mikrosatelitov v izraženem delu genoma hmelja. Izdelali bomo začetne oligonukleotide v celotni dolžini genov za CHS in VPS in iskali morebitni dolžinski polimorfizem alelov. V primeru homozigotnih dolžinskih alelov (monomorfima) bomo produktom PCR določili nukleotidno zaporedje in poiskali segregirajoč polimorfizem posameznih nukleotidov (SNP) v družini križancev. Iz teh podatkov bomo izvedli aplikacijo markerjev SNP na obeh genih (CHS in VPS) ter njun polimorfizem kartirali v obstoječo karto družine hmelja. Testirali bomo hipotezo, da katerikoli od navedenih dveh genov kaže na asociacijo s QTL-om z lastnostjo za produkcijo  $\alpha$ -kislin.

## 2 PREGLED OBJAV

### 2.1 OZNAKE IZRAŽENIH ZAPOREDIJ

Po centralni dogmi molekularne biologije (Crick, 1970) se DNA prepiše v RNA in nato prevede v protein. Po prepisu iz DNA najprej nastane pre-mRNA. Procesiranje le-te vključuje dodajanje posebnih struktur - kape na 5' koncu in poli-A repa na 3' koncu ter izrezovanje intronov in spajanje eksonov (Griffiths in sod., 2005). Dobimo molekule mRNA (angl. messenger RNA), ki predstavljajo transkriptom celice, t.j. skupina vseh mRNA, ki v določeni časovni točki nastajajo v izbrani populaciji celic (tkivu ipd.).

Molekule RNA so preveč nestabilne in tudi sicer neprimerne za nadaljnje postopke, zato jih z encimom reverzno transkriptazo prepišemo nazaj v molekule DNA, ki se sedaj imenujejo cDNA (angl. complementary DNA). Le-te kloniramo v primerne vektorje (najpogosteje plazmide) in naredimo knjižnice cDNA. Plazmidom z inserti cDNA nato določimo nukleotidno zaporedje v eni reakciji določanja nukleotidnega zaporedja (angl. single pass sequencing). Rezultat tega dela so zaporedja, ki jih imenujemo oznake izraženih zaporedij oz. krajše EST (angl. expressed sequence tags). Prvi članek o EST-jih, izoliranih iz človeških možganov, so leta 1991 objavili Adams in sod., od takrat pa se je razvilo mnogo aplikacij in orodij za delo z njimi.

Ker se klonom cDNA določa nukleotidno zaporedje v enem prehodu iz obeh smeri (3' ali 5'), so EST-ji razmeroma kratka (200 – 800 baznih parov) presežna 3' in 5' zaporedja (Nagaraj in sod., 2006). V začetku so bili favorizirani EST-ji iz 5' konca cDNA. Za ta zaporedja je namreč bolj verjetno, da vsebujejo zaporedja, ki kodirajo proteine kot pa 3' konci, ki pogosto vsebujejo še obsežne neprevedljive regije ali UTR (angl. untranslated regions) (Rudd, 2003). UTR so pomembni za posttranskripcijsko regulacijo izražanja genov, vključno z modulacijo transporta molekul mRNA iz jedra in učinkovitost translacije, za lokalizacijo znotraj celice in stabilnost (Mignone in sod., 2002).

Izboljšave v tehnikah priprave cDNA in napredovanje tehnologije določanja nukleotidnega zaporedja sta povzročila tudi razvoj EST-jev. Trenutno so namreč preferenčni tisti iz 3' konca, ker je večja verjetnost, da bodo vsebovali edinstveno sekvento in se zaradi tega lahko uporabijo za ločevanje paralognih genov. Določanje nukleotidnega zaporedja iz obeh smeri je vedno bolj pogosto (Rudd, 2003). Prav izboljšave v tehnologiji pa omogočajo določanje zaporedij celotne molekule cDNA, zaradi česar se izpusti pomemben korak – zbiranje EST-jev v gruče (angl. clusters), ki je lahko tudi vir napak (Wang in sod., 2004). Določanje zaporedij celotne cDNA bo počasi izpodrinilo EST-je.

Postopek pridobivanja EST-jev je pogosto avtomatiziran, zaporedja pa se ne preverjajo sproti, zato je kasneje potrebna vrsta postopkov *in silico* (več o tem v poglavju Materiali in metode). Tehnologija določanja nukleotidnega zaporedja EST-jev se izogne težavam, povezanim z velikostjo genoma in ponavlajočimi zaporedji (Rudd, 2003). Za razmeroma

nizko ceno jih dobimo veliko število. So široko uporabljeni pri številnih organizmih v različnih tkivih, razvojnih fazah in pod vrsto različnih biotskih in abiotskih stresov. Posebej pomembni so za nemodelne organizme, ki nimajo znanega nukleotidnega zaoredja celotnega genoma. Pri le-teh tudi relativno malo EST-jev prinese pomembne podatke (Malde in Jonassen, 2008).

Pri izdelavi EST-jev se srečamo z dvema težavama, in sicer z neenakomerno reprezentacijo genov v knjižnici in kakovostjo posameznega zaporedja. cDNA knjižnica je posnetek molekul mRNA znotraj celice ali tkiva v nekem trenutku, kar pomeni, da bodo geni z nizko stopnjo izražanja v knjižnici slabo zastopani, genov, ki se ne izražajo, pa sploh ne bo (Rudd, 2003). Tako bo v knjižnici veliko klonov genov s pogostim izražanjem, zato bomo dobili tudi več presežnih EST-jev. Po drugi strani pa je neenakomerna reprezentacija dobrodošel podatek o stopnji izražanja posameznih genov. Neenakomerno zastopanost posamezne mRNA v vzorcu lahko delno odpravijo s postopkom normalizacije (Bonaldo in sod., 1996).

Druga težava postane bolj očitna ob dejstvu, da je lahko v posameznem zaporedju EST lahko tudi več odstotkov napačnih nukleotidov (Rudd, 2003). Prvih in zadnjih 50-100 baznih parov EST-jev je običajno zelo slabe kakovosti zaradi tehnologije sekvenciranja (Nagaraj in sod., 2006), poleg tega pa so pogosto kontaminirani z vektorskimi in adapterskimi zaporedji (Pertea in sod., 2003; Chen in sod., 2007) in nenazadnje, tudi s strukturnimi ali regulatornimi molekulami RNA (Rudd, 2003).

Z EST-ji se v genomskih zaporedjih določajo geni, v pomoč so pri identifikaciji strukture genov ter raziskovanju tkivno specifičnega izražanja, alternativnega izrezovanja, iskanju polimorfizmov in polimorfizmov posameznih nukleotidov (SNP, angl. single nucleotide polymorphism). Uporabljo se pri poskusih z mikromrežami ter pomagajo pri analizi proteoma in identifikaciji proteinov, še posebej, če zaporedje celotnega genoma ni znano (Nagaraj in sod., 2006).

## 2.2 HALKON SINTAZA (CHS) IN VALEROFENON SINTAZA (VPS)

Za pivovarstvo so pomembne aromatične spojine, ki jih delijo v tri skupine, in sicer hmeljne smole, eterična olja in polifenolne spojine. Polifenolne spojine (tanini) se med varjenjem piva vežejo s proteini in jih oborijo. V splošnem imajo pozitiven vpliv na okus piva, stabilizirajo grenčico ter povečajo stabilnost piva in pene. Eterična olja predstavljajo od 0,5 do 1,5 % teže suhih storžkov in dajejo pivu vonj in okus. Delijo jih v v tri glavne skupine: terpene, oksigenirane derivate in spojine, ki vsebujejo žveplo. Hmeljne smole glede na topnost delimo na dve skupini, mehke in trde smole (Jakše, 2003).

Mehke smole so v pivovarstvu najbomembnejše. V to skupino sodijo  $\alpha$ -kisline (humulon, kohumulon in adhumulon) in  $\beta$ -kisline (lupulon, kolupulon in adlupulon), pri čemer je za

grenak okus piva pomembna izomerizacija predvsem  $\alpha$ -kislin.  $\alpha$ - in  $\beta$ -kisline se specifično sintetizirajo med razvojem ženskih socvetij (Slika 1) v žleznih trihomih.

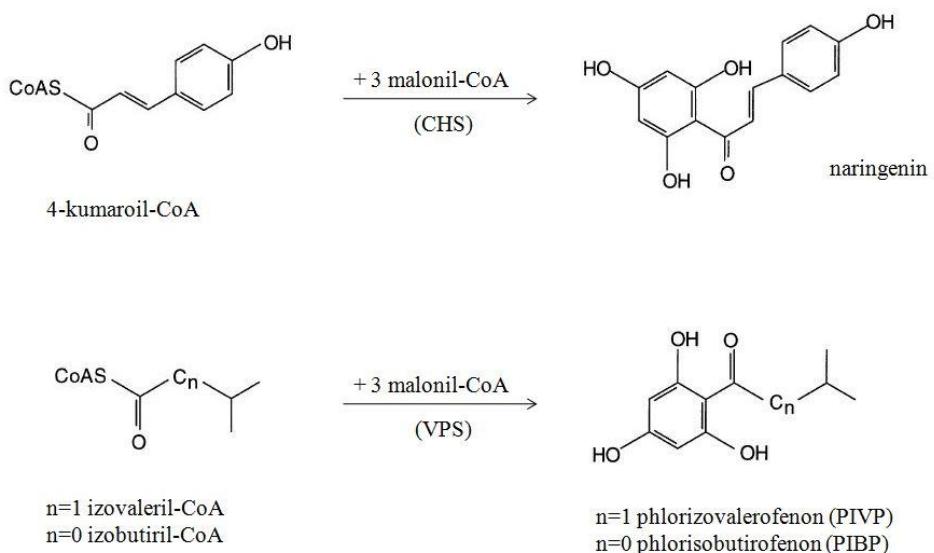


Slika 1: Ženski socvetji hmelja (»storžka«), ob osnovnem vretenu so lepo vidne lupulinske žleze (rumeno).

Halkon sintaza (EC 2.3.1.74) in valerofenon sintaza (EC 2.3.1.156) sta pomembna encima v sinteznih poteh snovi, ki vplivajo na aroma piva. Sta aciltransferazi, in sicer katalizirata zaporedno kondenzacijo treh acetatnih enot na neko začetno enoto. Rezultat je tetraketidni intermediat, ki se zvije v obroč. VPS je homolog CHS. Encima imata enako reakcijsko specifičnost, razlikujeta pa se v specifičnosti za substrat. (Paniego in sod., 1999). Raziskana je bila subcelularna lokalizacija CHS pri zoreči grozdni jagodi, in sicer se encim glede na razvojno stopnjo jagode nahaja na plastidih, grobem endoplazemskega retikulumu, steni vakuole ali na celični steni (Tian in sod., 2008).

