

Tartu Ülikool
Molekulaar- ja Rakubioloogia Instituut
Bioinformaatika õppetool

Kessy Abarenkov

**Liigispetsiifilised oligonukleotiidid resupinaatsete lehternahkiseliste
ITS järjestuste põhjal.**

Lõputöö

Juhendajad: prof. Urmas Kõljalg, prof. Mairo Remm

Tartu 2004

Sisukord

1. Sisukord	2
2. Sissejuhatus	3
2.1 Seeneliikide identifitseerimisest	3
2.2 Oligonukleotiidide disain	7
3. Materjalid ja meetodika	10
3.1 Kasutatud materjalid ja tarkvaraprogrammid	10
3.2 Oligote disainimise programmid	11
3.3 Disainimise meetodika	14
4. Tulemused	16
4.1 Disainitud oligod	16
4.2 Oligod sõltuvalt programmist	17
4.3 Oligod programmiga Probesel	17
5. Tulemuste arutelu	19
Kokkuvõte	22
Summary	23
Lühendid	
Kasutatud kirjandus	
Tabelid, joonised ja lisad	

Sissejuhatus

Seeneliikide identifitseerimisest

Teadusele uute seeneliikide kirjeldamiseks kasutatakse tänapäevani peamiselt fenotüübilisi tunnuseid. Botaanilise nomenklatuuri koodeks (*Greuter jt., 2000*), mis reguleerib ka seeneliikide kirjeldamist ja nimetamist, ei takista aga genotüübi tunnuste kaasamist uue taksoni kirjeldamisse. Seetõttu on uute seeneliikide kirjeldamiseks hakatud lisaks fenotüübilistele tunnustele kasutama ka tüüpliigi teatud lookuse nukleotiidsed järjestust (*Kõljalg ja Larsson, 1998*). Nukleotiidsete järjestuste abil koostatud liigikirjeldused võimaldavad hiljem keskkonna vm. proovidest pärit seene DNA-d määrata kasutades fülogeneetilist liigikontseptsiooni (*Taylor jt., 2000*). Kuna liigi nimi on seotud ainult ühe eksemplariga, s.o. tüübiga, siis on liigi molekulaarseks kirjeldamiseks oluline ka tüübi vastava lookuse sekveneerimine. Vastasel juhul ei saa olla kindel, et liigi nime kandev tüüpeksemplar ning liigi molekulaarseks kirjeldamiseks kasutatud eksemplarid kuuluvad samasse taksonisse. Seeneliikide tüüpeksemplarid on aga sageli 100-200 aastat vanad herbaareksemplarid, mistõttu DNA eraldamine ja teatud lookuse amplifitseerimine enamasti ebaõnnestub. Sellisel juhul on liigi molekulaarseks kirjeldamiseks oluline kasutada spetsialisti poolt määratud seente viljakehi.

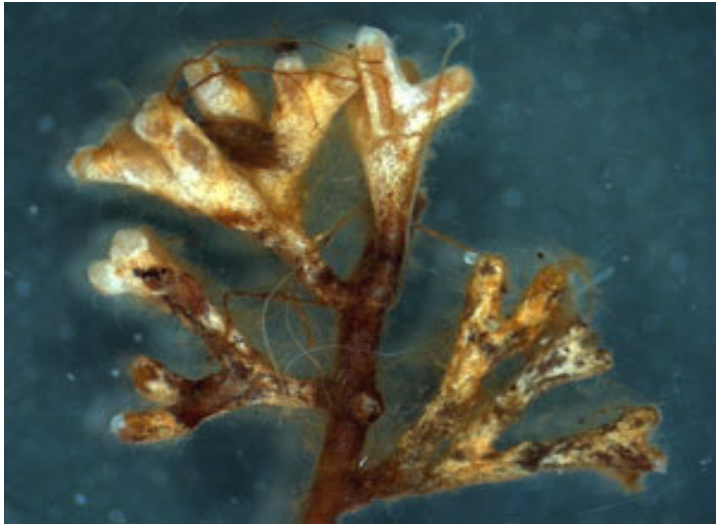
Molekulaarsete meetodite (PCR - polymerase chain reaction, DGGE - temperature-gradient gel electrophoresis, T-RFLP - terminal restriction fragment length polymorphism) kasutuselevõtt lubas organismide uurimiseks kasutada ribosomaalse RNA geeni (rDNA) või rRNA-d ennast, võimaldades eristada liike, mis muidu olid üheks arvatud. Sellega kaasnev järjestuste andmebaaside (EMBL, GenBank) teke ja areng on teinud võimalikuks võrrelda järjestusi üle maailma. See on oluliselt muutnud seente jt. mulla- ning veorganismide uurimist kuna nüüd on võimalik neid täpselt ja suhteliselt kiiresti identifitseerida. Peamiseks puuduseks on aga taksonite ebaühtlane esindatus ning valesti määratud liikide suur arv avalikes andmebaasides.

Nukleotiidsete järjestuste kasutamine liikide määramiseks on viimastel aastatel oluliselt mõjutanud ka taimejuurte mükoriisat moodustavate seente uurimist.

Mükoriisaks nimetatakse sümbioosi taimejuurte ja pinnast asustavate seente vahel. Seda iseloomustab mõlemasuunaline toitainete liikumine, kus seened omistavad taimedelt saadud süsinikku, varustades neid vastutasuks anorgaaniliste toitainete ning veega. Umbes 75% vaskulaarsetest taimeliikidest moodustavad mükoriisat. Mükoriisat moodustavad taimed on võrreldes teistega konkurentsivõimelisemad ning stressitaluvad, samas soodustavad nad seenesümbiontide ning teiste mükorisosfääri kuuluvate organismide kasvu. Mükoriisat moodustavad seened omavad tähtsat rolli taimedevahelisel ressursside jagamisel, soodustades

nende koeksistentsi looduses (*van der Heijden jt., 2003*).

Ektomükoriisa (EM) on mükoriisa spetsiaalne vorm, mis domineerib peamiselt puujuurtel boreaalsetes (põhjapoolkera-) ökosüsteemides. Teda iseloomustab taime juure väliste rakkude vahel olevate hüüfide olemasolu, mis moodustavad nn. Hartigi võrgustiku, ning paljudel juhtudel ka seenkoest moodustunud kesta või mantli olemasolu, mis võib juure täielikult katta. Mantlist tulenevad hüüfid, mis laienevad pinnasesse (vt. joonis1).



Joonis 1. *Tomentellopsis zygodesmoides*, autor: Urmas Kõljalg

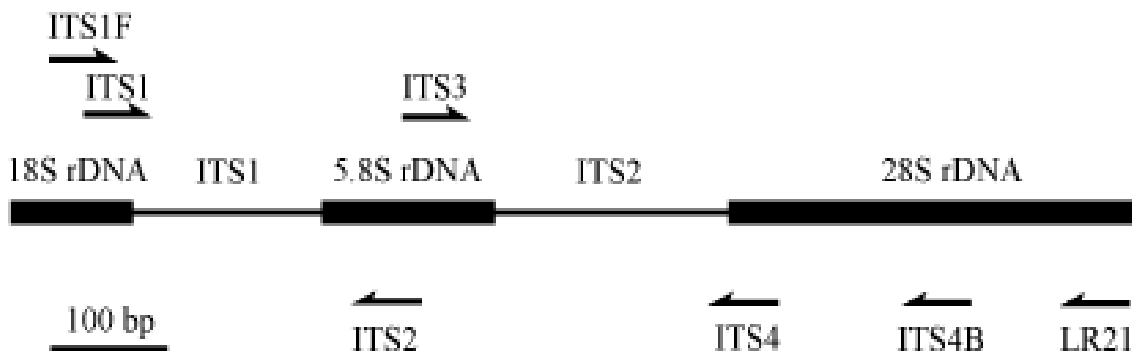
Peremeestaimedeks osutuvad tihti perekondadesse *Pinaceae* (männilised), *Fagaceae* (pöögilised), *Betulaceae* (kaselised) ja *Myrtaceae* (mürdilised) kuuluvad liigid. Seeneliikidest (peamiselt kandseened) üle 4000 teatakse moodustavat ektomükoriisat. Peremeestaime spetsiifilisus on varieeruv: nt. *Boletus betulicola* (kõivu-kivipuravik) vaid perekonna *Betula* liikidel või *Pisolithus arhizus* (nimetatakse ka *P. tinctorius*, eestikeelne nimi teadmata), mis moodustab ektomükoriisat 46 puuliigiga 8-st erinevast sugukonnast (*Sylvia D.M.*)

EM moodustavate seeneliikide identifitseerimiseks ja eristamiseks üksikutel mükoriisadel on kasutatud DNA järjestuste ja PCR-RFLP analüüsi (*Bruns jt., 1998, Kõljalg jt., 2000*). Kuigi seni on EM moodustavaid seeni kirjeldatud piiratud hulk, tõuseb see number uute liikide avastamisega pidevalt. On näidatud, et RFLP analüüsil on ka mõningaid puudusi. Näiteks jäävad mitmed eksemplarid identifitseerimata, kuna puudub avalik info just selle kindla RFLP tüübi kohta, ning mustri vähene info ei võimalda vahet teha ega analüüsida sarnase mustriga järjestusi. Seetõttu on oluline kaasata nende järjestuste sekveneerimine ning sekvenside võrdlemine (*Horton, 2002*). Antud meetod on oma töötappide (restriktsiooniensüümide lagundamine, produktide geelelektroforees) tõttu aeganõudev, võimaldades korraga analüüsida piiratud hulka proove. Nimetatud tehnika tundlik ka saastumise vastu, ning alati pole võimalik

PCR-i produktide väikest hulka geelil määrata.

