GTPB
The Gulbenkian Training Programme in Bioinformatics
(Since 1999)
Pedro Fernandes, Organiser
INSTITUTO GULBENKIAN DE CIÊNCIA

IB16S

Introductory Bioinformatics

12-16 December 2016

(Second 2016 run of this Course)

Basic Bioinformatics Sessions

Practical 4: Primer Design

Monday 5 December 2016

Primer Design

To determine the presence or absence of the mutation we have detected, a test based on restriction maps could be employed. This approach is investigated in one of the supplementary exercises at the end of this book. In that extra exercise, it is shown there is more than one restriction enzyme whose cut site is dependant upon the mutation. With a little more work (with the same programs), we could easily ascertain exact restriction fragment sizes expected for selected enzyme(s) with the mutation and without it. As long as the differences were sufficiently unsubtle, a **R**estriction **F**ragment **L**ength **P**olymorphism (**RFLP**) test could be designed.

For a variety of reasons, including the ready availability and ever decreasing cost of sequencing, this is typically not the preferred way to proceed. It is normally preferable to use PCR to isolate the region around the mutation and sequence all individuals under examination. To do this, the first step would be to design suitable PCR primers. One program, in many different forms, is almost exclusively used for this purpose. The program is **primer3**. It is free and can be downloaded and run under linux and windows (at least). It is available as part of the **EMBOSS** package (**eprimer3**) and from a number of websites, including at the **M**assachusetts **I**nstitute of **T**echnology (**MIT**)[1]:

```
http://frodo.wi.mit.edu/
```

This site is popular with many users offering complete control over the various options offered by **primer3**.

Another excellent **primer3** web interface developed in the Netherlands is available at:

```
http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi
```

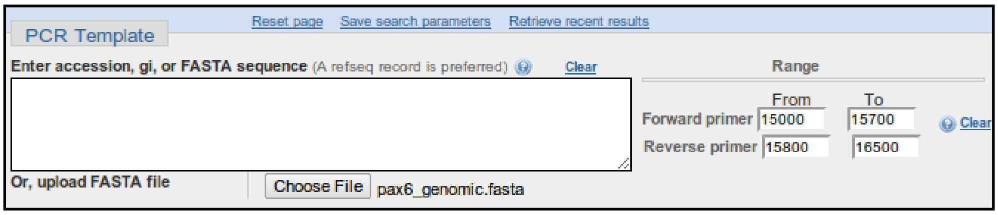
The site incorporates access to a **blast** search to check the uniqueness of the selected primers (important if unwanted PCR products are to be avoided).

Mostly because of its completely seamless inclusion of a **blast** search to compare potential primers with appropriate sequence collections, I suggest we here use **primer3** as implemented at the **NCBI**, even though it offers less than complete control over the execution of **primer3** itself. Go to:

```
http://www.ncbi.nlm.nih.gov
```

Click on the **BLAST** option. Select **Primer-BLAST** from the **Specialized BLAST** section.

Upload your genomic **PAX6** sequence using the **Browse** (or **Choose File**) button for the **PCR Template**.



Enter accession, gi, or FASTA sequence (A refseq record is preferred)

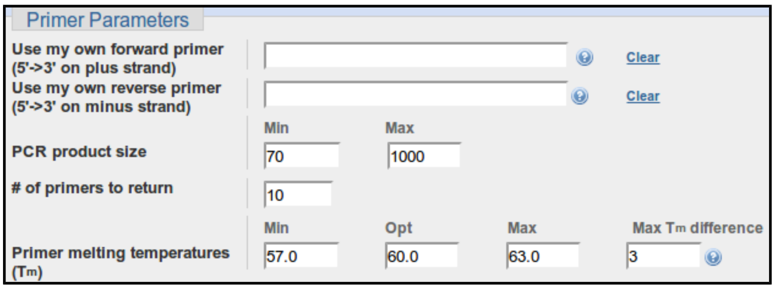
Or, upload FASTA file: pax6_genomic.fasta

	From	To
Forward primer	15000	15700
Reverse primer	15800	16500

You have established that the mutation of greatest interest is the **G/C** substitution at position **15714** of the genomic sequence copied from **Ensembl**. It is logical therefore to specify that this feature be included in the PCR product not too near either end. Accordingly, request the **Forward primer** to be chosen **From** the region starting at base pair **15000** and continuing **To** base pair **15700**. Set the range for the **Reverse primer** to be **From 15800** and **To 16500**.

The default **PCR product size** is specified in the **Primer Parameters** section as between **70** and **1000** base pairs. This seems fine.

I would not presume to advise you on the melting temperatures that were most suitable[2]. For this exercise, the defaults work splendidly.



Use my own forward primer (5'->3' on plus strand):

Use my own reverse primer (5'->3' on minus strand):

PCR product size: Min 70, Max 1000

# of primers to return: 10

Primer melting temperatures (Tm): Min 57.0, Opt 60.0, Max 63.0, Max Tm difference 3

By default, **primer-BLAST** will report the best **10** primer pairs it can find (**# of primers to return**). This is plenty for the exercise.

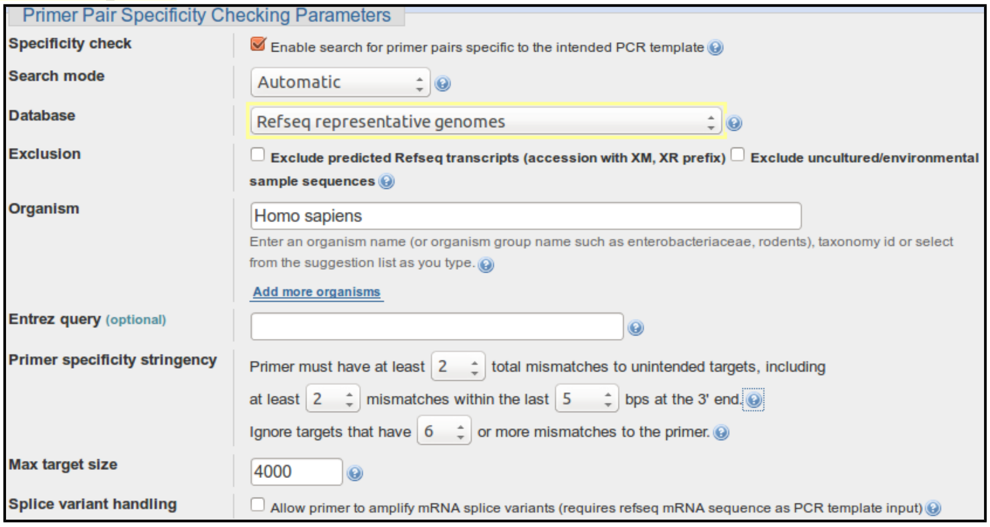
Do you think **10** primer pair suggestions is sufficient? If not, what number would you choose? ____________________

[1] The **MIT** now link to a newer version of **primer3** (**version 4.0.0**, soon **primer4** maybe?). Its URL is: **http://bioinfo.ut.ee/primer3/**. I have yet to investigate this version fully.