CHS iz treh molekul malonil-CoA in kumaroil-CoA sintetizira naringenin halkon, prekurzor številnih flavonoidov (Schröder J. in Schröder G., 1990) (Slika 2). Med produkti te poti sta ksantohumol, za katerega se zaradi antikarcinogenih lastnosti zanima tudi farmacevtska industrija (Miranda in sod., 1999; Gerhauser in sod., 2002), in 8-prenilnaringenin, eden najmočnejših fitoestrogenov (Milligan in sod., 1999), ki je prav tako zanimiv za medicino in farmacevtsko industrijo. Prvi opisani encim biosinteze poti

$\alpha$ - in  $\beta$ -kislin je bila valerofenon sintaza, ki katalizira kondenzacijo treh molekul malonil-CoA in izovaleril-CoA ali izobutiril-CoA (Slika 2). Med končnimi produkti te poti je humulon. (Paniego in sod., 1999; Okada in Ito, 2001).



Slika 2: Dve izmed reakcij, ki jih katalizirajo rastlinske poliketid sintaze (Paniego in sod., 1999).

V genomu hmelja so do sedaj našli skupno pet homologov, poimenovanih *chs\_H1* ali prava halkon sintaza (GenBank pristopna številka AJ304877) (Matoušek in sod., 2002a), *chs2* (AB061020), *chs3* (AB061022), *chs4* (AJ430353) (Novák in sod., 2003), in *vps* (AB047593) (Okada in Ito, 2001). Novák in sod. (2003) predvidevajo, da poleg že identificiranih petih homologov obstaja še šest dodatnih neidentificiranih, ki naj bi bili podobni *vps*. Izvedli so tudi primerjavo petih identificiranih homologov na nukleotidnem in aminokislinskem nivoju. Ugotovili so, da so si geni *vps*, *chs3* in *chs4* med sabo najbolj podobni, in sicer je podobnost *chs3* z *vps* na nukleotidnem nivoju 79%, na aminokislinskem pa 77%, *chs4* z *vps* pa 80% na nukleotidnem in 77% na aminokislinskem nivoju. Za primerjavo je podobnost med *chs\_H1* in *vps* na nukleotidnem nivoju 68%, na aminokislinskem pa 73%. Tudi introni genov *vps*, *chs3* in *chs4* so si bolj podobni po dolžini in nukleotidnem zaporedju.

Najvišjo stopnjo izražanja *chs\_H1* so Matoušek in sod. (2002b) ugotovili v žleznih trihomih med zorenjem storžkov, in sicer vsaj 100-krat višjo od drugih tkiv. Gen *vps* se specifično izraža v žleznih tkivih (Okada in sod., 2001), prav tako kot *chs4* (Novák in sod., 2003), medtem ko se *chs3* v testiranih kultivarjih ni izražal, ker je prekinjen in verjetno psevdogen (Novák in sod., 2003). To je v skladu z ugotovitvami Okade in sod. (2004), kjer protein iz gena *chs3* ni imel encimske aktivnosti in ni dal produkta. Ostali štirje proteini so kazali encimsko aktivnost, najvišjo CHS\_H1 (za produkt naringenin halkon in izovalerofenon) in VPS (za produkt izovalerofenon). *chs2* in *chs4* kot substrat uporabljata izovaleril-CoA ali izobutiril-CoA, a se reakcija predčasno prekine (Novák in sod., 2006).

## 2.3 MIKROSATELITI IN KARTIRANJE HMELJA

Pri hmelju so uporabili vrsto metod in molekularnih označevalcev, med njimi RFLP, RAPD, STS, SSR, ISSR in AFLP (Patzak, 2001; Patzak in sod., 2007). Molekularni markerji so uporabni za identifikacijo sort, nadzor nad avtentičnostjo in čistoto sort, za žlahtnjenje s pomočjo označevalcev (MAS, angl. marker assisted selection), identifikacijo somaklonske variabilnosti in genetske stabilnosti, za genetsko kartiranje, detekcijo in identifikacijo različnih patogenov ter najpogosteje za študije genetske raznolikosti (Patzak in sod., 2007).

Mikrosateliti ali SSR (angl. simple sequence repeat) so zaporedja DNA, ki so sestavljena iz ponovitev istega motiva nekaj baznih parov. Variabilnost izhaja iz različnega števila ponovitev motiva med dvema aleloma. So številni, kodominantni, večalelni, specifični za določen lokus in naključno razporejeni po genomu. Uporabni so predvsem za primerjave znotraj vrst. Negativna stran je, da vnaprej potrebujemo znano nukleotidno zaporedje za izdelavo začetnih oligonukleotidov in dejstvo, da so večinoma neprimerni za medvrstno primerjavo, kar pomeni, da jih je potrebno razviti za vsako vrsto posebej.

Mikrosatelite delijo v tri skupine (Goldstein in Schlötterer, 1999; Chambers in MacAvoy, 2000):

- popoln mikrosatelit je sestavljen samo iz ene ponovitve osnovnega motiva in ni prekinjen z drugo bazo (npr. CTCTCTCTCTCTCTCT)
- prekinjen ali nepopoln mikrosatelit vsebuje kraši vključek baz, ki se ne ujemajo z osnovnim motivom ponovitve (npr. CTCTCTCTGGGCTCT)
- sestavljen mikrosatelit gradita vsaj dva popolna ali prekinjena mikrosatelite z različnima osnovnima motivoma (npr. CTCTCTCTGAGAGAGAGA)

Mikrosateliti so že dolgo znan molekularni označevalec in še vedno zelo uporaben. Na hmelju so bili prvič uporabljeni l. 1996 (Brady in sod., 1996). Danes je znanih več kot 220 hmeljnih mikrosatelitnih lokusov in začetnih oligonukleotidov zanje (Jakše in sod. (2002); Hadonou in sod. (2004); Štajner in sod. (2005); Jakše in sod. (2008)).

Mikrosateliti so bili poleg markerjev AFLP tudi uporabljeni pri kartiraju hmelja (Čerenak in sod., 2006; Seefelder in sod., 2000). S kartiranjem so odkrili štiri kvantitativne lokuse za vsebnost  $\alpha$ -kislin (Čerenak in sod., 2006). Z isto problematiko so se ukvarjali še Koie in sod. (2005), a so uporabili le označevalce AFLP. Kvantificirali pa so akumulacijo  $\alpha$ -kislin,  $\beta$ -kislin, ksantohumola, humulena, razmerje med kohumulonom in  $\alpha$ -kislinami, razmerje med kolupulonom in humulenom idr.

### 3 MATERIAL IN METODE

#### 3.1 BIOINFORMACIJSKI DEL

Večina bioinformacijskega dela je potekala na operacijskem sistemu linux, Fedora 8.

##### 3.1.1 Urejanje zaporedij EST in zbiranje v gruče

Iz GenBanka smo pridobili 13390 EST-jev hmelja (CD527119- CD527124, CO653411- CO654000, ES437670- ES437798, ES652314- ES658722 in EX515309- EX521564), 1554 EST zaporedij pa je bilo v fazi obdelave še neobjavljenih (Jeltsch, 2008). Velika večina EST-jev iz GenBanka je določena iz trihomov hmeljnih storžkov, kjer naj bi potekala sinteza iskanih encimov.

EST-je smo zbrali v eni datoteki v formatu fasta. S skripto SeqClean (DFCI Gene Indices Software Tools), ki uporablja programa blastall in megablast (Zhang in sod., 2000), smo jim odrezali poliA repe in odstranili ostanke vektorskih zaporedij s primerjavo s podatkovno zbirko vektorskih zaporedij UniVec (UniVec, 2008). Na koncu smo odstranili EST-je, ki so bili krajiši od 80 bp (pred ali po odstranitvi vektorskih zaporedij), ki so bili prenizke kakovosti (več kot 3% manjkajočih baz) ali so vsebovali velik delež vektorskega zaporedja (in je bilo manj kot 80 bp EST-ja nevektorskega izvora).

S programskega paketom tgccl (DFCI Gene Indices Software Tools, ) s programoma CAP3 (Huang in Madan, 1999) in megablast smo izvedli zbiranje v gruče. Za vizualizacijo in pregled gruč je bil uporabljen program clview (DFCI Gene Indices Software Tools).

S pomočjo programskega paketa netblast, ki omogoča lokalno uporabo algoritmov BLAST, smo primerjali nekaj največjih nastalih soseg z zaporedji v GenBanku in našli soseg, ki je ustrezala genu za VPS. Naredili smo poravnavo gena za VPS in te gruče ter locirali intron v genu, kar je bilo pomembno za načrtovanje začetnih oligonukleotidov za pomnoževanje delov gena v verižni reakciji s polimerazo (PCR).

Ko smo prejeli rezultate določanja nukleotidnega zaporedja potomcev družine hmelja, smo odkrite SNP-e v genu za VPS primerjali tud z domnevнимi SNP-i v gruči.

##### 3.1.2 Analiza zaporedij EST s skripto MISA

S Perl skripto MISA (Thiel, 2001) smo v pridobljenih sosegah hmeljnih EST-jev poiskali mikrosatelite. Minimalne zahteve za priznanje nekega zaporedje kot mikrosatelite so podane v Preglednici 1.

Preglednica 1: Definicija mikrosatelita za iskanje po sosegah.

Število nukleotidov v motivu	2	3	4	5	6	7	8	9	10
Min. število ponovitev motiva	6	5	4	4	4	4	3	3	3

Iskali smo popolne, prekinjene in sestavljeni mikrosateliti. Lokus je bil definiran kot eden, če sta bila dva mikrosatelite v zaporedju med sabo oddaljena manj kot 200 bp.

### 3.2 RASTLINSKI MATERIJAL

Genetska analiza hmelja je običajno izvedena na prvi generaciji potomcev ( $F_1$ ), ki imajo iste starše (angl. full-sib). Starša morata biti heterozigotna do te mere, da je določen molekularni označevalci v heterozigotni obliki pri vsaj enem od njih. Ta tip analize se imenuje dvojno psevdotestno križanje.

