The analysis of the UniProt Data on MYB3R4_ARATH

Callum M. Simpson : 160303267

Abstract

Motivation: The MYB-type helix-turn-helix domain is a DNA-binding of around 55 amino acids that typically occurs in a tandem repeat in eukaryotic transcription factors. This protein appears in the organism Arabidopsis thaliana that is frequently used for biological research. In Arabidopsis thaliana, one of these proteins is MB3R4_ARATH. Like many other proteins, a UniPort record has been created. This paper aims to verify if the data stored in UniProt is correct using various computational techniques. I chose this protein out of the ones give (see appendix one) because it contained a 5/5 annotation score and has been manually reviewed meaning there is a lot to review and compare. This paper aims to verify if the data stored in UniPort is correct.

Contact: C.Simpson5@ncl.ac.uk

1 Introduction

Link to Uniport page - MB3R4_ARATH
https://www.ncbi.nlm.nih.gov/gene/831023

Arabidopsis thaliana (commonly known as Thale cress, mouse-ear cress or Arabidopsis) is a small flowering plant native to Eurasia and Africa. Arabidopsis was originally adopted as a model organism because of its usefulness for genetic experiments. Important features included a short generation time, with it taken about 8 - 12 weeks to complete its life cycle, small size that limited the requirement for growth facilities, and prolific seed production through self-pollination. (Speedbase, 2010)

Q94FL9 (MB3R4_ARATH) is a protein produced that produced by Arabidopsis thaliana which UniPort states has many different functions. It's a The transcription factor that binds 5'-AACGG-3` motifs in gene promoters (PubMed:21862669). It's involved in the regulation of cytokinesis (PubMed:17287251), cytokiness being the physical process of cell division which divides the cytoplasm of a parental cell into two daughter cells (scitable). It's also required for the for the maintenance of diploidy (PubMed:21862669). Diploidy being the condition of having two complete sets of chromosomes in the same cell nucleus. Source (Encyclopedia of Genetics, 2001). It's also involved in transcription regulation during induced end reduplication at a G.orontii infection site, thus promoting G.orontii growth and reproduction.

The UniPort is a database maintained by EMBL, which catalogues protein of a vast number of different organisms. The record for this protein is regularly updated and contains a 5 start annotation rating, all sections have been filled out with a lot of them being manually edit. I chose this over the other options (see appendix one for the others) due to its completeness meaning that there would be a lot more to analysis and compare. Most UniPort records get sections automatically generated from related proteins but this record seems to malually updated and filled in.

According to uniport the subcellular location of the protein is the nucleus. The Nucleus being a dense organelle present in most eukaryotic cells, typically a single rounded structure bounded by a double membrane, containing the genetic material. The nucleus is responsible for regular functions of the cell, such as administration and the processing of information(job nucleus)

This protein has homologues in many other eukaryotes organisms

2 Methods

Before starting my analysis, a check of the functions stated in UniPort was carried out to find out if they what was stated was accurate. This check was done by using Experiments Archive to cross-reference past experiments carried out on the protein to functions stated.

Further checks on the function where carried about by comparing the Family and domains stored in Uniport to the domains stated in Pfam and those stored in inter-pro.

Uniport states that the peptide length of this protein is 961 amino acids long in the following order.

```
MEAESSTPQERIPKLRHGRTSGPARRSTRGQWTAEEDEILRKAVHSFKGKNWKKIAEYFK
DRTDVQCLHRWQKVLNPELVKGPWTKEEDEMIVQLIEKYGPKKWSTIARFLPGRIGKQCR
ERWHNHLNPAINKEAWTQEEELLLIRAHQIYGNRWAELTKFLPGRSDNGIKNHWHSSVKK
KLDSYMSSGLLDQYQAMPLAPYERSSTLQSTFMQSNIDGNGCLNGQAENEIDSRQNSSMV
GCSLSARDFQNGTINIGHDFHPCGNSQENEQTAYHSEQFYYPELEDISVSISEVSYDMED
CSQFPDHNVSTSPSQDYQFDFQELSDISLEMRHNMSEIPMPYTKESKESTLGAPNSTLNI
DVATYTNSANVLTPETECCRVLFPDQESEGHSVSRSLTQEPNEFNQVDRRDPILYSSASD
RQISEATKSPTQSSSSRFTATAASGKGTLRPAPLIISPDKYSKKSSGLICHPFEVEPKCT
TNGNGSFICIGDPSSSTCVDEGTNNSSEEDQSYHVNDPKKLVPVNDFASLAEDRPHSLPK
HEPNMTNEQHHEDMGASSSLGFPSFDLPVFNCDLLQSKNDPLHDYSPLGIRKLLMSTMTC
MSPLRLWESPTGKKTLVGAQSILRKRTRDLLTPLSEKRSDKKLEIDIAASLAKDFSRLDV
MFDETENRQSNFGNSTGVIHGDRENHFHILNGDGEEWSGKPSSLFSHRMPEETMHIRKSL
EKVDQICMEANVREKDDSEQDVENVEFFSGILSEHNTGKPVLSTPGQSVTKAEKAQVSTP
RNQLQRTLMATSNKEHHSPSSVCLVINSPSRARNKEGHLVDNGTSNENFSIFCGTPFRRG
LESPSAWKSPFYINSLLPSPRFDTDLTIEDMGYIFSPGERSYESIGVMTQINEHTSAFAA
FADAMEVSISPTNDDARQKKELDKENNDPLLAERRVLDFNDCESPIKATEEVSSYLLKGC
B
```

The protein was searched in NCBI database and Pfam to get the AA chain from this protein from two different databases. All results where compared

via clustal omega to check alignment to see if they were the same. Another check was done with ProtParam to check if the protein lengths where similar.

EMBOSS Backtranseq and TBLASTN were used to obtain the nucleotide sequence of the protein, two different the nucleotide sequences where produced so Transeq was used to convert both these back to into a protein sequence. T-coffee was then used to compare the AA chain produced by Transeq to determine accuracy. The protein was then put through the ORF finder to discover the proteins relation to an open reading frame.

The genomic featuers of the gene where then discovered via using Ensembl plants (Ensembl plants where used as the gene I was looking at is a plant based eukaryotic gene). The Transcript where compared to one another and then the the genes nuclotide sequence was then imported into FGENSH to discover any other information.

To check Alignment a PSI-Blast was carried out on the protien to see how well it alligned with other protien. From the results given 5 protiens where chosen so a range of diffrent E-values could be alligned. The multiplae sequence aligment of the protiens where ran on cluster Omega and T-coffee. Both algiments where compared in Jal view to gain a better understanding of alignment.

To produce a phylogenetic tree a regular Blast p on the protien was performed to find all those that share a similarity to the protien. A blast tree was produced from this result. For a further test 5 protiens from different species where chosen. Clustal omega was used to produce a tree based of aligment and Sea view was then used to verfiy the polymetric tree of these organisms.

The network of interactions was discovered by using String to get a better understanding of the protien interacts with other proteins. Scores of each interaction where collected and compared.

To generate a 3d modle a Blast p of the protien was done against the PDB (protien data bank) To find similar protiens that had there crystals solved. The resulst gotten from blast p where gotten then looked up in the PDB via accession code to find . Stereochemical quality precheck was done on each of the selected values to check the radacandam plot for outliers. A 3D modle was produced using SWISS plot and the best result was put into Pymol to check alligment with its template crystal. A 3D moble was also produced from LOMETS. The modles where compared to ones held in UniPort.

3   Results

Checking Functionality

Using the experiments archive, I searched for the experiments on Arabidopsis that mentioned my protein. I found 3 experiments.

https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-15680/?query=MYB3R4
E-GEOD-15680 Talks about an experiment where a Arabidopsis was infect with Golovinomyces orontii. Arabidopsis was used due to it clearly defined stages of infection. It was discovered that cell cycle regulator

MYB3R4 exhibits altered expression at the site of infection. This back up the promoting G.orontii growth and reproduction function.

The research done in E-GEOD-29433
 (https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-29433/?query=MYB3R4). States that MB3R4_ARATH is believed to play important roles in cell cycle regulation through transcriptional regulation of G2/M phase-specific genes by binding to common cis-elements, called MSA elements. With the results demonstrating the importance of the MYB3R-MSA pathway for regulating G2/M-specific transcription. I believe that this tie it to being a transcription factor meaning suggesting that functions uni port states is correct.

What was also interesting was by doing this I discovered that Arabidopsis thaliana has 4 other proteins that are transcription factors of MYB3R (MYB3R-1, MYB3R-2, MYB3R-3, MYB3R-5). All are transcription factor that bind 5`-AACGG-3` motifs in gene promoters. Telling us that there are many proteins that can carry this function and it emphasize the importance of this task. MYB3R-4 seems to be the only one that the promoting G.orontii growth and reproduction though.

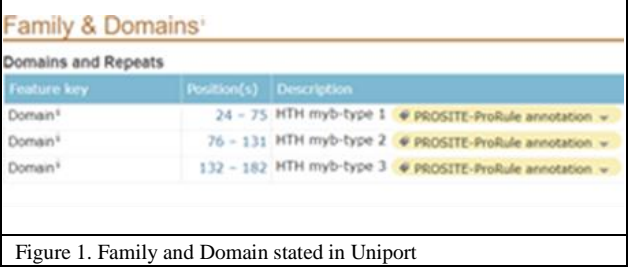Uniport states that the domain for this protein is are 3 HTH myb-types as seen in figure 1.

Figure 1. Family and Domain stated in Uniport

A Pfam search discovered (figure 2) the protein domain. According to Pfam the domain there were two Pfam domains a Myb-DNA-binding (29 –75) and Myb-DNA-bind-6 (84 - 144). Both relate to the function of the protein. It should be mentioned that this is only covers a small part of the protein with the rest being in discourse. However As we see in later chapters talking about similar proteins this range is where the other proteins would align with MYB3R4. This is probably due to MYB being the main function of this protein and they are around 50~ amino long. Whilst not in the Domains and repeats section these Pfams are under the family and domain databases records.
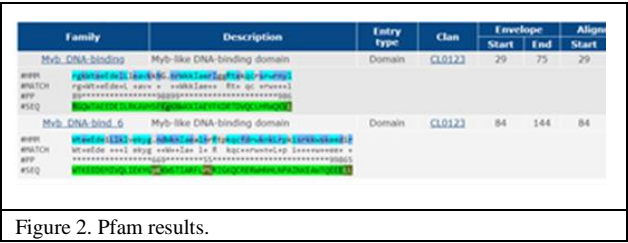

Figure 2. Pfam results.

When running an interpro search (figure 3) on the protein we discover that the protein is of a family type Myb-like. This supports the record in unipor. Similarly, we can see that in the domains that InterPro predicted that there are three different cases of HTH_MYB all at the locations that the record in Uni-port says they should be. Uniport does refer to the Homologous

superfamily that is Homeobox-like_sf and the other domain but in a spate section.

Figure 3. Inter pro results.

As we can see (figure 4) the Myb_DNA_Bind_6 PFAM exists in the InterPro results but not the first PFAM. What we can see is that there is no initial part of the sequence belongs to no protein domain suggesting that there is a higher chance of being subjective to some evaluation. This makes sense as lot of the uses of MB3R4 has to do with it being mutated and changed..