[2] My policy has been to not discuss parameters that pertain to the experimental conditions. In future versions of these notes, I will include discussion of some of these parameters. In the mean time, the ? buttons are very helpful. I would also suggest the **MIT** site (or the **Wageningen** site) for very readable explanations linked from every parameter. The full **primer3** manual can be found here.

In addition to running **primer3** to suggest primers, **Primer-BLAST** checks against the possibility of unwanted PCR products by comparing potential primers against an appropriate sequence database with **blast**.

In the **Primer Pair Specificity Checking Parameters** section, set the **Database** selection to **RefSeq representative genomes**. Leave the **Organism** set as **Homo sapiens**.



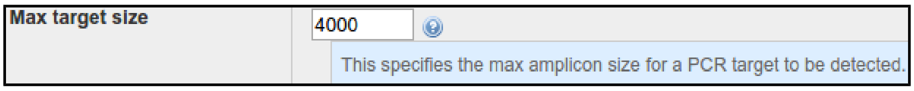
Primer Pair Specificity Checking Parameters

Specificity check	☑ Enable search for primer pairs specific to the intended PCR template ?
Search mode	Automatic ?
Database	Refseq representative genomes ?
Exclusion	□ Exclude predicted Refseq transcripts (accession with XM, XR prefix) □ Exclude uncultured/environmental sample sequences ?
Organism	Homo sapiens Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggestion list as you type. ? Add more organisms
Entrez query (optional)	?
Primer specificity stringency	Primer must have at least 2 total mismatches to unintended targets, including at least 2 mismatches within the last 5 bps at the 3' end. ? Ignore targets that have 6 or more mismatches to the primer. ?
Max target size	4000 ?
Splice variant handling	□ Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input) ?

You thus request each potential pair of PCR primers to be compared to the entire human genome. Thus unintended products of similar size to the intended product, can be identified.

The ideal conclusion is "just one product will be produced, on chromosome **11**, in the region of the **PAX6** gene".

Use the appropriate ? button to discover the purpose of the **Max target size** parameter.

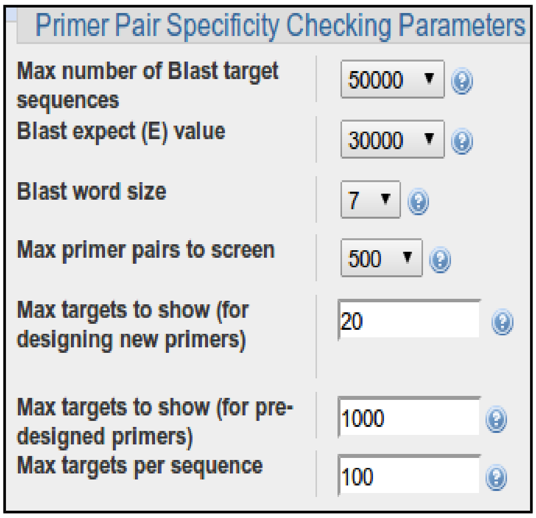


Max target size	4000 ? This specifies the max amplicon size for a PCR target to be detected.

This is a new parameter replacing a very different parameter, the purpose of which was somewhat less obvious. The reason for the **Max target size** parameter is surely pretty transparent, so maybe there is now less requirement to wake up its ? button? For the present, the maximum size of any proposed PCR product, in this instance, is **1,000** base pairs (the form default). So the greatest size of an unwanted product that might be a problem (the **Max target size**) must be small enough to potentially be mistaken for a real product of **1,000** base pairs. **4,000** base pairs seems a bit cautious to me? However, unless you feel strongly about the matter, accept the default value of **4000**.

What value would you choose here if you were looking for uncluttered results? ____________________

Before setting **primer-BLAST** going, click on the **Advanced parameters** button. Not really so **Advanced**? More **Avoidable** by those in a hurry. At the top are the **Primer Pair Specificity Checking Parameters** that control the way that **blast** is run. Note the ? buttons offering explanation.



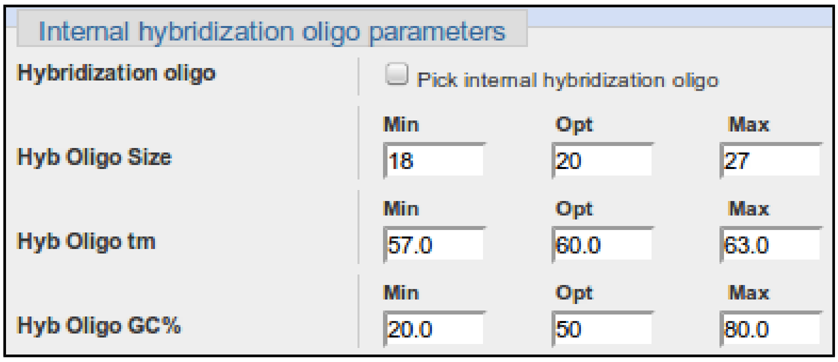
Primer Pair Specificity Checking Parameters

Max number of Blast target sequences	50000 ?
Blast expect (E) value	30000 ?
Blast word size	7 ?
Max primer pairs to screen	500 ?
Max targets to show (for designing new primers)	20 ?
Max targets to show (for pre-designed primers)	1000 ?
Max targets per sequence	100 ?

Note the very high default **Blast expect (E) value**, suggesting you will be interested in matches with your primers that might occur up too **30000** times by chance! This does make sense as the primers will be very short and so many good, even exact, "chance" matches might be expected against a large database.

Comment upon the small default value for the **Blast word size**? __________

____



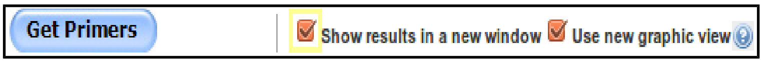
Internal hybridization oligo parameters

Hybridization oligo	□ Pick internal hybridization oligo		
	Min	Opt	Max
Hyb Oligo Size	18	20	27
	Min	Opt	Max
Hyb Oligo tm	57.0	60.0	63.0
	Min	Opt	Max
Hyb Oligo GC%	20.0	50	80.0

Note that you could get **primer-BLAST** to suggest an **Internal hybridisation oligo**, but decline the invitation this time.

