



Biotecnologia

ACH5545 Engenharia Genética

Atividades de Laboratório

2º Semestre 2024

Docente:

Felipe Chambergo (fscha@usp.br) - <https://sites.usp.br/lbbp>

Monitores:

Augusto Roldan Gonçalves - augusto.roldan@usp.br

Henrique dos Santos Hernandes - hernandesrique@usp.br

Servidores não-docentes:

Tec. Pedro Manoel dos Santos - pedroms@usp.br

Créditos: 4

Período: Quinta-feira (14h00 -18h00), Laboratório de Biotecnologia – Edifício A2, 1º andar

USP - 2024

Análise de Bioinformática
**Ferramentas computacionais (gDNA, cDNA,
sequência de aminoácidos e desenho de
primers)**

Bioinformática: Base de Datos



[COVID-19](#) [ABOUT](#) [PHONES](#) [CONTACTS](#)

Search JGI websites ...

[Our Science](#)

[Our Projects](#)

[Data & Tools](#)

[User Programs](#)

[News & Publications](#)

A BERKELEY LAB NATIONAL USER FACILITY



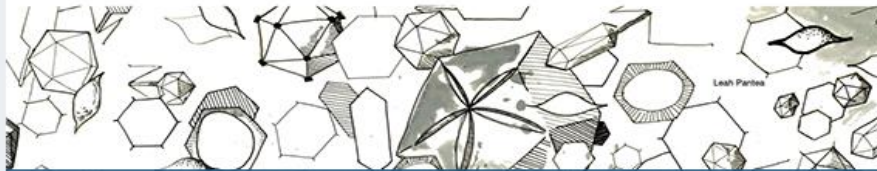
VEGA Viral EcoGenomics & Applications

April 8 - May 13, 2021

Shannon Bennett
Aude Bernheim
Edward Holmes
Mart Krupovic

Jens Kuhn
Karen Maxwell
Monir Moniruzzaman

David Pride
Simon Roux
Rachel Whitaker
Natalya Yutin



Don't miss our VEGA Symposium!

Latest News

FEBRUARY 22, 2021

[Green Algae Reveal One mRNA Encodes Many Proteins](#)

FEBRUARY 8, 2021

[An Age of CRAGE: Advances in Rapidly Engineering Non-model Bacteria](#)

JANUARY 27, 2021

[Fields of Breeders' Dreams: A Team Effort Toward Targeted Crop Improvements](#)

[LATEST JGI NEWS RELEASES](#)

Announcements

Read our [Statement on the Use of Genomics Data](#).



Our Science

How we're contributing to bioenergy research.

[LEARN MORE](#)



User Programs

Everything you need to know about getting a project started.

[LEARN MORE](#)

<https://jgi.doe.gov/>

NCBI Resources How To Sign in to NCBI

NCBI National Center for Biotechnology Information

COVID-19 Public health information

Notice of S April 24, 2022 the eRA website

Ending Structural Racism nih.gov/ending-structural-racism

All Databases

- Gene
- Genome
- GEO DataSets
- GEO Profiles
- GTR
- HomoloGene
- Identical Protein Groups
- MedGen
- MeSH
- NCBI Web Site
- NLM Catalog
- Nucleotide
- OMIM
- PMC
- PopSet
- Protein
- Protein Clusters
- Protein Family Models
- PubChem BioAssay
- PubChem Compound

Information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

Please note that eRA systems will be undergoing scheduled maintenance from 10am until 10pm Eastern US time on dependent services such as MyBibliography and Grant Reporting may be unavailable. More information is available on

and structural racism and achieve racial equity in the biomedical research enterprise.

NCBI National Center for Biotechnology Information advances science and health by providing access to genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

Data for SARS-CoV-2 variants now available at NCBI

23 Apr 2021
Looking for genomes for the B.1.1.7

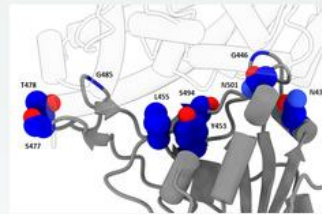
NCBI Home
Resource List (A-Z)
All Resources

- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials

In Focus

Comment on recent variants and spike protein changes

As seen on many occasions before, mutations are naturally expected for viruses and are most often simply neutral regional markers useful for contact tracing. The changes seen have rarely affected viral fitness and almost never affected clinical outcome but the detailed effects of these mutations remain to be determined fully. Changes in the spike protein have relevance for potential effects on both host receptor as well as antibody binding with possible consequences for infectivity, transmission potential and antibody and vaccine escape. Actual effects need to be measured and verified experimentally.



As has become evident, these few S gene mutations and some deletions are found in multiple genomic contexts (different clades in different countries) that may be an early indication for some potential advantage for these viruses but needs to be verified and does not necessarily mean change in clinical severity or transmission efficiency. [> read more](#)



Genomic epidemiology of hCoV-19



hCoV-19 data sharing via GISAID

1,225,793
submissions

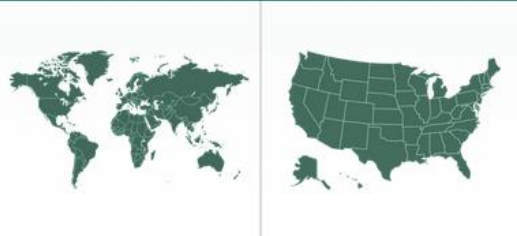
Enabled by data shared via GISAID

SEARCH AND SIMILARITY QUERY

Status of Detection Systems

University of Turin (Italy)
Diagnostic detection of 2019-nCoV by real-time

hCoV-19 Submission Tracking



hCoV-19 Tracking of Variants



Public-Private Partnerships of the GISAID Initiative

The GISAID Initiative involves public-private-partnerships between the Initiative's administrative arm Freunde of GISAID e.V., a registered non-profit association, and governments of the [Federal Republic of Germany](#), the official host of the GISAID platform, [Singapore](#) and the [United States of America](#), with support from private and corporate philanthropy.



hosted by the
Federal Republic
of Germany

GISAID Resources



Free Access Credentials
Register here and join thousands of researchers around the globe.



Frequently Asked Questions
Answers about the GISAID Initiative

Análise de Bioinformática

Atividades:

1- Na sequência de interesse, analisar, identifique/determine:

a- A sequência corresponde a um organismo procariótico ou eucariótico. Justifique.

b- Determinar a presença de exons/introns, região reguladora, região codificadora, tamanho e localização.

c- A fase de leitura aberta (ORF, open reading frame), presente na sequência.

d- Proteína(s) codificadas e determine: peso molecular, pI, composição de aminoácidos.

e- Faça uma figura mostrando:

(i) a sequência de nucleotídeos e de aminoácidos, determinada, e

(ii) a estrutura gênica prevista para o *locus* genômico que contém os genes identificados.

f- Desenhe os oligonucleotídeos (primers) correspondentes para a clonagem do gene(s), com a finalidade de obtenção e purificação da proteína recombinante. Utilize o vetor de expressão de sua preferência.

g- Faça uma figura mostrando a construção e o plasmídeo recombinante obtido a partir do item f.

Sequência de Interesse

>Problema (2320 bp)

CTTCTTTATTGGGTAATATACAGCCAGGCGGGGATGAAGCTCATTAGCCGCCACTCAAGGCTATAACAATGTTGCCAACTCTCCGGGC
TTTATCCTGTGCTCCCGAATACCACATCGTGATGATGCTTCAGCGCACGGAAGTCACAGACACCGCCTGTATAAAAGGGGGACTGT
GACCCTGTATGAGGCGCAACATGGTCTCACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCTCTGTGTATTGCACCATGATTGT
CGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGG
TAATTATGTGAACCCTCTCAAGAGACCCAAATACTGAGATATGTCAAGGGGCCAATGTGGTGGCCAGAATTGGTCTGGGTCCGACTT
GCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTGTCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGC
GCGCCGCGTCGACGACTTCTCGAGTATCCCCACAACATCCCGGTCGAGCTCCGCGACGCCTCCACCTGGTTCTACTACTACCAGA
GTACCTCCAGTCGGATCGGGAACCGCTACGTATTAGGCAACCCTTTTGTGGGGTCACTCCTTGGGCCAATGCATATTACGCCTCT
GAAGTTAGCAGCCTCGCTATTCCTAGCTTGACTGGAGCCATGGCCACTGCTGCAGCAGCTGTCGCAAAGGTTCCCTCTTTTATGTG
GCTGTAGGTCCTCCCGGAACCAAGGCAATCTGTTACTGAAGGCTCATCATTACTGCAGAGATACTTTGACAAGACCCCTCTCATG
GAGCAAACCTTGCCGACATCCGCACCGCCAACAAGAATGGCGGTAAGTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATC
GCGATTGCGCTGCCCTTGCCCTCGAATGGCGAATACTCTATTGCCGATGGTGGCGTCGCCAAATATAAGAATAATCGACACCATTCCG
TCAAATTGTCGTGGAATATTCCGATATCCGGACCCTCCTGGTTATTGGTATGAGTTTAAACACCTGCCTCCCCCCCCCTTCCCTTCT
TTCCCGCCGGCATCTTGTGCTTGTGCTAACTATTGTTCCCTCTTCCAGAGCCTGACTCTTGGCCAACTGGTGACCAACCTCGGTAC
TCCAAAGTGTGCCAATGCTCAGTCAGCCTACCTTGAGTGCATCAACTACGCCGTCACACAGCTGAACCTTCCAAATGTTGCGATGTA
TTTGACGCTGGCCATGCAGGATGGCTTGGCTGGCCGGCAAACCAAGACCCGGCCGCTCAGCTATTTGCAAATGTTTACAAGAAT
GCATCGTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTCGCCAACTACAACGGGTGGAACATTACCAGCCCCCATCGTACAC
GCAAGGCAACGCTGTCTACAACGAGAAGCTGTACATCCACGCTATTGGACCTCTTCTTGCCAATCACGGCTGGTCCAACGCCTTCTT
CATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGGTGCAATGTGATCGGCACCGGATT
TGGTATTCGCCCATCCGCAAACACTGGGGACTCGTTGCTGGATTGCTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACC
AGCGACAGCAGTGCGCCACGATTTGACTCCCACTGTGCGCTCCAGATGCCTTGCAACCGGCGCCTCAAGCTGGTGCTTGGTTCC
AAGCCTACTTTGTGACGCTTCTCACAACGCAAACCCATCGTTCCTGTAAGGCTTTTCGTGACCGGGCTTCAAACAATGATGTGCGAT
GGTGTGGTTCCCGGTTGGCGGAGTCTTTGTCTACTTTGGTTGTCTGTGCGAGGTCGGTAGACCGCAAATGAGCAACTGATGGATTG
TTGCCAGCGATACTATAATTCACATGGATGGTCTTTGTGCGATCAGTAGCTAGTGAGAGAGAGAGAACATCTATCCACAATGTCGAGT
GTCTATTAGACATACTCCGAGAATAAAGTCAACTGTGTCTGTGATCTAAAGATCGATTCCGGCAGTCGAGTAGCGTATAACAACCTCCG
AGTACCAGCAAAGCACGTCGTGACAGGAGCAGGGCTTTGCCAACTGCGCAACCTTGCTTGAATGAGGATACACGGGGTGCAAC
ATGGCTGACTGATCCATCGCAACCAAAATTTCTGTTTATAGATCAAGCTGGTAGATTCCAATTACTCCACCTCTT

a- A sequência corresponde a um organismo procariótico ou eucariótico? Justifique.

Ferramenta: Translate - ExPASy

ExPASy 🇨🇭 Translate

Programmatic access ↓

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 [Back to the top](#)

5'3' Frame 1

LLYVWIYSQAGMKLISRHSRLYNVANS PGFILCSRI PHRDDASAHGSHRHRLYKRGTVTLYEAQHGLTAAHLKRLVRSVSVYCTMIVGILTTLATLATLAASVPLEERQACSSVW-LCEPSQETQILRYVKGPMWWPELVGSDLLCFRKHMR
LLQRLLLPVSSRRCKLKLHARRVDDFSSIPHNIPVELRDASTWFYIYQSTSSRIGNRVVFRQPPFCWGHSLGQCILRL-S-QPRYS-LDWSHGHCSSCRKGSFLVYVAVGPGTKAICY-RLIIHCRDLDKTPLEQTLADIRRTANKNGN
YAGQFVVYDLDDRCAALASNGEYSIADGGVAKYKNIIDTIRQIVVEYSDIRTLVIGMSLNTCLPPFPSPFAGILSLC-LLFPLPEPDSLANLVNLTGPKCANASAYLECINAYVATQLNLPNVA MYLDAGHAGWLGWPANQDPAQOLF
ANVYKNASSPRRLGLATNVANVYNGNITSPPSYTQGNVAVNEKLYIHAIGPLLANHGWSNAFFIIDQGRSGKQPTQQQGDWCNVTGTGFGIRPSANTGDSLLDSFVWVKPGGECDDTSDSSAPRFDHSICALPDALQPAPQAGAWFQAYF
VQLLTNANPSFL-GFRDRASNNDVRWCGRSLAESLSTLVCCRSDVRK-ATDGLLPAIL-FTWVVFVDQ-LVREHRLSTMSSVY-TYSENKVCVCDLKIDSAVE-RITTPSTSKSTS-QEQGFANCATLLE-GYTGCNMAVLIHRNQFC
L-IKLVDSNYSTS

5'3' Frame 2

FFIG-YTARRG-SSLAATQGYTMLPTLRALSCAPEYHIVMMLQRTEVDTTACIKGGL-PCMRNVMVSQQLT-RGL-DHPLCIAP-LSAFSPRWLRWPHSQLVCL-RSGKLAQASGNYVNLKRPKY-DMSRGCCGGNWSGPTCCASGSTCV
YSNDYYSQCLPGAASSSSSTRAASTSRVSPPTSRSSSATPPGSTTRVPPVGGGTATYSGNPFVGVTPWANAYYASEVSSLAIPLSLTGAMATAAAA VAKVPSFMWL-VLPEPRQSVTEGSSPTAEILLTRPLSWSKPWPT SAPTRMAVT
MPDLSWCMTCRIAIALPLPRMANTLLPMVASPNIRTIISTPFVKLSWNIPISGPSWLLV-V-TPASPPPLPFPFPASCRCANYSCLFQSLTLLPTW-PTSVLQSVPMLSQPTLSASTTPSHS-TFQMLRCIWTLAMQDGLAGRQTKTRPLSYL
QMFTRMHRRLREL FADWQPMSPPTTGGTLPAPHRTRKATLSTTRCTSTLLDLFLPITAGPTPSSSLIKVDRESL PDSNSGETGAM-SAPDLVFAHPQTLGTRCWIRLSGSSQAASVTAPATAVRHDLTPTVRSQMP CNRRLKLVLSGKPTL
CSFSQTQTHRSCAFVTTGLQTMCCDGVVPGWRSCLLWLSVAGR-TANEQLMCCQRYNSHGWSLSISS--ERENIYQCRVSRIRHTPRIKSTVSVI-RSIRQSSSV-QLRVPAKARRDRSRALPTAQPLNEDTRGATWLY-SIATKISV
YRSSW-IPITPL

5'3' Frame 3

SLLGNIQPGGDEAH-PPLKAIQCCQLSGLYPVLPNTTS--CFSARKSQTPPV-KGDCDPV-GATWSHSSSPEEACKITLCVLHHDCHRSHHAGYAGHTRS-CASRGAASLLKRLVIM-TLSRDPNTEICQGANVVARIGRVLAVLPEAHAS
TPTTITPVSFPALQAQRPRAPRRRLEYPQHPGRAPRRLHLVLLPEYLOSDREPLRIQATLLGSLGPMHITPLKLAASLFLA-LEPWLLQQLSRFPLLCGCRS SRNQG NLLKAHHSLOQYS-QDP SHGANLGRH PHRQEWER-L
CRTVCGV-LAGSRLRCPCLEWRILYCRWRRQI-ELYRHHSSNCRGIFRYDPDPGYWYEFKHLPPPLPFLSRRHLVVVLTIVPSSRA-LSQPGDQPRYSKVCQCSVSLP-VHQLRRHTAEPSKCCDVFGRWPCRMAWLAGKPRPGRSAIC
KLCQECIVSESSRIGNQCRQLQRVEHYQPPIVHARQRCLOREAVHPRYWTSSCSQSRLVQRLLHH-SRSIGKAAARTATVGRVQCDRHRWIWYSPIRKHWGLVAGFVCLGQARRV-RHORQCATI-LPLCAPRCLATGASSWCLVPSLLC
AASHKRKPIVVRLS-PGFKQ-CAMVWFVPGGVVYFPGCLSQVGRPQMSN-WIVASDTI IHMDGLCRSVASERERTSIHNVECLLDILRE-SQLCL-SKDRFGSRVAYNNSEYQKHVVTGAGLCQLRNLA-MRIHGVOHGCTDPSQPKFLF
IDQAGRFQLLHD

3'5' Frame 1

KRWSNWNLPA-SINRNFCDGSGVQPCCTPCILIQARLSWQSPAPVTTFCFWYSELLYATRLPNRSLDHRHS-LYSRSMNRHSTLWIDVLSLSTADQRPSM-IIVSLATHQLLIGLPTCDRQPK-TKTPPTGNHTIAHHLKPGHES
LTGTMGLRL-EAAQSRGLTKHQLEAPVARHLGAHSGSQIVAHCCRWRHTRRLA-PRQTNPATSPQCLR MGEYQIRCRSHCTSLPTVAVR-AAFPIDLQ--RRRWT SRDQEEVQ-RGCTASRCRQRCLACTMGGW-CSTRCSWRHWLPIR
EELSETMHCKHLOIAER PGLPASQAILHGQRPNTOHLEGS AV-RRS-CTQGRLEHWHHTLEYRGWSPGWQESQALEE GTIVSTTRCRRERKGRGGGGRCLNSYQ-PGSGYRNI PRQFDEWCYSSYIWRHRHQ-SIRHSRQQRN
RDPASHTPQTVRHSYRHSWCRCRPRFAP-EGSCQEYLCSE--AFSNRLPWFREDLQPHKRGNLCDSCSSSGHSSQARNSEAA NFRVICIPRSDPNKVA-IRSGSRSDWRYSGSSRTRWRRRGARPGCCGGYSR SRRRGARGRA-AC
SAGKTLGVIVGVDCASGSTASRTRPILATTLAP-HISVFGSLERVHIITRRLSKLAAPLEAH-LRVWPA-PAW-ECRQSWCNTQRVILQASSGELL-DHVAPHTG SQSPFYTGVCDFRALKHHHLVDFVGTG-S-SPESWQHCIALSG--
ASSPPGCIPLNKE