Figure 4. Single Pfam

Confirming protein sequence

A aa chain for the protein where achieved from NCBI database and pfam. All results where compared via clustal omega. Alignment showed that they were all the same showing that the protein section in UniPort was correct. (Figure 5)

Another check was done with ProtParam (figure 6) and by using the protein sequence ProtParam detected that there where a total for 961 amino acids in the sequence so showing that uniport was correct.

Number of amino acids: 961

Molecular weight: 108236.22

Theoretical pI: 5.47

Figuer.6 ProtParam results

Figure.5 clustal omega alignment

Translation to nucleotide sequence

EMBOSS Backtranseq and TBLASTN were used to obtain the nucleotide sequence of the proteins. Two different nucleotide sequence where produced (see appendix 2 for emboss and appendix 3 for TBLASTN). Transeq was used to convert both these back to into a protein sequence. The two peptides produce where compared to the one in uniport producing the figure 7. As we can see all share a 100% alignment suggest that both nucleotide sequence are valid.

Figuer.8 Comparison of alignment via Transeq

The ORFfinder results of the nucleotide sequence are shown in figures 9 and 10. These show that the entire protein sits inside a single ORF upstream, with multiple ORFs existing within, going both upstream and downstream.
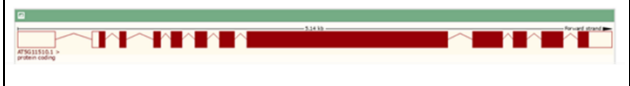
Figuer.9 ORFfinder results of Nucleotide Sequence
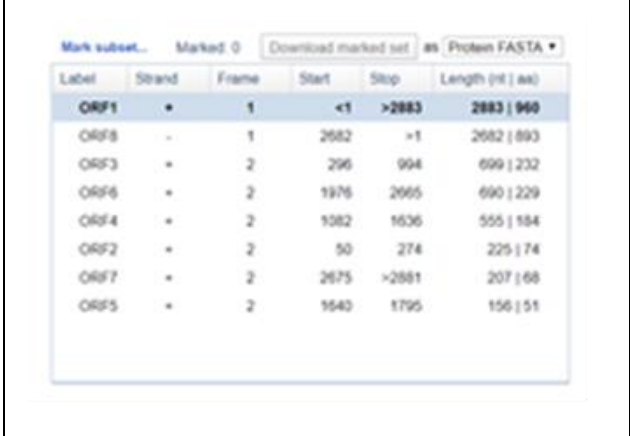
Figuer.10 Open reading Frames within Nucleotide sequence

Confirming protein sequence

I translated the raw nucleotide sequence of this gene's DNA into raw mRNA. From there I sequenced it into a sequence of amino acids.

As the gene I was looking at is a plant based eukaryotic gene I used the gene finder on Ensembl plants to work out the exons. Looking at the sequence at there are a total of 3 exons from the neighboring genes and 12 that where prime exons of the gene
There are 5 transcripts.

Transcript 1 : AT5G11510.3				
Exon	Coding Exon	Transcript length	Translation length	introns
12	11	3546	961	11

Description: this transcript details one large exon in the middle surrounded by a number of shorter exons

Transcript 2 : AT5G11510.1				
Exon	Coding Exon	Transcript length	Translation length	introns
12	11	3473	961	11

Description: this transcript details one large exon in the middle surrounded by a number of shorter exons

Transcript 3 : AT5G11510.4				
Exon	Coding Exon	Transcript length	Translation length	introns
11	11	3161	961	10

Description: this transcript details one large exon in the middle surrounded by a number of shorter exons

Transcript 4 : AT5G11510.2				
Exon	Coding Exon	Transcript length	Translation length	introns
8	7	2723	798	7

Description: this transcript details a start of a bunch of smaller exons then one large exons

Transcript 5 : AT5G11510.5				
Exon	Coding Exon	Transcript length	Translation length	introns
8	8	2319	772	7

Description: this transcript details a group of smaller exons then followed by one large one.

Comparison of the Transcripts is difficult due to their differences 1,2,3 shared a similar number of exons and a similar pattern however only 1 and 2 share the same start codons, 4 and 5 share the same stop codons.

I exported the nucleotide sequence from Ensemble with 1000 upstream and 1000 downstream into FGENESH gene detector. I retrieved the nucleotide sequence from Ensemble and imported it into the gene tector FGENESH 2.6. FGENESH. This detected the 15 exons from 3 different genes but no introns.(Fig 11)

Looking through this record I discovered that in gene 2 and exons 11 produced a result that has the same length as my protein.. (Fig 12)

After identified the coding region for the protein within the gene, I check the translation retrieved from FGNESH to the record in UNiPort protein sequence to see if there were any similarities between the two. The figure shows that there is 100% identity between the two suggesting that the consensus meaning that the protein sequence is correct. .(Fig 13) This shows the entire protein exists as on exon.

Figure. 11 FGENESH result

Figure. 12 Possible protein

Figure. 13 Protein comparision
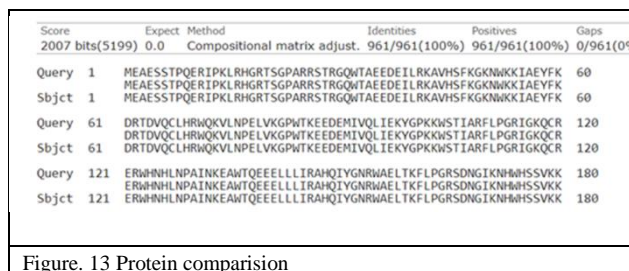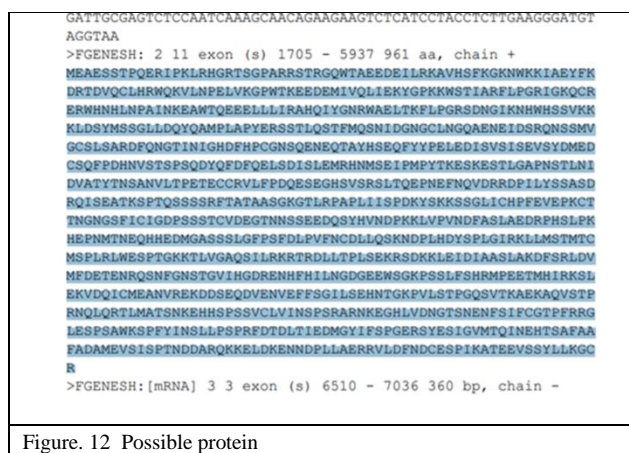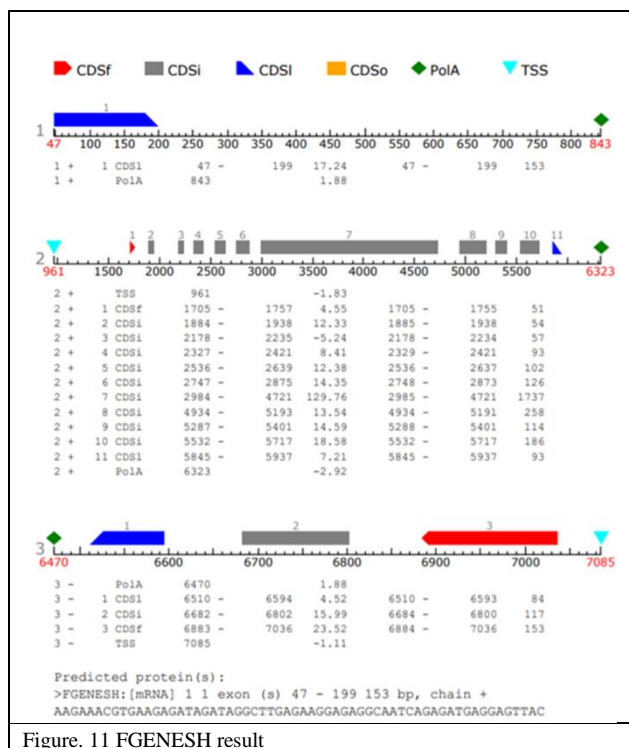
Alignment

Using PSI-BLAST tool I was able to get compare the alignment of Q94FL9 (MB3R4_ARATH) to other proteins. (fig 14). Looking at the top 5 proteins that are similar to Q94FL9 (MB3R4_ARATH) by E Value we discover that all par one are other Transcription factors MYB3R that belong to Arabidopsis thaliana. The protein in 4th place is a Transcription factor MYB3R-2 from a Oryza sativa subsp. Japonica (one of three major subspecies of rice). The top 196 proteins have some relation to a transcription function. This further backs up the functional analysis stated on the uniport page for MB3R4_ARATH. As we can see all seeme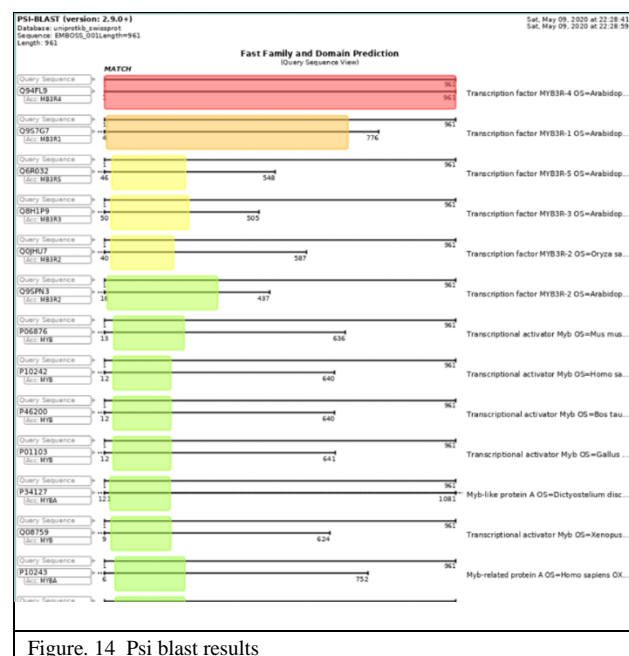d to be aligned around one core section. This is where the proteins domain is located further suggesting that all the similar protein share the same function/

Figure. 14 Psi blast results

The following proteins where picked for their varying E - values. These sequences where exported into Clustal omega (see appendix 4 for results) and T-coffe to Preform alignment (see appendix 5 for results). T-coffe produced a sequence that was 1295 in length whilst Cluster Omega produced an alignment of 1178.