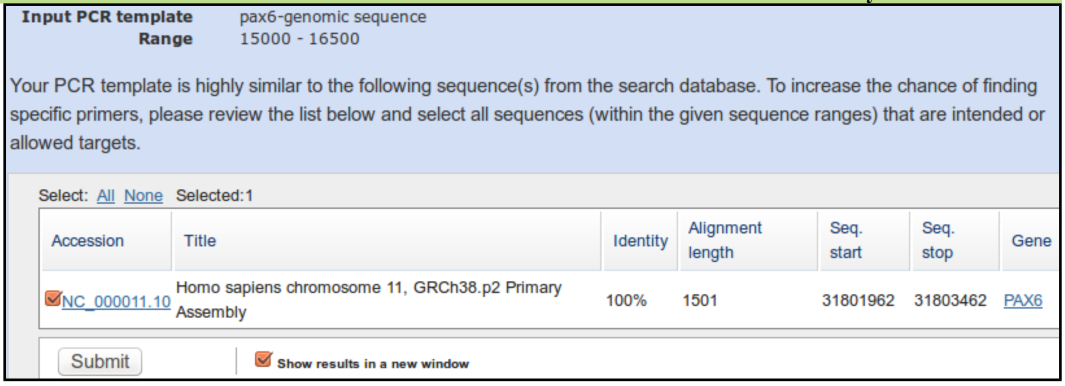
Accept all the **Advanced parameters** as they are. Ask **primer-BLAST** to **Show results in a new window**.

Click on the **Get Primers** button.



☑ Show results in a new window ☑ Use new graphic view ?

After a few moments of deep thought, **primer-BLAST** will notice that the template sequence you are using is **highly similar** (identical in fact) to part of an entry in the database being searched. Hardly surprising if one was to think about it.



Input PCR template pax6-genomic sequence
Range 15000 - 16500

Your PCR template is highly similar to the following sequence(s) from the search database. To increase the chance of finding specific primers, please review the list below and select all sequences (within the given sequence ranges) that are intended or allowed targets.

Select: All None Selected:1

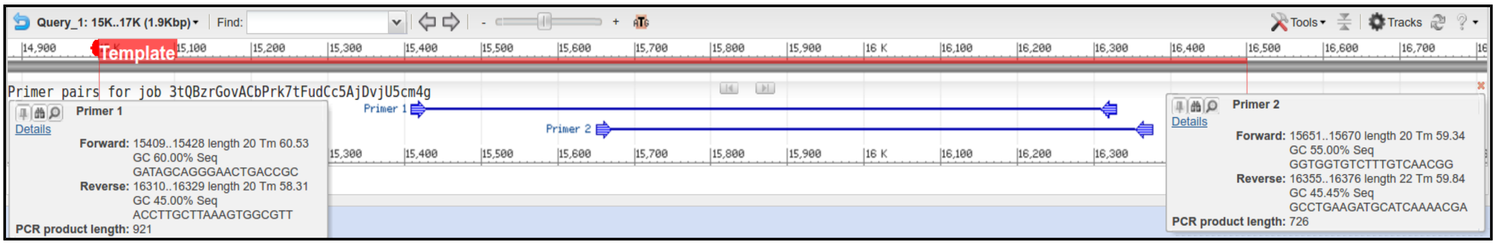
Accession	Title	Identity	Alignment length	Seq. start	Seq. stop	Gene
☑ NC_000011.10	Homo sapiens chromosome 11, GRCh38.p2 Primary Assembly	100%	1501	31801962	31803462	PAX6

☑ Show results in a new window

You are invited to select all listed regions (just one this time) where matches with primers are likely to be the intended product. In this case, that is the whole list of one, so click on the **All** button.

Every pair of primers that **primer3** selects ***must*** match this region of **Chromosome 11** as it is precisely the region investigated by **primer3** in the first place. This process avoids **blast** reporting intended products as unintended products. Finally, all is ready, so ask to **Show results in a new window** and then click on the **Submit** button.

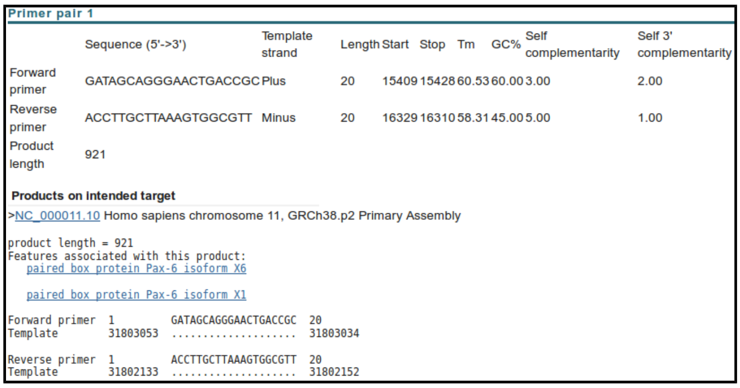
Once you have revelled in the opportunity to twiddle the fingers and scratch the ear(s) whilst **primers3** and **blast** go merrily about their appointed tasks, you will receive your results. These should look disarmingly like mine if all has gone well, in **Summary** and in **Detail**.



Primer pairs for job 3tQBzrGovACbPrk7tFudCc5AjDvjU5cm4g

Primer 1
Forward: 15409..15428 length 20 Tm 60.53 GC 60.00% Seq GATAGCAGGGAACTGACCGC
Reverse: 16310..16329 length 20 Tm 58.31 GC 45.00% Seq ACCTTGCTTAAAGTGGCGTT
PCR product length: 921

Primer 2
Forward: 15651..15670 length 20 Tm 59.34 GC 55.00% Seq GGTGGTGTCTTTGTCAACGG
Reverse: 16355..16376 length 22 Tm 59.84 GC 45.45% Seq GCCTGAAGATGCATCAAAACGA
PCR product length: 726



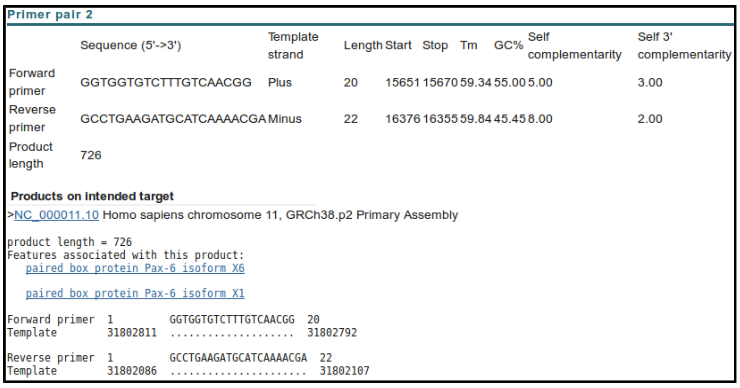
Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GATAGCAGGGAACTGACCGC	Plus	20	15409	15428	60.53	60.00	3.00	2.00
Reverse primer	ACCTTGCTTAAAGTGGCGTT	Minus	20	16329	16310	58.31	45.00	5.00	1.00
Product length	921								

Products on intended target

>NC_000011.10 Homo sapiens chromosome 11, GRCh38.p2 Primary Assembly

```
product length = 921
Features associated with this product:
   paired box protein Pax-6 isoform X6

   paired box protein Pax-6 isoform X1

Forward primer  1         GATAGCAGGGAACTGACCGC  20
Template        31803053  ....................  31803034

Reverse primer  1         ACCTTGCTTAAAGTGGCGTT  20
Template        31802133  ....................  31802152
```



Primer pair 2

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GGTGGTGTCTTTGTCAACGG	Plus	20	15651	15670	59.34	55.00	5.00	3.00
Reverse primer	GCCTGAAGATGCATCAAAACGA	Minus	22	16376	16355	59.84	45.45	8.00	2.00
Product length	726								

Products on intended target

>NC_000011.10 Homo sapiens chromosome 11, GRCh38.p2 Primary Assembly

```
product length = 726
Features associated with this product:
   paired box protein Pax-6 isoform X6

   paired box protein Pax-6 isoform X1

Forward primer  1         GGTGGTGTCTTTGTCAACGG  20
Template        31802811  ....................  31802792

Reverse primer  1         GCCTGAAGATGCATCAAAACGA  22
Template        31802086  ......................  31802107
```

Just **two** solutions met the default criteria for success used by **primer3**. Up to **10** were permitted[3]. Hovering over the graphical results will bring forth textual summaries. Try it. Note the rather ugly job identification! Clearly, the poetry generated for your results is extremely unlikely to be the same as illustrated.

