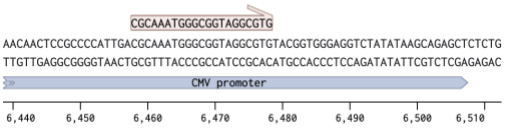
Answers Laboratory 6

Exercise 1

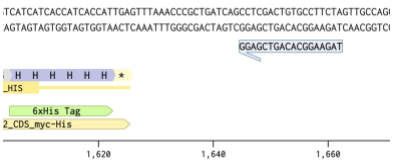
For sequencing of the recombinant plasmid CMV-for and BGH-rev were used.
a) In Benchling create and attach the standard sequencing primers to your recombinant plasmid pAmy2CDS-myc-HIS cloned last semester (reminder: go to Primers>Create Primers>Manual. If the primer is attached to the wrong location (red nucleotides): first Detach Primer, then go to Primers> Attach Existing, it will scan the sequence for complementary regions). Paste a screenshot of the attached primers into SLIMS.

Answer:
Attached primers can be shown in different views (one view in SLIMS is sufficient). No formal figure legends are requested for the screenshots.

Forward primer

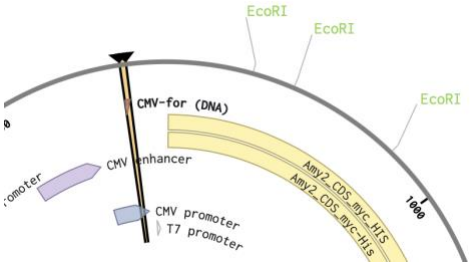


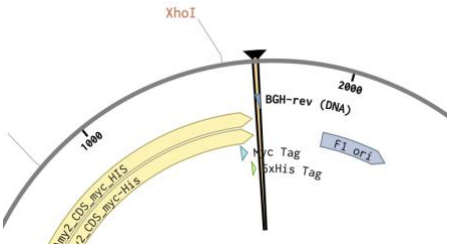
Reverse primer



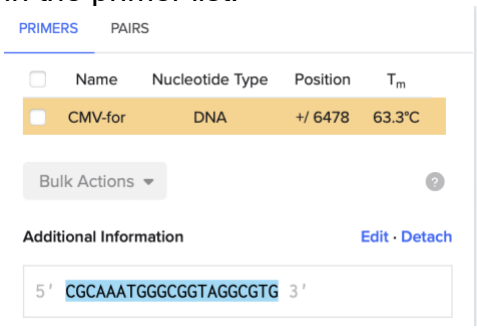
attached to the sequence:

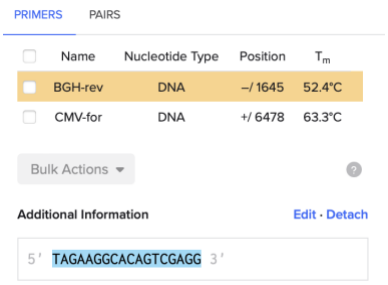
on the circular map (or linear map, not shown here):





in the primer list:





b) Knowing the distance between the two primers, would sequencing with a single primer be sufficient to verify the sequence of the recombinant Amy2-myc-His?

Answer: The distance between the forward and reverse primer is 1478 bp (or 1546 bp including the primers). Since the typical Sanger sequencing read length is approximately 800-1100 bases, sequencing with a single primer is in this case not sufficient to verify the entire sequence of recombinant Amy2-myc-His.

c) For Sanger sequencing you accidentally used the primer Mm-Amy2-for (used for PCR amplification in lab 3). What is the effect on the sequencing result? Is it sufficient to verify the sequence of the recombinant Amy2-myc-His sequence?

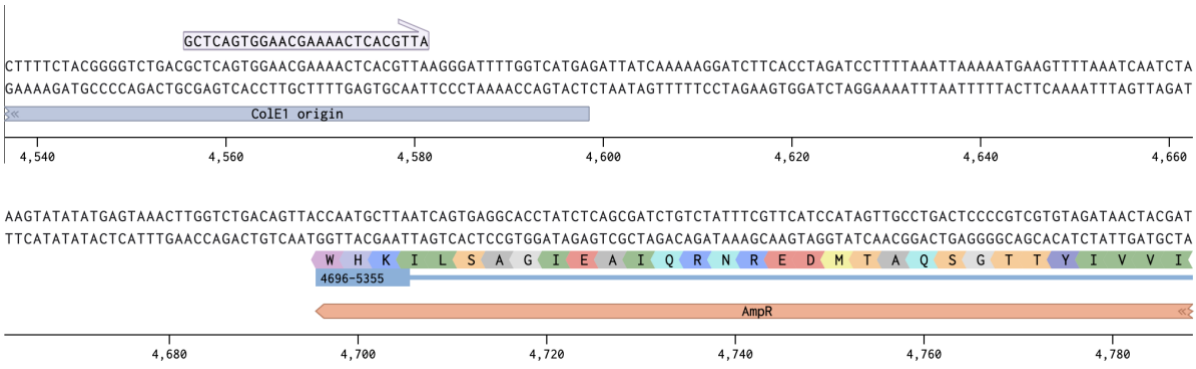
Answer: Only part of the recombinant Amy2-myc-His sequence will be analyzed, i.e the part that lies approximately 50 bp downstream of the primer. As a result no information is available about the regulatory sequences (Kozak, start codon) and N-terminus of the protein. Sequencing with the primer that was used for PCR is not sufficient to verify the entire sequence of recombinant Amy2-myc-His.

