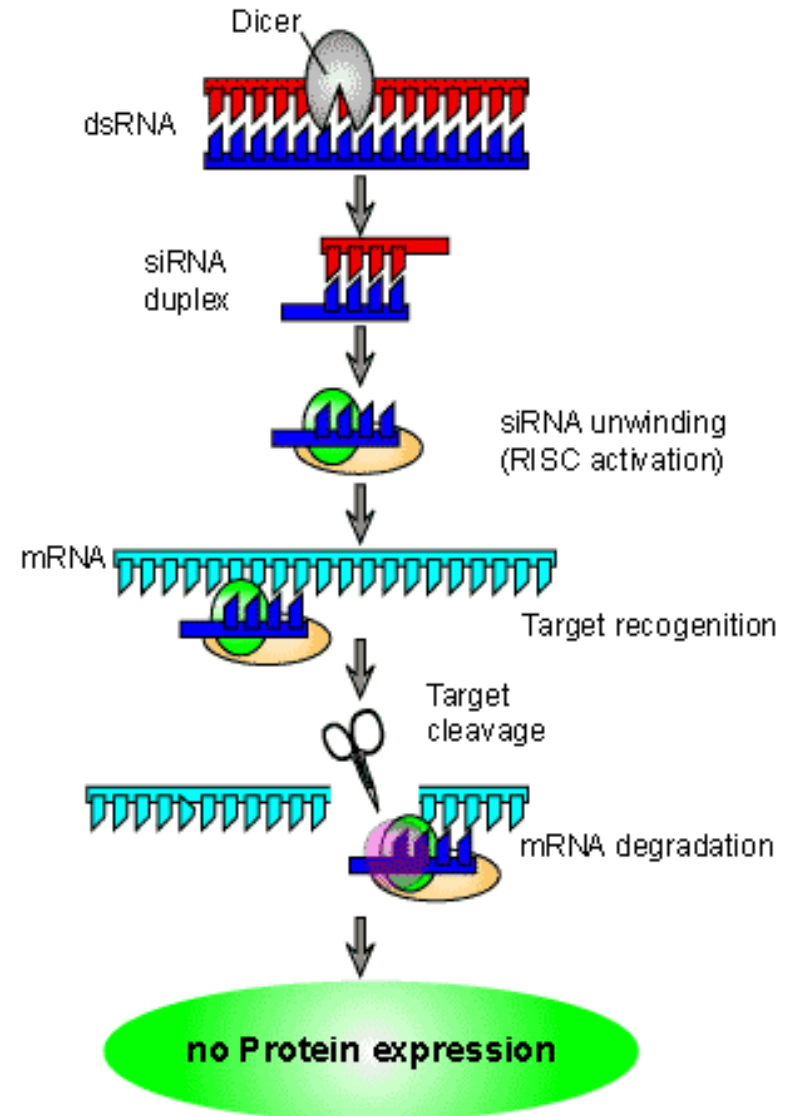


Kõrge efektiivsusega siRNA järjestuste leidmine

Age Tats

RNAi

- Geeni vaigistamise mehhanism mRNA tasemel
- 19-nt pikkused mRNA-ga komplementaarsed siRNA-d katkestavad valgusünteesi
- mRNA degradeeritakse



siRNA (*short interfering RNA*)

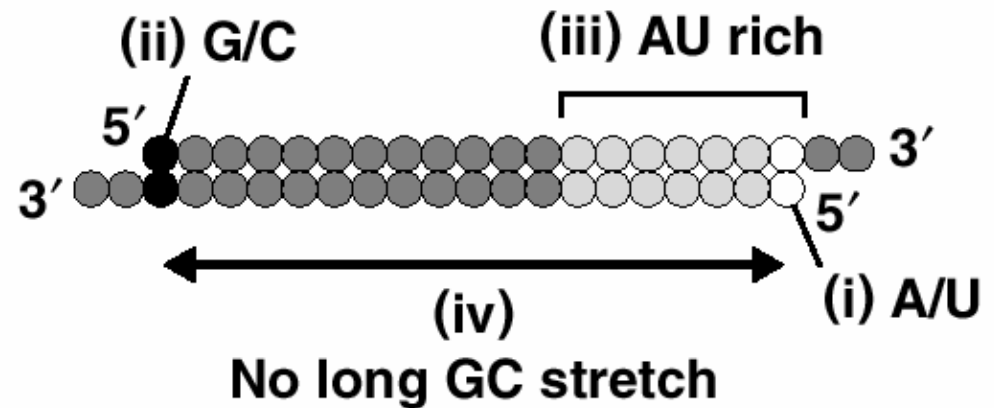
- Lühike, 10-25 nt RNA dupleks (21nt)
- Põhjustab mRNA järjestuse-spetsiifilist degradatsiooni
- 3' otsades üleulatuvad 2 nt-
- Imetajates:

(i) antisense ahela 5' ots A/U

(ii) Sense ahela 5' otsas G/C

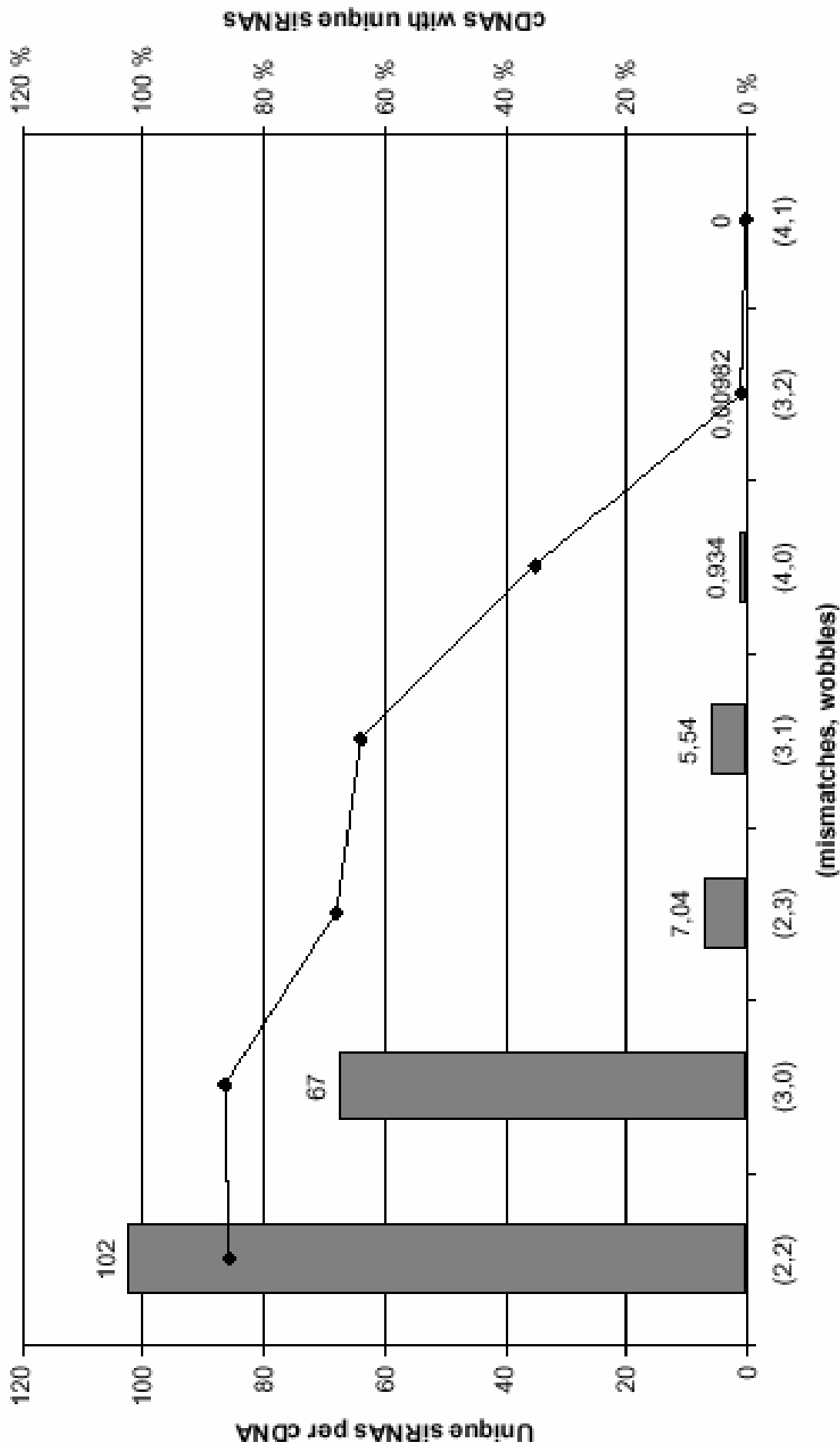
(iii) Vähemalt 5 A/U jääki antisense ahela 5' otsmises kolmandikus