Transcription factor MYB3R-1	http://www.uni-prot.org/uni-prot/Q9S7G7	4.0E-160
Transcription factor MYB3R-5 O	http://www.uni-prot.org/uni-prot/Q6R032	9.0E-82
Transcription factor MYB3R-2 OS=Oryza sativa subsp. japonica	http://www.uni-prot.org/uni-prot/Q0JHU7	1.0E-79
Transcriptional activator Myb OS=Bos taurus	http://www.uni-prot.org/uni-prot/P46200	1.0E-63

Myb protein OS=Drosophila melanogaster	http://www.uniprot.org/uniprot/P04197	3.0E-55

To gain a greater understanding of the alignment I viewed both alignment in Jal view to see as seen in the psi blast we saw that proteins that are like MB3R4_ARATH (full alignment for Clustal can be seen in appendix whereas T-blast can be seen in appendix 7) only seem to align in one area of the sequence.). That can be seen in figure 15 where between 100 to 250 all proteins line up near perfectly in both crustal and T-coffee. There are also smaller cluster where all a lot of the proteins share similarities however most of conservation and quality is poor for the rest of the sequence. Using my previous finding, I am confident to say that due to this patch of matching that the function shown in uniport if correct, however I will admit I'm a bit skeptical.
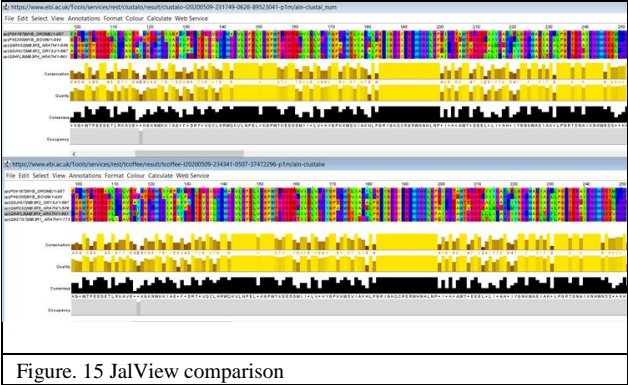

Figure. 15 JalView comparison

Phylogenetic Tree

I began by doing a Blast on the protein to produce a set of proteins that share similarities with my protein. This allowed me to produce a phylogenetic tree that can be seen in figure 15 that shows the similarities with other proteins (Max seq difference used was 0.35 for easy viewing). Though it's a bit blurred we can see that there is a lot that shares homologues with other proteins the share the MYB function.
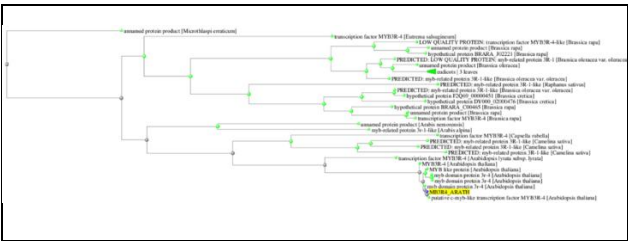

Figure. 15 Blast phylogenetic tress comparison

Further test where done but one a smaller scope, I picked 5 proteins related to MB3R4_ARATH but that existed in different organisms and aligned them via cluster Omega. Cluster omega and sea view where used to create phylogenetic trees based on the multiple alignment that it receives (not based on order sequences where added so does not necessarily reflect multiple alignment). Figure 16 shows Cluster image tree and figure 17 show sea view.

The other proteins
transcription factor MYB3R-4 [Capsella rubella] (pink shepherd's-purse close relation to
myb-related protein 3R-1-like, partial [Camelina sativa] (camelina)
transcription factor MYB3R-4 [Eutrema salsugineum](saltwater cress)
unnamed protein product [Arabis nemorensis] (no common name)
myb-related protein 3r-1-like [Arabis alpina] (gray rockcress)
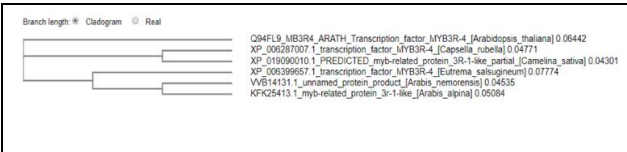
Figure. 16 Cluster Omega phylogenetic tree
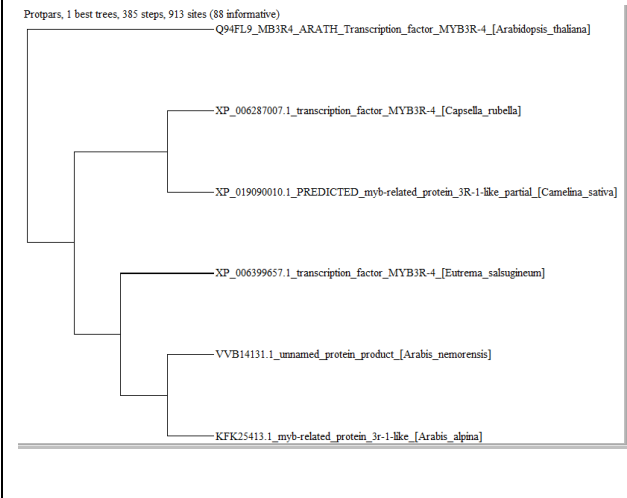
Figure. 17 Sea view phylogenetic tree

As you can see both graphs produced similar results. I expected a result like this from the given the chosen organisms. The Arabis plants are from the same family so I expected them to be on the same branch. However, though they are still related, the other organism are more related to one another than they are with Arabidopsis thaliana. Out of the two graphs I would chose the Seaview one as it a bit clear and sticks to the two leaves per deviation.

Network of interactions

String was used for protein sequence network As seen in the figure 18 there is a total of 11 nodes connected to one another with a total of 34 edges, 10 spanning off our protein. Looking at the results produced from the score we find that the protein has multiple weak connections to each other protein that when added together produce the average combined score of around 0.7 which suggests that there is a good probability that each interaction exists. This is good as many of the proteins that MYBR3-4 is connected to share a similar function to it. The fact that the majority of the proteins are interlinked to one another suggests that this is random, further suggesting that this network exists.

3D model

For clarity 2 models where produced using two different tools. One was created using Swiss model and the other was created using Loments' . Before I even started making the model, I ran a blast to find an appropriate template.
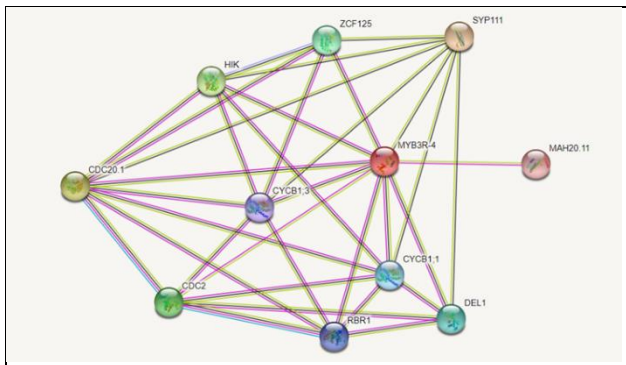
Figure. 18 Network of interactions

All results had a lower query number but and a decently low E-Value. The crystals where compared in PDB. Once records where check on PDB I produced a model using SWISS.

Out of the three models produced I chose the one belonging to template 1h88.1.c (see figure 20) is a Transcription factor of Homo sapiens and Mus musculus. It being a Transcription factor means its shares the same function as MB3R4_ARATH. Also as it had the lowest e-value. It suggests a better similarity to the protein. Plus it had the best overall Global estimates out of the three and the it's the only one that lies in the range when compared to other crystal structures.

Figure. 20 Swiss 3d model

Using Pymol the model that was produced by Swiss model was compared to the template it was based off (figure 20) to seen how aligned they are. It was discovered that the RMS given was 0.097 suggesting that these two structures where very similar globally. This model is very similar to the one produced in uniport.
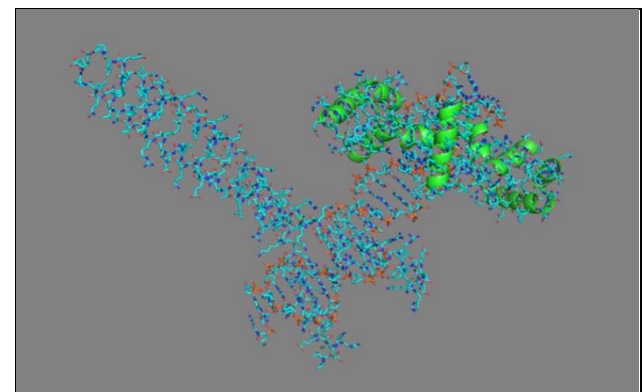
Figure. 21 Swiss model compared with template

3D model in Swiss plot

3D model
For clarity 2 models where produced using two different tools. One was created using Swiss model (https://swissmodel.expasy.org/) and the other was created using Lo-ments(https://zhanglab.ccmb.med.umich.edu/LOMETS/)'s . Before I even started making the model I ran a blast to find an appropriate crystal template. See figure 19 for results.

The top-rated structure rated by LOMENTS was the same model generated by Swiss plot. It produced a Norm score of 6.80, which was the highest by some margin suggesting that the model in uniport and the model made by Swiss is correct. A high ID suggest that that 1h88c is evolutionary related to my protein, the low COV score suggest that there is more than one domain in my protein which there is. However low coverage also suggest that it may be missing part of the protein.

Looking at the secondary structure from loments provided a highly confident results primarily made up of large sections of coils, a few alpha helix and a few very short beta strands.

All the full-length models produced by loments (see fig 24) (par 5) in my opinion have a very strong resemblance to the model produced by swiss

Figure. 19 Blast p templates
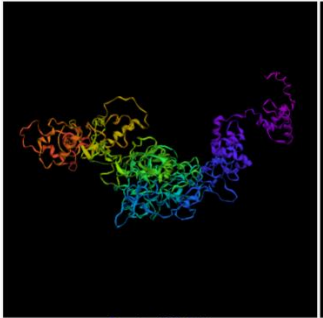
– model, although there are some inconsistence with the with purple secondary structure in some locations (in Swiss the blue and purple secondary sort of warp around each other whereas is most of these models there are the purple secondary structure spans out wards. While in most of helical structure of the predictions roughly match the ones in the Swiss model, they are just aligned at different angles. It's difficult to say which one is better, I have been having trouble generating a c score for these models so I actually can tell which is the most accurate.

1h88C
HHpred,FFAS3D,
SP3
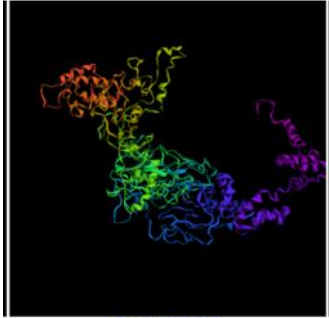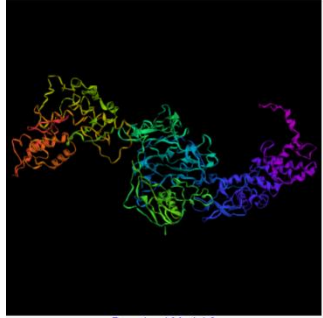
Figure. 22 Loments predicted structure
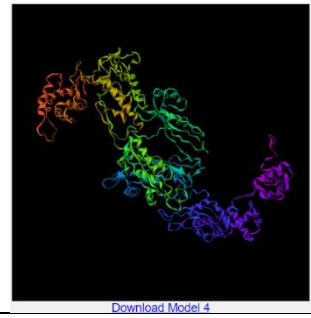
Rank	PDB Hit	ID1	ID2	Cov	Norm. Zscore	Download Alignment
1	1h88C	0.62	0.10	0.16	6.80	HHpred
2	2pffB	0.06	0.06	0.99	1.29	CEthreader
3	5wtjA	0.14	0.11	0.82	1.86	SparksX
4	1zlgA	0.10	0.06	0.68	1.64	MUSTER
5	1h88C	0.62	0.10	0.16	1.43	FFAS3D
6	3cnfB	0.07	0.06	0.92	5.23	Neff-MUSTER
7	6exnO	0.20	0.06	0.31	2.55	HHsearch
8	3j3iA	0.09	0.08	0.92	1.76	PROSPECTOR2
9	1h88C	0.62	0.10	0.16	3.10	SP3
10	4btgA	0.11	0.08	0.67	1.75	SparksX

Figure. 23 Loments top ten predicted structures.