Nii mikroobide, taimede kui seente koosluse ja fülogeneesi uurimiseks on viimasel ajal tihti kasutatud ribosomaalse RNA geene (*Alvarez jt., 2002, Kõljalg jt., 2000, 2002, Oh jt., 2002, Valinski jt., 2002*), selle organisatsioon seentel on toodud joonisel 2. Nende mõned eelised teiste geenide ees on:

- universaalsete praimerite olemasolu. Kuna ribosoomi subühikute RNA-d kodeerivad alad on väga konserveerunud, ei teki probleeme praimerite leidmisel nende ülesamplifitseerimiseks;
- järjestuste sobiv pikkus (ITS1-5.8S-ITS2 - 500-700 aluspaari) ja suur korduste arv genoomis;
- väike genoomisisene varieeruvus. Selle põhjuseks arvatakse olevat nn. kooskõlastatud evolutsiooni (*concerted evolution*, ingl. k.), kus erinevad geenikoopiad genoomis muudetakse järjestuselt sarnaseks sagedase ebavõrdse ristisirde ja geenikonversiooni tulemusena;
- genoomidevaheline varieeruvus. On näidatud, et ITS järjestused, mida on kasutatud ka käesolevas töös, on piisavalt varieeruvad liikide eristamiseks nende põhjal (*Grönberg jt., 2003*). Nad pole siiski sobilikud kõrgemal (sugukonna) tasemel liikide identifitseerimiseks, mille puhul kasutatakse konserveerunumaid mitokondriaalse või tuuma DNA regioone (*Bruns jt., 1998*).



Joonis 2. rDNA struktuur seentel

Paksu joonega on näidatud ribosoomi subühikuid kodeerivad alad 18S, 5.8S, 28S, peenema joonega nende vahel asuvad mittetransleeritavad alad ITS1 ja ITS2. Ära on toodud ka geeni amplifitseerimiseks kõige laiemalt kasutatavad PCR praimerid. Joonis pärineb internetiaadressilt: <http://unite.zbi.ee/primers.php3>.

Paari viimase aastaga on molekulaarsete meetodite hulgas intensiivselt levima hakanud mikrokiipide (*microarray*, ingl. k.) kasutamine geeniekspressioonianalüüsiks, bakterikoosluse kindlaks-määramiseks ning ka liikide identifitseerimiseks (*Fukushima jt., 2003, Pfunder jt., 2004*). Selle meetodi eelised näiteks RFLP (restriction fragment length

polymorphism) ees on väiksem aja- ja rahakulu. Üheks oluliseks osaks mikrokiibi loomisel liikide määramiseks on iga liigi (perekonna, rühma) jaoks unikaalsete oligote disain, mis annaks signaali vaid oma target-järjestusega seondudes. Proovide disainiga on seotud mitmed aspektid ja probleemid, millest käesolevas töös ka lähemalt juttu tuleb.

Töös vaadeldud seeneliike (perekonnad *Thelephora*, *Tomentella*, *Tomentellopsis*) on uurinud Urmas Kõljalg (2000, 2002), ning näidanud, et antud perekondadesse kuuluvad liigid on olulise tähtsusega EM moodustavad seened, mõjudes otseselt puude kasvule ja metsa produktiivsusele. Nende täpne identifitseerimine on vajalik ökoloogia-alastes uurimustes.

Antud perekondade fülogeneesianalüüsiks ja liikide identifitseerimiseks on kasutatud rDNA ITS (sisemised transkribeeritud vahemikud, *Internal Transcribed Spacer*, ingl. k.) järjestusi. Kuna suured avalikud andmebaasid nagu NCBI ja EMBL sisaldavad järjest enam vale taksoni nime kandvaid järjestusi ja puudub vajalik täpsem informatsioon konkreetse eksemplari kohta, oli vajalik luua lokaalne antud seenerühmade ITS sekventse sisaldav andmebaas, mis täidaks lüngad eelpool nimetatud andmebaasides. Antud andmebaasi loomist alustati 2002. aastal.

UNITE (a User-friendly Nordic ITS Ectomycorrhiza Database) on ektomükoriisat moodustavatele seentele spetsialiseeruv rDNA järjestuste andmebaas. Algselt hõlmab andmebaas vaid ITS regiooni järjestusi. Need on sekveneeritud avalikes herbariumites säilitatavatest viljakehadest, mis on korjatud ja määratud oma ala spetsialistide poolt. Lisaks järjestustele sisaldab andmebaas ka eksemplari- ja liigikirjeldusi ning illustratsioone. Järjestuste sarnasusotsingud UNITE vastu sooritatakse kasutades BLAST tarkvara, tulevikus on plaanis lisada fülogeneesil põhinevad otsingusüsteemid.

UNITE on MySQL-Linux platvormile ehitatud relatsiooniline andmebaas, millega luuakse ühendus PHP (PHP: Hypertext Preprocessor) ja perli (Practical Extraction and Report Language) skriptide abil. Hetkel (seisuga 22.05.04) on andmebaasis 24 tabelit ning 551 järjestust.

UNITE andmebaas on nähtaval internetiaadressil <http://unite.zbi.ee>.

Töö üheks eesmärgiks on välja selgitada, kas resupinaatsete lehternahkiseliste ITS järjestused võimaldavad luua liigispetsiifilisi oligonukleotiide, et valmistada liikide määramiseks mikrokiip. Kui mitte, siis milliseid DNA regioone saaks selleks otstarbeks kasutada. Töö teiseks eesmärgiks on katsetada erinevaid oligonukleotiidide disainimise programme ning leida nendest sobivaim eelpool mainitud ülesande lahendamiseks.

Oligonukleotiidide disain

Viimase paari aasta jooksul on üha enam populaarsust kogunud molekulaarbioloogia uus meetod, mikrokiip, mis on üldine lähenemisviis mitmete erinevate bioloogiliste probleemide lahendamiseks. Mikrokiip kujutab endast väikeste mõõtmetega klaasplaadile viidud spetsiifiliste DNA proovide hübridiseerimist uuritava target-DNA-ga, mis on näiteks fluorestseeruvalt märgistatud. Järgneb iga seondunud proovi signaali intensiivsuse määramine skänneriga. Antud tehnoloogia kasuks räägivad mõned eelised tavalise hübridisatsiooni ees: kõrgem lahutusvõime, tundlikkus, kiirem määramine, odavus, automat-sus.

Seni on mikrokiibi tehnoloogiat rakendatud järgmistes valdkondades:

- geeniekspressiooni analüüs, nt. normaalse ja kasvajakoe võrdlemine, uute geenide detekteerimine;
- liikide (nt. patogeenide) identifitseerimine, mikroorganismide koosluse kindlakstegemine keskkonnaproovidest;
- mutatsioonide detekteerimine, polümorfismide esinemine genoomses DNA-s.

Mikrokiipe võib jagada omaette klassidesse, näiteks koosluse genoomi kiip (CGA, Community Genome Array), funktsionaalse geeni kiip (FGA, Functional Gene Array) ja fülogeneetiline oligonukleotiidide kiip (POA, Phylogenetic Oligonucleotide Array) mikroorganismide puhul (Zhou, 2003). CGA-sid kasutatakse iseloomustamiseks mikroobikooslust teatud kultiveerimistingimustes, FGA-d on mõeldud looduslikus keskkonnas olevate mikroobipopulatsioonide füsioloogilise seisundi ja funktsionaalsete tegevuste jälgimiseks bioloogiliselt tähtsaid ensüüme kodeerivate geenide kaudu. Kõige huvipakkavam on aga POA mikroobi-koosluste koostise ja struktuuri analüüsiks kasutades rRNA geenide järjestusi. Antud kiip sobiks hästi ka seeneliikide kindlakstegemiseks kas keskkonnaproovist või PCR produktist. POA võib universaalsete praimerite olemasolul hõlpsasti siduda PCR-ga, mis tõstab tema tundlikkust veelgi. Siiski võivad POA kasutamisel ilmned mõned probleemid. Nimelt, kuna rRNA geenid on kõrgelt konserveerunud ning olemas kõigis mikroorganismides, võivad tekkida raskused erinevuste detekteerimisel. Zhou (2003) on hinnanud ühenukleotiidise valepaardumise (proovi enda pikkus 18-25 nukleotiidi) efekte hübridiseerumisel ning näidanud, et valepaardumise positsioon ja tüüp omavad suurt tähtsust sulamistemperatuurile ja signaaliintensiivsusele. Ühenukleotiidilise valepaardumisega oli signaaliintensiivsus 3 kuni 10 korda väiksem kui täieliku paardumise korral.