(iv) Rohkem kui 9 nt GC regiooni puudumine



Mismatch tolerance

- Teadlane: Maksimaalne *mismatch*’ide arv, mille juures siRNA ei kaota oma funktsionaalsust on 3
- Bioinformaatik-teadlane: 75% 359-st avaldatud siRNAst omas 3 *mismatch*’i juures off-target riski. Unikaalsemad järjestused **on** olemas ja neid saab otsida ka kiiremate algoritmidega, kui Smith-Waterman



Average number of unique siRNAs per human cDNA transcript
 Percentage of cDNA transcripts with unique siRNA

Off-target risk

siRNA primary target (Acc#/RefSeq)	siRNA name	siRNA off-target hit (Acc#/RefSeq) (mismatches/wobbles)
Lamin A/C (70 kDa lamin) (NM_170707)	Lamin A/C	NDUFS4 (NM_002495) <8 19> 5' -AACUGGAGU <u>U</u> CCAGAAGAGCA-3' : 3' -UUGACCUGAAGGUCUUCUUGU-5'

Miks BLAST ei sobi?

- Kui sõna pikkus on 7, jääb 21 nt-pikkuse ja 3 *mismatch*'iga järjestuste otsimisel leidmata ca 6% off-target järjestustest.
- 19-meeride puhul jääb leidmata 15%.

```
GAAGGCAGTCCAGTGAAAT (NM_000014)
||||| |||| | |||||
GAAGCCAGTACAGAGAAAT (NM_002827)
```

siDirect <http://design.RNAi.jp>

- *Mismatch tolerance* (mt) – minimaalne arv *mismatch*'e siRNA järjestuse ja iga off-target järjestuse vahel

GAAGGCAGTCCAGTGAAAT

| | | X | | | X | | X | | |

GAAGCCAGTACAGAGAAAT

siRNA sequence

off-target sequence

mt

G (input järjestus): ATGCTACGATCGTAGCTC

E (*G* alamstring): ATGC

$$\text{mt}(E, G) = 2$$

Mismatch tolerance (mt) – minimaalne arv *mismatch*'e siRNA järjestuse ja iga mitte-sihtmärk järjestuse vahel.

Longest common factor (lcf)