Download Model 1

Download Model 2

Download Model 3

Download Model 4

Download Model 5

Figure. 24 Loments predicted full models

What should also be mentioned is that Uniport had another model for

this protein in mod base (figure 25) . The model produced by loments and swiss are similar, neither are like the one stored in ModBase suggesting that one is incorrect. Another thing backing this up is even Mod base says its model for this protein in incorrect in a few ways.
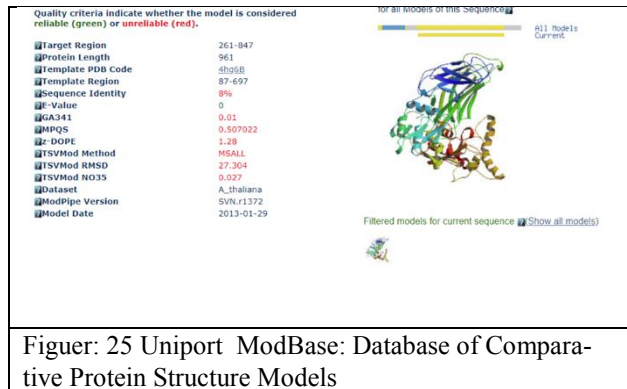
Figuer: 25 Uniport ModBase: Database of Comparative Protein Structure Models

4   Conclusion

Overall, I believe that the record in Uniport of this protein is correct, as I have used a variety of tool and techniques to manually check many different aspects of the report. I do believe that fact that the page has a 5-star annotation level does mean that the page has been thoroughly reviewed and manually annotated already however I wanted to further back up the findings with my own. Everything that I have research matches up on the Uni-port record.

From what I have found about this protein is that most its importance is a 150 peptide long parts of its total 961 AA length. This is the part in the peptide that the amino that is responsible for the MYB-type transcription. Everything about this model is tied to this chain of amino acids. how it aligns with other proteins that are similar to MYB3R4_ARATH are always and only frequently aligned by this region (as shown in the multi sequence analysis. This states how important this MYB-type transcription is as its present in many different proteins.

References

 https://en.wikipedia.org/wiki/Arabidopsis_thaliana
http://en.wikipedia.org/wiki/Cell_nucleus

Seedbase.2010.Sourcehttps://seedgenes.org/publications/Koornneef_Meinke_2010.pd
experiments/E-GEOD-15680.https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-15680/?query=MYB3R4
experimentsE-GEOD-29433(https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-29433/?query=MYB3R4
experimentsE-GEOD-52298(https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-52298/?query=MYB3R4)
job nuclus(https://www.reference.com/science/job-nucleus-c47d93d81c3b988e)
Internet Archive
    Experiments:https://experiments.archivelab.org/
EMBL_EBI(https://pfam.xfam.org/)
InterPro(https://www.ebi.ac.uk/interpro/)
Clustal Omega(https://www.ebi.ac.uk/Tools/msa/clustalo/)
ProtParam(https://web.expasy.org/protparam)
EMBOSS Backtranseq (https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/)
Blast(https://blast.ncbi.nlm.nih.gov/Blast.cgi)
EMBOSS Transeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq/)
ORFfinder(https://www.ncbi.nlm.nih.gov/orffinder/)
EnsemblePlanys(https://plants.ensembl.org/index.html)
FGENESH(https://omictools.com/fgenesh-tool)
SeaView(http://pbil.univ-lyon1.fr/software/seaview3.html)
T-Coffee(https://www.ebi.ac.uk/Tools/msa/tcoffee/)
Jalview(http://www.jalview.org/getdown/release/)
Sting(https://string-db.org/)
Lomets(https://zhanglab.ccmb.med.umich.edu/LOMETS/)
Swiss-model(https://swissmodel.expasy.org/)
PDB(https://www.rcsb.org/)
STRING(https://string-db.org/)
scitable (source https://www.nature.com/scitable/definition/cytokinesis-100/)

Appendix 1 – other proteins

A0A4D7FBD4_BACIU -
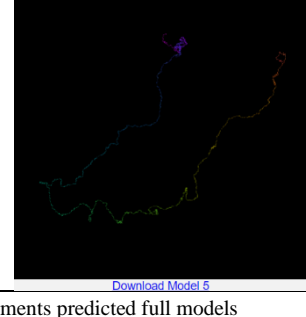https://www.uniprot.org/uniprot/A0A4D7FBD4.txt?version=1
https://www.ncbi.nlm.nih.gov/protein/QCJ17225.1/
The first protein I looked at was for Bacillus subtilis. The gene is FA024_08740. This is a spore forming facultative aerobe bacteria found mostly in soil and vegetation. It can produce and secrete antibiotics. When I did a search for it on uni Plot I got a result saying that on the 13/11/19 the page was made redundant. After some digging, I discovered that it only has one review on the 31/07/19.

B1Q270_CAEEL
https://www.uniprot.org/uniprot/B1Q270.txt
https://www.ncbi.nlm.nih.gov/gene/185674
This is protein of gene CELE_F42G4.5 is for the organism Caenorhabditis elegans (non-parasitic roundworms and threadworms). C.elegans has thousands of scientists from around the world working full time to investigate its biology however the protein I have been given is classified as an Uncharacterized protein on uniplot. The protein that I have been given has an unreviewed 1-star annotation score with only the protein predicted. From what I can tell this has only been referenced in one paper.

MB3R4_ARATH
https://www.uniprot.org/uniprot/Q94FL9.txt
https://www.ncbi.nlm.nih.gov/gene/831023
From the organism Arabidopsis thaliana, often called Mouse-ear cress or Thale cress, is a small flowering plant that is native to Africa. The protein given has a full 5-star annotation score and even gives experimental evidence at the protein level. The Protein has been referenced in several academic papers and the uni prot entry page being updated 153 times, with the last update being a few days before the creation of this report.

YOHD_ECOLI
https://www.uniprot.org/uniprot/P33366.txt
The inner membrane protein YohD of the Escherichia coli (strain K12). Escherichia coli is a type of bacteria that normally lives in the intestines, mostly harmless but can cause diarrhea if you eat contaminated food or drink. A quick look over the uniprot page shows that that the idle the protein entry has been be reviewed but it has only scored a 2-star annotation score. The entry page has been updated 127 but is missing a few sections such as entry and expression.

Appendix 2 – Emboss
```
ATGGAAGCTGAATCTTCTACTCCTCAAGAAAGAATTCCTAAGCTTAGACA
TGGAAGAACTTCTGGACCTGCTAGAAGATCTACTAGAGGACAATGGACTG
CTGAAGAAGATGAAATTCTTAGAAAGGCTGTTCATTCTTTTAAGGGAAAG
AATTGGAAGAAGATTGCTGAATATTTTAAGGATAGAACTGATGTTCAATG
TCTTCATAGATGGCAAAAGGTTCTTAATCCTGAACTTGTTAAGGGACCTT
GGACTAAGGAAGAAGATGAAATGATTGTTCAACTTATTGAAAAGTATGGA
CCTAAGAAGTGGTCTACTATTGCTAGATTTCTTCCTGGAAGAATTGGAAA
GCAATGTAGAGAAAGATGGCATAATCATCTTAATCCTGCTATTAATAAGG
AAGCTTGGACTCAAGAAGAAGAACTTCTTCTTATTAGAGCTCATCAAATT
TATGGAAATAGATGGGCTGAACTTACTAAGTTTCTTCCTGGAAGATCTGA
TAATGGAATTAAGAATCATTGGCATTCTTCTGTTAAGAAGAAGCTTGATT
CTTATATGTCTTCTGGACTTCTTGATCAATATCAAGCTATGCCTCTTGCT
CCTTATGAAAGATCTTCTACTCTTCAATCTACTTTTATGCAATCTAATAT
```

```
TGATGGAAATGGATGTCTTAATGGACAAGCTGAAAATGAAATTGATTCTA
GACAAAATTCTTCTATGGTTGGATGTTCTCTTTCTGCTAGAGATTTTCAA
AATGGAACTATTAATATTGGACATGATTTTCATCCTTGTGGAAATTCTCA
AGAAAATGAACAAACTGCTTATCATTCTGAACAATTTTATTATCCTGAAC
TTGAAGATATTTCTGTTTCTATTTCTGAAGTTTCTTATGATATGGAAGAT
TGTTCTCAATTTCCTGATCATAATGTTTCTACTTCTCCTTCTCAAGATTA
TCAATTTGATTTTCAAGAACTTTCTGATATTTCTCTTGAAATGAGACATA
ATATGTCTGAAATTCCTATGCCTTATACTAAGGAATCTAAGGAATCTACT
CTTGGAGCTCCTAATTCTACTCTTAATATTGATGTTGCTACTTATACTAA
TTCTGCTAATGTTCTTACTCCTGAAACTGAATGTTGTAGAGTTCTTTTTC
CTGATCAAGAATCTGAAGGACATTCTGTTTCTAGATCTCTTACTCAAGAA
CCTAATGAATTTAATCAAGTTGATAGAAGAGATCCTATTCTTTATTCTTC
TGCTTCTGATAGACAAATTTCTGAAGCTACTAAGTCTCCTACTCAATCTT
CTTCTTCTAGATTTACTGCTACTGCTGCTTCTGGAAAGGGAACTCTTAGA
CCTGCTCCTCTTATTATTTCTCCTGATAAGTATTCTAAGAAGTCTTCTGG
ACTTATTTGTCATCCTTTTGAAGTTGAACCTAAGTGTACTACTAATGGAA
ATGGATCTTTTATTTGTATTGGAGATCCTTCTTCTTCTACTTGTGTTGAT
GAAGGAACTAATAATTCTTCTGAAGAAGATCAATCTTATCATGTTAATGA
TCCTAAGAAGCTTGTTCCTGTTAATGATTTTGCTTCTCTTGCTGAAGATA
GACCTCATTCTCTTCCTAAGCATGAACCTAATATGACTAATGAACAACAT
CATGAAGATATGGGAGCTTCTTCTTCTCTTGGATTTCCTTCTTTTGATCT
TCCTGTTTTTAATTGTGATCTTCTTCAATCTAAGAATGATCCTCTTCATG
ATTATTCTCCTCTTGGAATTAGAAAGCTTCTTATGTCTACTATGACTTGT
ATGTCTCCTCTTAGACTTTGGGAATCTCCTACTGGAAAGAAGACTCTTGT
TGGAGCTCAATCTATTCTTAGAAAGAGAACTAGAGATCTTCTTACTCCTC
TTTCTGAAAAGAGATCTGATAAGAAGCTTGAAATTGATATTGCTGCTTCT
CTTGCTAAGGATTTTTTCTAGACTTGATGTTATGTTTGATGAAACTGAAAA
TAGACAATCTAATTTTGGAAATTCTACTGGAGTTATTCATGGAGATAGAG
AAAATCATTTTCATATTCTTAATGGAGATGGAGAAGAATGGTCTGGAAAG
CCTTCTTCTCTTTTTTCTCATAGAATGCCTGAAGAAACTATGCATATTAG
AAAGTCTCTTGAAAAAGGTTGATCAAATTTGTATGGAAGCTAATGTTAGAG
AAAAGGATGATTCTGAACAAGATGTTGAAAATGTTGAATTTTTTTCTGGA
ATTCTTTCTGAACATAATACTGGAAAGCCTGTTCTTTCTACTCCTGGACA
ATCTGTTACTAAGGCTGAAAAGGCTCAAGTTTCTACTCCTAGAAATCAAC
TTCAAAGAACTCTTATGGCTACTTCTAATAAGGAACATCATTCTCCTTCT
TCTGTTTGTCTTGTTATTAATTCTCCTTCTAGAGCTAGAAATAAGGAAGG
ACATCTTGTTGATAATGGAACTTCTAATGAAAATTTTTCTATTTTTTGTG
GAACTCCTTTTAGAAGAGGACTTGAATCTCCTTCTGCTTGGAAGTCTCCT
TTTTATATTAATTCTCTTCTTCCTTCTCCTAGATTTGATACTGATCTTAC
TATTGAAGATATGGGATATATTTTTTCTCCTGGAGAAAGATCTTATGAAT
CTATTGGAGTTATGACTCAAATTAATGAACATACTTCTGCTTTTGCTGCT
TTTGCTGATGCTATGGAAGTTTCTATTTCTCCTACTAATGATGATGCTAG
ACAAAAGAAGGAACTTGATAAGGAAAATAATGATCCTCTTCTTGCTGAAA
GAAGAGTTCTTGATTTTAATGATTGTGAATCTCCTATTAAGGCTACTGAA
GAAGTTTCTTCTTATCTTCTTAAGGGATGTAGA
```