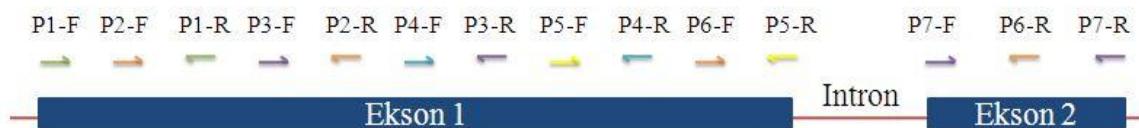
Določevanje nukleotidnega zaporedja in kartiranje je bilo izvedeno na družini križanja Magnum  $\times$  2/1. Magnum je nemški kultivar z visoko vsebnostjo  $\alpha$ -kislin, moški 2/1 pa je slovenski divji hmelj. Družino je ob sajenju nasada sestavljalo 111 potomcev  $F_1$  generacije, od tega 97 ženskih in 14 moških rastlin (Čerenak in sod., 2006), v našem primeru se je družina zaradi propadanja zmanjšala na 94 rastlin.

Določanje vsebnosti  $\alpha$ -kislin v treh zaporednih letih, izolacijo DNA družine ter kartiranje 96 označevalcev na materno in 70 na očetno karto in določitev štirih QTL-ov za producijo  $\alpha$ -kislin so opravili Čerenak in sod. (2006).

### 3.3 NAČRTOVANJE OLIGONUKLEOTIDNIH ZAČETNIKOV IN PCR VPS

S programom Primer3 (Rozen in Skaletsky, 2000) smo izbrali 7 parov začetnih oligonukleotidov, katerih teoretični produkti so skupaj pokrili celoten gen VPS. Razen talilne temperature ( $T_m$  min 52°C, optimalna  $T_m$  57°C,  $T_m$  max 60°C), in dolžine želenih produktov (ki smo jo spremnjali med 500 in 700 bp, glede na polimorfizme znotraj genske družine in položaj introna) smo uporabili privzete nastavitev programa. Glavni izliv načrtovanja začetnih oligonukleotidov je bilo iskanje mest prileganja, ki so bila polimorfna znotraj genske družine in značilna le za gen *vps*. Pomagali smo si s poravnavo že omenjenih zaporedij genske družine CHS s programom Clustal X 2.0.6 (Thompson in sod, 1997; Larkin in sod., 2007).

V preglednici 2 je predstavljenih 7 parov izdelanih začetnih oligonukleotidov. Posebno pozornost smo posvetili P6, ker se razteza preko introna, kjer smo pričakovali tudi največje število polimofizmov. Položaji oligonukleotidnih začetnikov so grafično predstavljeni na sliki 3.



Slika 3: Shema strukture gena za VPS in položajev oligonukleotidnih začetnikov.

PCR je potekal v volumnu 20 µL, in sicer je vsebovala 5,0 µL DNA s koncentracijo 4 ng/µL, 8,9 µL dH<sub>2</sub>O, 2,0 µL 10x PCR pufra z 1,5 mM MgCl<sub>2</sub>, 0,4 µL 25 mM MgCl<sub>2</sub> (skupna koncentracija MgCl<sub>2</sub> v reakciji je bila 2 mM), 1,6 µL 10 mM mešanice dNTP, po 1 µL 10 µM raztopine obeh začetnih oligonukleotidov ter 0,1 µL encima *Taq* polimeraze s koncentracijo 5 U/µL.

PCR smo izvajali v napravi za ciklično termostatiranje PE9700. Temperaturni profil pomnoževanja je bil sledeč:

- 5 minut pri 94°C
- 5 ciklov, 45 sekund pri 94°C, 30 sekund pri 59°C in 90 sekund pri 72°C
- 28 ciklov, 45 sekund pri 94°C, 30 sekund pri 54°C in 90 sekund pri 72°C
- 8 minut pri 72°C.

Preglednica 2: Lastnosti sedmih parov začetnih oligonukleotidov za pomnoževanje delov gena VPS.

Ime	Zaporedje	Dolžina zač. nukleotida (bp)	CG% <sup>1</sup>	P (bp) <sup>2</sup>	T <sub>m</sub> (°C) <sup>3</sup>
P1-F	ATCGATCAACACCCCTCCTAA	20	45,00	637	57,05
P1-R	GGCCAACGTGATGCTCTTAAT	20	45,00		56,90
P2-F	AATACTGGGTGCACTGTTTG	20	45,00	657	56,13
P2-R	TTCCGCATAATAGCACTTGA	20	40,00		57,02
P3-F	GCTGCTATAACCCAACACCTT	20	50,00	530	56,85
P3-R	GCTTAGCCGGAAAGTTCATA	20	45,00		57,17
P4-F	TATTGGCCAAACGTCAAAC	19	42,11	614	57,05
P4-R	AGCTCTCTGTTGGGTCAAGTT	20	50,00		56,51
P5-F	CTCGATGCCATAACTTCCTT	20	45,00	679	56,91
P5-R	GGAAAATCAGCTTGGTTGAA	20	40,00		57,79
P6-F	TAATGGCGTCCGTAACTGTA	20	45,00	613	56,81
P6-R	GCTATGCGAAGAACCTTTCC	20	45,00		56,75
P7-F	CAATCAATGCCATCAAAGAA	20	35,00	691	57,13
P7-R	CATGACATGTTCCCGTACTC	20	50,00		56,34

<sup>1</sup> Odstotek vsebnosti baz gvanina in citozina v začetnem nukleotidu.

<sup>2</sup> Predvidena dolžina produkta PCR.

<sup>3</sup> Predvidena talilna temperatura.

Vzorce smo analizirali s horizontalno agarozno elektroforezo v 0,5 × TBE elektroforeznem pufru (44,5 mM Tris, 44,5 mM borna kislina in 1 mM EDTA). 1,3% gel je bil pripravljen v 1× TBE elektroforetskem pufru. Elektroforeza je potekala pri konstantni napetosti 100 V. Uporabljali smo horizontalno elektroforetsko napravo SubCell Model 192 (Bio-Rad, ZDA).

Na gel smo nanesli mešanico 80% produkta PCR in 20% nanašalnega barvila (12,5%- (w/v) Ficoll tip 400, 0,2-odstoten (w/v) brom fenol modro). Poleg vzorcev pa smo na gel nanesli tudi 200 ng DNA markerja 100 bp lestvice (Promega) s fragmenti dolžin 100, 200, 300, 400, 500, 600, 700, 800, 900 in 1000 bp. Analizirano DNA smo obarvali z etidijevim

bromidom, ki smo ga predhodno dodali v agarozni gel v koncentraciji 10 µg/mL gela. Elektroforetske vzorce smo opazovali s transiluminatorjem TFM-30 (UVP Inc., Anglija) in jih fotografirali s pomočjo digitalnega fotografskega aparata Nikon.

Optimizacija pomnoževanja je potekala na 8 naključno izbranih vzorcih DNA potomcev križanja Magnum × 2/1. Para začetnih oligonukleotidov P3 in P6 sta dala pričakovani rezultat z osnovnim protokolom. Za par P7 smo temperaturo v prvi seriji ciklov s 59°C znižali na 57°C.

### 3.4 NAČRTOVANJE OLIGONUKLEOTIDNIH ZAČETNIKOV IN PCR CHS

V GenBanku smo poiskali dve nukleotidni zaporedji za CHS\_H1 (AM263199, AJ304877) in ju poravnali. Pri načrtovanju začetnih oligonukleotidov smo se želeli izogniti domnevno polimorfnim mestom in enemu intronu. S programom Primer3 smo s privzetimi nastavtvami izdelali 3 pare začetnih oligonukleotidov, ki so predstavljeni v Preglednici 3.

Preglednica 3: Lastnosti treh izbranih parov začetnih oligonukleotidov za pomnoževanje gena CHS.

Ime	Zaporedje	Dolžina zač. nukleotida (bp)	CG% <sup>1</sup>	P (bp) <sup>2</sup>	T <sub>m</sub> (°C) <sup>3</sup>
P33	CGACCGTCCATTCTCCTAC	20	55	598	59,55
P34	ACTCTGGAGGATGCAGTTGG	20	55		60,26
P35	CCAGACAAGACATGGTGGTG	20	55		60,00
P36	GGGTTCCCTCACAGAACATCA	20	50	1008	59,90
P37	GAAGTTCGCAAGGCTCAAAG	20	50		60,13
P38	TCAGCACCAATGATGAGAGC	20	50	858	59,95

<sup>1</sup>Odstotek vsebnosti baz gvanina in citozina v začetnem oligonukleotidu.

<sup>2</sup>Predvidena dolžina produkta PCR.

<sup>3</sup>Predvidena talilna temperatura.

PCR je potekal v 25 µL, in sicer je vsebovala 10 µL DNA s koncentracijo 4 ng/µL, 7,375 µL dH<sub>2</sub>O, 2,5 µL 10x PCR pufra z 1,5 mM MgCl<sub>2</sub>, 0,5 µL 25 mM MgCl<sub>2</sub> (skupna koncentracija MgCl<sub>2</sub> v reakciji je bila 2 mM), 2,0 µL 10 mM mešanice dNTP, po 1,25 µL 10 µM raztopine obih začetnih oligonukleotidov ter 0,125 µL encima *Taq* polimeraze s koncentracijo 5 U/µL.

PCR smo izvajali v napravi za ciklično termostatiranje PE9700. Temperaturni profil pomnoževanja je bil sledeč:

- 5 minut pri 94°C
- 35 ciklov, 45 sekund pri 94°C, 30 sekund pri 55°C, 90 sekund 72°C
- 8 minut pri 72°C

Produkti PCR so bili nanešeni na 1,2% agarozni gel z etidijevim bromidom (končna koncentracija etidijevega bromida je bila 10 µg/ml gela). Gel in vzorci so bili pripravljeni na isti način kot za VPS. Optimizacija reakcije PCR je potekala na 12 vzorcih družine.

### 3.5 DOLOČANJE NUKLEOTIDNEGA ZAPOREDJA

#### 3.5.1 Izolacija produktov PCR iz agaroznega gela (CHS)

Produkte PCR smo izrezali s skalpelom iz agaroznega gela in izolirali s kompletom DNA Gel Extraction Kit po navodilih proizvajalca (Fermentas). Koščku gela z DNA smo v 1,5 mL mikrocentrifugirki dodali raztopini iz kompleta: 1,0 mL »binding solution« in 111,0  $\mu$ L TBE »conversion buffer« ter 10 minut inkubirali v vodni kopeli na 55°C, da se je rezina gela povsem raztopila. Nato smo dodali 8,0  $\mu$ L suspenzije silicijevega prahu iz kompleta in 5 minut inkubirali na 55°C, vsaki 2 minuti pa smo vsebino mikrocentrifugirk temeljito premešali. Mikrocentrifugirke smo centrifugirali 5 sekund, da se je silicijev prah usedel, in odlili supernatant.

