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

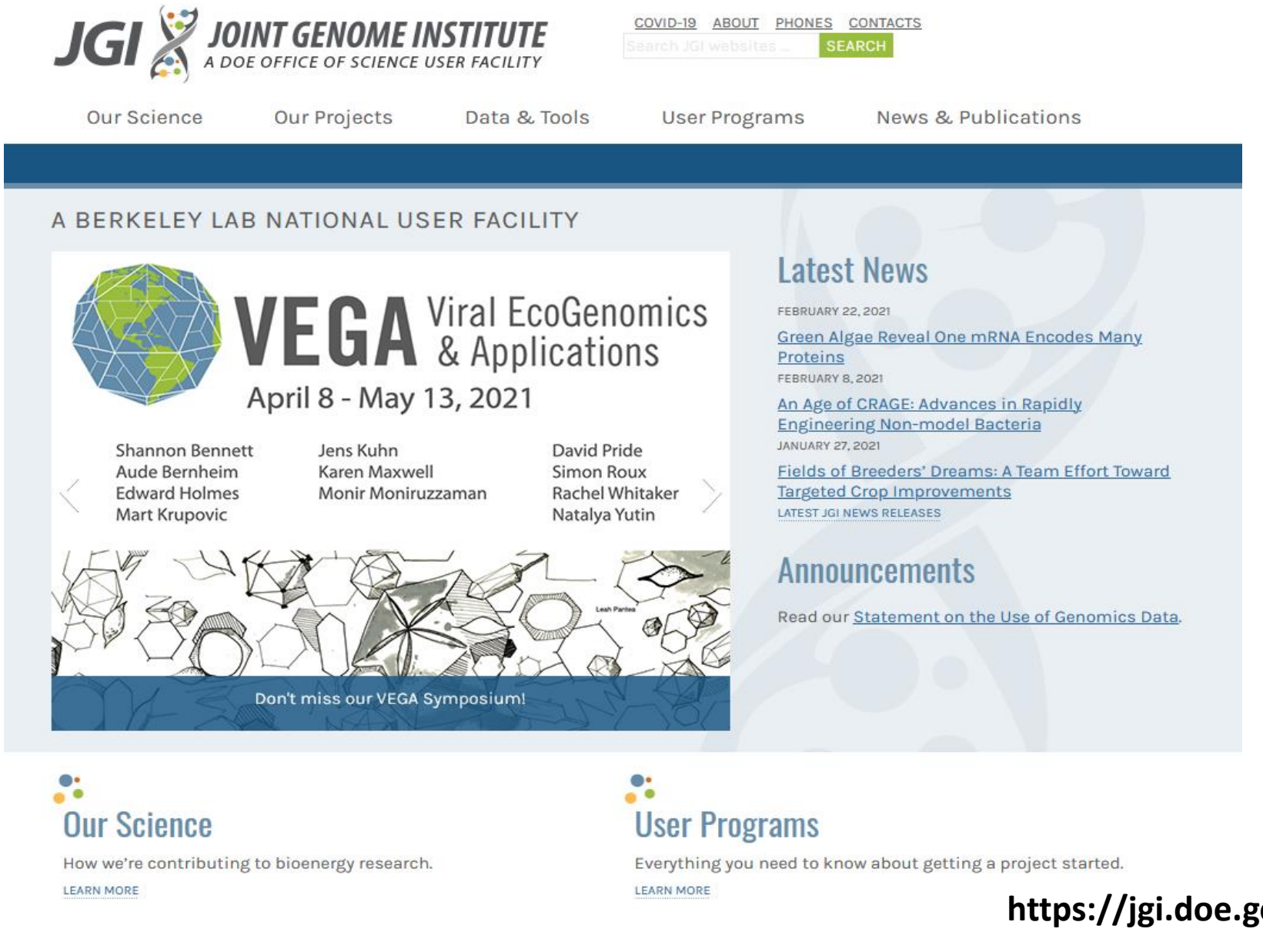
Servidores não-docentes:
Tec. Pedro Manoel dos Santos - pedroms@usp.br

Créditos: 4
Período: Quinta-feira (14h00 -18h00), Laboratório de Biotecnologia – Edifício A2, 1º andar

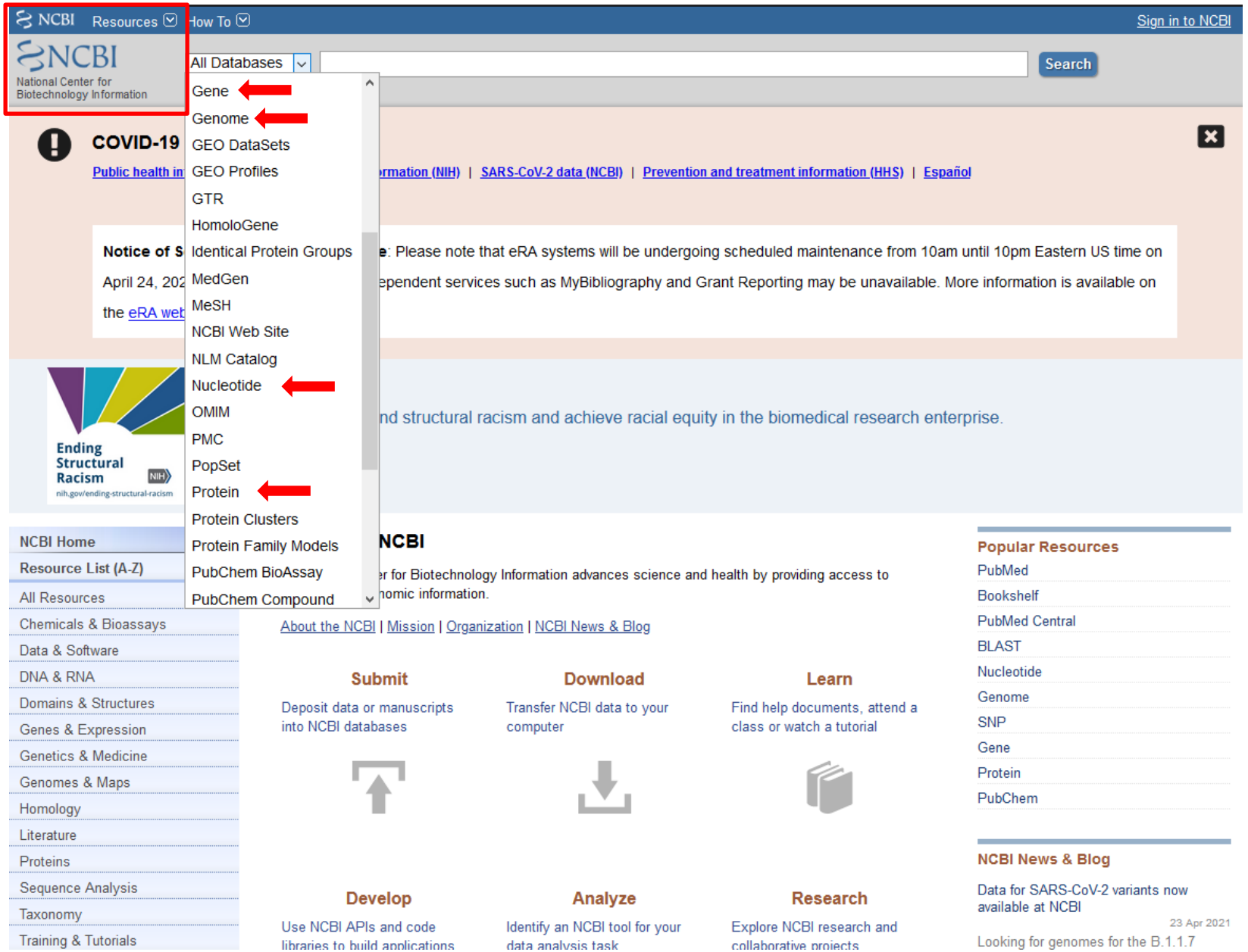
USP - 2024

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

Bioinformática: Base de Dados



https://jgi.doe.gov/



https://www.ncbi.nlm.nih.gov/

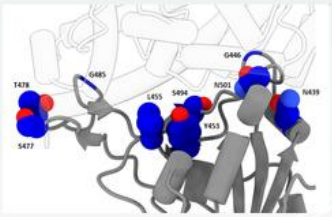
GISAID

Login

About us | Database Features | Events | Collaborations | References | Registration | Help

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 SNV blitn 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

GISAID Resources

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

Frequently Asked Questions
Answers about the GISAID Initiative

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

https://www.gisaid.org/

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.

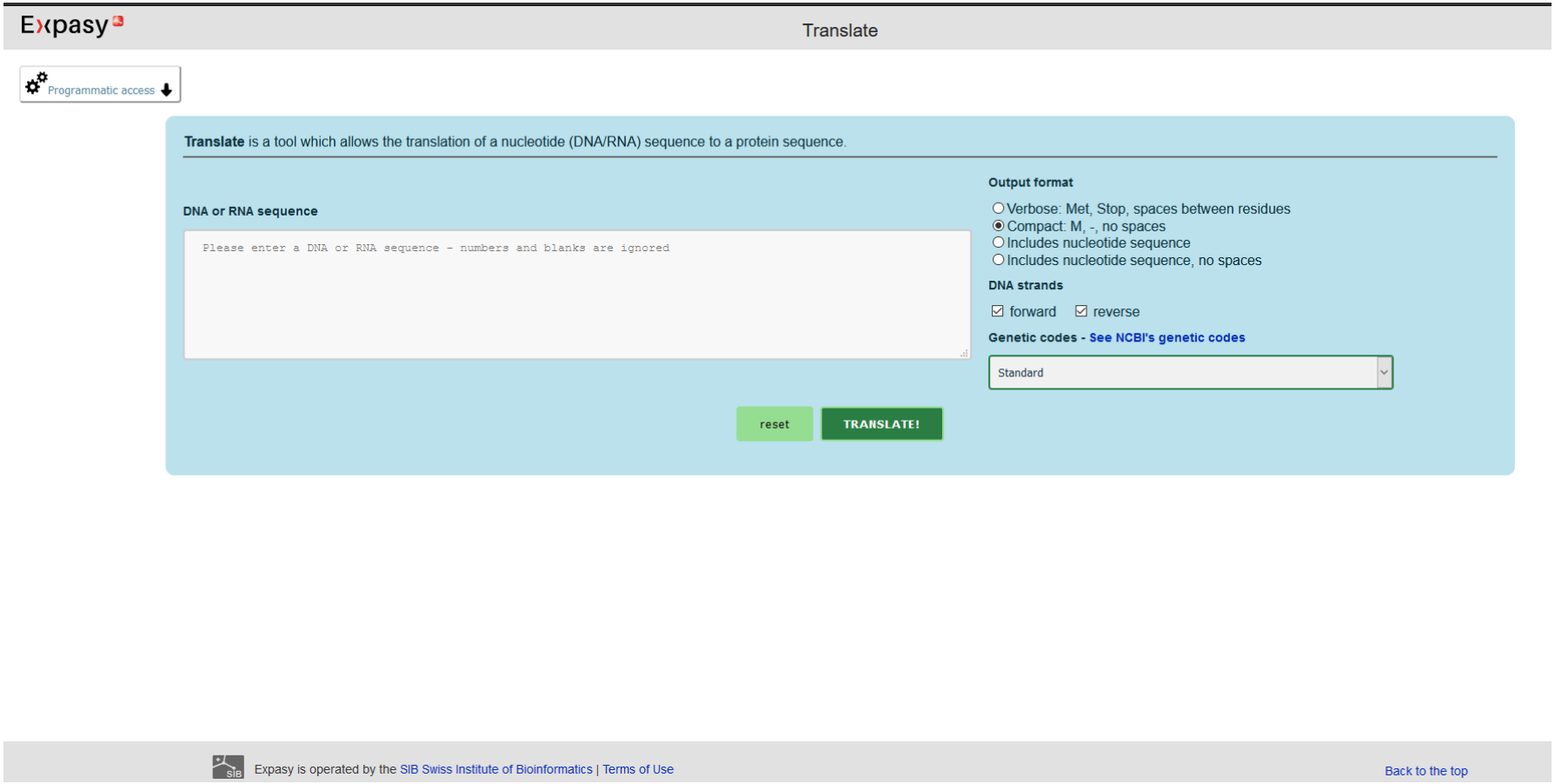
Sequência de Interesse

>Problema (2320 bp)

CTTCTTTATTGGGTAATATACAGCCAGGCGGGGATGAAGCTCATTAGCCGCCACTCAAGGCTATACAATGTTGCCAACTCTCCGGGC
TTTATCCTGTGCTCCCGAATACCACATCGTGATGATGCTTCAGCGCACGGAAGTCACAGACACCGCCTGTATAAAAGGGGGACTGT
GACCCTGTATGAGGCGCAACATGGTCTCACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCCTCTGTGTATTGCACCATGATTGT
CGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGG
TAATTATGTGAACCCTCTCAAGAGACCCAAATACTGAGATATGTCAAGGGGCCAATGTGGTGGCCAGAATTGGTCGGGTCCGACTT
GCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTGTCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGC
GCGCCGCGTCGACGACTTCTCGAGTATCCCCCACAACATCCCGGTCGAGCTCCGCGACGCCTCCACCTGGTTCTACTACTACCAGA
GTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCAACCCTTTTGTTGGGGTCACTCCTTGGGCCAATGCATATTACGCCTCT
GAAGTTAGCAGCCTCGCTATTCCTAGCTTGACTGGAGCCATGGCCACTGCTGCAGCAGCTGTCGCAAAGGTTCCCTCTTTTATGTG
GCTGTAGGTCCTCCCGGAACCAAGGCAATCTGTTACTGAAGGCTCATCATTCACTGCAGAGATACTCTTGACAAGACCCCTCTCATG
GAGCAAACCTTGGCCGACATCCGCACCGCCAACAAGAATGGCGGTAACTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATC
GCGATTGCGCTGCCCTTGCCTCGAATGGCGAATACTCTATTGCCGATGGTGGCGTCGCCAAATATAAGAACTATATCGACACCATTCG
TCAAATTGTCGTGGAATATTCCGATATCCGGACCCTCCTGGTTATTGGTATGAGTTTAAACACCTGCCTCCCCCCCCCCTTCCCTTCCT
TTCCCGCCGGCATCTTGTCGTTGTGCTAACTATTGTTCCCTCTTCCAGAGCCTGACTCTCTTGCCAACCTGGTGACCAACCTCGGTAC
TCCAAAGTGTGCCAATGCTCAGTCAGCCTACCTTGAGTGCATCAACTACGCCGTCACACAGCTGAACCTTCCAAATGTTGCGATGTA
TTTGGACGCTGGCCATGCAGGATGGCTTGGCTGGCCGGCAAACCAAGACCCGGCCGCTCAGCTATTTGCAAATGTTTACAAGAAT
GCATCGTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTCGCCAACTACAACGGGTGGAACATTACCAGCCCCCCATCGTACAC
GCAAGGCAACGCTGTCTACAACGAGAAGCTGTACATCCACGCTATTGGACCTCTTCTTGCCAATCACGGCTGGTCCAACGCCTTCTT
CATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGGTGCAATGTGATCGGCACCGGATT
TGGTATTCGCCCATCCGCAAACACTGGGGACTCGTTGCTGGATTCGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACC
AGCGACAGCAGTGCGCCACGATTTGACTCCCACTGTGCGCTCCCAGATGCCTTGCAACCGGCGCCTCAAGCTGGTGCTTGGTTCC
AAGCCTACTTTGTGCAGCTTCTCACAAACGCAAACCCATCGTTCCTGTAAGGCTTTCGTGACCGGGCTTCAAACAATGATGTGCGAT
GGTGTGGTTCCCGGTTGGCGGAGTCTTTGTCTACTTTGGTTGTCTGTCGCAGGTCGGTAGACCGCAAATGAGCAACTGATGGATTG
TTGCCAGCGATACTATAATTCACATGGATGGTCTTTGTCGATCAGTAGCTAGTGAGAGAGAGAGAACATCTATCCACAATGTCGAGT
GTCTATTAGACATACTCCGAGAATAAAGTCAACTGTGTCTGTGATCTAAAGATCGATTCGGCAGTCGAGTAGCGTATAACAACTCCG
AGTACCAGCAAAAGCACGTCGTGACAGGAGCAGGGCTTTGCCAACTGCGCAACCTTGCTTGAATGAGGATACACGGGGTGCAAC
ATGGCTGTACTGATCCATCGCAACCAAAATTTCTGTTTATAGATCAAGCTGGTAGATTCCAATTACTCCACCTCTT

