Rnaspl

Program for predicting exon-exon junction positions in cDNA sequences.

Recognition of exon-exon junctions in cDNA may be very useful for gene sequencing when starting with a sequence of cDNA clone. In a given cDNA sequence we need to select sites for PCR primers that (hopefully) lie in adjacent exons. Prediction is performed by linear discriminant function combining characteristics describing tipical sequences around exon-exon junctions.

Accuracy:

We can not predict exon-exon junction position with very high accuracy, because some important information is being lost during splicing. We predict positions marked by '*', where 75% of potential exon-exon junctions are localized. Additionally, we mark '-' positions where exon-exon junctions atr absent with probability about 90%. We recommend to select primer sequences in continuous '-' regions that do not cross '*' or ' ' positions.

Reference:

Solovyev V.V.,Salamov A.A., Lawrence C.B. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. (Nucl.Acids Res., 1994, 22, 24, 5156-5163).

RNASPL output:

```
First line - name of your sequence
Second line - your sequence
3d line - '*' shows potential exon-exon junction position (Pr > 0.75) '-'
shows position where exon-exon junction absent (Pr > 0.90) 'n' is nonanalyzed
flanking position
For example:
   HSACHG7       690 bp    DNA             PRI       18-DEC-1990
        10        20        30        40        50        60
ATGGCGGCGACGGCGAGTGCCGGGGCCGGCGGGATGGACGGGAAGCCCCGTACCTCCCCT
nnnnnnnnnnnnnnnnnnn--------  ---------*---- ----*----------
        70        80        90       100       110       120
AAGTCCGTCAAGTTCCTGTTTGGGGGCCTGGCCGGGATGGGAGCTACAGTTTTTGTCCAG
----- *----*--------- -- --------*-------  --------------- -
       130       140       150       160       170       180
CCCCTGGACCTGGTGAAGAACCGGATGCAGTTGAGCGGGGAAGGGGCCAAGACTCGAGAG
-----------*-*--- ---- ------ --*----- -----------*------ --
       190       200       210       220       230       240
TACAAAACCAGCTTCCATGCCCTCACCAGTATCCTGAAGGCAGAAGGCCTGAGGGGCATT
------ ---------- ----------------  ------------------------
       250       260       270       280       290       300
TACACTGGGCTGTCGGCTGGCCTGCTGCGTCAGGCCACCTACACCACTACCCGCCTTGGC
----- -- ---------------------------------------------
```