

Splicing Alternativo

Dr. Pablo Rodrigo Sanches

Departamento de Genética – FMRP/USP

psanches@usp.br

Algumas definições

- “O splicing alternativo é um processo pelo qual os exons de um transcrito primário são ligados de diferentes maneiras durante o processamento do RNA, levando à síntese de proteínas distintas....”;
- “O splicing alternativo é um processo que gera diferentes mRNA. Estes geralmente codificam diversos produtos proteicos a partir de um gene. Isto então aumenta drasticamente a capacidade codificante dos genes...”.

Modos alternativos de splicing

Exon skipping (uso alternativo de exon)



Alternative 5' splice sites (sítio alternativo 5')



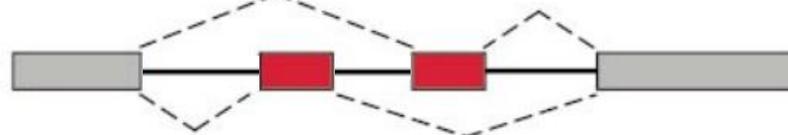
Alternative 3' splice sites (sítio alternativo 3')



Intron retention (retenção de intron)

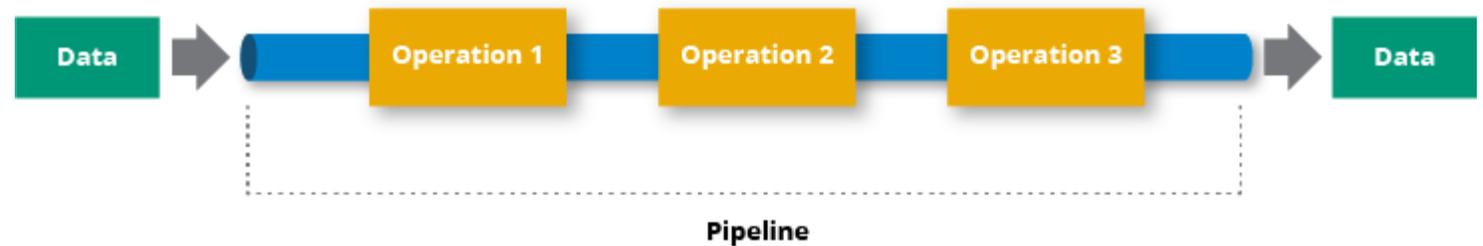


Mutually exclusive exons (exons mutuamente excludentes)

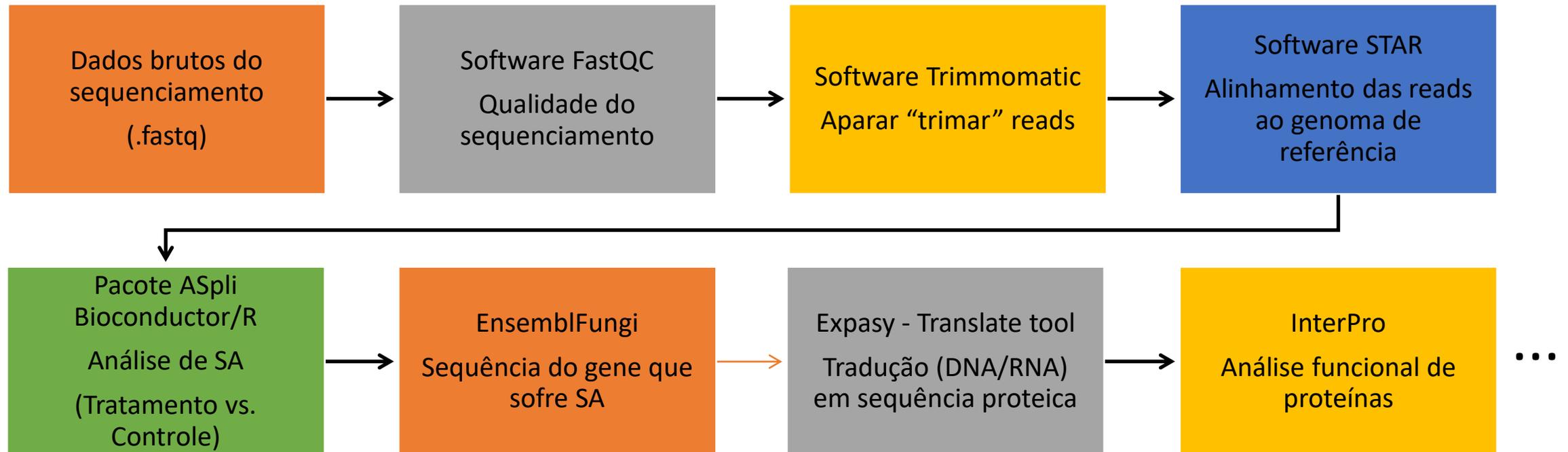


Na figura, exons constitutivos estão desenhados em cinza e regiões que sofrem SA em vermelho. Introns estão representados por linhas. Em tracejado indicam-se os eventos de splicing (modificado de Fardilha et al., 2008).