3'5' Frame 2

RGGVIGIYQLDL-TEILVAMDQYSHVAPRVSSFKQGC AVGKALLSRRAFAGTRSCYTLDDCRIDL-IDTVDVDFILGVCLIDTRHCG-MFSLSH-LLIDKDHPCEL-YRWQOISCSFAVYR PATDNQSRQLRQPQTPPSHIIIV-SPVTKA
LQERWVCCEKHLKVGLEPSTSLRRRLQGIWERTVGVKSWRTAVAGAVTLAAWLDPDKRIQQRVSPVCGWANTKSGADHIIAPVSPLLSGRLLSRSTLISDEEGVPAVIGKRSNSVDVQLLVVDVSVALRVWAGNVPVVVGDIGCQSA
KSSRRRCILVNICK-LSGRVLVCRPAKPSCMASVQIHRNIWKVQLCDGVVDALKVG-LSIGTLWSTEVGHQVGRVRLWKREQ-LAQRDAGGKREGGGEAGV-THTNNOEGPDIGIFHDNLTNGVDIVLIFGDATIGNRVFAIRGKSAI
AIRQVIHHLKSGIVTAILVGGADVQGLLHERGLVKSISAVNDEPSVTDCLGSGRTYSHIKEGTFATAAAAAMAPVKLG IARLLTSEA-YALAQVTPTKGLPEVAVVPDPTGGTLVVVEPGGVAELDRD VVGDTRVVDAAARVDELELA
APGRHWE--SLE-THVLPEAQVGPDPQFWPPHPLDISQYLGLLRGFT-LPDA-ASLPLL-RHTSCECGQRSQRGENADNHGAIHRG-SYKPLQVSCCETMLRLIQGHSPPFIQAVSVTSVR-SIITMWSGAQDKARRVGNIV-P-VAANE
LHPRLAVYYPICK

3'5' Frame 3

EVE-LESTSLIYKQFWRWISTAMLHPVYPHSSKVAQLAKPCSCHDVLVLLVGVVIRYSTAESIFRSQTQLTTFSEYV--TLDIVDRCSLSTSY-SKTIHVNYSIAGNPNVAHLRSTDLRQTKVKDKDSANREPHHRTSLFEARSRP
YRNDGFAFVRSCTK-AWNQAPA-GAGCKASGSAQWESNRGALLSLVPSHSPGLTQTNESNESPVFADGRI PNPVPI TLHQSPHCCCPVGC FDRP-SVMKKALDQP-LARRGPIAWMYSFSL-TALPCVYDGGGLV MFHPL-LATLVANPR
RALGDDAFL-TFANS-AAGSWFAGQPSHPAWPASKYIATFGRFSCVTA-LMHSR-AD-ALAHFGVPRLVTRLARESGSGRGNNS-HNDKMPAGKEGKGGGRQVFKLIPITRRVRISEYSTTI-RMVISI-FLYLATPPSAIEYSPFEARAAQS
RSGKSYTNCPA-LPPFLAVRMSARVCSMRGVL SRVSLQ-MMSLQ-QIALVPGGPTAT-KREPLRQLLQQWPWLQSS-E-RGC-LQRRNMHWPK E-PQOKGCLNT-RFP IRLVW--NQVEASRSTGMLW GILERSSTRRAWTSLSLQ
RREDTGSNSRWSRRMCFRKHKSQDPTNSGHHIGPLTYLSIWVS-EGSHNYQTLQACRSSRGTLAASVASVASVVRMPTIMVQYTEGDLTSLFR-AAVR PCCASVYRVTPLLYRCL-LPCA EASSRCGIREHRKPGELATLYSLEWRIMS
FIPAWLYITQ-R

Resposta:

b- Determinar presença de exons/introns, região reguladora, região codificadora, tamanho e localização.

Ferramenta: BLAST - NCBI

The screenshot shows the NCBI BLAST website. At the top left is the BLAST logo, and at the top right are navigation links: Home, Recent Results, Saved Strategies, and Help. The main content area is divided into several sections:

- Basic Local Alignment Search Tool:** A descriptive paragraph explaining that BLAST finds regions of similarity between biological sequences by comparing nucleotide or protein sequences to sequence databases and calculating statistical significance. A "Learn more" link is provided.
- NEWS:** A news box with a vertical "NEWS" label on the left. The text reads: "A new feature was added to Primer-BLAST. We now offer the ability for user to run primer-blast from NCBI assembly page..". The date and time are "Tue, 23 Feb 2021 12:00:00 EST", and there is a link for "More BLAST news...".
- Web BLAST:** A section with three main options:
 - Nucleotide BLAST:** Represented by a green box with a DNA double helix icon. The text below the icon says "nucleotide ► nucleotide".
 - blastx:** Represented by a blue arrow pointing right. The text inside says "translated nucleotide ► protein".
 - tblastn:** Represented by a blue arrow pointing left. The text inside says "protein ► translated nucleotide".
 - Protein BLAST:** Represented by a blue box with a protein ribbon icon. The text below the icon says "protein ► protein".
- BLAST Genomes:** A section with a search input field containing the placeholder text "Enter organism common name, scientific name, or tax id" and a "Search" button. Below the input field are four buttons: "Human", "Mouse", "Rat", and "Microbes".

[← Edit Search](#)[Save Search](#)[Search Summary ▾](#)[? How to read this report?](#)[▶ BLAST Help Videos](#)[↶ Back to Traditional Results Page](#)

Job Title	1
RID	3T2J2RAV016 <small>Search expires on 03-25 23:12 pm</small> Download All ▾
Program	BLASTX ? Citation ▾
Database	nr See details ▾
Query ID	lcl Query_770492
Description	1
Molecule type	dna
Query Length	2320
Other reports	?

Filter Results

Organism only top 20 will appear exclude[+ Add organism](#)

Percent Identity

E value

Query Coverage

 to to to [Filter](#)[Reset](#)

Descriptions

[Graphic Summary](#)[Alignments](#)[Taxonomy](#)

Sequences producing significant alignments

[Download](#) ▾[Select columns](#) ▾

Show

100 ▾

[?](#) select all 100 sequences selected[GenPept](#)[Graphics](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	glycoside hydrolase family 6 [Trichoderma reesei QM6a]	Trichoderma reesei QM6a	591	740	62%	0.0	91.20%	471	XP_006962580.1
<input checked="" type="checkbox"/>	Chain A_CELLOBIOHYDROLASE II CORE PROTEIN [Trichoderma reesei]	Trichoderma reesei	591	677	51%	0.0	91.20%	365	3CBH_A
<input checked="" type="checkbox"/>	Chain A_CELLOBIOHYDROLASE CEL6A (FORMERLY CALLED CBH II) [Trichoderma reesei]	Trichoderma reesei	591	674	51%	0.0	91.20%	363	1QK0_A
<input checked="" type="checkbox"/>	cellobiohydrolase II [synthetic construct]	synthetic construct	590	738	61%	0.0	91.20%	447	AER26911.1
<input checked="" type="checkbox"/>	cellbiohydrolase II [Trichoderma reesei]	Trichoderma reesei	590	739	62%	0.0	90.91%	471	ADC83999.1
<input checked="" type="checkbox"/>	Chain A_CELLOBIOHYDROLASE II [Trichoderma reesei]	Trichoderma reesei	589	675	51%	0.0	90.91%	365	1CB2_A
<input checked="" type="checkbox"/>	hypothetical protein TgHK011_002311 [Trichoderma gracile]	Trichoderma gracile	589	773	65%	0.0	91.20%	471	KAH0490860.1

glycoside hydrolase family 6 [Trichoderma reesei QM6a]

NCBI Reference Sequence: XP_006962580.1

[Download Protein](#) [FASTA](#) [Align/BLAST](#)[Go to:](#) [⌕](#)

LOCUS XP_006962580 471 aa linear PLN 05-FEB-2020
 DEFINITION glycoside hydrolase family 6 [Trichoderma reesei QM6a].
 ACCESSION XP_006962580
 VERSION XP_006962580.1
 DBLINK BioProject: [PRJNA225530](#)
 BioSample: [SAMN02746107](#)
 DBSOURCE REFSEQ: accession [XM_006962518.1](#)
 KEYWORDS RefSeq.
 SOURCE Trichoderma reesei QM6a

ORGANISM *Trichoderma reesei QM6a*
 Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;
 Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae;
 Trichoderma.
 REFERENCE 1 (residues 1 to 471)
 AUTHORS Martinez, D., Berka, R.M., Henriessat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, V., Chernikov, G., Coutinho, P.M., Cullen, D., Danchin, E.S., Grigoriev, I.V., Harris, F., Jackson, M., Kubicek, C.P., Han, C.S., Ho, I., Laxroo, L.F., de Leon, A.L., Magnuson, J.K., Merino, S., Mira, M., Nelson, B., Putnan, N., Robertse, B., Salamov, A.A., Schell, M., Terry, A., Trayer, M., Westerholm-Barvinen, A., Schoch, C.L., Yao, J., Barabote, R., Nelson, M.A., Detter, C., Bruce, D., Kuske, C.R., Xie, G., Richardson, F., Rokhsar, D.S., Lucas, S.M., Rubin, E.M., Dunn-Coleman, N., Ward, M. and Brettin, T.S.
 TITLE Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*)
 JOURNAL Nat. Biotechnol. 26 (5), 553-560 (2008)
 PUBMED [18454138](#)
 REMARK Erratum [Nat Biotechnol. 2008 Oct;26(10):1193.. Barbote, Ravi [corrected to Barabote, Ravi]]
 REFERENCE 2 (residues 1 to 471)
 CONSTRM NCBI Genome Project
 TITLE Direct Submission
 JOURNAL Submitted (01-FEB-2020) National Center for Biotechnology

FEATURES
 Location/Qualifiers
 source 1..471
 /organism="Trichoderma reesei QM6a"
 /strain="QM6a"
 /type_material="type material of Trichoderma reesei"
 /db_xref="taxon:431251"
 /chromosome="Unknown"
 Protein 1..471
 /product="glycoside hydrolase family 6"
 /calculated_mol_wt=49522
 Region 30..62
 /region_name="fCBD"
 /note="Fungal-type cellulose-binding domain: smart00236"
 /db_xref="CDD:197593"
 Region 132..434
 /region_name="Glyco_hydro_6"
 /note="Glycosyl hydrolases family 6; pfam01341"
 /db_xref="CDD:334499"
 CDS 1..471
 /locus_tag="TRIREDRAFT_72567"
 /coded_by="XM_006962518.1:53..1468"
 /note="cellulohydrolase II (Cel16A); N-terminal CBM1 module"
 /db_xref="GeneID:18488147"
 /db_xref="InterPro:IPR000254"
 /db_xref="InterPro:IPR001524"
 /db_xref="InterPro:IPR001524"

ORIGIN
 1 mivgilttla tlaclaasvp leergacsav wqggggqns gptccasgt cyvandyayq
 61 clpqaasasa straaatstr vsptrssaa atpppggett rrvpvgagta tyeqgnfvqv
 121 tpenayyas evsalaipal tgamataaaa vakvpsfmvl dtdktpime qladlirtan
 181 knngnyaqgf vvyldprdc aalaangeys iadggvakyk nydrtirqlv veyadrtll
 241 vlepdelanl vmlgtpkca naqsayleci nyavtqinlp nvamylagah agwlgvpaq
 301 dpaqlfanv ykmaaspral rqlatnvany ngwnitppp ytgqnavyne klyibaqlpl
 361 lanhgwanaf fitdqgrgk qptggqvgd wenvigtqfg lrpsantqds lldefvwwk
 421 ggeodgteds saprfdhca lpdalqpapq agawfgyfv qlitnanpef 1
 //

Change region shown ▾

Customize view ▾

Analyze this sequence ▾

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Show in Genome Data Viewer

Articles about the TRIREDRAFT_72567 gene ▾

Domain architecture divergence leads to functional divergence in bindi [J Biol Chem. 2020]

Genome sequencing and analysis of the biomass-degrading fungus [Nat Biotechnol. 2008]

See all...

Pathways for the TRIREDRAFT_72567 gene ▾

Metabolic pathways

Starch and sucrose metabolism

Reference sequence information ▾

RefSeq mRNA

See reference mRNA sequence for the TRIREDRAFT_72567 gene (XM_006962518.1).

PubMed (Weighted)

Related Structures (Summary)

Recent activity ▾

Turn Off Clear

glycoside hydrolase family 6 [Trichoderma reesei QM6a] Protein

Trichoderma reesei QM6a glycoside hydrolase family 6 (TRIREDRAFT_7) nucleotide

glycoside hydrolase family 7 [Trichoderma reesei QM6a] Protein

See more...

CDS

```

53..1468
/locus_tag="TRIREDRAFT_72567"
/note="cellobiohydrolase II (Cel6A); N-terminal CBM1
module"
/codon_start=1
/product="glycoside hydrolase family 6"
/protein_id="XP_006962580.1"
/db_xref="GeneID:18488147"
/db_xref="InterPro:IPRO00254"
/db_xref="InterPro:IPRO01524"
/db_xref="JGI:Trise2_72567"

```

Proteína

```

/translation="MIVGILTTLATLTLAASVPLEERQACSSVWQCGGQNWSGPTC
CASGSTCVYSNDYYSQLPGAASSSSSTRAASTSRVSPITSRSSSATPPPGSTTTRV
PFVGSATYSGNPFVGVTFWANAYASEVSSLAIPSLTGAMATAAAVAVKVPFVWML
DTLDKTPLEQLADIRITANKNGGNYAGQFVVVDLDRDCAALASNGEYSIADGGVAK
YKNYIDTIRQIVVEYSDIRILLVIEPDSLNLVNLGTPKCANAQSAYLECINYAVTQ
LNLFNVAMYLDAHGAGWLGWPNQDPAQLFANVYKNASSFRALRGLATNVNANYGNW
ITSPPSYTQGNVAVNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGGQQWGDWC
NVIGTGFGRPSANTGDSLLDSFVWVKPGGECGTSDSSAPRFDSDHCLPDLQPAQP
AGAWFQAYFVQLLINANPSFL"

```

ORIGIN

```

1 cacagcagct cacctgaaga ggcctgtaag atcacacctct gtgtattgca coattgattgt
61 cggcattctc accacgctgg ctacgctggc cacactogca gctagtgtgc ctctagagga
121 gggcgaagct tgcctcaagc tctggggcca atgtgggtgc cagaattggt cgggtccgac
181 ttgctgtgct tccggaagca catgctgcta ctccaacgac tattactccc agtgtctccc
241 cggcgctgca agctcaagct cgtcccagcg cgcgcgctcg acgactcttc gagtatcccc
301 cacaacatcc cggctogagct cgcgacgccc tccacctggt tctactacta ccagagtacc
361 tccagctcga tccggaaccg ctacgtatcc aggcaacct tttgttgggg tcactccttg
421 ggccaatgca tattacgcct ctgaagttag cagcctcgct attcctagct tgaactggagc
481 catggccact gctgcagcag ctgtcgcaaa ggttccctct tttatgtggc tagatactct
541 tgacaagacc cctctcatgg agcaaacctt ggcgcacatc cgcaccgcca acaagaatgg
601 cggtaactat gccggacagt ttgtgtgcta tgacttgccg gatcggcatt gcgctgacct
661 tgcctogaat ggccaatact ctattgcca tgggtgggct gccaaaata agaactatat
721 cgacaccatt cgtcaaatg tegtgaata ttccgatac cggaccctcc tggttattga
781 gcctgactct cttgccaaac tggtgaccaa cctcggctac ccaaagtgtg ccaatgctca
841 gtcagcctac cttgagtgca tcaactacgc cgtcacacag ctgaaccttc caaatgttg
901 gatgtatttg gacgctggcc atgcaggatg gcttggctgg cgggcaaac aagaccggc
961 cgctcagcta tttgcaaatg tttacaagaa tgcatcgtct ccgagagctc ttcgcggtt
1021 ggcaaccaat gtcgccaact acaacgggtg gaacattacc agccccccat cgtacacgca
1081 aggcaacgct gtctacaacg agaagctgta catccaagct attggacctc ttcttgccea
1141 tcacggctgg tccaaagcct tcttcaatca tgatcaaggt cgatcgggaa agcagcctac
1201 cggcaacgca cagtggggag actggtgcaa tgtgatcggc accggatttg gtattcgcct
1261 atccgcaaac actggggact cgttgcctga ttctgttctc tgggtcaagc caggcgggca
1321 gtgtgacggc accagcgaca gcagtgcgcc acgatttgac tcccactgtg cgtccocaga
1381 tgctcttgca cggcgctctc aagctggtgc ttggttccaa gctacttttg tgcagctct
1441 cacaaaacgca aaccatcgt tcoctgaagg cttctgtgac cgggcttcaa acaatgatg
1501 gcagtggtgt ggttcccggt tggcggagtc tttgtctact ttggttctct gtocaggtc
1561 ggtagaccgc aaatgagcaa ctgatggatt gttgccagcg atactataat tcaatggat
1621 ggtctttgtc gatcagtagc tagtgagaga gagagaacat ctatccaaa tgtcagatgt
1681 ctattagaca tactccgaga ataaagtcaa ctgtgtctgt gatct

```

//

gDNA ou cDNA???

