Analysis of a single DNA sequence

Sequence contamination

vector sequence: UniVec
(www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html)

verifying a restriction map:

REBASE database (rebase.neb.com)
Webcutter: www.firstmarket.com/cutter/cut2.html
commercial site: http://tools.neb.com/NEBcutter2/index.php

Designing PCR primers

biotools.umassmed.edu

Analyzing DNA composition (G+C content)

http://sites.univ-provence.fr/~wabim/english/logligne.html

http://bioweb2.pasteur.fr/intro-en.html

http://www.basic.northwestern.edu/biotools/oligocalc.html

Counting words in DNA sequences

2- or 3-letter words

www.genomatix.de/cgi-bin/tools/tools.pl

Counting long words in DNA sequences

regulatory sequence motifs (for n-letters, 2^{2n} different words)

DNA sequence analysis/codon usage, composition/wordcount

http://www.bioinformatics.org/sms2/dna_stats.html

Finding internal repeats

tandem repeats, inverted repeats
finding repeats is a tricky business

Dot-plot approach

Molecular Toolkit (http://arbl.cvmbs.colostate.edu/molkit)
click Dot Plots
click Make Plots

how to identify inverted repeats (reverse complement)

How to assess the significance of repeats

9 ATGC repeats in 3000-bp DNA sequence
the random probability of observing ATGC: 1/256
the expected number of ATGC in 3000-bp: 3000/256=11.7

Finding protein coding regions

ORF(open reading frame) in microbial DNA sequences
or eukaryotic mRNA sequences
Start codon (ATG)
Stop codon (TAA, TAG, TGA)

ORF finder: www.ncbi.nlm.nih.gov/gorf/gorf.html
a more sophisticated: GeneMark (opal.biology.gatech.edu/GeneMark/)

finding internal coding exons
MZEF at Cold Spring Harbor (argon.cshl.org/genefinder)
GenomeScan at MIT (genes.mit.edu/genomescan)

Gene expression analysis

Expression profile

NCBI (http://www.ncbi.nlm.nih.gov/)

NM_007409

click GEO profile in Link at the right corner

Transcription factor binding motifs

www.genome.ad.jp, Motif

paste and search for DNA

Assignments

Analyze your DNA sequence

- DNA composition (G+C content)
- internal repeats: Dot-plot approach and assessment
- codon usage

ORF find using cDNA

Internal coding exon find using genomic DNA

Expression profile

Search for promoter sequence: ~500 bp of the start codon