Utilizando dados de RNA-Seq na busca por eventos de SA



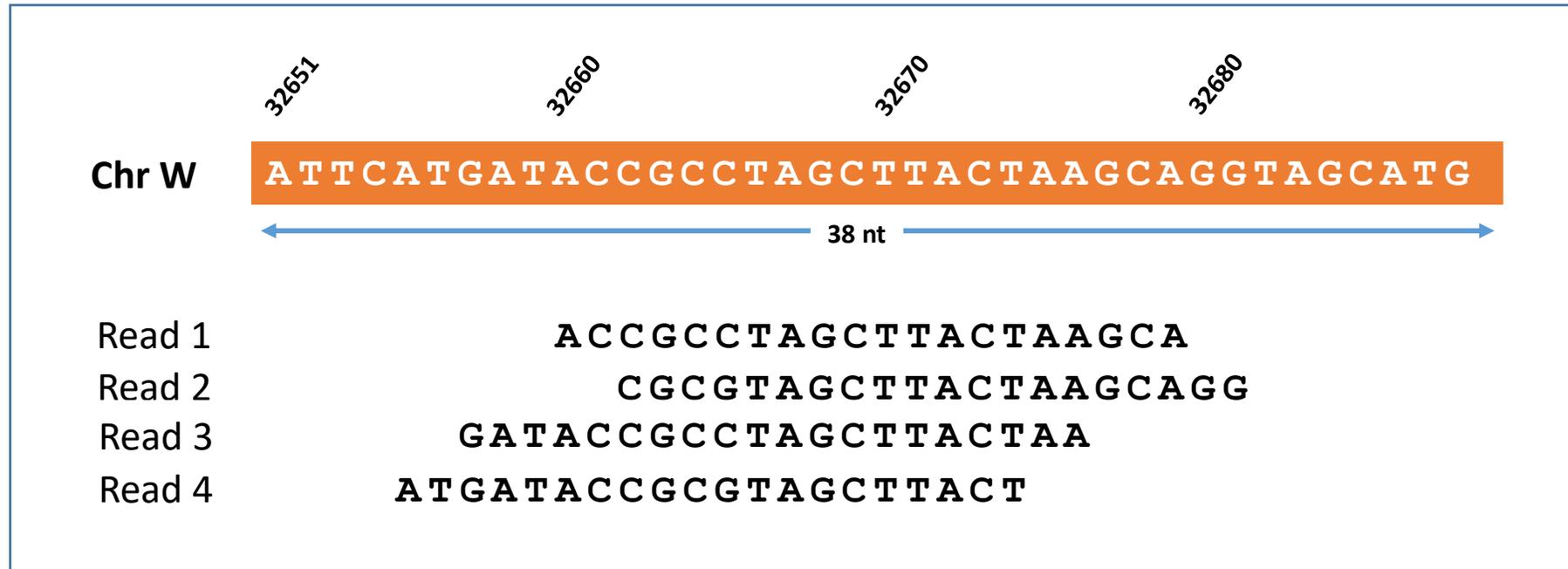
Exemplo de pipeline



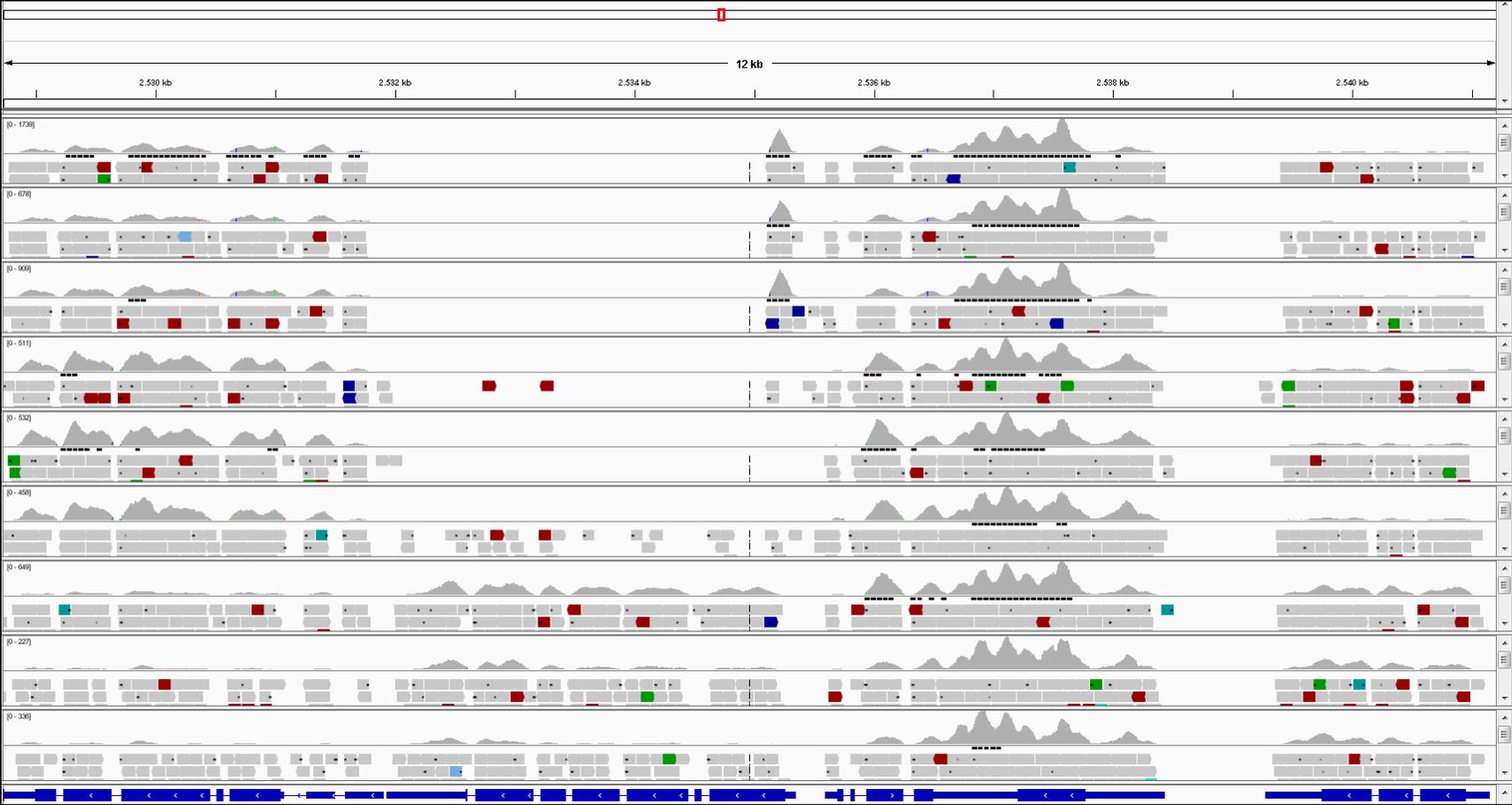
Vamos recordar...