Appendix 3 – T blast

```
atggaagctgagtcttcaacccctcaggaacggatcccgaaattgcgtcacgggaggactag-
    tggccctgcaagacgttcaactcgaggccaatggacagctgaggaggatgagatcttgagaaaagctgtccata
    gtttcaaaggaaaaaactggaagaagatagctgaatacttcaaggatcgaactgatgtccag-
    tgcctccacaggtggcaaaaggtccttaatcctgaacttgttaaagggccctggacgaaggaggaagatgaaat
    gatcgttcagttaatcgaaaaatatgggccgaagaagtggtccacattgtctcgattttac-
    ctggccgtattggaaagcaatgtagagagaggtggcacaatcatctcaatcctgcccataaacaaagaagcttgga
    ctcaagaagaagagctgcttctgattcgtgctcatcaaatatatggtaatagatgggctgag-
    ttaacaaagttcttgcctggaaggtcggataatggaattaaaaatcactggcacagttcagtcaagaagaaacttg
    attcctacatgtcgtctggcctttggatcagtaccaagccatgcctctagctccttatga-
    gagaagttctacattacaatcaactttatgcaaagtaatatagatggaaatggctgcctaaatggacaggcagag
    aacgaaatagacagccgtcaaaattcaagcatggttggctgctcactatctgcaaga-
    gactttcaaaatggaactatcaatattgggcatgatttccaccctgtgggaaactctcaagagaatgagcaaactgc
    gtatcattcagacgcagttttactatccggaattggaggatatctcag-
    tctccatctccgaagtctcttatgacatggaagactgttcacagtttcctgatcataatgtctcgacttctccaagcca
    ggattaccaatttgattttcaggaactatcagatatatcactcgagatgagacataa-
    tatgtcagaaataccaatgccatacacacaaggagagcaaggaatctactctgggagctccaaactccacattaaa
```

```
tattgatgtggctacttacaccaactcagcaaacgttttgacacctgaaacggaatgtt-
gcagggtacttttccctgatcaagagagtgaagggcacagtgtttctagatctttgacccaagagccaaacgagtt
taaccaggtagaccgtcgagatcctatcttgtattcatcagcttcagacagacaga-
tatcagaagctacaaaatctcctacgcagagctcttcttcaaggttcactgcaacggccgcctcaggaaaaggaa
ctcttcggccagctcccttaatcatatcaccagacaagtatagcaagaagtcatctg-
gattgatatgtcatccctttgaggtagaaccaaagtgcacaacaaatggaaatgggtagcttcatatgcatcggtgat
ccatcaagctccacttgtgttgatgaagggaccaataactccagcgaagaagatcaatcat-
atcatgtgaatgatcccaaaaagttggttccagtgaatgattttgcttctctggcagaagataggccacactctctac
ctaagcatgagccaaacatgacaaacgagcagcatcatgaggatatgggggg-
catcatcaagtcttggtttcccaagcttcgacttacctgtttttaactgtgatcttctacaatctaaaaatgatcctctaca
cgattatagcccacttggcatacgcaagctactgatgtccaccatgacgtgcatgag-
tcccccttcgactgtgggaatctcccactgggaaaaagacgttggttggtgcacagtcaatccttaggaaacggac
ccgtgatctgctgacaccgttatctgagaaaagaagtgataaaaagcttgaaattgata-
tagcagccagtctggcaaaagacttttcacgtttagatgtgatgtttgatgagactgagaatcggcaatctaattttg
gaaatagcactggtgtcatccatggagatagagaaaaccactttcacattttgaatggagac-
ggtgaagaatggagtggcaagccttcttcactctttagtcacagaatgccagaagaaaccatgcatatcagaaaat
cgcttgagaaagtagaccaaatttgcatggagccaaatgtcagggaaaaggatgat-
tctgaacaagatgttgagaatgttgaattttctctggtattctttctgagcacaataccggtaagccagtcctgtctac
tcctggtcagtcagttactaaagcagagaaggcgcaagtttcaaccccaagaaatcagtt-
gcaaagaactctcatggctacgtctaacaaagaacatcattcccccttcaagtgtttgtttagtgattaattcgccttctc
gagcgagaaacaaggagggtcatcttgttgacaatgggacaagcaatgagaattttag-
catattctgtggaactccattcaggagaggtcttgaatctccctcagcctggaaatctccattttacataaactctctct
tgcccagcccaaggtttgacacagatttaactatcgaggacatgggctacattttcag-
tcctggggagagaagctacgagagcataggagtgatgacacagataaatgaacatacaagtgcatttgcagcttt
tgcagatgcaatggaggtttcaatttcacctactaatgatgatgcgaggcagaa-
gaaggaattggataaagagaataatgatcctctactggccggagcgtcgagtgctggacttcaacgattgcgagtc
tccaatcaaagcaacagaagaagtctcatcctacctcttgaagggatgtagg
```

