BIOLOOGIA-GEOGRAAFIA TEADUSKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
BIOINFORMAATIKA ÕPPETOOL

Meelis Olev

PCR praimerite automaatne multipleksimine

Keskastme lõputöö

Juhendaja: Maido Remm, Prof., PhD

Tartu, 2003

Sisukord

Lühendid ja mõisted 4
I Teoreetiline osa 5
1 PCR 5
1.1 PCR praimerite disain 5
1.2 PCR praimerite disainiks kasutatavad programmid 6
2 PCR kvaliteeti mõjutavad parameetrid 6
2.1 Praimerite sulamistemperatuur 6
2.2 Kordusjärjestused 7
2.3 Polü-N traktid 7
2.4 Praimerite vahelised interaktsioonid 7
2.5 Produkti pikkus 7
2.6 Praimeri pikkus 7
2.7 GC nukleotiidide sisaldus ..... 8~~7~~
3 Seostumise arvutamise meetodid 8
3.1 Aukudeta joondus 8
3.2 Smith-Waterman 8
3.3 BLAST 9
3.4 Suffiksipuud 9
3.5 Räsialgoritmid (*hashing algorithms*) 9
3.6 Kui pikka ala praimerist arvestada? 9
3.7 Globaalne ja lokaalne joondus 9
4 PCR multipleks 9
4.1 Olulised parameetrid multipleksil 9
4.2 Multipleksi kvaliteeti mõjutavad tegurid 9
5 Alternatiivsed praimerite grupeerimise meetodid 9
5.1 xxx 9
II Praktiline osa 10
1 Programmi Multiplx üldine kirjeldus 10
2 Parameetrite ja käskude kirjeldus 10
2.1 Käsud 10

2.2 Võtmed 11
3 Skoori arvutamise meetodid 12
Viited ..... 15~~13~~

Lühendid ja mõisted

PCR - polümeraasahelreaktsioon (*polymerase chain reaction*)

praimer - ... (*primer*)

matriits - ... (*template*)

I Teoreetiline osa

1 PCR

PCR (Polymerase Chain Reaction e. polümeraasi ahelreaktsioon) on molekulaarbioloogias laialt rakendatav meetod, mida võib kasutada genotüpiseerimisel, punktmutatsioonide uurimisel, genoomse DNA kloneerimisel jm. Eesmärk on saada ühest geenilõigust palju koopiaid, mida tänu suurele hulgale on näiteks hiljem lihtsam detekteerida. PCR koosneb kolmest etapist, mida korratakse tsükliliselt:

- denaturatsioon – genoomse DNA puhul temperatuuril 93-95 °C. DNA kaksikahel lõhutakse 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 ja katkevad vesiniksidemed. Stabiilsemale kaksikahelale saab seostuda polümeraas ning algab matriitsi duplitseerimine.
- ekstensioon e. DNA süntees (70-75 °C). Polümeraas lisab dNTP-sid 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 tsüklis.

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

1.1 PCR praimerite disain

PCR teostamiseks on vaja teada ~~üles~~amplifitseeritava DNA järjestust~~i~~, mille alusel disainida praimerid – 15-30-meersed sünteetilised nukleotiidid. Praimerite disainil tuleb arvestada järgnevaga:

- praimerite sulamistemperatuur (T_m) peaks olema 50-70 °C
- praimerite GC sisaldus peaks olema 50-55%, väiksema GC sisalduse puhul tuleb sulamistemperatuuri tõstmiseks üle 50 °C disainida praimerid pikkusega üle 20 nukleotiidi
- vältida tuleks ühe või kahe nukleotiidi kordusi, eriti 3' otsas, kuna see võib põhjustada sekundaarseid hübridisatsioone

- praimeri 3’ ots on stabiilsem, kui seal asub G või C nukleotiid
- praimerid ei tohiks moodustada sekundaarstruktuure ning di- ja oligomeere

