

War Stories

Synonymous & Non-Synonymous substitutions

Estimate the number of syn (K_s) and non-syn (K_a) substitutions in order to identify if regions are under selection.

TIATPDDID--RLLSSNFLISQIVDLLAALITLISTFSAINLILTKSIFQWSTKILLFQN
TIACASIEQQRLRSSNFVIAQYIDLLCIVITFVTTYPAIQLVLNKSLFQWSTKMLILES
*** . . * : ** ***** : * : * : ***** . : * * : : : * : . * * : * : * . * * : * * * * * * : * : : : .

ACCATCGCCACCCCGGACGACATTGAT-----CGATTGCTCAGCTCAAATTTTCTAATC
ACCATAGCTTGTGCATCAATCATAGAACAGCAAAGGTTGCGTAGTTCAAATTTTGTAATT
***** ** * ***** ** * * * * * * * * * * * * * * * *

TCCCAAATCGTTGATCTCCTCGCCGCGCTGATAACCCTTATCTCTACATTCTCAGCGATA
GCTCAATACATTGATCTTCTTTGTATTGTTATCACTTTTGTGACCACCTATCCAGCAATT
* *

First do the alignment in protein space

Do alignment:

```
%ssearch34 gene1.fa gene2.fa >  
alignment.SSEARCH
```

Project the alignment back into DNA

```
use Bio::SeqIO;
use Bio::AlignIO;
use Bio::Align::Utilities qw(aa_to_dna_aln);
my $parser = new Bio::SearchIO(-file => 'rpt.fasta',
                               -format => 'fasta');

my $result = $parser->next_result;
my $hit     = $result->next_hit;
my $hsp    = $hit->next_hsp;
my $aln    = $hsp->get_aln();
my $seqparser = new Bio::SeqIO(-file => 'cds.fasta');
my %seqs;
while( my $seq = $seqparser->next_seq ) {
    $seqs{$seq->display_id} = $seq;
}
my $dnaaln = &aa_to_dna_aln($aln, \%seqs);
```

Calculate Ka/Ks

```
use Bio::Tools::Run::Phylo::PAML::Codeml;

my $codeml = new Bio::Tools::Run::Phylo::PAML::Codeml();
$codeml->alignment($dnaaln);
my ($rc,$parser) = $codeml->run();
my $result = $parser->next_result;
my $MLmatrix = $result->get_MLmatrix();
print "Ka = ", $MLmatrix->[0]->[1]->{'dN'}, "\n";
print "Ks = ", $MLmatrix->[0]->[1]->{'dS'}, "\n";
print "Ka/Ks = ", $MLmatrix->[0]->[1]->{'omega'}, "\n";
```

Designing primers for a gene family

- Gene family with a conserved protein domain, want to clone it from other species
- Use HMMer to identify the domain location in each gene
- Extract the domain, get the corresponding DNA
- Manually inspect alignment for conserved sites for primer choices.

Run HMMer

```
hmmsearch domain.hmm proteins.pep >  
proteins_domain.HMMER
```

Extracting the regions

- use `Bio::SearchIO` to parse the report
- Get the start/end of the hit
- use `Bio::Index::Fasta` to index sequence database
- Get sequences and sub sequences