

Ahne algoritm praimerite arvu miniseerimiseks korduvatel PCR reaktsioonidel

Meelis Olev

PCR

- Polümeraasahelreaktsioon amplifitseerib (võimendab) DNA järjestuse kahe praimeri vahel
- Üks praimeripaar ühele genoomi lõigule
- Praimeri hind on võrdeline tema pikkusega

Paljude praimeritega PCR

- Tihti on vaja amplifitseerida mitu genoomi lõiku
- Multipleks PCR – niipalju kui on amplifitseeritavaid lõike, on ka praimeripaare
- AP-PCR – suvalised 8-12 nukleotiidised praimerid

Vajadus vähendada praimerite arvu

- AP-PCR täiendus – praimerid asendada sekventsist tuletatutega ning võrrelda ennustatud ja eksperimentaalseid produkte
- Minimaalse sihtjärjestuste amplifitseerimiseks vajaliku praimerite hulga leidmine oleks majanduslikult kasulik
- Pärmis (6000 ORF-i) puhul on amplifitseeritavaid sihtjärjestusi liiga palju, et neid elektroforeesil eristada

Geeniekspressioonimustrite leidmine

- Viimastel aastatel on saadud geeniekspressioonimustreid oligo-kiipe ja DNA mikrokiipe
- Saadud andmed on suhteliselt ebatäpsed
- AP-PCR ja tema täiendatud variandid annavad täpsemaid andmed
- Koichiro Doi ja Hiroshi Imai meetod annab juba päris täpsed tulemused!

Teised praimerite selektsiooni programmid

- Nicodéme ja Steyart – multipleks PCR
- Pearson jt. - minimaalne praimerite hulk kõigi DNA sekventsides katmiseks, ei arvesta praimerite orientatsiooni
- Doi ja Imai – praimerite arvu minimiseerimine ühekordse PCR jaoks

Käesolev programm / artikkel

- Tegeleb praimerite valikuga korduva PCR jaoks
- Algul disainitakse primereid ühekaupa, kasutades ühekordse PCR praimerite leidmise ahnet algoritmi
- Hiljem täiendatakse algoritmi ning testitakse pärmis genoomi peal
- Niimoodi võideti praimerite hinnas 5 korda võrreldes multipleks PCR-ga

Bioloogilised piirangud praimeritele

- GC sisaldus 40-60%
- Pikkus – lühikesed (8-12 nukleotiidi)
- Paarikaupa / iseendaga komplementaarsus
 - 5'-GCCTAGGC-3'
 - 5'-GACAATGC-3', 5'-GCATTGTC-3'
- Produkti pikkus – 50-500 aluspaari
- Produktide pikkuste erinevus elektroforeesil eraldamiseks vähemalt 5