~~*** kirjutada lühidalt, kuidas programmid praimereid disainivad ***~~

1.2 PCR praimerite disainiks kasutatavad programmid

Suured molekulaarbioloogias kasutavad programmide paketid nagu EMBOSS ja GCG sisaldavad programme praimerite disainimiseks, vastavalt Prima ja Prime. Laialt tuntud ning ka veebibrauseri kaudu kasutav on Primer3. Erinevate tööpõhimõtete ja sisendandmetega on programmid PrimerGen (DOS), Primer (Stanford) (Sun Sparcstations), Primer (Whitehead) (Unix, Vms), Amplify (Mac), PrimerDesign (DOS), PC-Rare (Mac, Windows). Veebibrauseri kaudu kasutatavad on veel NetPrimer, GeneFisher, xprimer, Web Primer (Stanford), EMBL Primer Design. Eraldi võiks mainida Primer3 põhjal massgenotüpiseerimise jaoks arendatud programmi BioPrimer. *** kirjutada lühidalt, kuidas programmid praimereid disainivad ***

2 PCR kvaliteeti mõjutavad parameetrid

PCR kvaliteedi all tuleb mõista põhiliselt seda, kuivõrd efektiivselt seostuvad praimerid õigetesse kohtadesse ning seostumise stabiilsust. Kvaliteedi määravad põhiliselt järgnevalt toodud parameetrid.

2.1 Praimerite sulamistemperatuur

Lihtsustatud lähenemine praimerite sulamistemperatuuri arvutamisele on lisada 2 ºC iga A-T paari ning 4 ºC iga G-C paari kohta (Suggs et al, 1981). Seda nimetatakse ka Wallace’i reegliks. Enamasti kasutatakse täpsemat valemit ~~(~~

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

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. Iseendaga komplementaarse ahela puhul kasutatakse c/4 asemel valemis c. Sellist valemit kasutab oma töös näiteks programm Primer3.

--- uurida teiste programmide kohta, milliseid Tm arvutamise valemeid nad kasutavad

2.2 Kordusjärjestused

Kordusjärjestused, mis valku ei kodeeri (nt. Alu kordusjärjestused), moodustavad inimese genoomist 50% või üle selle. Et vältida mittespetsiifilist seondumist genoomis korduvatele lõikudele, testivad mitmed programmid (nt. PRIMER) praimeri komplementaarsust genoomsete kordusjärjestustega. või välditakse kordusi heksa/oktameeride sagedustabeli kaudu.

2.3 Polü-N traktid

Sarnaselt kordusjärjestustele põhjustab ka ühe nukleotiidi paljukordne järjest esinemine seostumise mittespetsiifilisust. Samuti korduvad di ja trinukleotiidid. Täpse äratundmise probleem.

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 praimereid 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

Lühemad PCR produktid seostuvad väiksema tõenäosusega valede APEX praimeritega. Liiga pikk ei tööta hästi PCRis. 100-1000 mõistlik vahemik. Samas (APEX) PCRi reaktsioonis soovitav omada sarnase pikkusega produkte, et tagada võrdsed kogused ja ühtlustada masina tsükli aegu.

--- mis põhjustel veel ebasobiv produkti pikkus halvendab PCR kvaliteeti?

2.6 Praimeri pikkus

Tavaliselt loetakse küllaldaseks praimeri pikkusi alates 18 nukleotiidist. Kindel 18 aluse pikkune oligonukleotiid esineb tõenäosusega 4^{18} = ~7 x 10^{10}, mis peaks seostuma ühte positsiooni enamikes eukarüootsetes genoomides. Lühemad praimerid seostuvad sellise suurusega genoomides suure tõenäosusega mitmesse kohta, kuid neid võib kasutada näiteks juhul, kui matriitsiks (*template*) on plasmiidne DNA. Kuna genoomi järjestus pole juhusliik, siis on tegelikult vaja pikemaid (soovitavalt AT-

rikkaid, et välistada mittespetsiifilist seondumist) praimereid. Praimeri pikkus mõjutab Tm'i.

2.7 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 valeseostumisi ja halvendada PCR kvaliteeti (Li et al, 1997). GC sisaldus mõjutab omakorda ka sulamistemperatuuri.

3 Seostumise arvutamise meetodid

PCR modelleerimisel tuleb arvutada praimerite seostumistei arvu ja seostumistugevust) teiste praimerite, produktide või genoomiga. Vahel arvutatakse välja seostumise tugevus absoluutskaalal, mõnikord aga lubatakse lihtsustusi ja loetakse võrdseks A-T ja G-C seondumise tugevus ning tihti arvutatakse seostumise tugevus suhtelisel skaalal, n.ö. skoorina. Grupeerimise seisukohalt on olulisim arvestada praimer-praimer interaktsioone. Seostumiste leidmisel kasutatakse järgnevaid meetodeid:

3.1 Aukudeta joondus

Aukudeta 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. Selline meetod on keerukusega O(m*n), kus m on ühe ja n teise järjestuse pikkus, kuid mõnel juhul võib tegeliku seondumise käigus osa nukleotiide paardumata jääda, mida antud algoritm ei arvesta.