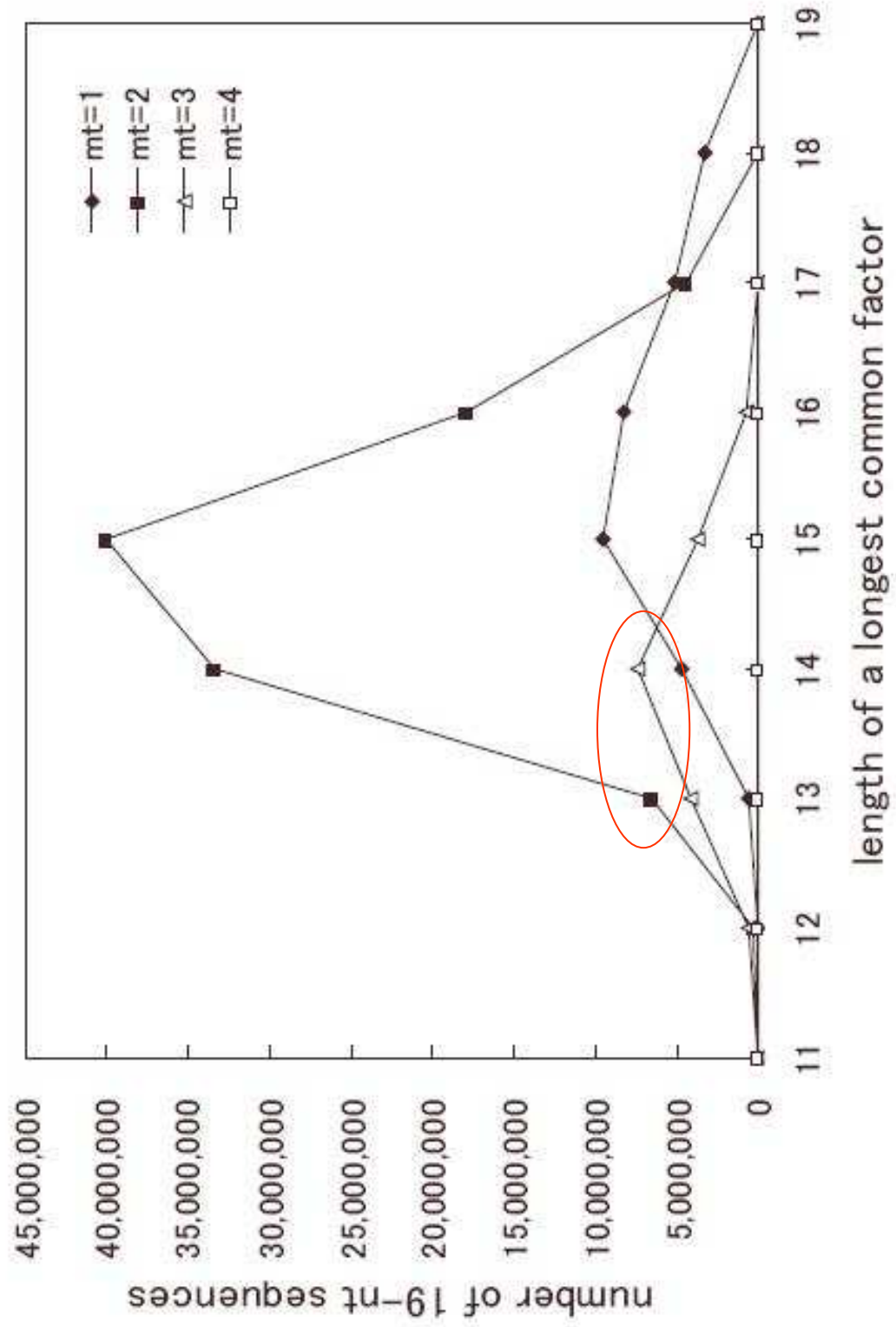
- Off-target riski vähendamiseks tuleb siRNA järjestus E valida nii, et pikim ühine faktor (lcf) E ja iga teise sama pika E -ga mittekattuva stringi vahel oleks väikseim kui 11.

G : ATGCTACGATCGTAGCTC

E : ATGC

$lcf(E, G)$ on AT või GC

$|lcf|=2$



Mt arvutamise algoritmid

- *Seed sequence* ~ *q-gram* – alamstring pikkusega q , mida kasutatakse mustriotsingutel filtrina.

ACCGTAGTA	TATGCTGCTG
X	X X X
AGCGTAGTA	TGTGCTCCAG

siRNA sequence

off-target sequence

Search phase: seed search trials

- Naiivne algoritm ja BYP meetod (Baeza-Yates & Perleberg) nõuavad, et vähemalt üks *seed* järjestus **sobitub täielikult** off-target järjestusega (maksimaalne pikkus 3)
- siDirect algoritm : vähemalt üks *seed* järjestus pikkusega l peab sobituma off-target järjestusega **maksimaalselt $s(\leq k)$ *mismatch*'i abil.** → st luuakse kõikvõimalikud stringid, kus $i(\leq s)$ tähte l positsiooni hulgas asendatakse ülejäänud kolme tähega.

s	number of seed search trials	expected number of seed hits	seed length
1	57	2,087,402	6
2	704	402,832	9
3	5,240	2,998,352	9

$$|G| = 1.5 \cdot 10^8$$

$$|E| = 19$$

$$k = 4$$

$$\frac{G}{4^l} \cdot \left[\frac{k+1}{s+1} \right] \cdot \sum_{i=0}^s 3^i \cdot \binom{l}{i}$$