Neither of your two suggested primer pairs should be associated with any unintended products, even with the very generous suggestion that products **4000** bases long should be considered a potential problem[4].

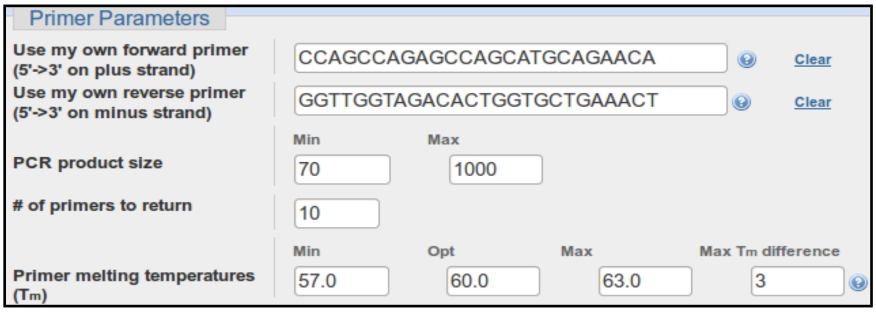
[3] Which rather makes mock of all the deep thought employed deciding upon the most sensible maximum number of predictions to be reported.

[4] This was not true until very recently. **Primer-BLAST** reported many more primer pair suggestions and quite a few unintended products for each. The previous parameter restriction the length of unintended products was somewhat more generous.

As well as suggesting primers for PCR (or other purposes) and (optionally) suggesting hybridisation oligos, **primer-BLAST** can be used to evaluate user-selected primers. Earlier, you saved a pair of primer sequences associated with **PAX6** when searching the nucleotide databases at the **NCBI**. It would be interesting to discover the product these might produce. To do this you need an unsullied **Primer-BLAST** page. Go again to:

`http://www.ncbi.nlm.nih.gov`

Click on the **BLAST** option. Select **Primer-BLAST** from the **Specialized BLAST** section. Upload your genomic **PAX6** genomic sequence using the **Browse (or Choose File)** button for the **PCR Template**.

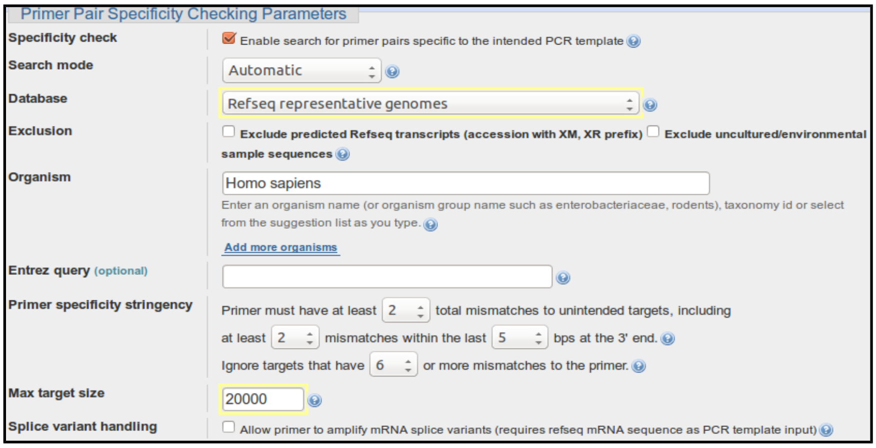


Primer Parameters

Use my own forward primer (5'->3' on plus strand)	CCAGCCAGAGCCAGCATGCAGAACA			Clear
Use my own reverse primer (5'->3' on minus strand)	GGTTGGTAGACACTGGTGCTGAAACT			Clear
	Min	Max		
PCR product size	70	1000		
# of primers to return	10			
	Min	Opt	Max	Max Tm difference
Primer melting temperatures (Tm)	57.0	60.0	63.0	3

Open up the file you made containing the primers from **GenBank** (pax6_primers.fasta) in a text editor.

Copy and **Paste** the two primer sequences into the **Use my own forward primer** and **Use my own reverse primer** boxes as appropriate.



Primer Pair Specificity Checking Parameters

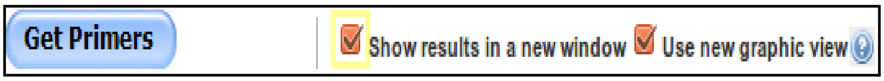
Specificity check	☑ Enable search for primer pairs specific to the intended PCR template
Search mode	Automatic
Database	Refseq representative genomes
Exclusion	☐ Exclude predicted Refseq transcripts (accession with XM, XR prefix) ☐ Exclude uncultured/environmental sample sequences
Organism	Homo sapiens Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggestion list as you type. Add more organisms
Entrez query (optional)	
Primer specificity stringency	Primer must have at least 2 total mismatches to unintended targets, including at least 2 mismatches within the last 5 bps at the 3' end. Ignore targets that have 6 or more mismatches to the primer.
Max target size	20000
Splice variant handling	☐ Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input)

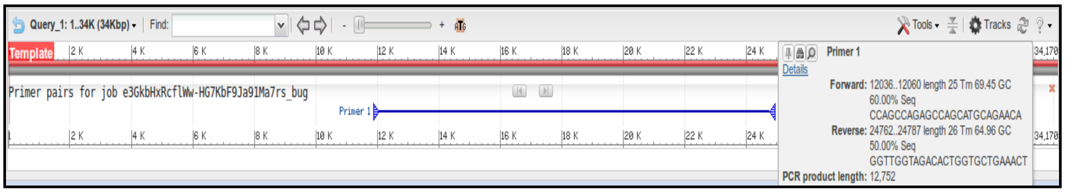
In the **Primer Pair Specificity Checking Parameters** section, set the **Database** selection to **RefSeq representative genomes**.

Leave the **Organism** as **Homo sapiens**.

Raise the **Max target size** parameter from **4000** to **20000**. You should check for enormous unintended products with this run of **Primer-BLAST**. The reasons for this will soon become apparent.