>gDNA – sequência Problema

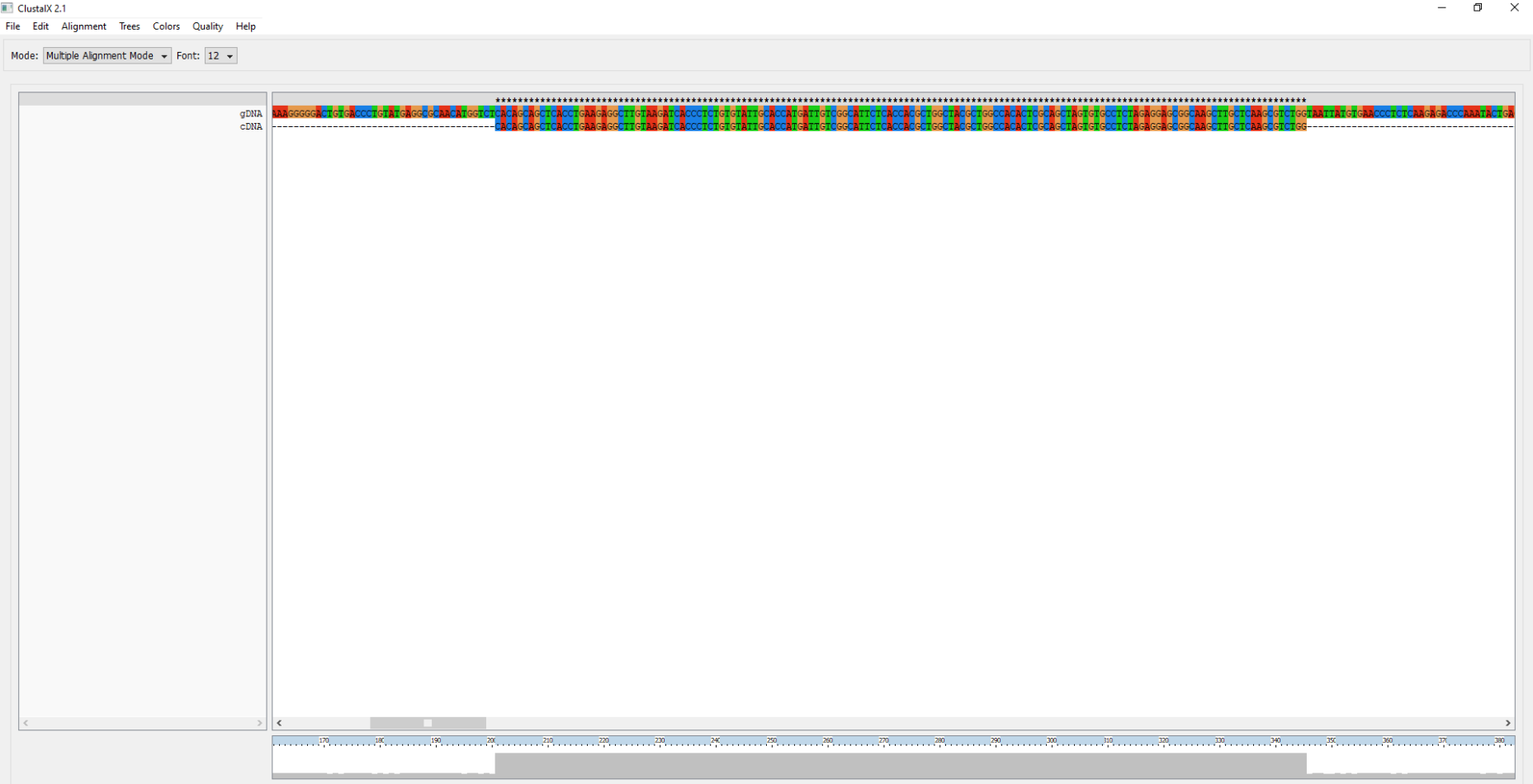
CTTCTTTATTGGGTAATATACAGCCAGGCGGGGATGAAGCTCATTAGCCGCCACTCAAGGCTATACAATGTTGCCAACTC
TCCGGGCTTTATCCTGTGCTCCCGAATACCACATCGTGATGATGCTTCAGCGCACGGAAGTCACAGACACCGCCTGTATA
AAAGGGGGACTGTGACCCTGTATGAGGCGCAACATGGTCTCACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCTCT
GTGTATTGCACCATGATTGTCGGCATTCTACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCCTCTAGAGGA
GCGGCAAGCTTGCTCAAGCGTCTGGTAATTATGTGAACCCTCTCAAGAGACCCAAATACTGAGATATGTCAAGGGGCCAA
TGTGGTGGCCAGAATTGGTCCGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCA
GTGTCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTGCACGACTTCTCGAGTATCCCCACAACATCCC
GGTCGAGCTCCGCGACGCTCCACCTGGTTCTACTACTACCAGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCA
GGCAACCCTTTTGTGGGGTCACTCCTTGGGCCAATGCATATTACGCCTCTGAAGTTAGCAGCCTCGCTATTCTAGCTT
GACTGGAGCCATGGCCACTGCTGCAGCAGCTGTCGCAAAGGTTCCCTCTTTTATGTGGCTGTAGGTCCTCCCGGAACCAA
GGCAATCTGTTACTGAAGGCTCATCATTCACTGCAGAGATACTCTTGACAAGACCCCTCTCATGGAGCAAACCTTGGCCG
ACATCCGCACCCGCAACAAGAATGGCGGTAAGTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATCGCGATTGCGCT
GCCCTTGCCTCGAATGGCGAATACTCTATTGCCGATGGTGGCGTCGCCAAATATAAGAACTATATCGACACCATTTCGTCA
AATTGTCGTGGAATATTCCGATATCCGGACCCTCCTGGTTATTGGTATGAGTTTAAACACCTGCCTCCCCCCCCCTTCC
CTTCCTTCCCGCCGGCATCTTGTGTTGTGCTAACTATTGTTCCCTCTCCAGAGCCTGACTCTCTTGCCAACCTGGTG
ACCAACCTCGGTACTCCAAAGTGTGCCAATGCTCAGTCAGCCTACCTTGAGTGCATCAACTACGCCGTACACAGCTGAA
CCTTCAAATGTTGCGATGTATTTGGACGCTGGCCATGCAGGATGGCTTGGCTGGCCGGCAAACCAAGACCCGGCCGCTC
AGCTATTTGCAAATGTTTACAAGAATGCATCGTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTCGCCAACTACAAC
GGGTGGAACATTACCAGCCCCCATCGTACACGCAAGGCAACGCTGTCTACAACGAGAAGCTGTACATCCACGCTATTGG
ACCTCTTCTGCCAATCACGGCTGGTCCAACGCCTTCTTCATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGAC
AGCAACAGTGGGGAGACTGGTGCATGTGATCGGCACCGGATTTGGTATTCGCCATCCGCAAACACTGGGGACTCGTTG
CTGGATTCGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACCAGCGACAGCAGTGCGCCACGATTTGACTCCCA
CTGTGCGCTCCAGATGCCTTGCAACCGGCGCTCAAGCTGGTGTGGTTCCAAGCCTACTTTGTGCAGCTTCTCACAA
ACGCAAACCCATCGTTCCTGTAAGGCTTTCGTGACCGGGCTTCAAACAATGATGTGCGATGGTGTGGTTCCCGTTGGCG
GAGTCTTTGTCTACTTTGGTTGTCTGTGCGCAGGTCGGTAGACCGCAAATGAGCAACTGATGGATTGTTGCCAGCGATACT
ATAATTCACATGGATGGTCTTTGTGATCAGTAGCTAGTGAGAGAGAGAGAACATCTATCCACAATGTCGAGTGTCTATT
AGACATACTCCGAGAATAAAGTCAACTGTGTCTGTGATCTAAAGATCGATTTCGGCAGTCGAGTAGCGTATAACAACCTCCG
AGTACCAGCAAAAGCACGTCGTGACAGGAGCAGGGCTTTGCCAACTGCGCAACCTTGCTTGAATGAGGATACACGGGGTG
CAACATGGCTGTAAGTATCCATCGCAACCAAATTTCTGTTTATAGATCAAGCTGGTAGATTCCAATTACTCCACCTCTT

>cDNA

CACAGCAGCT CACCTGAAGA GGCTTGTAAG ATCACCTCT GTGTATTGCA CCATGATTGTCGGCATTCTC ACCACGCTGG CTACGCTGGC
CACACTCGCA GCTAGTGTGC CTCTAGAGGAGCGGCAAGCT TGCTCAAGCG TCTGGGGCCA ATGTGGTGGC CAGAATTGGT CGGGTCCGAC
TTGCTGTGCT TCCGGAAGCA CATGCGTCTA CTCCAACGAC TACTACTCCC AGTGTCTTCCCGGCGTGCA AGCTCAAGCT CGTCCACGCG
CGCCGCGTCG ACGACTTCTC GAGTATCCCCACAACATCC CGGTCGAGCT CCGCGACGCC TCCACCTGGT TCTACTACTA CCAGAGTACC
TCCAGTCGGA TCGGGAACCG CTACGTATTC AGGCAACCCT TTTGTTGGGG TCACTCCTTGGGCCAATGCA TATTACGCCT CTGAAGTTAG
CAGCCTCGCT ATTCCTAGCT TGA CTGGAGCCATGGCCACT GCTGCAGCAG CTGTCGCAA GGTTCCTCT TTTATGTGGC TAGATACTCT
TGACAAGACC CCTCTCATGG AGCAAACCTT GGCCGACATC CGCACCGCCA ACAAGAATGGCGGTA ACTAT GCCGGACAGT TTGTGGTGTA
TGACTTGCCG GATCGCGATT GCGCTGCCCTTGCCCTCGAAT GGCGAATACT CTATTGCCGA TGGTGGCGTC GCCAAATATA AGAACTATAT
CGACACCATT CGTCAAATTG TCGTGAATA TTCCGATATC CGGACCCTCC TGGTTATTGAGCCTGACTCT CTTGCCAACC TGGTGACCAA
CCTCGGTACT CCAAAGTGTG CCAATGCTCAGTCAGCCTAC CTTGAGTGCA TCAACTACGC CGTCACACAG CTGAACCTTC CAAATGTTGC
GATGTATTTG GACGCTGGCC ATGCAGGATG GCTTGGCTGG CCGGCAAACC AAGACCCGGCCGCTCAGCTA TTTGCAAATG TTTACAAGAA
TGCATCGTCT CCGAGAGCTC TTCGCGGATTGGCAACCAAT GTCGCCA ACT ACAACGGGTG GAACATTACC AGCCCCCAT CGTACACGCA
AGGCAACGCT GTCTACAACG AGAAGCTGTA CATCCACGCT ATTGGACCTC TTCTTGCCAATCACGGCTGG TCCAACGCCT TCTTCATCAC
TGATCAAGGT CGATCGGGAA AGCAGCCTACCGGACAGCAA CAGTGGGGAG ACTGGTGCAA TGTGATCGGC ACCGGATTG GTATTCGCCC
ATCCGCAAAC ACTGGGGACT CGTTGCTGGA TTCGTTTGTG TGGGTCAAGC CAGGCGGCGAGTGTGACGGC ACCAGCGACA GCAGTGCGCC
ACGATTTGAC TCCCACTGTG CGCTCCAGATGCCTTGCAA CCGGCGCCTC AAGCTGGTGC TTGGTTCAA GCCTACTTTG TGCAGTTCT
CACAAACGCA AACCCATCGT TCCTGTAAGG CTTTCGTGAC CGGGCTTCAA ACAATGATGTGCGATGGTGT GGTTCCTGGT TGGCGGAGTC
TTTGTCTACT TTGGTTGTCT GTCGCAGGTCGGTAGACCGC AAATGAGCAA CTGATGGATT GTTGCCAGCG ATACTATAAT TCACATGGAT
GGTCTTTGTC GATCAGTAGC TAGTGAGAGA GAGAGAACAT CTATCCAAA TGTCGAGTGTCTATTAGACA TACTCCGAGA ATAAAGTCAA
CTGTGTCTGT GATCT

Alinhamento de sequências de DNA – gDNA x cDNA

Ferramenta: Clustal X



Visualização do Alinhamento de seqüências de DNA

Ferramenta: GeneDoc

GeneDoc - [Seqs]

File Project Edit Arrange Shade Groups Score Tree Reports Plot Window Help

C S G [Icons]

C Q P E S H I L D [Icons] M U

```
gDNA : * 120 * 140 * 160 * 180 * 200
       CCGAATACCACATCGTGTGATGCTTCAGCGGCACGGAAGTACACAGACACCGCCTGTATAAAAGGGGACTGTGACCCTGTATGAGCGGCAACATGGTCT : 200
cDNA : ----- : -

gDNA : * 220 * 240 * 260 * 280 * 300
       CACAGCAGCTCACTGAAGAGGCTTGAAGATCACCTCTGTGTATTGCACCATGATTTGTGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCA : 300
cDNA : CACAGCAGCTCACTGAAGAGGCTTGAAGATCACCTCTGTGTATTGCACCATGATTTGTGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCA : 100

gDNA : * 320 * 340 * 360 * 380 * 400
       GCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCCTCTGGTAATTATGTGAACCTCTCAAGAGACCCAAATACTGAGATATGTCAAGGGGCCAA : 400
cDNA : GCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCCTCTGG-----BGCCAA : 151

gDNA : * 420 * 440 * 460 * 480 * 500
       TGTGGTGGCCAGAATTGGTCGGGTCGGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTGTCTTCCCGGCGCTGCAA : 500
cDNA : TGTGGTGGCCAGAATTGGTCGGGTCGGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTGTCTTCCCGGCGCTGCAA : 251

gDNA : * 520 * 540 * 560 * 580 * 600
       GCTCAAGCTCGTCCACGGCGCGCGCTCGACGACTTCTCGAGTATCCCCACAACATCCCGGTGGAGCTCCGCGACGCTCCACCTGGTTCTACTACTAC : 600
cDNA : GCTCAAGCTCGTCCACGGCGCGCGCTCGACGACTTCTCGAGTATCCCCACAACATCCCGGTGGAGCTCCGCGACGCTCCACCTGGTTCTACTACTAC : 351

gDNA : * 620 * 640 * 660 * 680 * 700
       CAGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCCAACCTTTTGTGGGGTCACTCCTTGGGCCAATGCATATTACGCCCTCTGAAGTTAGC : 700
cDNA : CAGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCCAACCTTTTGTGGGGTCACTCCTTGGGCCAATGCATATTACGCCCTCTGAAGTTAGC : 451

gDNA : * 720 * 740 * 760 * 780 * 800
       AGCCTCGCTATTCTAGCTTGACTGGAGCCATGCCACTGCTGCAGCAGCTGTGCGAAAGGTTCCCTCTTTATGTGGCTGTAGGTCCCTCCCGGAACCAA : 800
cDNA : AGCCTCGCTATTCTAGCTTGACTGGAGCCATGCCACTGCTGCAGCAGCTGTGCGAAAGGTTCCCTCTTTATGTGGCT----- : 531

gDNA : * 820 * 840 * 860 * 880 * 900
       GGCAATCTGTACTGAAGGCTCATCATTCACTGCAGAGATACTCTTGACAAGACCCCTCTCATGGAGCAAACCTTGGCCGACATCCGCACCGCCAACAAG : 900
cDNA : -----AGATACTCTTGACAAGACCCCTCTCATGGAGCAAACCTTGGCCGACATCCGCACCGCCAACAAG : 595

gDNA : * 920 * 940 * 960 * 980 * 1000
       AATGGCGGTAACATATGCCGACAGTTTGTGGTGTATGACTTGCCGGATCCGGATTGCGCTGCCCTTGCCCTCGAATGGCGGAATACTCTATGCCGATGGTG : 1000
cDNA : AATGGCGGTAACATATGCCGACAGTTTGTGGTGTATGACTTGCCGGATCCGGATTGCGCTGCCCTTGCCCTCGAATGGCGGAATACTCTATGCCGATGGTG : 695

gDNA : * 1020 * 1040 * 1060 * 1080 * 1100
       GCGTCGCCAAATATAAGAACTATATCGACACCAATTCGTCAAATTGTCGTGGAATATCCGATATCCGGACCCCTCCTGGTTATTGGTATGAGTTTAAACAC : 1100
cDNA : GCGTCGCCAAATATAAGAACTATATCGACACCAATTCGTCAAATTGTCGTGGAATATCCGATATCCGGACCCCTCCTGGTTATTG----- : 779
```

>File >Export >-All sequences – Clustal (aln)

CLUSTAL W(1.60) multiple sequence alignment

gDNA CTTCTTTATTGGGTAATATACAGCCAGGCGGGGATGAAGCTCATTAGCCGCCACTCAAGG
cDNA -----

gDNA CTATACAATGTTGCCAACTCTCCGGGCTTTATCCTGTGCTCCCGAATACCACATCGTGAT
cDNA -----

gDNA GATGCTTCAGCGCACGGAAGTCACAGACACCGCCTGTATAAAAGGGGGACTGTGACCCTG
cDNA -----

gDNA TATGAGGCGCAACATGGTCTCACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCCCTCT
cDNA -----CACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCCCTCT

gDNA GTGTATTGCACCATGATTGTTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCA
cDNA GTGTATTGCACCATGATTGTTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCA

gDNA GCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGGTAATTATGTGAACCC
cDNA GCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGG-----




gDNA TCTCAAGAGACCCAAATACTGAGATATGTCAAGGGGCCAATGTGGTGGCCAGAATTGGTC
cDNA -----GGCCAATGTGGTGGCCAGAATTGGTC

gDNA GGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCA
cDNA GGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCA

gDNA GTGTCTTCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTTCGACGACTTCTCG
cDNA GTGTCTTCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTTCGACGACTTCTCG

c- A fase de leitura aberta (ORF, open reading frame), presente na sequência.