Üks olulisemaid küsimusi liikide määramise mikrokiibi valmistamisel ongi hetkel spetsiifiliste oligote disainimine. Optimaalsete proovide olemasolu on tähtis täpsemaks

määramiseks (vähendades taustaefekti) ning kiibi maksumuse vähendamiseks (üks proov mitme asemel ühe liigi/geeni kindlakstegemiseks).

On võimalik kasutada erineva pikkusega oligoid, lühikesi (20-25 nukleotiidi) või pikki (50 või 70 nukleotiidi). Pikad oligod arvatakse olevat kõrgema tundlikkusega, kuid neil on ka suurem risthübridisatsiooni (oligo paardumine vale, targetile sarnase järjestusega) oht (*Kane jt., 2000*).

Ideaalne oligo oleks selline, mis hübridiseeruks minimaalse energiaga antud hübridisatsioonitingimustel vaid oma target-järjestusega ja maksimaliseeriks hübridisatsiooni vabaenergia kõige sarnasema mitte-target-järjestusega (*Li ja Stormo, 2001*). Kahjuks pole vabaenergia hetkel pelgalt järjestuse pealt arvutatav. Näiteks sõltub see target-DNA täielikust struktuurist, mille täpseks ennustamiseks pole praegu piisavalt häid struktuuri ennustamise programme. Samuti sõltub vabaenergia iga proovi ja target-järjestuse vahel nende geenide kontsentratsioonist, veidi kõrgema hübridisatsioonienergiaga ja suurema kontsentratsiooniga DNA võib anda kiibil tugevama signaali ka valepaardumise korral lihtsalt sellepärast, et ta on suuremas hulgas. Erinevate DNA-de hulk proovis pole aga kaugeltki teada, just seda tegelikult uuritaksegi. Seetõttu tehakse vabaenergiat arvesse võtvates oligote disainimise programmides hübridisatsioonienergia arvutamisel mõned lihtsustused:

T_m arvutamise valemis

$$T_m = \Delta H / (\Delta S + R \ln C_t / 4) ,$$

kus ΔH - entalpia muutus, ΔS - entroopia muutus, R – Boltzmanni konstant, C_t - ahelate täielik molaarne kontsentratsioon, C_t (kuna ei ole teada) asendatakse 1×10^{-6} (*Kaderali ja Schliep, 2002, Li ja Stormo, 2001*).

Lisaks hübridisatsioonienergiale, mida ei arvesta kaugeltki mitte kõik oligote disainimise programmid (T_m arvutatakse pelgalt GC sisaldust ja oligo pikkust arvestades), on ka muid tingimusi (*Li ja Stormo, 2001*), mida kasutab nt. Affymetrix oma oligote selekteerimise kriteeriumina:

- GC sisaldus 40 ja 60 vahel;
- ükski nukleotiid ei tohi moodustada üle 50% kogu proovist;
- järjest paiknevate samade nukleotiidide summa ei tohi ületada 25% kogu proovi pikkusest;
- ei tohi olla üle 15 nukleotiidi pikkust pidevat kordust kogu proovi ulatuses;
- proov ei tohi paarduda iseendaga.

Lisaks äratoodud kriteeriumitele on risthübridisatsiooni (*cross hybridization*, ingl.

k.) ja taustasignaali (*background signals*, ingl. k.) vältimiseks väga oluline arvestada valepaardumiste arvu ja asukohta. Oh jt. (2002) ja Zhou (2003) on näidanud, et juba 1 valepaardumise korral on võimalik signaalide erinevusi detekteerida, samas Li ja Stormo (2001) väidavad, et alla 25 nukleotiidi pikkuse proovi korral on risthübridisatsiooni vältimiseks maksimaalne lubatud valepaardumiste arv 4, ning 50- ja 70-meersete oligote jaoks vastavalt 10 ja 20. Valepaardumiste mõju 50-meersete oligote korral varieeruva sarnasusega liikidel on eksperimentaalselt testinud Kane jt. (2000) ning näidanud, et:

- microarray spetsiifilisuse seisukohalt peaks mitte-target-järjestused olema < 75% identsed 50 nukleotiidses regioonis, millele on disainitud oligo;
- kui 50 nukleotiidne regioon on 50-75% ulatuses sarnane, ei tohi ta sisaldada pidevat >15 nukleotiidi pikkust järjestust.

Oligote disainimiseks on viimastel aastatel loodud ja tutvustatud palju programme. Kuna erineva eesmärgiga kiibid (kogu genoomi analüüs, liikide eristamine, ekspresioonianaalüüs) vajavad erinevaid oligoid, tuleb programmide seast endale valida just antud tööks sobiv. Tihti on programmi sisendiks DNA/geenide järjestused ning väljundiks optimaalsed oligod igale järjestusele (nt. OligoArray 2.0 (Rouillard jt., 2003), ProbeSelect (Li ja Stormo, 2001), Probesel (Kaderali ja Schliep, 2002)). Liikide eristamisel on tegemist aga liigi erinevatelt eksemplaridelt saadud DNA järjestustega, ning sobiv programm peab võimaldama disainida oligoid lähtudes mitmest eksemplarist, võttes sisendiks palju järjestusi. Sellisteks programmideks on näiteks PROBEmer (Emrich jt., 2003) ja PRIMROSE (Ashelford jt., 2002).

Materjalid ja meetodika

Kasutatud materjalid ja tarkvaraprogrammid

Uurimiseks ja oligote disainimiseks kasutati perekondadesse *Thelephora*, *Tomentella* ja *Tomentellopsis* kuuluvate liikide eksemplaridelt eraldatud ribosomaalse DNA ITS1-5.8S-ITS2 järjestusi, mis on enamikus esindatud UNITE andmebaasis. Liikide nimekiri, millele prooviti oligoid disainida, on ära toodud lisa 1, nende taksonoomiline kuuluvus on järgmine:

riik - *Fungi*, seened

hõimkond- *Basidiomycota*, kandseened

klass - *Hymenomyces*, eoslavaseened

selts - *Thelephorales*, lehternahkiselaadsed

sugukond - *Thelephoraceae*, lehternahkiselised

perekonnad - *Thelephora*, *Tomentella*, *Tomentellopsis*

Liikide valimiseks, millel testida oligote disainimist, kasutati kahte tingimust:

- liiki pidi esindama vähemalt 3 eksemplari;
- liigi eksemplarid pidid fülogeneesipuul moodustama monofüleetilise(d) rühma(d) ning eristuma teistest rühmadest piisavalt;

Nende tingimuste seadmise tingis asjaolu, et seente ITS järjestused on liigisiselt küllaltki varieeruvad (2-3%), ning vaid 1-2 sekveneeritud eksemplari põhjal disainitud oligod ei kataks kindlasti kogu liigi isendeid, mis veel määramata, pigem annaks nad suure arvu eksemplarispetsiifilisi oligoid, mille väljasorteerimine antud hetkel võimatuks osutuks ning mille kasutamine kiibil oleks ebapraktiline. Võivad ilmned ka taksonoomilised probleemid, kus üks eksemplar on valesti määratud ning paigutatud vale liigi alla. Ka see raskendaks neile oligote disainimist.

Sel teel valiti välja 4 liiki perekonnast *Tomentellopsis*: omaette rühmad *T. zygodesmoides* (3 eksemplariga) ja *T. bresadoliana* (4), ning fülogeneesipuul läbisegi paiknevad *T. echinospora* (7) ja *T. submollis* (6). Viimaste valimine toimus eeldusel, et neile on võimalik disainida ühised oligod.

Perekondade *Tomentella*-*Thelephora* fülogeneesipuult valiti 7 rühma:

- *Tomentella lapida* (5)

- *Tomentella stuposa* (5)
- *Tomentella bryophila* (5)
- *Thelephora americana (caryophyllea)* (5)
- *Tomentella sublilacina* (7)
- *Tomentella ellisii* (4)
- *Thelephora terrestris* (5) ja *Tomentella radiosa* (2)

Oligote disainimisel on kasutatud UNITE andmebaasis olevaid järjestusi (551), nende testimiseks NCBI andmebaasi ning programmi PROBEmer 'oligo check' moodulit. Proovid on disainitud 3 erineva programmiga - PROBEmer, PRIMROSE ja ProbeSel, mille pikem tutvustus on toodud järgmises peatükis. Järjestuste aligneerimiseks PRIMROSE' i jaoks kasutati programmi MAFFT (*Katoh et al. 2002*) 'Fast Fourier transformi' progressiivset meetodit (FFT-NS-2). Fülogeneesipuu (NJ) perekonna *Tomentellopsis* jaoks (lisa 2) ja näidisjärjestuste aligneering (lisa 8) on loodud programmiga ClustalX . Perekondade *Tomentella* ja *Thelephora* NJ puu (lisa 3) tegemisel on kasutatud programmi PAUP HKY85 evolutsioonimudelit.

Oligote disainimise programmid

Töös on oligote disainimiseks kasutatud 3 programmi, mis kõik on ülesandele erineva lähenemisviisiga:

1) **PRIMROSE** - programm fülogeneetiliste oligote disainimiseks.

Ashelford' i jt. (2002) poolt välja töötatud programm, mille põhimõtteks on võimaldada oligote disainimisel lähedastele liikidele kasutada 'degeneratiivseid' nukleotiide, ehk maksimaalselt 2 nukleotiidi, mis võivad eri target-järjestustel varieeruda. Programm on välja töötatud küll bakteriliikidele, võimaldades kasutada andmeid RDP (Ribosomal Database Projekt, <http://rdp.cme.msu.edu/html/>) andmebaasist, kuid lubab kasutamiseks alla-laadida ka oma andmebaasi mitmes erinevas formaadis.