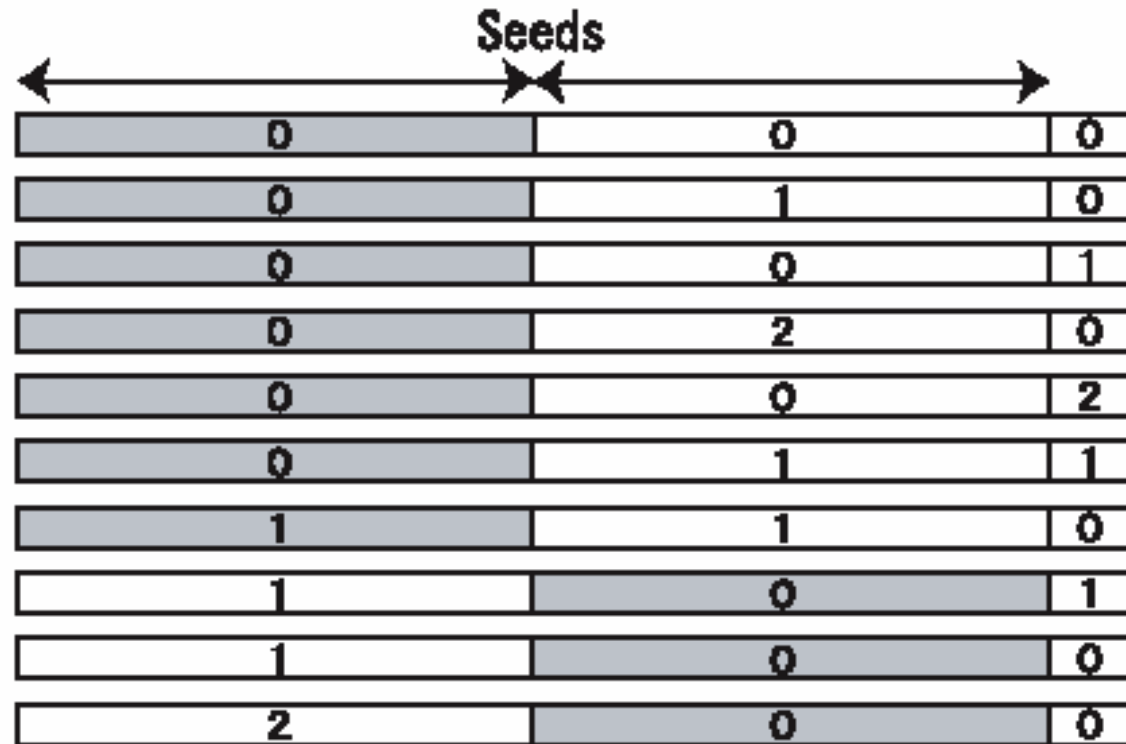
Seed'ide arv
järjestuses E
(19meeris)
Kõikvõimalike
lisa-seed
järjestuste arv