Exercise 2

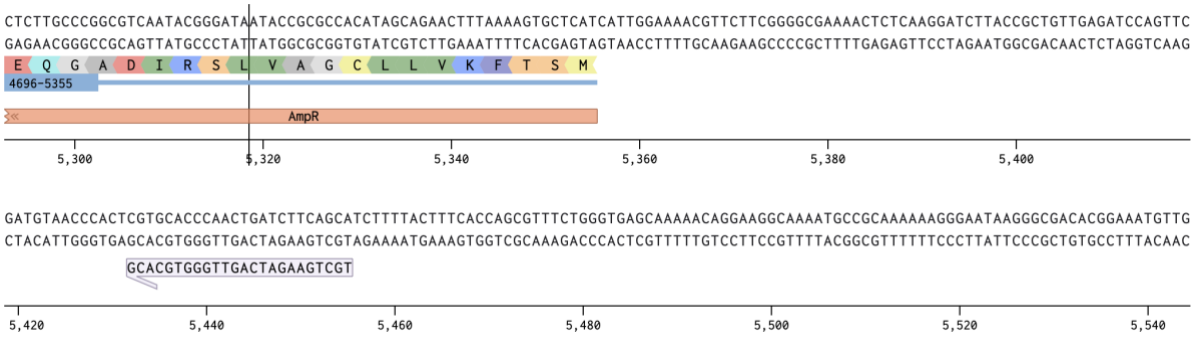
*You want to sequence the **ampicillin resistance gene** in the recombinant plasmid.*
a) Create primers using the Benchling 'wizard' tool that suggests suitable primers (standard primer parameters). Choose a suitable primer pair for sequencing in both directions. Paste a screenshot of the attached primers into SLIMS and indicate the nucleotide distance between the primer and the target sequence you want to sequence.

Answer: There are multiple solutions since many different primers can be designed. The primers should be
1) outside of the ampicillin resistance gene
2) with a minimum distance of 50 bp to the AMP coding sequence (CDS)

Here an example:
distance forward primer-AMP CDS: 114 bp (screenshot not requested)



distance reverse primer-AMP CDS: 76 bp (screenshot not requested)



b) Knowing the distance between the two primers, would sequencing with a single primer be sufficient to verify the sequence? Justify your answer.

Answer: In this case the distance between the forward and reverse primer is 850 bp (or 900 bp including the primers). The distance between the primer and the end of the AMP CDS is less than 800 bp. Since the typical Sanger sequencing read length is approximately 800-1100 bases, sequencing with a single primer is sufficient to verify the entire AMP CDS sequence.

Exercise 3

*Analyze the sequences using Benchling: align **both** sequencing files (.ab1 files on Moodle) to your recombinant plasmid pAmy2CDS-myc-HIS (open the sequence of the recombinant plasmid (=template) go to Alignment>Create New Alignment, upload the **chromatograms** use standard parameters).*

a) *To help you analyze Sanger sequencing data, first read: How to Interpret DNA Chromatograms (Moodle). Then check the beginning and end of the **chromatograms** for*

- *Baseline noise* (hardly visible in the chromatograms)
- *Loss of resolution* (visible at the end of the chromatograms)
- *Miscalled bases* ('N', several examples at the beginning or end of the reads)

This question is for understanding, no answers requested.

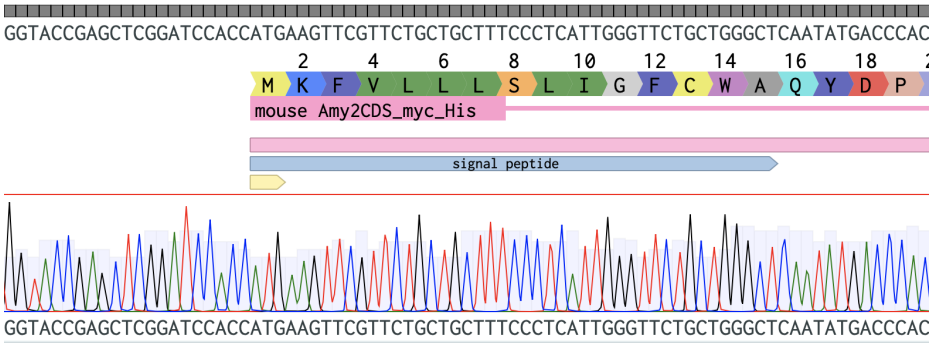
Sequence Analysis

*Analyze the sequences using **Benchling**. Upload the **chromatograms** (.ab1 files on Moodle). Align the sequencing files to the recombinant plasmid pAmy2CDS-myc-HIS (template) from last semester.*