Programm jookseb MS Windows 95/NT/2000 operatsioonisüsteemil, on vabalt kasutatav ning kodulehelt (<http://www.cf.ac.uk/biosi/research/biosoft/Primrose/index.html>) alla-laaditav.

PRIMROSE kasutab ühte kahest algoritmist olenevalt sellest, kas järjestused on aligneeritud või mitte. Aligneeritud järjestustel lubatakse kuni 2 'degeneratiivse' nukleotiidi esinemist.

Algoritm I ise on järgmine:

- luuakse maatriks joondatud target-järjestustega;
- loetakse kokku kõik A-d, T-d, G-d, C-d ja tühikud;
- vastavalt nendele luuakse konsensuspuu;
- eemaldatakse gäpid;
- konsensusjärjestuse pealt võetakse kõik etteantud pikkusega oligod ja sorteeritakse vastavalt 0, 1, 2 'degeneratiivseid' nukleotiide sisalduse järgi.

Algoritm II (joondamata järjestused):

- salvestatakse kõik etteantud pikkusega oligod järjestustest;
- jäetakse alles vaid unikaalsed;
- jäetakse alles vaid need oligod, mis sisalduvad kasutaja poolt määratud minimaalses arvus järjestustes.

Mõlemale algoritmile järgneb unikaalsuse kontroll:

- luuakse otsingustringid, kus degeneratiivsed nukleotiidid asendatakse lihtsa regulaaravaldisega, nt. $M = [AC]$, $R = [AG]$;
- Otsing mitte-target järjestuste vastu andmebaasis. Kui mitte-targetite arv ületab kasutaja poolt määratud arvu, oligo eemaldatakse;
- lõpus antakse kasutajale ülevaade, kui palju targeteid ja mitte-targeteid tabas, ning mis liikidest täpsemalt.

Programm ei kasuta T_m määramisel termodünaamikat, see arvutatakse järgmiselt:

$$T_m = 69,3 + 0,41 \times \%GC - 650/N,$$

kus N - oligo pikkus.

Programmi puuduseks võib lugeda seda, et ta ei arvesta valepaardumiste arvu.

2) **PROBEmer** - veebipõhine tarkvara, mis võimaldab kasutajal valida optimaalseid oligoid PCR aplifikatsiooniks ja microarray' de valmistamiseks.

Emrich jt. (2003) poolt avaldatud programm tagastab kasutajale oligod, mis esinevad tema poolt antud target-grupis, kuid ei esine mitte-target grupi järjestustel. Ta arvutab välja ka lähimad naabrid mitte-target grupis ja näitab nende aligneeringuid võimaliku risthübridisatsiooni kindlakstegemiseks.

Programm kasutab 'suffix array' andmestruktuuri ühiste alamstringide lokaliseerimiseks, seega ei ole piiratud kõrge sarnasusastmega järjestustele. Oligote optimeerimisel ei arvestata vaba-energiaga (kuna see pole autorite sõnul veel piisavalt hästi uuritud).

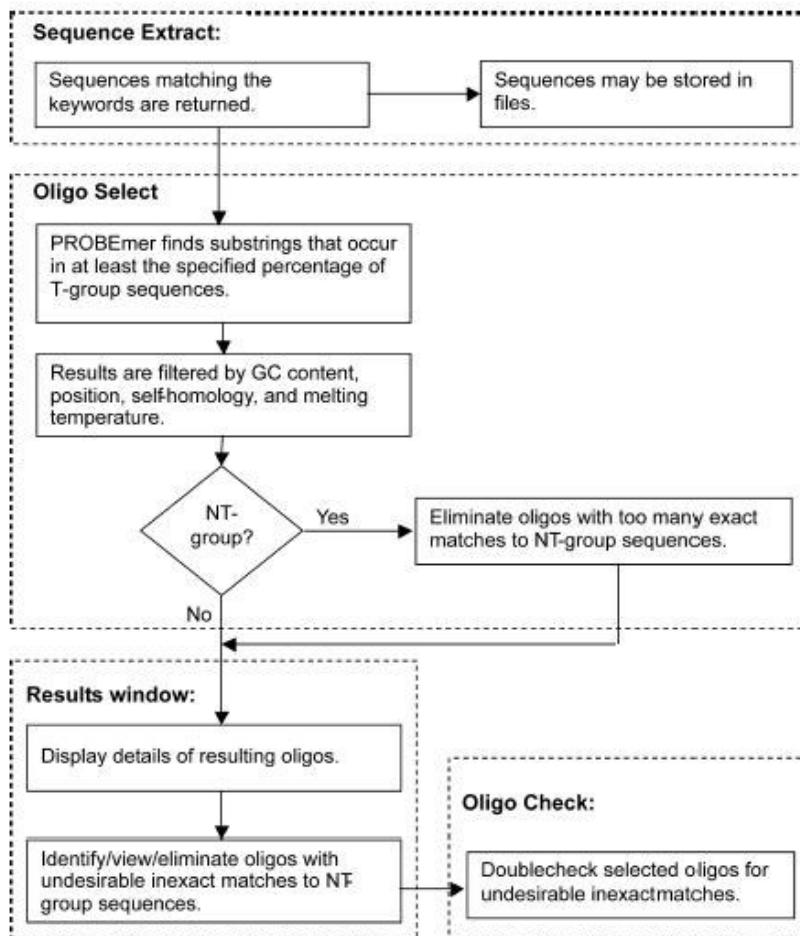


Figure 1. Flowchart showing major steps in PROBEmer.

Joonis 3. Illustratsioon programmi PROBEmer töötappidest (Emrich jt., 2003)

Algoritmi kirjeldus (vt. ka joonis 3):

- leitakse target - grupi järjestustele ühised proovid;
- need filtreeritakse GC sisalduse, T_m ja iseendaga komplementaarsuse suhtes;
- võrreldakse mitte-target-grupi järjestustega. Need proovid, mis esinevad liiga paljudes (arv kasutaja poolt määratud) mt-grupi järjestustes, eemaldatakse;
- allesjäänud proovid ja info nende kohta näidatakse ekraanil.

Ei arvesta samuti hübriidisatsioonienergiat T_m arvutamisel:

$$T_m = (4 \times \%GC) + (2 \times \%AT) - 5 \text{ (proovi pikkus kuni 20 nukleotiidi)}$$

$$T_m = 62,3 + 0,41 \times \%GC - (500/\text{oligo pikkus}) - 5 \text{ (proovi pikkus üle 20 nukleotiidi)}$$

Programmi puuduseks on asjaolu, et minimaalset valepaardumiste arvu mitte-target-grupi järjestustega ei saa ette anda, seega väljastab ta kõik oligod, mis etteantud T_m -ga, GC sisaldusega ja pikkusega.

3) ProbeSel & pickprb - programm proovide disainimiseks mikrokiibi jaoks

Kaderali ja Schliep (2002) avaldasid oligote disainimise programmi, mis võtab disainil arvesse ka vabaenergia olemasolu. Nad eeldavad, et on kahte sorti alustevahelisi interaktsioone: 'base pairing' (tänu vesiniksidemetetele) ning 'base stacking' (hüdrofoobsed mõjud ning Londoni dispersioonijõud), ja et aluste paari stabiilsus sõltub ainult tema eelnevatest ja järgnevatest naabritest. T_m arvutamisel kasutab programm oligote disaini juures ära toodud T_m arvutamise valemit. Programm võimaldab oligo disainida vaid ühest järjestusest lähtudes.

Algoritmi kirjeldus:

- luuakse üldine suffiksipuu;
- leitakse proovid, mis on unikaalsed ja täielikult komplementaarsed oma targetiga;
- filtreerimine (oligo pikkus, hübriidsatsioonitingimused);
- proove hinnatakse edasi kasutades 'Nearest Neighbour' termodünaamilist mudelit. Arvutatakse sulamistemperatuurid kõigi proovi-targeti kombinatsioonidele;
- sekundaarstruktuuri hindamine;
- väljastatakse proovid ja nende sulamistemperatuurid, oligote väljavalimine tuleks teha vastavalt nende tarkvara soovitudele.

Disainimise meetodika

Programmile PROBEmer tuli ette anda järjestused fasta-formaadis, kasutajal oli võimalik määrata minimaalne target-järjestuste arv, s.t. mitmes järjestuses antud oligo minimaalselt sisalduma pidi. Enamikul juhtudel pidi üks oligo kirjeldama vähemalt kahte eksemplari, erandiks nt. *Tomentellopsis zygoesmoides* (oligo pidi sisalduma 100%-s järjestustes), kus järjestused väga sarnased on. Reeglina pidi oligo olema antud grupi järjestuste suhtes unikaalne - ta ei tohtinud sisalduda üheski grupist väljaspool olevas järjestuses.

Programm PRIMROSE genereeris oligod aligneeritud järjestuste failist. Antud juhul ei olnud võimalik määrata minimaalselt kirjeldatavate eksemplaride arvu, seega tuli rühmasisesel jagunemisel puhul teha mitu aligneeringut ja kõiki eraldi töödelda.

