

BIOLOOGIA-GEOGRAAFIA TEADUSKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
BIOINFORMAATIKA ÕPPETOOL

Meelis Olev

Erinevate PCR praimerite grupeerimise efektiivsus

Bakalaureusetöö

Juhendajad: Mairo Remm, Prof., PhD,
Lauris Kaplinski, B. Sc.

Tartu, 2005

Sisukord

Sisukord	2
Lühendid ja mõisted	4
Sissejuhatus	5
I Teoreetiline osa	6
1.1 PCR praimerite disain.....	7
1.2 PCR praimerite disainiks kasutatavad programmid.....	8
2.1 Praimerite sulamistemperatuur.....	9
2.2 Kordusjärjestused.....	9
2.3 Polü-N traktid.....	9
2.4 Praimerite vahelised interaktsioonid.....	10
2.5 Produkti pikkus.....	10
2.6 Praimerite pikkus.....	10
2.7 Praimerite GC nukleotiidide sisaldus.....	11
2.8 Muud parameetrid.....	11
3.1 Indeliteta joondus.....	11
3.2 Smith-Waterman.....	12
3.3 BLAST.....	13
3.4 Suffiksipuud.....	13
3.5 Räsialgoritmid.....	14
3.6 Kui pikka ala praimerist arvestada.....	14
4.1 Olulised parameetrid grupeerimisel.....	15
4.2 Grupeerimise algoritm.....	18
5.1 PCR supressioonil baseeruv amplifitseerimine (Broude et al., 2001).....	18
5.2 DOP-PCR (Jordan et. al, 2001).....	19
II Praktiline osa	20
Eesmärk	20
Töövahendid	20
Grupeerimiseks vajaliku korduste arvu leidmine	20
Katse skeem.....	20
Hüpotees.....	21
Tulemused.....	21
Praimeri GC-sisalduse mõju grupeerimise efektiivsusele	21
Esimese katse skeem.....	21
Hüpotees.....	22
Tulemused.....	22
Teise katse skeem.....	22
Hüpotees.....	22
Tulemused.....	22
Praimeri ja produkti pikkuse mõju grupeerimise efektiivsusele	22
Hüpotees.....	23
Tulemused.....	23
Praimeri 3'-otsa mõju grupeerimise efektiivsusele	23
Hüpotees.....	23
Tulemused.....	23
Arutelu	24
Kokkuvõte	26

Grouping efficiency of different PCR primers	27
Tänu	28
Viited	29
Lisad	32

Lühendid ja mõisted

indel	joonduses mittekattuvate nukleotiidide piirkond (<i>gap</i>), n.ö. auk
matriits	ahel, mille pealt DNA polümeraas uue, komplementaarse ahela sünteesib (<i>template</i>)
PCR	polümeraasahelreaktsioon (<i>polymerase chain reaction</i>)
piirav väärtus	väärtus, millega leitud skoori võrreldakse; jah/ei otsus langetatakse selle järgi, kas väärtus on suurem või väiksem piiravast väärtusest (<i>cutoff</i>)
skoor	väärtus, sobivuse hinnang (<i>score</i>)
SNP	ühenukleotiidne polümorfism (<i>single nucleotide polymorphism</i>)

Sissejuhatus

Nii nagu tundub tänapäeval kummaline kohata bioloogi, kes ei tea midagi geenidest, ei saa ilmselt tõsiselt võtta ka biotehnoloogiaga tegelevat inimest, kellele midagi ei ütle täheks kombinatsioon PCR, eestikeelse vastega polümeraasahelreaktsioon. PCR abil on võimalik paljundada valitud lõike genoomist (eeldusel, et sellised lõigud seal eksisteerivad), et näiteks hiljem teha need nähtavaks elektrofooresil. Tüüpilises PCR reaktsioonis osalevad termopolümeraas (kuumaveebakterist saadud DNA-d paljundav valk), lähtematerjal (näiteks genoomne DNA), kaks praimerit (oligonukleotiidid, mis seonduvad paljundatava järjestuse otstele) ja PCR puhver. Inimese geenide analüüs vereproovist on tüüpiline näide, kus soovatakse genoomse materjali pealt paljundada e. amplifitseerida erinevaid lõike. See nõuab kas suurema hulga identsete praimerite reaktsioonisegusse lisamist või genoomse materjali piisava hulga korral iga lõigu jaoks eraldi reaktsiooni läbi viimist. Teine võimalus on kasutada n.ö. multipleks-PCR-i – mitu erinevat lõiku amplifitseerida ühes reaktsioonis. Selline reaktsioon seab aga omad piirangud amplifitseeritavatele lõikudele, praimerid ei tohi seonduda teiste praimeritega, samuti produktid produktidega ning praimerid produktidega. Antud piiranguid on võimalik arvesse võtta ning sobivad koos amplifitseeritavad lõigud leida arvutitarkvara abil. Samas võib tekkida küsimus, kas ei anna praimerite võimalikult efektiivselt gruppidesse jaotamist hõlbustada, valides juba praimeridisaini faasis välja multipleksimiseks sobivate parameetritega praimerid. Antud töös uuritaksegi praimerite grupeerimise efektiivsuse sõltuvust mitmesugustest praimerite parameetritest.

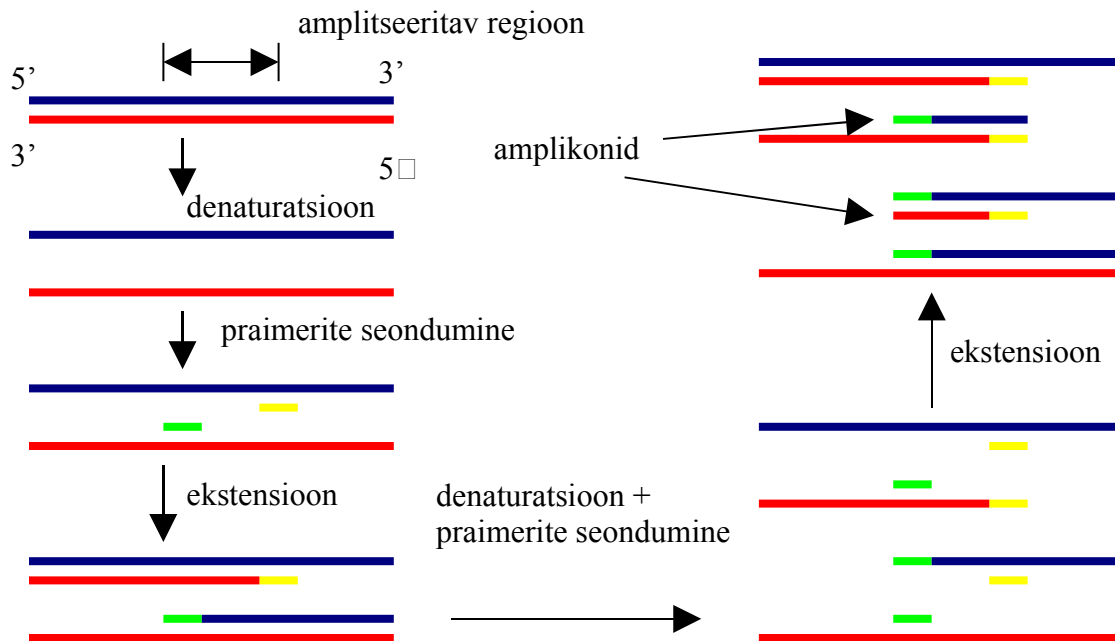
I Teoreetiline osa

1 PCR

PCR (Polymerase Chain Reaction e. polümeraasi ahelreaktsioon) on molekulaarbioloogias laialt rakendatav meetod, mida kasutatakse genotüüpiseerimisel, punktmutatsioonide uurimisel, genoomse DNA kloonimisel jm. PCR võimaldab saada ühest geenilõigust palju koopiaid, mille olemasolu tänu suurele hulgale on hiljem lihtsam detekteerida. PCR koosneb kolmest etapist, mida korratakse tsükliliselt:

- denaturatsioon – genoomse DNA puhul temperatuuril 93-95 °C. DNA kaksikahel “sulatakse lahti” kaheks üksikahelaks. Sellisel temperatuuril lõpevad ära ensümaatilised reaktsioonid, näiteks DNA süntees eelmisest tsüklist.
- annealing e. praimerite seostumine DNA-le (~50-70 °C). Üheahelalise praimeri ja üheahelalise matriitsi (*template*) vahel moodustuvad vesiniksidemed, komplementaarses piirkonnas moodustub kaksikahel. Stabiilsemale kaksikahelale saab seostuda polümeraas (tüüpiliselt kuumaveebakterist eraldatud nn. Taq-polümeraas) ning algab matriitsi duplitseerimine.
- ekstensioon e. DNA süntees (70-75 °C). Polümeraas lisab nukleotiidtrifosfaate praimeri 3' otsa. Praimerid, mida polümeraas on juba pikendanud, on stabiilsemad tänu tekkinud uutele vesiniksidemetele, ülejäänud praimerid matriitsi küljes ei püsi ning võivad seonduda teistele positsioonidele järgmises tsükli.