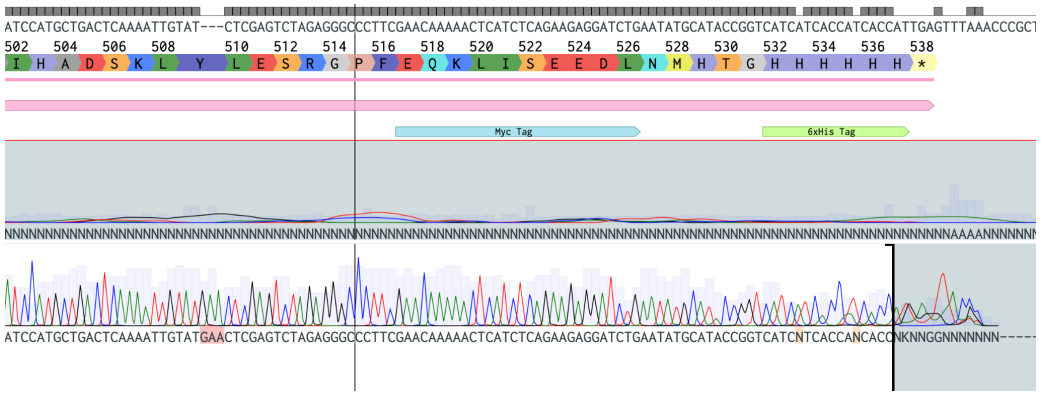
1. *Check aligned **chromatograms** (.ab1 files) for the presence of the following features:*

- ☐ *Cloning sites (e.g. BamHI and XhoI sites)*
- ☐ *Kozak sequence*
- ☐ *Start and stop codons (note that Amy2 stop codon (TAA) was modified to tyrosine (TAT) upon PCR amplification)*
- ☐ *Myc-epitope*
- ☐ *Histidine-tag (as the sequencing primer is close to the tag, the peaks are not well resolved and therefore it is not possible to read all histidines)*
- ☐ *Presence of mutations*
- ☐ *Paste relevant parts of the sequence alignments (nucleotide and amino acid sequence) into SLIMS and label the data.*

Beginning of recombinant sequence

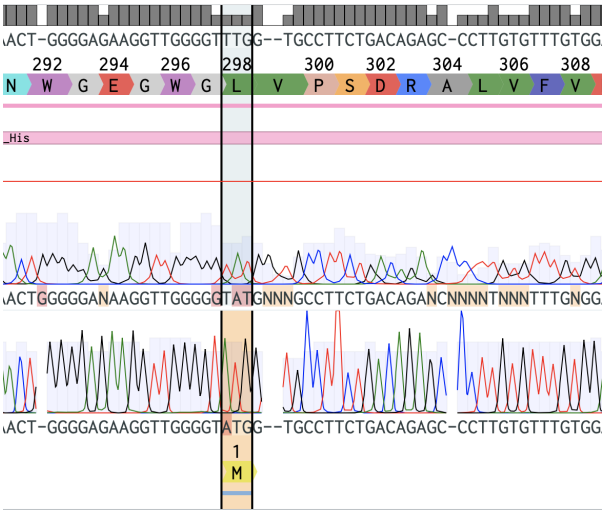


End of insert recombinant sequence (there is a mismatch in my template at GAA, will share updated version later). The Amy2 STOP codon was correctly mutated to generate the fusion protein containing the myc- and HIS-tag. Because the primer is close to the HIS-tag sequence some nucleotides are not well readable by the software (N), however by eye we see the color of the expected nucleotide.