Ferramenta: Translate - Expasy

caca	gca	gct	cac	ctg	aag	agg	ctt	gta	aga	tca	ccc	tct	gtg	tat	tgc	acc	atg	att	gtc
T	A	A	H	L	K	R	L	V	R	S	P	S	V	Y	C	T	M	I	V
ggc	att	ctc	acc	acg	ctg	gct	acg	ctg	gcc	aca	ctc	gca	gct	agt	gtg	cct	cta	gag	gag
G	I	L	T	T	L	A	T	L	A	T	L	A	A	S	V	P	L	E	E
cgg	caa	gct	tgc	tca	agc	gtc	tgg	ggc	caa	tgt	ggg	ggc	cag	aat	tgg	tcg	ggt	ccg	act
R	Q	A	C	S	S	V	W	G	Q	C	G	G	Q	N	W	S	G	P	T
tgc	tgt	gct	tcc	gga	agc	aca	tgc	gtc	tac	tcc	aac	gac	tat	tac	tcc	cag	tgt	ctt	ccc
C	C	A	S	G	S	T	C	V	Y	S	N	D	Y	Y	S	Q	C	L	P
ggc	gct	gca	agc	tca	agc	tcg	tcc	acg	cgc	gcc	gcg	tcg	acg	act	tct	cga	gta	tcc	ccc
G	A	A	S	S	S	S	S	T	R	A	A	S	T	T	S	R	V	S	P
aca	aca	tcc	cgg	tcg	agc	tcc	gcg	acg	cct	cca	cct	ggt	tct	act	act	acc	aga	gta	cct
T	T	S	R	S	S	S	A	T	P	P	P	G	S	T	T	T	R	V	P
cca	gtc	gga	tcg	gga	acc	gct	acg	tat	tca	ggc	aac	cct	ttt	ggt	ggg	gtc	act	cct	tgg
P	V	G	S	G	T	A	T	Y	S	G	N	P	F	V	G	V	T	P	W
gcc	aat	gca	tat	tac	gcc	tct	gaa	ggt	agc	agc	ctc	gct	att	cct	agc	ttg	act	gga	gcc
A	N	A	Y	Y	A	S	E	V	S	S	L	A	I	P	S	L	T	G	A
atg	gcc	act	gct	gca	gca	gct	gtc	gca	aag	ggt	ccc	tct	ttt	atg	tgg	cta	gat	act	ctt
M	A	T	A	A	A	A	V	A	K	V	P	S	F	M	W	L	D	T	L
gac	aag	acc	cct	ctc	atg	gag	caa	acc	ttg	gcc	gac	atc	cgc	acc	gcc	aac	aag	aat	ggc
D	K	T	P	L	M	E	Q	T	L	A	D	I	R	T	A	N	K	N	G
ggt	aac	tat	gcc	gga	cag	ttt	gtg	gtg	tat	gac	ttg	ccg	gat	cgc	gat	tgc	gct	gcc	ctt
G	N	Y	A	G	Q	F	V	V	Y	D	L	P	D	R	D	C	A	A	L
gcc	tcg	aat	ggc	gaa	tac	tct	att	gcc	gat	ggt	ggc	gtc	gcc	aaa	tat	aag	aac	tat	atc
A	S	N	G	E	Y	S	I	A	D	G	G	V	A	K	Y	K	N	Y	I
gac	acc	att	cgt	caa	att	gtc	gtg	gaa	tat	tcc	gat	atc	cgg	acc	ctc	ctg	ggt	att	gag
D	T	I	R	Q	I	V	E	Y	S	D	I	R	T	L	L	V	I	E	
cct	gac	tct	ctt	gcc	aac	ctg	gtg	acc	aac	ctc	ggt	act	cca	aag	tgt	gcc	aat	gct	cag
P	D	S	L	A	N	L	V	T	N	L	G	T	P	K	C	A	N	A	Q
tca	gcc	tac	ctt	gag	tgc	atc	aac	tac	gcc	gtc	aca	cag	ctg	aac	ctt	cca	aat	ggt	gcg
S	A	Y	L	E	C	I	N	Y	A	V	T	Q	L	N	L	P	N	V	A
atg	tat	ttg	gac	gct	ggc	cat	gca	gga	tgg	ctt	ggc	tgg	ccg	gca	aac	caa	gac	ccg	gcc
M	Y	L	D	A	G	H	A	G	W	L	G	W	P	A	N	Q	D	P	A
gct	cag	cta	ttt	gca	aat	ggt	tac	aag	aat	gca	tcg	tct	ccg	aga	gct	ctt	cgc	gga	ttg
A	Q	L	F	A	N	V	Y	K	N	A	S	S	P	R	A	L	R	G	L
gca	acc	aat	gtc	gcc	aac	tac	aac	ggg	tgg	aac	att	acc	agc	ccc	cca	tcg	tac	acg	caa
A	T	N	V	A	N	Y	N	G	W	N	I	T	S	P	P	S	Y	T	Q
ggc	aac	gct	gtc	tac	aac	gag	aag	ctg	tac	atc	cac	gct	att	gga	cct	ctt	ctt	gcc	aat
G	N	A	V	Y	N	E	K	L	Y	I	H	A	I	G	P	L	L	A	N
cac	ggc	tgg	tcc	aac	gcc	ttc	ttc	atc	act	gat	caa	ggt	cga	tcg	gga	aag	cag	cct	acc
H	G	W	S	N	A	F	F	I	T	D	Q	G	R	S	G	K	Q	P	T
gga	cag	caa	cag	tgg	gga	cag	tgg	tcg	aat	gtg	atc	ggc	acc	gga	ttt	ggt	att	cgc	cca
G	Q	Q	Q	W	G	D	W	C	N	V	I	G	T	G	F	G	I	R	P
tcc	gca	aac	act	ggg	gac	tcg	ttg	ctg	gat	tcg	ttt	gtc	tgg	gtc	aag	cca	ggc	ggc	gag
S	A	N	T	G	D	S	L	L	D	S	F	V	W	V	K	P	G	G	E
tgt	gac	ggc	acc	gac	agc	agt	gcg	cca	cga	ttt	gac	tcc	cac	tgt	gcg	ctc	cca	gat	
C	D	G	T	S	D	S	S	A	P	R	F	D	S	H	C	A	L	P	D
gcc	ttg	caa	ccg	gcg	cct	caa	gct	ggt	gct	tgg	ttc	caa	gcc	tac	ttt	gtg	cag	ctt	ctc
A	L	Q	P	A	P	Q	A	G	A	W	F	Q	A	Y	F	V	Q	L	L
aca	aac	gca	aac	cca	tcg	ttc	ctg	taa	ggc	ttt	cgt	gac	cgg	gct	tca	aac	aat	gat	gtg
T	N	A	N	P	S	F	L	-	G	F	R	D	R	A	S	N	N	D	V

CTTCTTTATTGGGTAATATACAGCCAGGCGGGGATGAAGCTCATTAGCCGCCACTCAAGG
CTATACAATGTTGCCAACTCTCCGGGCTTTATCCTGTGCTCCCGAATACCACATCGTGAT
GATGCTTCAGCGCACGGAAGTACAGACACCGCCTGTATAAAAAGGGGGACTGTGACCCTG
TATGAGGGCGCAACATGGTCTCACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCTCT
GTGTATTGCACCATGATTGTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCA
GCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGGTAATTATGTGAACCC
TCTCAAGAGACCCAAATACTGAGATATGTCAAGGGGCCAATGTGGTGGCCAGAATTGGTC
GGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCA
GTGTCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTGCAGGACTTCTCG
AGTATCCCCCACAACATCCCAGTCAAGCTCCGCGACGCTCCACCTGGTTCTACTACTAC
CAGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCAACCCTTTTGTGGGGT
CACTCCTTGGGCCAATGCATATTACGCCTCTGAAGTTAGCAGCCTCGCTATTCCCTAGCTT
GACTGGAGCCATGGCCACTGCTGCAGCAGCTGTGCAAAGGTTCCCTCTTTTATGTGGCT
GTAGGTCCTCCCGGAACCAAGGCAATCTGTTACTGAAGGCTCATCATTCACTGCAGAGAT
ACTCTTGACAAGACCCCTCTCATGGAGCAAACCTTGGCCGACATCCGCACCGCCAACAAG
AATGGCGGTAACACTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATCGCGATTGCGCT
GCCCTTGCCTCGAATGGCGAATACTCTATTGCCGATGGTGGCGTCGCCAAATATAAGAAC
TATATCGACACCATTGCTCAAATTGCTGGAATATTCCGATATCCGGACCCTCCTGGTT
ATTGATGAGTTTAAACACCTGCCTCCCCCCCCCTTCCCTTCCCTTCCCGCCGGC
TTGTCGTTGTGCTAACTATTGTTCCCTCTTCCAGAGCCTGACTCTCTTGCCAACCTGGTG
ACCAACCTCGGTACTCCAAAGTGTGCCAATGCTCAGTCAGCCTACCTTGAGTGCATCAAC
TACGCCGTCACACAGCTGAACCTTCAAATGTTGCGATGTATTTGGACGCTGGCCATGCA
GGATGGCTTGGCTGGCCGGCAAACCAAGACCCGGCCGCTCAGCTATTTGCAAATGTTTAC
AAGAATGCATCGTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTGCGCAACTACAAC
GGGTGGAACATTACCAGCCCCCATCGTACACGCAAGGCAACGCTGTCTACAACGAGAAG
CTGTACATCCACGCTATTGGACCTCTTCTTGCCAATCACGGCTGGTCCAACGCCTTCTTC
ATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGG
TGCAATGTGATCGGCACCGGATTTGGTATTCGCCCATCCGCAAACACTGGGGACTCGTTG
CTGGATTGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACCAGCGACAGCAGT
GCGCCACGATTTGACTCCCACTGTGCGCTCCAGATGCCTTGCAACCAGGCGCCTCAAGCT
GGTGCTTGGTTCCAAGCCTACTTTGTGTCAGCTTCTCACAAACGCAAACCCATCGTTCCCTG
TAAAGCTTTCGTGACCGGGCTTCAAACAATGATGTGCGATGGTGTGGTTCCCGGTTGGCG
GAGTCTTTGTCTACTTTGGTTGTCTGTCGAGGTCGGTAGACCGCAAATGAGCAACTGAT
GGATTGTTGCCAGCGATACTATAATTCACATGGATGGTCTTTGTGATCAGTAGCTAGTG
AGAGAGAGAGAACATCTATCCACAATGTCGAGTGTCTATTAGACATACTCCGAGAATAAA
GTCAACTGTGTCTGTGATCTAAAGATCGATTCCGGCAGTCGAGTAGCGTATAACAACCTCCG
AGTACCAGCAAAGCACGTCGTGACAGGAGCAGGGCTTTGCCAACTGCGCAAACCTTGCTT
GAATGAGGATACACGGGGTGCAACATGGCTGTACTGATCCATCGCAAACCAAATTTCTGT



d- Proteína(s) codificada(s) e determine: peso molecular, pI, composição de aminoácidos.

MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVVSYNDYYSQCLPGAASSSSSTRAASTTSRVSPPTSRSSSATPPPGSTTTTRVPPV
 GSGTATYSGNPFVGVTPWANAYASEVSSLAIPSLTGAMATAAAVAKVPSFMWLDLTKPLMEQTLADIRTANKNGGNYAGQFVVDLPDRDCAALASN
 GEYSIADGGVAKYKNYIDTIRQIVVEYSDIRLLVIEPDSLNLVTLNLGTPKCANAQSAYLEICINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAALFANV
 YKNASSPRALRGLATNVANYNGWNITSPSYTQGNVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQQWGDWCNVIGTGFIRPSANTGDSL
 LDSFVWVKPGGECGTSDSSAPRFDHICALPDALQPAPQAGAWFQAYFVQLLTNANPSFL

BLAST® » blastp suite » results for RID-89FZAEHH016 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

[← Edit Search](#) [Save Search](#) [Search Summary](#) [? How to read this report?](#) [BLAST Help Videos](#) [↶ Back to Traditional Results Page](#)

Job Title	Protein Sequence
RID	89FZAEHH016 <small>Search expires on 04-26 09:03 am</small> Download All ▼
Program	BLASTP ? Citation ▼
Database	nr See details ▼
Query ID	lcl Query_34109
Description	None
Molecule type	amino acid
Query Length	471
Other reports	Distance tree of results Multiple alignment MSA viewer ?

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to **E value** to **Query Coverage** to

[Filter](#) [Reset](#)

Descriptions | [Graphic Summary](#) | [Alignments](#) | [Taxonomy](#)

Sequences producing significant alignments [Download](#) New [Select columns](#) ▼ Show [?](#)

select all 100 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) New [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cov.	E value	Per. Ident.	Acc. Len.	Accession
<input checked="" type="checkbox"/>	glycoside hydrolase family 6 [Trichoderma reesei QM6a]	Trichoderma ree...	959	959	100%	0.0	100.00%	471	XP_006962580.1
<input checked="" type="checkbox"/>	cellbiohydrolase II [Trichoderma reesei]	Trichoderma reesei	958	958	100%	0.0	99.79%	471	ADC83999.1
<input checked="" type="checkbox"/>	cellobiohydrolase II [Trichoderma reesei]	Trichoderma reesei	956	956	100%	0.0	99.79%	471	AAG39980.1
<input checked="" type="checkbox"/>	cellobiohydrolase II [Trichoderma viride]	Trichoderma viride	956	956	100%	0.0	99.79%	471	AAQ76094.1
<input checked="" type="checkbox"/>	unnamed [Trichoderma reesei]	Trichoderma reesei	954	954	100%	0.0	99.58%	471	AAA72922.1
<input checked="" type="checkbox"/>	GH6 Cellobiohydrolase CEL6A/CBH2 [Trichoderma parareesei]	Trichoderma par...	951	951	100%	0.0	98.73%	471	OTA06465.1
<input checked="" type="checkbox"/>	cellobiohydrolase II [Trichoderma longibrachiatum]	Trichoderma long...	939	939	100%	0.0	97.88%	470	ACZ34301.1

ProtParam

User-provided sequence:

Molecular weight: 49653.35

Theoretical pI: 5.11

Amino acid composition:

[CSV format](#)

Ala (A)	60	12.7%
Arg (R)	14	3.0%
Asn (N)	30	6.4%
Asp (D)	21	4.5%
Cys (C)	12	2.5%
Gln (Q)	21	4.5%
Glu (E)	10	2.1%
Gly (G)	40	8.5%
His (H)	4	0.8%
Ile (I)	17	3.6%
Leu (L)	37	7.9%
Lys (K)	10	2.1%
Met (M)	5	1.1%
Phe (F)	12	2.5%
Pro (P)	32	6.8%
Ser (S)	47	10.0%
Thr (T)	38	8.1%
Trp (W)	12	2.5%
Tyr (Y)	21	4.5%
Val (V)	28	5.9%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 31

Total number of positively charged residues (Arg + Lys): 24

Atomic composition:

Carbon	C	2194
Hydrogen	H	3354
Nitrogen	N	594
Oxygen	O	691
Sulfur	S	17

Formula: C₂₁₉₄H₃₃₅₄N₅₉₄O₆₉₁S₁₇

Total number of atoms: 6850

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient	98040
Abs 0.1% (=1 g/l)	1.974, assuming all pairs of Cys residues form cystines

Ext. coefficient	97290
Abs 0.1% (=1 g/l)	1.959, assuming all Cys residues are reduced

e- Faça uma figura mostrando:

- (i) mostrando a estrutura gênica prevista para o *locus* genômico que contém os genes identificados, e
- (ii) a Sequência de nucleotídeos e de aminoácidos, determinada

A.

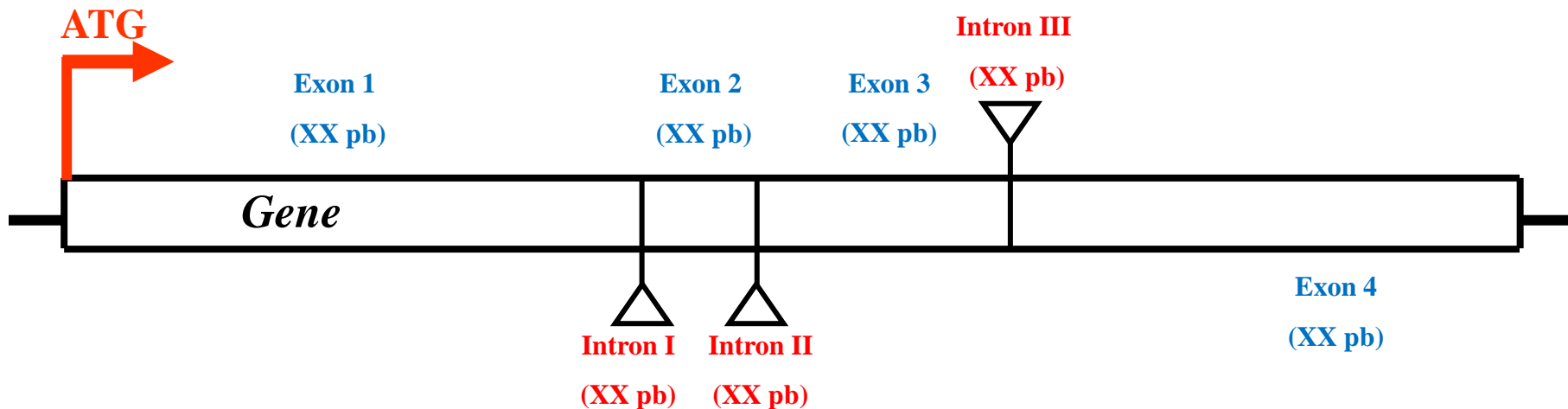


Figura X. Estrutura gênica (A) e sequência de nucleotídeos e de aminoácidos (B) prevista para o *locus* que contém o gene XXXXX do fungo filamentoso *Trichoderma reesei*. Os códons de iniciação e terminação estão indicados em negrito. Os íntrons identificados na sequência codificadora das proteínas estão representados por letras minúsculas. A numeração à esquerda corresponde à posição dos nucleotídeos em relação à adenina do códon de iniciação da proteína citrato sintase; os aminoácidos estão numerados a direita.

B.

