Scipio Documentation

Contents

- General
- Options
- Examples
- Activity diagrams

General

WebScipio offers an advanced option to map cDNA data (EST or mRNA) on the exon-intron gene structure reconstructed by Scipio. Based on this mapping alternative splice forms are predicted.

Options

Data

Different cDNA data files can be choosen for different species. An autocompletion field offers the possibility to search for a species. Then an EST or mRNA data file needs to be selected.

Sequence type

In the first step of the cDNA alignment the cDNA data is either aligned to the coding DNA of the gene, the translation of the coding DNA or the whole genomic DNA (including introns). In the second step it is always aligned to the genomic DNA.

BLAT Tile Size

Sets the size of match that triggers an alignment. Usually between 8 and 12. Default is 11 for DNA and 5 for protein.

BLAT Step Size

Spacing between tiles. Default is Tile Size.

BLAT Min. Identity

Sets minimum sequence identity (in percent) between cDNA sequence and gene sequence.

BLAT Min. Score

Sets minimum score. This is the matches minus the mismatches minus some sort of gap penalty. If the the min. score is set to a high value, only long cDNA hits are taken into account.

BLAT Min. Match

Sets the number of tile matches. Usually set from 2 to 4. Default is 2 for nucleotide, 1 for protein.

BLAT Max. Gap

Sets the size of maximum gap between tiles in a clump. Usually set from 0 to 3. Default is 2. Only relevent for minMatch > 1.

BLAT One Off

If set to 1 this allows one mismatch in tile and still triggers an alignments.

Examples

Drosophila melanogaster Myosin 7B gene with alternatively spliced exons

Organism	Drosophila melanogaster
Genome file	chromosome v 5.0.0
YAML-file for upload	ScipioResult_Dm_DSCAM.yaml

Query sequence	MSEFVRQHGEYVWVKPQNTTSEFAVPFGARIVRTEKTQTLVCDDRNKQFWVPAGDVLKAM HITSQEDVEDMITLGDLQEYTILRNLQNRYAKQLIYTYTGSMLVAINPYQILPIYTNREI QLYRNKSLAELPPHIFAISDNAFQRLQRLKENQCVVISGESGAGKTESTKLILQYLAAIS GKHSWIEQQIIEANPIMEAFGNAKTVRNDNSSRFGKYIEIRFTPQGAIQGARIQQYLLEK SRIVFQSRDERNYHIFYCMLAGLSTAERERLKLQEQSPSQYHYLAQGGCFTLPGRGDAKD FADIRAAMKVLSFKPEEVWSILSLLAAILHLGNLRFTATEVANLATAEIDDTPNLQRVAQ LLGIPISALNAALTQRTIFVHGEHVTTSLSKEAAIEGRDAFVKSLYDGIFVRIVRRINET INKQVDQPMNSIGVLDIFGFENFDNNSFEQLCINYANENLQQFFVGHIFKMEQDEYQNEH INWQHIEFQDNQQILDLIGMKPMNLMSLIDEESKFPKGTDQTLLEKLHVQHGNRSIYVKG KTTQTSLFGIRHYAGVVMYNPLGFLEKNRDSFSGDLRTLVQRSTNKYLVDIFPHEMPMDT AKKQPTLCVKFRNSLDMLMRTLSQAHPYFIRCIKPNEYKEPKNFDKELCVRQLRYSGMME TARIRRAGYPIRHAYRAFVERYRLLVPPVGPLEQCDCRKLARQICEVALPADSDRQYGKT KLFLRDEDDASLELQRSQLMLKSIVTIQRGIRRVLFRRYMKRYREAIITVQRYWRGRLQR RKYQVMRQGFHRLGACIAAQQLTTKFTMVRCRTIKLQALSRGYLVRKDFQKKLLERRKQN QLKKEELLKLAKMKEAEELLRLQQLKEQKEREQREQQEKRLQEEQRLKAEAAARNALAMA AVQQKRRTKPVKQEAPKAPTLQARNSLPPPPTTLIVAAPLPTRPASAVTRINTIPESPGT IDVESSKQMVDDVFRFLNDEPDAALRKLNNISSGDTIRLPKSVPNNIDTSDFSYLKYAAT YFGGGATAQHERKPLKKSLLKHEHPIDEMASKAIWLTILRFMGDLPDVVSSPTLHVFDNE NLMSDLASLLNTSDSYKPRLFVRQSQRRIPKPLASGEKEAQEFYQHWLNVPTSHLEKIHF IIGHGIIKNSLRDEILAQICKQLYLNPSRSSYSRGWLLLSLCLSCFPPSKEFEPHLRSFM KQGTAQLQATPSLQRLERTLVNGPRCQPPSLFELHAIRGRHPLRLDIHLMDGQQRRLQVD AASTAREAVNQLCQGMGLTDTFGFGLVMSLNGKLMPLGAGQEHVLDAISECEQRQLDAPW KLYIRKEMFATWYDPSMDPKATQLIYKQILNGLKCGEYRCRSEKDIAMVCALACFVEYGP GEILRLKPSEITAFVPSDLLAPGERAIENWSRLIAATYEKSSYVKEEQNDLLLEAQKRAK EDICLFAHLSWPMRHSRLFEVVRKEGPKLQSDELMLGINSAGLFLIDETEQVLASCCFSE VLKVHVESDDKLHVMTFQHVNFVLQCSSAQDANEVINYMLDNLRQRSSYGVALDPVVEGD LEDCLVLNPGDLIEFEAGVTGAQLMAGNAQDCYRGCVNGQWGQFLAGNVRVLATLTKPSE KLQDILREGRFQEPPKPTPRANYSRRRQHNISQLAESHFREPLDSDKAPLSKFSPEPLKA PLLKAVVKVPPLFQQALVMHHHILKYMGDIARSNLPVNTDLIFQPALQHPLLCDELYCQL MKQLSDNPSSESEKRGWDLLYLATGLVAPSVLVMRELIILLRMRADALADACLKRLKRSL AQGQRKKAPHLIEVEGIQQRCLHIYHKIYFPDDTVEAFEIESHTRGAELIADIAQRLELK SPVGYSIFLKTGDRVYAMPEEEFVFDFITQLIYWLRQQRTIRSISDGQYQLHFMRKLWLN NHPGEDLNGDMIFSYPQELHKYLKGYYPIDCEQASRLAILVYSADHDVSLQRLPEVLTRL IPEDLIPLQTVAEWRQQILPKVHRDHLTEDHAKILFLQELSHFACFGSTFFVVKQQNDDA LPETLLIAINSTGFHMLDPTTKEILRSYEYSQLGIWSSGKNHFHIRFGNMIGASKLLCST TQGYKMDDLLASYVKYFNEHE
Expert options	Gap to Close = 15 aa
Search for Alternatively Spliced Exons	defaults Enable search!