--- seletada näite varal algoritm lahti?

3.2 Smith-Waterman

Smith-Watermani algoritm on nn. dünaamilise programmeerimise algoritm.

--- samuti näide tuua

3.3 BLAST

3.4 Suffiksipuud

3.5 Räsialgoritmid (*hashing algorithms*)

3.6 Kui pikka ala praimerist arvestada?

3.7 Globaalne ja lokaalne joondus

4 PCR multipleks

~~4.1 Olulised parameetrid multipleksil~~4.1 Olulised parameetrid grupeerimisel

4.2 ~~Multipleksi~~ Grupeerimise kvaliteeti mõjutavad tegurid

5 Alternatiivsed praimerite grupeerimise meetodid

5.1 xxx

II Praktiline osa

EESMÄRK

TULEMUSED

1 Programmi Multiplx üldine kirjeldus

Programmi Multiplx esialgne versioon töötab Unix-keskkonnas käsumoodis. Programmile antakse parameetritena ette ülesanne ning tööks vajalikud failid. Enamikel parameetritel on vaikeväärtused, mida saab muuta käsurealt, mõnel juhul ka konfiguratsioonifailist.

2 Parameetrite ja käskude kirjeldus

Järgnevalt on toodud programmi viimases versioonis kasutatavad käsud. Programmi põhikasutuseks on mõeldud käsk *gg*, aga teised käsud võivad vajalikud olla näiteks üksikute etappide eraldi teostamiseks, konfiguratsiooni ja sisendfailist lugemise testimiseks.

2.1 Käsud

la Arvutab kahe parameetrina ette antud nukleotiidijärjestuse lokaalse joonduse skoori

```
$ multiplx la ATGCTGT GATCTGA
3
```

ga Arvutab kahe parameetrina ette antud nukleotiidijärjestuse globaalse joonduse skoori

```
$ multiplx ga ATGCTGT GATCTGA
2
```

ea Arvutab kahe parameetrina ette antud nukleotiidijärjestuse lokaalse nn. lõppjoonduse (joondatav ala ei tohi asuda järjestuse keskel)

```
$ multiplx ea ATGCTGT GATCTGA
2
```

ld Näitab kahe parameetrina ette antud nukleotiidijärjestuse lokaalset joondust

```
$ multiplx ld ATGCTGT GATCTGA
CTG
CTG
```

gd Näitab kahe parameetrina ette antud nukleotiidijärjestuse globaalset joondust

```
$ multiplx gd ATGCTGT GATCTGA
CTGT
CTGA
```

ea Näitab kahe parameetrina ette antud nukleotiidijärjestuse nn. lokaalset lõppjoondust

```
$ multiplx ed ATGCTGT GATCTGA
CTGT
CTGA
```

er Näitab infot sisendfaili kohta (multiplx.dat või –i parameetriga etteantav)

```
$ multiplx er
multiplx.dat: 1278 item(s)
```

ei Kuvab valitud numbritega ridadest andmed

```
$ multiplx ei 1 2 3
/.../
```

pt Koostab nõutud skooride tabeli. Esimene parameeter näitab skoori tüüpi, neli järgmist parameetrit on optsionaalsed ja näitavad esimese ja teise indeksi miinimum- ja maksimumväärtust. Skooride tüübid on järgmised:

1 – praimer/praimer lokaalne joondus (PRIMER_SELF_ANY)
2 – praimer/praimer globaalne joondus (PRIMER_SELF_END2)
3 – praimer/produkt lokaalne joondus (PRIMER_PRODUCT_ANY)
4 – praimer/praimer 3'-lõppjoondus (PRIMER_SELF_END1)
5 – primer/produkt 3'-lõppjoondus
6 – produkti pikkuste vahe
>=7 – plugin-ide arvutavad skoorid ja tuleviku rakendused

```
$ multiplx pt 4 1 3 1 3
6       3       4
3       4       3
4       3       3
```