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

Ferramenta: Translate - Expasy



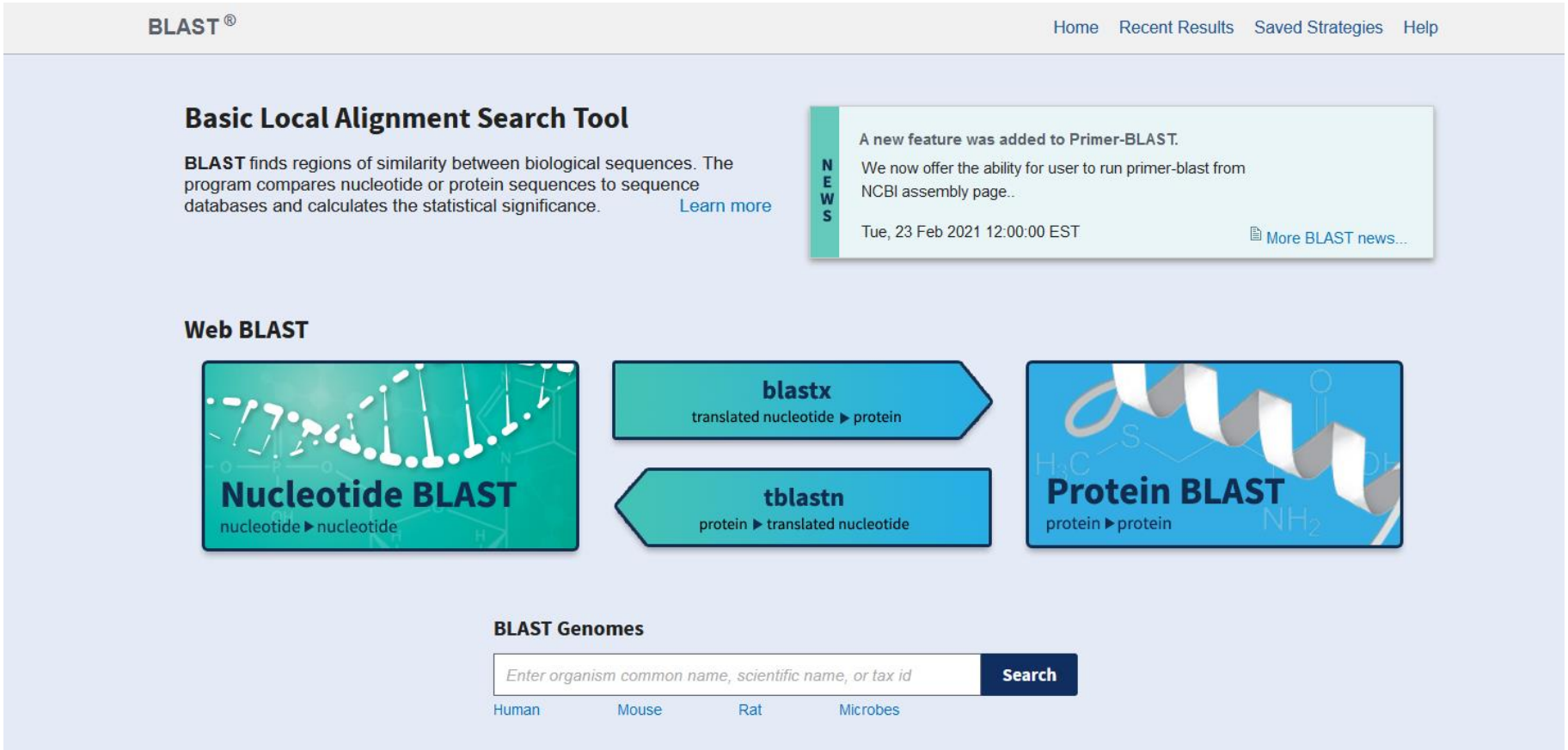
https://web.expasy.org/translate/


5'3' Frame 1
LLYWVIYSQAGMKLISRHSRLYNVANSPGFILCSRIPHRDDASAHGSHRHRLYKRGTVTLYEAQHGLTAAHLKRLVRSPSVYCTMIVGILTTLATLATLAASVPLEERQACSSVW-LCEPSQETQILRYVKGPMWWPELVGSDLLCFRKHMR
LLQRLLLPVSSRRCKLKLVHARRVDDFSSIPHNIPVELRDASTWFYYYQSTSSRIGNRYVFRQPFCWGHSLGQCILRL-S-QPRYS-LDWSHGHCCSSCRKGSLFYVAVGPPGTKAICY-RLIIHCRDTLDKTPLMEQTLADIRTANKNGGN
YAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIGMSLNTCLPPPFPSFPAGILSLC-LLFPLPEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLF
ANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPAPQAGAWFQAYF
VQLLTNANPSFL-GFRDRASNNDVRWCGSRLAESLSTLVVCRRSVDRK-ATDGLLPAIL-FTWMVFVDQ-LVREREHLSTMSSVY-TYSENKVNCVCDLKIDSAVE-RITTPSTSKSTS-QEQGFANCATLLE-GYTGCNMAVLIHRNQNFC
L-IKLVDSNYSTS
5'3' Frame 2
FFIG-YTARRG-SSLAATQGYTMLPTLRALSCAPEYHIVMMLQRTEVTDTACIKGGL-PCMRRNMVSQQLT-RGL-DHPLCIAP-LSAFSPRWLRWPHSQLVCL-RSGKLAQASGNYVNPLKRPKY-DMSRGQCGGQNWSGPTCCASGSTCV
YSNDYYSQCLPGAASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWL-VLPEPRQSVTEGSSFTAEILLTRPLSWSKPWPTSAPPTRMAVT
MPDSLWCMTCRIAIALPLPRMANTLLPMVASPNIRTISTPFVKLSWNIPISGPSWLLV-V-TPASPPPSLPFPPASCRCANYCSLFQSLTLLPTW-PTSVLQSVPMLSQPTLSASTTPSHS-TFQMLRCIWTLAMQDGLAGRQTKTRPLSYL
QMFTRMHRLRELFADWQPMSPTTTGGTLPAPHRTRKATLSTTRSCTSTLLDLFLPITAGPTPSSSLIKVDRESSLPDSNSGETGAM-SAPDLVFAHPQTLGTRCWIRLSGSSQAASVTAPATAVRHDLTPTVRSQMPCNRRLKLVLGSKPTL
CSFSQTQTHRSCKAFVTGLQTMMCDGVVPGWRSLCLLWLSVAGR-TANEQLMDCCQRYYNSHGWSLSISS--ERENIYPQCRVSIRHTPRIKSTVSVI-RSIRQSSSV-QLRVPAKARRDRSRALPTAQPCLNEDTRGATWLY-SIATKISV
YRSSW-IPITPPL
5'3' Frame 3
SLLGNIQPGGDEAH-PPLKAIQCCQLSGLYPVLPNTTS--CFSARKSQTPPV-KGDCDPV-GATWSHSSSPEEACKITLCVLHHDCRHSHHAGYAGHTRS-CASRGAASLLKRLVIM-TLSRDPNTEICQGANVVARIGRVRLAVLPEAHAS
TPTTITPSVFPALQAQARPRAPRRRLLEYPPQHPGRAPRRLHLVLLLPEYLQSDREPLRIQATLLLGSLLGPMHITPLKLAASLFLA-LEPWPLLQQLSQRFPLLCGCRSSRNQGNLLLKAHHSLQRYS-QDPSHGANLGRHPHRQQEWR-L
CRTVCGV-LAGSRLRCPCLEWRILYCRWWRRQI-ELYRHHSSNCRGIFRYPDPPGYWYEFKHLPPPPLPFLSRRHLVVVLTIVPSSRA-LSCQPGDQPRYSKVCQCSVSLP-VHQLRRHTAEPSKCCDVFGRWPCRMAWLAGKPRPGRSAIC
KCLQECIVSESSSRIGNQCRQLQRVEHYQPPIVHARQRCLQREAVHPRYWTSSCQSRLVQRLLHH-SRSIGKAAYRTATVGRLVQCDRHRIWYSPIRKHWGLVAGFVCLGQARRRV-RHQRQQCATI-LPLCAPRCLATGASSWCLVPSLLC
AASHKRKPIVPVRLS-PGFKQ-CAMVWFPVGGVFVYFGCLSQVGRPQMSN-WIVASDTIIHMDGLCRSVASERERTSIHNVECLLDILRE-SQLCL-SKDRFGSRVAYNNSEYQQKHVVTGAGLCQLRNLA-MRIHGVQHGCTDPSQPKFLF
IDQAGRFQLLHL
3'5' Frame 1
KRWSNWNLPA-SINRNFGCDGSVQPCCTPCILIQARLRSWQSPAPVTTCFCWYSELLYATRLPNRSLDHRHS-LYSRSMSNRHSTLWIDVLSLSLATDRQRPSM-IIVSLATIHQLLICGLPTCDRQPK-TKTPPTGNHTIAHHCLKPGHES
LTGTMGLRL-EAAQSRLGTKHQLEAPVARHLGAHSGSQIVAHCCRWCRHTRRLA-PRQTNPATSPQCLRMGEYQIRCRSHCTSLPTVAVR-AAFPIDLDQ--RRRWTSRDWQEEVQ-RGCTASRCRQRCLACTMGGW-CSTRCSWRHWLPIR
EELSETMHSCKHLQIAERPGLGLPASQAILHGQRPNTSQHLEGSAV-RRS-CTQGRLTEHWHTLEYRGWSPGWQESQALEEGTIVSTTTRCRRERKGRGGGGRCLNSYQ-PGGSGYRNIPRQFDEWCRYSSYIWRRHHRQ-SIRHSRQGQRN
RDPASHTPQTVRHSYRHSCWRCGCRPRFAP-EGSCQEYLCSE--AFSNRLPWFREDLQPHKRGNLCDSCCSSGHGSSQARNSEAANFRGVICIGPRSDPNKRVA-IRSGSRSDWRYSGSSRTRWRRRGARPGCCGGYSRSRRRGARGRA-AC
SAGKTLGVIVVGVDACASGSTASRTRPILATTLAP-HISVFGSLERVHIITRRLSKLAAPLEAH-LRVWPA-PAW-ECRQSWCNTQRVILQASSGELL-DHVAPHTGSQSPFYTGGVCDFRALKHHHDVVFGSTG-SPESWQHCIALSGG--
ASSPPGCILPNKE
3'5' Frame 2
RGGVIGIYQLDL-TEILVAMDQYSHVAPRVSSFKQGCAVGKALLLSRRAFAGTRSCYTLLDCRIDL-ITDTVDFILGVCLIDTRHCG-MFSLSH-LLIDKDHPCEL-YRWQQSISCSFAVYRPATDNQSRQRLRQPGTTPSHIIV-SPVTKA
LQERWVCVCEKLHKVGLEPSTSLRRRLQGIWERTVGVKSWRTAVAGAVTLAAWLDPDKRIQQRVPSVCGWANTKSGADHIAPVSPLLLSGRLLSRSTLISDEEGVGPAVIGKKRSNSVDVQLLVVDSVALRVRWGAGNVPPVVVGDIGCQSA
KSSRRRCILVNICK-LSGRVLVCRPAKPSCMASVQIHRNIWKVQLCDGVVDALKVG-LSIGTLWSTEVGHQVGKRVRLWKREQ-LAQRQDAGGKGREGGGEAGV-THTNNQEGPDIGIFHDNLTNGVDIVLIFGDATIGNRVFAIRGKGSAI
AIRQVIHHKLSGIVTAILVGGADVGQGLLHERGLVKSISAVNDEPSVTDCLGSGRTYSHIKEGTFATAAAAVAMAPVKLGIARLLTSEA-YALAQGVTPTKGLPEYVAVPDPTGGTLVVVEPGGGVAELDRDVVGDTREVVDAARVDELELA
APGRHWE--SLE-THVLPEAQQVGPDQFWPPHWPLDISQYLGLLRGFT-LPDA-ASLPLL-RHTSCECGQRSQRGENADNHGAIHRG-SYKPLQVSCCETMLRLIQGHSPPFIQAVSVTSVR-SIITMWYSGAQDKARRVGNIV-P-VAANE
LHPRLAVYYPIKK
3'5' Frame 3
EVE-LESTSLIYKQKFWLRWISTAMLHPVYPHSSKVAQLAKPCSCHDVLLLVLGVVIRYSTAESIFRSQTQLTLFSEYV--TLDIVDRCSLSLTSY-STKTIHVNYSIAGNNPSVAHLRSTDLRQTTKVDKDSANREPHHRTSLFEARSRKP
YRNDGFAFVRSCTK-AWNQAPA-GAGCKASGSAQWESNRGALLSLVPSHSPPGLTQTNESSNESPVFADGRIPNPVPITLHQSPHCCCPVGCFPDRP-SVMKKALDQP-LARRGPIAWMYSFSL-TALPCVYDGGLVMFHPL-LATLVANPR
RALGDDAFL-TFANS-AAGSWFAGQPSHPAWPASKYIATFGRFSCVTA-LMHSR-AD-ALAHFGVPRLVTRLARESGSGRGNNS-HNDKMPAGKEGKGGGRQVFKLIPITRRVRISEYSTTI-RMVSI-FLYLATPPSAIEYSPFEARAAQS
RSGKSYTTNCPA-LPPFLLAVRMSAKVCSMRGVLSRVSLQ-MMSLQ-QIALVPGGPTAT-KREPLRQLLQQWPWLQSS-E-RGC-LQRRNMHWPKE-PQQKGCLNT-RFPIRLEVLW---NQVEASRSSTGMLWGILEKSSTRRAWTSLSLQ
RREDTGSNSRWSRRMCFRKHSKSDPTNSGHHIGPLTYLSIWVS-EGSHNYQTLEQACRSSRGTLAASVASVASVVRMPTIMVQYTEGDLTSLFR-AAVRPCCASYRVTVPLLYRRCL-LPCAEASSRCGIREHRIKPGELATLYSLEWRLMS
FIPAWLYITQ-R

Resposta:

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

Ferramenta: BLAST - NCBI



https://blast.ncbi.nlm.nih.gov/Blast.cgi

Edit Search | Save Search | Search Summary | How to read this report? | BLAST Help Videos | Back to Traditional Results Page

Job Title	1
RID	3T2J2RAV016 *Search expires on 03-25 23:12 pm* 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