Mõlema programmi tulemused filtreeriti järgmiste näitajate suhtes:

- 5 ühesuguse nukleotiidi järjestikune paiknemine. Kõik oligod, mis sisaldasid 5 või enamat sama nukleotiidi järjest, eemaldati edasisest uurimisest;
- GC sisalduse piir 40 – 60 °C . GC sisaldust oli võimalik ette anda vaid programmi PROBEmer puhul;

- iseendaga paardumine ('self complementarity' – võimalike homodimeeride ja juuksenõela struktuuride moodustumise skoor) alla 5. Ka selle võis ette anda vaid programmis PROBEmer.
- 50-meersete oligote puhul eemaldati setist kõik oligod,
 1. mille blasti skoor mitte-targetiga UNITE andmebaasis oli üle 65;
 2. mis sisaldas üle 25 nukleotiidi pikkust ühtivat järjestuse alamstringi.
- 25-meersete oligote puhul eemaldati need proovid,
 1. millel oli vähem kui 5 valepaardumist lokaalses aligneeringus UNITE andme-baasi mitte-target järjestusega;
 2. mis sisaldas üle 17 nukleotiidi pikkust valepaardumisteta alamstringi.

Lõpuks valiti allesjäänud oligote setist need, mille GC % oli vahemikus 48 – 52 (kuna T_m arvutati nendes programmides vaid GC % - le toetudes, oli neil ka sarnane sulamistemperatuur) ning testiti programmi PROBEmer mooduli 'oligo check' abil võimaliku risthübridisatsiooni kindlakstegemiseks.

Programm ProbeSel on mõeldud oligote disainimiseks vaid ühe target-järjestuse põhjal. Programmi töö testimiseks valiti eelnevate programmide tulemustele toetudes 4-st suhteliselt sarnaste järjestustega rühmast igast üks järjestus, eeldades, et sobivate oligote olemasolul sobivad nad ka rühma teistele eksemplaridele. Valituks osutusid järgmised eksemplarid:

Tomentellopsis zygodesmoides – TAA159775

Tomentellopsis bresadoliana – RS04299

Thelephora caryophyllea – TAA172626

Tomentella lapida – TAA181555

Programmile oli võimalik ette anda mitte-target-järjestuste fail, mida oligote genereerimisel arvestatakse, kuid millele endile proove ei disainita. Vastava faili moodustasid UNITE andmebaasi järjestused, millest olid eemaldatud kõik töödeldava järjestusega ühte rühma kuuluvad eksemplarid. Saadud oligotest võeti sarnase hübridisatsioonienergiaga (vahe 1 °C) proovid.

Kõik allesjäänud oligod testiti viimases etapis NCBI andmebaasi vastu, et välistada hübridiseerumine teiste organismide DNA-ga.

Tulemused

Disainitud oligod

Kuigi esialgse parameetrina oli programmide PROBEmer ja PRIMROSE puhul oligote GC sisalduseks määratud 40-60 %, valiti lõplikuks analüüsiks nende seast GC %-ga 48-52 leitud oligod, mis hübriidiseerusid sarnasel temperatuuril. Samuti pidid oligod vastama kriteeriumitele, mis on ära toodud oligote disaini metoodikat kirjeldavas peatükis (vt. lk 14). Antud GC% juures leidsid 25-meersed oligod 9-le eksemplaride rühmale 12-st. Kolmele rühmale ei õnnestunud nimetatud programmidega oligoid disainida (vt. üldine tabel disainitud oligotest arvuliselt, lisa 4). Rühma kõiki eksemplare kirjeldavad oligod leiti liikidele *T. zygodesmoides*, *T. bresadoliana*, *Th. americana (caryophyllea)*, *T. lapida* ning *T. submollis* ja *T. echinospora* ühiselt. *T. zygodesmoides* ja *Th. americana (caryophyllea)* on liigisiselt niivõrd konserveerunud, et neile õnnestus leida kõiki eksemplare kirjeldavad oligod ilma ühegi valepaardumiseta mõne eksemplari suhtes. Teiste puhul leiti oligod, mis kirjeldavad vaid osasid eksemplare (vähemalt kahte) või sisaldavad 'mismatch' nukleotiide (maksimaalselt kahte) target-järjestustega seondumisel (kõik arvulised andmed on ära toodud tabelis lisa 4).

Antud kahe programmi poolt leitud sobivad oligod oli võimalik jaotada gruppidesse oma start-positsiooni järgi järjestustes. Enamike oligote algus jäi vahemikku 75-150, 350-400 või 475-525 (vt. graafik, lisa 7). Positsioonide vahemikud 75-150 ja 475-525 jäävad vastavalt ITS1 ja ITS2 järjestuse keskele, vahemik 350-400 aga moodustab 5.8S kodeeriva ala ülemineku ITS2 järjestuseks (vt. näidisjärjestused erinevate liikide eksemplaridest, lisa 8). Reeglina ongi oligod moodustunud suhteliselt konserveerunud ja variaablist alast kõrvuti. Kirjeldatu tuleb selgelt ilmsiks ka liigile *T. bresadoliana* disainitud oligote juures (vt. 25-meersed oligod, lisa 5), kus sarnaseima mitte-target-järjestuse 'mismatch' positsioonid asetuvad enamikel juhtudel oligo ühte otsa, millele vastandub näiteks oligote [Tom.bres.25_65](#) ja [Tom.bres.25_68](#) puhul 5.8S lõpu konserveerunud ala.

50-meersed oligod õnnestus eelnevalt ära toodud kriteeriume arvesse võttes disainida vaid kahele liigile: *T. zygodesmoides* ja *T. bresadoliana* kahele kõrgelt konserveerunud järjestusele. Pikki oligoid iseloomustab lühikestele sarnane struktuur, kus konserveerunud ala külgneb variaabli alaga, ning start-positsioonid mahuvad lühikestega samadesse piiridesse (vt. 50-meersed oligod liigile *T. bresadoliana*, lisa 5).

Oligod sõltuvalt programmist

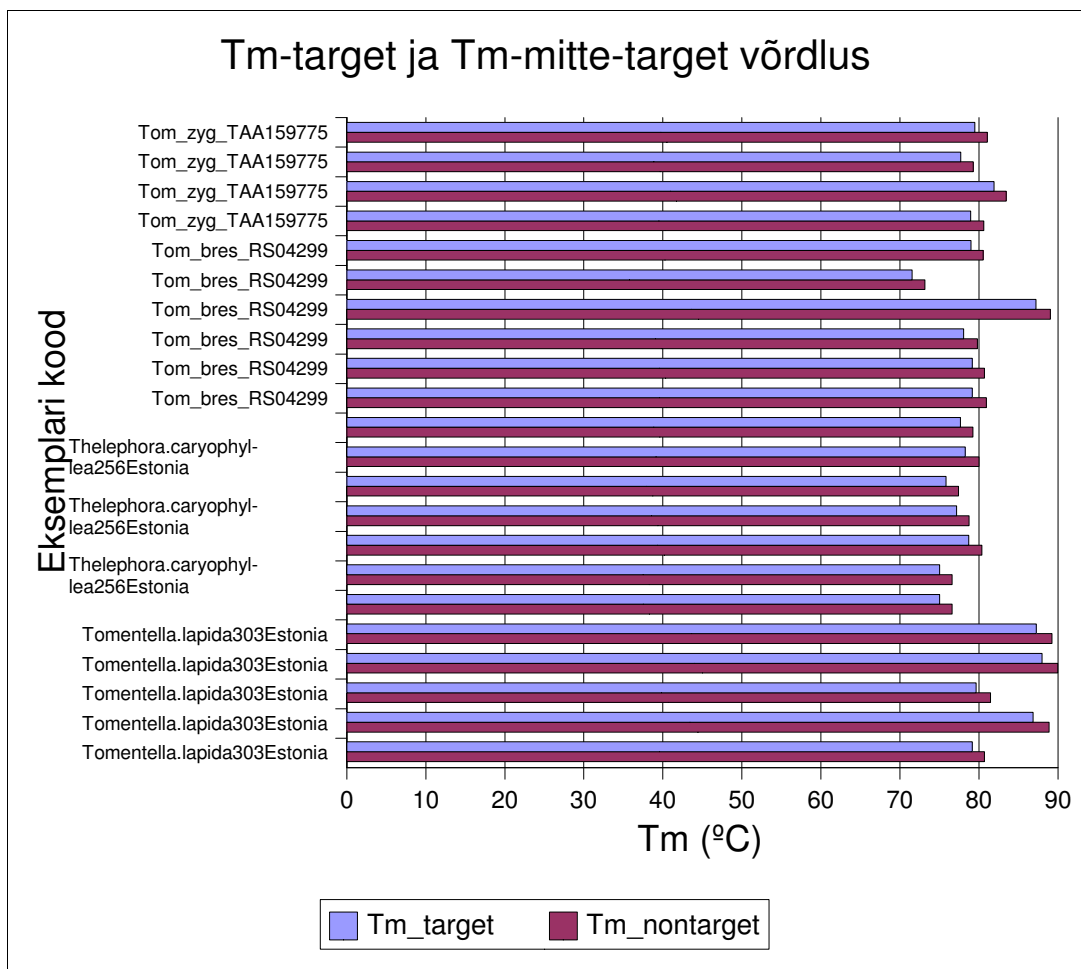
Mõlema programmiga viidi oligote disain läbi samasuguseid parameetrite väärtusi