Appendix 4 – Crustal omega alignment

```
sp|P04197|MYB_DROME     MASAST-------ENGEELMN--YG--SNSDSEESEYSENEDTQVCDKDSQQNSNADSGY  49
sp|P46200|MYB_BOVIN     --------------------MARRPRHSIYSSDED-------D----------         16
sp|Q6R032|MB3R5_ARATH   MSSSSSNPPVCSPEKEERSEMKIEIQCMENKQPLAASCSSASE--------GS------GCF  47
sp|Q8LHU7|MB3R2_ORYSJ   --------------MGAMAMVEQEGCVENRQPLAASSSSVSD-------GSSYGGGGGGL   39
sp|Q94FL9|MB3R4_ARATH   ------------------------------------------------------------   0
sp|Q9S7G7|MB3R1_ARATH   ------------------------------------------------------------   0

sp|P04197|MYB_DROME     PLDSPELQDSKTTG----QKGQNKSGKTSIGAVHPNYGFGKRWSKSEDVLLKQLVET-HG  104
sp|P46200|MYB_BOVIN     -----------EDIEMCDHDYDGL-LP-KSGKRH---LGKTRWTREEDEKLKKLVEQNGT   60
sp|Q6R032|MB3R5_ARATH   FLKSPEIATPAT--------VSSFPRRTSGPMRR----AKGGNTPEEDETLRRAVEKYKG   95
sp|Q8LHU7|MB3R2_ORYSJ   AQHSPPVVSSSAN--------SISGLRRTSGPIRR----AKGGNTPEEDETLRKAVEAYKG   87
sp|Q94FL9|MB3R4_ARATH   --MEAESSTPQE-------RIPKLRHGRTSGPARR---STRGQNTAEEDEILRKAVHSFKG   49
sp|Q9S7G7|MB3R1_ARATH   --MKREMKAPTTPLESLQGDLKGKQGRTSGPARR----STKGQNTPEEDEVLCKAVERFQG   55

sp|P04197|MYB_DROME     ENWEIIGPHFKDRLEQQVQQRWAKVLNPELIKGPWTRDEDDMVIKLVRNFGPKKWTLIAR  164
sp|P46200|MYB_BOVIN     DDNKVIANYLPNRTDVQCQHRWQKVLNPELIKGPWTKEEDQRVIELVQKYGPKRWSVIAK  120
sp|Q6R032|MB3R5_ARATH   KRWKKIAEFFPERTEVQCLHRWQKVLNPELVKGPWTQEEDDKIVELVKKYGPAKWSVIAK  155
sp|Q8LHU7|MB3R2_ORYSJ   RNWKKIAECFPYRTEVQCLHRWQKVLNPELIKGPWTQEEDDQIIDLVKKYGPTKWSVIAK  147
sp|Q94FL9|MB3R4_ARATH   KNWKKIAEYFKDRTDVQCLHRWQKVLNPELVKGPWTKEEDEMIVQLIEKYGPKKWSTIAR  109
sp|Q9S7G7|MB3R1_ARATH   KNWKKIAECFKDRTDVQCLHRWQKVLNPELVKGPWSKEEDNTIIDLVEKYGPKKWSTISQ  115

sp|P04197|MYB_DROME     YLNGRIGKQCRERWHNHLNPNIKKTAWTEKEDEIIYQAHLELGNQWAKIAKRLPGRTDNA  224
sp|P46200|MYB_BOVIN     HLKGRIGKQCRERWHNHLNPEVKKTSWTEEEDRIIYQAHKRLGNRWAEIAKLLPGRTDNA  180
sp|Q6R032|MB3R5_ARATH   SLPGRIGKQCRERWHNHLNPGIRKDAWTVEEESALMNSHRMYGNKWAEIAKVLPGRTDNA  215
sp|Q8LHU7|MB3R2_ORYSJ   ALPGRIGKQCRERWHNHLNPEIRKDAWTTEEEQALINAHRIYGNKWAEIAKVLPGRTDNS  207
sp|Q94FL9|MB3R4_ARATH   FLPGRIGKQCRERWHNHLNPAINKEAWTQEEELLLIRAHQIYGNRWAELTKFLPGRSDNG  169
sp|Q9S7G7|MB3R1_ARATH   HLPGRIGKQCRERWHNHLNPGINKNAWTQEEELTLIRAHQIYGNKWAELMKFLPGRSDNS  175

sp|P04197|MYB_DROME     IKNHWNSTMRRKYDVERRSVNASGSDLKSSRTHLITLIKSGGISKCMNNWQHNKESGGE-  283
sp|P46200|MYB_BOVIN     IKNHWNSTMRRKVEQE---GYLQESSK-----ASQPAVT-TSFQKNSHLHGFTHA------  226
sp|Q6R032|MB3R5_ARATH   IKNHWNSSLKKKLEFYLATGNLPPPAS----KFIVLK----DIADGDRDSK---------  258
sp|Q8LHU7|MB3R2_ORYSJ   IKNHWNSSLRKKQDMYNTSNNMVVPKL------LVHD-----KFKDKPKLMAN--------  249
sp|Q94FL9|MB3R4_ARATH   IKNHWHSSVKKKLDSYMSSGLLDQYQA----MPLAPYER-SSTL-QSTFMQSNIDGNGCL  223
sp|Q9S7G7|MB3R1_ARATH   IKNHWNSSVKKKLDSYYASGLLDQCQS----SPLIALQN-KSIASSSSWMHSNGDEGSSR  230

sp|P04197|MYB_DROME     ---AVNKSENADGASVTA----------VKGGD------------------------  303
sp|P46200|MYB_BOVIN     ---------------------PPSA-----------QLPPAG-----QPSVNSDY  244
sp|Q6R032|MB3R5_ARATH   ----------QS--------SATKPFKD------------------SDSLTQTSS----  277
sp|Q8LHU7|MB3R2_ORYSJ   EGH--------LD-------LNKAPIIN-------------------SKDQPGTAH----  271
sp|Q94FL9|MB3R4_ARATH   NGQAENEIDSRQNSSMVGCSLSARDFQN----GTIN-IGHDFHPCGNSQENEQTAYHSEQ  278
sp|Q9S7G7|MB3R1_ARATH   PGVDAEESECSQASTVFS--QSTNDLQDEVQRGNEEYYMPEFHS-GTEQQIS-NAASHAE  286

sp|P04197|MYB_DROME     ------------------------------------------LAQESQDDHQK-----------G  315
sp|P46200|MYB_BOVIN     PYYHISEAQN-VSSHVPYPVALHVNIVNVPQPAAAAIQRHYND-E--DPE---KEKRIKE  297
sp|Q6R032|MB3R5_ARATH   --------------------------GNT--DSNEVGRDHF------------------  290
sp|Q8LHU7|MB3R2_ORYSJ   -----------------------RSN--CSGFLSRSSL------------------  284
sp|Q94FL9|MB3R4_ARATH   FYYPELEDISVSISEVSYDKEDCSQFPDHN--VSTSPSQDYQFDFQELSDI---------S  328
sp|Q9S7G7|MB3R1_ARATH   PYYPSFKDVKIVVPEISCETECSKKFQNLN--CSHELRTTTAT-EDQLPGVSNDAKQDRG  343
```

```
sp|P04197|MYB_DROME     SNLAHLSMQHLIKLTMPRQTPIILKRTRKHIPETH--------HQAGCSSSETFNQEEAA 367
sp|P46200|MYB_BOVIN     LELLLLMSTENEL----KGQQALPTQNHTCSYPGWHSTTIADHTRPHGDSAPVSCLE-EHH 352
sp|Q6R032|MB3R5_ARATH   -----------------DSSSA----------------------------LLE-EVA 301
sp|Q0JHU7|MB3R2_ORYSJ   -----------------PTAQP----------------------------LT----- 291
sp|Q94FL9|MB3R4_ARATH   LE-MRHNMSEI------PMPYT----------------------------KE-SKE 348
sp|Q9S7G7|MB3R1_ARATH   LELLTHNMDNG-----GKNQA-----------------------------LQQ-DFQ 365


sp|P04197|MYB_DROME     GNARSRPPSSPVIS------PIKSLPFS-------PSHFLKSPCLTTFEDMDLRASTPVT 414
sp|P46200|MYB_BOVIN     --STPSLPADPGSLPEESASPARCMIFHQSTILDNVKNLLEFAETLQFIDSFLNTSNNHE 410
sp|Q6R032|MB3R5_ARATH   ----------------------ASRRIG-----VNEYA------CSPVEYKPQL------- 322
sp|Q0JHU7|MB3R2_ORYSJ   ----------------------SREAS-----VVDG--------SAVTLVAQALESDSVRG 317
sp|Q94FL9|MB3R4_ARATH   --STLGAPNSTLNIDVATYTNSANVL-----TPETE------CCRVLFPDQESEGHSVSR 395
sp|Q9S7G7|MB3R1_ARATH   --SSVRLSDQP-FLSNSDTDPEAQTL-----ITDEE------CCRVLFPDNMKDSSTSSG 411


sp|P04197|MYB_DROME     KV-----------YNR----------VGMEIKKEM--E-TSSIE-TPHKSQLGP--RT 445
sp|P46200|MYB_BOVIN     NLDLEMPSLTSTPLNGHKLTVTTPFHRDQTVKIQKENTIFRTPAIKRSILEGS--PR--T 466
sp|Q6R032|MB3R5_ARATH   ---------PNLEPISE-E-------------------VRINSKAYF------------ 340
sp|Q0JHU7|MB3R2_ORYSJ   KG-------LEIDSVHEKG---------------------LEVNSAPDH------------T 339
sp|Q94FL9|MB3R4_ARATH   SLTQEP--NEFNQVDRRD------------P-------ILYSSASDRQISEATKSPTQSS 434
sp|Q9S7G7|MB3R1_ARATH   EQGR----NRVDPQNG-K---------G-------SLCSQAAETHAHETGKVPAL-- 445


sp|P04197|MYB_DROME     PTPFKKALAAIGKKRDGRRYEPSSPSSLVEDLAEIIHEEHLSNSLTANNS----KNMGAA 501
sp|P46200|MYB_BOVIN     PTPFKHALTAQEIKYGPLKMLFQTPSHLVEDLQDEIKQESDESGIVAEFQENGQPLLKKI 526
sp|Q6R032|MB3R5_ARATH   --ERSIQRKVE-NGFGT------------------------------------ 354
sp|Q0JHU7|MB3R2_ORYSJ   GNSHTIQLEAA-PSKGE---------------------------------------- 355
sp|Q94FL9|MB3R4_ARATH   SSRFT--ATAA-SGKGTLRPAPLIISPDKYS-------KKSSGLICHPF---------E 474
sp|Q9S7G7|MB3R1_ARATH   --PWH--PSSS-EGLAGHNCVPLLDSDLKDSLLP--RNDSNAPIQGCRLF--------G 489


sp|P04197|MYB_DROME     DQNSTLSTEYNAQSPPHMKRARKSLLSTWSSNHP---------------YNAGSAKRIQPF 547
sp|P46200|MYB_BOVIN     KQEVESPTD---------KAGNFFCSNHWEGDSLN---------TQLFT-----QASPV 562
sp|Q6R032|MB3R5_ARATH   ----PK------------------------HGNLYYKSPLDYYFPSEADLQ-- 377
sp|Q0JHU7|MB3R2_ORYSJ   -AELSL-----------------------KNEARSLGPLCYQIPNMEDVVPV 383
sp|Q94FL9|MB3R4_ARATH   -VEPKCTTN---------GNGSFICIGDPSSSTCVDEGTNNSSEEDQSYHVNDPKKLVPV 524
sp|Q9S7G7|MB3R1_ARATH   ATELECKTD---------TNDGFIDTYGHVTSHGNDDNGGFPEQQGLSYIPKDSLKLVPL 540


sp|P04197|MYB_DROME     ETE---TPSKFLTSPG--DIL----KD--TLC------SEQDLPFD---------EGRKE 581
sp|P46200|MYB_BOVIN     ADMPNILTSSVLMTPVSED------EDNVLKAFTVPKSRSLASPLQPC------------ 604
sp|Q6R032|MB3R5_ARATH   ------------------HMYG--YE--C--------------------- 384
sp|Q0JHU7|MB3R2_ORYSJ   SSS----------LFSDHLTG--NHTSE---------------------- 399
sp|Q94FL9|MB3R4_ARATH   NDFASLAEDRPHSLPKHEPNMTNEQHHEDMGASSSLGFPSFDLPVFNCDLLQSKNDPLHD 584
sp|Q9S7G7|MB3R1_ARATH   NSFSSPSRVNKIYFPIDDKPAE---KDKGALCYEPPRFPSADIPFFSCDLVPSNSDLRQE 597


sp|P04197|MYB_DROME     NRPFHNRRINKYRGGLTYDHVIDPKWARVACGKSRDQMFMEEQAYACLKNLSCISR---- 637
sp|P46200|MYB_BOVIN     --------------N----GAWESASCGKTDDQMTASGQSRKYVN----------- 631
sp|Q6R032|MB3R5_ARATH   ----GCSPGAA--SPVSLMTTPCNKDSGLTATRSPESFLR-EAARTFPNTPSIFRKRRKV 437
sp|Q0JHU7|MB3R2_ORYSJ   ----HCGDDIL--SPAGCT-TPPPTKGKLTSQLSVDSILK-SAANSFPGTPSILKRKRD 451
sp|Q94FL9|MB3R4_ARATH   YSPLGIRKLLM--STMTCM-SPLRLWESPTGK-----------KTLVGAQSILRKRTRD 629
sp|Q9S7G7|MB3R1_ARATH   YSPFGIRQLMI--SSMNCT-TPLRLWDSPCHDRSPDVMLN-DTAKSFSGAPSILKKRHRD 653
```

```
sp|P04197|MYB_DROME     ------------SLNFEKQKCLVNSFDRFGSL---------------------- 657
sp|P46200|MYB_BOVIN     ----------------------AFSTRTLVM---------------------- 640
sp|Q6R032|MB3R5_ARATH   VLAAKTDAVVVV----NGV---VKEVDRKEESKDM---RK---SL----LLETTDNCSDD 480
sp|Q0JHU7|MB3R2_ORYSJ   KSTPVSASEMKI----SGS---NT--DRFYTPMGM---EP---ATATPESFKT-TSFLSL 495
sp|Q94FL9|MB3R4_ARATH   LLTPLSEKRSDKKLEIDIAASLAKDFSRLDVMFDETENRQSNFGNSTGVIHGDRENHFH- 688
sp|Q9S7G7|MB3R1_ARATH   LLSPVLDRRKDKKLKRAATSSLANDFSRLDVMLDEGDDCMTSRPSESPE---DK-NICAS 709


sp|P04197|MYB_DROME     ---------------------------------------------------- 657
sp|P46200|MYB_BOVIN     ---------------------------------------------------- 640
sp|Q6R032|MB3R5_ARATH   EELGLGNGNAFNLSPPYRLRAKRTAVIKSR--QLEFTSEKEKQPDNEIEFTSAKEKQPDNE 538
sp|Q0JHU7|MB3R2_ORYSJ   GSLDGSVKSFDVSPQYRARSKRMALTKTVEKQLDFSSDGLDTCGSEILNSSCNNSQSTLS 555
sp|Q94FL9|MB3R4_ARATH   -ILNGDGEEWSGKP-SSLFSHRMPEETM---HIRKSLEKVDQICMEANVREKDDSEQDVE 743
sp|Q9S7G7|MB3R1_ARATH   PSIARDN--RNCAS-ARLYQEMIPIDEE----PKETLESGGVTSMQNEN-GCNDGGASAK 761


sp|P04197|MYB_DROME     ---------------------------------------------------- 657
sp|P46200|MYB_BOVIN     ---------------------------------------------------- 640
sp|Q6R032|MB3R5_ARATH   IK-----------T----SEEDKPV-------------------------- 548
sp|Q0JHU7|MB3R2_ORYSJ   IT------------EAPKLKEKEHAVQL----------ENLTKNFAHTTNLDVT------ 587
sp|Q94FL9|MB3R4_ARATH   NVEFFSGILSEHNTGKPVLSTPGQSVTKAEKAQVSTPRNQLQRTLHATSNKEHHSPSSVC 803
sp|Q9S7G7|MB3R1_ARATH   NV--------------------------------SPSLSLHI----------- 771


sp|P04197|MYB_DROME     ---------------------------------------------------- 657
sp|P46200|MYB_BOVIN     ---------------------------------------------------- 640
sp|Q6R032|MB3R5_ARATH   ---------------------------------------------------- 548
sp|Q0JHU7|MB3R2_ORYSJ   ---------------------------------------------------- 587
sp|Q94FL9|MB3R4_ARATH   LVINSPSRARNKEGHLVDNGTSNENFSIFCGTPFRRGLESPSANKSPFYINSLLPSPRFD 863
sp|Q9S7G7|MB3R1_ARATH   -----------------------------------------IW---------- 773


sp|P04197|MYB_DROME     ---------------------------------------------------- 657
sp|P46200|MYB_BOVIN     ---------------------------------------------------- 640
sp|Q6R032|MB3R5_ARATH   ---------------------------------------------------- 548
sp|Q0JHU7|MB3R2_ORYSJ   ---------------------------------------------------- 587
sp|Q94FL9|MB3R4_ARATH   TDLTIEDMGYIFSPGERSYESIGVMTQINEHTSAFAAFADAMEVSISPTNDDARQKKELD 923
sp|Q9S7G7|MB3R1_ARATH   ---------------------------------------------------- 773


sp|P04197|MYB_DROME     ----------------------------------   657
sp|P46200|MYB_BOVIN     ----------------------------------   640
sp|Q6R032|MB3R5_ARATH   ----------------------------------   548
sp|Q0JHU7|MB3R2_ORYSJ   ----------------------------------   587
sp|Q94FL9|MB3R4_ARATH   KENNDPLLAERRVLDFNDCESPIKATEEVSSYLLKGCR   961
sp|Q9S7G7|MB3R1_ARATH   ----------------------------------   773
```