Usedlini smo dodali 500  $\mu$ L ledeno mrzlega pufra za spiranje iz kompleta, temeljito premešali, da se je le-ta popolnoma resuspendirala, centrifugirali 5 sekund in odlili supernatant. Ta korak smo trikrat ponovili. Ko smo zadnjič odstranili supernatant, smo še enkrat centrifugirali in odpipetirali preostalo tekočino, da smo odstranili ves etanol. Usedlino smo nato 10-15 minut sušili na zraku.

DNA smo potem eluirali v 30,0  $\mu$ L pufra TE in 5 minut inkubirali na 55°C. Mikrocentrifugirko smo centrifugirali. Supernatant z raztopljenou DNA smo odpipetirali v novo 1,5 mL mikrocentrifugirko in jo do nadaljnje uporabe shranili na -20°C.

#### 3.5.2 Neposredno čiščenje produktov PCR (VPS)

Uporabili smo komplet ExoSAP-IT kit proizvajalca USB Corp., ki temelji na encimski razgradnji neporabljenih začetnih oligonukleotidov in deoksinukleotidov trifosfatov iz PCR reakcije z encimoma alkalna fosfataza rakov (angl. shrimp alkaline phosphatase) in eksonukleaza I.

5,0  $\mu$ L reakcije PCR smo dodali 2,0  $\mu$ L Exo-SAP IT encimov in inkubirali 15 minut pri 37°C za razgradnjo začetnih oligonukleotidov in deoksinukleotidov trifosfatov, nato pa še 15 minut na 80°C, da smo inaktivirali encime.

#### 3.5.3 Reakcija za določanje nukleotidnega zaporedja

Sekvenčna reakcija je bila prirejena po protokolu Biotehnoškega centra Univerze v Wisconsinu (ZDA): 3,5  $\mu$ L DNA (očištene v prejšnjih korakih), 0,2  $\mu$ L 10  $\mu$ M za produkt PCR specifičnega začetnega oligonukleotida, 0,5  $\mu$ L BigDye 3.1 sekvenčne mešanice proizvajalca Applied Biosystems, 2  $\mu$ L 5x BigDye pufra in 3,8  $\mu$ L dH<sub>2</sub>O.

Temperaturni profil je bil sledeč in se je izvajal na napravi za ciklično termostatiranje PE9700:

- 3 minute pri 96°C

- 50 ciklov: 10 sekund pri 96°C, 10 sekund pri 50°C in 4 minute pri 60°C
- 7 minut pri 72°C

### **3.5.4 Čiščenje reakcije za določanje nukleotidnega zaporedja in določevanje nukleotidnega zaporedja**

Za čiščenje (odstranjevanje nevgrajenih fluorescenčnih terminatorjev) smo uporabili reagent CleanSEQ (Agencourt Bioscience Corp.) po navodilih proizvajalca s polovično količino magnetnih delcev. Reakciji za določanje nukleotidnega zaporedja smo dodali 10,0  $\mu$ L vode, 5,0  $\mu$ L reagenta CleanSEQ beads (magnetni delci) in 80  $\mu$ L 80% etanola. Raztopino smo premešali s pipetiranjem, nato pa postavili na magnetno ploščo. Počakali smo 3 minute, da je magnetna plošča privlekla magnetne delce na rob epruvete, in odpipetirali tekočino. Dodali smo 200  $\mu$ L 80% etanola in nato odstranili čim več tekočine. Vzorce smo odstranili z magnetne plošče. Dodali smo 15,0  $\mu$ L vode in počakali, da so se magnetni delci resuspendirali. Vzorce smo postavili nazaj na magnetno ploščo, počakali 3 minute in odpipetirali po 10,0  $\mu$ L vzorcev v ABI kompatibilno mikrotitrsko ploščo. Tako pripravljenim vzorcem smo določili nukleotidno zaporedje v laboratoriju Katedre za genetiko, animalno biotehnologijo, imunologijo, splošno živinorejo in konjerejo (na Oddelku za zootehniko Biotehniške fakultete Univerze v Ljubljani) na kapilarni napravi ABI3130XL v 36 cm kapilarji.

### **3.5.5 Obdelava rezultatov določanja nukleotidnega zaporedja**

Za VPS in CHS smo najprej določili nukleotidno zaporedje 16 oz. 8 vzorcem, za vsak začetni oligonukleotid. Rezultate (kromatograme) smo pregledali s programom CodonCode Aligner verzija 2.0.6 in poiskali domnevna mesta polimorfizmov posameznih nukleotidov. Na podlagi tega smo se odločili, s katerimi oligonukleotidnimi začetniki bomo določali zaporedje na celotni družini Magnum  $\times$  2/1. Za gen *chs\_H1* je bil to par P33 in P34, za *vps* pa para P6 in P7.

## **3.6 KARTIRANJE GENOV ZA CHS IN VPS**

Segregacijo označevalcev smo testirali s testom  $\chi^2$  v Excelovi preglednici.

Kartiranje smo izvedli s programom JoinMap 3.0 (Van Ooijen in Voorrips, 2001). Fazo (angl. phase) označevalca je na podlagi frekvence rekombinacij določil program. Za izdelavo karte smo uporabili minimalno vrednost LOD (angl. logarithm of odds) 5,0 za sestavljanje skupin označevalcev. Za pretvorbo podatkov o rekombinaciji v razdalje na karti smo uporabili Kosambijevo funkcijo (Kosambi, 1944).

## 4 REZULTATI

### 4.1 BIOINFORMACIJSKI DEL

#### 4.1.1 Rezultati urejanja in zbiranja v gruče

Po začetni obdelavi dobljenih EST-jev (odstranjevanje vektorskih zaporedij) je od 14944 ostalo 14731 zaporedij. Odstranili smo 213 zaporedij, vzroki so podani na Sliki 4. Vektorska zaporedja so bila najdena in nato odstranjena v 4683 EST-jih, kar je tretjina vseh EST zaporedij.



Slika 4: Vzroki odstranitve skupno 213 EST zaporedij pred nadaljnjem delom.

Preostalih 14731 nukleotidnih zaporedij je programski paket tgicl združil v 1318 gruč, ostalo pa je 76 posameznih zaporedij. Primer grafične predstavitev nastale gruče CL4Contig2 v programu clview je predstavljen na Sliki 5.

Pet sosesk je bilo daljših od 2000 bp (3431 bp, 2146 EST-jev v gruči; 4426 bp, 796 EST-jev v gruči; 3357 bp, 414 EST-jev v gruči; 2034 bp; 2132 bp), 80 med 1500 in 2000 in 53 med 1000 bp in 1500 bp.



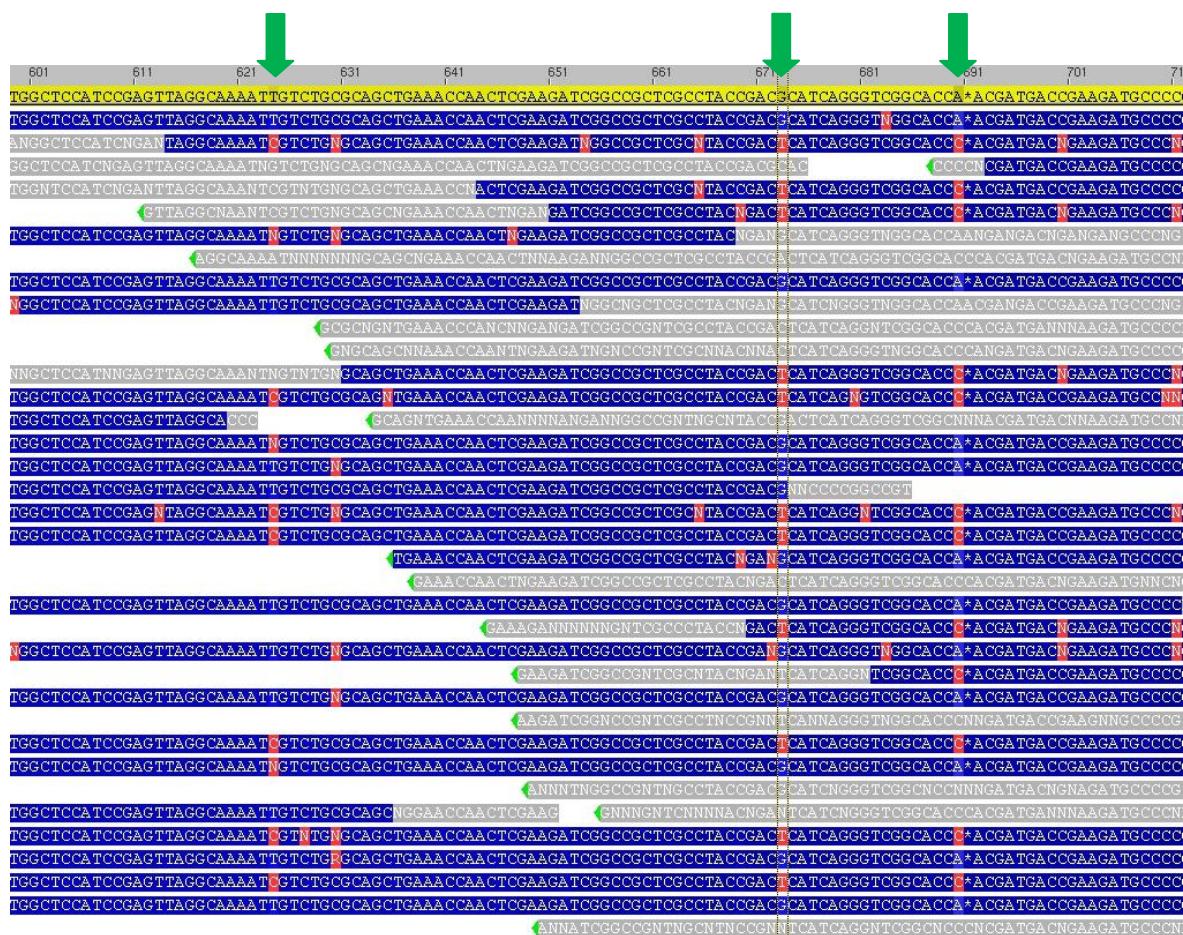
Slika 5: Primer rezultata zbiranja v gruče, prikazan je CL4Contig2, ki je transkript gena za VPS. Vsaka vodoravna črta ponazarja en EST v gruči.