Tsükli korratakse 25-45 korda (Glick, Pasternack, 1998).



Joonis 1. Polümeraasahelreaktsiooni skeem

1.1 PCR praimerite disain

PCR teostamiseks on vaja teada amplifitseeritava DNA järjestust, mille alusel leida e. disainida praimerid – 15-30-meersed sünteetilised oligonukleotiidid. Praimerite disainil tuleb arvestada järgnevaga:

- praimerite sulamistemperatuur (T_m) peaks olema 55-80 °C (Innis, Gelfand, 1990)
- praimerite GC sisaldus peaks olema 50-60% (Innis, Gelfand, 1990), väiksema GC sisalduse puhul tuleb sulamistemperatuuri tõstmiseks disainida pikemad praimerid
- vältida tuleks ühe või kahe nukleotiidi kordusi, eriti 3' otsas, kuna see võib põhjustada sekundaarseid hübridisatsioone (seondumisi iseenda või mõne teise reaktsioonis osaleva praimeriga)
- praimer 3' ots on stabiilsem (seondub tugevamini matriitsiga ja annab parema PCR kvaliteedi), kui seal asub G või C nukleotiid (Innis, Gelfand, 1990)

- praimerid ei tohiks moodustada sekundaarstruktuure (nt. juuksenõelastruktuur) ning di- ja oligomeere (Rachlin et. al, 2005)

1.2 PCR praimerite disainiks kasutatavad programmid

Suured molekulaarbioloogias kasutatavad programmide paketid nagu EMBOSS (<http://emboss.sourceforge.net/>, Windows, Unix, Mac, vabavara) ja GCG (<http://www.accelrys.com/about/gcg.html>, Unix, kommertstarkvara) sisaldavad programme praimerite disainimiseks, vastavalt Prima ja Prime. Laialt tuntud ning ka veebibrauseri kaudu kasutatav on Primer3 (Rozen, Skaletsky, http://frodo.wi.mit.edu/primer3/primer3_code.html, Unix, Windows, vabavara) (veebi kaudu kasutatav variant http://fokker.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), varasemates versioonis nimega Primer. Erinevate tööpõhimõtete ja sisendandmetega on ka näiteks programmid PC-Rare (Griffais, <http://bioinformatics.weizmann.ac.il/software/PC-Rare/>, Mac, Windows, vabavara), PRIDE (<http://www.uni-duesseldorf.de/urz/hardware/server/pride/PRIDE.html>, Unix, kommertstarkvara), PerlPrimer (<http://perlprimer.sourceforge.net/>, Windows, Unix, Mac, vabavara). Veebibrauseri kaudu kasutatavad on veel NetPrimer (<http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html>), GeneFisher (<http://www.genefisher.de/>), Web Primer (<http://seq.yeastgenome.org/cgi-bin/web-primer>). Toodud nimekirja võiks pikalt jätkata.

Optimaalse praimeripaari leidmine nõuab mitmete parameetrite optimeerimist. Viimase kümne aasta jooksul on kujunenud välja kindel hulk põhiparameetreid, mida kasutavad paljud programmid (Burpo, 2001). Neist on täpsemalt juttu allpool. Enamik programme leiavad praimerite parameetrite tegelikud väärtused ning suunavad neid, et leida optimaalne praimer kindlaks kasutusjuhtumiks (Kampke, et al, 2001). Kuid näiteks PRIDE kasutab kvalitatiivset kasutaja sisendit (näiteks sulamistemperatuuri ette andmine omadussõnana “madal”, “optimaalne”, “väga kõrge”) ning arvutustes hägusat loogikat (Haas, et al, 1998).

2 PCR kvaliteeti mõjutavad parameetrid

PCR kvaliteet on määratud õige produkti hulga ja valede produktide puudumisega. Kvaliteeti mõjutavad põhiliselt järgnevalt toodud parameetrid.

2.1 Praimerite sulamistemperatuur

Lihtsustatud lähenemine praimerite sulamistemperatuuri arvutamisele on lisada 2 °C iga A-T paari ning 4 °C iga G-C paari kohta (Suggs et al, 1981). Seda nimetatakse ka Wallace'i reegliks.

Enamasti kasutatakse täpsemat valemit

$$T_m = \frac{H}{S + R \ln(c/4)} - 273.15^\circ C + 16.6 \log_{10} [K^+]$$

kus H on nukleotiidide hübridiseerumisest tekkinud entalpia muutude summa, S analoogiliselt entroopia muutude summa, R gaasikonstant (8,31 J/mol), c ahelate keskmine molaarne kontsentratsioon. Sellist valemit kasutab oma töös näiteks programm Primer3.

Täpsema sulamistemperatuuri arvutamiseks vajalike termodünaamiliste parameetrite leidmiseks kasutavad paljud programmid nn. lähimate naabrite meetodit (Breslauer, et al, 1986). See tähendab, et kogu ahela H ja S saadakse liites kokku kõikide naabernukleotiidide paaride jaoks määratud ΔH ja ΔS .

Praimeripaari korral peab praimerite sulamistemperatuur erinema vähesel määral, et tagada võimalikult ühtlane erisuunaliste praimerite seondumine matriitsiga, samuti dissotsieerumine matriitsist.

2.2 Kordusjärjestused

Kordusjärjestused (nt. Alu), moodustavad inimese genoomist 50% või üle selle. Et vältida mittespetsiifilist seondumist genoomis korduvatele lõikudele, testivad mitmed programmid (nt. PRIMER) praimerite komplementaarsust genoomsete kordusjärjestustega või välditakse kordusi hekso/oktameeride sagedustabeli kaudu.

2.3 Polü-N traktid

Sarnaselt kordusjärjestustele põhjustab ka ühe nukleotiidi paljukordne järjest esinemine seostumise mittespetsiifilisust. Samuti korduvad di ja trinukleotiidid. Tüüpiliselt esineb matriitsis pikki komplementaarseid kordusi, pikemaid kui praimeris ning mitmes kohas, mis võimaldab praimeril seonduda paljudesse erinevatesse kohtadesse.

2.4 Praimerite vahelised interaktsioonid

Ühe praimeripaariga (*single-plex*) reaktsioonis tuleb kontrollida praimeri seondumist iseenda ja vastaspraimeriga ning 3' otsa seostumist enda ja vastaspraimeri 3' poole järjestusega, multipleks-reaktsioonis kõiki primereid omavahel. Iseenda või mõne teise reaktsioonis osaleva praimeriga hübridiseerudes väheneb *template* DNA'ga seostuvate praimerite kontsentratsioon ning võivad tekkida praimerite dimeerid (Kämpke et al, 2000) või sekundaarstruktuurid nt. juuksenõelastruktuur (*hairpin*).

2.5 Produkti pikkus

Mida lühem on PCR produkt, seda väiksem on tõenäosus, et tekib soovimatu seondumine teiste produktide või praimeritega. Sobiv PCR produkti pikkus oleneb genoomsete järjestuste puhul asukohast, näiteks mikrosatelliitide puhul on eelistatud kuni 150 aluspaari, heterodupleksi analüüside korral 250-500 aluspaari pikkune produkt (Wallace, unpublished). PCR reaktsioonis on soovitav omada sarnase pikkusega produkte võrdsete koguste tagamiseks ja masina tsükli aegade ühtlustamiseks, multipleks-PCR puhul võimaldab aga erinev produktide pikkus neid kergemini eristada. Üldjuhul jääb sobiv produkti pikkuse vahemik 100 ja 1000 nukleotiidi vahele.