kasutades, siiski ei olnud tulemused identsed. Kuuel juhul üheksast pakkusid sobivad oligod välja mõlemad programmid, *T. stuposa* ja *T. echinospora* + *T. submollis*' e puhul leidis sobivad oligod vaid PROBEmer ning *T. lapida* puhul vaid PRIMROSE (vt. üldine tabel, lisa 4). Iga rühma puhul, millele mõlemad programmid oligod leidsid, leidsid kahele programmile ühised oligod (vt. halli taustaga oligod *T. bresadoliana*' le, lisa 5). Unikaalsete oligote arv programmi PRIMROSE puhul oli sõltuvuses järjestuste sarnasusest, näiteks leiti antud programmiga rohkem unikaalseid oligoid liikidele *T. zygoesmoides* (14 vs 2) ja *T. americana (caryophyllea)* (5 vs 2) (arvud toodud tabelis lisa 4), mille eksemplaride ITS järjestused on liigi piires küllaltki konserveerunud. Seda seletab asjaolu, et PRIMROSE leiab oligod aligneeritud järjestuste pealt lubades maksimaalselt kahe varieeruva nukleotiidi esinemist. Mida konserveerunud järjestused, seda tõenäolisem on leida ühiseid oligoid. Põhjused, miks sobivad oligod leidis üks või teine programm, on erinevad. Näiteks leidis vaid 1 oligo liigile *T. lapida*, mis kirjeldas nelja eksemplari ning sisaldas ka ühte 'degeneratiivset' nukleotiidi kolmel target-järjestusel. Seega oleks antud oligot olnud võimatu leida PROBEmer programmil, kuna sel juhul oleks oligo pidanud sisalduma vähemalt kolmes järjestuses ning nendega 100%-liselt paarduma. Antud näitest tulebki välja üks olulisi erinevusi kahe programmi vahel. Teisest küljest võib programmi PROBEmer puhul alati vähendada oligo poolt kirjeldatavate järjestuste arvu, kui PRIMROSE' i puhul peab ta sisalduma kõigis järjestustes. See tekitab olukorra, kus üks programm leiab oligo näiteks 2-le eksemplarile, teine programm aga sama oligo 4-le eksemplarile ühe 'mismatch' nukleotiidiga. Antud juhul on vahe mitesobivate lisaoligote tekkes, mida esimesel juhul kindlasti suurem number oleks.

Oligod programmiga ProbeSel

Kuna ProbeSel võimaldab disainida oligod igale järjestusele individuaalselt, pidades kõiki teisi potentsiaalseteks mitte-target-järjestusteks, testiti teda 4 järjestusega, mis valiti välja eelmainitud kahe programmi tulemustele toetudes. Eeldati, et heade oligote korral sobivad nad ka liigi teistele eksemplaridele. Programm leidis oligod igale etteantud järjestusele, mis olid minimaalselt ühe valepaardumisega ning kirjeldasid vähemalt kahte samasse rühma kuuluvat eksemplari maksimaalselt ühe varieeruva nukleotiidiga. Proovide hübridisatsiooni-energiate vahe target-järjestuse ja sarnaseima mitte-target-järjestuse vahel oli enamasti < 2 . Nende hulgast valiti välja oligod, mille T_m -de vahe oli vähemalt $1,5\text{ }^{\circ}\text{C}$ (vt. joonis 4.) Oligod T_m vahemikus $78.25 - 79.16\text{ }^{\circ}\text{C}$ on toodud lisa 6. Võrreldes teiste programmidega, kus väljundiks oli enamasti 200-300 oligot rühma kohta, andis Probesel vähemalt $1,5\text{ }^{\circ}\text{C}$ hübridisatsioonienergiate vahega 4-8 oligot

eksemplari kohta.



Joonis 4. Programmi ProbeSel poolt disainitud oligote hübridisatsioonienergiate võrdlus target (Tm-target)- ja mitte-target (Tm-nontarget)-järjestustega

Tulemuste arutelu

Seente määramiseks nii kultuurist kui keskkonnaproovidest (taimejuurtest, mullast) on PCR-RFLP tehnikale lisaks välja pakutud mitmeid erinevaid meetodeid, näiteks hübriidiseerimine SSOP-dega (Sequence-Specific Oligonucleotide Probes) Oh jt. (2003) poolt, kus kasutatakse puud-lagundavate seente identifitseerimiseks nailonmembraanile kinnitatud 15-25 aluspaarilisi lühikesi oligoid, mis hübriidiseeritakse märgitud DNA proovidega (rDNA ITS järjestused). Antud meetod võimaldab analüüsida palju erinevaid proove ühel membraanil. Valinski jt. (2002) on seenekoosluse koostise analüüsimiseks välja pakkunud samuti array-põhise meetodi OFRG (Oligonucleotide Fingerprinting of rRNA Genes), kus rDNA kloonid sorteeritakse hübriidisatsiooniekspereimentide seeriat kasutades taksonoomilistesse klastritesse, iga eksperimendi jaoks kindel oligonukleotiid. Proovide hübriidiseerumine kindlate oligotega ning mitteseondumine teistega annab ülevaate olemasolevatest liikidest. Mõlema meetodi puhul kasutati oligote valmistamiseks ITS järjestusi, mis näitab, et nende kasutamine liikide määramiseks on vähemalt mõnede liikide puhul õigustatud. Oligonukleotiidide valimisel eksperimendi läbiviimiseks kasutati aligneeritud järjestuste visuaalset analüüsi, ega arvestatud seondumise termodünaamikat. Kasutati suhteliselt lühikesi proove (10-25) ning eeldati nende mitteseondumist target-DNA-ga kindlatel hübriidisatsioonitingimustel juba ühe valepaardumise korral. Samas on näidatud, et 20-25 aluspaari pikkuste oligote korral on sarnaste järjestuste/liikide puhul suur tõenäosus rishübriidisatsiooni tekkeks vähem kui 5 varieeruva nukleotiidiga oligote puhul, mistõttu proov võib anda vale signaali.

Töös vaadeldud liikidele oli kõigile võimalik disainida 1, 2, või isegi 3 valepaarduva nukleotiidiga 25-meersed oligod, kuid võimaliku rishübriidisatsiooni ning taustaefektide vältimiseks seati eesmärgiks leida oligod, mis oleksid vähemalt 5 erineva nukleotiidiga kõige sarnasemast mitte-target-järjestusest. 12-st rühmast 9-le (75%) kirjeldatud tingimuste puhul oligod ka leiti. 100%-liselt kõigis liigi eksemplarides sisalduvad oligod oli võimalik disainida liikidele *T. zygodesmoides* ja *Th. americana* (*Th. caryophyllea*), liigi kõigi eksemplaride piires 1 varieeruva nukleotiidiga oligod leiti 4-le liigile: *T. zygodesmoides*, *Th. americana* (*Th. caryophyllea*), *T. bresadoliana* ja *T. lapida*. Kõik nimetatud liigid moodustavad fülogeneesipuul ka ühtsed monofüleetilised klastrid. Samasugused lähestikku asetsevate eksemplaride rühmad on ka liikidel *T. sublilacina* ja *Th. terrestris* + *T. radiosa*, millele ei õnnestunud oligoid leida. Tegemist on sõsarrühmadega, mis ei ole teineteisest ja ka teistest lähedal asetsevatest liikidest veel piisavalt kaugele evolutsioneerunud. Tulevikus tasub kindlasti proovida leida neile ühised oligod, nagu seda tehti liikide *T. echinospora* ja *T. submollise* puhul.

Neid kahte kirjeldavad 25-meersed proovid, mis sisalduvad maksimaalselt 1 'degeneratiivse' nukleotiidiga kõigis eksemplarides, ka leiti. Teiste liikide puhul, nagu ka fülogeneesipuudelt (lisad 2 ja 3) näha, on tegemist liiga suure liigisisese varieeruvusega, et leida neid kõiki ühendavad head oligod. Seetõttu disainiti oligod vaid osale (2-4 järjestust hõlmavad) eksemplaridest.

Töö üheks eesmärgiks oli tutvuda ka olemasolevate oligote disainimise programmidega, mida juba mitu aastat erinevate töögruppide poolt küllaltki palju avaldatud on. Kuna pidevalt täienevad teadmised hübriidiseerumise termodünaamikast, suureneb praktiliste kogemuste pagas ning otsitakse kiiremaid ja täpsemaid algoritme vastavate probleemide lahendamiseks, on oodata uute programmide väljatöötamise suurenemist ka järgnevatel aastatel. Aastani 2004 polnud avaldatud või on olnud kasutajatele kättesaamatu programm, mis leiaks oligod tervele rühmale järjestustele ja arvestaks selle juures seondumise termodünaamikat T_m arvutamisel. Hiljuti tutvustasid Matveeva jt. (2004) HIV-viiruse (Human Immunodeficiency Virus) jaoks loodud programmi, mis just eelpoolmainitud probleemi arvestab. Programm leiab aligioneeritud järjestuste grupi baasil piirkonnad, mis on termodünaamiliselt kõige sobivamad hübriidiseerumiseks, kuna 'mismatch' positsioonid on mõnedes DNA dupleksites destabiliseerivamad kui teistes. Siiski annaks ka antud programmi lähedaste liikide identifitseerimiseks vajaminevate oligote leidmiseks kohandada, kuna ta ei arvesta mitte-target-grupi olemasolu, mis vähendaks sobivate piirkondade arvu tunduvalt.