-101 AGCATCATGCAGCTCCCATAGCTAACCGCTTCTTCTGCCCCGCCATTTCAGGAGCAGCTCGCCTTGTTTCATTTATCCTCACCTTTGCC
TCGTAGTAGTACGTCGAGGGTATCGATTGGCGAAGAAGACGGGGCGGTAAGTCTCGTCGAGCGGAACAAGTAAATAGGAGTGAAACCG

-11 CACACCAAATCATGGCTCTCAACCTCACCTCGTCGGCTCGAGCCTGCGCTCCTTCAAGgtcagtcgagtcctccttatagccttctat
GTGTGGTTTAG M A L N L I S S A R A L R S F K 16

80 cgagaccatttgacatagactgacactcctcccgtctacagCCCTACACCGCGCGCCCTCCTCGCCAACGCGCGGATGCTACTCA
I N T R O N I P Y T R A A L L A N A A R C Y S 32

170 ACCGCTGAGCCCACCTCAAGACGACGCTCAAGAGCGTCATCCCTGAGAAGCGCGAGCTGCTGAAGAAGTCAAGGCCACCGGACGCAAG
T A E P D L K T I L K S V I P E K R E L L K K V K A H G S K 62

260 GTCATTGGCGAGGTCAAGGTTGAGAACCATTGGCGGCATGCGCGGGCTCAAGGCCATGGTCTGGGAGGGCTCCGTGCTCGACCCCAAC
V I G E V K V E N T I G G M R G L K A M V W E G S V L D P N 92

350 GAGGGCATCCGCTTCCACGGCCGCACCATCAAGGACTGCCAGAAGGAGCTGCCAAGGGCAAGACGGGCACCGAGATGCTGCCGAGGGC
E G I R F H G R T I K D C Q K E L P K G K T G T E M L P E A 122

440 AIGTTTTGGCTGCTGTTGACCGGCCAGGTGCCCTCCGTCAACCAGTCCGCCAGTCTCCCGCGAGCTGGCCTCCAGACCCAGATCCCC
M F W L L L T G Q V P S V N Q V R Q F S R E L A S Q T Q I P 152

530 GCCTTCGTCAACAGGATGCTCGACGATTTCCCAAGGATCTGCACCCCATGACCCAGTTCGCCATTGCCGTCTCGGCCCTCAACTACGAG
A F V N R M L D D F P K D L H P M T Q F A I A V S A L N Y E 182

620 TCCAAGTTTGCAAAGGCCTACGAGAAGGGCCTCGCCAAGGCCGACTACTGGGAGCCACCTTTGACGACAGCATCTCGCTGCTCGCCAAG
S K F A K A Y E K G L A K A D Y W E P T F D D S I S L L A K 212

710 CTGCCCACCATGCGCCCAAGATCTACCAGAACTCTTACCGCGCGGGCGGCCCTCCCTGCCGAGGTCGACCTTGAGCAGGATTGGTCA
L P T I A A K I Y Q N S Y R G G G A L P A E V D L E Q D W S 242

800 TACAACCTTGTGCCATGCTCGGCAAGGGCGCAAGGAGAACGAGGACTTCCAGGACCTCCTCCGTCTACCTTGCCCTCCACGGCGAC
Y N F A A M L G K G G K E N E D F Q D L L R L Y L A L H G D 272

890 CACGAGGGTGGCAATGTGTCTGCCACGCCACTCACCTTGTGGTAGTGCCCTGAGTGACCCCTTCTGTCTTACAGCGCTGGTCTCCAG
H E G G N V S A H A T H L V G S A L S D P F L S Y S A G L Q 302

980 GGCTGGCCGCTCCTTACAGGgtaaagctgtccactatactataatcaatcttctatacaaaagttgattgtaaatcgttcgctgct
G L A G P L H G **I N T R O N II** 310

1070 tatgacgatagACTTGGCCGCCAGGAAGTTCCTCGCTGGATCCTGCAGATGAAGGAGGCCATCCCCGCCAACTACACCGAGCAGGATGTC
L A A Q E V L R W I L Q M K E A I P A N Y T E Q D V 336

1160 CACGACTACCTCTGGTCCACCCTCAACTCGGGCCGCGTCTGTCGGCGGCTACGGACACGCCGTTCTGCGCAAGCCCGACCCCTCGATTGAG
H D Y L W S T L N S G R V V P G Y G H A V L R K P D P R F E 366

1250 GCTCTCATGGAATATGCCGCTTCCCGCCCCGCGATTGCCAAGGACCCCGTCTTCCAGCTGGTTGAGAAGAAGCGCGCATCGCCCCGAG
A L M D Y A A S R P A I A K D P V F Q L V E K N S R I A P E 396

1340 GTGCTCAAGAAGCACGGCAAGACCAAGAACCCTACCCCAACGTCGACAGCAGCTCCGGCGTCCTCTCCACCCTACGGCTTCCACGAG
V L K K H G K T K N P Y P N V D S S S G V L F H H Y G F H E 426

1430 ACGCTCTACTACACGGCCACCTTTGGTGTTCGCGTGGTCTCGGTCCTCTGGCTCAGCTCATCTGGGACCGCGCCCTGGGCTGCCCCATT
T L Y Y T A T F G V S R G L G P L A Q L I W D R A L G L P I 456

1520 GAGCGCCCCAAGAGCATCAACCTCGAGGGTATTCTGAAGCAGGTGGAGAGCAGCTAA
E R P K S I N L E G I L K Q V E S S * 474

f- Desenhe os oligonucleotídeos correspondentes para a clonagem do gene(s), com a finalidade de obtenção e purificação da proteína recombinante. Utilize o vetor de expressão de sua preferência.

>Celobiohidrolase II – *Trichoderma reesei* (1416 pb)

```
ATGATTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCT
GGGGCCAATGTGGTGGCCAGAATTGGTCGGGTCGGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTG
TCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTCGACGACTTCTCGAGTATCCCCACAACATCCCGGTTCGAGCTCCGCG
ACGCCTCCACCTGGTTCTACTACTACCAGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCAACCCTTTTGTGGGGTCACTCC
TTGGGCCAATGCATATTACGCCTCTGAAGTTAGCAGCCTCGCTATTCCTAGCTTGACTGGAGCCATGGCCACTGCTGCAGCAGCTGTCGCAA
AGGTTCCCTCTTTTATGTGGCTAGATACTCTTGACAAGACCCCTCTCATGGAGCAAACCTTGGCCGACATCCGCACCGCCAACAAGAATGGC
GGTAACTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATCGCGATTGCGCTGCCCTTGCCCTCGAATGGCGAATACTCTATTGCCGATGG
TGCGTCCGCAAATATAAGA ACTATATCGACACCATTTCGTCAAATTGTCGTGGAATATTCCGATATCCGGACCCTCCTGGTTATTGAGCCTGA
CTCTTTGCCAACCTGGTGACCAACCTCGGTACTCCAAAGTGTGCCAATGCTCAGTCAGCCTACCTTGAGTGCATCAACTACGCCGTCACAC
AGCTGAACCTTCCAAATGTTGCGATGTATTTGGACGCTGGCCATGCAGGATGGCTTGCGTGGCCGGCAAACCAAGACCCGGCCGCTCAGC
TATTTGCAAATGTTTACAAGAATGCATCGTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTCGCCAACTACAACGGGTGGAACATTACC
AGCCCCCATCGTACACGCAAGGCAACGCTGTCTACAACGAGAAGCTGTACATCCACGCTATTGGACCTCTTCTTGCCAATCACGGCTGGTC
CAACGCCTTCTTCATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGGTGCAATGTGATCGGCAC
CGGATTTGGTATTGCCCCATCCGCAAACACTGGGGACTCGTTGCTGGATTGCTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCAC
CAGCGACAGCAGTGCGCCACGATTTGACTCCCACTGTGCGCTCCAGATGCCTTGCAACCGGCGCCTCAAGCTGGTGCTTGGTTCCAAGC
CTACTTTGTGCAGCTTCTCACAAACGCAAACCCATCGTTCTGTAA
```

- Conversão a DNA dupla fita

Ferramenta: Conversor

ATGATTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGGGGCCAAT
TACTAACAGCCGTAAGAGTGGTGCACCCGATGCGACCCGGTGTGAGCGTGCATCACACGGAGATCTCCTCGCCGTTTCAACGAGTTTCGCAGACCCCGGTTA

GTGGTGGCCAGAATTGGTCCGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTGTCTTCCCGGCGCTGCAAG
CACCACCCGGTCTTAACCAGCCCAGGCTGAACGACACGAAAGGCCCTTCGTGTACGCAGATGAGGTTGCTGATAATGAGGGTACAGAAAGGGCCGCGACGTTT

CTCAAGCTCGTCCACGCGCGCCGCGTGCAGCAGCTTCTCGAGTATCCCCACAACATCCCGGTGAGCTCCGCGACGCCTCCACCTGGTTCTACTACTACC
GAGTTCGAGCAGGTGCGCGCGGCGCAGCTGCTGAAGAGCTCATAGGGGGTGTGTAGGGCCAGCTCGAGGCGCTGCGGAGGTGGACCAAGATGATGATGG

AGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCAACCCTTTTGGTGGGGTCACTCCTTGGGCCAATGCATATTACGCCTCTGAAGTTAGCA
TCTCATGGAGGTCAGCCTAGCCCTTGGCGATGCATAAGTCCGTTGGGAAAAACAACCCAGTGAGGAACCCGGTTACGTATAATGCGGAGACTTCAATCGT

GCCTCGCTATTCTAGCTTGACTGGAGCCATGGCCACTGCTGCAGCAGCTGTGCGAAAGGTTCCCTCTTTTATGTGGCTAGATACTCTTGACAAGACCCC
CGGAGCGATAAGGATCGAACTGACCTCGGTACCGGTGACGACGTCGTGACAGCGTTCGAGGGGAGAAAATACACCGATCTATGAGAAGTGTCTGGGG

TCTCATGGAGCAAACCTTGGCCGACATCCGCACCGCCAACAAGAATGGCGGTAAGTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATCGCGATTGC
AGAGTACCTCGTTTGAACCGGCTGTAGGCGTGGCGGTTGTTCTTACCGCCATTGATACGGCCTGTCAAACACCACATACTGAACGGCCTAGCGCTAACG

GCTGCCCTTGCCTCGAATGGCGAATACTCTATTGCCGATGGTGGCGTCGCCAAATATAAGAAGTATATCGACACCATTTCGTCAAATTGTCGTGGAATATT
CGACGGGAACGGAGCTTACCCTTATGAGATAACGGCTACCACCGCAGCGGTTTATATTCTTGATATAGCTGTGGTAAGCAGTTTAAACAGCACCTTATAA

CCGATATCCGGACCCTCCTGGTTATTGAGCCTGACTCTCTTGCCAACTGGTGACCAACCTCGGTACTCCAAAGTGTGCCAATGCTCAGTCAGCCTACCT
GGCTATAGGCCTGGGAGGACCAATAAAGTCCGACTGAGAGAACGGTTGGACCCTGGTTGGAGCCATGAGGTTTACACCGTTACGAGTCAGTCGGATGGA

TGAGTGCATCAACTACGCCGTACACAGCTGAACCTTCCAAATGTTGCGATGTATTTGGACGCTGGCCATGCAGGATGGCTTGGCTGGCCGGCAAACCAA
ACTCACGTAGTTGATGCGGCAGTGTGTCGACTTGAAGGTTTACAACGCTACATAAACCTGCGACCCGGTACGTCTACCGAACCGACCGGCCGTTTGGTT

GACCCGGCCGCTCAGCTATTTGCAAATGTTTACAAGAATGCATCGTCTCCGAGAGCTCTTCCGCGGATTGGCAACCAATGTCGCCAACTACAACGGGTGGA
CTGGGCCGCGAGTCGATAAACGTTTACAATGTTCTTACGTAGCAGAGGCTCTCGAGAAGCGCCTAACCGTTGGTTACAGCGGTTGATGTTGCCACCT

ACATTACCAGCCCCCATCGTACACGCAAGGCAACGCTGTCTACAACGAGAAGCTGTACATCCACGCTATTGGACCTCTTCTTGCCAATCACGGCTGGTC
TGTAATGGTCCGGGGGTAGCATGTGCGTTCCGTTGCGACAGATGTTGCTCTTCGACATGTAGTGGCGATAACCTGGAGAAGAAGGTTAGTGCCGACCAG

CAACGCCTTCTTCATCACTGATCAAGGTGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGGTGCAATGTGATCGGCACCGGATTTGGT
GTTGCGGAAGAAGTAGTACTAGTTCCAGCTAGCCCTTTCGTGCGATGGCCTGTCGTTGTACCCCTCTGACCAGTTACTAGCCGTGGCCTAAACCA

ATTCGCCATCCGCAAACACTGGGGACTCGTTGCTGGATTGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACCAGCGACAGCAGTGCGCCAC
TAAGCGGGTAGGCGTTTGTGACCCCTGAGCAACGACCTAAGCAAACAGACCCAGTTCGGTCCGCGGCTCACACTGCCGTGGTTCGCTGTCGTCACGCGGTG

GATTTGACTCCCACTGTGCGCTCCAGATGCCTTGCAACCGGCGCCTCAAGCTGGTGTGTTGGTTCCAAGCCTACTTTGTGCGCTTCTCACAAACGCAAA
CTAAACTGAGGGTGACACGCGAGGGTCTACGGAACGTTGGCCGCGGAGTTGACACCACGAACCAAGGTTTCGGATGAAACACGTCGAAGAGTGTGCGTTT

CCCATCGTTCCTGTAA
GGGTAGCAAGGACATT

https://www.bioline.com/media/calculator/01_12.html

- Análise de enzimas de restrição

Ferramenta: RestrictionMapper

Name	Sequence	Site Length	Overhang	Frequency	Cut Positions
BsaAI	YACGTR	6	blunt	1	332
EcoRV	GATATC	6	blunt	1	705
HindII	GTYRAC	6	blunt	1	227
NaeI	GCCGGC	6	blunt	1	890
NruI	TCGCGA	6	blunt	1	593
SnaBI	TACGTA	6	blunt	1	332
SspI	AATATT	6	blunt	1	697
AvaI	CYCGRG	6	five_prime	1	236
BclI	TGATCA	6	five_prime	1	1119
BsePI	GCGCGC	6	five_prime	1	216
BspI407I	TGTACA	6	five_prime	1	1055
BstEII	GGTNACC	6	five_prime	1	750
BtgZI	GCGATG	6	five_prime	1	862
Esp3I	CGTCTC	6	five_prime	1	950
HindIII	AAGCTT	6	five_prime	1	74
NarI	GGCGCC	6	five_prime	1	1342
NcoI	CCATGG	6	five_prime	1	428
Sall	GTCGAC	6	five_prime	1	225
SapI	GCTCTTC	7	five_prime	1	962
SgrDI	CGTCGACG	8	five_prime	1	225
TatI	WGTACW	6	five_prime	1	1055
TfiI	GAWTC	5	five_prime	1	1237
XbaI	TCTAGA	6	five_prime	1	60
XhoI	CTCGAG	6	five_prime	1	236
AlwNI	CAGNNCTG	6	three_prime	1	448
BciVI	GTATCC	6	three_prime	1	252
BseRI	GAGGAG	6	three_prime	1	79
BsgI	GTGCAG	6	three_prime	1	1399
BsrDI	GCAATG	6	three_prime	1	1182
Eco57I	CTGAAG	6	three_prime	1	410

Noncutters: AarI, AatII, AbsI, AclI, AflIII, AflIII, AgeI, AgsI, AlfI, Aloi, Apal, ApaLI, Apol, ArsI, AscI, AsuII, AvrII, BamHI, BarI, BbvCI, BdaI, BglI, BglII, Bpu10I, BsaBI, BseSI, BseYI, BspHI, BspMI, BtrI, ClaI, CspCI, DraII, DraIII, DrdI, Eam1105I, EciI, Eco31I, Eco47III, EcoNI, EcoRI, Fall, Faul, FseI, FspAI, HaeIV, HpaI, KpnI, MauBI, MfeI, MluI, MslI, NdeI, NheI, NmeAIII, NotI, OliI, PacI, PaeI, PflMI, PfoI, PmaCI, PmeI, PpiI, PpuMI, PshAI, PstI, PstII, PspXI, PstII, RsrII, SacII, SanDI, Scal, PstI-Scel, SfiI, SgfI, SgrAI, SmaI, SpeI, SphI, SrfI, Sse8387I, StuI, SwaI, TaqII, TspGWI, Tth111I, VspI, XcmI, XhoII, XmnI

f- Desenhe os oligonucleotídeos correspondentes para a clonagem do gene(s), com a finalidade de **obtenção e purificação da proteína recombinante**. Utilize o **vetor de expressão** de sua preferência.

Proteína Recombinante

Isto é possível por:

- Universalidade do código genético.
- Similaridade da maquinaria de transdução (ribossomos)
- Rápido avanço das técnicas de biologia molecular e/ou engenharia genética: Clonagem de DNA e Sequenciamento de DNA, enzimas para clivagem, ligação, sínteses de moléculas de DNA, RNA, sínteses de nucleotídeos, etc.