Type common name, binomial, taxid or group name

+ Add organism

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

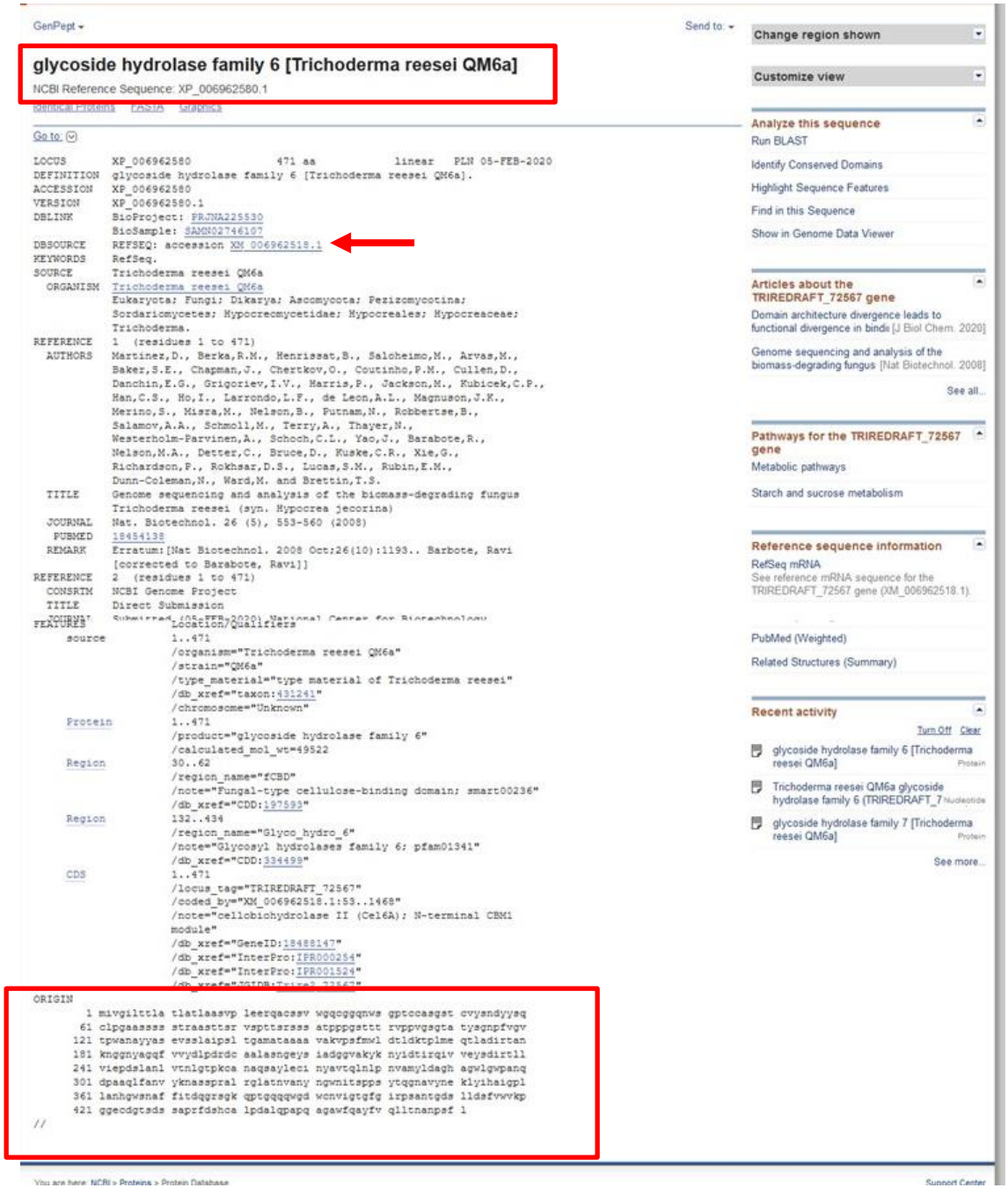
Filter Reset

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download 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
glycoside hydrolase family 6 [Trichoderma reesei QM6a]	Trichoderma reesei QM6a	591	740	62%	0.0	91.20%	471	XP_006962580.1
Chain A, CELLOBIOHYDROLASE II CORE PROTEIN [Trichoderma reesei]	Trichoderma reesei	591	677	51%	0.0	91.20%	365	3CBH_A
Chain A, CELLOBIOHYDROLASE CEL6A (FORMERLY CALLED CBH II) [Trichoderma ...	Trichoderma reesei	591	674	51%	0.0	91.20%	363	1QK0_A
cellobiohydrolase II [synthetic construct]	synthetic construct	590	738	61%	0.0	91.20%	447	AER26911.1
cellbiohydrolase II [Trichoderma reesei]	Trichoderma reesei	590	739	62%	0.0	90.91%	471	ADC83999.1
Chain A, CELLOBIOHYDROLASE II [Trichoderma reesei]	Trichoderma reesei	589	675	51%	0.0	90.91%	365	1CB2_A
hypothetical protein TgHK011_002311 [Trichoderma gracile]	Trichoderma gracile	589	773	65%	0.0	91.20%	471	KAH0490860.1



GenPept ▾ Send to: ▾

glycoside hydrolase family 6 [Trichoderma reesei QM6a]

NCBI Reference Sequence: XP_006962580.1

Identical Proteins FASTA Graphics

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., Henrissat,B., Saloheimo,M., Arvas,M.,
            Baker,S.E., Chapman,J., Chertkov,O., Coutinho,P.M., Cullen,D.,
            Danchin,E.G., Grigoriev,I.V., Harris,P., Jackson,M., Kubicek,C.P.,
            Han,C.S., Ho,I., Larrondo,L.F., de Leon,A.L., Magnuson,J.K.,
            Merino,S., Misra,M., Nelson,B., Putnam,N., Robbertse,B.,
            Salamov,A.A., Schmoll,M., Terry,A., Thayer,N.,
            Westerholm-Parvinen,A., Schoch,C.L., Yao,J., Barabote,R.,
            Nelson,M.A., Detter,C., Bruce,D., Kuske,C.R., Xie,G.,
            Richardson,P., 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)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (05-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:431241"
                     /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="cellobiohydrolase II (Cel6A); N-terminal CBM1
                     module"
                     /db_xref="GeneID:18488147"
                     /db_xref="InterPro:IPR000254"
                     /db_xref="InterPro:IPR001524"
                     /db_xref="JGIDB:Trire2_72567"
ORIGIN
        1 mivgilttla tlatlaasvp leerqacasv wgqcgggnws gptccasgst cvysndyysq
       61 clpgaasssa straasttsr vspttsrsss atpppgsttt rvppvgsgta tysgnpfvgv
      121 tpwanayyas evsslaipsl tgamataaaa vakvpsfmwl dtldktplme qtladirtan
      181 knqgnyagqf vvydlpdrdc aalasngeys iadggvakyk nyidtirqiv veysdirtll
      241 viepdslanl vtnlgtpkca naqsayleci nyavtqlnlp nvamyldagh agwlgwpanq
      301 dpaaqlfanv yknasspral rglatnvany ngwnitspps ytqgnavyne klyihaigpl
      361 lanhgwsnaf fitdqgrsgk qptqqqqwgd wcnvigtgfg irpsantgds lldafvwvkp
      421 ggecdgtsds saprfdshca lpdalqpapq agawfqayfv qlltnanpsf l
//
```

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...

You are here: NCBI › Proteins › Protein Database

Support Center

Proteína