Programm ProbeSel leidis oligod kõigile 4-le ekseplarile, mis katsetamiseks valiti. Selle eelis kahe teise programmi ees, mis BS-i (seondumistugevus, *Binding Strength*, ingl. k.) ei arvestanud, oli suure hulga mittedobivate oligote väljaviskamine. Kui teised programmid andsid oligod isegi ühealuspaarilise erinevusega teiste mitte-target-grupi järjestustest, siis loomulik on eeldada, et enamikul juhtudel on ühest valepaardumisest tekkinud seondumise erinevus liiga väike, et oligo reaalselt kahte liiki eristada võiks. Probesel programmi puhul oli võimalik välja filtreerida oligod, mis olid liiga väikese hübriidisatsioonienergiatega vahega target- ja mitte-target-grupi järjestuste vahel. Kindlasti tasub oligote disainil arvestada seostumisenergiat, ning kuigi hübriidisatsioonienergia arvutamine proovi ja lahuses olevate järjestuste vahel ei ole hetkel veel päris täpne ega kindlalt defineeritud, ning iga programm võtab arvesse erinevaid mõjusid (sekundaarstruktuuri mõju, järjestuse pikkuse, naabernukleotiidide ja kontsentratsioonide mõju), siis sellel alal tehtav suur töö ning järjest suurenev praktiline kogemus mikroarray' de abil lubab üha paremate ja reaalselt töötavate oligote disaini uute kiiremate programmidega.

Antud töös jäi lahtiseks küsimus, kui palju võib oligo sisaldada varieeruvaid nukleotiide, et oleks detekteeritavad õiged signaalid micro-array' l, et ei toimuks risthübriidisee-

rumist sarnaste mitte-target-grupi järjestustega, kuid et antud oligoga seonduksid ka 1-2 'mismatch' nukleotiidiga samasse liiki kuuluvad ekseplariid. Kuna kirjandusest sellele ühtset vastust ei antud, tuleb järgmises tööetapis leitud oligod laboris testida. Loomulikult oleks ITS-järjestusi kasutades, arvestades nende varieeruvust liigi piires, lihtsam leida oligod 2-3 'mismatch' nukleotiidiga, kui vähemalt 5-ga. Samas ei saa väita, et teisel juhul oleks võimatu kõiki neid seenerühmi katta, kuna töös uurimata jäänud liikide (milles pole sekveneeritud üle 2-3 eksemplari) esin-datus herbaariumites ja andmebaasides lähitulevikus kindlasti tõuseb ning annab võimaluse paremate oligote leidmiseks.

Lühikeste (25 nt) ja pikkade (50 nt) oligote võrdlemisel ilmnes, et ITS-järjestuste puhul on sobivam kasutada lühemaid. Ka siin on põhjuseks liigisisene varieeruvus. Kuna üks osa oligost on perekonna (tihti isegi sugukonna) sees küllaltki konserveerunud, peab teine osa olema piisavalt erinev, samas liigisiselt ühine. Sellised piirkonnad on aga tihti väga lühikesed. Grönberg jt. (2003) on ektomükoriisat moodustavate liikide määramiseks hübriidisatsioonil kasutanud nii lühemaid (20-25 nt) kui ka pikemaid (124-151 nt) oligoid, ning tõdenud, et tüvede ja liikide eristamiseks sobivad just lühemad oligod, kusjuures 124-151 aluspaarised oligod hübriidiseerusid küll erinevate tüvede PCR-iga ülesamplifitseeritud ITS-järjestustega, kuid signaalid kogu genoomse DNA vastu olid mitte-spetsiifilised.

Seega tuleks edasises töös kindlasti proovida lühemate oligote disaini ka teistele perekondade *Thelephora*, *Tomentella* ja *Tomentellopsis* liikidele, millele peaks eelnema juba antud töös tutvustatud oligote testimine. Ka tasub pidevalt kursis olla uute programmidega oligote disainiks ning teiste tööühmade tulemustega selles vallas. Kuigi pole kindel, kas liikide määramise microarray resupinaatsete lehternahkiste jaoks ka realselt teoks saab, siis ITS järjestuste sobivuse hindamisel liikide määramiseks, mida hetkel ka aktiivselt kasutatakse, võib liigispetsiifiliste oligote leidmise õnnestumine oma mõju avaldada.

Kokkuvõte

Käesolev töö kirjeldab liigispetsiifiliste oligonukleotiidide disaini resupinaatsete lehternahkiselistes ITS (Internal Transcribed Spacer) järjestuste põhjal, mida võib tulevikus kasutada mikrokiibi valmistamiseks liikide määramiseks. Eesmärgiks seati välja selgitada, kas liigispetsiifilised oligod nimetatud järjestuste põhjal on sellesse seenerühma kuuluvate liikide puhul võimalikud, ning leida antud ülesande lahendamiseks sobivaim programm.

Vaadeldud on 3 perekonda (*Thelephora*, *Tomentella*, *Tomentellopsis*) kuuluvat 12 seente rühma, mis moodutavad fülogeneesipuul ühtsed klastrid. Oligote disainiks on kasutatud kolme erineva lähenemisviisiga programmi: PRIMROSE lubab oligotes 'degeneratiivsete' nukleotiidide olemasolu võimaldades spetsiaalselt fülogeneetiliste oligote disaini, ProbeSel arvutab seondumise termodünaamikat kasutades kõige madalama hübridisatsioonienergiaga oligod, PROBEmer leiab oligod target- ja mitte-target-järjestuste baasil.

Nimetatud programmidega õnnestus disainida 25-meersed oligod 75%-le vaadeldud rühmadest ning 50-meersed kahele rühmale. Oligote leidmisel seati neile küllaltki ranged kriteeriumid, et vältida hilisemaid probleeme, mis nende kasutamisel sarnaste liikide eristamisel ilmnedavad võivad. Töö käigus leiti, et antud liikide puhul on sobivamad lühikesed (25-nukleotiidised) oligod, mis on enamasti kindla struktuuriga (konserveerunud ja variaabel ala kõrvuti) ning langevad oma start-positiooniga target-järjestustel kindlatesse rühmadesse. Konserveerunud järjestuste puhul leiti kõiki eksemplare kirjeldavad oligod ilma ühegi valepaardumiseta, suurema liigisisese varieerumise puhul oli vajalik mitme oligo disain, mis kirjeldaksid erinevaid liigi eksemplare. Mõnedele sõsarliikidele on võimalik disainida oligod, mis sisalduvad mõlema liigi eksemplaride ITS järjestustes.

Erinevate programmide katsetamisel ilmnas, et tõhusam on hübridisatsioonienergiaid arvestava programmi kasutamine, mida senini veel spetsiaalselt antud probleemi lahendamiseks avaldatud pole. Teiste programmidega langesid disainitud oligod suures osas ühte, unikaalsete oligote rohkus oli PRIMROSE' i puhul sõltuvuses järjestuste sarnasusest.

Mõlemad töö eesmärgid saavutati, mis annab eelduse oligote disainimiseks ka teistele uuritud seenerühma liikidele, ning lubab tulevikus nende kasutamist liikide eristamiseks mikrokiibil.

Summary

This work describes designing species-specific oligonucleotides for detection and species identification on microarray, based upon ITS (Internal Transcribed Spacer) sequences from resupinate theleporoid fungi. The aim of this work was to find out if these sequences from given species are suitable for designing species-specific oligos, and to find appropriate program for the given task.

12 studied groups belonging to 3 families form unified clusters on phylogenetic tree. There were three programs, each with different approach, used to design oligonucleotides. PRIMROSE, a program for designing phylogenetic oligos, allows 'degenerative' nucleotides to be used within the probe, Probesel calculates hybridization free energies for each target-probe duplex and recommends oligos with the highest free energy difference between probe-target and probe-non-target duplexes. PROBEmer finds oligos for given target-sequences taking into account non-target-sequences.

Above-named programs were able to give group-specific 25-mer oligos for 75% of studied groups and 50-mer oligos for two groups. There were quite strict criteria set for oligos to avoid problems that might occur later, trying to distinguish between similar species. It was shown that for these species short oligos (25 nucleotides) are more suitable than the long ones (50 nucleotides). They mostly have specific structure (conserved area alongside with more variable area), and their start positions on target-sequences fall into three defined groups. It was possible to find oligos for conserved sequences without any mismatch nucleotides, which describe all specimens inside one group. For groups with higher intra-specific variability, more than one oligo had to be designed. For some of the closely related species it is possible to find oligos that will describe most of the specimens under the given two species.

Testing different designing programs showed that it is more effective to use program which considers hybridization free energies. The other programs gave many identical oligos, but also unique ones. Unique oligos from PRIMROSE depended on the similarity of aligned sequences.