Ask **primer-BLAST** to **Show results in a new window**. Click on the **Get Primers** button.





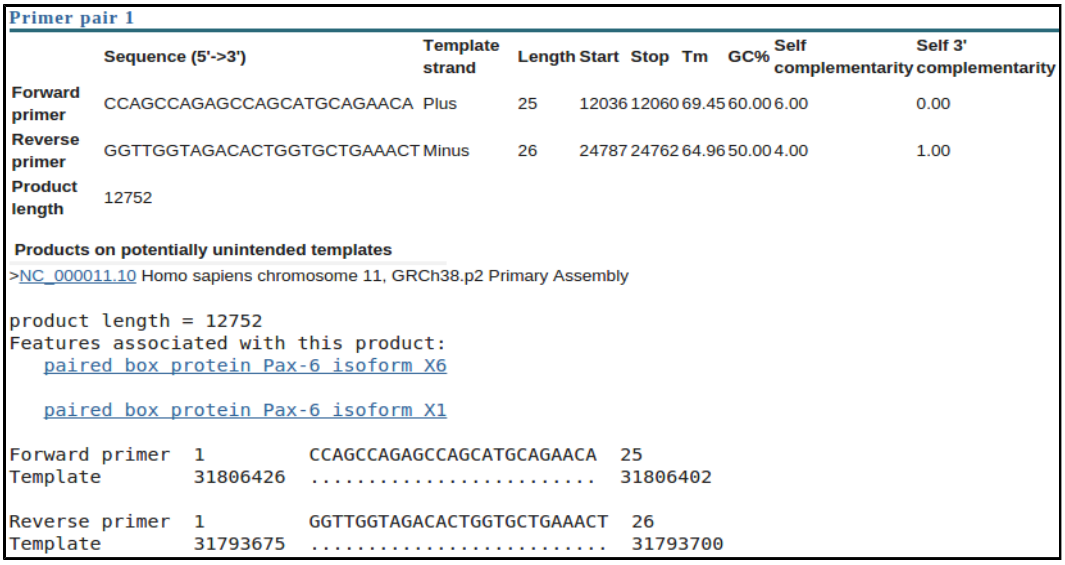
Primer pairs for job e3GkbHxRcflWw-HG7KbF9Ja91Ma7rs_bug

Primer 1

Forward: 12036..12060 length 25 Tm 69.45 GC 60.00% Seq CCAGCCAGAGCCAGCATGCAGAACA

Reverse: 24762..24787 length 26 Tm 64.96 GC 50.00% Seq GGTTGGTAGACACTGGTGCTGAAACT

PCR product length: 12,752



Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CCAGCCAGAGCCAGCATGCAGAACA	Plus	25	12036	12060	69.45	60.00	6.00	0.00
Reverse primer	GGTTGGTAGACACTGGTGCTGAAACT	Minus	26	24787	24762	64.96	50.00	4.00	1.00
Product length	12752								

Products on potentially unintended templates

>NC_000011.10 Homo sapiens chromosome 11, GRCh38.p2 Primary Assembly

```
product length = 12752
Features associated with this product:
   paired box protein Pax-6 isoform X6

   paired box protein Pax-6 isoform X1

Forward primer  1         CCAGCCAGAGCCAGCATGCAGAACA  25
Template        31806426  .........................  31806402

Reverse primer  1         GGTTGGTAGACACTGGTGCTGAAACT  26
Template        31793675  ..........................  31793700
```

After a short thrill filled pause, you will receive a result that should again looks more that a trifle like mine.

Seemingly a fine match. Even the single **potentially unintended product** reported is actually the **intended product**. For some reason, **Primer-BLAST** does not distinguish between intended products and unintended ones when investigating user specified primers[5]?

Success! However, applying a small measure of sober reflection, one has to wonder at a PCR product of **12,752** base pairs? I suspect that to be just a tad on the boastful side of probable[6]? Clearly, **primer-BLAST** is convinced, but maybe a look at the references that came with these primer sequences would be advised before accepting this result at face value.

[5] I have asked the guys at **NCBI** to explain. No full answer as yet, further prodding required. Prodded last **2016.04.02**.

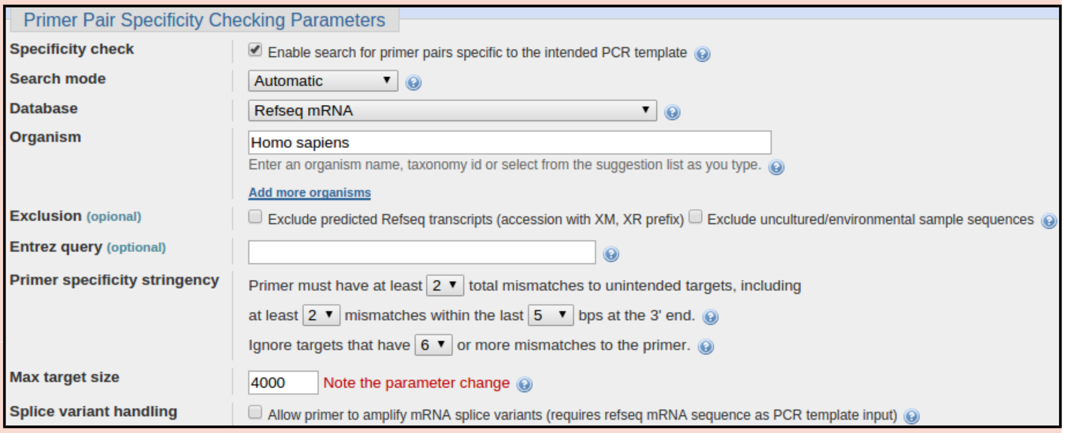
[6] Apparently, such a PCR product is possible! However, above **5,000** base pairs would be slow, require very close attention and be prone to errors.

Unfortunately, the only paper referenced does not explain what might be going on particularly clearly. However, there is a hint that the primers you saved were designed for use with mRNA/cDNA data. Therefore it might be interesting to run **primer-BLAST** one last time with pax6_cdna.fasta as the **PCR Template**.

Simply move back to your last **primer-BLAST** launch page. This time, load pax6_cdna.fasta as the **PCR Template**.

In the **Primer Pair Specificity Checking Parameters** section, set the **Database** selection set to **Refseq mRNA** and leave the organism set to **Homo sapiens**.

Set the **Max target size** back to its default value of **4000**, you should expect much smaller mRNA products this time, so no need for extending this maximum beyond **4000**.