```
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:IPR000254"
                /db_xref="InterPro:IPR001524"
                /db_xref="JGIDB:Trire2_72567"
                /translation="MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTC
                CASGSTCVYSNDYYSQCLPGAASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRV
                PPVGSGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWL
                DTLDKTPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAK
                YKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQ
                LNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRALRGLATNVANYNGWN
                ITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQQQWGDWC
                NVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPAPQ
                AGAWFQAYFVQLLTNANPSFL"
ORIGIN
        1 cacagcagct cacctgaaga ggcttgtaag atcaccctct gtgtattgca ccatgattgt
       61 cggcattctc accacgctgg ctacgctggc cacactcgca gctagtgtgc ctctagagga
      121 gcggcaagct tgctcaagcg tctggggcca atgtggtggc cagaattggt cgggtccgac
      181 ttgctgtgct tccggaagca catgcgtcta ctccaacgac tattactccc agtgtcttcc
      241 cggcgctgca agctcaagct cgtccacgcg cgccgcgtcg acgacttctc gagtatcccc
      301 cacaacatcc cggtcgagct ccgcgacgcc tccacctggt tctactacta ccagagtacc
      361 tccagtcgga tcgggaaccg ctacgtattc aggcaaccct tttgttgggg tcactccttg
      421 ggccaatgca tattacgcct ctgaagttag cagcctcgct attcctagct tgactggagc
      481 catggccact gctgcagcag ctgtcgcaaa ggttccctct tttatgtggc tagatactct
      541 tgacaagacc cctctcatgg agcaaacctt ggccgacatc cgcaccgcca acaagaatgg
      601 cggtaactat gccggacagt ttgtggtgta tgacttgccg gatcgcgatt gcgctgccct
      661 tgcctcgaat ggcgaatact ctattgccga tggtggcgtc gccaaatata agaactatat
      721 cgacaccatt cgtcaaattg tcgtggaata ttccgatatc cggaccctcc tggttattga
      781 gcctgactct cttgccaacc tggtgaccaa cctcggtact ccaaagtgtg ccaatgctca
      841 gtcagcctac cttgagtgca tcaactacgc cgtcacacag ctgaaccttc caaatgttgc
      901 gatgtatttg gacgctggcc atgcaggatg gcttggctgg ccggcaaacc aagacccggc
      961 cgctcagcta tttgcaaatg tttacaagaa tgcatcgtct ccgagagctc ttcgcggatt
     1021 ggcaaccaat gtcgccaact acaacgggtg gaacattacc agccccccat cgtacacgca
     1081 aggcaacgct gtctacaacg agaagctgta catccacgct attggacctc ttcttgccaa
     1141 tcacggctgg tccaacgcct tcttcatcac tgatcaaggt cgatcgggaa agcagcctac
     1201 cggacagcaa cagtggggag actggtgcaa tgtgatcggc accggatttg gtattcgccc
     1261 atccgcaaac actggggact cgttgctgga ttcgtttgtc tgggtcaagc caggcggcga
     1321 gtgtgacggc accagcgaca gcagtgcgcc acgatttgac tcccactgtg cgctcccaga
     1381 tgccttgcaa ccggcgcctc aagctggtgc ttggttccaa gcctactttg tgcagcttct
     1441 cacaaacgca aacccatcgt tcctgtaagg ctttcgtgac cgggcttcaa acaatgatgt
     1501 gcgatggtgt ggttcccggt tggcggagtc tttgtctact ttggttgtct gtcgcaggtc
     1561 ggtagaccgc aaatgagcaa ctgatggatt gttgccagcg atactataat tcacatggat
     1621 ggtctttgtc gatcagtagc tagtgagaga gagagaacat ctatccacaa tgtcgagtgt
     1681 ctattagaca tactccgaga ataaagtcaa ctgtgtctgt gatct
//
```

gDNA ou cDNA???

>**gDNA – sequência Problema**

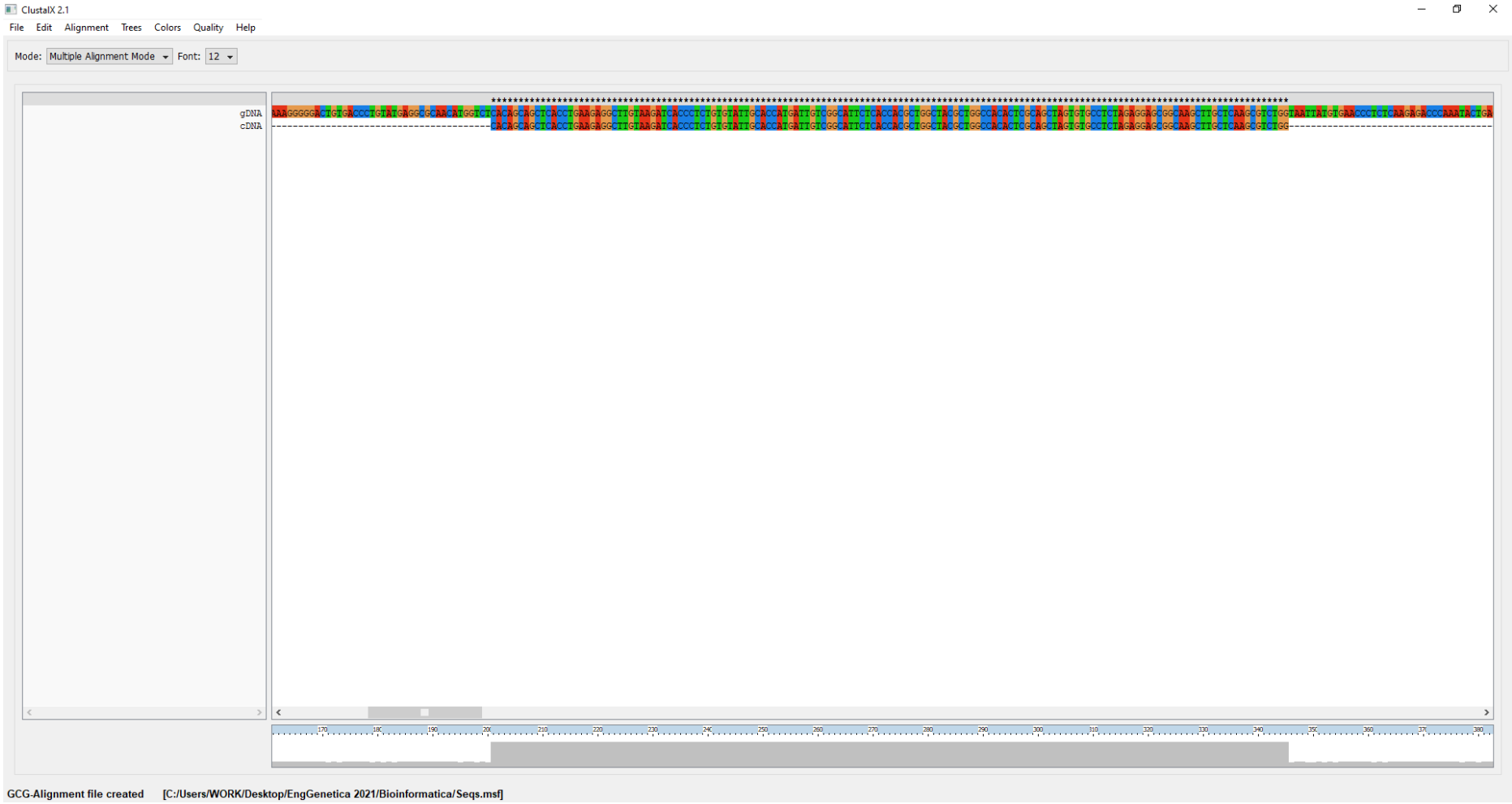
CTTCTTTATTGGGTAATATACAGCCAGGCGGGGATGAAGCTCATTAGCCGCCACTCAAGGCTATACAATGTTGCCAACTC
TCCGGGCTTTATCCTGTGCTCCCGAATACCACATCGTGATGATGCTTCAGCGCACGGAAGTCACAGACACCGCCTGTATA
AAAGGGGGACTGTGACCCTGTATGAGGCGCAACATGGTCTCACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCCTCT
GTGTATTGCACCATGATTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCCTCTAGAGGA
GCGGCAAGCTTGCTCAAGCGTCTGGTAATTATGTGAACCCTCTCAAGAGACCCAAATACTGAGATATGTCAAGGGGCCAA
TGTGGTGGCCAGAATTGGTCGGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCA
GTGTCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTCGACGACTTCTCGAGTATCCCCCACAACATCCC
GGTCGAGCTCCGCGACGCCTCCACCTGGTTCTACTACTACCAGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCA
GGCAACCCTTTTGTTGGGGTCACTCCTTGGGCCAATGCATATTACGCCTCTGAAGTTAGCAGCCTCGCTATTCCTAGCTT
GACTGGAGCCATGGCCACTGCTGCAGCAGCTGTCGCAAAGGTTCCCTCTTTTATGTGGCTGTAGGTCCTCCCGGAACCAA
GGCAATCTGTTACTGAAGGCTCATCATTCACTGCAGAGATACTCTTGACAAGACCCCTCTCATGGAGCAAACCTTGGCCG
ACATCCGCACCGCCAACAAGAATGGCGGTAACTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATCGCGATTGCGCT
GCCCTTGCCTCGAATGGCGAATACTCTATTGCCGATGGTGGCGTCGCCAAATATAAGAACTATATCGACACCATTCGTCA
AATTGTCGTGGAATATTCCGATATCCGGACCCTCCTGGTTATTGGTATGAGTTTAAACACCTGCCTCCCCCCCCCCTTCC
CTTCCTTTCCCGCCGGCATCTTGTCGTTGTGCTAACTATTGTTCCCTCTTCCAGAGCCTGACTCTCTTGCCAACCTGGTG
ACCAACCTCGGTACTCCAAAGTGTGCCAATGCTCAGTCAGCCTACCTTGAGTGCATCAACTACGCCGTCACACAGCTGAA
CCTTCCAAATGTTGCGATGTATTTGGACGCTGGCCATGCAGGATGGCTTGGCTGGCCGGCAAACCAAGACCCGGCCGCTC
AGCTATTTGCAAATGTTTACAAGAATGCATCGTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTCGCCAACTACAAC
GGGTGGAACATTACCAGCCCCCCATCGTACACGCAAGGCAACGCTGTCTACAACGAGAAGCTGTACATCCACGCTATTGG
ACCTCTTCTTGCCAATCACGGCTGGTCCAACGCCTTCTTCATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGAC
AGCAACAGTGGGGAGACTGGTGCAATGTGATCGGCACCGGATTTGGTATTCGCCCATCCGCAAACACTGGGGACTCGTTG
CTGGATTCGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACCAGCGACAGCAGTGCGCCACGATTTGACTCCCA
CTGTGCGCTCCCAGATGCCTTGCAACCGGCGCCTCAAGCTGGTGCTTGGTTCCAAGCCTACTTTGTGCAGCTTCTCACAA
ACGCAAACCCATCGTTCCTGTAAGGCTTTCGTGACCGGGCTTCAAACAATGATGTGCGATGGTGTGGTTCCCGGTTGGCG
GAGTCTTTGTCTACTTTGGTTGTCTGTCGCAGGTCGGTAGACCGCAAATGAGCAACTGATGGATTGTTGCCAGCGATACT
ATAATTCACATGGATGGTCTTTGTCGATCAGTAGCTAGTGAGAGAGAGAGAACATCTATCCACAATGTCGAGTGTCTATT
AGACATACTCCGAGAATAAAGTCAACTGTGTCTGTGATCTAAAGATCGATTCGGCAGTCGAGTAGCGTATAACAACTCCG
AGTACCAGCAAAAGCACGTCGTGACAGGAGCAGGGCTTTGCCAACTGCGCAACCTTGCTTGAATGAGGATACACGGGGTG
CAACATGGCTGTACTGATCCATCGCAACCAAAATTTCTGTTTATAGATCAAGCTGGTAGATTCCAATTACTCCACCTCTT

>cDNA

CACAGCAGCT CACCTGAAGA GGCTTGTAAG ATCACCCTCT GTGTATTGCA CCATGATTGTCGGCATTCTC ACCACGCTGG CTACGCTGGC
CACACTCGCA GCTAGTGTGC CTCTAGAGGAGCGGCAAGCT TGCTCAAGCG TCTGGGGCCA ATGTGGTGGC CAGAATTGGT CGGGTCCGAC
TTGCTGTGCT TCCGGAAGCA CATGCGTCTA CTCCAACGAC TATTACTCCC AGTGTCTTCCCGGCGCTGCA AGCTCAAGCT CGTCCACGCG
CGCCGCGTCG ACGACTTCTC GAGTATCCCCCACAACATCC CGGTCGAGCT CCGCGACGCC TCCACCTGGT TCTACTACTA CCAGAGTACC
TCCAGTCGGA TCGGGAACCG CTACGTATTC AGGCAACCCT TTTGTTGGGG TCACTCCTTGGGCCAATGCA TATTACGCCT CTGAAGTTAG
CAGCCTCGCT ATTCCTAGCT TGACTGGAGCCATGGCCACT GCTGCAGCAG CTGTCGCAAA GGTTCCCTCT TTTATGTGGC TAGATACTCT
TGACAAGACC CCTCTCATGG AGCAAACCTT GGCCGACATC CGCACCGCCA ACAAGAATGGCGGTAACTAT GCCGGACAGT TTGTGGTGTA
TGACTTGCCG GATCGCGATT GCGCTGCCCTTGCCTCGAAT GGCGAATACT CTATTGCCGA TGGTGGCGTC GCCAAATATA AGAACTATAT
CGACACCATT CGTCAAATTG TCGTGGAATA TTCCGATATC CGGACCCTCC TGGTTATTGAGCCTGACTCT CTTGCCAACC TGGTGACCAA
CCTCGGTACT CCAAAGTGTG CCAATGCTCAGTCAGCCTAC CTTGAGTGCA TCAACTACGC CGTCACACAG CTGAACCTTC CAAATGTTGC
GATGTATTTG GACGCTGGCC ATGCAGGATG GCTTGGCTGG CCGGCAAACC AAGACCCGGCCGCTCAGCTA TTTGCAAATG TTTACAAGAA
TGCATCGTCT CCGAGAGCTC TTCGCGGATTGGCAACCAAT GTCGCCAACT ACAACGGGTG GAACATTACC AGCCCCCCAT CGTACACGCA
AGGCAACGCT GTCTACAACG AGAAGCTGTA CATCCACGCT ATTGGACCTC TTCTTGCCAATCACGGCTGG TCCAACGCCT TCTTCATCAC
TGATCAAGGT CGATCGGGAA AGCAGCCTACCGGACAGCAA CAGTGGGGAG ACTGGTGCAA TGTGATCGGC ACCGGATTTG GTATTCGCCC
ATCCGCAAAC ACTGGGGACT CGTTGCTGGA TTCGTTTGTC TGGGTCAAGC CAGGCGGCGAGTGTGACGGC ACCAGCGACA GCAGTGCGCC
ACGATTTGAC TCCCACTGTG CGCTCCCAGATGCCTTGCAA CCGGCGCCTC AAGCTGGTGC TTGGTTCCAA GCCTACTTTG TGCAGCTTCT
CACAAACGCA AACCCATCGT TCCTGTAAGG CTTTCGTGAC CGGGCTTCAA ACAATGATGTGCGATGGTGT GGTTCCCGGT TGGCGGAGTC
TTTGTCTACT TTGGTTGTCT GTCGCAGGTCGGTAGACCGC AAATGAGCAA CTGATGGATT GTTGCCAGCG ATACTATAAT TCACATGGAT
GGTCTTTGTC GATCAGTAGC TAGTGAGAGA GAGAGAACAT CTATCCACAA TGTCGAGTGTCTATTAGACA TACTCCGAGA ATAAAGTCAA
CTGTGTCTGT GATCT

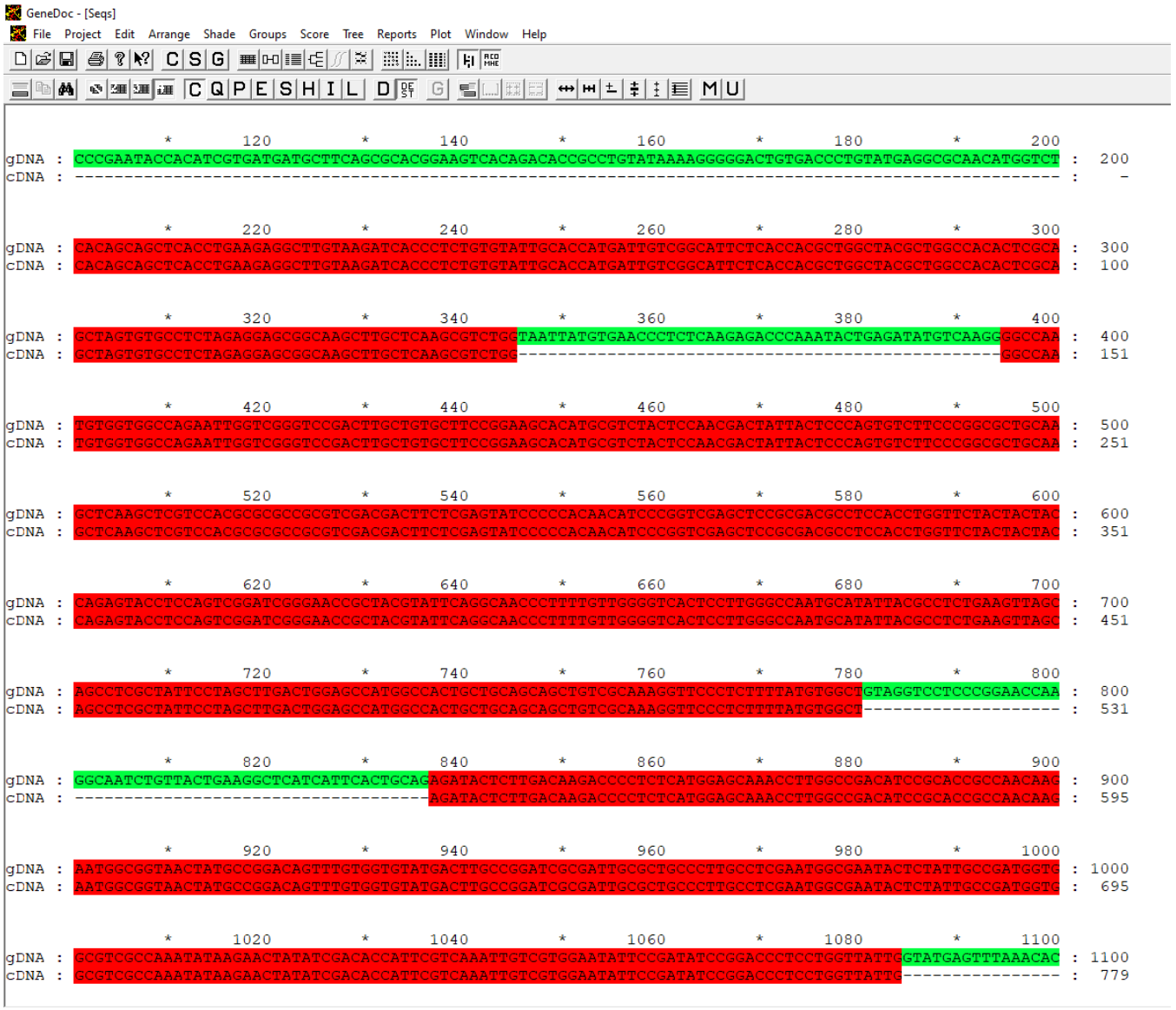
Alinhamento de sequências de DNA – gDNA x cDNA

Ferramenta: Clustal X



Visualização do Alinhamento de sequências de DNA

Ferramenta: GeneDoc

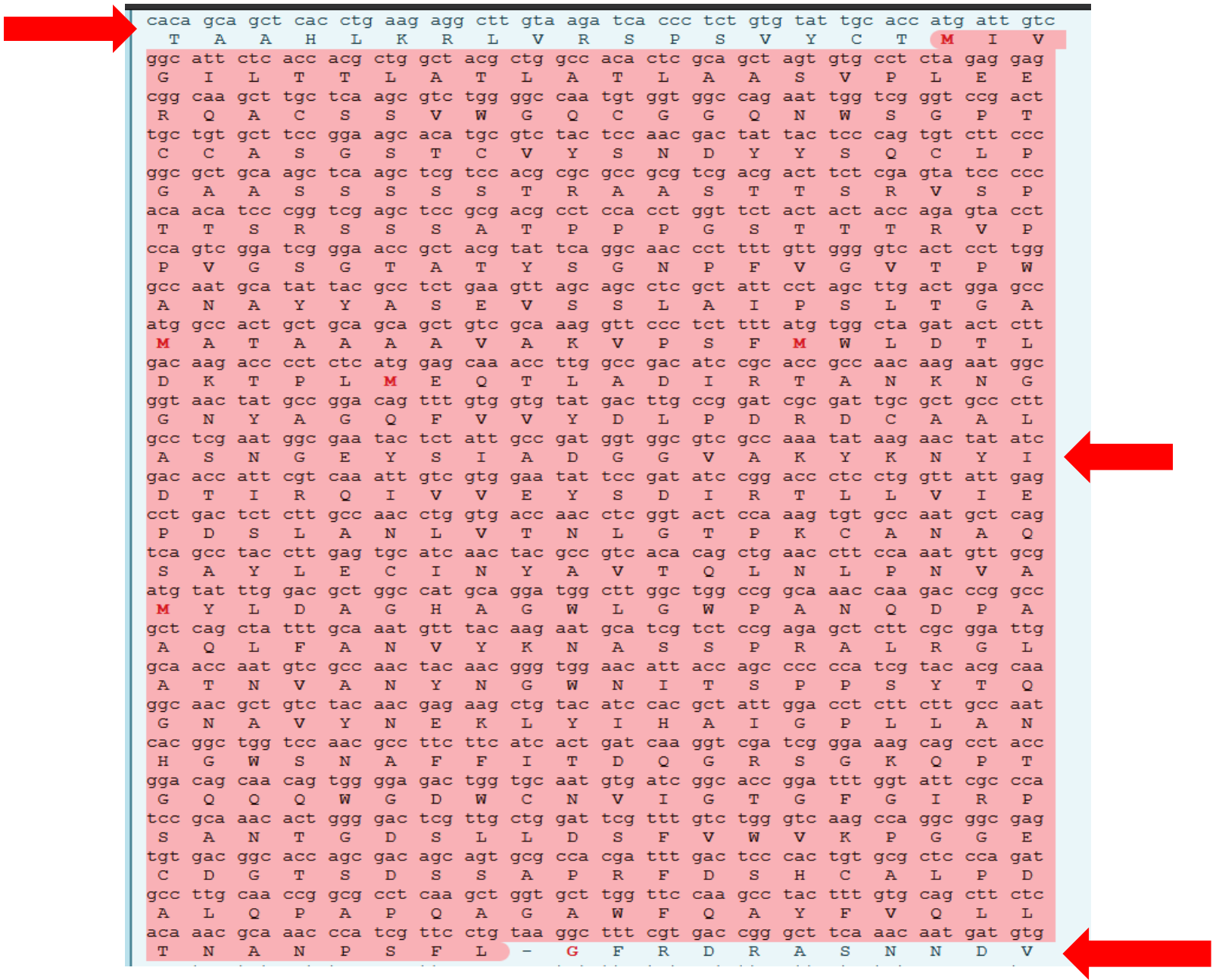


>File >Export >-All sequences – Clustal (aln)