#### 4.1.2 Rezultati skripte MISA

Od 1394 zaporedjih jih je 235 (tj. 16,8%) vsebovalo skupno 277 mikrosatelitov. Najpogosteji so bili dinukleotidni mikrosateliti (131), ki so jim sledili trinukleotidni mikrosateliti (118). Mikrosatelitov s širimi ali več nukleotidi v motivu je bilo po 10 ali manj. Najpogosteji dinukleotidni mikrosateliti (glede na komplementarnost sekvenc) so bili AG/CT, sledili so jim AT/AT, najmanj pa je bilo AC/GT. Število posameznih tipov mikrosatelitov je podano v Prilogi A.

#### 4.1.3 Primerjava najdenih SNP-ov

Ko smo prejeli rezultate določanja nukleotidnega zaporedja, smo najdene SNP-e v genu za VPS primerjali s SNP-i v gruči. Od 16 SNP-ov, ki smo jih odkrili pri določanju nukleotidnega zaporedja, jih je bilo v gruči opaznih 11, od tega trije zelo izrazito. Primer je predstavljen na Sliki 6. Včasih je znak za prisotnost SNP-a mogoče slutiti tudi, ko se v zaporedju nahaja N in sta dva enako visoka vrhova zmedla program za določanje vrednosti baze (angl. base caller), ki je iz kromatogramov določal nukleotidno zaporedje.



Slika 6: Gruča EST-jev, katerih sosekska predstavlja gen za VPS. Označena so tri polimorfna mesta, ki smo jih pred tem potrdili v naših produktih PCR.

#### 4.2 PCR IN DOLOČANJE NUKLEOTIDNEGA ZAPOREDJA VPS

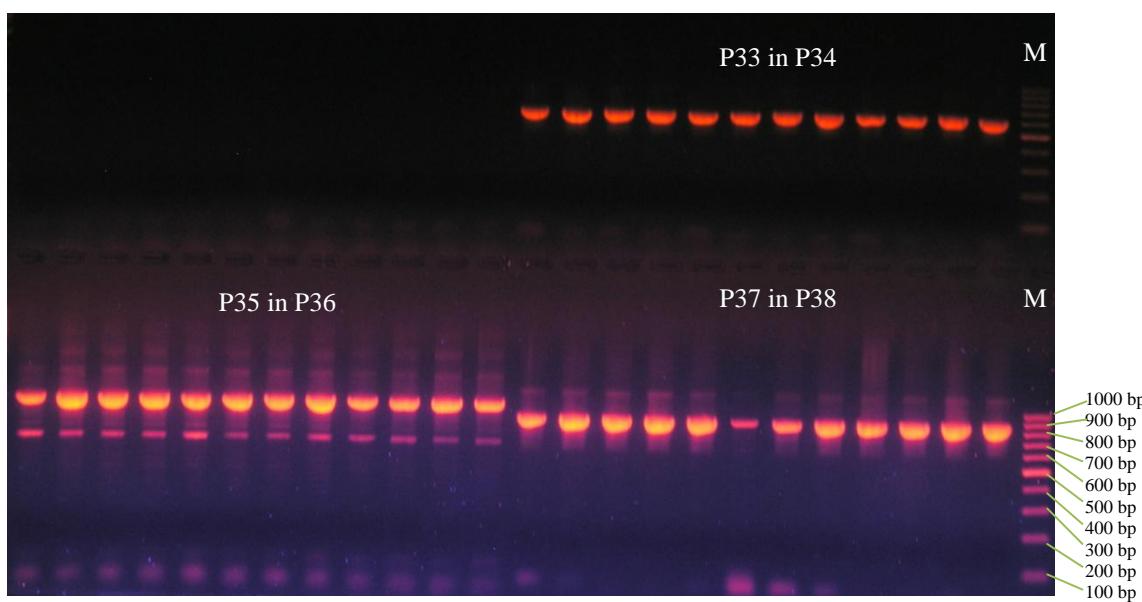
Optimizacija je potekala na 8 naključno izbranih vzorcih DNA potomcev križanja Magnum × 2/1. Kombinaciji začetnih oligonukleotidov P3 in P6 sta dali pričakovani rezultat z osnovnim protokolom. Za par P7 smo temperaturo v prvi seriji ciklov s 59°C znižali na 57°C. Skupno smo torej uspešno optimirali tri pare začetnih oligonukleotidov.

Par začetnih oligonukleotidov P7 je pokazal 8 SNP-ov. Določanje zaporedja s P6-R je pokazalo 6 SNP-ov in indel, s P7-F pa en SNP in isti indel. Zaporedje, narejeno s P3-R, je bilo slabše kakovosti, v zaporedju s P3-F pa smo našli en SNP. Na podlagi tega smo se odločili, da bomo določili nukleotidno zaporedje PCR produktom, pridobljenim s pari začetnih nukleotidov P6 in P7, in sicer z začetnima nukleotidoma P6-R in P7-R.

Genotipske vrednosti lokusa so priložene v Prilogi B.

#### 4.3 PCR IN DOLOČANJE NUKLEOTIDNEGA ZAPOREDJA CHS

Optimizacija je potekala na 12 vzorcih DNA za tri kombinacije začetnih nukleotidov. Produkti PCR so predstavljeni na Sliki 7.

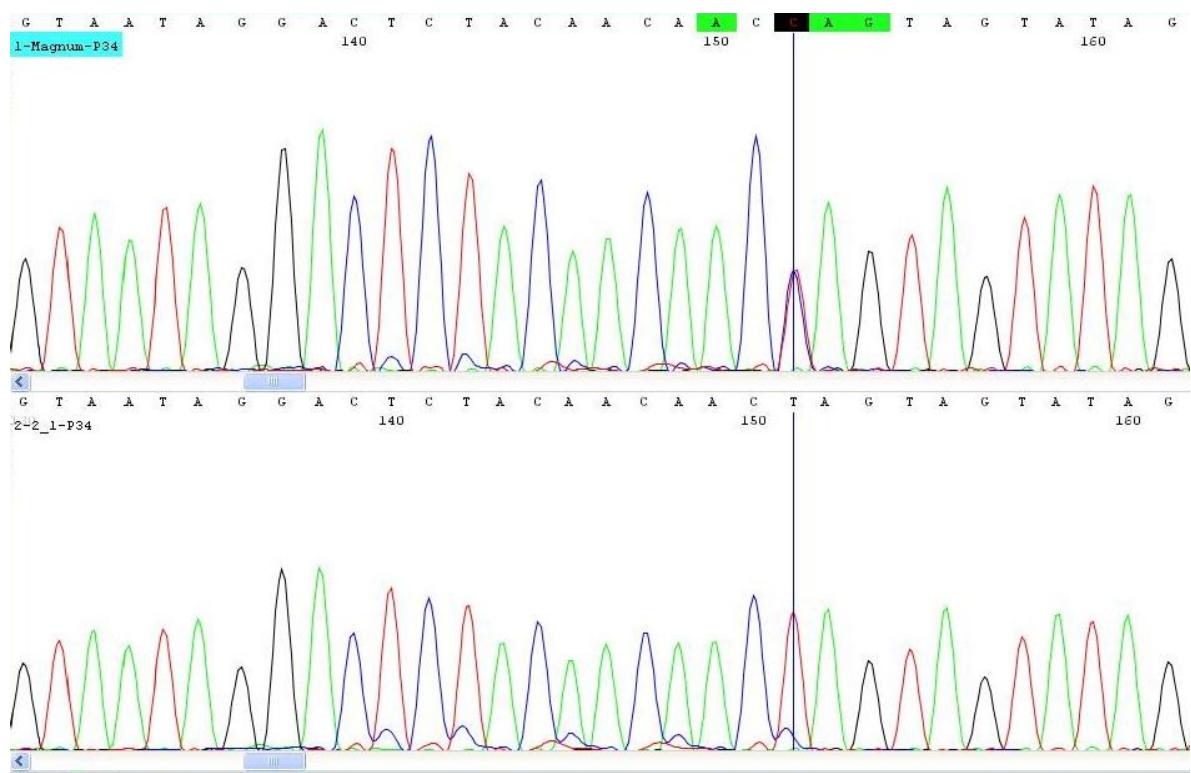


Slika 7: Slika agarognega gela s produkti PCR. Zgoraj desno: produkt PCR z začetnima oligonukleotidoma P33 in P34; spodaj levo produkt PCR z začetnima oligonukleotidom P35 in P36, spodaj desno pa s P37 in P38.