2.6 Praimerite pikkus

Tavaliselt loetakse küllaldaseks praimeri pikkusi alates 18 nukleotiidist. Kindel 18 aluse pikkune oligonukleotiid esineb keskmiselt korra $4^{18} = \sim 7 \times 10^{10}$ (70 Gb¹) pikkuses järjestuses, ta ei tohiks seostuda rohkem kui ühte positsiooni enamikes eukarüootsetes genoomides. Samas ei ole genoomi järjestus juhuslik, mis võib nõuda pikemaid primereid, et vältida mittespetsiifilist seondumist. Lühemad praimerid seostuvad sellise suurusega genoomides suure tõenäosusega mitmesse kohta, kuid neid võib kasutada näiteks juhul, kui matriitsiks (*template*) on plasmiidne DNA.

Praimeri pikkus mõjutab ka sulamistemperatuuri, iga lisatud nukleotiid tõstab seda. Pikemad praimerid nõuavad suuremat AT-sisaldust, et sulamistemperatuur jääks mõistlikku vahemikku.

¹ 1 Gb (*gigabase*) on miljard nukleotiidi

2.7 Praimerite GC nukleotiidide sisaldus

Praimeri nukleotiidide sisaldus praimeri 3' otsas määrab seostumise spetsiifilisuse DNA järjestusega. On soovitatud kasutada kõrge GC-sisaldusega 3' otsa (enamasti 50-55%), kuid on näidatud, et kõrge GC-sisaldusega 3'ots võib põhjustada valseostumisi ja halvendada PCR kvaliteeti (Li et al, 1997). GC sisaldus mõjutab samuti sulamistemperatuuri.

2.8 Muud parameetrid

Amplifikatsiooni spetsiifilisus multipleks-PCR reaktsioonis oleneb ka PCR puhvri ja praimerite kontsentratsioonist ning kasutatavatest temperatuuridest. Puhvrina kasutatakse näiteks KCl-sisaldavat PCR puhvrit (Henegariu, et al, 1997).

3 Seostumise arvutamise meetodid

PCR modelleerimisel tuleb arvutada praimerite seostumiste arvu ja seostumistugevust teiste praimerite, produktide või genoomiga. Olulisimad on praimer-praimer interaktsioonid. Vahel arvutatakse välja seostumise tugevus absoluutskaalal (sideme hüdrolüüsi energia e , termodünaamiline vaba energia muut ΔG), mõnikord aga lubatakse lihtsustusi ja loetakse võrdseks A-T ja G-C seondumise tugevus ning tihti arvutatakse seostumise tugevus suhtelisel skaalal, n.ö. skoorina.

Seostumiste leidmisel kasutatakse järgnevaid meetodeid:

3.1 Indeliteta joondus

Indeliteta joonduse (*gapless alignment*) leidmine on kõige lihtsam seostumise arvutamise meetod. Esimese järjestuse positsiooni teise suhtes varieeritakse ning iga positsiooni jaoks loetakse komplementaarsed ja mittekomplementaarsed nukleotiidid, vastavalt millele suurendatakse või vähendatakse skoori. Eraldi väärtused võivad olla teadmata nukleotiidide jaoks. Eeldusel, et tugevam seostumine annab suurema skoori, loetakse lõplikuks seostumise tugevuseks maksimaalne leitud väärtus. Mõnel juhul võib tegeliku seondumise käigus osa nukleotiide paardumata jääda, mida antud algoritm ei arvesta.

A							
T							
C							
G							
A							
A							
T							
	A	C	A	G	T	T	A

Tabel 1. Indeliteta joondus. Kahe järjestuse (5'-ATCGAAT-3', 3'-ACAGTTA-5') seonduvad positsioonid on esitatud sirglõigu ristumisena.

3.2 Smith-Waterman

Smith-Watermani algoritm (Smith, Waterman, 1981) on nn. dünaamilise programmeerimise (*dynamic programming*) algoritm. Optimaalseima joonduse leidmiseks arvutatakse „headust“ näitav punktisumma (*score*) kahe järjestuse iga võimaliku positsiooni kombinatsiooni kohta ning lõpuks leitakse koostatud tabeli väärtustest maksimum. Maksimumi leidumise kohalt tagasi suundudes (*backtracking*) on võimalik leida optimaalne joondus.

A							
T							
C							
G							
A							
A							
T							
	A	C	A	G	T	T	A

Tabel 2. Smith-Watermani joondus. Kahe järjestuse (5'-ATCGAAT-3', 3'-ACAGTTA-5') seonduvad positsioonid on esitatud sirglõigu ristumisena.

Punktisumma ei ole üheselt määratud, vaid sõltub algoritmile antavatest parameetritest – tüüpiliselt positiivne „preemia“ seondumise (*match*) jaoks ning negatiivne „karistus“ mitteseondumise (*mismatch*) ja augu (*gap*) eest.

Algoritmi töökiirust on võimalik tõsta, kui piirata võimalikku augu pikkust.

3.3 BLAST

BLASTi (*Basic Local Alignment Search Tool*) algoritmi võrreldes Smith-Watermani algoritmiga iseloomustab väiksem tundlikkus (*sensitivity*) ja suurem töökiirus. Kasutades lühikestel sõnadel baseeruvat heuristikat, püüab BLAST ligikaudu realiseerida nn. maksimaalsete segmendipaaride (*maximal segment pairs*) algoritmi. Maksimaalsete segmendipaaride joonduste statistikud on arusaadavad ja seega on võimalik arvutada olulisustõenäosus (Altschul et al, 1997). Ent selline joondus ei saa sisaldada indeleid, mistõttu ka BLAST eeldab, et indelitate arvestamata jätmine oluliselt joondust ei mõjuta.

Tüüpiline BLASTi kasutus on antud etalonjärjestusega sarnaste järjestuse otsimine kandidaatjärjestuste seast, näiteks genoomi andmebaasist. BLASTi annab kasutada praimerite ja produktide vahelise seondumistugevuse arvutamiseks, kui suur täpsus ei ole nõutav.

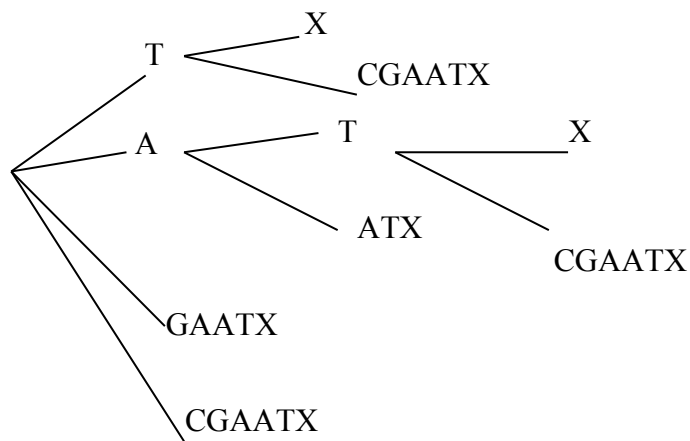
Praeguseks on arendatud ka täiustatud versioone BLASTist, mis on parema tundlikkuse või kiirusega ning võivad lubada ka indeleid joonduses. Nimetada võiks näiteks MEGABLASTi.

3.4 Suffiksipuud

Suffiksipuud (*suffix tree*) on üks võimalik efektiivne järjestuste esitamise viis puukujulise graafina. Nagu nimigi ütleb, sisaldab suffiksipuu sinna lisatava stringi kõiki suffikseid ehk alamstringe, mis algavad suvalisest kohast, ent lõpevad stringi lõpus. Lihtsamal juhul väljuvad igast tipust kaared tema alamtippudesse vastavalt sellele, milline on lisatud stringi või tema suffiksi järgmine täht. Keerulisemal (ja efektiivsemal) juhul hoitakse kaartel viiteid lisatud stringi alamstringi alguspositsioonile ja pikkusele. Samuti võivad tipud sisaldada viiteid teistele puus esinevatele tippudele.