```
CLUSTAL W(1.60) multiple sequence alignment


gDNA            CTTCTTTATTGGGTAATATACAGCCAGGCGGGGATGAAGCTCATTAGCCGCCACTCAAGG
cDNA            ------------------------------------------------------------

gDNA            CTATACAATGTTGCCAACTCTCCGGGCTTTATCCTGTGCTCCCGAATACCACATCGTGAT
cDNA            ------------------------------------------------------------

gDNA            GATGCTTCAGCGCACGGAAGTCACAGACACCGCCTGTATAAAAGGGGGACTGTGACCCTG
cDNA            ------------------------------------------------------------

gDNA            TATGAGGCGCAACATGGTCTCACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCCTCT
cDNA            --------------------CACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCCTCT

gDNA            GTGTATTGCACCATGATTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCA
cDNA            GTGTATTGCACCATGATTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCA

gDNA            GCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGGTAATTATGTGAACCC
cDNA            GCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGG---------------

gDNA            TCTCAAGAGACCCAAATACTGAGATATGTCAAGGGGCCAATGTGGTGGCCAGAATTGGTC
cDNA            ----------------------------------GGCCAATGTGGTGGCCAGAATTGGTC

gDNA            GGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCA
cDNA            GGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCA

gDNA            GTGTCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTCGACGACTTCTCG
cDNA            GTGTCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTCGACGACTTCTCG
```

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

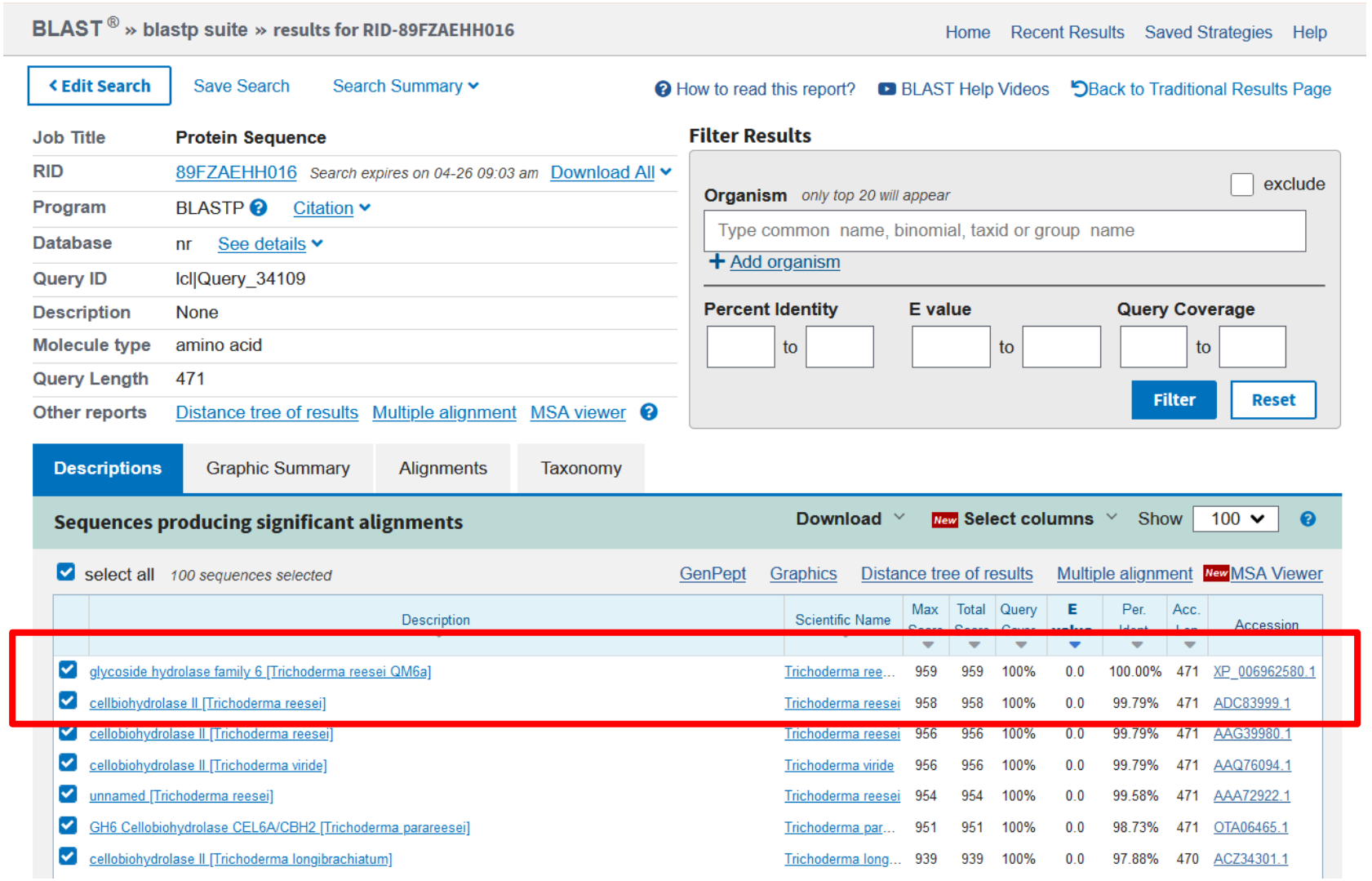
Ferramenta: Translate - Expasy




CTTCTTTATTGGGTAATATACAGCCAGGCGGGGATGAAGCTCATTAGCCGCCACTCAAGG
CTATACAATGTTGCCAACTCTCCGGGCTTTATCCTGTGCTCCCGAATACCACATCGTGAT
GATGCTTCAGCGCACGGAAGTCACAGACACCGCCTGTATAAAAGGGGGACTGTGACCCTG
TATGAGGCGCAACATGGTCTCACAGCAGCTCACCTGAAGAGGCTTGTAAGATCACCCTCT
GTGTATTGCACCATGATTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCA
GCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGGTAATTATGTGAACCC
TCTCAAGAGACCCAAATACTGAGATATGTCAAGGGGCCAATGTGGTGGCCAGAATTGGTC
GGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCA
GTGTCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTCGACGACTTCTCG
AGTATCCCCCACAACATCCCGGTCGAGCTCCGCGACGCCTCCACCTGGTTCTACTACTAC
CAGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCAACCCTTTTGTTGGGGT
CACTCCTTGGGCCAATGCATATTACGCCTCTGAAGTTAGCAGCCTCGCTATTCCTAGCTT
GACTGGAGCCATGGCCACTGCTGCAGCAGCTGTCGCAAAGGTTCCCTCTTTTATGTGGCT
GTAGGTCCTCCCGGAACCAAGGCAATCTGTTACTGAAGGCTCATCATTCACTGCAGAGAT
ACTCTTGACAAGACCCCTCTCATGGAGCAAACCTTGGCCGACATCCGCACCGCCAACAAG
AATGGCGGTAACTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATCGCGATTGCGCT
GCCCTTGCCTCGAATGGCGAATACTCTATTGCCGATGGTGGCGTCGCCAAATATAAGAAC
TATATCGACACCATTCGTCAAATTGTCGTGGAATATTCCGATATCCGGACCCTCCTGGTT
ATTGGTATGAGTTTAAACACCTGCCTCCCCCCCCCCTTCCCTTCCTTTCCCGCCGGCATC
TTGTCGTTGTGCTAACTATTGTTCCCTCTTCCAGAGCCTGACTCTCTTGCCAACCTGGTG
ACCAACCTCGGTACTCCAAAGTGTGCCAATGCTCAGTCAGCCTACCTTGAGTGCATCAAC
TACGCCGTCACACAGCTGAACCTTCCAAATGTTGCGATGTATTTGGACGCTGGCCATGCA
GGATGGCTTGGCTGGCCGGCAAACCAAGACCCGGCCGCTCAGCTATTTGCAAATGTTTAC
AAGAATGCATCGTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTCGCCAACTACAAC
GGGTGGAACATTACCAGCCCCCCATCGTACACGCAAGGCAACGCTGTCTACAACGAGAAG
CTGTACATCCACGCTATTGGACCTCTTCTTGCCAATCACGGCTGGTCCAACGCCTTCTTC
ATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGG
TGCAATGTGATCGGCACCGGATTTGGTATTCGCCCATCCGCAAACACTGGGGACTCGTTG
CTGGATTCGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACCAGCGACAGCAGT
GCGCCACGATTTGACTCCCACTGTGCGCTCCCAGATGCCTTGCAACCGGCGCCTCAAGCT
GGTGCTTGGTTCCAAGCCTACTTTGTGCAGCTTCTCACAAACGCAAACCCATCGTTCCTG
TAAGGCTTTCGTGACCGGGCTTCAAACAATGATGTGCGATGGTGTGGTTCCCGGTTGGCG
GAGTCTTTGTCTACTTTGGTTGTCTGTCGCAGGTCGGTAGACCGCAAATGAGCAACTGAT
GGATTGTTGCCAGCGATACTATAATTCACATGGATGGTCTTTGTCGATCAGTAGCTAGTG
AGAGAGAGAGAACATCTATCCACAATGTCGAGTGTCTATTAGACATACTCCGAGAATAAA
GTCAACTGTGTCTGTGATCTAAAGATCGATTCGGCAGTCGAGTAGCGTATAACAACTCCG
AGTACCAGCAAAAGCACGTCGTGACAGGAGCAGGGCTTTGCCAACTGCGCAACCTTGCTT
GAATGAGGATACACGGGGTGCAACATGGCTGTACTGATCCATCGCAACCAAAATTTCTGT

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

MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPV
GSGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASN
GEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANV
YKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSL
LDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL



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: $C_{2194}H_{3354}N_{594}O_{691}S_{17}$
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.

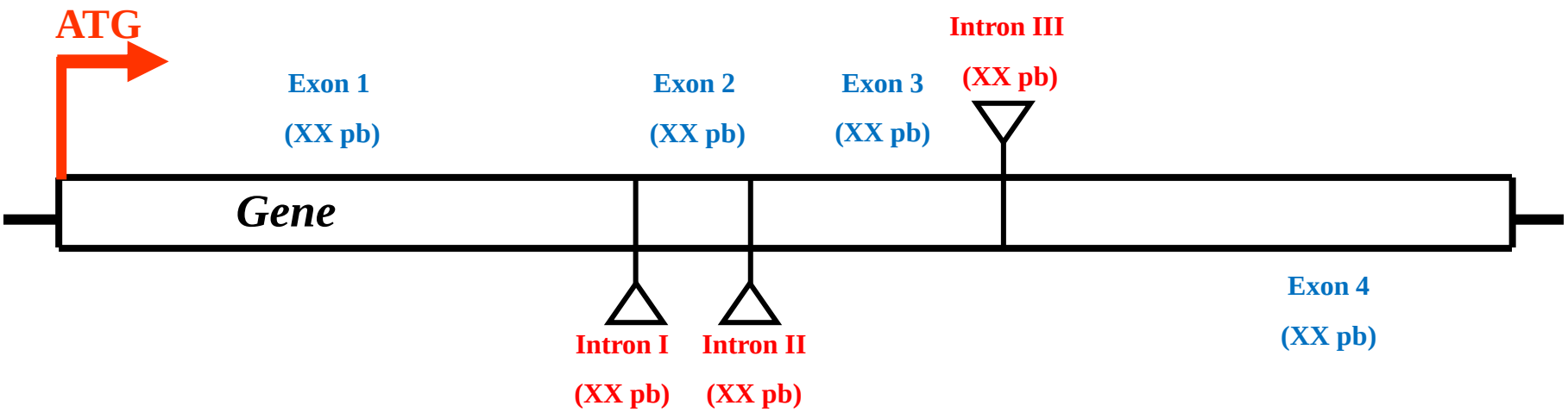


Figura X. Estrutura gênica (A) e sequência de nucleotídeos e de aminoácidos (B) prevista para o *locus* que contém o gene XXXXX do fungo filamentoso *Trichoderma reesei*. Os códons de iniciação e terminação estão indicados em negrito. Os íntrons identificados na sequência codificadora das proteínas estão 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.

```
                AGCATCATCATGCAGCTCCCATAGCTAACCGCTTCTTCTGCCCCGCCATTCAGGAGCAGCTCGCCTTGTTCATTTATCCTCACCTTTGCC
-101            TCGTAGTAGTACGTCGAGGGTATCGATTGGCGAAGAAGACGGGGCGGTAAGTCCTCGTCGAGCGGAACAAGTAAATAGGAGTGGAAACGG

                CACACCAAATCATGGCTCTCAACCTCACCTCGTCGGCTCGAGCCCTGCGCTCCTTCAAGgtcagtcgagtctcctcttatagccttctat
-11             GTGTGGTTTAG M  A  L  N  L  T  S  S  A  R  A  L  R  S  F  K                                16

80              cgagacccatttgacatagactgacactcctcccgtctacagCCCTACACCCGCGCCGCCCTCCTCGCCAACGCCGCGCGATGCTACTCA
                     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             ACCGCTGAGCCCGACCTCAAGACGACGCTCAAGAGCGTCATCCCTGAGAAGCGCGAGCTGCTGAAGAAGGTCAAGGCCCACGGCAGCAAG
                T  A  E  P  D  L  K  T  T  L  K  S  V  I  P  E  K  R  E  L  L  K  K  V  K  A  H  G  S  K   62

260             GTCATTGGCGAGGTCAAGGTTGAGAACACCATTGGCGGCATGCGCGGGCTCAAGGCCATGGTCTGGGAGGGCTCCGTGCTCGACCCCAAC
                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             GAGGGCATCCGCTTCCACGGCCGCACCATCAAGGACTGCCAGAAGGAGCTGCCCAAGGGCAAGACGGGCACCGAGATGCTGCCCGAGGCC
                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             ATGTTTTGGCTGCTGTTGACCGGCCAGGTGCCCTCCGTCAACCAGGTCCGCCAGTTCTCCCGCGAGCTGGCCTCCCAGACCCAGATCCCC
                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             GCCTTCGTCAACAGGATGCTCGACGATTTCCCCAAGGATCTGCACCCCATGACCCAGTTTGCCATTGCCGTCTCGGCCCTCAACTACGAG
                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             TCCAAGTTTGCAAAGGCCTACGAGAAGGGCCTCGCCAAGGCCGACTACTGGGAGCCCACCTTTGACGACAGCATCTCGCTGCTCGCCAAG
                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             CTGCCCACCATTGCCGCCAAGATCTACCAGAACTCTTACCGCGGCGGCGGCGCCCTCCCTGCCGAGGTCGACCTTGAGCAGGATTGGTCA
                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             TACAACTTTGCTGCCATGCTCGGCAAGGGCGGCAAGGAGAACGAGGACTTCCAGGACCTCCTCCGTCTCTACCTTGCCCTCCACGGCGAC
                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             CACGAGGGTGGCAATGTGTCTGCCCACGCCACTCACCTTGTTGGTAGTGCCCTGAGTGACCCCTTCCTGTCTTACAGCGCTGGTCTCCAG
                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             GGTCTGGCCGGTCCTCTTCACGGgtaagctgtccacttatacttacaatcaatcttcctatacaaacgttgattgctaatcgttcgcgct
                G  L  A  G  P  L  H  G                    I N T R O N   I I                              310

1070            tatgacgatagACTTGCCGCCCAGGAAGTTCTCCGCTGGATCCTGCAGATGAAGGAGGCCATCCCCGCCAACTACACCGAGCAGGATGTC
                          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            CACGACTACCTCTGGTCCACCCTCAACTCGGGCCGCGTCGTGCCCGGCTACGGACACGCCGTTCTGCGCAAGCCCGACCCTCGATTCGAG
                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            GCTCTCATGGACTATGCCGCTTCCCGCCCCGCGATTGCCAAGGACCCCGTCTTCCAGCTGGTTGAGAAGAACAGCCGCATCGCCCCCGAG
                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            GTGCTCAAGAAGCACGGCAAGACCAAGAACCCCTACCCCAACGTCGACAGCAGCTCCGGCGTCCTCTTCCACCACTACGGCTTCCACGAG
                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            ACGCTCTACTACACGGCCACCTTTGGTGTTTCGCGTGGTCTCGGTCCTCTGGCTCAGCTCATCTGGGACCGCGCCCTGGGTCTGCCCATT
                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
GGGGCCAATGTGGTGGCCAGAATTGGTCGGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTG
TCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTCGACGACTTCTCGAGTATCCCCCACAACATCCCGGTCGAGCTCCGCG
ACGCCTCCACCTGGTTCTACTACTACCAGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCAACCCTTTTGTTGGGGTCACTCC
TTGGGCCAATGCATATTACGCCTCTGAAGTTAGCAGCCTCGCTATTCCTAGCTTGACTGGAGCCATGGCCACTGCTGCAGCAGCTGTCGCAA
AGGTTCCCTCTTTTATGTGGCTAGATACTCTTGACAAGACCCCTCTCATGGAGCAAACCTTGGCCGACATCCGCACCGCCAACAAGAATGGC
GGTAACTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATCGCGATTGCGCTGCCCTTGCCTCGAATGGCGAATACTCTATTGCCGATGG
TGGCGTCGCCAAATATAAGAACTATATCGACACCATTCGTCAAATTGTCGTGGAATATTCCGATATCCGGACCCTCCTGGTTATTGAGCCTGA
CTCTCTTGCCAACCTGGTGACCAACCTCGGTACTCCAAAGTGTGCCAATGCTCAGTCAGCCTACCTTGAGTGCATCAACTACGCCGTCACAC
AGCTGAACCTTCCAAATGTTGCGATGTATTTGGACGCTGGCCATGCAGGATGGCTTGGCTGGCCGGCAAACCAAGACCCGGCCGCTCAGC
TATTTGCAAATGTTTACAAGAATGCATCGTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTCGCCAACTACAACGGGTGGAACATTACC
AGCCCCCCATCGTACACGCAAGGCAACGCTGTCTACAACGAGAAGCTGTACATCCACGCTATTGGACCTCTTCTTGCCAATCACGGCTGGTC
CAACGCCTTCTTCATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGGTGCAATGTGATCGGCAC
CGGATTTGGTATTCGCCCATCCGCAAACACTGGGGACTCGTTGCTGGATTCGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCAC
CAGCGACAGCAGTGCGCCACGATTTGACTCCCACTGTGCGCTCCCAGATGCCTTGCAACCGGCGCCTCAAGCTGGTGCTTGGTTCCAAGC
CTACTTTGTGCAGCTTCTCACAAACGCAAACCCATCGTTCCTGTAA