Result

1 gi|116010444|ref|NT_033779.4| (10516bp)

All cDNA hits merged:

cDNA hits:
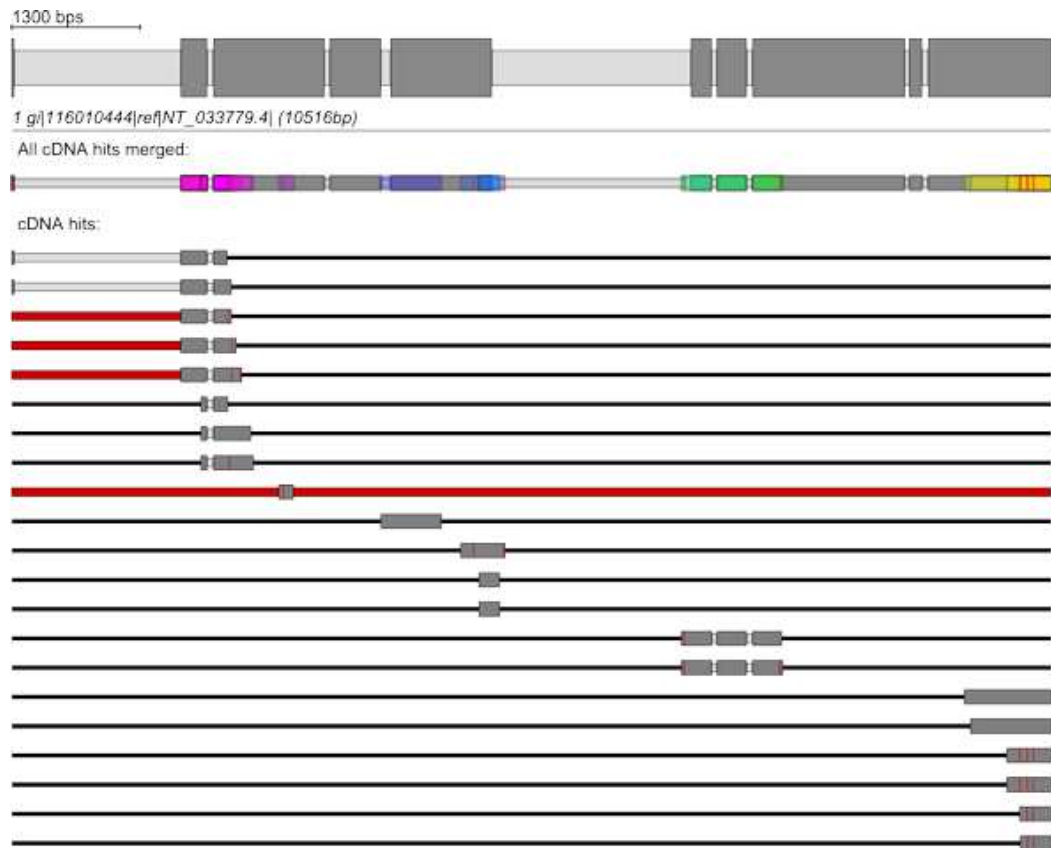
Activity diagrams

Detailed diagram of the cDNA search algorithm

Start

Run Scipio → Scipio result (YAML)

cDNA data (FASTA)

Run BLAT → 1st BLAT result (PSL)

Extract FASTA entries → 1st extracted cDNA data (FASTA)

Run BLAT → 2nd BLAT result (PSL)

Extract FASTA entries → 2nd extracted cDNA data (FASTA)

Loop: For each hit

Loop: End

Read BLAT result

Convert coordinates

Combine blocks

Loop: For each block

Loop: End

Extract subsequences of blocks

Tranlate subsequences

Determine mismatches

Identify and mark alternatively spliced exons

Store hits in ScipioResult

Generate exon-intron gene structure picture → Exon-intron gene structure picture

Generate alignment output → Alignment output

Generate files for download →

| YAML |
| GFF |
| SVG |
| Log |

End