Deve-se obter plasmídeos para clonagem que garantissem um alto nível de expressão da proteína de interesse (Vetor de Expressão)

1- Clonagem do cDNA da proteína de interesse num vetor de expressão

- Seleção do vetor e proteína a expressar

2- Transformação e seleção de Bactérias ou células competentes

- Seleção de organismo hospedeiro e técnica de transformação

3- Testes de Expressão

- Identifique o clone que expressa a proteína recombinante recombinante

4- Produção em larga escala da proteína recombinante

- Empregando um grande volumem de cultura

5- Recuperação e análises da proteína recombinante

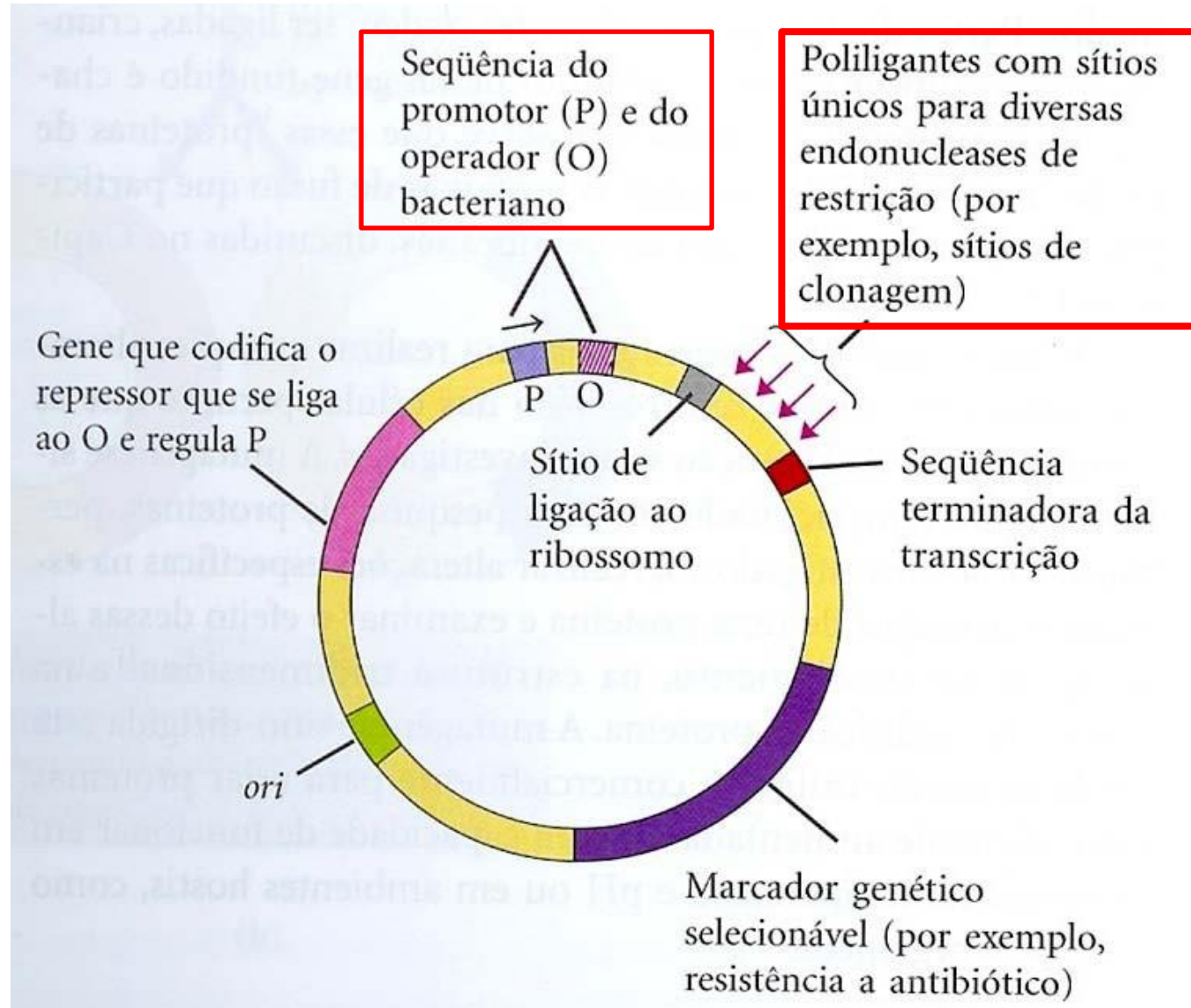
- Técnicas para recuperar a proteína a partir da cultura

6- Purificação

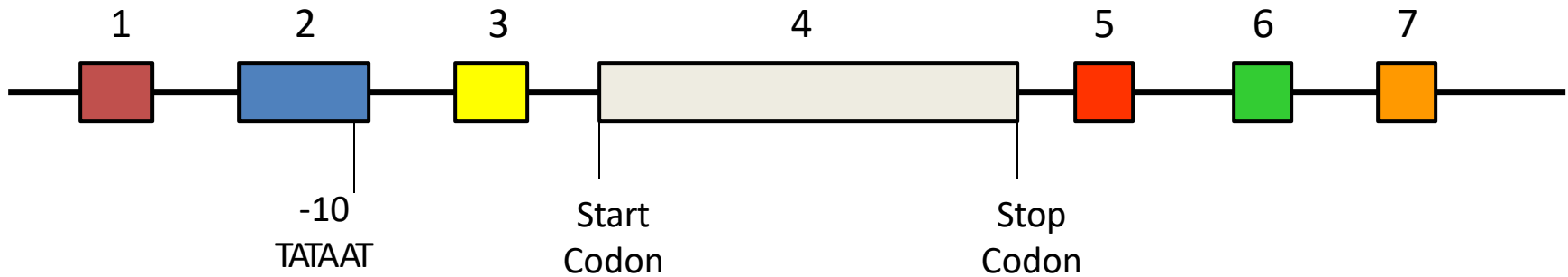
- Técnicas de purificação de proteínas

7- Aplicação

Vetor de Expressão

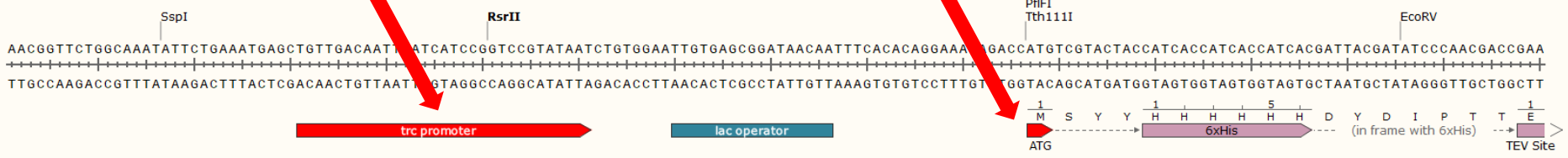
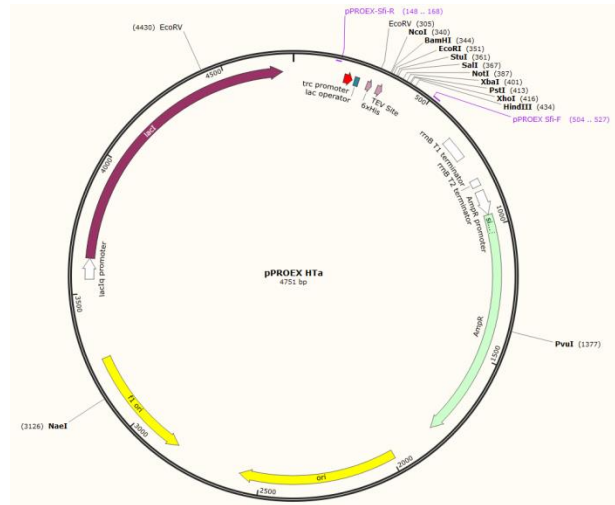


Elementos de um vetor de expressão procariótico



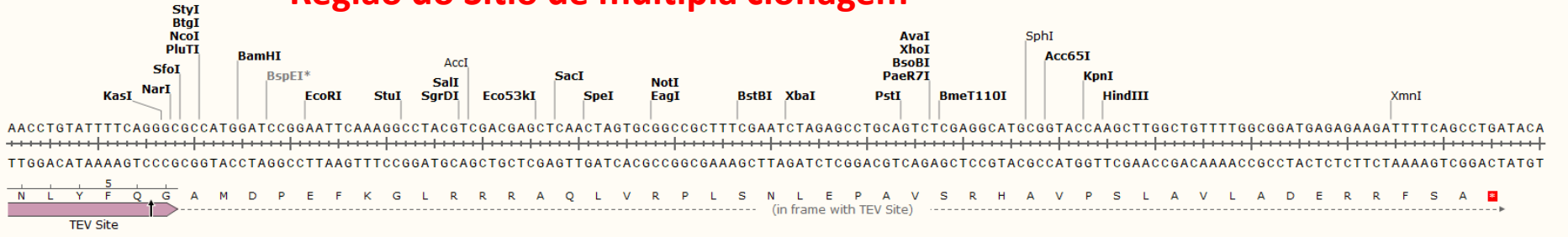
- 1- **Regulador** do promotor: Proteína que modula o promotor
- 2- **Promotor**: Deve ser forte (lac, trp, tac, λp^L , gene 10 do fago T7)
- 3- **Sequência Shine-dalgarno**: Sítio de ligação do ribossomo, (RBS).
- 4- **Região codificadora**: sítios de múltipla clonagem
- 5- **Terminador** de transcrição: Estabiliza o mRNA
- 6- **Marcador genético** (antibiótico de seleção)
- 7- **Ori**: Origem de replicação do plasmídeo.

Vetor de expressão pPROEX-Hta



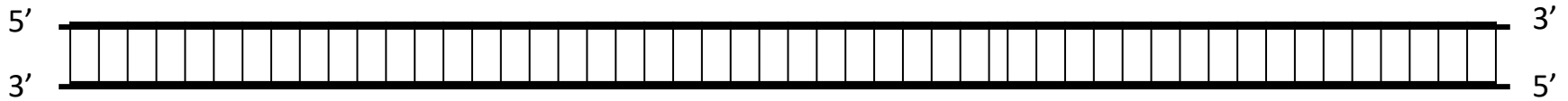
TTGCCAAG
pPROEX-Sfi-R

Região do Sítio de múltipla clonagem



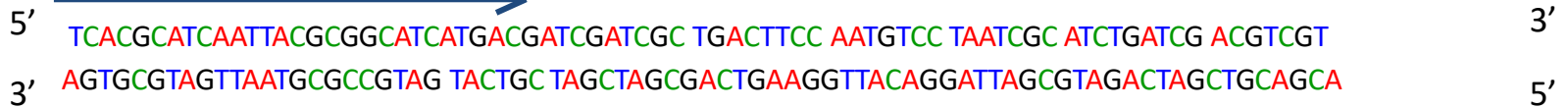
Noncutters: AarI, AatII, AbsI, AclI, AflIII, AflIII, AgeI, AgsI, AlfI, Alol, Apal, ApaLI, Apol, Arsl, AscI, AsuII, AvrII, **BamHI**, BarI, BbvCI, BdaI, BglI, BglII, Bpu10I, BsaBI, BseSI, BseYI, BspHI, BspMI, BtrI, ClaI, CspCI, DraII, DraIII, DrdI, Eam1105I, EciI, Eco31I, Eco47III, EcoNI, **EcoRI**, Fall, Faul, FseI, FspAI, HaeIV, HpaI, **KpnI**, MauBI, MfeI, MluI, MslII, NdeI, NheI, NmeAIII, NotI, OsiI, PacI, PasI, PflMI, PfoI, PmaCI, PmeI, Ppil, PpuMI, PshAI, PstI, PspI, PI-PspI, PspXI, PstI, RsrII, SacII, SanDI, Scal, PI-Scel, SfiI, SgfI, SgrAI, SmaI, SpeI, SphI, SrfI, Sse8387I, StuI, SwaI, TaqII, TspGWI, Tth111I, VspI, XcmI, XhoII, XmnI

DNA e desenho de Oligonucleotídeos/primers



Forward (5')

TCACGCATCAATTACGCGG

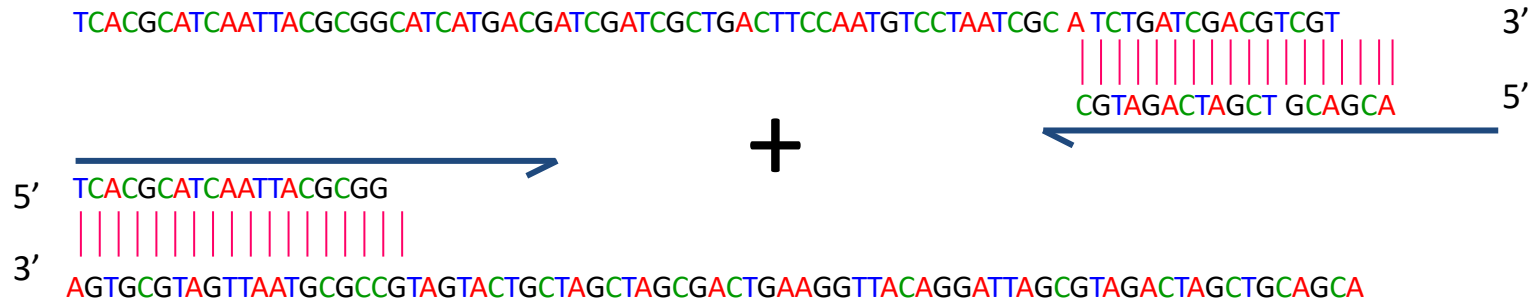


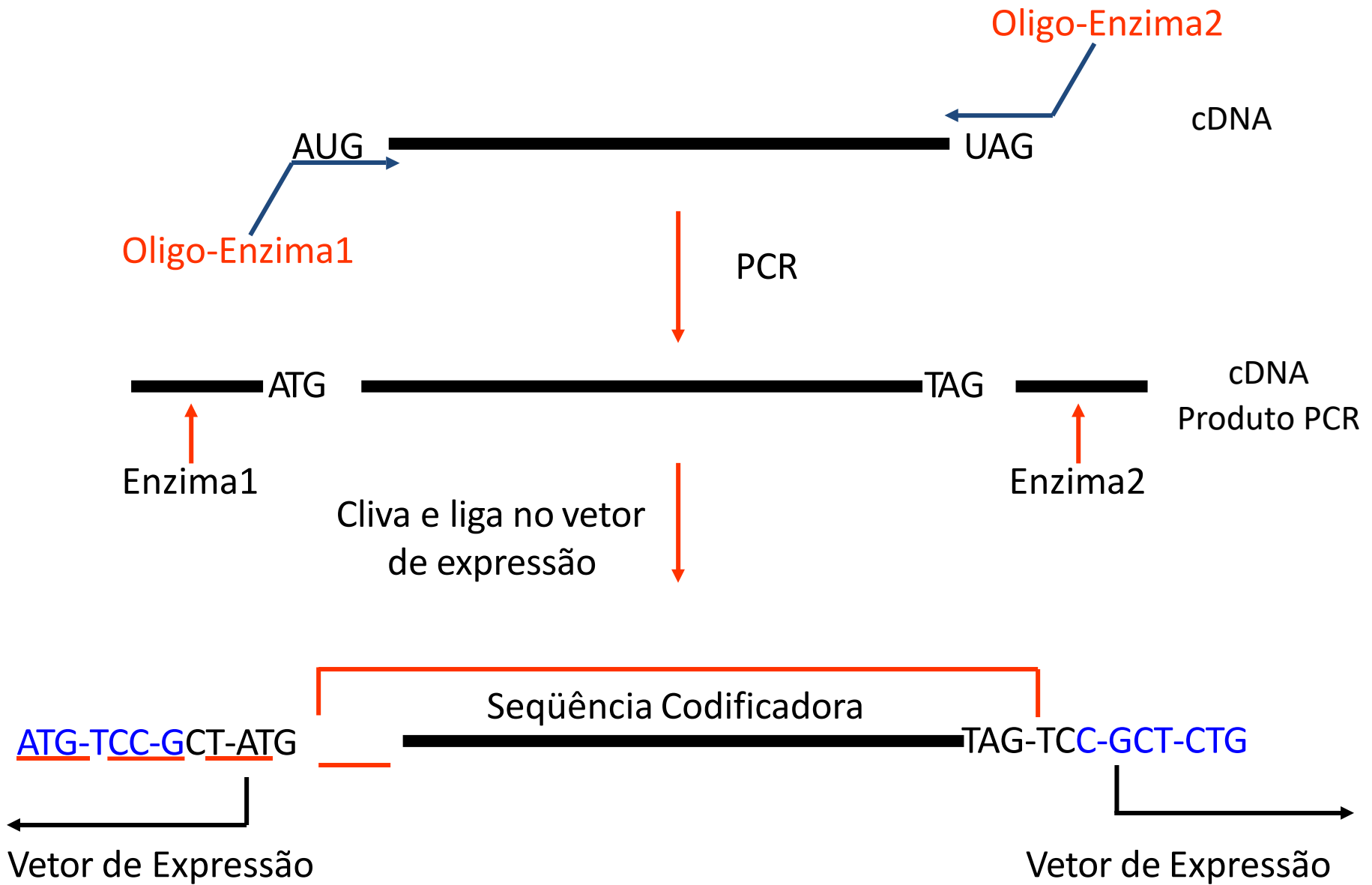
CGTAGACTAGCTGCAGCA

Reverso (3')



DNA e alinhamento de Oligonucleótidos





Manter o passo de leitura !!!!!!!

BamHI – 5' - G[↓] GATCC – 3'
3' – CCTAG[↑] G – 5'

KpnI – 5' - G[↓] GTACC – 3'
3' – CCATG[↑] G – 5'

5' -----AAC CTG TAT TTT CAG GGC GCC ATG GAT CCG GAA TTC AAA GGC----- 3'
N L Y F Q G A M D P E F K G

BamHI



5' - **ATGATTGTCGGCATTCTC**ACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGGGGCCAAT
3' - TACTAACAGCCGTAAGAGTGGTGCACCGATGCGACCGGTGTGAGCGTCGATCACACGGAGATCTCCTCGCCGTTCGAACGAGTTCGCAGACCCCGTTA

GTGGTGGCCAGAATTGGTCGGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTGTCTTCCCGGCGCTGCAAG
CACCACCGGTCTTAACCAGCCCAGGCTGAACGACACGAAGGCCTTCGTGTACGCAGATGAGGTTGCTGATAATGAGGGTCACAGAAGGGCCGCGACGTTG

CTCAAGCTCGTCCACGCGCGCCGCGTGCAGACTTCTCGAGTATCCCCACAACATCCCGGTGCGAGCTCCGCGACGCCTCCACCTGGTTCTACTACTACC
GAGTTCGAGCAGGTGCGCGCGGCGCAGCTGCTGAAGAGCTCATAGGGGGTGTGTTAGGGCCAGCTCGAGGCGCTGCGGAGGTGGACCAAGATGATGATGG

AGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTACAGGCAACCCTTTTGTGGGGTCACTCCTTGGGGCCAATGCATATTACGCCTCTGAAGTTAGCA
TCTCATGGAGGTCAGCCTAGCCCTTGGCGATGCATAAGTCCGTTGGGAAAAACAACCCAGTGAGGAACCCGTTACGTATAATGCGGAGACTTCAATCGT

CAACGCCTTCTTCATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGGTGCAATGTGATCGGCACCGGATTTGGT
GTTGCGGAAGAAGTAGTGACTAGTTCCAGCTAGCCCTTTCGTCGGATGGCCTGTGCTTGTACCCCTCTGACCACGTTACACTAGCCGTGGCCATAACCA

ATTCGCCCATCCGCAAACACTGGGGACTCGTTGCTGGATTGCTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACCAGCGACAGCAGTGCGCCAC
TAAGCGGGTAGGCGTTTGTGACCCTGAGCAACGACCTAAGCAAACAGACCCAGTTCGGTCCGCGGCTCACACTGCCGTGGTCGCTGTGTCACGCGGTG

GATTTGACTCCCAGTGTGCGCTCCAGATGCCTTGAACCGGCGCCTCAAGCTGGTGTGCTTGGTTCCAAGCCTACTTTGTGCAGTCTTCAAAAACGCAA
CTAAACTGAGGGTGACACGCGAGGGTCTACGGAACGTTGGCCGCGGAGTTCGACCACGAACCAAGGTTCCGGATGAAACACGTCGAAGAGTGTTCGCTT