- Conversão a DNA dupla fita

Ferramenta: Conversor

ATGATTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGGGGCCAAT
TACTAACAGCCGTAAGAGTGGTGCGACCGATGCGACCGGTGTGAGCGTCGATCACACGGAGATCTCCTCGCCGTTCGAACGAGTTCGCAGACCCCGGTTA

GTGGTGGCCAGAATTGGTCGGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTGTCTTCCCGGCGCTGCAAG
CACCACCGGTCTTAACCAGCCCAGGCTGAACGACACGAAGGCCTTCGTGTACGCAGATGAGGTTGCTGATAATGAGGGTCACAGAAGGGCCGCGACGTTC

CTCAAGCTCGTCCACGCGCGCCGCGTCGACGACTTCTCGAGTATCCCCCACAACATCCCGGTCGAGCTCCGCGACGCCTCCACCTGGTTCTACTACTACC
GAGTTCGAGCAGGTGCGCGCGGCGCAGCTGCTGAAGAGCTCATAGGGGGTGTTGTAGGGCCAGCTCGAGGCGCTGCGGAGGTGGACCAAGATGATGATGG

AGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCAACCCTTTTGTTGGGGTCACTCCTTGGGCCAATGCATATTACGCCTCTGAAGTTAGCA
TCTCATGGAGGTCAGCCTAGCCCTTGGCGATGCATAAGTCCGTTGGGAAAACAACCCCAGTGAGGAACCCGGTTACGTATAATGCGGAGACTTCAATCGT

GCCTCGCTATTCCTAGCTTGACTGGAGCCATGGCCACTGCTGCAGCAGCTGTCGCAAAGGTTCCCTCTTTTATGTGGCTAGATACTCTTGACAAGACCCC
CGGAGCGATAAGGATCGAACTGACCTCGGTACCGGTGACGACGTCGTCGACAGCGTTTCCAAGGGAGAAAATACACCGATCTATGAGAACTGTTCTGGGG

TCTCATGGAGCAAACCTTGGCCGACATCCGCACCGCCAACAAGAATGGCGGTAACTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGATCGCGATTGC
AGAGTACCTCGTTTGGAACCGGCTGTAGGCGTGGCGGTTGTTCTTACCGCCATTGATACGGCCTGTCAAACACCACATACTGAACGGCCTAGCGCTAACG

GCTGCCCTTGCCTCGAATGGCGAATACTCTATTGCCGATGGTGGCGTCGCCAAATATAAGAACTATATCGACACCATTCGTCAAATTGTCGTGGAATATT
CGACGGGAACGGAGCTTACCGCTTATGAGATAACGGCTACCACCGCAGCGGTTTATATTCTTGATATAGCTGTGGTAAGCAGTTTAACAGCACCTTATAA

CCGATATCCGGACCCTCCTGGTTATTGAGCCTGACTCTCTTGCCAACCTGGTGACCAACCTCGGTACTCCAAAGTGTGCCAATGCTCAGTCAGCCTACCT
GGCTATAGGCCTGGGAGGACCAATAACTCGGACTGAGAGAACGGTTGGACCACTGGTTGGAGCCATGAGGTTTCACACGGTTACGAGTCAGTCGGATGGA

TGAGTGCATCAACTACGCCGTCACACAGCTGAACCTTCCAAATGTTGCGATGTATTTGGACGCTGGCCATGCAGGATGGCTTGGCTGGCCGGCAAACCAA
ACTCACGTAGTTGATGCGGCAGTGTGTCGACTTGGAAGGTTTACAACGCTACATAAACCTGCGACCGGTACGTCCTACCGAACCGACCGGCCGTTTGGTT

GACCCGGCCGCTCAGCTATTTGCAAATGTTTACAAGAATGCATCGTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTCGCCAACTACAACGGGTGGA
CTGGGCCGGCGAGTCGATAAACGTTTACAAATGTTCTTACGTAGCAGAGGCTCTCGAGAAGCGCCTAACCGTTGGTTACAGCGGTTGATGTTGCCCACCT

ACATTACCAGCCCCCCATCGTACACGCAAGGCAACGCTGTCTACAACGAGAAGCTGTACATCCACGCTATTGGACCTCTTCTTGCCAATCACGGCTGGTC
TGTAATGGTCGGGGGGGTAGCATGTGCGTTCCGTTGCGACAGATGTTGCTCTTCGACATGTAGGTGCGATAACCTGGAGAAGAACGGTTAGTGCCGACCAG

CAACGCCTTCTTCATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGGTGCAATGTGATCGGCACCGGATTTGGT
GTTGCGGAAGAAGTAGTGACTAGTTCCAGCTAGCCCTTTCGTCGGATGGCCTGTCGTTGTCACCCCTCTGACCACGTTACACTAGCCGTGGCCTAAACCA

ATTCGCCCATCCGCAAACACTGGGGACTCGTTGCTGGATTCGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACCAGCGACAGCAGTGCGCCAC
TAAGCGGGTAGGCGTTTGTGACCCCTGAGCAACGACCTAAGCAAACAGACCCAGTTCGGTCCGCCGCTCACACTGCCGTGGTCGCTGTCGTCACGCGGTG

GATTTGACTCCCACTGTGCGCTCCCAGATGCCTTGCAACCGGCGCCTCAAGCTGGTGCTTGGTTCCAAGCCTACTTTGTGCAGCTTCTCACAAACGCAAA
CTAAACTGAGGGTGACACGCGAGGGTCTACGGAACGTTGGCCGCGGAGTTCGACCACGAACCAAGGTTCGGATGAAACACGTCGAAGAGTGTTTGCGTTT

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
Bsp1407I	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
SalI	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	CAGNNNCTG	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, AflII, AflIII, AgeI, AgsI, AlfI, AloI, ApaI, ApaLI, ApoI, 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, FalI, FauI, FseI, FspAI, HaeIV, HpaI, KpnI, MauBI, MfeI, MluI, MslI, NdeI, NheI, NmeAIII, NotI, OliI, PacI, PasI, PflMI, PfoI, PmaCI, PmeI, PpiI, PpuMI, PshAI, PsiI, PI-PspI, PspXI, PsrI, RsrII, SacII, SanDI, ScaI, PI-SceI, SfiI, SgfI, SgrAI, SmaI, SpeI, SphI, SrfI, Sse8387I, StuI, SwaI, TaqII, TspGWI, Tth111I, VspI, XcmI, XhoII, XmnI

http://www.restrictionmapper.org/

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.

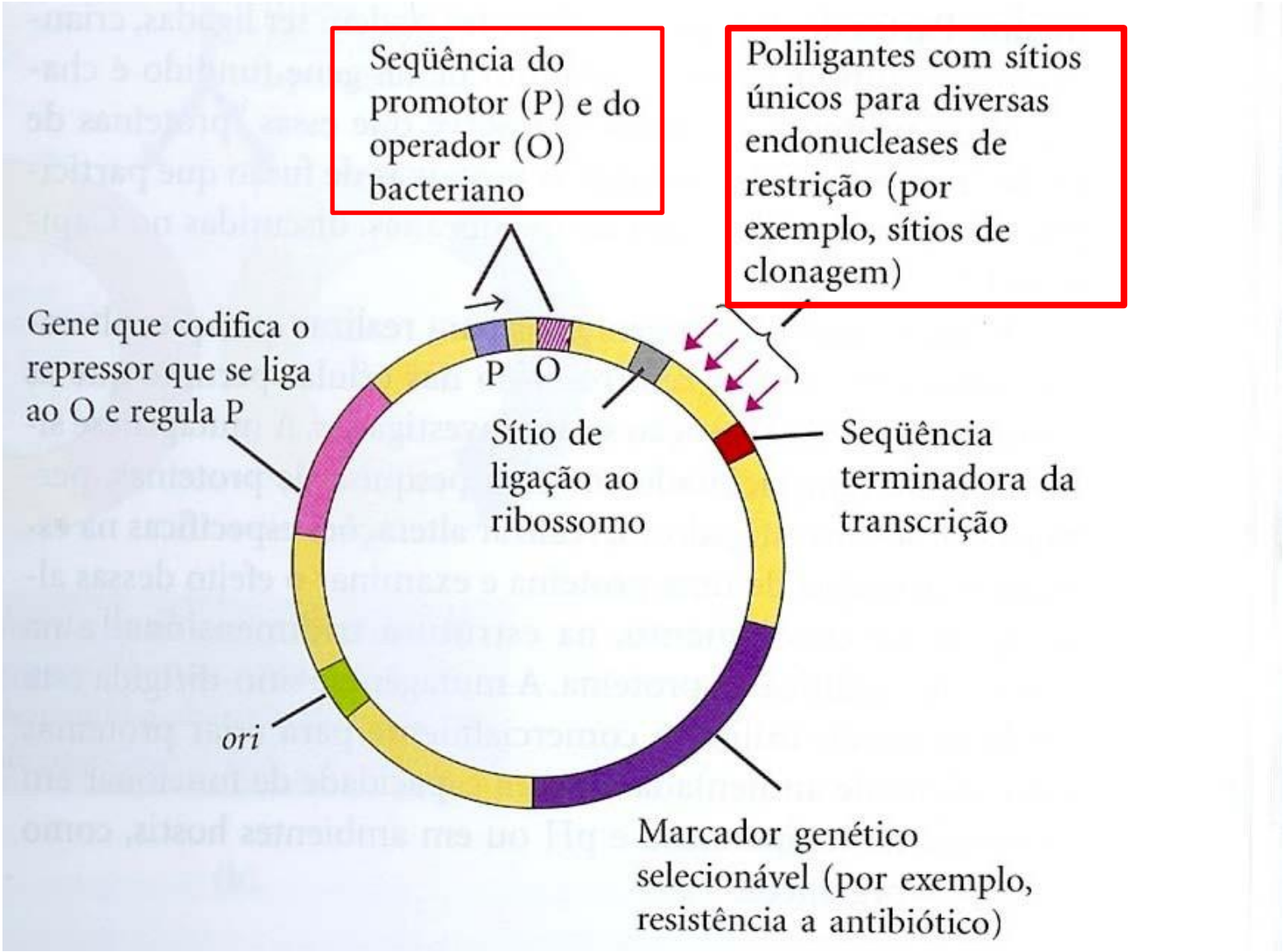
Proteína Recombinante

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.

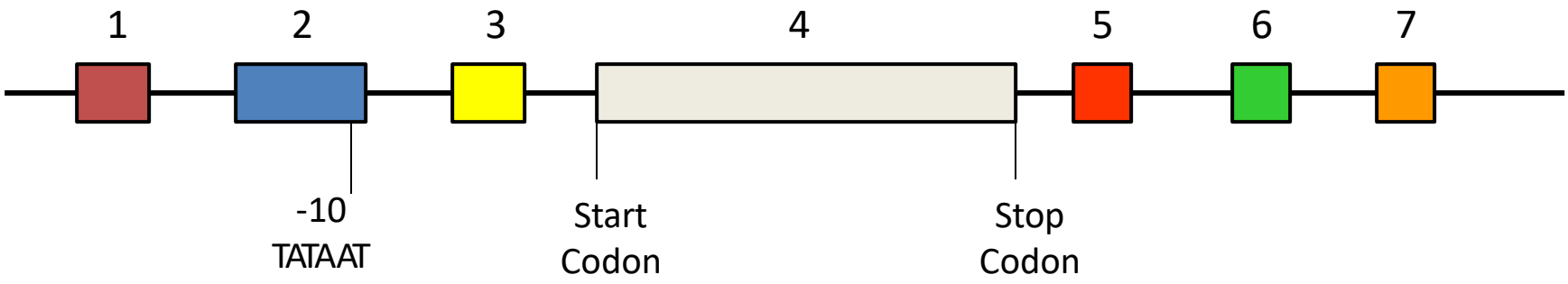
Deve-se obter plasmídeos para clonagem que garantissem um alto nível de expressão da proteína de interesse (Vetor de Expressão)

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

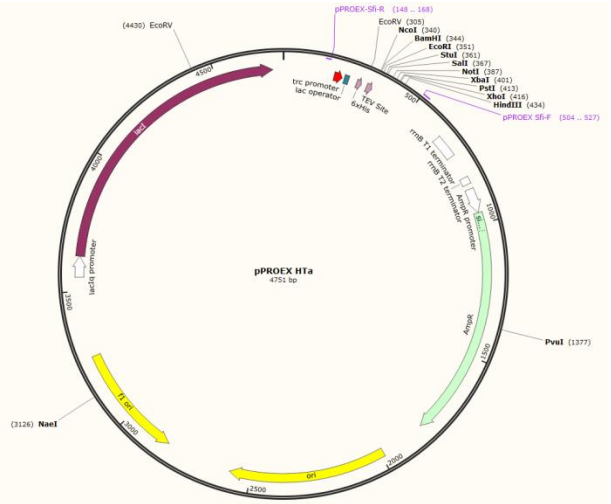


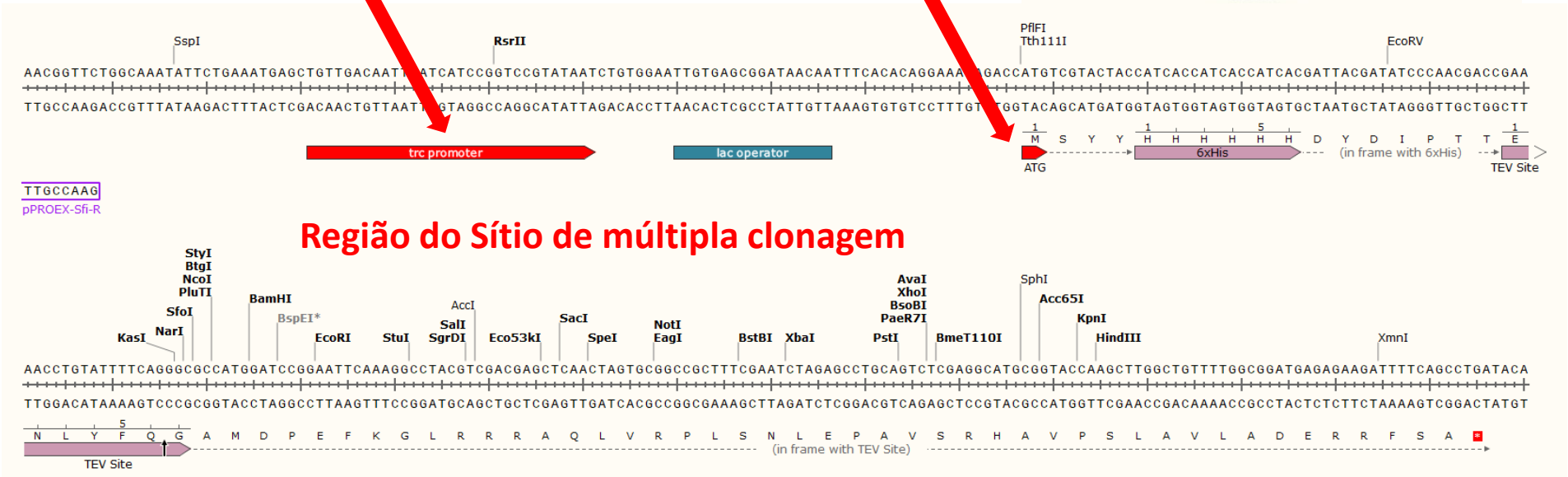
Elementos de um vetor de expressão procariótico



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: Sitio 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 plasmideo.