Ülesande püstitus

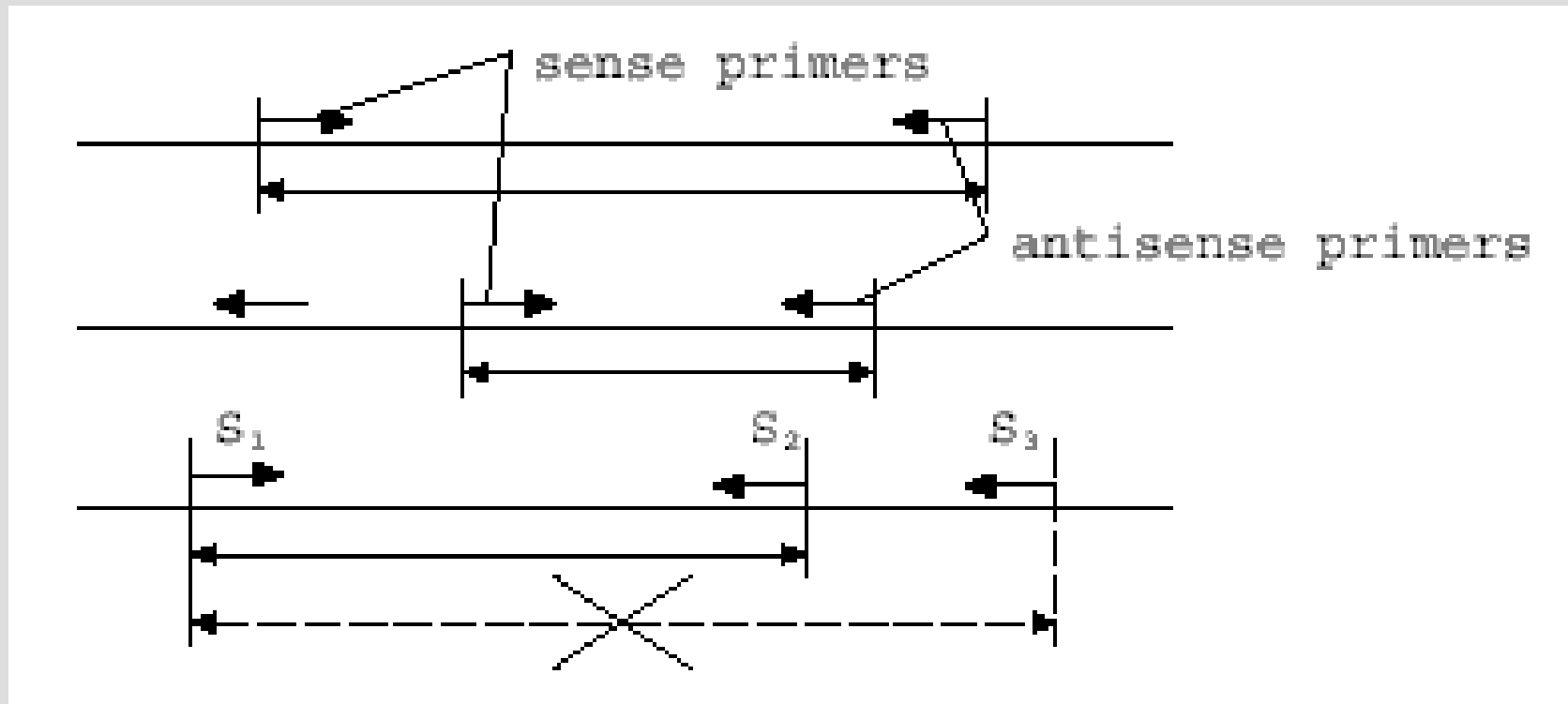
- PSP (praimerite valiku probleem)
 - Leida minimaalne praimerite alamhulk S' , mis sisaldab kõiki DNA järjestusi
 - Minimiseerida $|S'|$ s.t. $\forall j \exists S_i, S_i' \in S', (j, p_1, r) \in S_i,$
 $(j, p_2, l) \in S_i', p_2 - p_1 + 1 > 0$
 - $(j, p_1, r) \in S_i$ – praimer i sobib DNA järjestuse j positsiooni p orientatsiooniga paremale
 - $(j, p_2, l) \in S_i'$ – praimer i' sobib DNA järjestuse j positsiooni p orientatsiooniga vasakule

Ülesande püstitus 2

- Teoreem 1. PSP jaoks leidub ahne algoritm polünomiaalse keerukusega $O(\ln mn)$,
 $m = \max(m_1, \dots, m_n)$
 - m_1, \dots, m_n – amplifitseeritavate DNA järjestuste pikkused
 - n – amplifitseeritavate DNA järjestuste arv
- Teoreem 2. PSP-d ei saa hinnata polünomiaalse aja jooksul faktoriga $(1-\varepsilon)\ln n$ v.a. juhul $NP \subset TIME(n^{O(\log \log n)})$

Ülesande püstitus 3

- Amplifikatsiooni kitsendus kodeeriva ja mittekodeeriva ahela praimeride vahel



Ülesande püstitus 4

- PSPMDLC – leida minimaalne praimerite alamhulk, mis amplifitseerib kõik vajalikud DNA järjestused nii, et amplifitseeritud DNA jupid on pikkusega vähemalt k ja erinevusega vähemalt d .
- Ühekordne PCR ei amplifitseeri kõiki sihtjärjestusi
- PSPME – minimiseerime $\sum_e |S^e|$
 - S^e – praimeripaaride hulk
- Autorid võtavad $d=1$

Algoritmide raamistik

- A1 – algoritm praimerite arvu minimiseerimiseks ühekordse PCR puhul
 - Lõpptingimus: leitud 100 50-500 nukl. pikkust amplifitseeritud segmenti
- Mitmekordse PCR puhul
 - 1. Rakendada A1 kuni tema lõpptingimuseni
 - 2. Eemaldada uued DNA sihtjärjestused, mille jaoks A1 praimerid leidis
 - 3. Kui kõigi järjestuste jaoks on vähemalt üks praimeripaar või A1 uutele sihtjärjestustele primereid ei andnud, siis lõpetada, muidu pöörduda sammuni 1