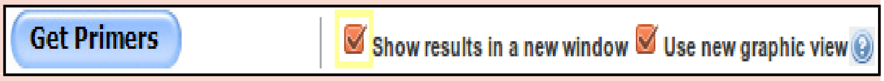


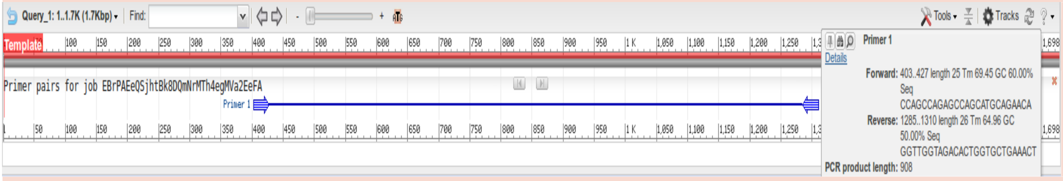
Primer Pair Specificity Checking Parameters

Specificity check	☑ Enable search for primer pairs specific to the intended PCR template
Search mode	Automatic ▾
Database	Refseq mRNA ▾
Organism	Homo sapiens Enter an organism name, taxonomy id or select from the suggestion list as you type. Add more organisms
Exclusion (opional)	□ Exclude predicted Refseq transcripts (accession with XM, XR prefix) □ Exclude uncultured/environmental sample sequences
Entrez query (optional)	
Primer specificity stringency	Primer must have at least 2 ▾ total mismatches to unintended targets, including at least 2 ▾ mismatches within the last 5 ▾ bps at the 3' end. Ignore targets that have 6 ▾ or more mismatches to the primer.
Max target size	4000 Note the parameter change
Splice variant handling	□ Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input)

These selections suppose that the design of PCR product was for selection from a library of all human cDNAs.

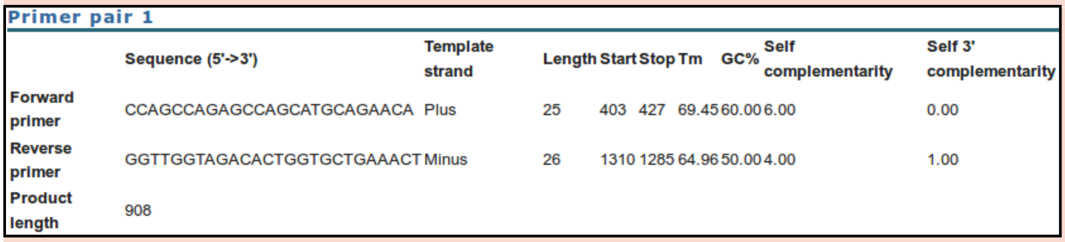
Ask **primer-BLAST** to **Show results in a new window**. Click on the **Get Primers** button.





Primer pairs for job EBrPAEeQSjhtBk8DQmNrMTh4egMVa2EeFA

Primer 1
Forward: 403..427 length 25 Tm 69.45 GC 60.00% Seq CCAGCCAGAGCCAGCATGCAGAACA
Reverse: 1285..1310 length 26 Tm 64.96 GC 50.00% Seq GGTTGGTAGACACTGGTGCTGAAACT
PCR product length: 908



Primer pair 1

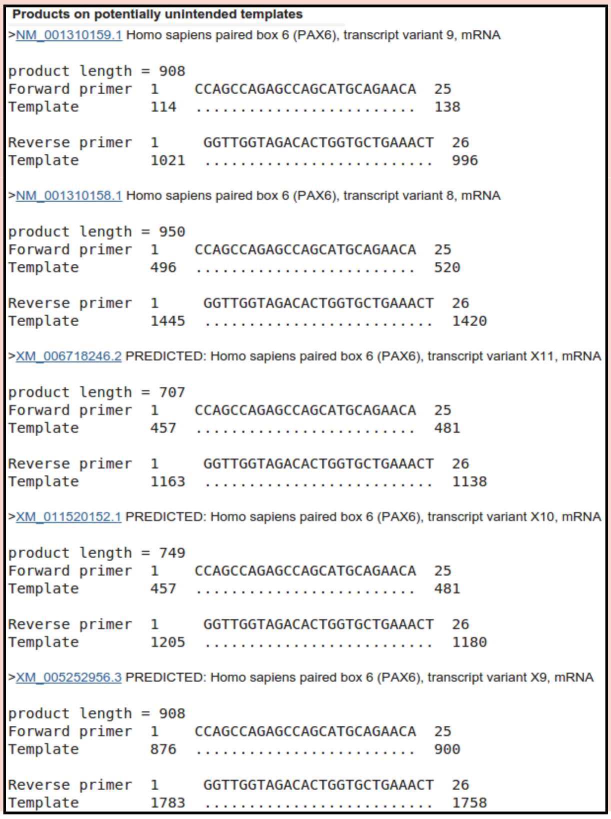
	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CCAGCCAGAGCCAGCATGCAGAACA	Plus	25	403	427	69.45	60.00	6.00	0.00
Reverse primer	GGTTGGTAGACACTGGTGCTGAAACT	Minus	26	1310	1285	64.96	50.00	4.00	1.00
Product length	908								

The result is a much more reasonable **Product length** of just **908** base pairs,reinforcing the theory that these primers were indeed designed for use with a cDNA library.

Before moving on, afford a quick glance at the report offered concerning possible unintended products. Here **primer-BLAST** warns against human mRNAs that might be cloned along with the intended target.

The first thing to note is that the intended target is not generated from a **RefSeq** mRNA. It comes from an mRNA taken from an **aniridia** patient directly. Therefore, there is no unintended product that we can ignore because it is really the intended product discovered by a different route, even though no filtering of the **RefSeq** database was undertaken.

All the unintended products could/would potentially be generated by the primers under investigation and have the potential to cause confusion. If you look down the list, you should conclude that the **16** unintended products come from **16** of the **24 RefSeq PAX6** transcripts first noted by **GeneCards** and then confirmed later by **blast**.



Products on potentially unintended templates

>NM_001310159.1 Homo sapiens paired box 6 (PAX6), transcript variant 9, mRNA

```
product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        114   .........................  138

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1021  ..........................  996
```

>NM_001310158.1 Homo sapiens paired box 6 (PAX6), transcript variant 8, mRNA

```
product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        496   .........................  520

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1445  ..........................  1420
```

>XM_006718246.2 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X11, mRNA

```
product length = 707
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        457   .........................  481

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1163  ..........................  1138
```

>XM_011520152.1 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X10, mRNA

```
product length = 749
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        457   .........................  481

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1205  ..........................  1180
```

>XM_005252956.3 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X9, mRNA

```
product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        876   .........................  900

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1783  ..........................  1758
```

9 of the **11 NM_** good quality transcripts are detected. **7** of the **13** poorer quality **XM_** **"PREDICTED"** transcripts are also present. So **16** of the **24 PAX6** transcript sequences in **RefSeq** were detected.

Why do you suppose **blast** did not pick up all the transcripts? ______________________________

____

Note that the intended product is **908** base pairs long. Note that all the unintended products except two, near the top of the list are either **908** long or **950** long. A difference of **42**.

How would you tell quickly which isoform was represented by each mRNA listed here? ______________________________

____

Some fairly redundant questions to finish this section. I think I have already answered them all. But maybe you might wish to differ?