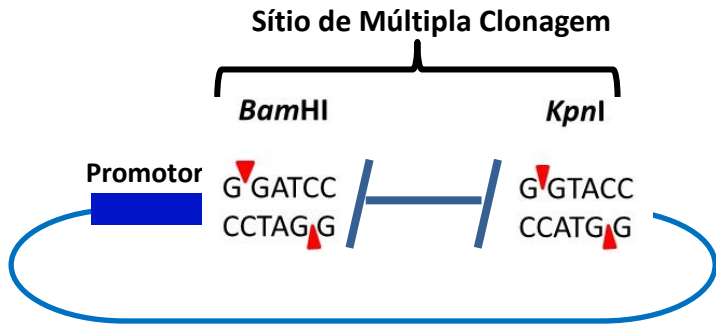
CCCATCGTTCCTGTAA – 3'

GGGTAGCAAGGACATT – 5'

BamHI

5' - TTT CAG GGC GCC ATG GAT CCG **ATG ATT GTC GGC ATT CTC** -- 3'
F Q G A M D P M I V G I L

Estratégia de Clonagem no Vetor de Expressão



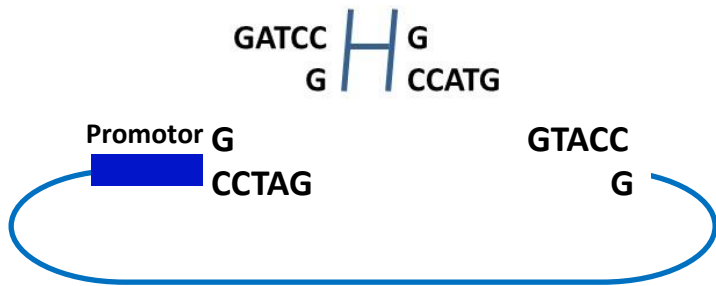
Vetor de Expressão



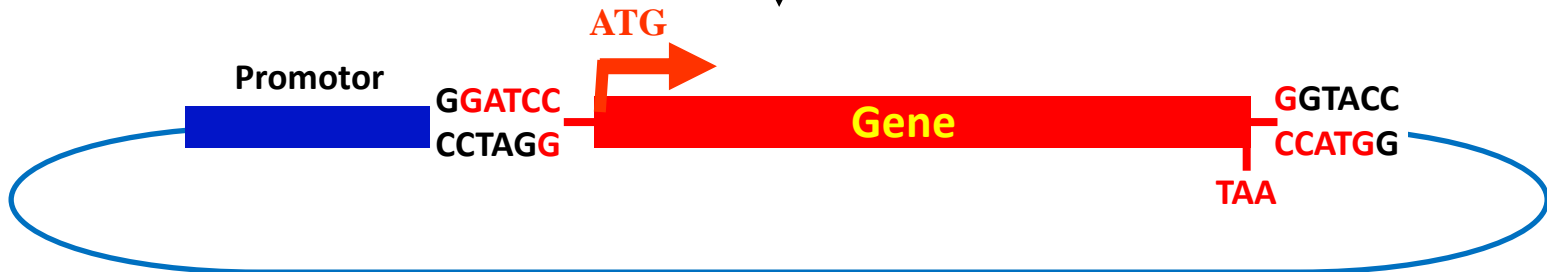
Gene de Interesse



Clivagem com enzimas de restrição *Bam*HI e *Kpn*I



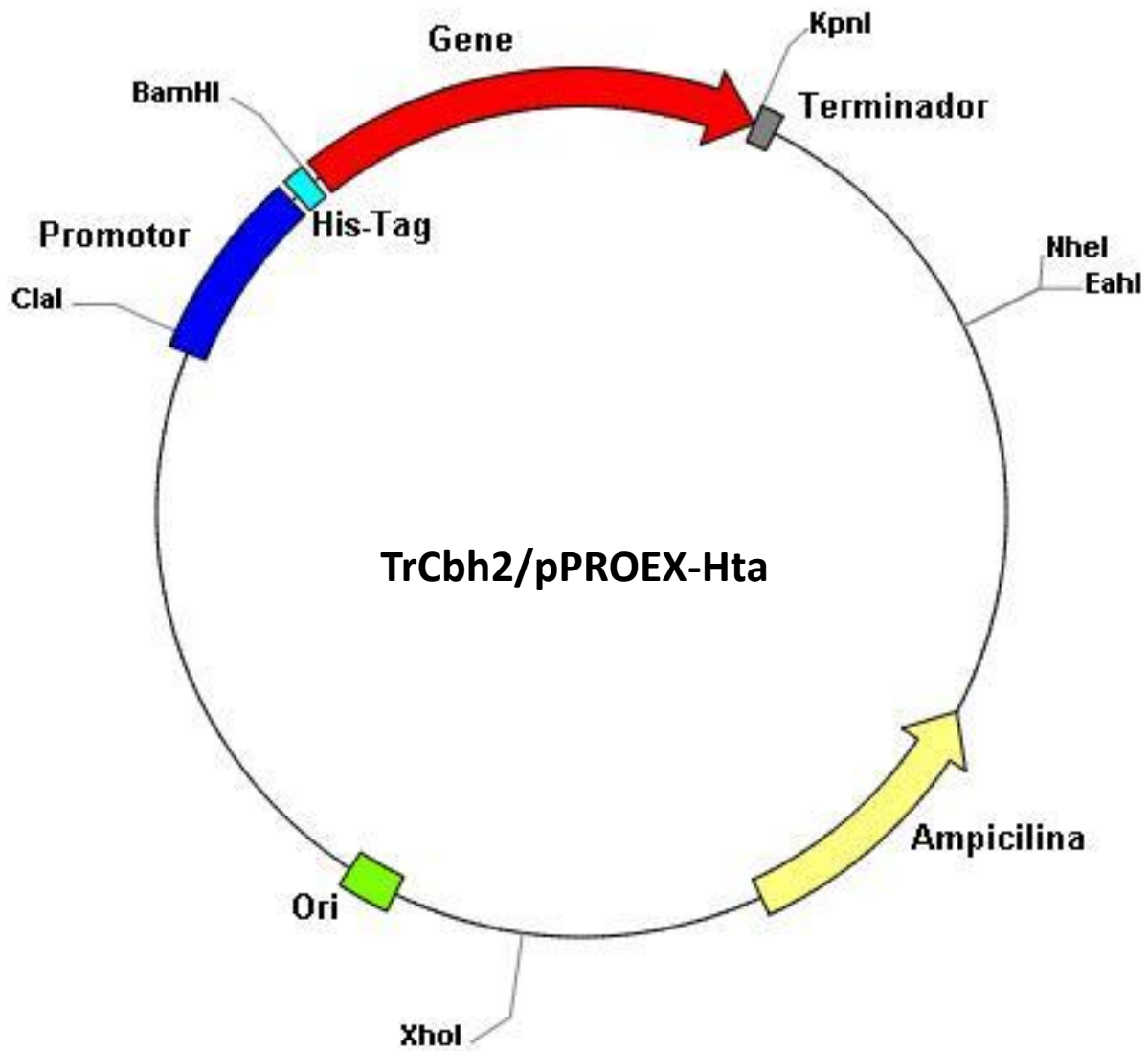
Ligação com enzima DNA Ligase



Vetor de Expressão/Gene de Interesse

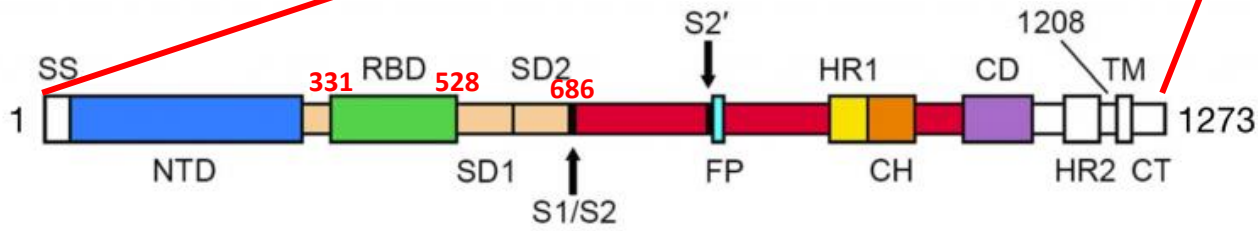
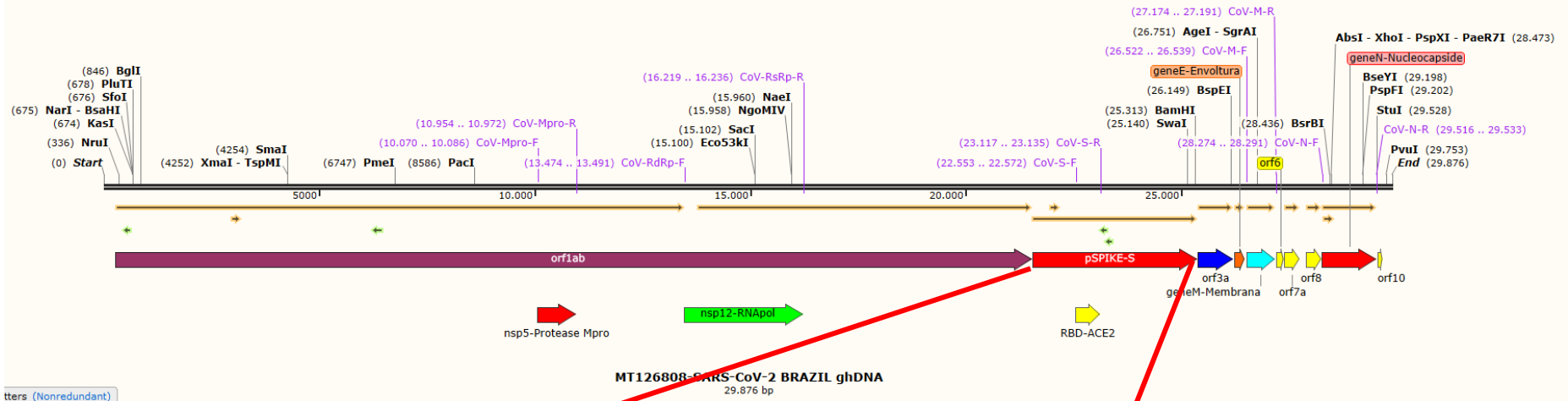
g- Faça uma figura mostrando a construção e o plasmídeo recombinante obtido a partir do item f.

Plasmid Map	
NoeClone 3.0 Demo	Virtual cloning laboratory.
pDRAW32 1.1.129	Scientific software for the molecular biologist
SimVector 4.60 Demo	tool for drawing publication and vector catalog quality maps
ApE 2.0.49	A Plasmid Editor
BVTech Plasmid 5.1 Demo	DNA plasmid drawing software
XPlasMap 0.99	a DNA mapping program for Mac OS
CGView 1.0	Circular Genome Viewer
PlasmaDNA 1.4.2	A free, cross-platform PLAsmid MANipulation program
Plasmidomics 0.2	Plasmid Drawing Program.
SnapGene Viewer 3.1.2	Create, Browse, and Share richly Annotated DNA Sequence.
pLOT 1.0.10h	Plasmid mapping program
Online Tools	
PlasMapper 2.0	automatically generates and annotates plasmid maps using only the plasmid DNA sequence as input.
NetPlasmid	Online Plasmid map drawing
Savvy v0.1	Draw plasmid map tool.
EZ PLASMID MAP V1.9	Free online plasmid draw/plotter program.
WebDSV	Free Online DNA Sequence Editor and Plasmid Drawing program.



Mapa esquemático do plasmídeo para expressão recombinante da enzima TrCbh2

GenBank: MT126808 SARS-CoV-2 - Brasil



SS - signal sequence
 NTD - N-terminal domain
 RBD - receptor-binding domain
 SD1 - subdomain 1
 SD2 - subdomain 2
 S1/S2 = S1/S2 protease cleavage site
 S2' = S2' protease cleavage site

FP = fusion peptide
 HR1 = heptad repeat 1
 CH = central helix
 CD = connector domain
 HR2 = heptad repeat 2
 TM = transmembrane domain
 CT = cytoplasmic tail

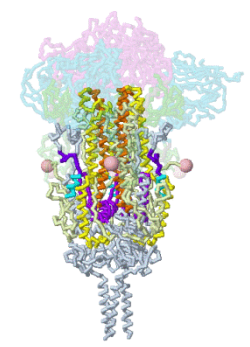
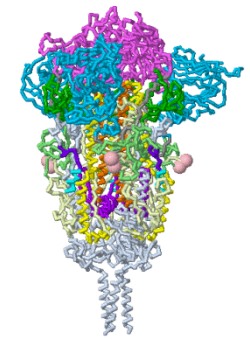


Figure adapted from Wrapp et al. 2020

Jmol

Jmol

Sequência de DNA do gene da Proteína S de SARS-CoV-2 (3822 pb)

ATGTTTGTTCCTTCTGTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAACCAGAACTCAATTACCCCCTGCATACACTAATTCTTTCACACGTGGTGTATTACC
CTGACAAAGTTTTAGATCCTCAGTTTTACATCAACTCAGGACTTGTCTTACCTTTCTTTCCAATGTTACTTGGTTCCATGCTATACATGTCTCTGGGACCAATGGTACTA
AGAGGTTTGATAACCCTGTCTACCATTTAATGATGGTGTATTATTTGCTTCCACTGAGAAGTCTAACATAATAAGAGGCTGGATTTTTGGTACTACTTTAGATTCTGAAGAC
CCAGTCCCTACTTATTGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTGTAATGATCCATTTTTGGGTGTTTATTACCACAAAAACAACAAAAGTTG
GATGGAAAGTGAGTTCAGAGTTTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCAGCCTTTTCTTATGACCTTGAAGGAAAACAGGGTAATTTCAAAAATCTT
AGGGAATTTGTGTTAAGAATATTGATGGTTATTTAAAAATATAATTCTAAGCACACGCCTATTAATTTAGTGCGTGATCTCCCTCAGGGTTTTTCGGCTTTAGAACCATTGGT
AGATTTGCCAATAGGTATTAACATCACTAGGTTTCAAACCTTTACTTGTCTTACATAGAAGTTATTTGACTCCTGGTGATTCTTCTCAGGTTGGACAGCTGGTGTCTGCAGCTT
ATTATGTGGGTTATCTTCAACCTAGGACTTTTCTATTTAAAATATAATGAAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTCTCAGAAACAAAGTGTAC
GTTGAAATCCTTCACTGTAGAAAAAGGAATCTATCAAACCTTCTAACTTTAGAGTCCAACCAACAGAATCTATTGTTAGATTTCTAATATTACAACTTGTGCCCTTTGGTG
AAGTTTTTAACGCCACCAGATTTGCATCTGTTTATGCTTGGAAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCTATATAATTCCGCATCATTTTTCCACTTTT
AAGTGTATGGAGTGTCTCTACTAAATTAATGATCTCTGCTTTACTAATGTCTATGCAGATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAA
CTGGAAAGATTGCTGATTATAATTATAAATTACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAAATTCTAACAACTTGTATTCTAAGGTTGGTGGTAATTATAATTACCT
GTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGGAGAGATATTTCAAACCTGAAATCTATCAGGCCGGTAGCACACCTTGAATGGTGTGAAAGGTTTTAATTGTTAC
TTTCTTTACAATCATATGGTTTCCAACCCACTAATGGTGTGGTTACCAACCATACAGAGTAGTAGTACTTTCTTTGAACTTCTACATGCACCAGCAACTGTTTGTGGACC
TAAAAAGTCTACTAATTTGGTTAAAAACAATGTGTCAATTTCAACTTCAATGGTTTAAACAGGCACAGGTGTTCTTACTGAGTCTAACAAAAAGTTTCTGCCTTTCCAACAAT
TTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTTGAGATTTTGACATTACACCATGTTCTTTTGGTGGTGTGAGTGTATAACACCAGGAAC
AAATACTTCTAACCAGGTTGCTGTTCTTTATCAGGATGTTAACTGCACAGAAGTCCCTGTTGCTATTTCATGCAGATCAACTTACTCTACTTGGCGTGTATTCTACAGGTT
CTAATGTTTTTCAAACACGTGCAGGCTGTTAATAGGGGCTGAACATGTCAACAACCTCATATGAGTGTGACATACCCATTGGTGCAGGTATATGCGCTAGTTATCAGACTC
AGACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATCCATCATTGCCTACACTATGTCACTTGGTGCAGAAAATTCAGTTGCTTACTCTAATAACTCTATTGCCATA
CCCACAAATTTTACTATTAGTGTACCACAGAAATTCTACCAGTGTCTATGACCAAGACATCAGTAGATTGTACAATGTACATTTGTGGTATTCAACTGAATGCAGCAATC
TTTTGTTGCAATATGGCAGTTTTTGTACACAATTAACCGTGCTTTAACTGGAATAGCTGTTGAACAAGACAAAAACACCCAAGAAGTTTTTGCACAAGTCAAACAAATTTA
CAAAACACCACCAATTAAGATTTTTGGTGGTTTTAATTTTTCAAAATATTACCAGATCCATCAAACCAAGCAAGAGGTCAATTTATTGAAGATCTACTTTTCAACAAAGTG
ACACTTGCAGATGCTGGCTTCATCAAACAATATGGTATTGCCTTGGTGTATTTGCTGCTAGAGACCTCATTTGTGCACAAAAAGTTTAAACGGCCTTACTGTTTTGCCACCTT
TGCTCACAGATGAAATGATTGCTCAATACACTTCTGCACTGTTAGCGGGTACAATCACTTCTGGTTGGACCTTTGGTGCAGGTGCTGCATTACAAATACCATTTGCTATGCA
AATGGCTTATAGGTTAATGGTATTGGAGTTACACAGAATGTTCTCTATGAGAACCAAAAATGATTGCCAACCAATTTAATAGTGCTATTGGCAAAAATTCAGACTCACTT
TCTCCACAGCAAGTGCCTTGGAAAACCTCAAGATGTGGTCAACCAAAATGCACAAGCTTTAAACACGCTTGTTAAACAACCTTAGCTCCAATTTTGGTGCATTTCAAGTG
TTTTAAATGATATCCTTTACGCTTGTACAAAAGTTGAGGCTGAAGTGCAAATTTGATAGGTTGATCACAGGCAGACTTCAAAGTTTGCAGACATATGTGACTCAACAATTAAT
TAGAGCTGCAGAAATCAGAGCTTCTGCTAATCTTGTGCTACTAAAATGTGAGAGTGTGACTTGGACAATCAAAAAGAGTTGATTTTTGTGGAAAGGGCTATCATCTTAT
GTCCTTCCCTCAGTCAAGCCTCATGGTGTAGTCTTCTTGCATGTGACTTATGTCCCTGCACAAGAAAAGAATTCACAACCTGCTCCTGCCATTTGTCATGATGGAAAAGCA
CACTTTTCTCGTGAAGGTGTCTTTGTTTCAAATGGCACACACTGGTTTGAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACACATTTGTGTCTGGTA
ACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCTGAATTAGACTCATTCAAGGAGGAGTTAGATAAATATTTAAGAATCATACATCACC
AGATGTTGATTTAGGTGACATCTCTGGCATTAAATGCTTCAGTTGTAAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTGCCAAGAATTTAAATGAATCTCTCATCGAT
CTCCAAGAACTTGGAAAAGTATGAGCAGTATATAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTTGAATGCCATAGTAATGGTGACAATTATGCTTTGCTGT
ATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTTGTGGATCCTGCTGCAAATTTGATGAAGACGACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACA
TAA