Algoritm ühekordse PCR jaoks

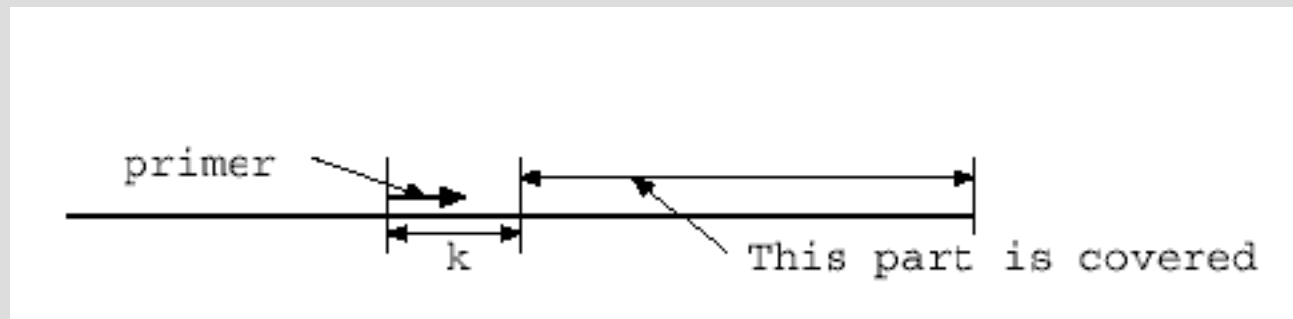
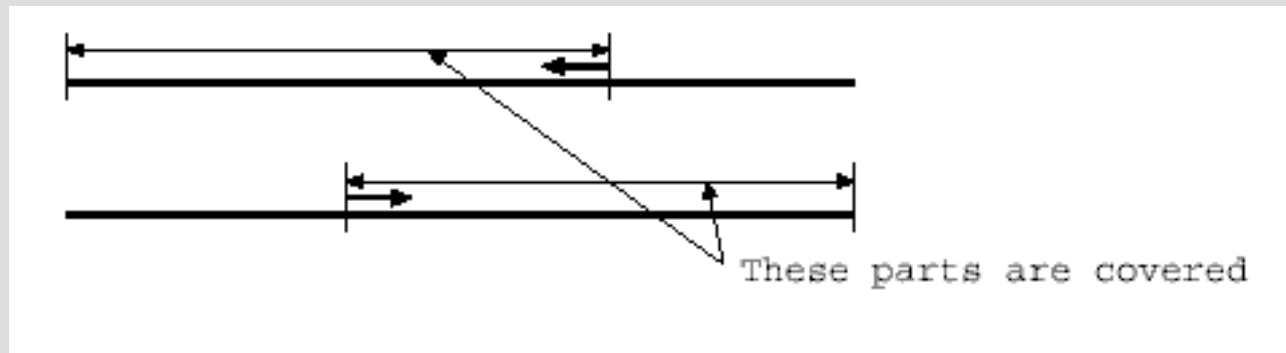
- Ahne algoritm
 - 1. Skaneerida DNA järjestusi ja leida kandidaatpraimerid
 - 2. Valida praimer S_i , mis hõlmab suurima hulga katmata elemente ning eemaldada kandidaatpraimerite hulgast
 - 3. Korrata sammu 2 kuni kõik elemendid on kaetud või rohkem kandidaatpraimereid pole

Praimerite valiku tingimused

- 1. sammul
 - GC sisaldus
 - praimeri pikkus
- 2. sammul
 - paarikaupa/iseendaga komplementaarsus
 - amplifitseeritud segmendi minimaalne pikkus
 - amplifitseeritud segmendi pikkuste erinevus (segmente pikkustega väljaspool 50-500 nukl. ei arvestata)
 - katvus... vt. järgmine slaid

Praimeri katvus

- Modifitseeritud katvus – amplifitseeritava osa minimaalpikkuse k alla jäävat juppi ei arvestata



Algoritmi täiendus

- Eelistatult on juba identifitseeritud DNA järjestuste segmentide arv väike
 - Seega valime praimerid, mis amplifitseerivad eelnevalt identifitseeritud DNA järjestustest kõige vähem
- Identifitseeritud DNA järjestuste puhul ei tule pikki amplifitseeritud segmente eristada
 - Rakendame amplifitseeritud segmentide minimaalset pikkust ainult identifitseerimata DNA järjestustel

Seniste laienduste tulemused

- Mõlemal juhul elementide katvuse parameeter $k=50$, amplifitseeritud segmendi minimaalpikkus $t=50$
 - $n_d(50,500)$ – DNA segmentide arv pikkustega 50-500, mis erinevad teistest amplifitseeritud segmentidest

extension	no	yes
# experiments	42	31
# primers	292	944
average # of primers	7.0	30.5
$n_d(50,500)$	2034	3128

Algoritmi täiendus 2

- Erinevalt ühekordsest PCR-st eelistada praimereid, mis katavad palju DNA järjestusi
- Ahne algoritm
 - 1. Skaneerida DNA järjestusi ja leida kandidaatpraimerid, mis rahuldavad bioloogilisi tingimusi
 - 2. Valida praimer S_i , kui S_i ja juba leitud praimerid amplifitseerivad suurima hulga katmata DNA järjestusi. Võrdsuse korral praimer S_i , mis katab suurima hulga katmata elementidest
 - Korrata sammu 2 kuni kõik elemendid on kaetud või rohkem kandidaatpraimereid pole