Otsingukatsete arv

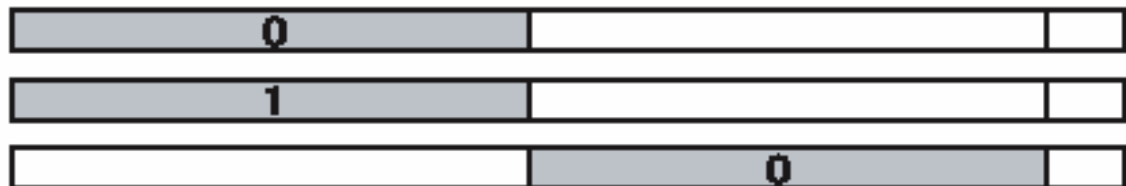
$|E|=19, k=2, s=1$

$l = 9$

Katseid 56



Katseid 29



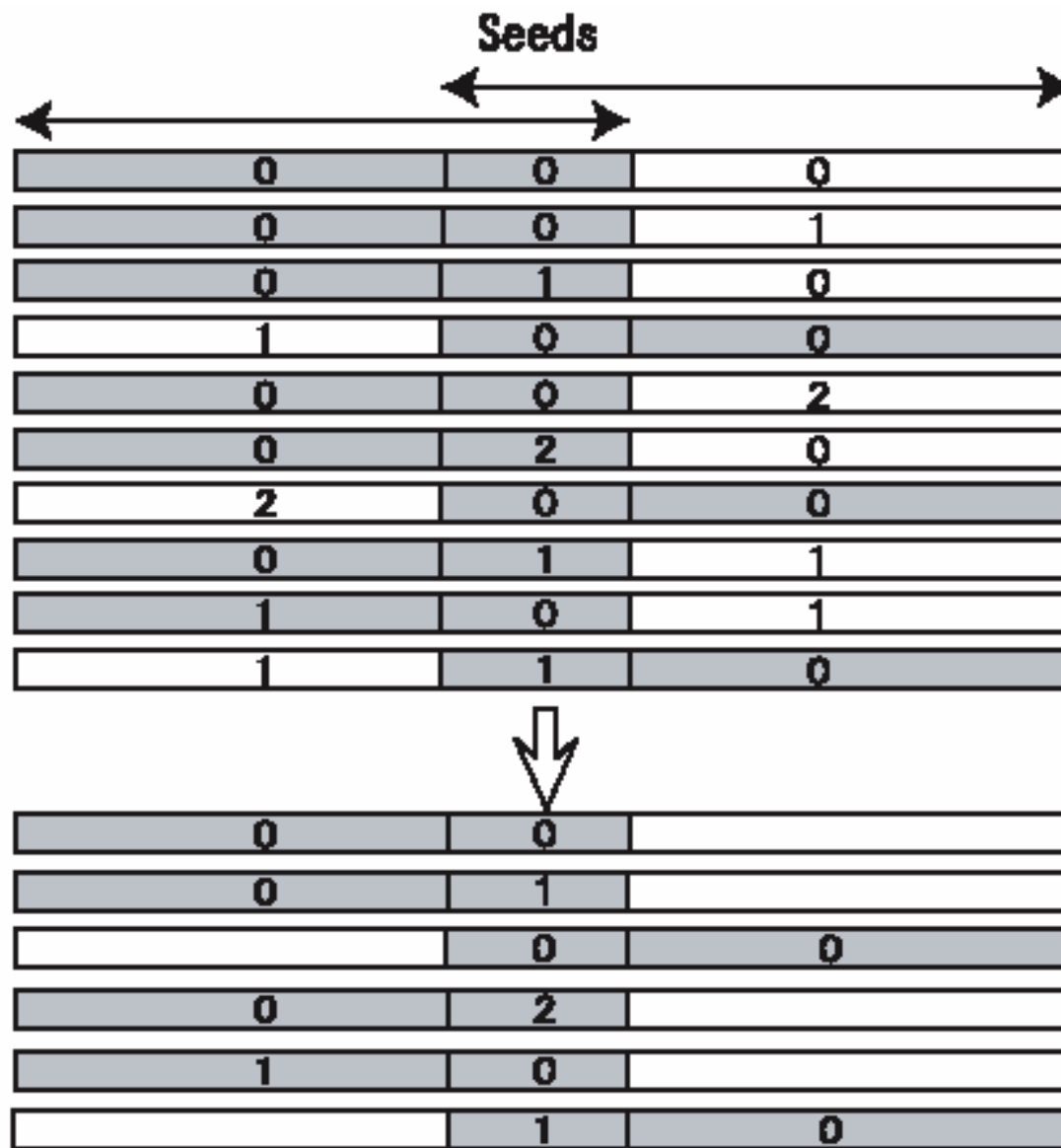
Kattuvad *seed* järjestused

- 19 nt järjestuse puhul saab kasutada ainult ühte mittekatuvat *seed* järjestust, mille pikkus on >9 .
- Kattuvate järjestuste kasutamine vähendab sooritusaega