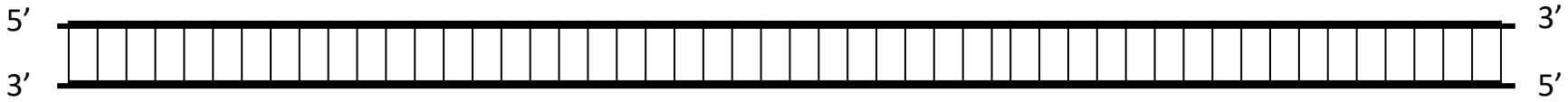
Vetor de expressão pPROEX-Hta

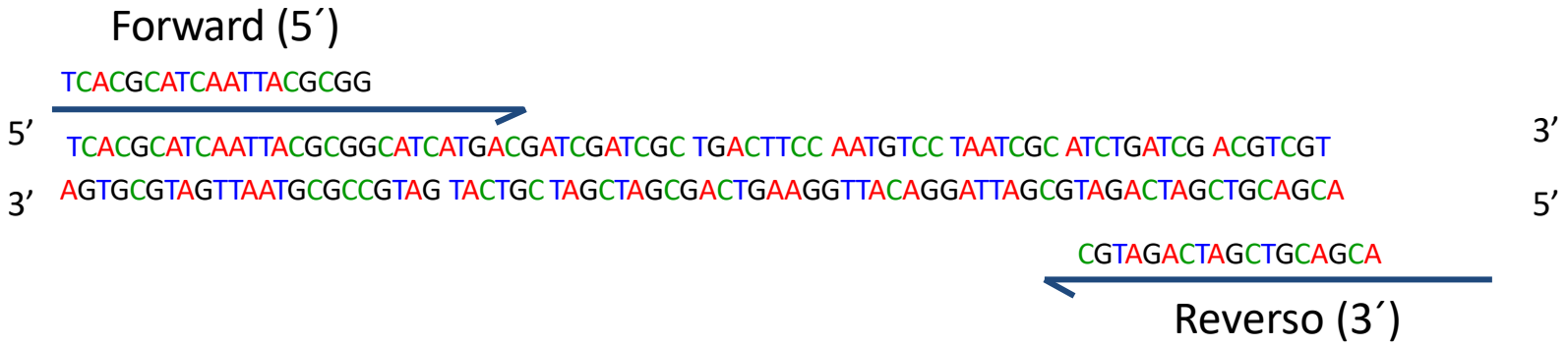




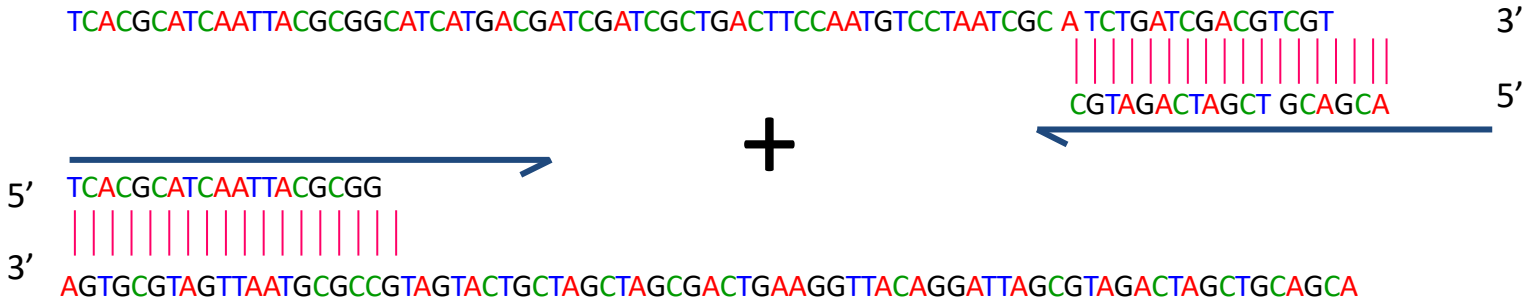
Noncutters: AarI, AatII, AbsI, AclI, AflII, AflIII, AgeI, AgsI, AlfI, AloI, ApaI, ApaLI, ApoI, ArsI, AscI, AsuII, AvrII, **BamHI**, BarI, BbvCI, BdaI, BglI, BglII, Bpu10I, BsaBI, BseSI, BseYI, BspHI, BspMI, BtrI, ClaI, CspCI, DraII, DraIII, DrdI, Eam1105I, EciI, Eco31I, Eco47III, EcoNI, **EcoRI**, FalI, FauI, FseI, FspAI, HaeIV, HpaI, **KpnI**, MauBI, MfeI, MluI, MslI, NdeI, NheI, NmeAIII, NotI, OliI, PacI, PasI, PflMI, PfoI, PmaCI, PmeI, PpiI, PpuMI, PshAI, PsiI, PI-PspI, PspXI, PsrI, RsrII, SacII, SanDI, ScaI, PI-SceI, SfiI, SgfI, SgrAI, SmaI, SpeI, SphI, SrfI, Sse8387I, StuI, SwaI, TaqII, TspGWI, Tth111I, VspI, XcmI, XhoII, XmnI

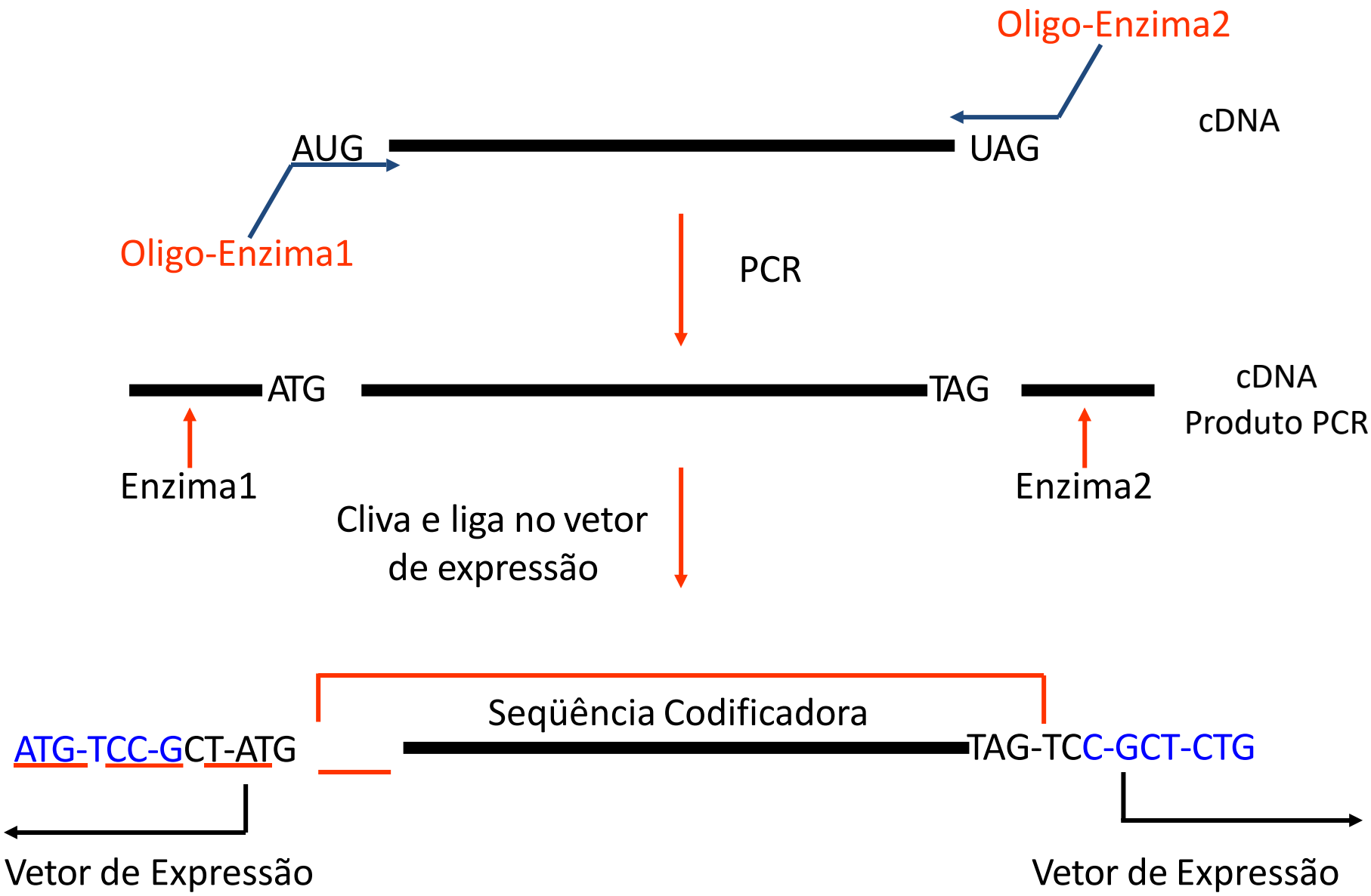
DNA e desenho de Oligonucleotídeos/primers





DNA e alinhamento de Olignucleótidos





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

BamHI – 5’- G GATCC – 3’
3’ – CCTAG G – 5’

KpnI – 5’ - G GTACC - 3’
3’ – CCATG G – 5’

BamHI

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

5’ - ATGATTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCCTCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGGGGCCAAT
3’ - TACTAACAGCCGTAAGAGTGGTGCGACCGATGCGACCGGTGTGAGCGTCGATCACACGGAGATCTCCTCGCCGTTCGAACGAGTTCGCAGACCCCGGTTA

GTGGTGGCCAGAATTGGTCGGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCCCAGTGTCTTCCCGGCGCTGCAAG
CACCACCGGTCTTAACCAGCCCAGGCTGAACGACACGAAGGCCTTCGTGTACGCAGATGAGGTTGCTGATAATGAGGGTCACAGAAGGGCCGCGACGTTC

CTCAAGCTCGTCCACGCGCGCCGCGTCGACGACTTCTCGAGTATCCCCCACAACATCCCGGTCGAGCTCCGCGACGCCTCCACCTGGTTCTACTACTACC
GAGTTCGAGCAGGTGCGCGCGGCGCAGCTGCTGAAGAGCTCATAGGGGGTGTTGTAGGGCCAGCTCGAGGCGCTGCGGAGGTGGACCAAGATGATGATGG

AGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCAACCCTTTTGTTGGGGTCACTCCTTGGGCCAATGCATATTACGCCTCTGAAGTTAGCA
TCTCATGGAGGTCAGCCTAGCCCTTGGCGATGCATAAGTCCGTTGGGAAAACAACCCCAGTGAGGAACCCGGTTACGTATAATGCGGAGACTTCAATCGT

CAACGCCTTCTTCATCACTGATCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGGTGCAATGTGATCGGCACCGGATTTGGT
GTTGCGGAAGAAGTAGTGACTAGTTCCAGCTAGCCCTTTCGTCGGATGGCCTGTCGTTGTCACCCCTCTGACCACGTTACACTAGCCGTGGCCTAAACCA

ATTCGCCCATCCGCAAACACTGGGGACTCGTTGCTGGATTCGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACCAGCGACAGCAGTGCGCCAC
TAAGCGGGTAGGCGTTTGTGACCCCTGAGCAACGACCTAAGCAAACAGACCCAGTTCGGTCCGCCGCTCACACTGCCGTGGTCGCTGTCGTCACGCGGTG

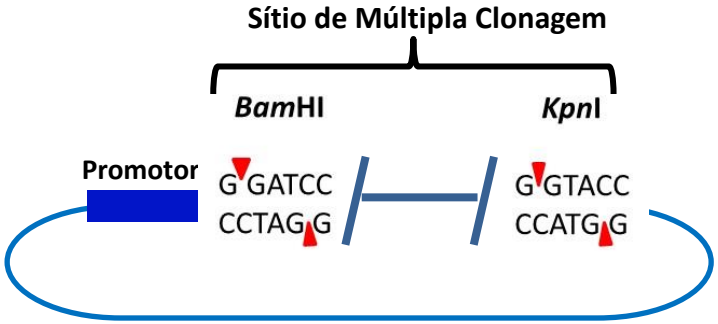
GATTTGACTCCCACTGTGCGCTCCCAGATGCCTTGCAACCGGCGCCTCAAGCTGGTGCTTGGTTCCAAGCCTACTTTGTGCAGCTTCTCACAAACGCAAA
CTAAACTGAGGGTGACACGCGAGGGTCTACGGAACGTTGGCCGCGGAGTTCGACCACGAACCAAGGTTCGGATGAAACACGTCGAAGAGTGTTTGCGTTT

CCCATCGTTCCTGTAA – 3’
GGGTAGCAAGGACATT – 5’

BamHI

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

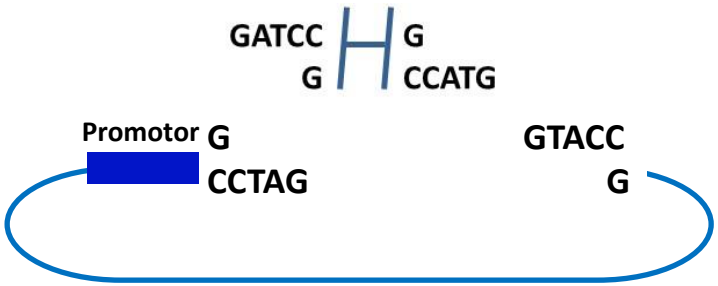
Estratégia de Clonagem no Vetor de Expressão