jt Ühendab skooride tabelid, mis on saadud nt. pt käsuga, kokku üheks tabeliks, kasutades antud nivooväärtusi (*cutoff*)

```
$ multiplx pt 1 1 3 1 3 > table.1
$ multiplx pt 2 1 3 1 3 > table.2
$ multiplx pt 3 1 3 1 3 > table.3
$ multiplx jt table.1 5 table.2 2 table.3 15
6       0       1
0       5       0
1       0       8
```

jg Jaotab (eeldatavasti käsu jt tulemusena saadud) antud tabeli põhjal praimeripaarid gruppideks. Teise parameetriga antakse ette maksimaalne skoori väärtus, mille korral praimeripaarid võivad veel ühte gruppi kuuluda

```
$ multiplx jt table.1 5 table.2 2 table.3 9 > table.j
$ multiplx jg table.j 0
Group 1: 1 2
Group 2: 3
```

gg Moodustab ja ühendab skooritabelid (mälus, ajutisi faile ei tekitata) ning grupeerib, andes väljundiks sisendi, kuhu on esimesse tulpa lisatud grupi number. Erinevus jg käsust on veel see, et ette antakse grupi minimaalne ja maksimaalne suurus ning vajadusel ka maksimaalne väljavisatavate praimeripaaride arv

```
$ multiplx gg pt1=5 pt2=2 pt3=9 min=1 max=2 drop=0
1       8_9GH1  367     GGAGACCTGTAGTCAGAGCCAAAAA
/.../
```

2.2 Võtmed

-e n Lubab joonduses maksimaalselt n auku (*gap*), vaikimisi 1

-g x Annab skoori joonduses oleva augu (*gap*) pikenemisel ühe nukleotiidi võrra,

	vaikimisi 2
-i file	Määrab sisendfaili nime, vaikimisi multiplx.dat. - failinimena tähendab standardsisendit
-m x	Määrab joondamisel skoori nukleotiidide seostumisele (*match*), vaikimisi 1
-n x	Määrab joondamisel skoori nukleotiidide mittesobivusele (*mismatch*), vaikimisi –1
-o file	Määrab väljundfailinime, vaikimisi standardväljund (-)
-r n	Proovib grupeerimisel kasutada n erinevat järjestust, vaikimisi 0 s.t. juhuslikke järjestusi ei kasutata
-u x	Määrab skoori tundmatu nukleotiidi või tundmatute nukleotiidide seostumisele, vaikimisi –0.25

3 Skoori arvutamise meetodid

...

4 Grupeerimismetoodika

5 Tulemused (*proof of principle*)

Arutelu

Kokkuvõte (kahes keeles)

Viited

1. Michael A. Innis, David H. Gelfand, John J. Sninsky "PCR Applications. Protocols For Functional Genomics" Academic Press, 1999
2. Glick, B. and Pasternack, J. Molecular Biotechnology: Principles and Applications of Recombinant DNA. ASM Press, Washington, D.C., 1998
3. Suggs, S., et al. ICN-UCLA Symposia on Developmental Biology Using Purified Genes. Academic Press Inc., New York, NY, Vol 23, pp683-693., 1981
4. Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. Nucleic Acids Research. 24, 4501-4505.
5. 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
6. ----------
7. ---- !!!!!!!!!!sellest peale tulevad artiklid, millele pole viidatud, aga mida on töös kasutatud!!!!!!!!!
8. ----------
9. F. John Burpo "A critical review of PCR primer design algorithms and cross-hybridization case study", Biochemistry 218, 2001
10. Programmi Primer Premier abiinfofail http://www.premierbiosoft.com/primerdesign/index.html
11. A. Kurg, Molekulaarse biotehnoloogia loengukonspekt, 2003 http://www.biotech.ebc.ee/biotech.materjalid.html
12. Praimerite disainimise programmide lühikirjeldused http://bip.weizmann.ac.il/mb/bioguide/pcr/software.html
13. PCR kirjeldus ja joonised http://allserv.rug.ac.be/~avierstr/principles/pcr.html
14. Sulamistemperatuurist rääkimine http://www.sigma-genosys.com/oligo_meltingtemp.asp
15. Primer3 sulamistemperatuuri kohta http://binfo.ym.edu.tw/shen/meeting/Primer3_help.htm

16. viide kordusjärjestuse kohta - http://www.ornl.gov/TechResources/Human_Genome/publicat/primer2001/4.html
17. jätkub...