Is the number of "**potentially unintended products**" as you would you expect, given the evidence from **GeneCards**, **Ensembl** and **blast**?______________________________

____

For all the "**potentially unintended products**", the selected primers match exactly. Can you explain this? ______________________________

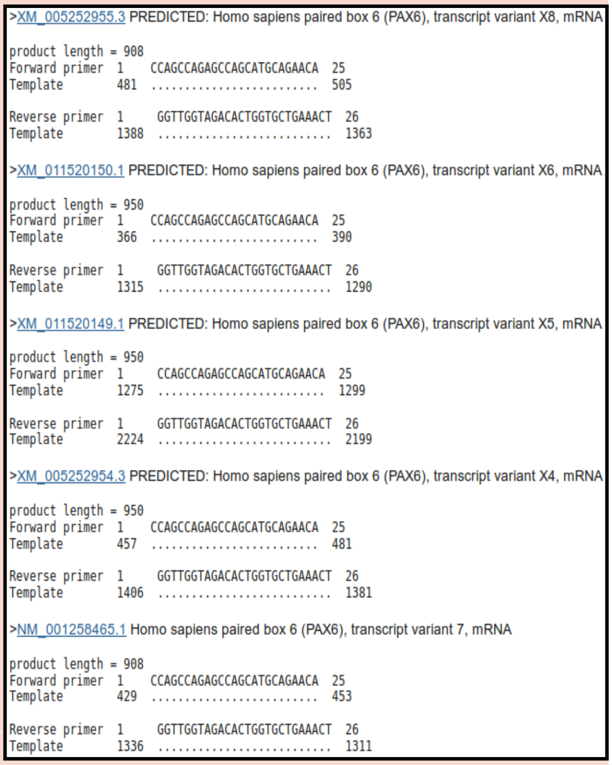
____

The "**potentially unintended products**" are of different sizes. Can you explain the difference between the possible product lengths? ___

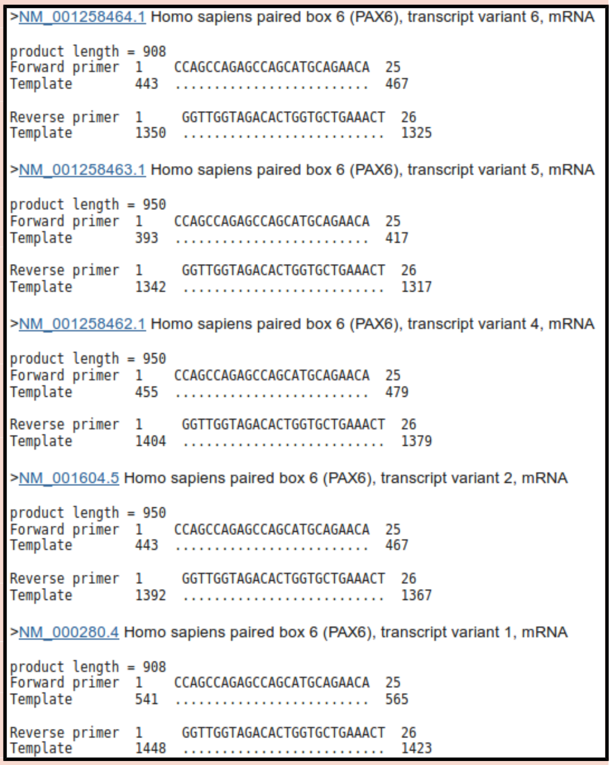
____

Are the numbers of "**potentially unintended products**" of each possible length consistent with your **blast** results? ________________

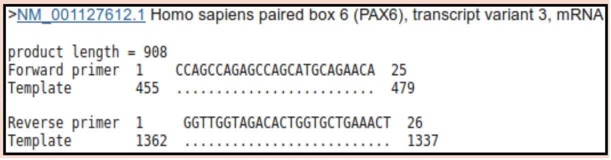
____



```
>XM_005252955.3 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X8, mRNA

product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        481   .........................  505

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1388  ..........................  1363

>XM_011520150.1 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X6, mRNA

product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        366   .........................  390

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1315  ..........................  1290

>XM_011520149.1 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X5, mRNA

product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        1275  .........................  1299

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        2224  ..........................  2199

>XM_005252954.3 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X4, mRNA

product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        457   .........................  481

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1406  ..........................  1381

>NM_001258465.1 Homo sapiens paired box 6 (PAX6), transcript variant 7, mRNA

product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        429   .........................  453

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1336  ..........................  1311
```



```
>NM_001258464.1 Homo sapiens paired box 6 (PAX6), transcript variant 6, mRNA

product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        443   .........................  467

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1350  ..........................  1325

>NM_001258463.1 Homo sapiens paired box 6 (PAX6), transcript variant 5, mRNA

product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        393   .........................  417

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1342  ..........................  1317

>NM_001258462.1 Homo sapiens paired box 6 (PAX6), transcript variant 4, mRNA

product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        455   .........................  479

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1404  ..........................  1379

>NM_001604.5 Homo sapiens paired box 6 (PAX6), transcript variant 2, mRNA

product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        443   .........................  467

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1392  ..........................  1367

>NM_000280.4 Homo sapiens paired box 6 (PAX6), transcript variant 1, mRNA

product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        541   .........................  565

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1448  ..........................  1423
```



```
>NM_001127612.1 Homo sapiens paired box 6 (PAX6), transcript variant 3, mRNA

product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        455   .........................  479

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1362  ..........................  1337
```

DPJ – 2016.12.05

Model Answers to Questions in the Instructions Text.

Notes:

For the most part, these "**Model Answers**" just provide the reactions/solutions I hoped you would work out for yourselves. However, sometime I have tried to offer a bit more background and material for thought? Occasionally, I have rambled off into some rather self indulgent investigations that even I would not want to try and justify as pertinent to the objective of these exercises. I like to keep these meanders, as they help and entertain me, but I wish to warn you to only take regard of them if you are feeling particularly strong and have time to burn. Certainly not a good idea to indulge here during a time constrained course event!

Where things have got extreme, I am going to make two versions of the answer. One starting:

Summary:

Which has the answer with only a reasonably digestible volume of deep thought. Read this one.

The other will start:

Full Answer:

Beware of entering here! I do not hold back. Nothing complicated, but it will be long and full of pedantry.

This makes the Model answers section very big. **BUT**, it is not intended for printing or for reading serially, so I submit, being long and wordy does not matter. Feel free to disagree.

From your investigations of **Primer Design**

Do you think **10** primer pair suggestions is sufficient? If not, what number would you choose?

Until very recently, the default here was **5**. That seemed rather low to me. I included this question to solicit opinion rather than to impart knowledge. A default of **10** seems more in line with my instincts, but people who use this program seriously mostly tell me that they can select suitable primers from the first **2** or **3** suggestions of the program. So, **5** would seem a good choice and **10** would be moving towards cautiously overdoing things.

On the whole, informed opinion suggests that **10** suggestions will be more than enough in most circumstances.

