

Анализа појединачних ДНК секвенци

Циљеви часа

- Научите како да манипулишете ДНК секвенце на компјутеру, нађете рестрикциону мапу, и умножите ваше секвенце са PCR-ом
- Користите методе за предикцију гена, разумете њихов потенцијал и ограничења

План

1. Чишћење ДНК секвенце од контаминације
2. Налажење рестрикционе мапе на компјутеру
3. Одређивање ДНК прајмера
4. Налажење секвенци у ДНК које кодирају протеине

Клонирање ДНК секвенци

- Да би могли да секвенцујете геноме, ДНК секвенце се често клонирају у вектор (нпр плазмид)
- Може да се деси да се секвенце вектора помешају са вашом ДНК секвенцом.
- Пре него што радите са вашом ДНК секвенцом, увек треба да је очистите са VecScreen – ом.

Paste in your new sequence.

NCBI

PubMed Entrez BLAST OMIM Taxonomy Structure

NCBI Home Page

Contamination
Definition
Sources
Consequences
Detection

VecScreen
Overview
Example
Search Parameters
Match Categories
Interpretation

Screen A Sequence Using VecScreen

Enter your query sequence below as an Accession, GI or [FASTA](#).

```
>gi|4099437|gb|U67251.1|NVU67251 Mustela vison GT  
dinucleotide repeat, chromosome 7q2.1  
AAACANACDBCGBTCNGAAVACNATNGBATHTHAACAVTDPATCNAAAACNGTBCTABATCN  
CGAAGCCNGGGGATCCTTTTAAATTTAAAAATGAAGTTTTAAATVAAATNNAAGDATANAGGA  
GTCTGANAGTNACVAAATGGGTAATCAGTAAGGNACCNATNENAGGATCNGTCTTTGAAGN  
TAAATNTGGTATGNATAAACATGTATATATACATATOGGGTATGBATADGDATGGGTATATA
```

Run VecScreen Clear Input

Пример

- Испитајте да ли је ген под GenBank Accession number U87251 контаминисан са неким од вектора
- Која је препорука да се разреши ова контаминација?

Одређивање рестрикционе мапе

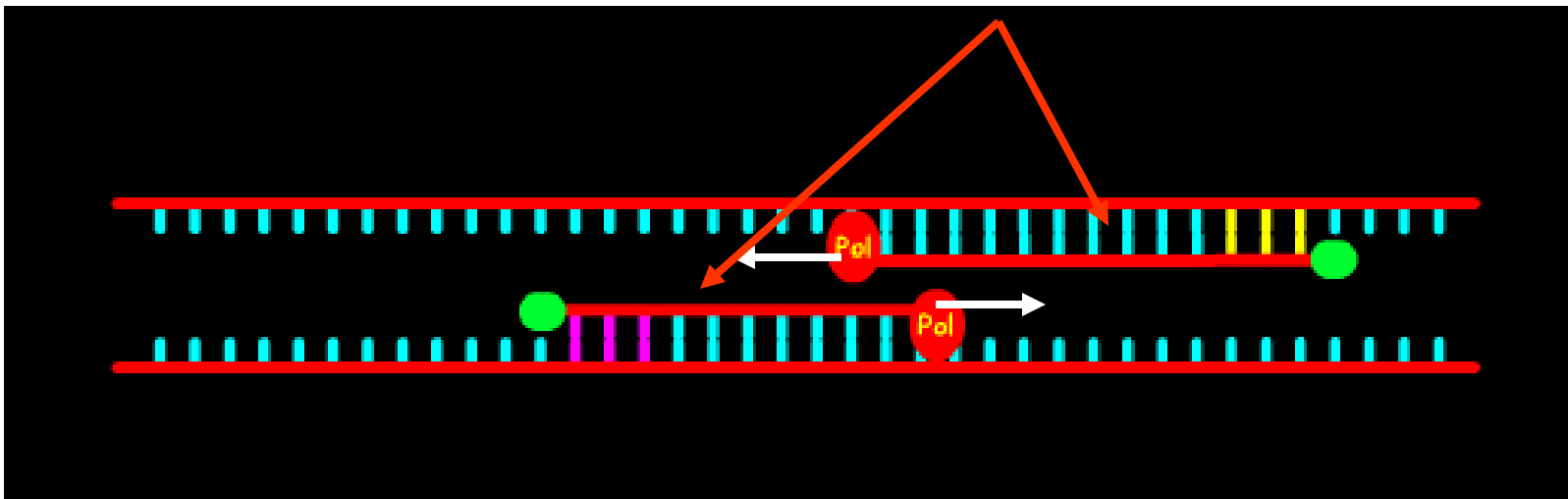
- Могуће је исећи ДНК секвенце користећи рестрикционе ензиме
- Сваки тип рестрикционог ензима препознаје и сече одговарајућу ДНК секвенцу:
 - EcoR1: GAATTC
 - BamH1: GGATCC
- Постоји више од 900 различитих рестрикционих ензима, од којих сваки сече одговарајућу ДНК секвенцу.
- Рестрикциона мапа наводи сва потенцијална места за сечење на ДНК молекулу
- Можете направити рестрикциону мапу са www.firtsmarket.com/cutter