Appendix 5 – T-Blast

```
sp|P04197|MYB_DROME     MASASTENGEELMNYGSNSDSEESEYSENEDTQVCDKDSQQ-------NS
sp|P46200|MYB_BOVIN     MARRPRHSIY--------------------------------------SSDE
sp|Q0JHU7|MB3R2_ORYSJ   MGAMAMVEQE----------------------GCVENRQPLAASSSSVS
sp|Q6R032|MB3R5_ARATH   MSSSSNPPVCSP---------EKEERSEMKIEIQCMENKQPLAASCSSAS
sp|Q94FL9|MB3R4_ARATH   MEAESSTPQ-----------------------------------------
sp|Q9S7G7|MB3R1_ARATH   MKREMKAPTT----------------------------------------
                        *


sp|P04197|MYB_DROME     N------ADSGYPLDSPELQDSKTTG-QKGQNKSGKTSIGAVHPNYGFGK
sp|P46200|MYB_BOVIN     D-------------DEDIEMCDHDY-DGLLPKSGKRHL--------GKT
sp|Q0JHU7|MB3R2_ORYSJ   DGSSYGGGGGGLAQMSPPVSSSANSI-SGLRRTSGPIRR--------AKG
sp|Q6R032|MB3R5_ARATH   E------GSGCFFLKSPEIATPATVS-SFPRRTSGPMRR--------AKG
sp|Q94FL9|MB3R4_ARATH   --------------------ERIP---KLRHGRTSGPARR----S---TRG
sp|Q9S7G7|MB3R1_ARATH   ------------P-----LESLQGDLKGKQGRTSGPARR----S---TKG
                               .**


sp|P04197|MYB_DROME     RWSKSEDVLLKQLVET-HGENWEIIGPHFKDRLEQQVQQRWAKVLNPELI
sp|P46200|MYB_BOVIN     RWTREEDEKLKKLVEQNGTDDWKVIANYLPNRTDVQCQHRWQKVLNPELI
sp|Q0JHU7|MB3R2_ORYSJ   GWTPEEDETLRKAVEAYKGRNWKKIAECFPYRTEVQCLHRWQKVLNPELI
sp|Q6R032|MB3R5_ARATH   GWTPEEDETLRRAVEKYKGKRWKKIAEFFPERTEVQCLHRWQKVLNPELV
sp|Q94FL9|MB3R4_ARATH   QWTAEEDEILRKAVHSFKGKNWKKIAEYFKDRTDVQCLHRWQKVLNPELV
sp|Q9S7G7|MB3R1_ARATH   QWTPEEDEVLCKAVERFQGKNWKKIAECFKDRTDVQCLHRWQKVLNPELV
                        *: .** *: :.    *: *.  : *: *  :** ******:


sp|P04197|MYB_DROME     KGPWTRDEDDMVIKLVRNFGPKKWTLIARYLNGRIGKQCRERWHNHLNPN
sp|P46200|MYB_BOVIN     KGPWTKEEDQRVIELVQKYGPKRWSVIAKHLKGRIGKQCRERWHNHLNPE
sp|Q0JHU7|MB3R2_ORYSJ   KGPWTQEEDDQIIDLVKKYGPTKWSVIAKALPGRIGKQCRERWHNHLNPG
sp|Q6R032|MB3R5_ARATH   KGPWTQEEDDKIVELVKKYGPAKWSVIAKSLPGRIGKQCRERWHNHLNPG
sp|Q94FL9|MB3R4_ARATH   KGPWTKEEDEMIVQLIEKYGPKKWSTIARFLPGRIGKQCRERWHNHLNPA
sp|Q9S7G7|MB3R1_ARATH   KGPWSKEEDNTIIDLVEKYGPKKWSTISQHLPGRIGKQCRERWHNHLNPG
                        ****:;:**: :: .*..;:** :*: *:: *  ****************


sp|P04197|MYB_DROME     IKKTANTEKEDEIIYQAHLELGNQWAKIAKRLPGRTDNAIKNHWNSTMRR
sp|P46200|MYB_BOVIN     VKKTSWTEEEDRIIYQAHKRLGNRWAEIAKLLPGRTDNAIKNHWNSTMRR
sp|Q0JHU7|MB3R2_ORYSJ   IRKDAWTTEEEQALINAHRIYGNKWAEIAKVLPGRTDNSIKNHWNSSLRK
sp|Q6R032|MB3R5_ARATH   IRKDAWTVEEESALMNSHRMYGNKWAEIAKVLPGRTDNAIKNHWNSSLKK
sp|Q94FL9|MB3R4_ARATH   INKEANTQEEELLLIRAHQIYGNRWAELTKFLPGRSDNGIKNHWHSSVKK
sp|Q9S7G7|MB3R1_ARATH   INKNANTQEEELTLIRAHQIYGNKWAELMKFLPGRSDNSIKNHWNSSVKK
                        :.* :** :*: :.;:*  **;**:: * ****:**.*****:*;:::


sp|P04197|MYB_DROME     KYDVE--RRSV-NASGSDLKSSRTHLITLIKSGGISKCMNNMQHNKESG-
sp|P46200|MYB_BOVIN     KVEQEGYLQESSKASQ----PA---------------------------
sp|Q0JHU7|MB3R2_ORYSJ   KQDMY--NTSN-NMVV----PKL---L----------------------
sp|Q6R032|MB3R5_ARATH   KLEFY--LATG-NLPP----PASK-FIVLK-DI----------------
sp|Q94FL9|MB3R4_ARATH   KLDSY--MSSG-LLDQ----YQAMPLAPYE-RSS-TLQSTFMQSNIDGNG
sp|Q9S7G7|MB3R1_ARATH   KLDSY--YASG-LLDQ----CQSSPLIALQ-NKSIASSSSNMHSNGDEGS
                        * :
```

```
sp|P04197|MYB_DROME     ---GEAVNKSENADGASVT--AVKGGDLAQESQ------------------
sp|P46200|MYB_BOVIN     --------------------VTTSFQKNSHLNGF-----T---------
sp|Q0JHU7|MB3R2_ORYSJ   --------------------VHDKFKDKPKLMAM------EGHLDLN
sp|Q6R032|MB3R5_ARATH   -----ADGDRDSKQSS-------ATKPFKDSDSLTQT-----SSGNTDSN
sp|Q94FL9|MB3R4_ARATH   CLNGQAENEIDSRQNSSMVGCSLSARDFQNGTINIGH-----D-------
sp|Q9S7G7|MB3R1_ARATH   SRPGVDAEESECSQASTVF--SQSTNDLQDEVQRGNEEYYMPE-------
                                              : .


sp|P04197|MYB_DROME     ---------------DDHQK----------------------------
sp|P46200|MYB_BOVIN     HAPPSAQLPPAGQPSVNSDYPYYHI-------------SEAQNV----
sp|Q0JHU7|MB3R2_ORYSJ   KAP-IINSKDQPGTAH---------------------RSNCSGFL---
sp|Q6R032|MB3R5_ARATH   -----------EVGRDHFDSS-SALLEEVAASR-RIGVNEYACSPVE---
sp|Q94FL9|MB3R4_ARATH   FHP-CGNSQENEQTAYHSEQFYYPELEDISVSISEVSYDMEDCSQFPDHN
sp|Q9S7G7|MB3R1_ARATH   FH--SGTEQQISNAASHAEP-YYPSFKDVKIVVPEISCETECSKKFQNLN


sp|P04197|MYB_DROME     ---------------GSNLAHLSM---------------
sp|P46200|MYB_BOVIN     -------------------------------------
sp|Q0JHU7|MB3R2_ORYSJ   -------------------------------------
sp|Q6R032|MB3R5_ARATH   -------------------------------------
sp|Q94FL9|MB3R4_ARATH   VSTSPSQDYQFDFQELSDIS-LEM------------
sp|Q9S7G7|MB3R1_ARATH   CSH-------------------ELRTTTATEDQLPGVSNDAKQDRGLELL


sp|P04197|MYB_DROME     QH-LIKLTMP---------------------
sp|P46200|MYB_BOVIN     ---SSHVPYPV--------------------AL-------------HVN
sp|Q0JHU7|MB3R2_ORYSJ   --SRSSLPTAQPLTSREAS--------VVDGSA-------------VTL
sp|Q6R032|MB3R5_ARATH   --YKPQLPNLEPIS------------------------
sp|Q94FL9|MB3R4_ARATH   RHNMSEIPMPYTKESKESTLGAPNSTLNIDVAT--------------YTN
sp|Q9S7G7|MB3R1_ARATH   THNMDN-------------GGKNQALQQDFQSSVRLSDQPFLSNSDTDP


sp|P04197|MYB_DROME     -----------------------------RQTPIILKR
sp|P46200|MYB_BOVIN     IVNVPQPAAAAIQRHYNDEDPEK-------------EKRIKELELLLMST
sp|Q0JHU7|MB3R2_ORYSJ   VAQALE-------------------------SDSVRGKG--LEID
sp|Q6R032|MB3R5_ARATH   -----E------------------------EVRINSKA--YFER
sp|Q94FL9|MB3R4_ARATH   SANVLTPETECCRVLFPDQESEGHSVSRSLTQEPNEFNQVDRRDPILYSS
sp|Q9S7G7|MB3R1_ARATH   EAQTLITDEECCRVLFPDNMKDSSTS--SGEQGRNMVDPQNG-KGSLCSQ


sp|P04197|MYB_DROME     -TRKHIPETHHQAG---CSSS-----E----------------
sp|P46200|MYB_BOVIN     ENELKGQQALPTQN---HTCSYPGWHSTTIADHTRPH-------------
sp|Q0JHU7|MB3R2_ORYSJ   SVHEKGLEVNSAPD---HTGN------SWTIQLEAAPSKG--------
sp|Q6R032|MB3R5_ARATH   SIQRKVENGFGTPK---H------------------
sp|Q94FL9|MB3R4_ARATH   ASDRQISEATKSPT---QSSS-----SR-FTATAASGKGTLRPAPLIISP
sp|Q9S7G7|MB3R1_ARATH   AAETHAHETGKVPALPWHPSS-------------SEGLAGHNCVPLLDSD
                            : :


sp|P04197|MYB_DROME     -------------------------------------