- O que é o formato SAM/BAM?
 - SAM é o formato genérico para armazenar grandes quantidades de alinhamentos de nucleotídeos.
 - São arquivos tabulados em formato de texto que contém as informações sobre os alinhamentos das leituras (reads) à sequência-alvo (ex. genoma de referência).
 - BAM é o formato de arquivo comprimido, tornando-o mais compacto.
 - Arquivos SAM podem ser visualizados por editores de texto, enquanto arquivos BAM não podem.
 - Arquivos BAM podem ser facilmente usados por diferentes programas de análise de sequências.

Exemplo – Esquema para Arquivo SAM



Visualização gráfica de arquivos BAM



Script ASpli

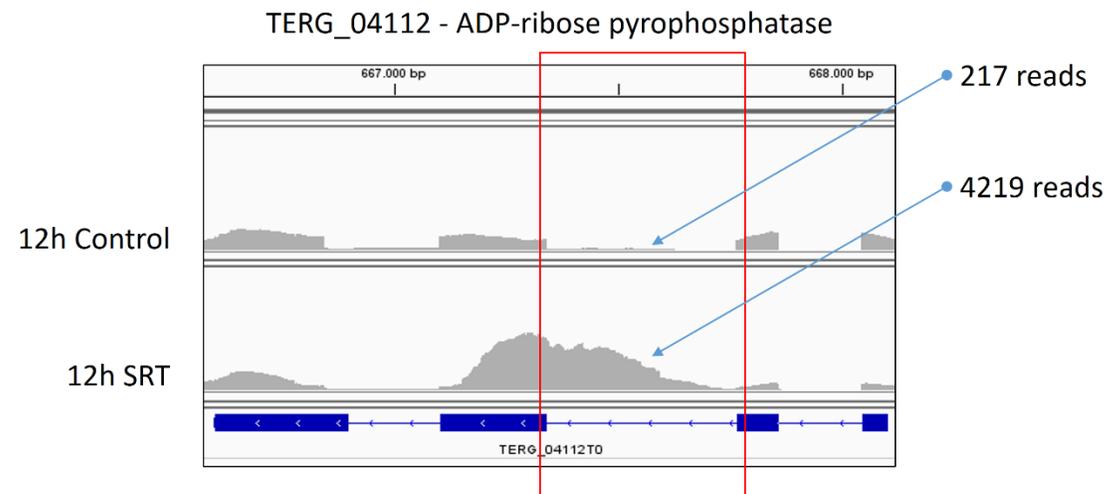
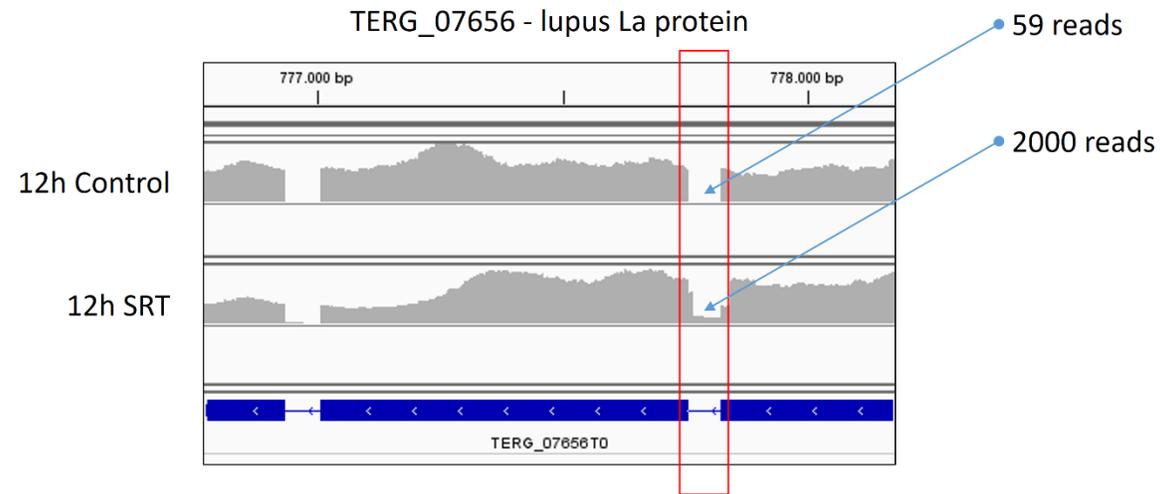
Mais informações em: <https://bioconductor.org/packages/release/bioc/html/ASpli.html>

```
1 library(ASpli)
2 library(GenomicFeatures)
3 setwd("C:/Users/lgmm/Desktop/Desktop/RNASeq_sertralina/ASpli/3h")
4 TxDb <- makeTxDbFromGFF(file="../gtf/trichophyton_rubrum_cbs_118892_2_exons.gtf",format="gtf")
5 features <- binGenome( TxDb )
6 geneCoord <- featuresg( features )
7 binCoord <- featuresb( features )
8 junctionCoord <- featuresj( features )
9 binMetadata <- mcols( binCoord )
10 features
11 bamFiles <- c( "bam/TRSAB3h_I.bam", "bam/TRSAB3h_II.bam", "bam/TRSAB3h_III.bam", "bam/TRSABStr3h_
12 bamFiles
13 targets <- data.frame( row.names = c("TRSAB3h_I","TRSAB3h_II","TRSAB3h_III", "TRSABStr3h_I","TRS
14 targets
15 getConditions( targets )
16 bam <- loadBAM(targets)
17 memory.limit(9999999999)
18 counts <- readCounts (features, bam, targets, cores = 1, readLength = 150, maxISize = 50000 )
19 GeneCounts <- countsg(counts)
20 GeneRd <- rdsg(counts)
21 BinCounts <- countsb(counts)
22 BinRd <- rdsb(counts)
23 JunctionCounts <- countsj(counts)
24 writeCounts(counts=counts, output.dir = "results")
25 writeRds(counts=counts, output.dir = "results")
26 eliCounts <- countseli(counts)
27 ie2Counts <- countsie2(counts)
```