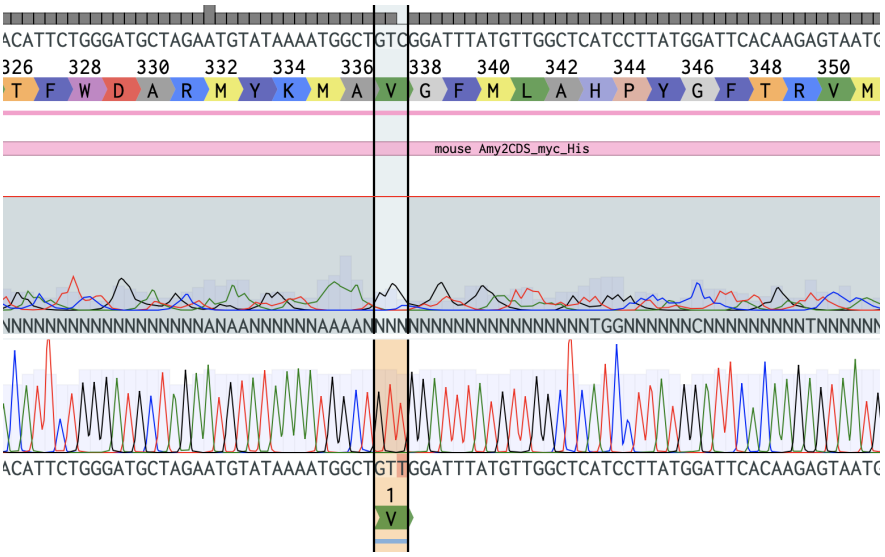


Mutations detected inside the Amy2 CDS

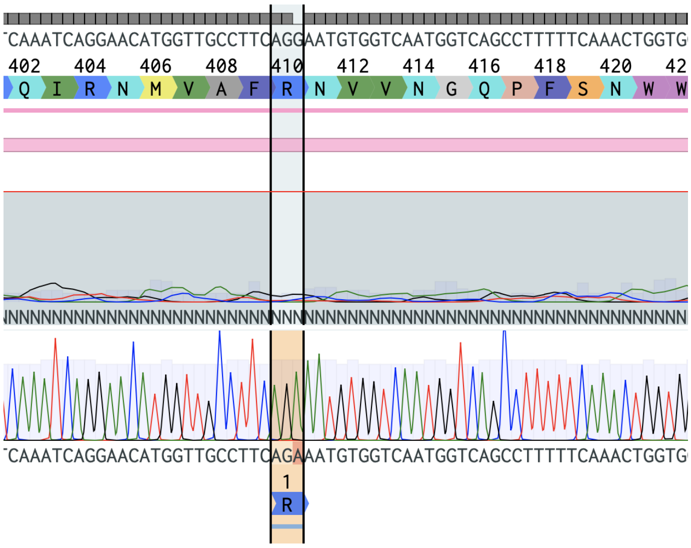
1. Single nucleotide change T to A (first position of codon) results in amino acid change L to M (two non polar amino acids).



2. Single nucleotide change C to T (third position of codon), conservation of amino acid (silent mutation).



3. Single nucleotide change G to A (third position of codon), conservation of amino acid (silent mutation).

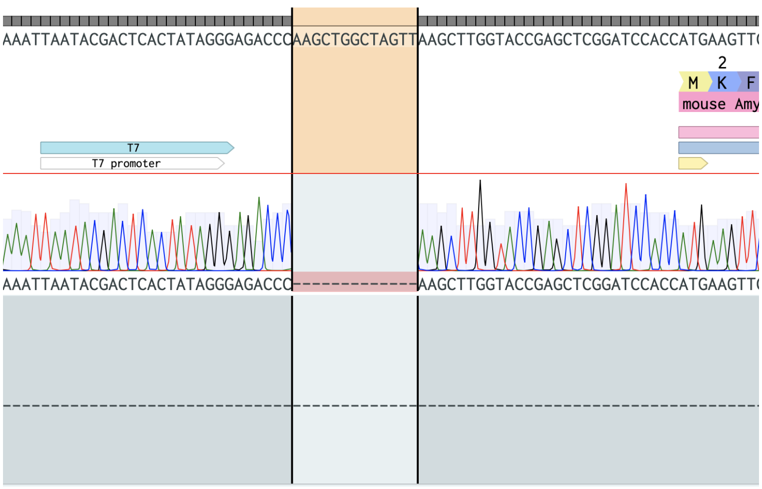


There are several read errors 'N' detected (only in reverse sequence). These are NOT mutations.

Are the sequences correct? Does your plasmid encode the expected fusion protein and regulatory sequences?

Yes, the recombinant sequence Amy2CDS-myc-HIS was generated. It can be used for protein expression. There is one mutation in the fusion protein (we have tested this variant in the lab and it does not affect activity).

Note that there are 13 nucleotides missing upstream of the ATG start codon. This does not affect expression; *please ignore this for the purpose of the course*.