Suffiksipuu rakendusi informaatikas ja bioinformaatikas on mitmeid. Üldistatud suffiksipuu on võimalik otsida näiteks pikim ühine alamjada ajahinnanguga $O(m)$, kus m on üldistatud suffiksipuu suurus. Samuti on üldistatud suffiksipuust võimalik otsida

kõikidest sinna lisatud stringidest üheaegselt keerukusega $O(n + k)$, kus n on otsitava sõne pikkus ning k selle sõne esinemiste arv. Suffiksipuudel baseeruvalt on võimalik koostada kogu genoomi joondus, nagu seda teevad programmid MUMmer (Delcher, et al., 2002) ja Multiple Genome Aligner (Höhl, et al., 2002). Samuti on suffiksipuude kasutamine efektiivne mikrokiibile seotud oligonukleotiidsete signatuuride leidmisel geeniekspressiooni uurimisel (Kaderali, Schliep, 2002).



Joonis 2. Näide suffiksipuust ATCGAAT

3.5 Räsialgoritmid

Räsialgoritmid (*hashing algorithms*) võimaldavad kiirendada otsinguid andmebaasidest. Räsimine (*hashing*) jaotab järjestuse suhteliselt lühikesteks sõnadeks ning nende põhjal on võimalik koostada otsingut hõlbustav „indeks“. Alamjärjestuste asemel võivad indeksis olla ka nn. sõrmejäljed (*fingerprint, message digest*), mis saadakse spetsiaalset räsifunktsiooni kasutades. Räsimise kombineerimisel teiste meetoditega on võimalik sarnaste järjestuste otsingu keerukus viia lineaarseks.

3.6 Kui pikka ala praimerist arvestada

Püüdes arvutuslikult ennustada praimerite seondumist genoomi, võib piisava täpsuse anda juba praimeris osalise järjestuse kasutamine. On näidatud, et vähemalt 13 nukleotiidi arvestamine praimeris 3' lõpust tagab arvestatava kvaliteedi (Andreson et al.,

unpublished). Praimeri osalise ja seega lühema järjestuse kasutamine võimaldab efekti arvutuskiiruses, näiteks suuremõtmelise genotüüpiseerimise korral.

4 PCR multipleks

Tüüpilise ühe praimeripaariga (*single-plex*) PCR reaktsiooni asemel võib samas keskkonnas korraga läbi viia mitme erineva lõigu üheaegse amplifitseerimise. See võimaldab säästa genoomset materjali ning on leidnud kasutust näiteks deletsioonide, mutatsioonide ja polümorfismide analüüsil (Henegariu, et al., 1997). Ühes reaktsioonis amplifitseeritavate lookuste arv on siiski piiratud, kuna vajalik on vältida ebasoovitavaid interaktsioone erinevate praimerite ja produktide vahel. Kui vajalikku arvu lookusi ei ole võimalik amplifitseerida ühes reaktsioonis, on siiski võimalus lasta arvutil minimiseerida vajalike reaktsioonide hulka. Praimeripaaride jaotamisest (tüüpiliselt võimalikult väikesesse hulka) gruppidesse tulebki järgnevalt juttu.

4.1 Olulised parameetrid grupeerimisel

Olulised grupeerimise parameetrid võib jaotada:

4.1.1 Energia piirväärtused

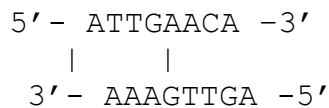
Kui oligonukleotiidide seostumisenergia ei ole väga suur, toimub paardumine piisavalt nõrgalt, et mitte alandada liiga palju PCR multipleks-reaktsiooni kvaliteeti. Seostumisenergiat arvutatakse mitmel viisil ning erinevatest oligonukleotiididest või nende lõikudest lähtuvalt. Mõned võimalikud variandid on:

1. **Praimer-praimer interaktsioonid, mõlemad otsad**, teisiti nimetatuna globaalne joondus. Seostumistugevuse arvutamiseks võrreldakse erinevaid joondusi, mis peavad sisaldama mõlema praimeri 3' otsa. Sellise joonduse korral ei oma tähtsust suhteliselt lühike keskel asetsev komplementaarne regioon, kuna suuremas osas mittekomplementaarsete praimerite korral jääb joondatud ala otste lähedal suhteliselt lühikeseks. Tegelikult tuleb kontrollida ka praimeri seondumist iseendaga, kuna praimer esineb reaktsioonis kindlasti mitmekordselt.

T								
G								
A								
A								
C								
A								
	A	A	A	G	T	T	G	A

Tabel 3. Globaalne kahe praimeri joondus - joondusesse peab jääma mõlemi praimeri 3' ots. Antud juhul ei arvestata joonduse leidmisel indelite võimalusega, kuna kasutatud MultiPLX versioon indeleid ei toetanud.

Joondus:



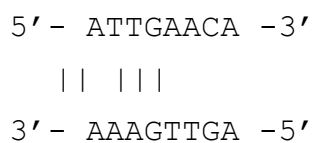
2. Praimer-praimer interaktsioonid, üks ots. Sellisel juhul nõutakse, et joondusesse jääks vähemalt ühe praimeri 3' ots.

Kaks võimalikku joondust, olenevalt sellest, kumba praimeri otsa joondusesse sissejäämist nõutakse:



3. Praimer-praimer interaktsioonid, suvaline regioon ehk lokaalne joondus. Sellisel juhul leitakse kõige tugevamini seonduv piirkond, mis võib sisalda ka praimeri 3' otsa, kuid ei pruugi.

Joonduse näide:



4. **Praimer-produkt interaktsioonid, üks ots.** Kasutatakse sama meetodit, mis praimer-praimer interaktsiooni arvutamisel.
5. **Praimer-produkt interaktsioonid, suvaline regioon.** Kasutatakse sama meetodit, mis praimer-praimer interaktsiooni arvutamisel.

Nagu eelpool mainitud, võib seostumisenergia asemel kasutada tinglikke skooore ja neile vastavaid *cutoff* väärtusi.

4.1.2 Produktide pikkuse erinevuse piirväärtused

Soovitav on omada ühes grupis primereid, mille produktide pikkuste erinevused ei ole liiga väikesed ega ka mitte liiga suured. Erineva pikkusega produktide korral toimuvad reaktsiooni komponendid erinevate kiirustega ja see raskendab multipleks-PCR-i. Kui produkti soovitakse nähtavaks teha elektroforeesil, nõuab suur produkti pikkuste erinevus suuremat skaalat. Kui aga reaktsiooni käigus tekib kaks väga sarnase pikkusega produkti, muudab see nende elektroforeesil eristamise raskeks .

4.1.3 Praimerite sulamistemperatuuride erinevus

Ideaaljuhul peaksid kõik praimeripaarid multipleks-PCR reaktsioonis vastavaid lookusi amplifitseerima sarnase efektiivsusega (Elnifro et. al, 2000). Sulamistemperatuuride erinevus põhjustab ka ebavõrdse amplifitseerimiseefektiivsuse, mis nõuab omakord praimerite kontsentratsiooni varieerimist.

4.1.4 Piirangud grupi suurusele

Ühes multipleks-PCR reaktsioonis võib korraga võimendada tüüpiliselt 2-14 (Henegariu, unpublished), mõnel juhul kuni 26 lookust (Lin et. al, 1996). Kinnitamata andmetel on õnnestunud reaktsioon läbi viia isegi ligikaudu 100 lookusega. Nii suure arvu lookuste kasutamine õnnestub siiski vaid parimatel juhtudel või spetsiaalvõtteid kasutades (neist räägib ptk. 5), tihti tuleb piirduda 5-10 lookusega (Hacia et al., 1998). Suure arvu lookuste puhul kannatab PCR kvaliteet, kuna siis on tarvis suuremat hulka primereid, see tähendab aga enamaid kitsendusi keskkonnatingimustele, samuti muutub märgatavaks paljude nõrgemate interaktsioonide mõju. Samas on eesmärk tavaliselt minimiseerida reaktsioonide arvu ning sellest tulevalt maksimeerida lookuste hulka.