Sequencia de aminoácidos da Proteína S de SARS-CoV-2 (1273 aa)

10 20 30 40 50 60 70 80 90 100 110 120
MFVFLVLLPL VSSQCVNLT RTQLPPAYTN SFTRGVVYYPD KVFRSSVLHS TQDLFLPFSS NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNI IRGWIFGTTL DSKTQSLLIIV

130 140 150 160 170 180 190 200 210 220 230 240
NNATNVVIKV CEFQFCNDPF LGVYYHKNNK SWMESEFRVY SSANNCTFEY VSQPFMLDLE GKQGNFKNLR EFVFKNIDGY FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT

250 260 270 280 290 300 310 320 330 340 350 360
LLALHRSYLT PGDSSSGWTA GAAAYVGYL QPRTFLLYN ENGTITDAVD CALDPLSETK CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPFGE VFNATRFASV YAWNKRKISN

370 380 390 400 410 420 430 440 450 460 470 480
CVADYSVLYN SASFSTFKCY GVSPTKLNLD CFTNVYADSF VIRGDEVRFI APGQTGKIAD YNYKLPDDFT GCVIAWNSNN LDSKVGGNYN YLYRLFRRKS LKPFERDIST EIYQAGSTPC

490 500 510 520 530 540 550 560 570 580 590 600
NGVEGFNCYF PLQSYGFQPT NGVGYQPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN FNFNGLTGTG VLTESNKKFL PFQQFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSVITP

610 620 630 640 650 660 670 680 690 700 710 720
GTNTSNQVAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL IGAEHVNNSY ECDIPIGAGI CASYQTQNS PRRARSVASQ SIIAYTMSLG AENSVAYSNN SIAIPTNFTI

730 740 750 760 770 780 790 800 810 820 830 840
SVTTEILPVS MTKTSVDCTM YICGDSSTCS NLLLYGYSFC TQLNRALTGI AVEQDKNTQE VFAQVKQIYK TPIKDFGGF NFSQILPDP S KPSKRSFIED LFNKVTLD AGFIKQYGD

850 860 870 880 890 900 910 920 930 940 950 960
LGDIAARDLI CAQKFNGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM QMAYRFNGIG VTQNVLYENQ KLIANQFNSA IGKIQDSLSS TASALGKLQD VVNQNAQALN

970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080
TLVKQLSSNF GAISVVLNDI LSRLDKVEAE VQIDRLITGR LQSLQTYVTQ QLIRAAEIRA SANLAATKMS ECVLQSKRV DFCGKGYHLM SFPQSAPHGV VFLHVTYVPA QEKNEFTTAPA

1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
ICHDGKAHFP REGVFVSNGT HWFVTQRNFY EPQIITDNT FVSGNCDVVI GIVNNTVYDP LQPELDSFKE ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLID

1210 1220 1230 1240 1250 1260 1270
QELGKYEQYI KWPWYIWLGF IAGLIAIVMV TIMLCCMTSC CSCLKGCCSC GSCCKFDED SEPVLKGVKL HYT

Sequencia de DNA e aminoácidos da Proteína S de SARS-CoV-2 (1273 aa)

atg	ttt	gtt	ttt	ctt	gtt	tta	ttg	cca	cta	gtc	tct	agt	cag	tgt	gtt	aat	ctt	aca	acc
M	F	V	F	L	V	L	L	P	L	V	S	S	Q	C	V	N	L	T	T
aga	act	caa	tta	ccc	cct	gca	tac	act	aat	tct	ttc	aca	cgt	ggt	gtt	tat	tac	cct	gac
R	T	Q	L	P	P	A	Y	T	N	S	F	T	R	G	V	Y	Y	P	D
aaa	gtt	ttc	aga	tcc	tca	gtt	tta	cat	tca	act	cag	gac	ttg	ttc	tta	cct	ttc	ttt	tcc
K	V	F	R	S	S	V	L	H	S	T	Q	D	L	F	L	P	F	F	S
aat	gtt	act	tgg	ttc	cat	gct	ata	cat	gtc	tct	ggg	acc	aat	ggt	act	aag	agg	ttt	gat
N	V	T	W	F	H	A	I	H	V	S	G	T	N	G	T	K	R	F	D
aac	cct	gtc	cta	cca	ttt	aat	gat	ggt	gtt	tat	ttt	gct	tcc	act	gag	aag	tct	aac	ata
N	P	V	L	P	F	N	D	G	V	Y	F	A	S	T	E	K	S	N	I
ata	aga	ggc	tgg	att	ttt	ggt	act	act	tta	gat	tcg	aag	acc	cag	tcc	cta	ctt	att	gtt
I	R	G	W	I	F	G	T	T	L	D	S	K	T	Q	S	L	L	I	V
aat	aac	gct	act	aat	gtt	gtt	att	aaa	gtc	tgt	gaa	ttt	caa	ttt	tgt	aat	gat	cca	ttt
N	N	A	T	N	V	V	I	K	V	C	E	F	Q	F	C	N	D	P	F
ttg	ggt	gtt	tat	tac	cac	aaa	aac	aac	aaa	agt	tgg	atg	gaa	agt	gag	ttc	aga	gtt	tat
L	G	V	Y	Y	H	K	N	N	K	S	W	M	E	S	E	F	R	V	Y
tct	agt	gcg	aat	aat	tgc	act	ttt	gaa	tat	gtc	tct	cag	cct	ttt	ctt	atg	gac	ctt	gaa
S	S	A	N	N	C	T	F	E	Y	V	S	Q	P	F	L	M	D	L	E
gga	aaa	cag	ggt	aat	ttc	aaa	aat	ctt	agg	gaa	ttt	gtg	ttt	aag	aat	att	gat	ggt	tat
G	K	Q	G	N	F	K	N	L	R	E	F	V	F	K	N	I	D	G	Y
ttt	aaa	ata	tat	tct	aag	cac	acg	cct	att	aat	tta	gtg	cgt	gat	ctc	cct	cag	ggt	ttt
F	K	I	Y	S	K	H	T	P	I	N	L	V	R	D	L	P	Q	G	F
tcg	gct	tta	gaa	cca	ttg	gta	gat	ttg	cca	ata	ggt	att	aac	atc	act	agg	ttt	caa	act
S	A	L	E	P	L	V	D	L	P	I	G	I	N	I	T	R	F	Q	T
tta	ctt	gct	tta	cat	aga	agt	tat	ttg	act	cct	ggt	gat	tct	tct	tca	ggt	tgg	aca	gct
L	L	A	L	H	R	S	Y	L	T	P	G	D	S	S	S	G	W	T	A
ggt	gct	gca	gct	tat	tat	gtg	ggt	tat	ctt	caa	cct	agg	act	ttt	cta	tta	aaa	tat	aat
G	A	A	A	Y	Y	V	G	Y	L	Q	P	R	T	F	L	L	K	Y	N
gaa	aat	gga	acc	att	aca	gat	gct	gta	gac	tgt	gca	ctt	gac	cct	ctc	tca	gaa	aca	aag
E	N	G	T	I	T	D	A	V	D	C	A	L	D	P	L	S	E	T	K
tgt	acg	ttg	aaa	tcc	ttc	act	gta	gaa	aaa	gga	atc	tat	caa	act	tct	aac	ttt	aga	gtc
C	T	L	K	S	F	T	V	E	K	G	I	Y	Q	T	S	N	F	R	V

caa cca aca gaa tct att gtt aga ttt cct aat att aca aac ttg tgc cct ttt ggt gaa
Q P T E S I V R F P N I T N L C P F G E
gtt ttt aac gcc acc aga ttt gca tct gtt tat gct tgg aac agg aag aga atc agc aac
V F N A T R F A S V Y A W N R K R I S N
tgt gtt gct gat tat tct gtc cta tat aat tcc gca tca ttt tcc act ttt aag tgt tat
C V A D Y S V L Y N S A S F S T F K C Y
gga gtg tct cct act aaa tta aat gat ctc tgc ttt act aat gtc tat gca gat tca ttt
G V S P T K L N D L C F T N V Y A D S F
gta att aga ggt gat gaa gtc aga caa atc gct cca ggg caa act gga aag att gct gat
V I R G D E V R Q I A P G Q T G K I A D
tat aat tat aaa tta cca gat gat ttt aca ggc tgc gtt ata gct tgg aat tct aac aat
Y N Y K L P D D F T G C V I A W N S N N
ctt gat tct aag gtt ggt ggt aat tat aat tac ctg tat aga ttg ttt agg aag tct aat
L D S K V G G N Y N Y L Y R L F R K S N
ctc aaa cct ttt gag aga gat att tca act gaa atc tat cag gcc ggt agc aca cct tgt
L K P F E R D I S T E I Y Q A G S T P C
aat ggt gtt gaa ggt ttt aat tgt tac ttt cct tta caa tca tat ggt ttc caa ccc act
N G V E G F N C Y F P L Q S Y G F Q P T
aat ggt gtt ggt tac caa cca tac aga gta gta gta ctt tct ttt gaa ctt cta cat gca
N G V G Y Q P Y R V V V L S F E L L H A
cca gca act gtt tgt gga cct aaa aag tct act aat ttg gtt aaa aac aaa tgt gtc aat
P A T V C G P K K S T N L V K N K C V N
ttc aac ttc aat ggt tta aca ggc aca ggt gtt ctt act gag tct aac aaa aag ttt ctg
F N F N G L T G T G V L T E S N K K F L
cct ttc caa caa ttt ggc aga gac att gct gac act act gat gct gtc cgt gat cca cag
P F Q Q F G R D I A D T T D A V R D P Q
aca ctt gag att ctt gac att aca cca tgt tct ttt ggt ggt gtc agt gtt ata aca cca
T L E I L D I T P C S F G G V S V I T P
gga aca aat act tct aac cag gtt gct gtt ctt tat cag gat gtt aac tgc aca gaa gtc
G T N T S N Q V A V L Y Q D V N C T E V
cct gtt gct att cat gca gat caa ctt act cct act tgg cgt gtt tat tct aca ggt tct
P V A I H A D Q L T P T W R V Y S T G S
aat gtt ttt caa aca cgt gca ggc tgt tta ata ggg gct gaa cat gtc aac aac tca tat
N V F Q T R A G C L I G A E H V N N S Y
gag tgt gac ata ccc att ggt gca ggt ata tgc gct agt tat cag act cag act aat tct
E C D I P I G A G I C A S Y Q T Q T N S

cct cgg cgg gca cgt agt gta gct agt caa tcc atc att gcc tac act atg tca ctt ggt
P R R A R S V A S Q S I I A Y T M S L G
gca gaa aat tca gtt gct tac tct aat aac tct att gcc ata ccc aca aat ttt act att
A E N S V A Y S N N S I A I P T N F T I
agt gtt acc aca gaa att cta cca gtg tct atg acc aag aca tca gta gat tgt aca atg
S V T T E I L P V S M T K T S V D C T M
tac att tgt ggt gat tca act gaa tgc agc aat ctt ttg ttg caa tat ggc agt ttt tgt
Y I C G D S T E C S N L L L Q Y G S F C
aca caa tta aac cgt gct tta act gga ata gct gtt gaa caa gac aaa aac acc caa gaa
T Q L N R A L T G I A V E Q D K N T Q E
gtt ttt gca caa gtc aaa caa att tac aaa aca cca cca att aaa gat ttt ggt ggt ttt
V F A Q V K Q I Y K T P P I K D F G G F
aat ttt tca caa ata tta cca gat cca tca aaa cca agc aag agg tca ttt att gaa gat
N F S Q I L P D P S K P S K R S F I E D
cta ctt ttc aac aaa gtg aca ctt gca gat gct ggc ttc atc aaa caa tat ggt gat tgc
L L F N K V T L A D A G F I K Q Y G D C
ctt ggt gat att gct gct aga gac ctc att tgt gca caa aag ttt aac ggc ctt act gtt
L G D I A A R D L I C A Q K F N G L T V
ttg cca cct ttg ctc aca gat gaa atg att gct caa tac act tct gca ctg tta gcg ggt
L P P L L T D E M I A Q Y T S A L L A G
aca atc act tct ggt tgg acc ttt ggt gca ggt gct gca tta caa ata cca ttt gct atg
T I T S G W T F G A G A A L Q I P F A M
caa atg gct tat agg ttt aat ggt att gga gtt aca cag aat gtt ctc tat gag aac caa
Q M A Y R F N G I G V T Q N V L Y E N Q
aaa ttg att gcc aac caa ttt aat agt gct att ggc aaa att caa gac tca ctt tct tcc
K L I A N Q F N S A I G K I Q D S L S S
aca gca agt gca ctt gga aaa ctt caa gat gtg gtc aac caa aat gca caa gct tta aac
T A S A L G K L Q D V V N Q N A Q A L N
acg ctt gtt aaa caa ctt agc tcc aat ttt ggt gca att tca agt gtt tta aat gat atc
T L V K Q L S S N F G A I S S V L N D I
ctt tca cgt ctt gac aaa gtt gag gct gaa gtg caa att gat agg ttg atc aca ggc aga
L S R L D K V E A E V Q I D R L I T G R
ctt caa agt ttg cag aca tat gtg act caa caa tta att aga gct gca gaa atc aga gct
L Q S L Q T Y V T Q Q L I R A A E I R A

tct gct aat ctt gct gct act aaa atg tca gag tgt gta ctt gga caa tca aaa aga gtt
S A N L A A T K M S E C V L G Q S K R V
gat ttt tgt gga aag ggc tat cat ctt atg tcc ttc cct cag tca gca cct cat ggt gta
D F C G K G Y H L M S F P Q S A P H G V
gtc ttc ttg cat gtg act tat gtc cct gca caa gaa aag aac ttc aca act gct cct gcc
V F L H V T Y V P A Q E K N F T T A P A
att tgt cat gat gga aaa gca cac ttt cct cgt gaa ggt gtc ttt gtt tca aat ggc aca
I C H D G K A H F P R E G V F V S N G T
cac tgg ttt gta aca caa agg aat ttt tat gaa cca caa atc att act aca gac aac aca
H W F V T Q R N F Y E P Q I I T T D N T
ttt gtg tct ggt aac tgt gat gtt gta ata gga att gtc aac aac aca gtt tat gat cct
F V S G N C D V V I G I V N N T V Y D P
ttg caa cct gaa tta gac tca ttc aag gag gag tta gat aaa tat ttt aag aat cat aca
L Q P E L D S F K E E L D K Y F K N H T
tca cca gat gtt gat tta ggt gac atc tct ggc att aat gct tca gtt gta aac att caa
S P D V D L G D I S G I N A S V V N I Q
aaa gaa att gac cgc ctc aat gag gtt gcc aag aat tta aat gaa tct ctc atc gat ctc
K E I D R L N E V A K N L N E S L I D L
caa gaa ctt gga aag tat gag cag tat ata aaa tgg cca tgg tac att tgg cta ggt ttt
Q E L G K Y E Q Y I K W P W Y I W L G F
ata gct ggc ttg att gcc ata gta atg gtg aca att atg ctt tgc tgt atg acc agt tgc
I A G L I A I V M V T I M L C C M T S C
tgt agt tgt ctc aag ggc tgt tgt tct tgt gga tcc tgc tgc aaa ttt gat gaa gac gac
C S C L K G C C S C G S C C K F D E D D
tct gag cca gtg ctc aaa gga gtc aaa tta cat tac aca **taa**
S E P V L K G V K L H Y T -

Exercício

Desenhe os oligonucleotídeos/primers para clonagem e expressão de:

- 1- Proteína S de SARS-CoV-2
- 2- RBD da proteína S de SARS-CoV-2

Região de Múltipla Clonagem do vetor de Expressão:

BamHI EcoRI StuI Sall XbaI XhoI KpnI HindIII
ATGGATCCGGAATTCAAAGGCCTACGTCGACGAGCTCAACTAGTGC GGCCGCTTTCGAAATCTAGAGCCTGCAGTCTCGAGGCATGCGGTACCAAGCTTGGCTGTTTTG
M D P E F K G L R R R A Q L V R P L S N L E P A V S R H A V P S L A V L

Enzimas que não clivam o gene: AatII, AbsI, Bcgl, BciVI, BglI, BplI, Bpu10I, BsePI, BseYI, BsrBI, BtgZI, BtsI, DralI, DrdI, Esp3I, FseI, FspAI, HaeII, HaeIV, Hgal, Hpy99I, KpnI, MauBI, MfeI, MluI, NaeI, NarI, NheI, NotI, NruI, PacI, PvuI, RsrII, SacI, SacII, Sall, SanDI, SapI, P1-Scel, SfiI, SgfI, SgrAI, SgrDI, SmaI, SnaBI, SphI, SrfI, StuI, TaqII, TauI, XbaI, XhoI

BamHI: G↓GATCC
EcoRI: G↓AATTC
StuI: AGG↓CCT
Sall: G↓TCGAC
XbaI: T↓CTAGA
XhoI: C↓TCGAG
KpnI: GGTAC↓C
HindIII: A↓AGCTT



Obrigado

fscha@usp.br

USP – 2º Semestre 2024