Exercise 4

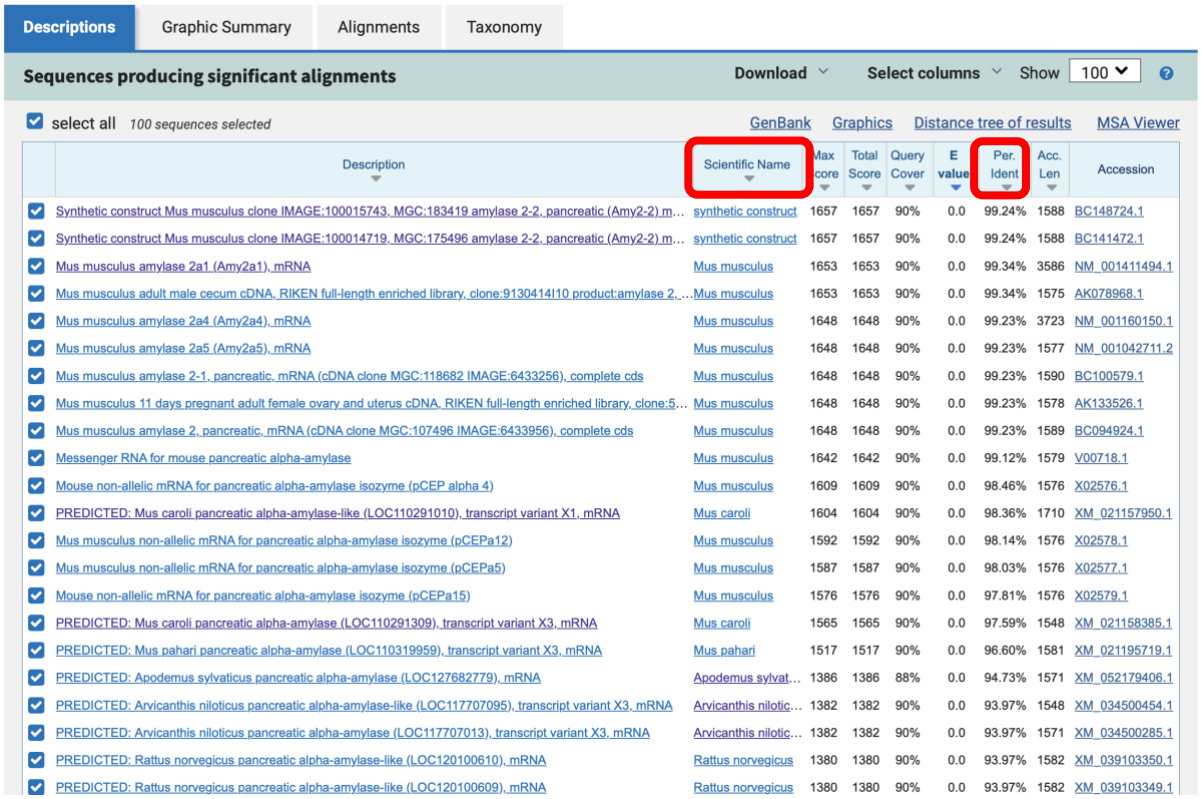
Use nucleotide BLAST to analyse the sequence file of the forward sequencing primer CMV-for (Sequence1.fasta on Moodle). Choose default settings, i.e database: Standard> Nucleotide Collection (nr/nt). Look at the results: by default the table 'Descriptions' is sorted by E value, other criteria may be used depending on the context.

a) Name four distinct top ranked species in column 'scientific name'. What does this tell us about the conservation of the sequenced gene in other rodents (see percent identity)?

Answer:
The four top ranked species are:
1. Mus musculus
2. Mus caroli
3. Mus pahari
4. Apodemus sylvaticus

The % identity is between 94-99%, thus the Amy2 sequence is conserved in other rodents.

The output looks like this (screenshot not requested)



Descriptions | Graphic Summary | Alignments | Taxonomy

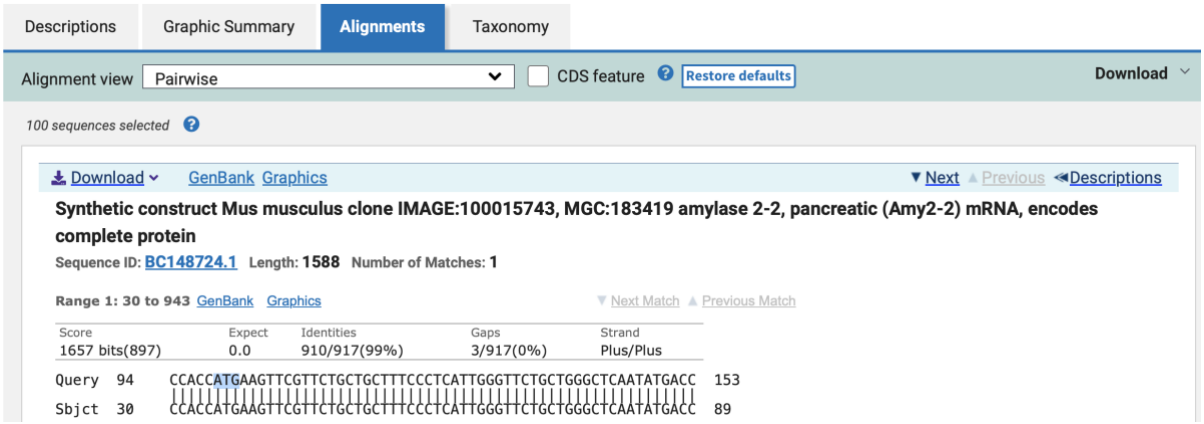
Sequences producing significant alignments — Download — Select columns — Show 100