Resultado ASpli

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	id	event	Gene Product Name	gene_coordinates	start	end	length	logFC	pvalue	Sab3h_I	Sab3h_II	Sab3h_III	Str3h_I	Str3h_II	Str3h_III
2	TERG_07560:I002	-	hypothetical protein	Supercontig2.10:529914-532726	532542	532600	59	4,02	5,47E-07	0	0	0	5	6	9
3	TERG_00243:E007	IR	pathogenesis associated protein Cap20, putative (T. verrucosum)	Supercontig2.1:637124-640572	640225	640272	48	2,88	2,63E-52	206	174	126	1279	1540	2277
4	TERG_08430:I002	-	Neurofilament heavy polypeptide (T. tonsurans)	Supercontig2.14:193767-201907	195916	195978	63	2,77	3,35E-20	4	4	4	112	57	81
5	TERG_08333:I002	-	1-pyrroline-5-carboxylate dehydrogenase	Supercontig2.13:250906-253214	251460	251519	60	2,70	8,07E-19	6	4	2	26	29	34
6	TERG_04145:E010	IR	ATP synthase subunit beta, mitochondrial	Supercontig2.4:761445-764665	763650	763721	72	2,12	8,09E-52	348	321	393	1439	1546	2248
7	TERG_06358:I007	-	dicer (T. tonsurans)	Supercontig2.7:1029769-1035122	1E+06	1E+06	54	2,04	1,56E-05	1	3	2	7	4	8
8	TERG_03566:I003	-	GYF domain-containing protein (T. equinum)	Supercontig2.3:2195805-2200733	2E+06	2E+06	66	2,02	7,97E-10	5	5	12	25	34	58
9	TERG_07169:I002	-	hypothetical protein	Supercontig2.9:491247-493680	492424	492482	59	2,01	1,42E-04	2	2	0	3	11	9
10	TERG_07200:I003	-	C2 domain-containing protein (T. tonsurans)	Supercontig2.9:590060-594795	593696	593755	60	1,91	2,06E-29	24	30	22	171	163	189
11	TERG_03599:I002	-	metalloproteinase (T. equinum)	Supercontig2.3:2274093-2275292	2E+06	2E+06	58	1,90	7,80E-04	0	0	2	8	8	7
12	TERG_07409:I001	-	amino acid permease (T. equinum)	Supercontig2.10:137576-139612	138029	138116	88	1,88	5,07E-09	8	9	2	86	37	73
13	TERG_00683:I002	-	hypothetical protein	Supercontig2.1:1762701-1765796	2E+06	2E+06	81	1,87	1,11E-03	0	2	1	8	3	4
14	TERG_02789:I002	-	hypothetical protein	Supercontig2.3:209282-211522	209633	209704	72	1,86	1,10E-05	5	13	13	23	34	123
15	TERG_08846:I001	-	hypothetical protein	Supercontig2.5:852946-854767	854122	854170	49	1,82	1,82E-03	0	2	0	7	5	8
16	TERG_00286:I001	-	ABC transporter (T. tonsurans)	Supercontig2.1:730323-732834	731173	731230	58	1,77	4,77E-03	0	0	3	2	7	15
17	TERG_07656:I002	-	lupus La protein (T. equinum)	Supercontig2.10:776774-778176	777760	777825	66	1,73	1,96E-36	90	96	98	318	304	399
18	TERG_02301:I002	-	hypothetical protein	Supercontig2.2:1916640-1917726	2E+06	2E+06	61	1,72	1,37E-04	0	3	3	6	10	11
19	TERG_00759:I001	-	conidiophore development protein HymA (T. tonsurans)	Supercontig2.1:2018858-2020612	2E+06	2E+06	53	1,68	2,04E-09	9	5	7	22	21	23
20	TERG_07418:I001	-	MFS multidrug transporter (T. equinum)	Supercontig2.10:159907-161796	160646	160704	59	1,67	8,34E-11	6	19	12	50	75	82
21	TERG_06000:I001	-	glycolate oxidase, subunit GlcD	Supercontig2.7:26377-28835	26797	26898	102	1,67	2,69E-10	6	9	11	44	28	34
22	TERG_01443:I004	-	ABC multidrug transporter (T. tonsurans)	Supercontig2.1:3800869-3807381	4E+06	4E+06	62	1,66	3,85E-04	2	2	3	16	4	16
23	TERG_01938:I001	-	hypothetical protein	Supercontig2.2:1034813-1036840	1E+06	1E+06	56	1,66	3,48E-14	19	27	23	67	123	93
24	TERG_08902:I002	-	hypothetical protein	Supercontig2.7:1011556-1012239	1E+06	1E+06	83	1,65	1,07E-03	0	1	4	10	4	12
25	TERG_07753:I001	-	hypothetical protein	Supercontig2.11:7488-8035	7583	7634	52	1,64	1,63E-02	0	1	0	6	4	8
26	TERG_02845:I002	-	cercosporin toxin biosynthesis protein (T. equinum)	Supercontig2.3:351162-353375	352305	352369	65	1,63	2,56E-07	5	13	12	19	26	44
27	TERG_07695:I001	-	pH-response regulator protein palA/RIM20 (T. tonsurans)	Supercontig2.10:879998-882289	880729	880790	62	1,61	4,04E-04	3	2	8	6	15	28
28	TERG_00334:I001	-	hypothetical protein	Supercontig2.1:850186-853919	851156	851210	55	1,60	1,78E-13	26	28	26	83	108	163
29	TERG_00220:I003	-	vacuolar assembly protein (T. equinum)	Supercontig2.1:566700-570748	568054	568106	53	1,58	3,35E-03	0	2	1	5	7	7
30	TERG_05678:I001	-	topoisomerase 1-associated factor 1 (T. equinum)	Supercontig2.6:727160-730848	728206	728258	53	1,57	1,68E-03	0	13	8	19	24	6
31	TERG_04580:E010	IR	NADP-specific glutamate dehydrogenase	Supercontig2.4:1840673-1843592	2E+06	2E+06	76	1,57	2,37E-47	527	490	558	923	1031	1128
32	TERG_05540:I002	-	cytochrome P450 monooxygenase, putative (T. verrucosum)	Supercontig2.6:354838-357252	355788	355858	71	1,56	2,99E-03	0	2	1	7	8	7
33	TERG_04580:I005	-	NADP-specific glutamate dehydrogenase	Supercontig2.4:1840673-1843592	2E+06	2E+06	70	1,55	5,05E-40	359	394	463	665	794	866
34	TERG_00757:I004	-	hypothetical protein	Supercontig2.1:2015019-2015909	2E+06	2E+06	136	1,53	4,76E-04	7	3	1	10	5	11

Exemplo – Retenção de Intron



EnsemblFungi – Busca por gene

The screenshot displays the EnsemblFungi interface for the gene **TERG_07656** in **Trichophyton rubrum CBS 118892 (ASM15142v1)**. The browser address bar shows the URL: `fungi.ensembl.org/Trichophyton_rubrum_cbs_118892_gca_000151425/Gene/Summary?g=TERG_07656;r=supercont2.10:776774-778176;t=EGD91437;db=core`.