Пример

- Преко Entrez Gene лоцирајте ген за T7 bacteriophage DNA polymerase
- Запамтите секвенцу овог гена као FASTA фајл
- Upload – ујте секвенцу овог гена на Webcutter и помоћу њега нађите рестрикциону мапу за ову секвенцу

Дизајнирање PCR прајмера

- Polymerase Chain Reaction (PCR) је метод за умножавање ДНК
- PCR се користи за многе примене, укључујући
 - Клонирање гена
 - Форенсичка анализа
 - Тест очинства
- PCR умножава ДНК између два “сидра”
- Ова сидра се зову **PCR primer-и**



Дизајнирање PCR прајмера

- PCR прајмери су типично дуги 20 нуклеотида
- Прајмери морају добро да се хибридизују са ДНК
- На biotools.umassmed.edu, можете да нађете најбољу локацију за прајмере:
 - Најстабилнији
 - Најдужа екстензија

Paste your sequence here.

Primer3: WWW primer tool
pick primers from a DNA sequence

[disclaimer](#) [cautions](#) [bugs? suggestions?](#) [Questions?](#)

Paste source sequence below (5'→3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINEs, etc.) or use a [Miscellaneous Library \(repeat library\)](#). [NONE]

```
TGTATTCTTTTATCATGATGAAAGGAAAAGTCCACTCTCCGGCAAGCTTTTATATTCCTGCTCGTGTATT  
CCATATAGAGGGGCTTGGCTTGGATTGAAATTCGACCGCTAAGGATGTGATTTTATTCAGGATTTGATAGAA  
AAAGAAAGCTTTATACCTACTACTTGGCTTAGAGCTATAGGTAAGATCTGAGAGAAATTTATATAATTTTA  
TTATAATTTCAGTAACTTATAAGCTTGTAAATAAAGGTTGGCGGTTAAATTTTATACCTCAGCATATT  
ACTGCTCATCCGTTTAAACAAGTGTATTTAGTAGATGCAGATACCCGGAATATTCTACTGAAAGCCAGGCAAA  
AAATTAAGCTCCGCTTGGCTTAAAAAATATCTCGGGGAAAGGCTTAAATAATATTTTAGTAGCTCATGAAAC  
TTTAAATCGCCAAATATCTATCCGAAGATTAAAGAGATCCTGCAAGCCGATGAAATTTAOC AAAAATCGGT
```

Pick left primer or use left primer below. Pick hybridization probe (internal oligo) or use oligo below. Pick right primer or use right primer below (5'→3' on opposite strand).

Pick Primers Reset Form

Sequence Id: A string to identify your output

Targets: E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or mark the [source sequence](#) with [and]:
e.g. ...ATCT[CCCC]TCAT.. means that primers must flank the central CCCC.

Excluded Regions: E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the [source sequence](#) with < and >: e.g. ...ATCT<CCCC>TCAT.. forbids primers in the central CCCC.

Product Size Min: 100 Opt: 200 Max: 1000

Number To Return: 5 Max 3' Stability: 9.0

Max Mispriming: 12.00 Pair Max Mispriming: 24.00

Pick Primers Reset Form

Click here to get your primers.

Пример

- Дизајнирајте PCR прајмере за амплификацију T7 DNA polymerase гена.

Манипулација ДНК секвенце

- Састав ДНК је доста промењив
- Стабилност ДНК секвенци зависи од њеног **G+C састава** (укупан гуанин и цитозин)
- Висок G+C састав чини веома стабилне ДНК молекуле
- На интернету имате програме који вам омогућавају да одредите нуклеотидни састав ваше ДНК секвенце, да одредите комплементарни ланац, да екстракујете ДНК секвенцу са одређеним координатама
- www.genomatrix.de/cgi-bin/tools/tools.pl

Пример

- За секвенцу T7 ДНК полимеразе:
 - a) Одреди GC састав
 - b) Нађи инверзни комплемент овој секвенци.
Запамти као FASTA фајл.
 - c) Издвоји ДНК секвенцу дужине 500
полазећи од нуклеотида са бројем 300.
Запамти као FASTA фајл

Пример

- Одредите GC састав ДНК секвенце гена T7 DNA polymerase

Предвиђање гена

- Вероватно најважнија анализа која се може извршити на ДНК секвенцама је предвиђање гена
- Предвиђање гене захтева различите методе за еукариоте и прокариоте
- Предикција гена код еукариота је компликована постојањем интрона

Предвиђање гена у прокариотским геномима

- У прокариотима, гени који кодирају протеине су непрекидни
 - Нема интрона
- Предвиђање гена који кодирају протеине у прокариотима се сматра решеним проблемом
 - Можете да очекујете тачност од 99%

Налажење гена у прокариотима са GeneMark - ом

- GeneMark је најпознатији програм за откривање гена у микробима
- GeneMark може
 - Да нађе кратке протеине
 - Детектује и гене који се преклапају
 - Идентификује најбољи старт кодон
- GeneMark користи “hidden Markov Models”
- Користите exon.gatech.edu/GeneMark