select all 100 sequences selected — GenBank | Graphics | Distance tree of results | MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Synthetic construct Mus musculus clone IMAGE:100015743, MGC:183419 amylase 2-2, pancreatic (Amy2-2) m...	synthetic construct	1657	1657	90%	0.0	99.24%	1588	BC148724.1
Synthetic construct Mus musculus clone IMAGE:100014719, MGC:175496 amylase 2-2, pancreatic (Amy2-2) m...	synthetic construct	1657	1657	90%	0.0	99.24%	1588	BC141472.1
Mus musculus amylase 2a1 (Amy2a1), mRNA	Mus musculus	1653	1653	90%	0.0	99.34%	3586	NM_001411494.1
Mus musculus adult male cecum cDNA, RIKEN full-length enriched library, clone:9130414I10 product:amylase 2,...	Mus musculus	1653	1653	90%	0.0	99.34%	1575	AK078968.1
Mus musculus amylase 2a4 (Amy2a4), mRNA	Mus musculus	1648	1648	90%	0.0	99.23%	3723	NM_001160150.1
Mus musculus amylase 2a5 (Amy2a5), mRNA	Mus musculus	1648	1648	90%	0.0	99.23%	1577	NM_001042711.2
Mus musculus amylase 2-1, pancreatic, mRNA (cDNA clone MGC:118682 IMAGE:6433256), complete cds	Mus musculus	1648	1648	90%	0.0	99.23%	1590	BC100579.1
Mus musculus 11 days pregnant adult female ovary and uterus cDNA, RIKEN full-length enriched library, clone:5...	Mus musculus	1648	1648	90%	0.0	99.23%	1578	AK133526.1
Mus musculus amylase 2, pancreatic, mRNA (cDNA clone MGC:107496 IMAGE:6433956), complete cds	Mus musculus	1648	1648	90%	0.0	99.23%	1589	BC094924.1
Messenger RNA for mouse pancreatic alpha-amylase	Mus musculus	1642	1642	90%	0.0	99.12%	1579	V00718.1
Mouse non-allelic mRNA for pancreatic alpha-amylase isozyme (pCEP alpha 4)	Mus musculus	1609	1609	90%	0.0	98.46%	1576	X02576.1
PREDICTED: Mus caroli pancreatic alpha-amylase-like (LOC110291010), transcript variant X1, mRNA	Mus caroli	1604	1604	90%	0.0	98.36%	1710	XM_021157950.1
Mus musculus non-allelic mRNA for pancreatic alpha-amylase isozyme (pCEPa12)	Mus musculus	1592	1592	90%	0.0	98.14%	1576	X02578.1
Mus musculus non-allelic mRNA for pancreatic alpha-amylase isozyme (pCEPa5)	Mus musculus	1587	1587	90%	0.0	98.03%	1576	X02577.1
Mouse non-allelic mRNA for pancreatic alpha-amylase isozyme (pCEPa15)	Mus musculus	1576	1576	90%	0.0	97.81%	1576	X02579.1
PREDICTED: Mus caroli pancreatic alpha-amylase (LOC110291309), transcript variant X3, mRNA	Mus caroli	1565	1565	90%	0.0	97.59%	1548	XM_021158385.1
PREDICTED: Mus pahari pancreatic alpha-amylase (LOC110319959), transcript variant X3, mRNA	Mus pahari	1517	1517	90%	0.0	96.60%	1581	XM_021195719.1
PREDICTED: Apodemus sylvaticus pancreatic alpha-amylase (LOC127682779), mRNA	Apodemus sylvat...	1386	1386	88%	0.0	94.73%	1571	XM_052179406.1
PREDICTED: Arvicanthis niloticus pancreatic alpha-amylase-like (LOC117707095), transcript variant X3, mRNA	Arvicanthis nilotic...	1382	1382	90%	0.0	93.97%	1548	XM_034500454.1
PREDICTED: Arvicanthis niloticus pancreatic alpha-amylase (LOC117707013), transcript variant X3, mRNA	Arvicanthis nilotic...	1382	1382	90%	0.0	93.97%	1571	XM_034500285.1
PREDICTED: Rattus norvegicus pancreatic alpha-amylase-like (LOC120100610), mRNA	Rattus norvegicus	1380	1380	90%	0.0	93.97%	1582	XM_039103350.1
PREDICTED: Rattus norvegicus pancreatic alpha-amylase-like (LOC120100609), mRNA	Rattus norvegicus	1380	1380	90%	0.0	93.97%	1582	XM_039103349.1

b) Click on the top ranked alignment. Can you find the start and stop codon? Justify your answer.

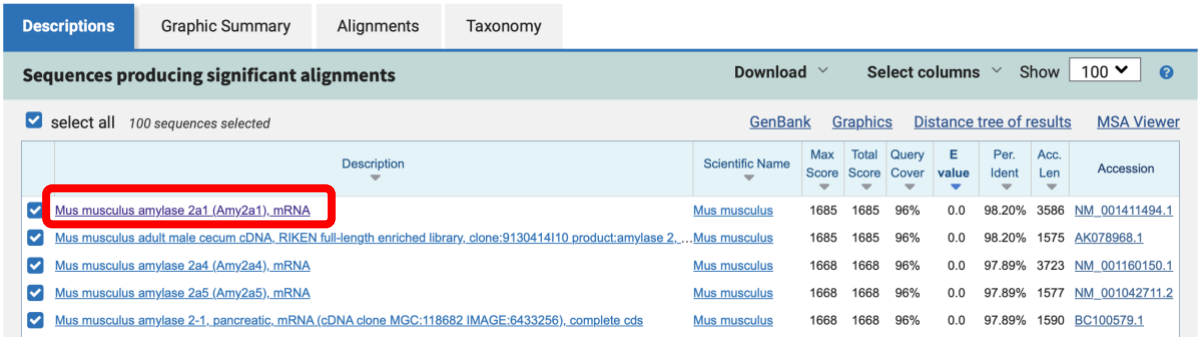
Answer: The start codon can be found at the beginning of the alignment. However, the stop codon is not included in the sequence since it is too far away from the forward sequencing primer. Thus it the stop codon cannot be found in the alignment.
The output looks like this, with ATG highlighted (screenshot not requested)