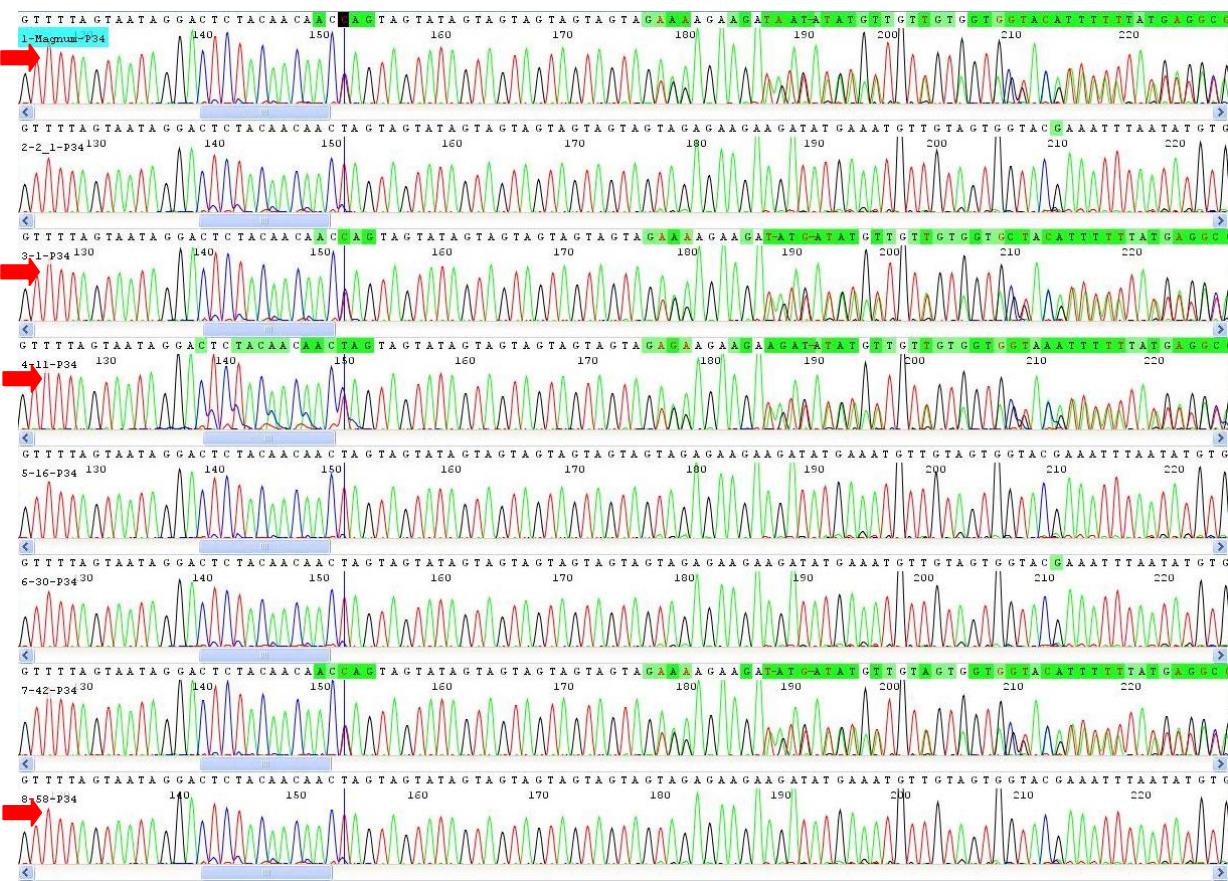
Produkt PCR z začetnima nukleotidoma P33 in P34, kateremu smo nukleotidno zaporedje določili z začetnim oligonukleotidom P33, je nakazoval prisotnost polimorfnega mesta z zaporedjem CCCT<sub>10</sub>. To zaporedje se nahaja v skrajnem 5' delu gena, ki bi lahko bil še 5'UTR.

Produkt PCR, kateremu smo nukleotidno zaporedje določili z začetnim oligonukleotidom P34, je nakazoval prisotnost SNP-a za zaporedjem ACTCTACAACAC. Magnum je na tem mestu heterozigoten za citozin in timin, 2/1 pa je homozigot za timin (Slika 8). Pri analizi 6 potomcev smo potrdili prisotnost homozigotnih in heterozigotnih rastlin.

Dva nukleotida za omenjenim SNP-om se začne mikrosatelit z motivom GTA, ki ima pri homozigotnem 2/1 6 ponovitev. Magnum je tudi tu heterozigot. En alel ima 6 ponovitev, drugi pa 5, kar povzroči porušitev zaporedja (Slika 9). Od tega mesta je pri vseh heterozigotnih rastlinah zaporedje neberljivo (zaradi dveh zaporedij različnih dolžin). Zaporedje homozigotov za 6 ponovitev pa ostane berljivo.



Slika 8: Del grafičnega prikaza rezultatov določanja nukleotidnega zaporedja produkta PCR dela gena za CHS z začetnim oligonukleotidom P34. Označen je SNP – pri Magnumu (zgornja vrsta) sta vidna dva vrhova (moder in rdeč) za citozin in timin, pri 2/1 pa je viden le en rdeč vrh za timin. Desno od SNP-a je viden tudi začetek mikrosatelita GTA.



Slika 9: Primer analize potomcev polimorfnega mesta gena *chs\_H1* iz slike 8. Viden je celoten trinukleotiden mikrosatelit GTA in porušenje zaporedja desno od njega pri heterozigotnih rastlinah (prvo, tretje, četrto in sedmo zaporedje; označena so z rdečimi puščicami).

V regiji, ki smo jo pomnožili z začetnima oligonukleotidoma P35 in P36, nismo zasledili polimorfizmov. Pri regiji, pomnoženi z P37 in P38, pa pri določanju nukleotidnega zaporedja z obeh strani nismo dobili berljivega zaporedja.

Na podlagi teh podatkov smo se odločili, da bomo na družini Magnum  $\times$  2/1 pomnožili regijo s P33 in P34 in produktu določili zaporedje z oligonukleotidnim začetnikom P34. Genotipske vrednosti lokusa so priložene v Prilogi B.

#### 4.4 REZULTATI KARTIRANJA

Oba gena, CHS in VPS, sta bila heterozigotna pri Magnumu (ženski starš) in homozigotna za 2/1 (tip križanja ab  $\times$  aa). Test  $\chi^2$ , predstavljen v Preglednici 4, ni pokazal odstopanj od pričakovanih Mendlovih razmerij.

Oba lokusa sta se pri kartirjanju uvrstila na skupini povezanosti 8 (Magnum-8). Genetske razdalje v tej skupini povezanosti so prikazane v Preglednici 5 in grafično na Sliki 10.

Čerenak in sod. (2006) so štiri domnevne QTL-e za vsebnost  $\alpha$ -kislin umestili na skupine povezanosti 1, 3 (dva QTL-a) in 9, kar pomeni, da niti gen za CHS niti gen za VPS nista v asociaciji z njimi.

Preglednica 4: Test  $\chi^2$  za segregacijo obeh analiziranih genov pri družini Magnum  $\times$  2/1.

lokus	Število posameznikov		$\chi^2$	df <sup>1</sup>	p <sup>2</sup>
	aa	ab			
VPS	52	38	2,18	1	0.140
CHS	43	47	0,18	1	0.671

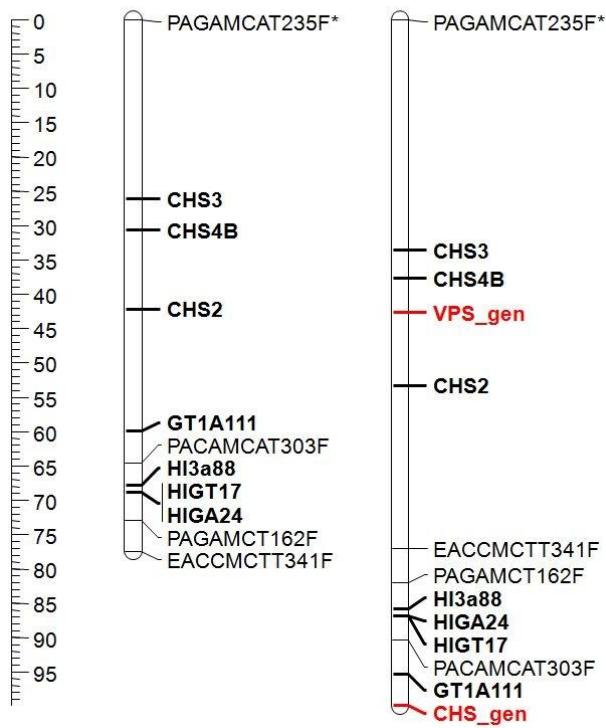
<sup>1</sup> stopnja prostosti

<sup>2</sup> verjetnost

Preglednica 5: Genetske razdalje med označevalci v skupini povezanosti 8 (Magnum-8). Ostali markerji iz te skupine izvirajo iz dela Čerenak in sod. (2006).

Št.	Položaj (cM)	Lokus
1	0.00	PAGAMCAT235F*
2	33.65	CHS3
3	37.73	CHS4B
4	42.74	VPS_gene
5	53.29	CHS2
6	76.96	EACCMCTT341F
7	82.03	PAGAMCT162F
8	85.76	Hl3a88
9	86.84	HIGA24
10	86.90	HIGT17
11	90.28	PACAMCAT303F
12	95.39	GT1A111
13	99.92	CHS_gene

**Magnum-8      Magnum-8 (CHS+VPS)**



Slika 10: Skupina povezanosti Magnum-8. Na levi strani je karta pred (Čerenak in sod., 2006), na desni pa po dodatku novih markerjev.

## 5 RAZPRAVA IN SKLEPI

### 5.1 EST

Presenetljivo veliko število zaporedij iz GenBank ni imelo odstranjenih vektorskih zaporedij. O težavah s tem poročajo že dolgo časa (Lamperti in sod., 1992; Seluja in sod., 1999). Kritično ocenjevanje postopkov, povezanih z obdelavo EST-jev pred vnosom v podatkovne zbirke, pa je še vedno pomanjkljivo, verjetno zaradi prevelike izbire in nepoznavanja lastnosti orodij (Nagaraj in sod., 2006).

Rezultati zbiranja EST-jev v gruče bodo dostopni na internetu. Zanimiva in uporabna aplikacija izvedenega zbiranja v gruče bi bilo iskanje domnevnih SNP-ov izbranih genov, kot smo to opravili za VPS. Pogoj je dovolj veliko število EST sekvenc, ki se bo v prihodnosti nedvomno povečevalo. Omejitev glede sorodnosti kultivarjev, za katere lahko uporabimo to ekstrapolacijo, bo potrebno natančneje preučiti.

EST-ji vsekakor ostajajo pomembno orodje, še posebej v sodelovanju z bioinformacijskimi orodji. Pred kratkim so npr. z analizo EST-jev iz žleznih trihomov našli encim v isti biosintezi poti kot CHS (Nagel in sod., 2008).

### 5.2 MIKROSATELITI

Brady in sod. (1996) so predpostavili naslednjo pogostnost mikrosatelitskih ponovitev v hmeljnem genomu (od najvišje do najnižje): GA, GT, TAA, CAA, GACA, CAC, GATA in GAA. Prvi je bil najpogosteji tudi pri nas, tu pa se podobnosti končajo. Glavna razlika je, da smo mi delali s kodirajočimi zaporedji, zato so bile pričakovane višje frekvence trinukleotidnih mikrosatelitov, kjer mutacija mikrosatelita vpliva na odvzem ali dodatek ene aminokisline in ne na porušitev bralnega okvirja proteina.

V pridobljenih soseskah smo našli skupno 277 mikrosatelitov, ki so lahko uporabno orodje pri kartiranju.