4.2 Grupeerimise algoritm

Oma olemuselt võib grupeerimist vaadelda kui orienteerimata graafi jagamist minimaalseks hulgaks klikkideks (alamgraafideks, mis on täisgraafid s.t. kõik tipud on seotud kõigi teiste antud alamgraafi tippudega), mis on NP-täielik probleem (Garey, Johnson, 1979). Võimalik on ka lähenemine, mis kasutab multi-tippudega graafe ehk graafe, mis võib olla mitmes erinevas seisundis (Rachlin et. al, 2005). Igal juhul on optimaalset lahendust praktikas kasutatavate andmehulkade korral leida liiga arvutusmahukas ning tuleb rahul olla lihtsustatud algoritmi järgi leitud lähendiga.

5 Alternatiivid praimerite grupeerimisele

Eelnevalt vaadeldud grupeerimise korral kasutatakse praimeritena lühikesi oligonukleotiide, kaks praimerit iga lookuse kohta, ent võimalike interaktsioonide kontrollimine muutub keeruliseks kui neid on palju (Wang et. al, 1998). Erinevad autorid on kirjeldanud teistsuguseid lähenemisi mitmete lookuste üheaegseks amplifitseerimiseks.

5.1 PCR supressioonil baseeruv amplifitseerimine (Broude et al., 2001)

Autorid on arendanud meetodi, mis võimaldab multipleks-PCR-s lookuse kohta tavapärase kahe praimerit asemel kasutada ühte praimerit ning lisaks ühte praimerit kogu reaktsiooni kohta s.t. n lookuse amplifitseerimiseks läheb tarvis $n+1$ praimerit. Meetod põhineb nn. PCR supressiooni efektil (Lukyanov et al., 1994). Kasutatakse nn. adaptorpraimerit, mis seostub geenifragmentide lõppu ligeeritud n.ö. adaptoriga (Rosenthal, Jones, 1990). Adaptor on u. 40 aluspaari pikk ning kõrge GC sisaldusega. Pärast iga denaturatsioonisammu formeeruvad iseendaga komplementaarsetest fragmentide lõppudest dupleksid ning juuksenõelastruktuurid (*hairpin structures*). Adapterile vastavad praimerid (A-praimerid) ei replikeeri taolisi DNA fragmente, ent juuksenõelastruktuuri üheaahelalises silmuses (*loop*) asuva DNA-ga komplementaarsed praimerid (T-praimerid) küll. Saadud produkti amplifitseerivad nii A- kui T-praimerid. Seega toimub PCR ainult nõutavat sihtjärjestust (*target*) sisaldavates fragmentides. A-praimer aga on kõigile sihtjärjestustele.

Antud meetodit on õnnestunud kasutatud kuni 14 lookusega multipleks-PCR jaoks, eesmärgiga saada analüüse tsüstilist fibroosi põdevate inimeste genotüüpiseerimiseks.

Tüüpilise multipleks-PCR korral oleks nii suure arvu lookuste kasutamine tihti komplitseeritud (Hacia et al., 1998).

5.2 DOP-PCR (Jordan et. al, 2001)

Üks lähenemine unikaalsete genoomse järjestuse lõikude amplifitseerimiseks on kasutada PCR praimeritena lookustevaheliste kordusjärjestustega (näiteks Alu-järjestus) komplementaarseid järjestusi. Niimoodi saavutatud amplifitseeritavate genoomsete järjestuste hulk on siiski piiratud. Suurema amplifitseeritavate järjestuste hulga annab esialgu kogu genoomi amplifitseerimiseks loodud meetod degeneratiivsete oligonukleotiidipraimeritega PCR, DOP-PCR. DOP-PCR puhul kasutatakse praimereid, kus umbes üks kolmandik oligonukleotiidi keskel asuvatest positsioonidest on degeneratiivsed. Oligonukleotiid, mille 3' otsas on 6 unikaalset nukleotiidi, amplifitseerib efektiivselt suure osa eukarüootsest genoomist. Seda meetodit on kasutatud muuhulgas inimese ja hiire SNP-de genotüpiseerimiseks. Autorid on näidanud täielikult sekveneeritud genoomide puhul võimalikkust leida optimeeritud hulk DOP-PCR produkte.

II Praktiline osa

Eesmärk

1. Hinnata optimaalseks grupeerimiseks vajalikku iteratsioonide arvu, kui praimerid grupeeritakse ahne algoritmi (*greedy algorithm*) alusel juhuslikus järjekorras.
2. Leida praimerite grupeerimise efektiivsuse sõltuvus praimerite ja produktide pikkustest ning praimerite GC sisaldusest.
3. Leida praimerite grupeerimise efektiivsuse sõltuvus 1-2 nukleotiidist praimerid 3' otsas.

Töövahendid

1. Sobivate praimerite leidmiseks kasutati programmi PRIMER3 (Rozen et al. 2000), mis on kättesaadav aadressil <http://fokker.wi.mit.edu/primer3/>.
2. Grupeerimiseks kasutati programmi MultiPLX (Kaplinski et al. 2005), mis on kättesaadav Tartu Ülikooli Bioinformaatika Õppetooli serverist aadressil <http://bioinfo.ebc.ee/download/>.

Grupeerimiseks vajaliku korduste arvu leidmine

Katse skeem

1. Inimese genoomist 12000 SNP ümbrusest 2500 nukleotiidi pikkuse järjestuse leidmine (järjestused Fasta formaadis).
2. Järjestuste konverteerimine Fasta formaadist PRIMER3 sisendiks.
3. Praimerite leidmine kõigile järjestustele (saadi 11988 praimeripaari). Praimeritele esitatud nõudmised olid GC-sisaldus 45-55%, praimerite pikkus 18-24 ja produktide pikkus 100-500 nukleotiidi ning sulamistemperatuur 57-63 °C.
4. Järjestuste konverteerimine MultiPLX sisendformaati.
5. Järjestuste jagamine sisenditeks suurustega 10, 30, 100, 300, 1000 ja 3000 praimeripaari.
6. Kõikide sisendite grupeerimine praimereid juhuslikus järjekorras gruppidesse jagades iteratsioonide arvuga 1, 3, 10, 100, 300, 1000, 3000, 10000 ja 30000. MultiPLX parameetrid olid järgmised:

```
> cmultiplx -primers praimeritefail -calcscores 12345  
-savescores skooridefail
```

```
> cmultiplx -primers praimeritefail -scores skooridefail  
-calcgroups 10000 10000 -groupiter iteratsioonidearv  
-initialorder random -savegroups väljundfail
```

7. Gruppides esineva keskmise praimerite arvu leidmine.

Hüpotees

Suurema korduste arvuga saadakse parem tulemus, ent mingist korduste arvust alates tulemust enam parandada pole võimalik (tulemus on optimaalne).

Tulemused

Nagu näha, saavutatakse 10 praimeripaari grupeerimisel tõenäoline maksimum 3000 iteratsiooni juures (lisad, graafik 1). Suurema hulga praimeripaaride grupeerimisel näib soovitatav olevat läbi viia üle 30 000 iteratsiooni, mis on aga arvutuslikult väga mahukas.

Praimeri GC-sisalduse mõju grupeerimise efektiivsusele

Katse sai läbi viidud kahel viisil. Esimeses katses koostati erineva GC-sisaldusega praimerite hulgad ning võrreldi nende grupeerimise efektiivsust. Teises katses püüti imiteerida praktikas ettetulevat situatsiooni, erineva GC-sisaldusega primereid juhuslikult grupeerides ning üritati leida seaduspärasusi erineva GC-sisaldusega praimerite paigutamises gruppidesse.