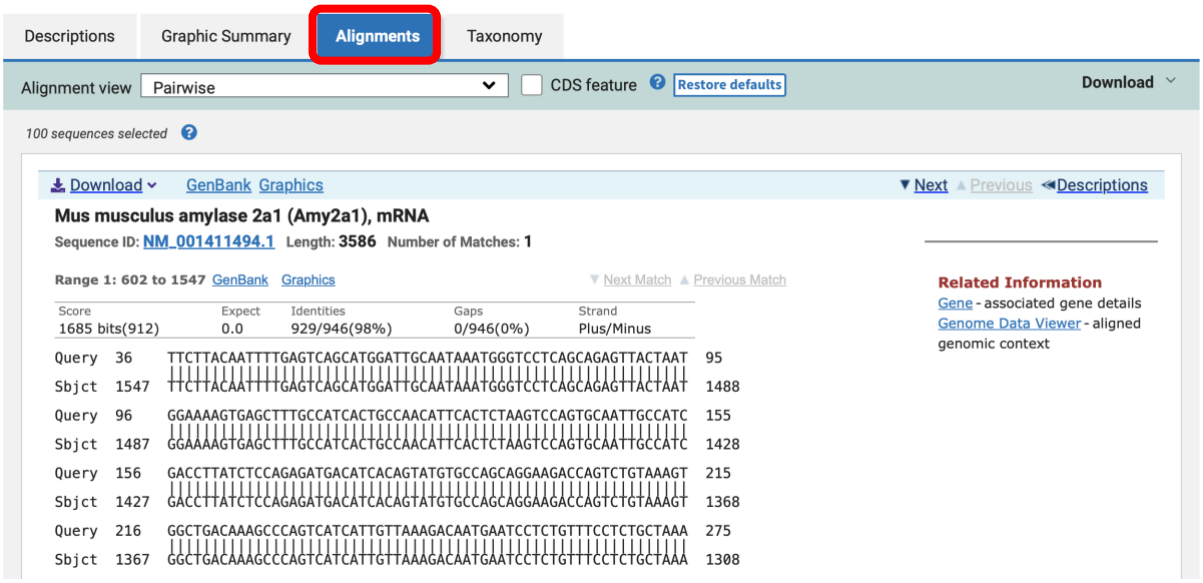


Exercise 5
Use nucleotide BLAST to analyse the sequence file of the reverse sequencing primer BGH-rev (Sequence2.fasta on Moodle).
a) Choose again default settings, i.e. database: Standard> Nucleotide Collection (nr/nt). Click on the top ranked alignment. Can you find the start and stop codon? Justify your answer.

Answer: The start codon is too far away from the reverse sequencing primer, thus the start codon cannot be found in the alignment.
The mutated stop codon (recombinant His-myc fusion protein) is included in the alignment, but it is difficult to find (by eye, no annotation as in Benchling).

The output looks like this (screenshot not requested)