Gene-based displays sidebar (left):

- Summary
- Splice variants
- Transcript comparison
- Gene alleles
- Sequence
- Secondary Structure
- Gene families
- Literature
- Fungal Compara
- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues
- Pan-taxonomic Compara
- Gene Tree
- Orthologues
- Ontologies
- GO: Biological process
- GO: Cellular component
- GO: Molecular function
- Phenotypes
- Genetic Variation
- Variant table
- Variant image
- Structural variants
- Gene expression
- Pathway
- Molecular interactions
- Regulation
- External references
- Supporting evidence
- ID History
- Gene history

Gene: TERG_07656

Description: hypothetical protein

Location: [SuperContig_supercont2.10:776,774-778,176](#) reverse strand.
ASM15142v1:GG700657.1

About this gene: This gene has 1 transcript ([splice variant](#)).

Transcripts: [Hide transcript table](#)

Table: Show/hide columns (1 hidden) | Filter

Name	Transcript ID	bp	Protein	Biotype	Flags
-	EGD91437	1263	420aa	Protein coding	Ensembl Canonical

Summary

Gene type: Protein coding

Annotation method: Protein coding genes annotated in [ENA](#)

Summary description: Go to [Region in Detail](#) for more tracks and navigation options (e.g. zooming)

Genes track: 21.40 kb Forward strand

Genes track details (from left to right):

- 768kb: KFL62772 > protein coding
- 770kb: EGD91433 > protein coding
- 772kb: EGD91434 > protein coding
- 774kb: EGD91436 > protein coding
- 776kb: EGD91438 > protein coding
- 778kb: EGD91437 > protein coding
- 780kb: KFL62779 > protein coding
- 782kb: KFL62773 > protein coding
- 784kb: KFL62789 > protein coding
- 786kb: KFL62775 > protein coding

Buttons at the bottom: Configure this page, Custom tracks

Constitutivo vs. Retenção-Intron

>TERG_07656

```
ATGGCGGAAGAGCAGAAAGTAGCTGCGGCCGTGGATGCGACCGCCGATAATGCCGCCGCA
GAGCAGGATGTGAAGGAAGTCCTGGCTGAGCTCAAATCTGACGAAGCCAGCAAGCAGGAC
AGCGCGGATGCTGAAAAGGCCGAAGAAGAAAAGATCGTTGCGGCTGCTAAGAGGCTAGGC
GAAGAAGCTCTGTCAAATGAGACTGCAAAAGAAGCGTCAGAGACACAAAAGGGGCGTGCC
AGCGGCCGTGGCCGTGGCCGTGGAGGAGTGCGCATCAACTACCGTGACAATATCAAATCG
GATCCATCCTCCCTGGAGGAGACAGATGATCCTGTTGAGATCAGAAAGCAGGTGAGTGTG
GCCGCTCTGGCTTTTCCATGTTGTATGCAAGTTGGACTTTACTCACCTTTGACAGGTT
GAATTCTACTTCTCTGACTCCAACCTACCAATGGACAAGTTCCTCCTCTCAAAGTTGGC
GGTAGCGAGAACAGGCCAGTTGAGCTCGCTCTTCTTCATTCGTTCAAGCGAATGCGCCGT
TTCCAGCCTTTTCAGCGCTATTGTTGAAGCCCTCAAGAGCTCAGAGCTCGTCGAACTGGTA
GATGACGACAAAGCTGTGCGCCGAAGGTCCCTCTTCCAGACACCATCAAGGAGACAGCC
GATTCATCAGCCGTCAAATATTCGAGGATAAAGCCATGCATCGCAGCATATACGCCAAG
GGATTCGGTCCAGAGGAACCCAGTACCCAGTTTGACATTGAAGCCTTCTTTACTCCCTAT
GGCCCCACCAACGCTGTTCTGACTGAAAGCCTGCAAAAGGCATTTCTGGCAGTTGAACCAAAGCCG
AAGTGAAGGGCACCACAGAATTAATTATCAAGAGCAAGAAGCAGTACTGTGACGAGAAG
ATCAAGGAGATAGAAGCTGGTCGTCTGAAGCCTAGTGACCGGTCTAGTGGACGAGGTGGT
CGAGGAGGCCGCGGAGGCCGTGGAGGTCTGGTGGACGGGGCGGTGAGGTGGCCGTGGT
AATGGTCGGGATCGTAGTGACCGAAACAACGGTGCCCAAGTAAAAGAAGAAGCCAGGCA
AAACGCCCTGAACCAGAGAAAGATAGCCGGTATGTTCTTGCAAGTTCGCAATTAGTGGGT
TACAGCCCGCATAGGCACAATTTAACTTGGCTTTTGATATAGCGCTGTTCTGTCTATCC
AGGTATCCAGTAAACAAGGCTGAATCTCAACAAAATGGAGGCGCAATGGCCAAAACGCA
GCCGTGAAGAAGACTCTGGGCCCAAGGCTGATGGCAATACTGAGGAGAGACCAGCTAAGA
AGGTTGATGCTAAAGATAGTTAA
```