Esimese katse skeem

1. Eelmises katses kasutatud 2500 nukleotiidi pikkustele genoomilõikudele praimerite leidmine. Praimerite keskmine GC-sisaldus varieerus vahemikus $30-65 \pm 2$ 5 nukleotiidi kaupa, ülejäänud parameetrid olid samad eelmises katses kasutatuga.
2. Praimerite jagamine 384 praimerite suurustesse juhuslikesse gruppidesse. Iga GC-sisaldusega praimerite jaoks leiti 10 gruppi.
3. Praimerite grupeerimine, kasutades programmi MultiPLX “suurima sõprade arvu” meetodit. Antud meetod on enamasti efektiivsem juhuslikust gruppidesse paigutamisest, kuna esimeses järjekorras grupeeritakse praimerid, mis sobivad kokku võimalikult väheste ülejäänutega. MultiPLX parameetrid olid järgmised:

```
> cmultiplx -primers praimeritefail -calcscores 12345  
-savescores skooridefail  
> cmultiplx -primers praimeritefail -scores skooridefail  
-calcgroups 10000 10000 -initialorder friends -savegroups
```

väljundfail

4. Keskmise grupi suuruse leidmine.

Hüpotees

GC-sisalduse muutudes muutub grupeerimise efektiivsus – keskmine praimerite arv grupis suureneb või väheneb.

Tulemused

Praimerite GC-sisaldusega üle 50% halveneb märgatavalt nende grupeeritavus (Graafik 2).

Teise katse skeem

1. Eelmises katses leitud erineva GC-sisaldusega praimeripaaride n.ö. segamine ühte faili.
2. Antud failis sisalduvate praimeripaaride jagamine kümneks 384 praimeripaari sisaldavateks sisendiks.
3. Praimerite grupeerimine. MultiPLX parameetrid samad, mis esimesel katsel.
4. Erinevate GC-sisaldusega praimerite leidmine ning gruppide, kus nad asuvad, keskmise suuruse leidmine. Tegu on kaalutud keskmisega – kui antud parameetritega praimer asub grupis mitu korda, võetakse see ka arvesse keskmise arvutamisel.

Hüpotees

Olenevalt praimerite GC-sisaldusest paikneb praimer tõenäolisemalt suurematesse või väiksematesse gruppidesse.

Tulemused

Nagu lisas olevast tabelist selgub, võib suurema GC-sisaldusega praimerite jaotumist väiksematesse gruppidesse aimata küll, kuid erinevus keskmistes gruppide suurustes on minimaalne (Tabel 4).

Praimeri ja produkti pikkuse mõju grupeerimise efektiivsusele

Sarnaselt GC-sisaldusele sai läbi viidud kaks katseseeriat, varieerides esimesel juhul praimerite pikkust kuueteistkümnest nukleotiidist kahekümne nelja nukleotiidini kahe kaupa, teisel juhul produkti pikkust vahemikus 100-1000 nukleotiidini 100 nukleotiidini kaupa, jättes GC-sisalduse vahemikku 45-55%. Tm väärtus oli piiratud 57-63 °C

olenemata praimerite pikkustest, mistõttu pikemate praimerite korral oli tõenäolisem madalam GC-sisaldus. MultiPLX parameetrid olid samad, mis eelmise katse korral.

Hüpotees

Praimeri või produkti pikkuse suurenedes muutub grupeerimise efektiivsus.

Tulemused

Seos praimerite pikkuse ja grupeerimise efektiivsuse vahel joonistus esimeses katses üsna selgelt välja – praimerite pikendamise korral kuni 24 nukleotiidini korral oli grupis keskmiselt rohkem primereid ehk grupeerimise efektiivsus oli märgatavalt suurem (Graafik 3). 26 nukleotiidi korral oli efektiivsus veidi tagasi langenud. Produkti pikkuse varieerimisel oli vastupidine mõju – pikem produkt tähendas madalamat grupeerimise efektiivsust, optimaalne praimerite pikkus oli 100 nukleotiidi (Graafik 4).

Teise katse korral ehk juhusliku pikkuse või produkti pikkusega primereid grupeerides seaduspära välja ei tulnud (Tabelid 5 ja 6).

Praimeri 3'-otsa mõju grupeerimise efektiivsusele

Varieeritud sai ühte nukleotiidi mõlema praimerite 3'-otsast. MultiPLX parameetrid samad, mis GC sisalduse mõju leidmise katse korral.

Hüpotees

Praimerite grupeerimise efektiivsus on erinev sarnaste, komplementaarsete või erinevate ja mittekomplementaarsete otste korral.

Tulemused

Esimese katse skeemi järgi sama tüüpi primereid gruppidesse paigutades olid erinevused märgatavad – kõige efektiivsemalt grupeerusid adenosiiniga lõppevad praimerid. Kõige ebaefektiivsem oli variant, kus üks praimer lõppes T- ja teine G-nukleotiidiga (Tabel 7). Teist katse skeemi järgides märgatavaid erinevusi ei ilmnenud (Tabel 8).

Arutelu

Siiani ei ole publitseeritud meetodit, mis suudaks mõistliku ajakuluga leida praimeripaaride grupeerimisel optimaalse lahendi, tõenäoliselt sellist algoritmi ka ei eksisteeri. Seega on eesmärk saada optimaalsele lahendile võimalikult lähedane tulemus ehk leida võimalikult väikese gruppide arvuga praimeripaaride jaotus. Esimesest katseseeriast selgub, et praktikas kasutatava suurusega hulga praimeripaaride juhuslik grupeerimine nõuab optimaalsele lahendile piisavalt lähedase tulemuse saamiseks üle 30 000 iteratsiooni, see on aga arvutuslikult väga mahukas. Sellepärast on ka teistes katsetes kasutatud nn. suurima sõprade arvu meetodit, mille puhul puudub iteratsioonide arvu mõju ja saadakse oluliselt väiksema ajakuluga tüüpiliselt vähemalt sama hea tulemus.

Edasistest tulemustest selgub, et teatud parameetritega praimereid grupeerimises kasutades on võimalik suurema tõenäosusega saada suurem grupeerimise efektiivsus ehk väiksem gruppide arv.

1. Praimerid GC-sisaldusega üle 50% on märgatavalt halvemini grupeeritavad. See võib tuleneda sellest, et G ja C nukleotiidide vahelise side on tugevam kui A ja T puhul.
2. Praimerid pikkusega 24 nukleotiidi olid kõige paremini grupeeritavad, väiksema pikkusega praimerid grupeerusid madalama multipleksi efektiivsusega. Ilmselt on põhjus lühemate praimerite mittespetsiifilisuses – märgatav osa praimerist võib seonduda võõra produktiga, kuna produktide pikkused olid antud katses juhuslikud. 24 nukleotiidist pikemad praimerid aga võisid halvemini grupeeruda tänu nende tõenäolisemale omavahel seondumisele. Võimalik, et pikemate praimerite puhul oleks grupeeritavuse hindamisel tulnud lubada tugevamat omavahelist seondumist ehk kasutada suuremat *cutoff* väärtust, sest ka spetsiifiline seondumine on neil eeldatavasti tugevam.
3. Produktide pikkuse ja grupeerimise efektiivsuse vahel oli üsna selge seos – pikemad produktid grupeerusid väiksema efektiivsusega. Ilmselt pakkus produktide suurem pikkus rohkem võimalusi võõraste praimerite või produktidega seondumiseks.
4. Kõige efektiivsemalt grupeerusid praimeripaarid, milles mõlema praimeri 3' otsas asus adenosiin. Kirjanduses on näidatud, et A-nukleotiid ja T-nukleotiid 3' otsas seondub G- või C-otsast nõrgemalt teiste praimerite või genoomiga, seega peaks grupeerimise

efektiivsus sellisel juhul kõrgem olema. See mõju ei ole antud juhul küll ilmselt ainus, kuna praimeripaarid, mille mõlemas otsas oli C-nukleotiid, grupeerusid efektiivsemalt kui kahe 3' T-nukleotiid otsaga praimeripaarid. Seega tuleks põhjuseid otsida võimalikust nukleotiidide ja dinukleotiidide jaotumisest genoomi SNP-de ümber.

Võib arvata, et saadud tulemuste järgi peaksid olema eriti efektiivsed väikese GC-sisaldusega (30%), 24 nukleotiidi pikkused 100-nukleotiidise produkti pikkusega praimerid, mille 3'-otsas on A nukleotiid. Seda tüüpi praimerid on praktikas üsna ebastabiilsed, hoolimata nende pikkusest, tänu väikesele GC-sisaldusele ja A-nukleotiidile 3'-otsas.