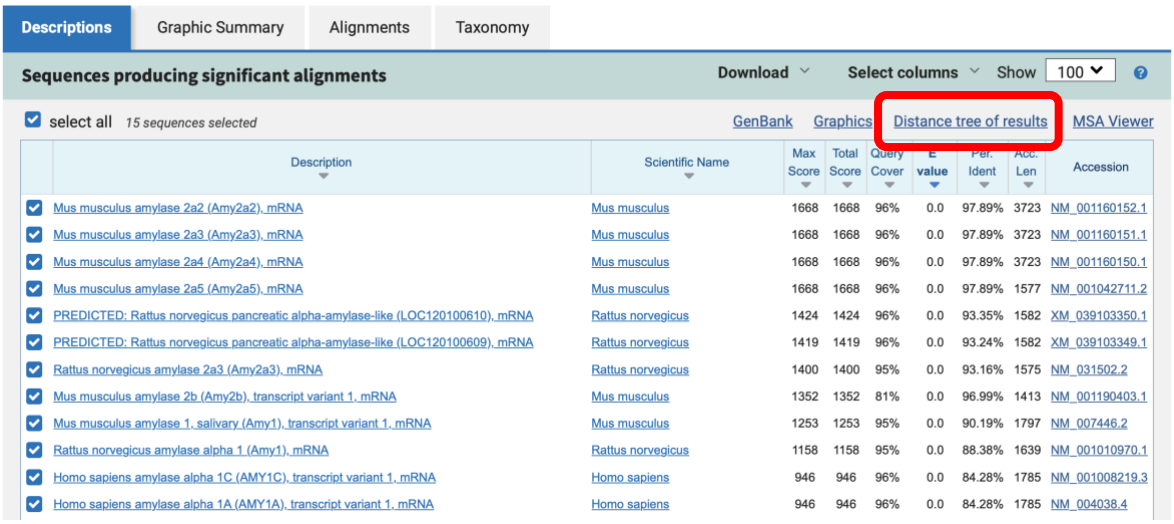




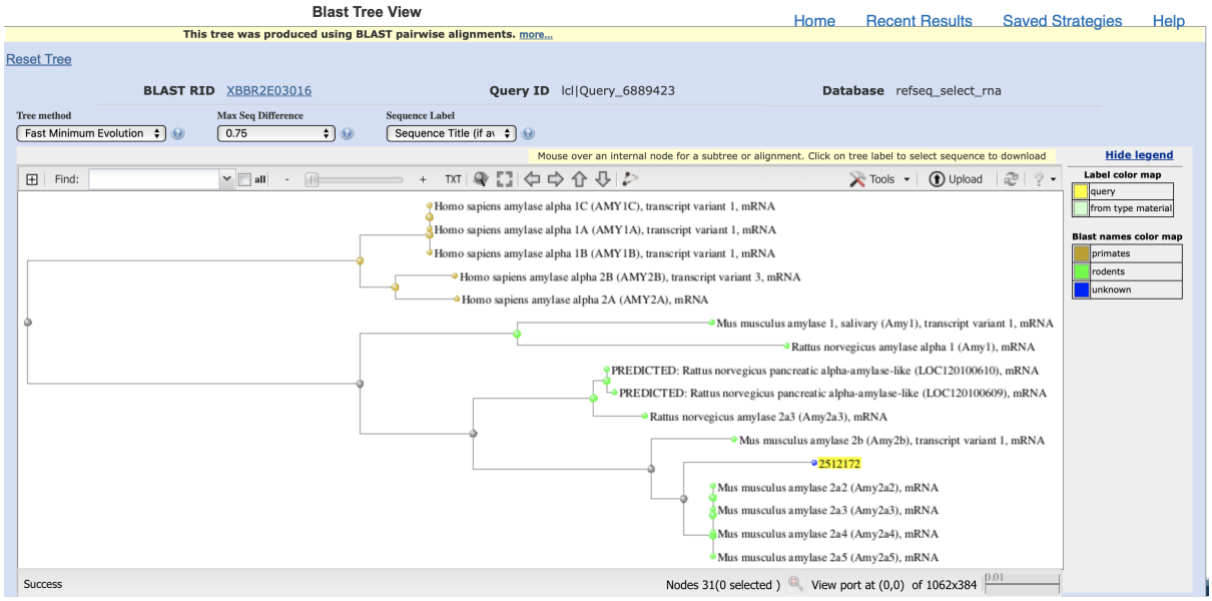
b) Now use nucleotide BLAST with modified search criteria. Choose database: Standard> RefSeq Select RNA sequences (refseq_select) in drop-down menu. What is the percent identity (see column 7 labelled Per. Ident) between mouse Amy2 and the closest sequence identified
-in rat?
-in human?
Paste the 'Distance Tree of Results' into SLIMS.

Answer:
The percent identity between (the partial) mouse Amy2 and
-rat is: 93 %
-human is: 84 %.
Note this analysis should be done with the full length coding sequence.

The output looks like this (screenshot not requested)



The 'Distance Tree of Results' should be pasted into SLIMS (with short figure legend).


Blast Tree View
Home
Recent Results
Saved Strategies
Help
This tree was produced using BLAST pairwise alignments. more...
Reset Tree
BLAST RID XBBR2E03016
Query ID lcl|Query_6889423
Database refseq_select_rna
Tree method
Fast Minimum Evolution
Max Seq Difference
0.75
Sequence Label
Sequence Title (if av
Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download
Hide legend
Label color map
query
from type material
Blast names color map
primates
rodents
unknown
Find:
all
TXT
Tools
Upload
Homo sapiens amylase alpha 1C (AMY1C), transcript variant 1, mRNA
Homo sapiens amylase alpha 1A (AMY1A), transcript variant 1, mRNA
Homo sapiens amylase alpha 1B (AMY1B), transcript variant 1, mRNA
Homo sapiens amylase alpha 2B (AMY2B), transcript variant 3, mRNA
Homo sapiens amylase alpha 2A (AMY2A), mRNA
Mus musculus amylase 1, salivary (Amy1), transcript variant 1, mRNA
Rattus norvegicus amylase alpha 1 (Amy1), mRNA
PREDICTED: Rattus norvegicus pancreatic alpha-amylase-like (LOC120100610), mRNA
PREDICTED: Rattus norvegicus pancreatic alpha-amylase-like (LOC120100609), mRNA
Rattus norvegicus amylase 2a3 (Amy2a3), mRNA
Mus musculus amylase 2b (Amy2b), transcript variant 1, mRNA
2512172
Mus musculus amylase 2a2 (Amy2a2), mRNA
Mus musculus amylase 2a3 (Amy2a3), mRNA
Mus musculus amylase 2a4 (Amy2a4), mRNA
Mus musculus amylase 2a5 (Amy2a5), mRNA
Success
Nodes 31(0 selected)
View port at (0,0) of 1062x384
0.01