Järjestuse

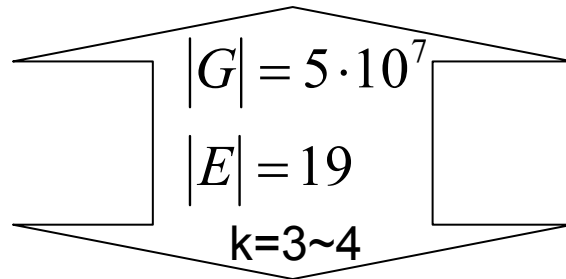
pikkus

ületab $\lfloor |E|/2 \rfloor$



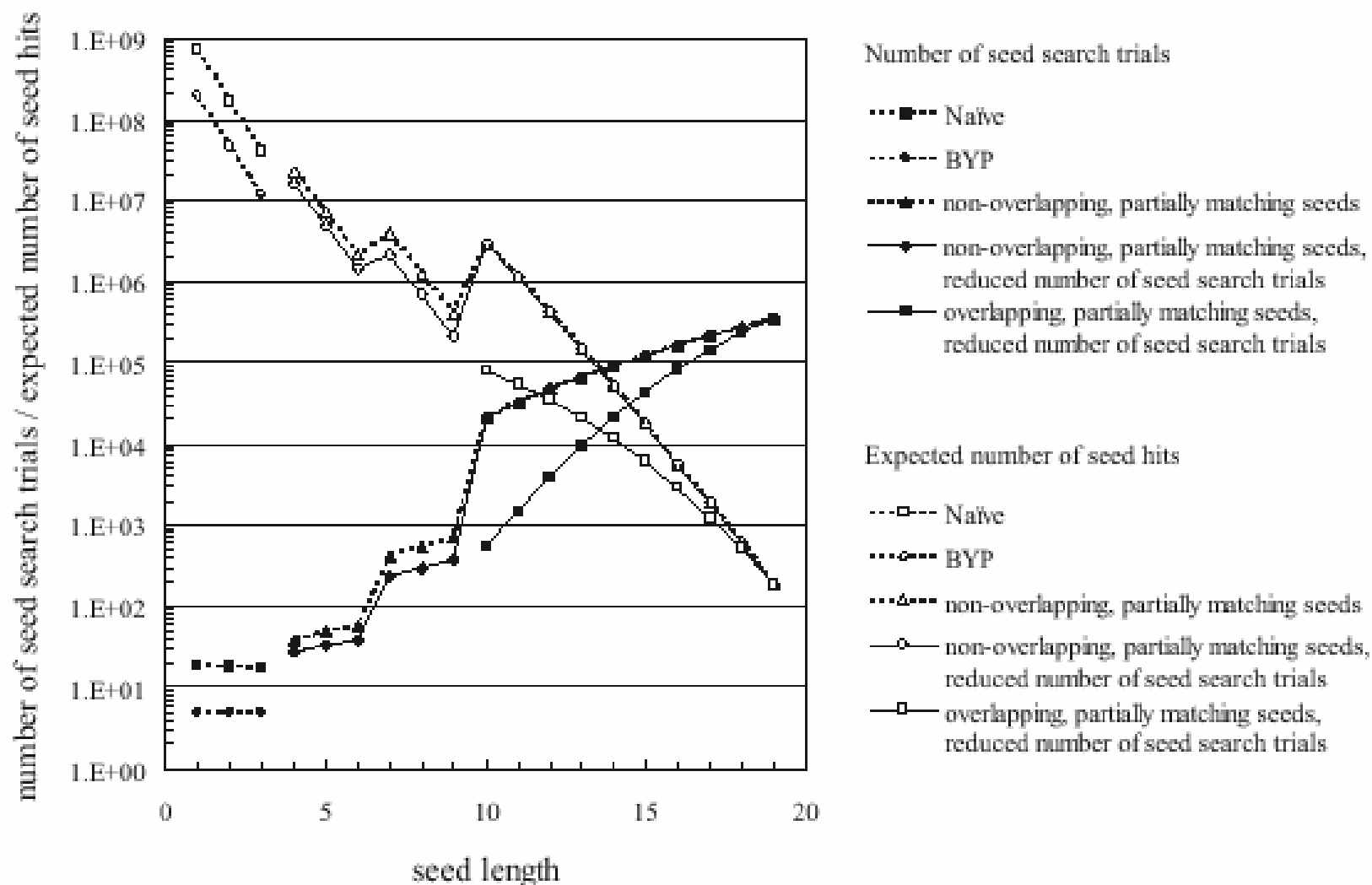
Algoritmide võrdlus

- Naiivne algoritm
- BYP meetod
- Mittekattuvad, osaliselt sobituvad *seed* järjestused



- Mittekattuvad, osaliselt sobituvad *seed* järjestused vähendatud otsingukatsete arvuga
- Kattuvad osaliselt sobituvad *seed* järjestused otsingukatsete arvuga

Algoritmidde vörðlus



Järeldus

- Parimad on osaliselt sobituvad 13 või 14 nt pikkused seed järjestused.
- Kattuvad *seed* järjestused annavad alati parema tulemuse, kui mittekattuvad