sp|P46200|MYB_BOVIN     ----------------------------GDSAPVSCLEE
sp|Q0JHU7|MB3R2_ORYSJ   -------------------------------EAE------
sp|Q6R032|MB3R5_ARATH   -------------------------------------
sp|Q94FL9|MB3R4_ARATH   DKYSK------KSSGLICHPFE-VEPKCTTNGNGSFICIGDPSSSTCVDE
sp|Q9S7G7|MB3R1_ARATH   LKDSLLPRNDSNAPIQGCRLFGATELECKTDTNDGFIDTYGHVTSHGNDD
```

```
sp|P04197|MYB_DROME     ------------------------------------------------
sp|P46200|MYB_BOVIN     HHSTPSLPADPGSLPEESASPARCMIFHQSTILDNVKNLLEF--------
sp|Q0JHU7|MB3R2_ORYSJ   ------------------------------------------------
sp|Q6R032|MB3R5_ARATH   ------------------------------------------------
sp|Q94FL9|MB3R4_ARATH   GTNNSSEE----------------DQSYHVNDPKKLVPVNDFASLAE
sp|Q9S7G7|MB3R1_ARATH   NGGFPEQQ----------------GLSYIPKDSLKLVPLNSFSSPSR


sp|P04197|MYB_DROME     ------------------------------------------------
sp|P46200|MYB_BOVIN     ------------------------------------------AET
sp|Q0JHU7|MB3R2_ORYSJ   ------------LSLKNEARSLGPLCYQIPNME---------DVVPVSSS
sp|Q6R032|MB3R5_ARATH   ------------------GNLYYKSPL-D---------YYFPSEAD
sp|Q94FL9|MB3R4_ARATH   DRPHSLPKHEPNMTNEQHHEDMGAS---SSLGFPSFDLPVFNCDLLQSKND
sp|Q9S7G7|MB3R1_ARATH   VNKIYFP-----IDDKPAEKDKGALCYEPPRFPSADIPFFSCDLVPSNSD


sp|P04197|MYB_DROME     --------------------------TFNQEEAAG-NARSRPPSSPVIS
sp|P46200|MYB_BOVIN     LQFIDSF-LNTSNNHENLDLEMPSLT-STPLNGHKLTVTTPFHRDQTVKI
sp|Q0JHU7|MB3R2_ORYSJ   L-FSDHLTGNHTSEHCGDDILSPAGC-TTPPPTKGK-LTSQLSVDSILKS
sp|Q6R032|MB3R5_ARATH   L-Q--HMYG--YECGCSPGAASPVSLMTTPCNKDSG-LTATRSPESFLRE
sp|Q94FL9|MB3R4_ARATH   P-L--HDY---SPLGIRKLLMSTMTC-MSPLRLWES-P-----------T
sp|Q9S7G7|MB3R1_ARATH   L-R--QEY---SPFGIRQLMISSMNC-TTPLRLWDS-PCHDRSPDVMLND


sp|P04197|MYB_DROME     -PIKSLPFSPSHF----LKSPCLTTFEDMDLRASTPVTKVYNRVGMEIKK
sp|P46200|MYB_BOVIN     QKENTIFRTPAIKRSILEGS--------------------------
sp|Q0JHU7|MB3R2_ORYSJ   -AANSFPGTPSI----LKRR------------------------
sp|Q6R032|MB3R5_ARATH   -AARTFPNTPSI----FRKR------------------------
sp|Q94FL9|MB3R4_ARATH   -GKKTLVGAQSI----LRKR------------------------
sp|Q9S7G7|MB3R1_ARATH   -TAKSFSGAPSI----LKKR------------------------
                                   .:: : :


sp|P04197|MYB_DROME     EMETSSIETPHKSQLGPRTPTPFKKALAAIGKKRDGRRYEPSSPS--SL-
sp|P46200|MYB_BOVIN     -----------------PRTPTPFKHALTAQEIKYGPLKMLPQTPS--HL-
sp|Q0JHU7|MB3R2_ORYSJ   -----------K-----RDKSTPVSASEM--K----ISG-SNT
sp|Q6R032|MB3R5_ARATH   -----------R-----KVVLAAKTDAVV--V----VNGVVKEV
sp|Q94FL9|MB3R4_ARATH   -----------T-----RDLLTPLSEKRSDKKLEIDIAA--SL-
sp|Q9S7G7|MB3R1_ARATH   -----------H-----RDLLSPVLDRRKDKKLKRAATS--SL-
                                 : : :.  :


sp|P04197|MYB_DROME     -----VEDLAEI--IHEEHLSN------SLTANNS-------------KM
sp|P46200|MYB_BOVIN     -----VEDLQDE--I-------------------------
sp|Q0JHU7|MB3R2_ORYSJ   DRFYTPMGM-EP----ATATPESFKTT-------------------
sp|Q6R032|MB3R5_ARATH   DRKEESKDM-RKSLLLETTDNC-----S-------------
sp|Q94FL9|MB3R4_ARATH   ------AKDFSRLDVMFDETENR-----QS-NFGNSTGVIHGDRENHFHIL
sp|Q9S7G7|MB3R1_ARATH   ------ANDFSRLDVMLDEGDDC-----M------------
                            .:
```

```
sp|P04197|MYB_DROME     --------NSTLSTE------------------------
sp|P46200|MYB_BOVIN     --KQESDESGIVAEFQENGQPLLKKIKQEV---------ESPTDKAGNFFC
sp|Q0JHU7|MB3R2_ORYSJ   ------------------------------------
sp|Q6R032|MB3R5_ARATH   ------------------------------------
sp|Q94FL9|MB3R4_ARATH   DVENVEFFSGILSEHN-TGKPVLSTFGQSVTKAEKAQVSTPRNQLQRTLM
sp|Q9S7G7|MB3R1_ARATH   ------------------------------------


sp|P04197|MYB_DROME     -----------------YNAQSPPHMKRARKSLLS-TW-------
sp|P46200|MYB_BOVIN     SNHW-EGDS-LN--TQLFTQASPVA--DMPNILTSS----------
sp|Q0JHU7|MB3R2_ORYSJ   -----SLGSLDGSVKSFDVSPQYRARSKRMALTKT-----------V-
sp|Q6R032|MB3R5_ARATH   ------ELG-LN--GNAFNLSPPYRLRAKRTAVIKS----------
sp|Q94FL9|MB3R4_ARATH   ATSNKEHHS-PS--SVCLVINSPSRARNKEGHLVDNGTSNENFSIFCGTP
sp|Q9S7G7|MB3R1_ARATH   ------SES-PE--DKN-ICASPSIARDNRNCASAR----------
                                              .


sp|P04197|MYB_DROME     -SSNH------------------------PYNAGSAKRIQPFETET
sp|P46200|MYB_BOVIN     ---V------------------------L--M-TPVSEDE-----
sp|Q0JHU7|MB3R2_ORYSJ   -EKQL-----------------------D--FSSDGLDTC-----
sp|Q6R032|MB3R5_ARATH   --RQL-----------------------E--FTSEKEKQP-----
sp|Q94FL9|MB3R4_ARATH   FRRGLESPSAWKSPFYINSLLPSPRFDTDLTIED--M-GYIFSPG-----
sp|Q9S7G7|MB3R1_ARATH   ---LY-----------------------QE--M-IPIDEEP-----


sp|P04197|MYB_DROME     PSKFLTSPGDILKDTLCSEQDLPFDEGRKENRPFHNRRINKYRGGLTYDH
sp|P46200|MYB_BOVIN     -------------------------DNVLKAFTVPKSRSLASPLQ
sp|Q0JHU7|MB3R2_ORYSJ   ---------------------------GSEI
sp|Q6R032|MB3R5_ARATH   ---------------------------DNEI
sp|Q94FL9|MB3R4_ARATH   ---------------------------ERSY
sp|Q9S7G7|MB3R1_ARATH   ---------------------------KETL


sp|P04197|MYB_DROME     VIDPKWARVAC--GKSRDQMF------------------MEEQAYAC
sp|P46200|MYB_BOVIN     PCNGAWESASC--GKTDDQMT------------------ASGQSRKY
sp|Q0JHU7|MB3R2_ORYSJ   ------LNSSCNNSQSTLSIT------------------EAPKLKEK
sp|Q6R032|MB3R5_ARATH   ------EFTSAKEKQPDNEIK------------------TSEEDKP-
sp|Q94FL9|MB3R4_ARATH   ------ESIGVMTQINEHTSAFAAFADAMEVSISPTNDDARQKKELDKEN
sp|Q9S7G7|MB3R1_ARATH   ------ESGGVTSMQNENGCN------------------DGGASAKN
                              .


sp|P04197|MYB_DROME     LKN--------LSCISRSLNFEKQKCLVNSFDRFGSL------
sp|P46200|MYB_BOVIN     VNA--------F-ST-RTLVM--------------------
sp|Q0JHU7|MB3R2_ORYSJ   EHAVQLENLTKNFAHTTNLDVT--------------------
sp|Q6R032|MB3R5_ARATH   ----------------------V--------------
sp|Q94FL9|MB3R4_ARATH   NDP-------LLAERRVLDFNDCESPIKATEEVSSYLLKGCR
sp|Q9S7G7|MB3R1_ARATH   VSP-------S----LSLHI--------------IW---
```
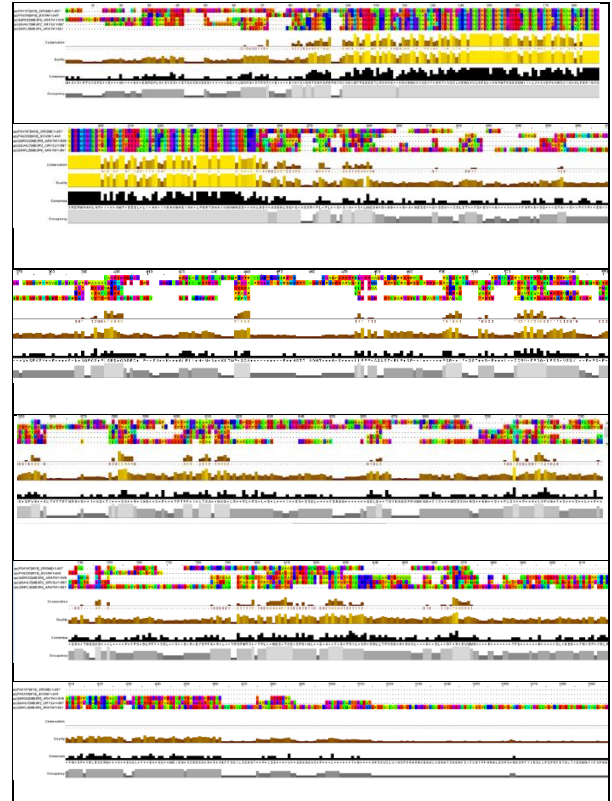
Appendix 6 – Clustal jal view alignment

Appendix 7 – Tblast jal-veiw.