Juhuslikult disainitud praimereid gruppidesse paigutades seaduspära ei ilmnenud, küll aga ilmnis huvitav seaduspära – kasutades varieeruva pikkusega praimereid, kus teised parameetrid muutuvad vähe, on grupeerimise efektiivsus märgatavalt suurem, kui näiteks varieeruva produkti pikkusega praimereid grupeerides. Ilmselt on selline tulemus saadud tänu sellele, et pikad produktid (üle 500 nukleotiidi) seonduvad kergesti teiste produktide ja praimeritega.

Saadud tulemusi annab kasutada multipleksi jaoks praimereid disainides, kuid täpsemad tulemuste tagamaad vajaksid kindlasti põhjalikumat uurimist. Näiteks tuleks analüüsida inimese genoomi ühenukleotiidsete polümorfismide ümber asuva piirkonna nukleotiidide sisaldust.

Kokkuvõte

PCR ehk polümeraasahelreaktsioon on tänapäeval laialt kasutatav meetod tööks genoomse materjaliga. Rohkem kui ühe paari praimeritega PCR läbi viimist, eesmärgiga amplifitseerida mitu lookust, nimetatakse multipleks-PCR-iks. Multipleks-PCR korral halvendavad reaktsiooni kvaliteeti praimerite ja produktide omavahelised interaktsioonid, samuti praimerite interaktsioonid võõraste produktidega. Olenevalt praimerite omadustest on teoreetiliselt võimalik PCR läbi viia kuni 14 praimeripaariga. Antud töös on uuritud, kuidas erinevad praimerite parameetrid mõjutavad ühes reaktsioonis kasutatavat praimeripaaride arvu ehk praimeripaaride grupi suurust. Selle jaoks on läbi viidud eksperimendid, kus kõigepealt proovitakse grupeerida hulk sarnaste omadustega primereid, hiljem aga juhuslikke primereid ning jälgitud nende jaotumist erinevatesse gruppidesse. Uuritud on grupi suuruse sõltuvust praimerite pikkusest, GC-sisaldusest, produkti pikkusest ja 3'-otsas asuvast ühest nukleotiidist. Nagu selgub, on teatud parameetritega primereid grupeerides võimalik saada parem multipleksimise tase ehk suuremad grupid. Grupeerimise kvaliteet halveneb märgatavalt GC-sisalduse tõusmisel üle 50%, samuti produkti pikkuse kasvades. Kõige sobivamaks praimerite pikkuseks grupeerimises osutub 24 nukleotiidi. Kõige efektiivsemalt grupeeruvad praimeripaarid, millel mõlema praimerite 3' otsas asub adenosiin. Samuti selgub, et suur produkti pikkuste varieeruvus halvendab märgatavalt grupeerimise kvaliteeti, erinevalt näiteks praimerite pikkuste muutumisest.

Grouping efficiency of different PCR primers

Meelis Olev

Resume

PCR (polymerase chain reaction) is a widely used method for working with genomic material nowadays. So called multiplex-PCR uses more than two primers and is used to amplify more than one loci. In case of multiplex-PCR interactions between primers and products must be considered. Depending on properties of primers it is theoretically possible to use up to 14 primer pairs per reaction. This work is about how different properties of primers influence number of primers amplifiable in one reaction. For this purpose experiments are made: at first primers with similar properties, later random primers are grouped. Research covers how group size depends on primer length, GC content, product length and nucleotide at the 3' end. As experiment shows, primers with specific parameters constitute bigger groups, their multiplex level is higher. Quality of grouping declines noticeably when rising GC-content over 50 percent. That holds also when raising length of primer. The best primer for multiplexing is 24 nucleotides long and with adenosine in 3' end. It also dawns that great variability of length of product decreases grouping quality, differently from variability of length of primer.

Tänu

Tänu prof. Mairo Remmi ja Lauris Kaplinski juhendamise eest, samuti Reidar Andresoni.

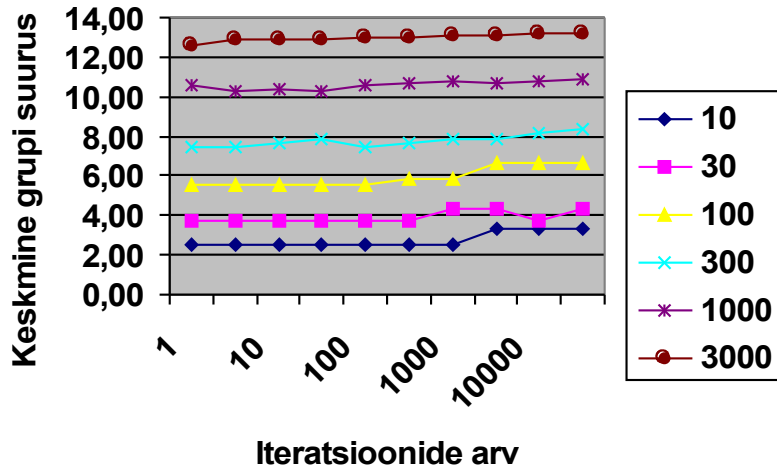
Viited

1. Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Research* 25(17), 3389-3402, 1997
2. Andreson R., Erinevate *in silico* meetodite võrdlus PCR praimerite kvaliteedi parandamiseks, magistritöö, Tartu Ülikool, 2001
3. Breslauer K. J., Frank R., Blocker H., Marky L. A. Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci U.S.A.* 83(11), 3746-3750, 1986
4. Broude N. E., Zhang L., Woodward K., Englert D., Cantor C. R. Multiplex allele-specific target amplification based on PCR suppression. *Proc Natl Acad Sci USA* 98 (1), 206-11, 2001.
5. Brownie J., Shawcross S., Theaker J., Whitcombe D., Ferrie R., Newton C., Little S. The elimination of primer-primer accumulation in PCR. *Nucleic Acids Research* 25 (16), 3235-3241, 1997
6. Burpo F. J. "A critical review of PCR primer design algorithms and cross-hybridization case study", *Biochemistry* 218, 2001
7. Delcher A. L., Kasif S., Fleischmann R. D., Peterson J., White O., Salzberg S. L. Alignment of Whole Genomes. *Nucleic Acids Research* 30(11), 2478-2483, 2002
8. Dong S., Wang E., Hsie L., Cao Y., Chen X., Gingeras T. R. Flexible use of high-density oligonucleotide arrays for single-nucleotide polymorphism discovery and validation. *Genome Research* 11:1418-1424, 2001
9. Elnifro E. M. Ashshi, A. M. Cooper R. J., Klapper P. E. Multiplex PCR: Optimization and Application in Diagnostic Virology. *Clinical Microbiology Reviews*, 559-570, Oct, 2000
10. Glick B., Pasternack J. *Molecular Biotechnology: Principles and Applications of Recombinant DNA*. ASM Press, Washington, D.C., 1998
11. Garey M. R., Johnson D.S. *Computers and Intractability*. W. H. Freeman, San Francisco, 1979
12. Haas S., Vingron M., et al. Primer Design for large scale sequencing. *Nucleic Acid Research* 20(17), 4567-4573, 1998
13. Hacia J. G., Sun B., Hunt N., Edgemon K., Mosbrook D., Robbins C., Fodor S. P., Tagle D. A., Collins F. S. Strategies for mutational analysis of the large