>TERG_07656_MRNA_CONSTITUTIVO

```
ATGGCGGAAGAGCAGAAAGTAGCTGCGGCCGTGGATGCGACCGCCGATAATGCCGCCGCA
GAGCAGGATGTGAAGGAAGTCCTGGCTGAGCTCAAATCTGACGAAGCCAGCAAGCAGGAC
AGCGCGGATGCTGAAAAGGCCGAAGAAGAAAAGATCGTTGCGGCTGCTAAGAGGCTAGGC
GAAGAAGCTCTGTCAAATGAGACTGCAAAAGAAGCGTCAGAGACACAAAAGGGGCGTGCC
AGCGGCCGTGGCCGTGGCCGTGGAGGAGTGCGCATCAACTACCGTGACAATATCAAATCG
GATCCATCCTCCCTGGAGGAGACAGATGATCCTGTTGAGATCAGAAAGCAGGTGAATTC
TACTTCTCTGACTCCAACCTACCAATGGACAAGTTCCTCCTCTCCAAAGTTGGCGTAGC
GAGAACAGGCCAGTTGAGCTCGCTCTTCTTCATTGTTCAAGCGAATGCGCCGTTTCCAG
CCTTTTCAGCGCTATTGTTGAAGCCCTCAAGAGCTCAGAGCTCGTCGAACTGGTAGATGAC
GACAAAGCTGTGCGCCGAAGGTCCCTCTTCCAGACACCATCAAGGAGACAGCCGATTC
TCAGCCGTCAAATATTCGAGGATAAAGCCATGCATCGCAGCATATACGCCAAGGGATT
GGTCCAGAGGAACCCAGTACCCAGTTTGACATTGAAGCCTTCTTTACTCCCTATGGCCCC
ACCAACGCTGTTCGTCTTAGACGTGCCATGGATAAGACATTCAAGGGTAGCGCTGTTGTC
GAGTTTGAGACTGAAGACCTGCAAAAGGCATTTCTGGCAGTTGAACCAAAGCCGAAGTGG
AAGGGCACCAAGAATTACTTATCAAGAGCAAGAAGCAGTACTGTGACGAGAAGATCAAG
GAGATAGAAGCTGGTCTGTAAGCCTAGTGACCGGTCTAGTGGACGAGGTGGTCCAGGA
GGCCGCGGAGGCCGTGGAGGTCTGGTGGACGGGGCGGTGAGGTGGCCGTGGTAAATGGT
CGGGATCGTAGTGACCGAAACAACGGTGCCCAAGTAAAAGAAGAAGCCAGGCAAAACGC
CCTGAACCAGAGAAAGATAGCCGCGCTGTTCTGTCTATCCAGGTATCCAGTAAACAAGCT
GAATCTCAACAAAATGGAGGCGCAATGGCCAAAACGCAAGCCGTGAAGAAGACTCTGGG
CCCAAGGCTGATGGCAATACTGAGGAGAGACCAGCTAAGAAGGTTGATGCTAAAGATAGT
TAA
```

>TERG_07656_MRNA_INTRON1_RETIDO

```
ATGGCGGAAGAGCAGAAAGTAGCTGCGGCCGTGGATGCGACCGCCGATAATGCCGCCGCA
GAGCAGGATGTGAAGGAAGTCCTGGCTGAGCTCAAATCTGACGAAGCCAGCAAGCAGGAC
AGCGCGGATGCTGAAAAGGCCGAAGAAGAAAAGATCGTTGCGGCTGCTAAGAGGCTAGGC
GAAGAAGCTCTGTCAAATGAGACTGCAAAAGAAGCGTCAGAGACACAAAAGGGGCGTGCC
AGCGGCCGTGGCCGTGGCCGTGGAGGAGTGCGCATCAACTACCGTGACAATATCAAATCG
GATCCATCCTCCCTGGAGGAGACAGATGATCCTGTTGAGATCAGAAAGCAGGTGAGTGTG
GCCGCTCTGGCTTTTCCATGTTGTATGCAAGTTGGACTTTACTCACCTTTGACAGGTT
GAATTCTACTTCTCTGACTCCAACCTACCAATGGACAAGTTCCTCCTCTCAAAGTTGGC
GGTAGCGAGAACAGGCCAGTTGAGCTCGCTCTTCTTCATTCGTTCAAGCGAATGCGCCGT
TTCCAGCCTTTTCAGCGCTATTGTTGAAGCCCTCAAGAGCTCAGAGCTCGTCGAACTGGTA
GATGACGACAAAGCTGTGCGCCGAAGGTCCCTCTTCCAGACACCATCAAGGAGACAGCC
GATTCATCAGCCGTCAAATATTCGAGGATAAAGCCATGCATCGCAGCATATACGCCAAG
GGATTCGGTCCAGAGGAACCCAGTACCCAGTTTGACATTGAAGCCTTCTTTACTCCCTAT
GGCCCCACCAACGCTGTTCTGACTGAAAGCCTGCAAAAGGCATTTCTGGCAGTTGAACCAAAGCCG
AAGTGAAGGGCACCACAGAATTAATTATCAAGAGCAAGAAGCAGTACTGTGACGAGAAG
ATCAAGGAGATAGAAGCTGGTCTGTAAGCCTAGTGACCGGTCTAGTGGACGAGGTGGT
CGAGGAGGCCGCGGAGGCCGTGGAGGTCTGGTGGACGGGGCGGTGAGGTGGCCGTGGT
AATGGTCGGGATCGTAGTGACCGAAACAACGGTGCCCAAGTAAAAGAAGAAGCCAGGCA
AAACGCCCTGAACCAGAGAAAGATAGCCGCGCTGTTCTGTCTATCCAGGTATCCAGTAA
AAGGCTGAATCTCAACAAAATGGAGGCGCAATGGCCAAAACGCAAGCCGTGAAGAAGAC
TCTGGGCCCAAGGCTGATGGCAATACTGAGGAGAGACCAGCTAAGAAGGTTGATGCTAAA
GATAGTTAA
```

Expasy - Translate tool

The screenshot shows a web browser window with the URL `web.expasy.org/translate/`. The page title is "Translate" and the Expasy logo is in the top left. The main heading is "Translate tool". Below this, a light blue box contains the tool's description: "Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence." The interface is divided into several sections: "DNA or RNA sequence" with a text input field and instructions; "Output format" with radio button options for "Verbose", "Compact", "Includes nucleotide sequence", and "Includes nucleotide sequence, no spaces"; "DNA strands" with checked checkboxes for "forward" and "reverse"; and "Genetic codes" with a dropdown menu set to "Standard". At the bottom of the tool box are "reset" and "TRANSLATE!" buttons. The footer includes the SIB logo, the text "Expasy is operated by the SIB Swiss Institute of Bioinformatics", and links for "Terms of Use" and "Privacy policy". A "Back to the top" link is in the bottom right corner.

