

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

## Gene recognition via spliced alignment

ArticleInfo		
ArticleID	:	3607
ArticleDOI	:	10.1186/gb-2000-1-1-reports233
ArticleCitationID	:	reports233
ArticleSequenceNumber	:	98
ArticleCategory	:	Web report
ArticleFirstPage	:	1
ArticleLastPage	:	4
ArticleHistory	:	RegistrationDate : 2000-1-10 Received : 2000-1-10 OnlineDate : 2000-3-17
ArticleCopyright	:	BioMed Central Ltd2000
ArticleGrants	:	
ArticleContext	:	130591111

Todd Richmond

## Abstract

The PROCUSTES server provides a method for determining protein-coding sequences in genomic DNA.

## Content

The PROCUSTES server provides a method for determining protein-coding sequences in genomic DNA. The main difference between PROCUSTES and other gene-finding programs is that PROCUSTES allows the user to supply a related protein sequence, which the program then uses to define the best multi-exon structure for the predicted protein. The resulting prediction is often much better than that produced by other programs, especially for genes with many introns.

## Navigation

It is somewhat difficult to find the basic page for submitting sequences ([Gene recognition via spliced alignment](#)). The main page contains reference information and a simple explanation of how the program works. Once you locate the basic submission page, however, you can bookmark it separately. You can submit a genomic sequence up to 180,000 base pairs long and a maximum of 10 related protein sequences. There are only a few options to worry about. You can choose some parameters that the program uses for aligning the related proteins with the predicted protein, and select the minimum intron size you expect. You can also choose to specify whether or not you believe that the sequence being analysed contains a full gene, or one that is incomplete at either the 5' or 3' end. You can also specify the organism, though the choices are currently limited to human and mammalian, *Drosophila*, monocot plants, dicot plants or yeast. The site warns, however, that only the parameters for human and mammalian sequence have been extensively tested and optimized.

## Reporter's comments

# Timeliness

Last updated 2 January 1997.

## Best feature

The ability to use a related sequence to determine the gene structure for an unknown gene is a powerful tool. Even distantly related proteins can be extremely useful in predicting exons in unknown sequence. The program outputs a combined graphic showing the predicted gene structures from all related proteins submitted, as well as a separate table of exons, sequence alignments, and predicted protein sequence for each related sequence, with a confidence score for each related sequence.

## Worst feature

PROCRUSTES uses a very strict definition for splice sites, which can cause problems. The set of candidate exons is constructed by selection of all blocks between candidate acceptor and donor sites (that is, between an AG dinucleotide at an intron-exon boundary and a GU dinucleotide at an exon-intron boundary). As a result, if there are any deviations from this, the program will either fail to find the correct exons, or define exons of the wrong length. As slight deviations are fairly common, this is a major drawback.

## Wish list

Allow the user to submit up to ten related sequences in a single FASTA-formatted file. Currently, each related sequence has to be cut and pasted into the web form separately. Allow the integration of organism-specific splice-site prediction programs (like [NetGene2](#)) to increase the accuracy of the program. Fully optimize the parameters for filtering exons for organisms other than mammals. Allow the integration of partial cDNA sequence information when this data is available.

## Related websites

There are a number of gene prediction websites, including [GENSCAN](#), [Grail](#), [GeneMark](#) and [Genie](#).

# Table of links

[PROCRUSTES](#)

[Gene recognition via spliced alignment](#)

[NetGene2](#)

[GENSCAN](#)

[Grail](#)

[GeneMark](#)

[Genie](#)

## References

1. [PROCRUSTES](#).