Gene Prediction Results

Information on input sequence

Sequence title: Wed Sep 25 13:56:18 EDT 2002
Length: 5040 bp
G+C percentage: 31.41 %

Parse predicted by GeneMark.hmm 2.0

GeneMark.hmm PROKARYOTIC (Version 2.1)
Sequence file name: sequence, RBS: N
Model file name: heuristic_no_rbs.mat
Model organism: Heuristic_model
Wed Sep 25 13:56:28 2002

Predicted genes

Gene #	Strand	LeftEnd	RightEnd	Gene Length	Class
1	+	1	822	822	1
2	+	1039	1356	318	1
3	+	1367	2116	750	1
4	+	2117	2893	777	1
5	+	2890	>5040	2151	1

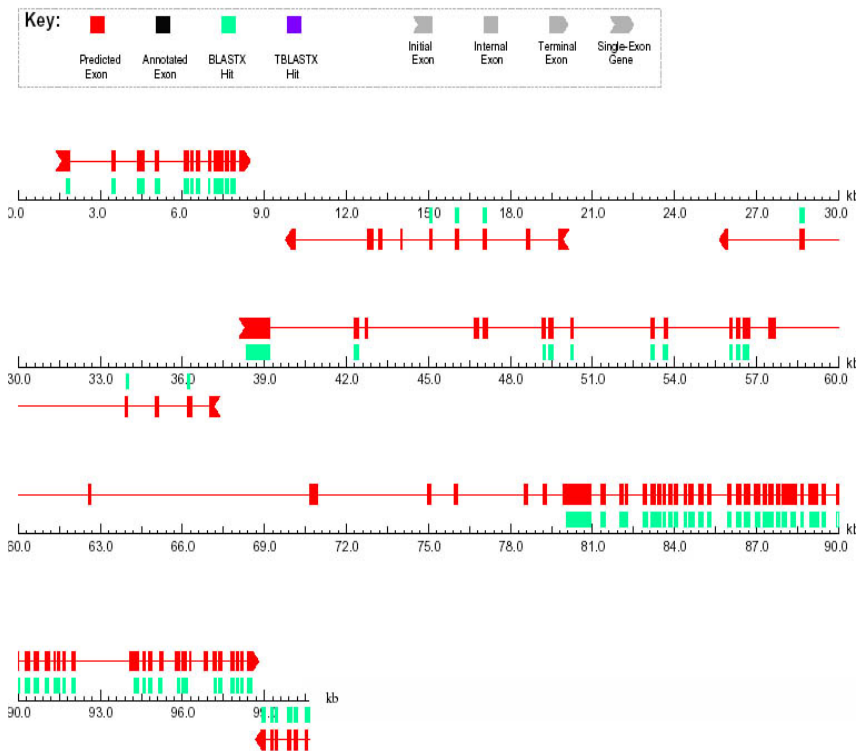
Предвиђање гена у еукариотима

- Веома је тешко предвидети гене у еукариотима (нпр. човеку)
- Тачно предвиђање гена у еукариотима је још увек отворен проблем
 - ENSEMBL садржи 21,662 гена за геном човека
 - Врло је вероватно да постоји више гена од тога, који су пропуштени у предвиђањима
- Са аутоматским методима може се очекивати тачност од 70% на геному човека
- Експериментална информација је и даље потребна да би се тачно предвидели гени у еукариотима.

Налажење гена у еукариотима са GenomeScan - ом

- GenomeScan је најпознатији програм за одређивање гена у еукариотима
- GenomeScan најбоље ради са
 - Дугачким егзонима
 - Генима са ниским GC садржајем
- GenomeScan користи
 - Hidden Markov Models
 - Претраге помоћу хомологије
- Може да се укључи и експериментална информација
- Користите genes.mit.edu/genomescan

GenomeScan predicted genes in sequence gb



Задаци за вежбу

- 1) Odredite помоћу VecScreen да ли је секвенцовани ген бактериофага T7 контаминиран са вектором. У случају да сте за клонирање користили вектор pBR322, одговорите на следећа питања:
 - a) Са којим вектором је секвенца контаминирана
 - b) Дајте препоруку да се разреши ова контаминација
 - c) Одговорите на исто питање у случају да је за клонирање коришћен вектор pUC19

- 2) Преко Entrez Gene лоцирајте ген за T7 bacteriophage RNA polymerase. Запамтите секвенцу овог гена као FASTA фајл и помоћу Webcutter нађите рестрикциону мапу за ову секвенцу. Изаберите два ензима који ће да исеку ДНК секвенцу на почетку и на крају.

- 3) За претходну секвенцу дизајнирајте PCR прајмере

- 4) За ДНК секвенцу из задатка 2):
 - a) одредите њен GC састав
 - b) нађите инверзни комплемент ове секвенце
 - c) издвојте секвенцу дужине 500, полазећи од нуклеотида број 200

- 5) Користите GenMark да предвидите гене у геному бактериофага Хр10