Algoritmi täiendus 3

- DNA järjestuste arvu asemel kasutame kaalude summat, kuna pikki DNA järjestusi on lihtsam amplifitseerida
 - Ühe elemendi kaal $1/(\text{elementi sisaldava DNA järjestuse pikkus})$
 - DNA järjestuse kaal $1/(\text{DNA järjestuse pikkus})$

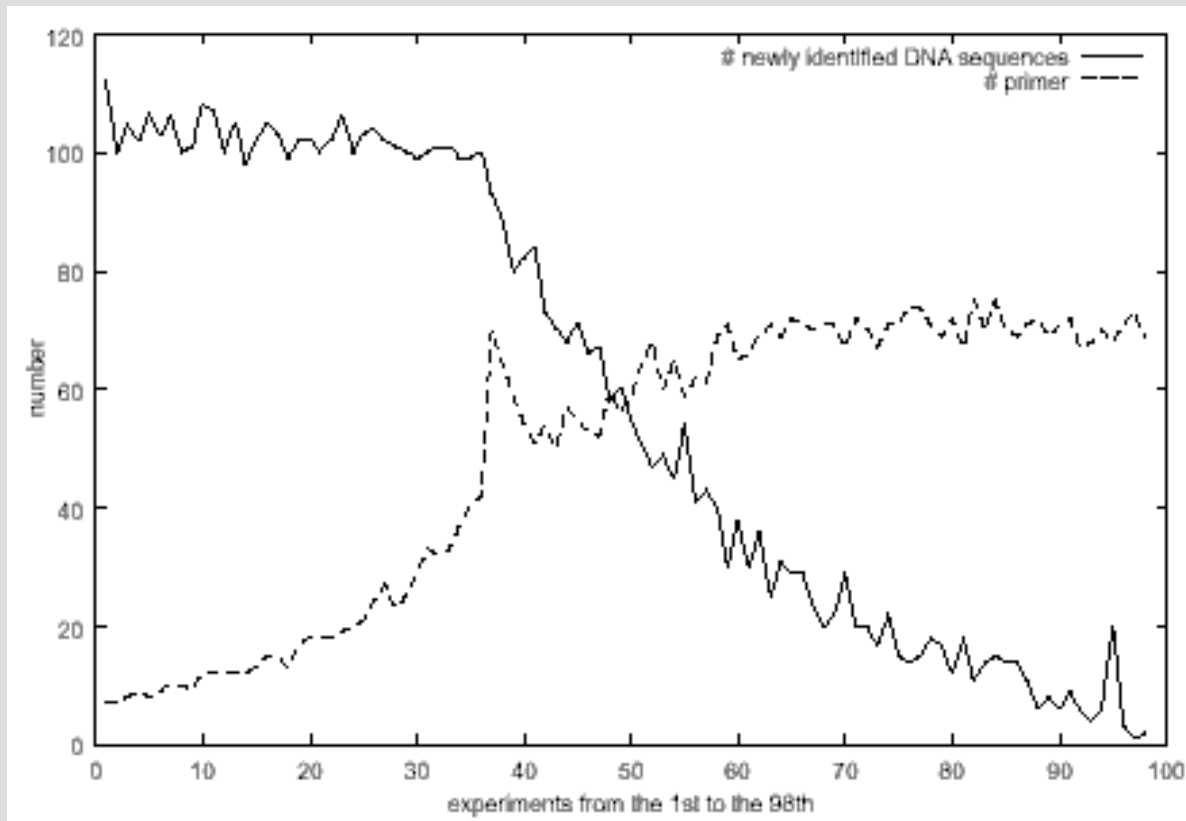
changing the order of priority	no	yes
# experiments	31	98
# primers	944	4853
average # of primers	30.5	49.5
$n_d(50, 500)$	3128	5750

Tulemused 2

- Praimerite arv 85% identifitseeritud DNA järjestuste arvust
 - 90% kõigist sihtjärjestustest
 - 42% praimerite arvust, mida vajab multipleks PCR
 - Praimeri pikkuste lühiduse tõttu hind 20% multipleks PCR omast
- Ligikaudu 470 järjestust eristamatud
 - Multipleks PCR ei suuda amplifitseerida üle 100 DNA järjestuse, ent suudab u. 350 järjestust, mida antud algoritm ei suuda
 - Parema tulemuse võiks saada GC sisalduse ja praimeri pikkuse tingimusi muutes

Tulemused 3

- Pidevjoon – uued identifitseeritud DNA järjestused, punktir – praimerite arv



Tulemused 4

- Algul leitakse head praimerid, lõpus on aga praimerite efektiivsus halvem kui multipleks PCR-I
- Efektiivsuse tõstmiseks eemaldame praimerid, mis ei amplifitseeri äsja identifitseeritud DNA järjestusi
 - Detailid jäävad antud artiklist välja

Kokkuvõte

- vt. slaidi „Tulemused 2“