### 5.3 UPORABA SNP

Pri obeh analiziranih genih smo potrdili prisotnost SNP-ov pri Magnumu in potrdili označevalce SNP kot označevalce z najvišjim deležem polimorfizma v genomih. Pri genu VPS smo odkrili kar 15 polimorfizmov v dolžini 1111 bp, kar je 1 SNP na vsakih 74 bp za to regijo in precej odstopa od pričakovane vrednosti, ki naj bi bila približno 1 SNP na 100 bp (Edwards in Mogg, 2001). Zanimivo je, da je kodogeni del imel več SNP-ov kot intron – v intronu smo namreč našli le dva SNP-a.

### 5.4 KARTIRANJE GENOV ZA CHS IN VPS

Uspešno smo kartirali dva hmeljna gena v že obstoječo gensko karto družine Magnum × 2/1. V nasprotju s postavljenou hipotezo ne VPS ne CHS nista v asociaciji s QTL-om za  $\alpha$ -

kisline. V asociaciji s tem QTL-om bi lahko bil kateri izmed ostalih encimov iste biosinteze poti, a je to malo verjetno, ker sta CHS in VPS ključna v tem procesu. Mogoče je tudi, da je s QTL-om asociirani kateri izmed regulatornih proteinov, saj je sama ekspresija gena *vps* je natančno uravnavana (Sugiyama in sod., 2006).

Oba gena sta kartirana v isto skupino povezanosti (skupina 8, Slika 10), kamor so predhodno kartirali tudi druge gene iz družine CHS (Čerenak in sod., 2006). Znano je, da tudi pri rastlinah obstajajo gruče genov za biosinteze poti (Field in Osbourn, 2008). Ker je ta biosinteza pot ekonomsko zanimiva in zaradi domnevne prisotnosti še šestih do sedaj neidentificiranih homologov družine CHS (Novák in sod., 2003), je skupina povezanosti 8 je vsekakor vredna bolj podrobnih študij.

Ne glede na to, da nismo odkrili povezanosti s QTL-i za vsebnost  $\alpha$ -kislin, smo pripomogli k saturiranju genske karte, kar bo v pomoč pri nadalnjem delu.

## 6 POVZETEK

14944 hmeljnih EST-jem, ki so večinoma izhajali iz trihomov, smo porezali vektorska zaporedja, odstranili prekratka zaporedja in zaporedja nizke kakovosti ter preostanek zbrali v gruče s programskim paketom tgc1. Dobili smo 1318 gruč, ostalo pa je 76 posameznih zaporedij. Pregledali smo vsebnost mikrosatelitov v nastalih soseskah. 235 sosek je vsebovalo skupno 277 mikrosatelitov. Najpogosteji so bili dinukleotidni (131) in trinukleotidni (118) mikrosateliti.

Soseske smo primerjali z zaporedji v GenBanku. Sosesko, ki je ustrezala genu za valerofenon sintazo (VPS), smo primerjali z zaporedjem za gen, da smo locirali intron. Izdelali smo sedem parov začetnih oligonukleotidov, katerih produkti so skupaj pokrivali celoten gen. Reakcijo za PCR smo optimizirali za tri pare in rezultate preverili na agaroznem gelu. Produkte reakcije PCR smo očistili neposredno, z encimi za razgradnjo začetnih oligonukleotidov in deoksinukleotidov trifosfatov.

Za izdelavo začetnih oligonukleotidov za halkon sintazo (CHS) smo uporabili zaporedja, ki so se nahajala v GenBanku. Izdelali smo tri pare začetnih nukleotidov in produkte PCR reakcije ločili na agaroznem gelu. Iz gela smo nato izolirali fragmente.

Za oba gena smo izvedli reakcijo za določanje nukleotidnega zaporedja in tudi samo določanje le-tega na manjši skupini vzorcev DNA hmeljne družine Magnum  $\times$  2/1. Glede na prisotnost domnenih SNP-ov smo izbrali začetne nukleotide, s katerimi smo nato določali zaporedje vzorcem DNA celotne družine. Pri genu za VPS smo odkrili kar 15 polimorfizmov v dolžini 1111 bp, kar je 1 SNP na vsakih 74 bp za to regijo in precej odstopa od pričakovane vrednosti.

Na podlagi najdenih SNP-ov in njihove segregacije pri hmeljni družini Magnum  $\times$  2/1 smo gena kartirali na skupino povezanosti Magnum-8, kar pomeni, da ne gen za VPS ne gen za CHS ne ležita na QTL-u za vsebnost  $\alpha$ -kislin.

Domnevne SNP-e, katere smo našli z določanjem zaporedja produktov PCR gena za VPS, smo primerjali z EST-ji v gruči za VPS in ugotovili, da je več kot polovico SNP-ov mogoče določiti že iz poravnav EST-jev v gruče, kar pospeši in poceni kartiranje.

## 7 VIRI

Adams M. D., Kelley J. M., Gocayne J. D., Dubnick M., Polymeropoulos M. H., Xiao H., Merril C. R., Wu A., Olde B., Moreno R. F., Kerlavage A. R., McCombie W. R., Venter J. C., Go 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252: 1651-1656

BLAST 2.2.17. 2008. National Center for Biotechnology Information.  
<ftp://ftp.ncbi.nih.gov/blast/executables/LATEST/> (1. mar 2008)

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

Brady J. L., Scott N. S., Thomas M. R. 1996. DNA typing of hops (*Humulus lupulus*) through application of RAPD and microsatellite marker sequences converted to sequence tagged sites (STS). *Euphytica*, 91: 277-284

Chambers G. K., MacAvoy E. S. 2000. Microsatellites: consensus and controversy. Comparative Biochemistry and Physiology - Part B, 126: 455-576

Chen Y., Lin C., Wang C., Wu H., Hwang P. 2007. An optimized procedure greatly improves EST vector contamination removal. *BMC Genomics*, 8: 416: 11 str.

Crick F. 1970. Central Dogma of Molecular Biology. *Nature*, 227: 561-563

Čerenak A., Satovic Z., Javornik B. 2006. Genetic mapping of hop (*Humulus lupulus L.*) applied to the detection of QTLs for alpha-acid content. *Genome*, 49: 485-494

DFCI Gene Indices Software Tools. Computational Biology and Functional Genomics Laboratory, Dana Farber Cancer Institute.  
<http://compbio.dfci.harvard.edu/tgi/software/> (1. mar. 2008)

Edwards K. J., Mogg R. 2001. Plant genotyping by analysis of single nucleotide polymorphisms. V: Plant genotyping: The DNA fingerprinting of plants. Henry R. J. (ed.). New York, CABI: 1 – 5

Field B., Osbourn A. E. 2008. Metabolic diversification—Independent assembly of operon-like gene clusters in different plants. *Science*, 320, 5875: 543-547

Gerhauser C , Alt A., Heiss E., Gamal-Eldeen A., Klimo K., Knauf J., Neumann I., Scherf H., Frank N., Bartsch H., Becker H. in sod. 2002. Cancer chemopreventative activity of xanthohumol, a natural product derived from hop. *Molecular Cancer Therapeutics*, 1, 11: 959-969

Goldstein D. B., Schlötterer C. 1999. Microsatellites – evolution and applications. Oxford, Oxford University Press: 352

- Griffiths A., Wessler S. R., Lewontin R. C., Gelbart W. M., Suzuki D. T., Miller J. H.  
2005. Introduction to genetic analysis, 8th edition. New York, W. H. Freeman: 782 str.
- Hadonou A. M., Walden R., Darby P. 2004. Isolation and characterizaton of polymorphic  
microsatellites for asessment of genetic variation of hops (*Humulus lupulus L.*).  
Molecular Ecology Notes, 4: 280-282
- Huang X., Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome  
Research, 9: 868-877
- Jakše J. 2003. Izolacija mikrosatelinih lokusov na kromosomski DNA hmelja (*Humulus lupulus L.*) z obogatitvenim postopkom. Doktorska disertacija, Ljubljana, Univerza v Ljubljani: 141 str.
- Jakše J., Bandelj D., Javornik B. 2002. Eleven new microsatellites for hop (*Humulus lupulus L.*) Molecular Ecology Notes, 2: 544-546
- Jakše J., Luthar Z., Javornik B. 2008. New polymorphic dinucleotide and trinucleotide  
microsatellite loci for hop *Humulus lupulus L.* Molecular Ecology Notes, 8, 4: 769-772
- Jeltsch J. 2008. »1554 hmeljnih EST-jev«. (osebni vir, februar 2008)
- UniVec. 2008.  
<ftp://ftp.ncbi.nih.gov/pub/UniVec/> (1. mar. 2008)
- Koie K., Inaba A., Okada Y., Kaneko T., Ito K.. 2005. Construction of the genetic linkage  
map and QTL analysis on hop (*Humulus lupulus L.*). V: Proceedings of the first  
international humulus symposium. Hummer K. E. in Henning J. A. (ed.). Leuven,  
Belgija, International Society for Horticultural Science: 59-65
- Kosambi, D. D. 1944. The estimation of map distance from recombination values. Annals  
of Eugenics, 12: 172-175
- Lamperti E. D., Kittelberger J. M., Smith T. F., Villa-Komaroff L. 1992. Corruption of  
genomic databases with anomalous sequence. Nucleic Acid Research, 20, 11: 2741-2747
- Larkin M. A., Blackshields G., Brown N. P., Chenna R., McGgettigan P. A., McWilliam H.,  
Valentin F., Wallace I. M. , Wilm A., Lopez R., Thompson J. D., Gibson T.J., Higgins  
D.G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics, 23, 21, 2947-2948
- Malde K., Jonassen I. 2008. Repeats and EST analysis for new organisms. BMC  
Genomics, 9, 23 :7 str.
- Matoušek J., Novák P., Patzak J., Bříza J., Krofta K. 2002a. Analysis of true chalcone  
synthase from *Humulus lupulus L.* and biotechnology aspects of medicinal hops.  
Rostlinná Výroba, 48, 1: 7-14