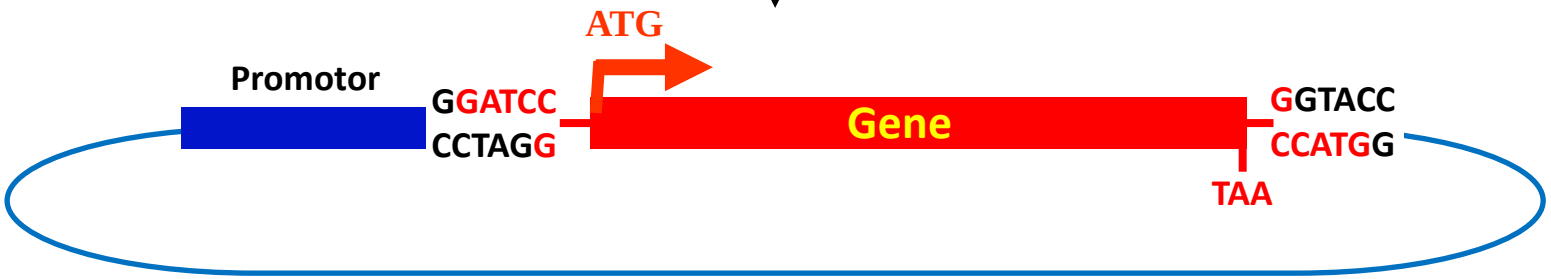
Vetor de Expressão

Gene de Interesse

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



Ligação com enzima DNA Ligase

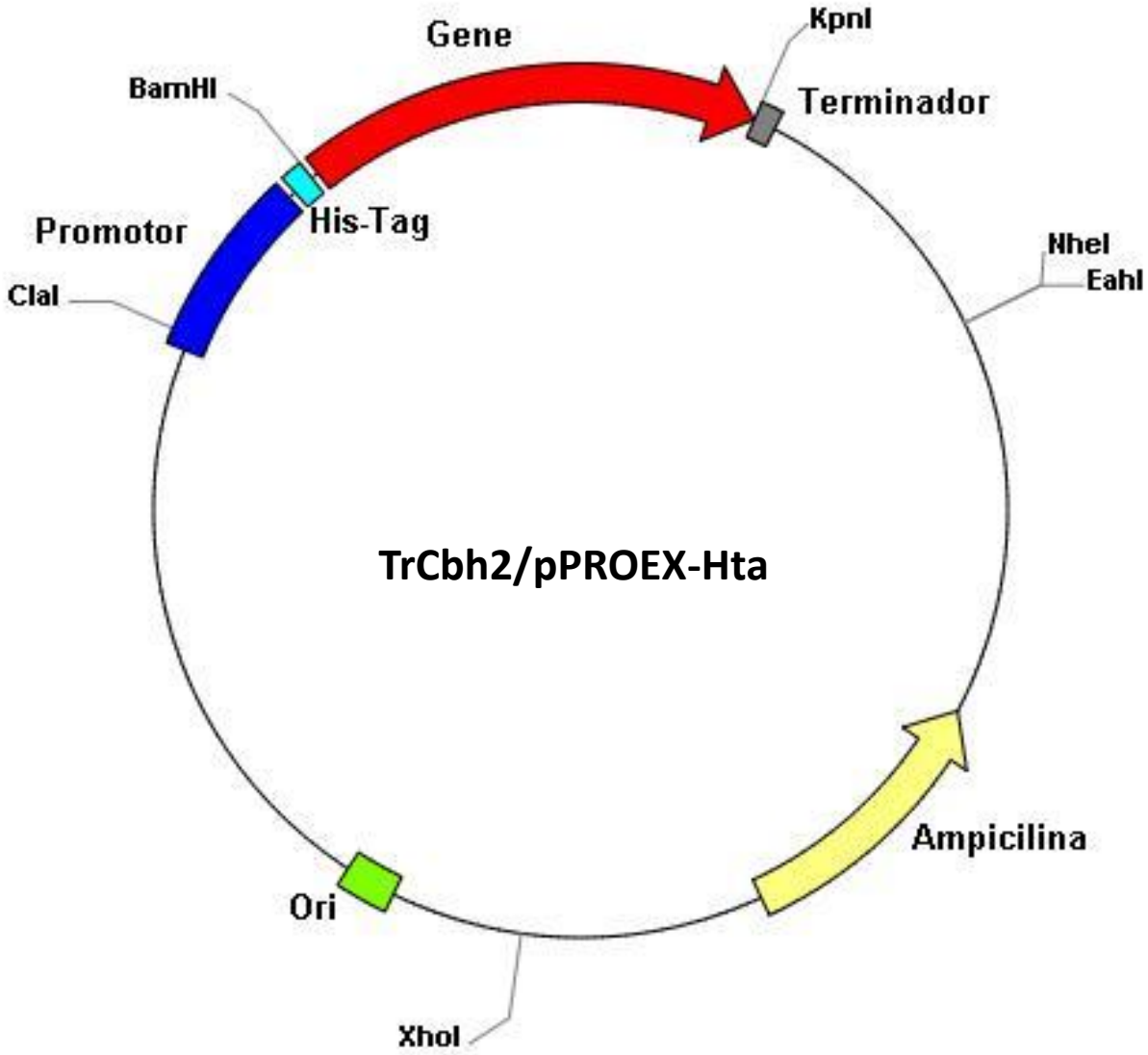


Vetor de Expressão/Gene de Interesse

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

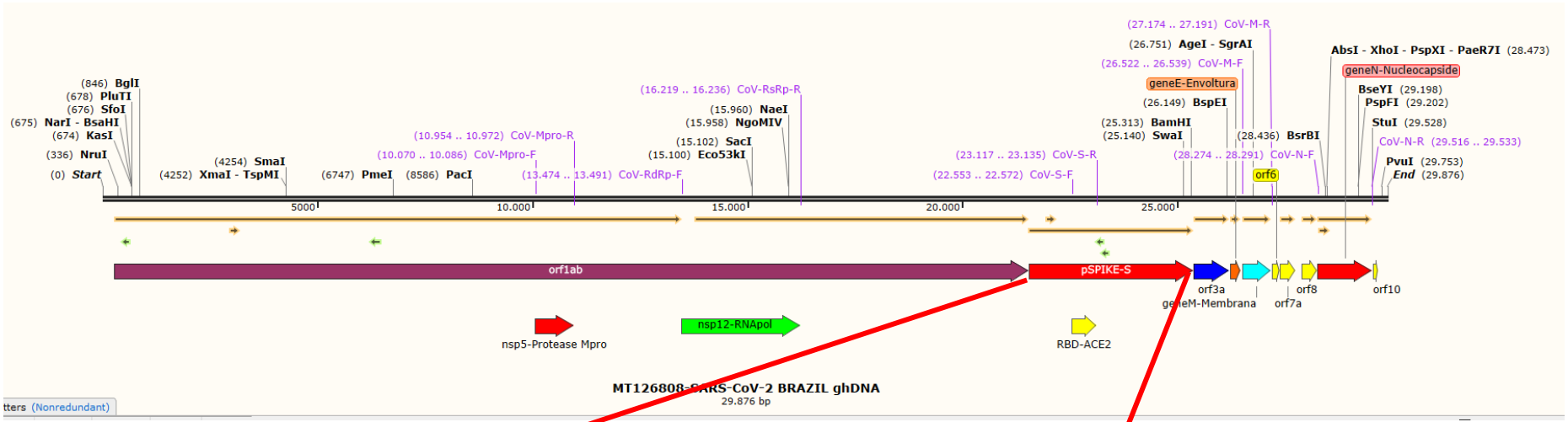
Plasmid Map	
NoeClone 3.0 Demo	Virtual cloning laboratory.
pDRAW32 1.1.129	Scientific software for the molecular biologist
SimVector 4.60 Demo	tool for drawing publication and vector catalog quality maps
ApE 2.0.49	A Plasmid Editor
BVTech Plasmid 5.1 Demo	DNA plasmid drawing software
XPlasMap 0.99	a 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.

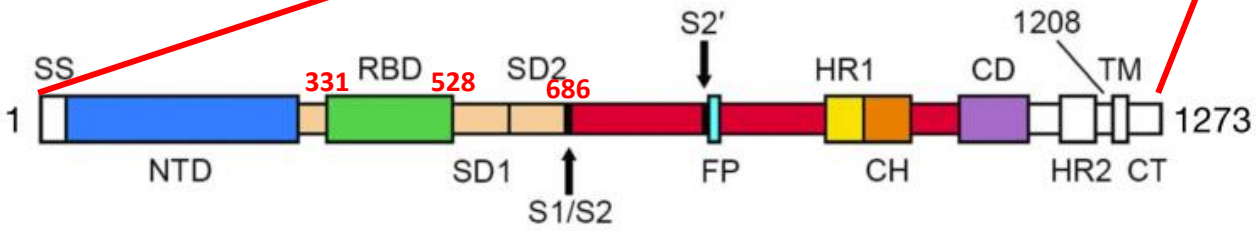
http://en.bio-soft.net/plasmid.html



Mapa esquemático do plasmídeo para expressão recombinante da enzima TrCbh2

GenBank: MT126808 SARS-CoV-2 - Brasil

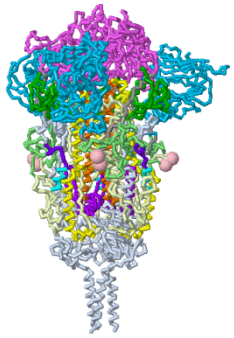


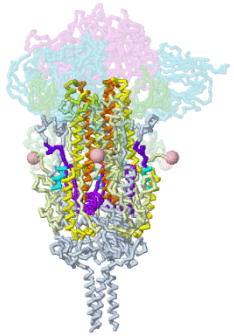


SS - signal sequence
NTD - N-terminal domain
RBD - receptor-binding domain
SD1 - subdomain 1
SD2 - subdomain 2
S1/S2 = S1/S2 protease cleavage site
S2' = S2' protease cleavage site

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





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

ATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAACCAGAACTCAATTACCCCCTGCATACACTAATTCTTTCACACGTGGTGTTTATTACC
CTGACAAAGTTTTCAGATCCTCAGTTTTACATTCAACTCAGGACTTGTTCTTACCTTTCTTTTCCAATGTTACTTGGTTCCATGCTATACATGTCTCTGGGACCAATGGTACTA
AGAGGTTTGATAACCCTGTCCTACCATTTAATGATGGTGTTTATTTTGCTTCCACTGAGAAGTCTAACATAATAAGAGGCTGGATTTTTGGTACTACTTTAGATTCGAAGAC
CCAGTCCCTACTTATTGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTTGTAATGATCCATTTTTGGGTGTTTATTACCACAAAAACAACAAAAGTTG
GATGGAAAGTGAGTTCAGAGTTTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCAGCCTTTTCTTATGGACCTTGAAGGAAAACAGGGTAATTTCAAAAATCTT
AGGGAATTTGTGTTTAAGAATATTGATGGTTATTTTAAAATATATTCTAAGCACACGCCTATTAATTTAGTGCGTGATCTCCCTCAGGGTTTTTCGGCTTTAGAACCATTGGT
AGATTTGCCAATAGGTATTAACATCACTAGGTTTCAAACTTTACTTGCTTTACATAGAAGTTATTTGACTCCTGGTGATTCTTCTTCAGGTTGGACAGCTGGTGCTGCAGCTT
ATTATGTGGGTTATCTTCAACCTAGGACTTTTCTATTAAAATATAATGAAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTCTCAGAAACAAAGTGTAC
GTTGAAATCCTTCACTGTAGAAAAAGGAATCTATCAAACTTCTAACTTTAGAGTCCAACCAACAGAATCTATTGTTAGATTTCCTAATATTACAAACTTGTGCCCTTTTGGTG
AAGTTTTTAACGCCACCAGATTTGCATCTGTTTATGCTTGGAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAATTCCGCATCATTTTCCACTTTT
AAGTGTTATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTTACTAATGTCTATGCAGATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAAA
CTGGAAAGATTGCTGATTATAATTATAAATTACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAATTCTAACAATCTTGATTCTAAGGTTGGTGGTAATTATAATTACCT
GTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAGATATTTCAACTGAAATCTATCAGGCCGGTAGCACACCTTGTAATGGTGTTGAAGGTTTTAATTGTTAC
TTTCCTTTACAATCATATGGTTTCCAACCCACTAATGGTGTTGGTTACCAACCATACAGAGTAGTAGTACTTTCTTTTGAACTTCTACATGCACCAGCAACTGTTTGTGGACC
TAAAAAGTCTACTAATTTGGTTAAAAACAAATGTGTCAATTTCAACTTCAATGGTTTAACAGGCACAGGTGTTCTTACTGAGTCTAACAAAAAGTTTCTGCCTTTCCAACAAT
TTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTTGAGATTCTTGACATTACACCATGTTCTTTTGGTGGTGTCAGTGTTATAACACCAGGAAC
AAATACTTCTAACCAGGTTGCTGTTCTTTATCAGGATGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTGGCGTGTTTATTCTACAGGTT
CTAATGTTTTTCAAACACGTGCAGGCTGTTTAATAGGGGCTGAACATGTCAACAACTCATATGAGTGTGACATACCCATTGGTGCAGGTATATGCGCTAGTTATCAGACTC
AGACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATCCATCATTGCCTACACTATGTCACTTGGTGCAGAAAATTCAGTTGCTTACTCTAATAACTCTATTGCCATA
CCCACAAATTTTACTATTAGTGTTACCACAGAAATTCTACCAGTGTCTATGACCAAGACATCAGTAGATTGTACAATGTACATTTGTGGTGATTCAACTGAATGCAGCAATC
TTTTGTTGCAATATGGCAGTTTTTGTACACAATTAAACCGTGCTTTAACTGGAATAGCTGTTGAACAAGACAAAAACACCCAAGAAGTTTTTGCACAAGTCAAACAAATTTA
CAAAACACCACCAATTAAAGATTTTGGTGGTTTTAATTTTTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCATTTATTGAAGATCTACTTTTCAACAAAGTG
ACACTTGCAGATGCTGGCTTCATCAAACAATATGGTGATTGCCTTGGTGATATTGCTGCTAGAGACCTCATTTGTGCACAAAAGTTTAACGGCCTTACTGTTTTGCCACCTT
TGCTCACAGATGAAATGATTGCTCAATACACTTCTGCACTGTTAGCGGGTACAATCACTTCTGGTTGGACCTTTGGTGCAGGTGCTGCATTACAAATACCATTTGCTATGCA
AATGGCTTATAGGTTTAATGGTATTGGAGTTACACAGAATGTTCTCTATGAGAACCAAAAATTGATTGCCAACCAATTTAATAGTGCTATTGGCAAAATTCAAGACTCACTT
TCTTCCACAGCAAGTGCACTTGGAAAACTTCAAGATGTGGTCAACCAAAATGCACAAGCTTTAAACACGCTTGTTAAACAACTTAGCTCCAATTTTGGTGCAATTTCAAGTG
TTTTAAATGATATCCTTTCACGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTGATAGGTTGATCACAGGCAGACTTCAAAGTTTGCAGACATATGTGACTCAACAATTAAT
TAGAGCTGCAGAAATCAGAGCTTCTGCTAATCTTGCTGCTACTAAAATGTCAGAGTGTGTACTTGGACAATCAAAAAGAGTTGATTTTTGTGGAAAGGGCTATCATCTTAT
GTCCTTCCCTCAGTCAGCACCTCATGGTGTAGTCTTCTTGCATGTGACTTATGTCCCTGCACAAGAAAAGAACTTCACAACTGCTCCTGCCATTTGTCATGATGGAAAAGCA
CACTTTCCTCGTGAAGGTGTCTTTGTTTCAAATGGCACACACTGGTTTGTAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACACATTTGTGTCTGGTA
ACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCTGAATTAGACTCATTCAAGGAGGAGTTAGATAAATATTTTAAGAATCATACATCACC
AGATGTTGATTTAGGTGACATCTCTGGCATTAATGCTTCAGTTGTAAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTGCCAAGAATTTAAATGAATCTCTCATCGAT
CTCCAAGAACTTGGAAAGTATGAGCAGTATATAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTTGATTGCCATAGTAATGGTGACAATTATGCTTTGCTGT
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 VSSQCVNLTT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS TQDLFLPFFS NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNI IRGWIFGTTL DSKTQSLLIV

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

       250        260        270        280        290        300        310        320        330        340        350        360
LLALHRSYLT PGDSSSGWTA GAAAYYVGYL QPRTFLLKYN ENGTITDAVD CALDPLSETK CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPFGE VFNATRFASV YAWNRKRISN

370        380        390        400        410        420        430        440        450        460        470        480
CVADYSVLYN SASFSTFKCY GVSPTKLNDL CFTNVYADSF VIRGDEVRQI APGQTGKIAD YNYKLPDDFT GCVIAWNSNN LDSKVGGNYN YLYRLFRKSN 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 CASYQTQTNS PRRARSVASQ SIIAYTMSLG AENSVAYSNN SIAIPTNFTI

       730        740        750        760        770        780        790        800        810        820        830        840
SVTTEILPVS MTKTSVDCTM YICGDSTECS NLLLQYGSFC TQLNRALTGI AVEQDKNTQE VFAQVKQIYK TPPIKDFGGF NFSQILPDPS KPSKRSFIED LLFNKVTLAD AGFIKQYGDC

       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 GAISSVLNDI LSRLDKVEAE VQIDRLITGR LQSLQTYVTQ QLIRAAEIRA SANLAATKMS ECVLGQSKRV DFCGKGYHLM SFPQSAPHGV VFLHVTYVPA QEKNFTTAPA

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

      1210       1220       1230       1240       1250       1260       1270
QELGKYEQYI KWPWYIWLGF IAGLIAIVMV TIMLCCMTSC CSCLKGCCSC GSCCKFDEDD 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 SalI XbaI XhoI KpnI HindIII

ATGGATCCGGAATTCAAAGGCCTACGTCGACGAGCTCAACTAGTGCGGCCGCTTTCGAATCTAGAGCCTGCAGTCTCGAGGCATGCGGTACCAAGCTTGGCTGTTTTG

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, BcgI, BciVI, BglI, BplI, Bpu10I, BsePI, BseYI, BsrBI, BtgZI, BtsI, DraII, DrdI, Esp3I, FseI, FspAI, HaeII, HaeIV, HgaI, Hpy99I, KpnI, MauBI, MfeI, MluI, NaeI, NarI, NheI, NotI, NruI, PacI, PvuI, RsrII, SacI, SacII, SalI, SanDI, SapI, PI-SceI, SfiI, SgfI, SgrAI, SgrDI, SmaI, SnaBI, SphI, SrfI, StuI, TaqII, TauI, XbaI, XhoI

*Bam*HI: G↓GATCC
*Eco*RI: G↓AATTC
*Stu*I: AGG↓CCT
*Sal*I: G↓TCGAC
*Xba*I: T↓CTAGA
*Xho*I: C↓TCGAG
*Kpn*I: GGTAC↓C
*Hind*III: A↓AGCTT

Obrigado

fscha@usp.br