Expasy - Translate tool

web.expasy.org/translate/

Translate

Home Programmatic Access Contact

Translate tool

Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

DNA or RNA sequence

Please enter a DNA or RNA sequence - numbers and blanks are ignored

Output format

- Verbose: Met, Stop, spaces between residues
- Compact: M, -, no spaces
- Includes nucleotide sequence
- Includes nucleotide sequence, no spaces

DNA strands

- forward
- reverse

Genetic codes - [See NCBI's genetic codes](#)

Standard

reset TRANSLATE!

Expasy is operated by the SIB Swiss Institute of Bioinformatics
[Terms of Use](#) | [Privacy policy](#)

[Back to the top](#)

Constitutivo vs. Retenção-Intron

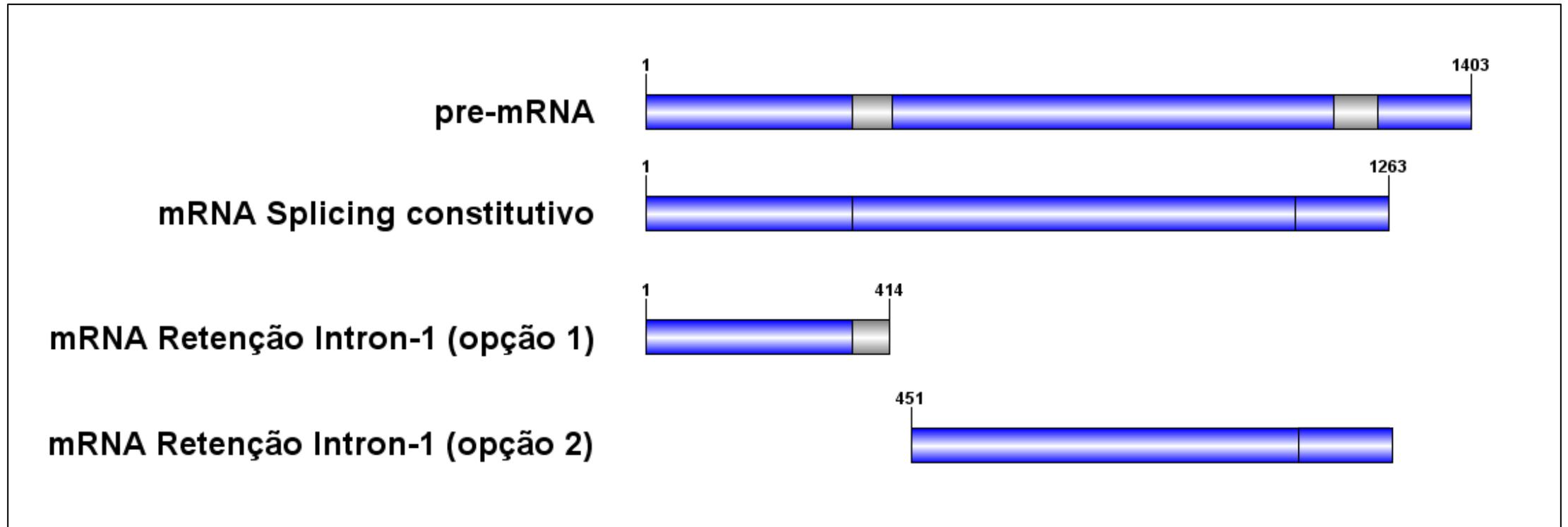
5'3' Frame 1

MAEEQKVAAAVDATADNAAAEQDVKEVLAEELKSDEASKQDSADA EKAE E EKIVAAAKRLGEEALSNETAKEASETQKGRGSGRGRGRGGVVRINYRDNIKSDPSSLEETDDPVEIRKQVEFYFSDSNLPMDKFLLSKVGGSENRPVELALLHSFKRMRRFQPFS AIVEALKSSELVELVDDDKAVRRKVPLPDTIKETADSSAVKIFEDKAMHRSIYAKGFGPEEPSTQFDIEAFFTPYGPTNAVRLRRAMDKTFKGSVFVEFETEDLQKAFLAVEPKPKWKGTTELLIKSKKQYCDEKIKEIEAGRLKPSDRSSGRGGRRGGRRGGRRGGRRGGRRGNGRDRSDRNNGAQVKEEAQAKRPEPEKDSRAVPVIQVSSNKAESQONGGANGQKRSREEDSGPKADGNTEERPAKKVDAKDS-