- Matoušek J., Novák P., Bříza J., Patzak J., Niedermeierová H. 2002b. Cloning and characterization of *chs*-specific DNA and cDNA sequences from hop (*Humulus lupulus* L.). *Plant Science*, 162, 1007-1018
- Mignone F., Gissi C., Liuni S., Pesole G. 2002. Untranslated regions of mRNAs. *Genome Biology*, 3, 3:reviews0004.1-0004.10
- Milligan S. R., Kalita J. C., Heyerick A., Rong H., De Cooman L., De Keukeleire D. 1999. Identification of a potent phytoestrogen in hops (*Humulus lupulus* L.) and beer. *Clinical Endocrinology & Metabolism*, 84, 6: 2249-2252
- Miranda C. L., Stevens J. F., Helmrich A., Henderson M. C., Rodriguez R. J., Yang Y. H., Deinzer M. L., Barnes D. W., Buhler D. R.. 1999. Antiproliferative and cytotoxic effects of prenylated flavonoids from hop (*Humulus lupulus*) in human cancer cell lines. *Food and Chemical Toxicology*, 37, 4: 271-285
- Murakami A., Darby P., Javornik B., Pais M. S. S., Seigner E., Lutz A., Svoboda P. 2006. Molecular phylogeny of wild hops, *Humulus lupulus* L. *Heredity*, 97, 1: 66-74
- Nagaraj S. H., Gasser R. B., Ranganathan S. 2006. A hitchhiker's guide to expressed sequence tags (EST) analysis. *Briefings in Bioinformatics*, 8, 1: 6-21
- Nagel J., Culley L. K., Lu Y., Liu E., Matthews P. D., Stevens J. F., Page J. E. 2008. EST analysis of hop glandular trichomes identifies an O-methyltransferase that catalyzes the biosynthesis of xanthohumol. *The Plant Cell*, 20: 186-200
- Novák P., Matoušek J., Bříza J. 2003. Valerophenone synthase-like chalcone synthase homologues in *Humulus lupulus*. *Biologia plantarum*, 46, 3: 375-381
- Novák P., Krofta K., Matoušek J. 2006. Chalcone synthase homologues from *Humulus lupulus*: some enzymatic properties and expression. *Biologia Plantarum*, 50, 1: 48-54
- Okada Y., Ito K. 2001. Cloning and analysis of valerophenone synthase gene expressed specifically in lupulin gland of hop (*Humulus lupulus* L.). *Bioscience, Biotechnology and Biochemistry*, 65, 1: 150-155
- Okada Y., Sano Y., Kaneko T., Abe I., Noguchi H., Ito K.. 2004. Enzymatic reactions by five chalcone synthase homologs from hop (*Humulus lupulus* L.). *Bioscience, Biotechnology and Biochemistry*, 68, 5: 1142-1145
- Paniego N. B., Zuurbier K. W. M., Fung S., van der Heijden R., Scheffer J. J. C., Verpoorte R. 1999. Phlorisovalerophenon synthase, a novel polyketide synthase from hop (*Humulus lupulus* L.) cones. *European Journal of Biochemistry*, 262, 2: 612-616
- Patzak J. 2001. Comparison of RAPD, STS, ISSR and AFLP molecular methods used for assessment of genetic diversity in hop (*Humulus lupulus* L.). *Euphytica*, 121: 9-18

- Patzak J., Vrba L., Matoušek J. 2007. New STS molecular markers for assessment of genetic diversity and DNA fingerprinting in hop (*Humulus lupulus L.*). *Genome*, 50, 1: 15-25
- Pertea G., Huang X., Liang F., Antonescu V., Sultana R., Karamycheva S., Lee Y., White J., Cheung F., Parvizi B., Tsai J., Quackenbush J. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 19, 5: 651-652.
- Rozen S. in Skaletsky H. J. 2000. Primer3 on the WWW for general users and for biologist programmers. V: *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Krawetz S. in Misener S. (ed.). Totowa, NJ, Humana Press: 365-386
- Rudd S. 2003. Expressed sequence tags: alternative or complement to whole genome sequences? *Trends in Plant Science*, 8, 7: 321-329
- Seefelder S., Ehrmaier H., Schweizer G., Seigner E. 2000. Male and female genetic linkage of hops, *Humulus lupulus*. *Plant Breeding*, 119, 3: 249-255
- Seluja A. G., Farmer A., McLeod M., Harger C., Schad P. A. 1999. Establishing a method of vector contamination identification in database sequencing. *Bioinformatics*, 15, 2: 106-110
- Schröder J., Schröder G. 1990. Stilbene and chalcone synthases: related enzymes with key functions in plant-specific pathways. *Zeitschrift für Naturforschung*, 45c: 1-8
- Sugiyama R., Oda H., Kurosaki F. 2006. Two distinct phases of glandular trichome development in hop (*Humulus lupulus L.*). *Plant Biotechnology*, 23: 439-496
- Štajner N., Jakše J., Kozjak P., Javornik B. 2005. The isolation and characterization of microsatellites in hop (*Humulus lupulus L.*). *Plant Science*, 168: 213-221
- Thiel, T. MISA - Microsatellite identification tool. 2002. Plant Genome Resources Center, Leibniz Institute of Plant Genetics and Crop Plant Research (14. maj 2002). <http://pgrc.ipk-gatersleben.de/misa/> (3. junij 2008)
- Tian L., Wan S., Pan Q., Zheng Y., Huang W. 2008. A novel plastid localization of chalcone synthase in developing grape berry. *Plant Science* 175: 431–436
- Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., Higgins D. G. 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acid Research*, 25, 24: 4876-4882
- Van Ooijen J. W., Voorrips R. E. 2001: JoinMap Version 3.0, Software for the calculation of genetic linkage maps. Wageningen, Nizozemska, Plant Research International: 51 str.

Wang J. Z., Lindsay B. G., Leebens-Mack J., Cui L., Wall K., Miller W. C., dePamphilis C. W. 2004. EST clustering error evaluation and correction. *Bioinformatics*, 20, 17: 2973-2984

## ZAHVALA

Najlepše se zahvaljujem mentorju doc. dr. Jerneju Jakšetu za številne nasvete, vse posredovano znanje in potrpljenje.

Za pregled dela se zahvaljujem članu komisije izr. prof. dr. Gregorju Anderluhu.

Zahvala gre še prof. dr. Jean-Marcu Jeltschu (Univerza Louis Pasteur v Strasbourg) za dobljena zaporedja EST pred objavo in prof. dr. Zlatku Šatoviću (Univerza v Zagrebu) za pomoč pri kartirjanju.

Za pomoč v trenutkih zmede in družbo med čakanjem na rezultate se zahvaljujem zaposlenim na Katedri za genetiko, biotehnologijo in žlahtnjenje rastlin, še posebej pa Tjaši Gril.

Za pomoč pri reševanju ugank linuxa, potrpežljivost, razumevanje in vzpodbudo pri delu se zahvaljujem Boštjanu. Iskrena zahvala pa gre tudi staršem.

PRILOGA A

Frekvenca tipov mikrosatelitov v analiziranih soseskah, upoštevajoč komplementarnost zaporedij.

## PRILOGA B

Genotipske vrednosti lokusa CHS in VPS za družino Magnum × 2/1, ki služijo kot vhodne vrednosti za program JoinMap. Vrednost lm pomeni, da je rastlina heterozigot, vrednost ll pa, da je homozigot.

Št. vzorca	Rastlina	CHS	VPS
	magnum (mati)	lm	lm
	2/1 (oče)	ll	ll
1	279/1	lm	ll
2	279/2	ll	ll
3	279/3	ll	ll
4	279/4	ll	lm
5	279/5	lm	ll
6	279/7	lm	ll
7	279/8	ll	ll
8	279/9	ll	lm
9	279/10	ll	lm
10	279/11	lm	lm
11	279/12	lm	ll
12	279/13	--	--
13	279/14	ll	lm
14	279/15	lm	lm
15	279/16	ll	ll
16	279/18	--	--
17	279/19	--	--
18	279/20	ll	ll
19	279/22	lm	ll
20	279/23	--	--
21	279/24	lm	ll
22	279/25	lm	ll
23	279/27	lm	ll
24	279/28	ll	ll
25	279/29	ll	lm
26	279/30	ll	lm
27	279/31	lm	ll
28	279/32	lm	lm
29	279/33	lm	lm
30	279/34	lm	ll
31	279/35	ll	ll
32	279/36	lm	ll
33	279/37	ll	lm
34	279/38	--	--
35	279/39	lm	ll
36	279/40	lm	ll
37	279/41	--	--
38	279/42	lm	ll

Se nadaljuje.

Nadaljevanje.

Št. vzorca	Rastlina	CHS	VPS
39	279/43	--	--
40	279/44	lm	ll
41	279/45	--	--
42	279/46	lm	ll
43	279/47	lm	ll
44	279/48	ll	ll
45	279/49	ll	lm
46	279/50	lm	ll
47	279/51	ll	lm
48	279/52	--	--
49	279/53	ll	lm
50	279/54	ll	lm
51	279/56	lm	lm
52	279/57	lm	lm
53	279/58	ll	lm
54	279/59	ll	ll
55	279/60	ll	ll
56	279/61	lm	ll
57	279/62	ll	lm
58	279/63	lm	ll
59	279/64	ll	ll
60	279/65	ll	lm
61	279/66	ll	lm
62	279/67	ll	lm
63	279/68	lm	lm
64	279/69	lm	ll
65	279/70	ll	lm
66	279/72	ll	ll
67	279/73	ll	ll
68	279/75	lm	ll
69	279/76	lm	ll
70	279/77	lm	lm
71	279/80	ll	ll
72	279/84	lm	ll
73	279/85	ll	lm
74	279/87	ll	lm
75	279/89	ll	lm
76	279/92	lm	lm
77	279/93	lm	lm
78	279/94	lm	ll
79	279/96	ll	lm
80	279/101	lm	lm
81	279/103	lm	ll
82	279/104	ll	ll
83	279/106	--	--

Se nadaljuje.

### Nadaljevanje.

Št. vzorca	Rastlina	CHS	VPS
84	279/107	lm	ll
85	279/110	ll	ll
86	279/111	--	--
87	279/112	ll	lm
88	279/113	lm	ll
89	279/115	--	--
90	279/116	ll	lm
91	279/120	lm	ll
92	279/121	--	--
93	279/122	ll	ll
94	279/123	lm	ll
95	279/124	ll	lm
96	279/125	lm	lm
97	279/126	ll	lm
98	279/91 - m	ll	lm
99	279/100 - m	lm	ll
100	279/102 - m	lm	ll
102	280/30 - m	lm	ll
103	280/31 - m	--	--
104	280/33 - m	lm	ll
105	280/34 - m	--	--
108	280/45 - m	--	--
110	280/49 - m	--	--
111	280/50 - m	--	--
112	280/52 - m	--	--
113	280/53 - m	lm	ll
114	280/54 - m	--	--
115	280/55 - m	--	--