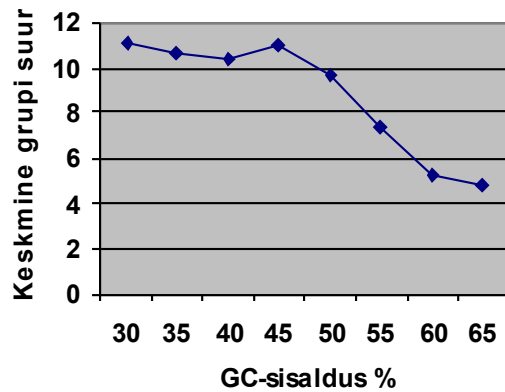
- multiexon ATM gene using high-density oligonucleotide arrays. *Genome Research* 8 (12), 1245-1258, 1998
14. Henegariu O., Heeremaa N. A., Dlouhy S. R., Vance G. H., Vogt P. H. Multiplex PCR: Critical Parameters and Step-by-Step Protocol. *Biotechniques* 23, 504-511, 1997
 15. Höhl M., Kurtz S., Ohlebusch E. Efficient multiple genome alignment. *Bioinformatics* 18, 312S-320S, 2002
 16. Innis M.A., Gelfand D.H. Optimization of PCRs. *PCR Protocols*, Academic Press, New York, 1990
 17. Jordan B., Charest A., Dowd J. F., Plumenstiel J. P., Yeh R., Osman A., Housman D. E., Landers J. E. Genome complexity reduction for SNP genotyping analysis. *PNAS* 99 (5), 2942-2947, 2002
 18. Kaderali L, Schliep A. Selecting signature oligonucleotides to identify organisms using DNA arrays. *Bioinformatics* 18, 1340-1349, 2002
 19. Kampke T., Kieninger M., Mecklenbug M. Efficient primer design algorithms. *Bioinformatics* 17(3): 214-225, 2001
 20. Kaplinski L., Andreson R., Puurand T., Remm M. MultiPLX: automatic grouping and evaluation of PCR primers. *Bioinformatics* 21(8), 1701-2, 2005
 21. Kennedy G. C., Matsuzaki H., Dong S., Liu W. M., Huang J., Liu G., Su X., Cao M., Chen W., Zhang J. et al. Large-scale genotyping of complex DNA. *Nat. Biotechnol.* 21:1233-1237, 2003
 22. Li D., Vijg J. Multiplex co-amplification of 24 retinoblastoma gene exons after pre-amplification by long-distance PCR. *Nucleic Acids Research* 24 (3), 538-539, 1996
 23. Li P., Kupfer K. C., Davies C. J., Burbee D., Evans G. A., Garner H.R. PRIMO: A Primer Design Program That Applies Base Quality Statistics for Automated Large-Scale DNA Sequencing. *Genomics* 40(3): 476-85, 1997
 24. Lin Z., Cui X., Li H. Multiplex Genotype Determination at a Large Number of Gene Loci. *Proc Natl Acad Sci USA* 93(6): 2582-2587, March 19, 1996
 25. Liu W.-m., Di X., Yang G., Matsuzaki H., Huang J., Mei R., Ryder T. B., Webster T. A., Dong S., Liu G, et al. Algorithms for large scale genotyping microarrays. *Bioinformatics* 19: 2397-2403, 2003
 26. Lukyanov, S. A. et al. *Bioorganic Chemistry (Russia)* 20, 701-704, 1994

27. Matsuzaki H., Loi H., Dong S., Tsai Y., Fang J., Law J., Di X., Liu W., Yang G., Liu G., Huang J., Kennedy G. C., Ryder T. B., Marcus G. A., Walsh P. S., Shiver M. D., Puck J. M., Jones K. W., Mei R. Parallel Genotyping of Over 10,000 SNPs Using a One-Primer Assay on a High-Density Oligonucleotide Array. *Genome Research* 14:414-425, 2004
28. Michael A. Innis, David H. Gelfand, John J. Sninsky "PCR Applications. Protocols For Functional Genomics" Academic Press, 1999
29. Rachlin J., Ding C., Cantor C., Kasif S. Computational tradeoffs in multiplex PCR assay design for SNP genotyping. *BMC Genomics* 6, 102-113, 2005
30. Rozen S., Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols in the series Methods in Molecular Biology*. Humana Press, Totowa, NJ, 365-386, 2000
31. Schuber A. P., Grondin V. J., Klinger K. W. A simplified procedure for developing multiplex PCRs. *Genome Research* 5 (5), 488-493, 1995
32. Siebert P. D., Chenchik A., Kellogg D. E., Lukyanov K. A., Lukyanov S. A. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research* 23 (6), 1087-1088, 1995
33. Smith T. F., Waterman M. S. Identification of common molecular subsequences. *Journal of Molecular Biology* 147 (1), 195-197, 1981
34. Suggs S., et al. ICN-UCLA Symposia on Developmental Biology Using Purified Genes. Academic Press Inc., New York, NY, Vol 23, pp683-693., 1981
35. Sugimoto N., Nakano S., Yoneyama M., Honda K. Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Research*. 24, 4501-4505, 1996
36. Wang, D. G., Fan J. B., Siao C. J., Berno A., Young P., Sapolsky R., Ghandour G., Perkins N., Winchester E., Spencer J., et al. *Science* 280, 1077-1082, 1998

Lisad



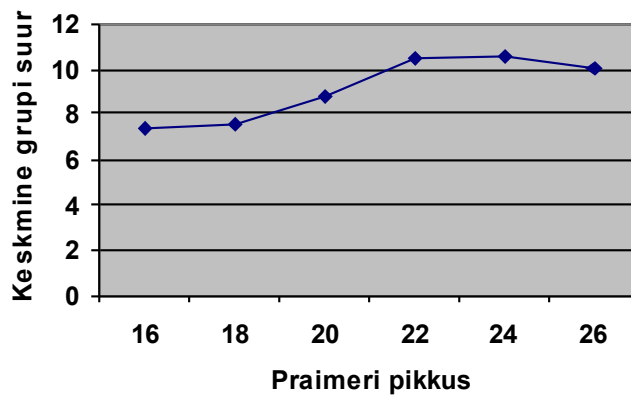
Graafik 1. Keskmine grupi suurus olenevalt grupeerimise iteratsioonide arvust. Erinevad jooned tähistavad erinevaid katseid erineva grupeeritavate praimerite arvuga.



Graafik 2. Keskmine grupi suurus olenevalt praimerite GC-sisaldusest

GC%	30	35	40	45	50	55	60	65
Grupi suurus	10,22	10,20	10,21	10,22	10,20	10,19	10,18	10,16

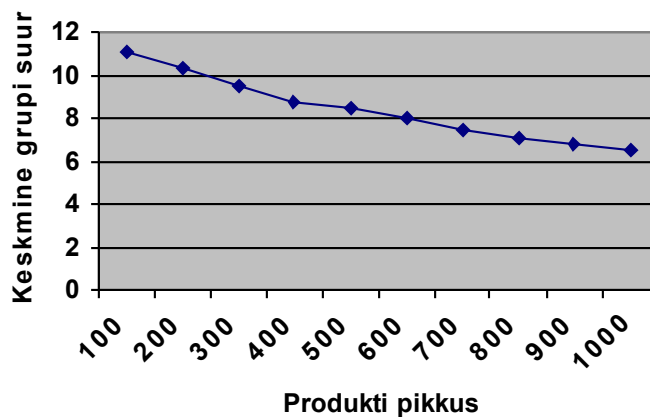
Tabel 4. GC-sisalduse mõju praimerite jaotuvusele erineva suurusega gruppidesse.



Graafik 3. Keskmise grupi suuruse sõltuvus praimeri pikkusest.

Praimeri pikkus	16	18	20	22	24	26
Grupi suurus	11,00	10,99	11,02	11,00	11,01	10,98

Tabel 5. Erineva pikkusega praimerite jaotumine erineva suurusega gruppidesse.



Graafik 4. Keskmise grupi suuruse sõltuvus produkti pikkusest.

Produkti pikkus	100	200	300	400	500	600	700	800	900	1000
Grupi suurus	8,87	8,88	8,88	8,88	8,89	8,87	8,89	8,87	8,89	8,87

Tabel 6. Erineva produkti pikkusega praimerite jaotumine erineva suuruse gruppidesse.

Esimese praimer 3'-ots	A	A	A	A	T	T	T	G	G	C
Teise praimer 3'-ots	A	T	G	C	T	G	C	G	C	C
Grupi suurus	11,8	10,3	10,2	10,7	9,8	8,7	9,8	9,2	9,0	10,2

Tabel 7. Erinevate 3'-otstega praimerite grupeerimise efektiivsus.

Esimese praimer 3'-ots	A	A	A	A	T	T	T	G	G	C
Teise praimer 3'-ots	A	T	G	C	T	G	C	G	C	C
Grupeerimise efektiivsus	10,20	10,22	10,18	10,20	10,23	10,20	10,19	10,19	10,19	10,21

Tabel 8. Erineva 3' nukleotiidiga praimerite jaotumine erineva suurusega gruppidesse.