5'3' Frame 1

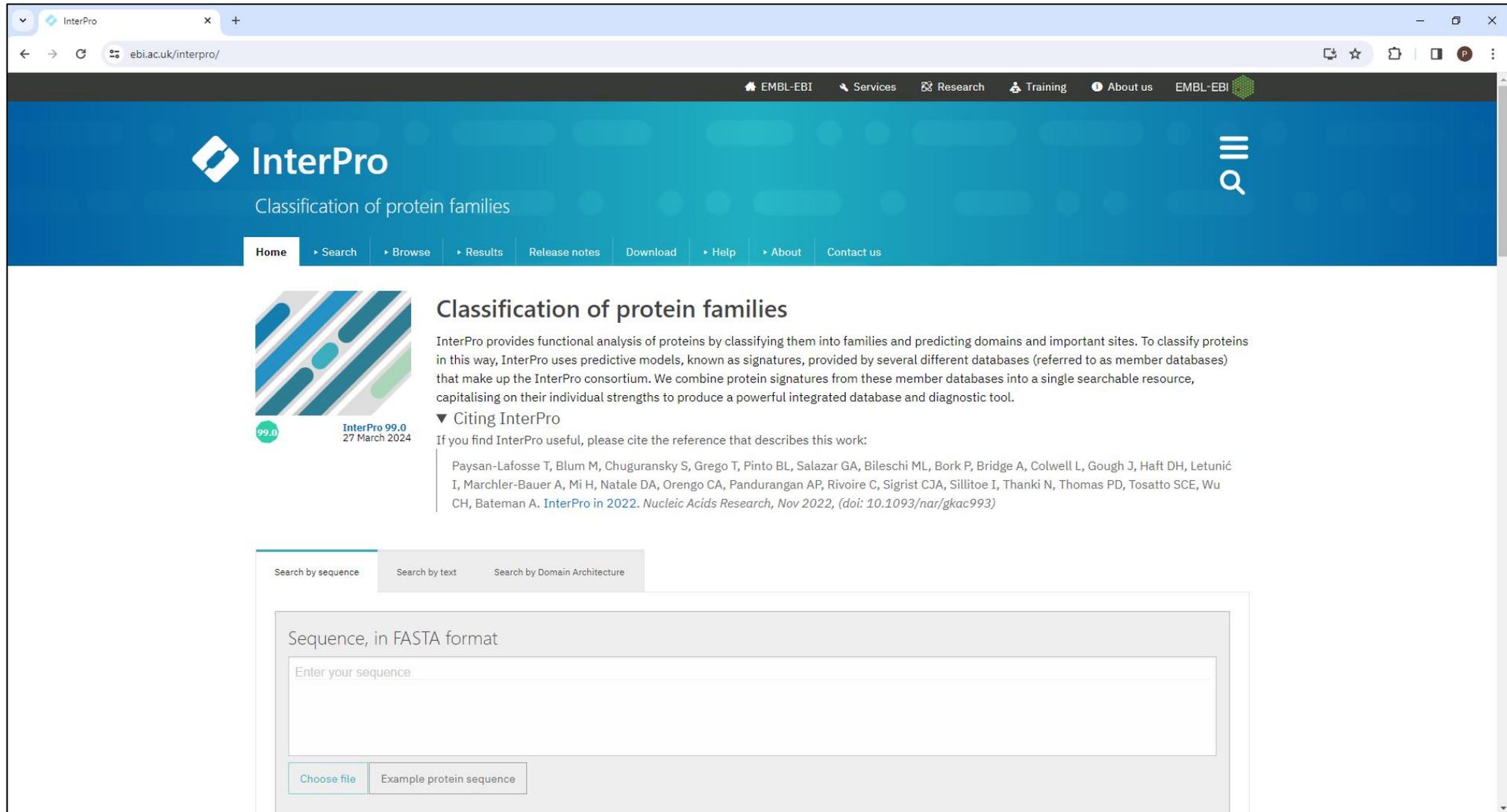
MAEEQKVAAAVDATADNAAAEQDVKEVLAEELKSDEASKQDSADA EKAE E EKIVAAAKRLGEEALSNETAKEASETQKGRGSGRGRGRGGVVRINYRDNIKSDPSSLEETDDPVEIRKQVSVAA SGFSMLYASWTLTLL-QVEFYFSDSNLPMDKFLLSKVGGSENRPVELALLHSFKRMRRFQPFS AIVEALKSSELVELVDDDKAVRRKVPLPDTIKETADSSAVKIFEDKAMHRSIYAKGFGPEEPSTQFDIEAFFTPYGPTNAVRLRRAMDKTFKGSVFVEFETEDLQKAFLAVEPKPKWKGTTELLIKSKKQYCDEKIKEIEAGRLKPSDRSSGRGGRRGGRRGGRRGGRRGNGRDRSDRNNGAQVKEEAQAKRPEPEKDSRAVPVIQVSSNKAESQONGGANGQKRSREEDSGPKADGNTEERPAKKVDAKDS-

Constitutivo vs. Retenção-Intron



(Desenvolvido no Software IBS)

Interpro



The screenshot shows the InterPro website homepage. The browser address bar displays "ebi.ac.uk/interpro/". The navigation menu includes "Home", "Search", "Browse", "Results", "Release notes", "Download", "Help", "About", and "Contact us". The main heading is "InterPro Classification of protein families". A secondary heading reads "Classification of protein families". The introductory text states: "InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. To classify proteins in this way, InterPro uses predictive models, known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro consortium. We combine protein signatures from these member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool." Below this is a section titled "Citing InterPro" with the following text: "If you find InterPro useful, please cite the reference that describes this work: Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. *InterPro in 2022. Nucleic Acids Research, Nov 2022, (doi: 10.1093/nar/gkac993)*". At the bottom, there is a search interface with three tabs: "Search by sequence", "Search by text", and "Search by Domain Architecture". The "Search by sequence" tab is active, showing a text input field labeled "Sequence, in FASTA format" with the placeholder "Enter your sequence". Below the input field are two buttons: "Choose file" and "Example protein sequence".

InterPro
Classification of protein families

Home Search Browse Results Release notes Download Help About Contact us

Classification of protein families

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. To classify proteins in this way, InterPro uses predictive models, known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro consortium. We combine protein signatures from these member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool.

▼ Citing InterPro

If you find InterPro useful, please cite the reference that describes this work:

Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. *InterPro in 2022. Nucleic Acids Research, Nov 2022, (doi: 10.1093/nar/gkac993)*

Search by sequence Search by text Search by Domain Architecture

Sequence, in FASTA format

Enter your sequence

Choose file Example protein sequence

Constitutivo vs. Retenção-Intron

TERG_07656_PROTEINA_CONSTITUTIVA



TERG_07656_PROTEINA_INTRON1_RETIDO (opção 1)

Length	137 amino acids
Actions	  
Status	✓ finished
Expires ⓘ	Tue Apr 09 2024
Protein family membership	
None predicted	

TERG_07656_PROTEINA_INTRON1_RETIDO (opção 2)



Outros Exemplos