Both objectives that were set, were also attained, which gives a good prospect to find oligos to other species from the same group of studied fungi, and makes it possible to use them on microarray to determine fungal community composition.

Käesoleva töö raames avaldatakse tänud töö juhendajatele prof. Urmas Kõljalale ja prof. Mairo Remmile, ning Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituudi Bioinformaatika õppetooli inimestele, kes on töö valmimisele kaasa aidanud.

Kasutatud kirjandus

1. Alvarez, I., J.F. Wendel. (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phyl. Evol.*, 29, 417-434.
2. Ashelford, K. E., A.J. Weightman, J.C. Fry. (2002) PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acid Res.*, 30, 3481-3489.
3. Bruns, T.D., T. M. Szaro, M. Gardes, K. W. Cullings, J. J. Pan, D. L. Taylor, T. R. Horton, A. Kretzer, M. Garbelotto and Y. Li. (1998) A sequence database for the identification of ectomycorrhizal basidiomycetes by phylogenetic analysis. *Mol. Ecol.*, 7, 257-272.
4. Cheung, V.G., M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati, G. Childs. (1999) Making and reading microarrays. *Nature Genet. Suppl.*, 21, 15-19.
5. Dahllöf, I. (2002) Molecular community analysis of microbial diversity. *Curr. Opin. Biotech.*, 13, 213-217.
6. Emrich, S.J., M. Lowe, A.L. Delcher. (2003) PROBEMER: a web-based software tool for selecting optimal DNA oligos. *Nucleic Acid Res.*, 31, 3746-3750.
7. Fukushima, M., K. Kakinuma, H. Hayashi, H. Nagai, K. Ito and R. Kawaguchi. (2003) Detection and Identification of *Mycobacterium* Species Isolates by DNA Microarray. *J. Clin. Microbiol.*, 41, 2605-2615.
8. Greuter, W., J. McNeill, F. R. Barrie, H.-M. Burdet, V. Demoulin, T. S. Filgueiras, D. H. Nicolson, P. C. Silva, J. E. Skog, P. Trehane, N. J. Turland, D. L. Hawksworth. (2000) International Code of Botanical Nomenclature (St Louis Code). *Regnum Vegetabile* 138. Koeltz Scientific Books, Königstein. (elektrooniline versioon <http://www.bgbm.fu-berlin.de/iapt/nomenclature/code/SaintLouis/0000St.Luistitle.htm>)
9. Grönberg, H., L. Paulin and R. Sen. (2003) ITS probe development for specific detection of *Rhizoctonia* spp. and *Suillus bovinus* based on southern blot and liquid hybridization-fragment length polymorfism. *Mycol. Res.*, 107, 428-438.
10. Haas, S.A., M. Hild, A.P.H. Wright, T. Hain, D. Talibi and M. Vingron. (2003) Genome-scale design of PCR primers and long oligomers for DNA microarrays. *Nucleic Acid Res.*, 31, 5576-5581.
11. Horton, T.R. (2002) Molecular approaches to ectomycorrhizal diversity studies: variation in ITS at a local scale. *Plant and Soil*, 244, 29-39.

12. Kaderali, L. & Schliep, A. (2002) Selecting signature oligonucleotides to identify organisms using DNA arrays. *Bioinformatics*, 18, 1340-1349.
13. Kane, M.D., T.A. Jatkoe, C.R. Stumpf, J. Lu, J.D. Thomas and S.J. Madore. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acid Res.*, 28, 4552-4557.
14. Katoh, K., K. Mizawa, K. Kuma and T. Miyata. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acid Res.*, 30, 3059-3066.
15. Kõljalg, U., A. Dahlberg, A. F. S. Taylor, E. Larsson, N. Hallenberg, J. Stenlid, K.-H. Larsson, P. M. Fransson, O. Karen and L. Jonsson. (2000) Diversity and Abundance of resupinate theleporoid fungi as ectomycorrhizal symbionts in Swedish boreal forests. *Mol. Ecol.*, 9, 1985-1996.
16. Kõljalg, U. & Larsson, E. (1998). *Pseudotomentella ochracea* sp. nov., based on morphological and molecular data. *Folia Cryptog. Estonica*, 33, 53-56.
17. Kõljalg, U., H. Tammi, S. Timmonen, R. Agerer, R. Sen. (2002) ITS rDNA sequence-based phylogenetic analysis of *Tomentellopsis* species from boreal and temperate forests and the identification of pink-type ectomycorrhizas. *Mycol. Progr.*, 1, 81-92.
18. Li, F., G.D. Stormo. (2001) Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics*, 17, 1067-1076.
19. Lipshutz, R.J., S.P.A. Fodor, T.R. Gingeras, D.J. Lockhart. (1999) High density synthetic oligonucleotide arrays. *Nature Genet. Suppl.*, 21, 20-24.
20. Matveeva, O.V., B.T. Foley, V.A. Nemtsov, R.F. Gesteland, S. Matsufuji, J.F. Atkins, A.Y. Ogurtsov, S.A. Shabalina. (2004) Identification of regions in multiple sequence alignments thermodynamically suitable for targeting by oligonucleotides: application to HIV genome. *BMC Bioinformatics*, 5:44.
21. Nilsson, R.H., K.-H. Larsson, B.M. Ursing. (2004) Galaxie-CGI scripts for sequence identification through automated phylogenetic analysis. *Bioinformatics*, 20, 1-6.
22. Oh, S.K., D.P. Kamdem, D.E. Keathley and K.-H. Han. (2003) Detection and Species Identification of Wood-Decaying Fungi by Hybridization of Immobilized Sequence-Specific Oligonucleotide Probes with PCR-Amplified Fungal Ribosomal DNA Internal Transcribed Spacers. *Holzforschung*, 57, 346-352.

23. Pfunder, M., O. Holzgang and E. Frey. (2004) Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I vs. cytochrome b as genetic markers. *Mol. Ecol.*, 13, 1277-1286.
24. Rouillard*, J.-M., M. Zuker and E. Gulari. (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acid Res.*, 31, 3057-3062.
25. Stears, R.L., T. Martinsky and M. Schena. (2003) Trends in microarray analysis. *Nat. Med. Review*, 9.
26. Sylvia, D.M. Overview of Mycorrhizal Symbioses (<http://dmsylvia.ifas.ufl.edu/mycorrhiza.htm>).
27. Taylor, J.W., D.J. Jacobson, S. Kroken, T. Kasuga, D.M. Geiser, D.S. Hibbett and M.C. Fisher. (2000) Phylogenetic Species Recognition and Species Concepts in Fungi. *Fungal Gen. and Biol.*, 31, 21-32.
28. Troesch, A., H. Nguyen, C.G. Miyada, S. Desvarenne, T.R. Gingeras, P.M. Kaplan, P. Cros and C. Mabilat. (1999) *Mycobacterium* Species Identification and Rifampin Resistance Testing with High-Density DNA Probe arrays. *J. Clin. Microbiol.*, 37, 49-55.
29. Valinski, L., G.D. Vedova, T. Jiang and J. Borneman. (2002) Oligonucleotide Fingerprinting of rRNA Genes for Analysis of Fungal Community Composition. *Appl. Environ. Microbiol.*, 68, 5999-6004.
30. Van der Heijden, M.G.A., A. Wiemken and I.R. Sanders. (2003) Different arbuscular mycorrhizal fungi alter coexistence and resource distribution between co-occurring plant. *New Phyt.*, 157, 569-578.
31. Volkohov, D., A. Rasooly, K. Chumakov and V. Chizhikov. (2002) Identification of *Listeria* Species by Microarray-Based Assay. *J. Clin. Microbiol.*, 40, 4720-4728.
32. Zhou, J., D.K. Thompson. (2002) Challenges in applying microarrays to environmental studies. *Curr. Opin. Biotech.* 13, 204-207.
33. Zhou, J. (2003) Microarrays for bacterial detection and microbial community analysis. *Curr. Opin. Biotech.*, 6, 288-294.

Töös esinevad lühendid

PCR – polümeraasi ahelreaktsioon, *polymerase chain reaction*, ingl. k.

TGGE – temperatuuri gradiendi geel-elektroforees, *temperature gradient gel electrophoresis*, inkl. k.

RFLP - restriktiooni fragmendi pikkuse polümorfism, *restriction fragment length polymorphism*, ingl. k.

T-RFLP – terminaalne restriktiooni fragmendi pikkuse polümorfism, *terminal restriction fragment length polymorphism*, ingl. k.

ITS – sisemine transkribeeritud vahemik, *internal transcribed spacer*, ingl. k.

PHP – PHP: Hypertext Preprocessor.

Perl – Practical Extraction and Report Language.

CGA – koosluse genoomi kiip, *community genome array*, ingl. k.

FGA – funktsionaalse geeni kiip, *functional gene array*, ingl. k.

POA – fülogeneetiline oligonukleotiidide kiip, *phylogenetic oligonucleotide array*, ingl. k.

SSOP – järjestuse-spetsiifilised oligonukleotiidid, *sequence-specific oligonucleotide probes*, ingl. k.

OFRG – rRNA geenide tüpeerimine oligonukleotiididega, *oligonucleotide fingerprinting of rRNA genes*, ingl. k.

BS - seondumistugevus, *binding strength*, ingl. k.