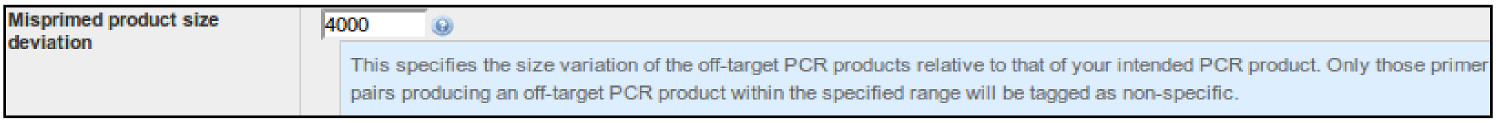
What value would you choose here if you were looking for uncluttered results?

Summary:

Clearly, the smaller the number chosen, the shorter will be the list of spurious products. However, pick something too small and you risk including unintended product(s) that could cause confusion. The size selected must be sufficient that larger unwanted PCR product(s) could easily be spotted by other means (simply by size?).

Full Answer:

Well, mostly for me, and just in case you were curious, when I first wrote the question, the parameter was very different and not so easy to understand. Pure self indulgence, I know, but here is the history. The parameter explained itself, via the button, thus:



Misprimed product size deviation 4000

This specifies the size variation of the off-target PCR products relative to that of your intended PCR product. Only those primer pairs producing an off-target PCR product within the specified range will be tagged as non-specific.

I interpreted this to mean that only **blast** predicted products of up to **X**+**4,000** base pairs, where **X** base pairs is the length of the intended target, will be given any regard. It is thus assumed that a difference of **4,000** base pairs between an intended PCR product (predicted by **primer3**) and a spurious product (detected by **blast**) can easily be detected simply by size difference.

Of course this parameter also will reject unwanted **blast** predicted products that are less than **X–4,000** base pairs will be given any regard. Given the largest possible **primer3** suggestion will be **1,000** base pairs (the form setting for the exercise specifies products of between **100**[7] and **1,000** base pairs), this is hardly an issue here.

Comment upon the small default value for the **Blast word size**?

By default, **blast** will be looking for aligned exactly matching blocks of **7** nucleotides when identifying where a primer might match a database entry. The entire primer match with the template sequence does not have to be exact for the primer to be acceptable. The entire primer is typically only around **20** bases long. And word size much more that **7** would clearly miss too much to be effective.

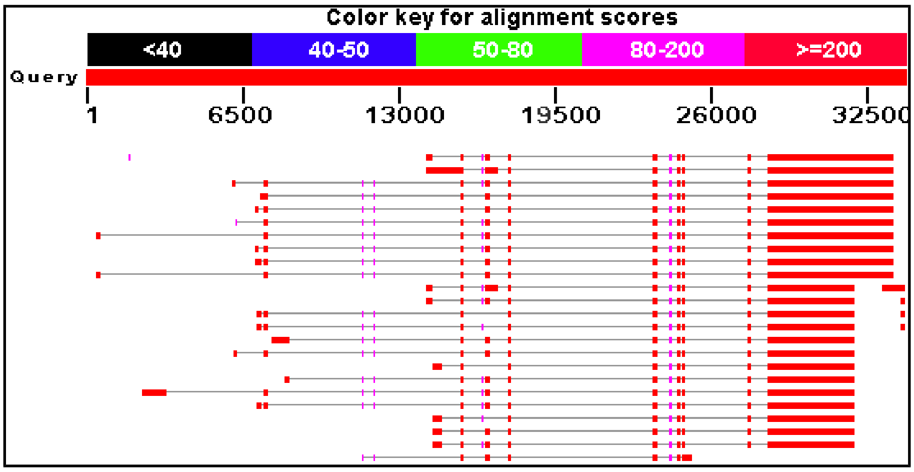
[7] The form explicitly declares a minimum of **70**, but the ranges from which the **forward** & **reverse** primers must come (**15000-15700** & **15800-16500**) make the smallest possible **primer3** prediction **100** base pairs long.

Why do you suppose **blast** did not pick up all the transcripts?

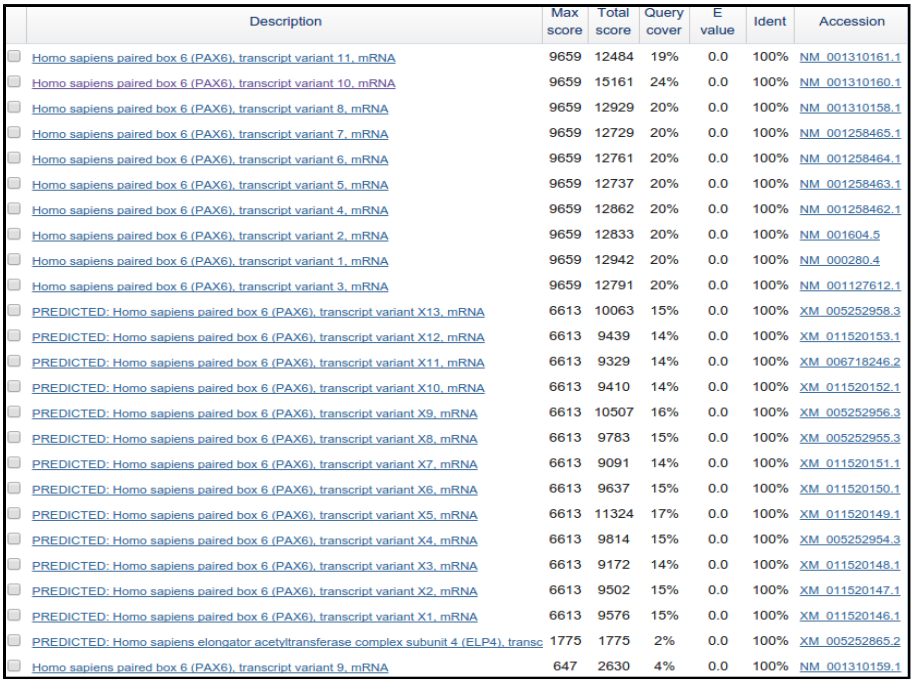
Summary:

Well, the simple answer is that the transcripts that were not detected as unwanted products cannot include either the forward primer, or the reverse primer, or both. This is, almost, the only possible explanation.

Full Answer:



Of course, for this run, you did specify that you were not interested in products longer than **4,000** base pairs, so it could be that one or more products were possible but longer than that? I suspect this would only be feasible if there were retained introns involved, but previous **blast** results do not suggest this to be the case. I would say the only possible candidate for an over-length product might be the second hit down in the graphical representation generated previously by **blast**. The first and third exons from the left look a bit bloated, but not really sufficiently to cause a problem.



	Description	Max score	Total score	Query cover	E value	Ident	Accession
□	Homo sapiens paired box 6 (PAX6), transcript variant 11, mRNA	9659	12484	19%	0.0	100%	NM_001310161.1
□	Homo sapiens paired box 6 (PAX6), transcript variant 10, mRNA	9659	15161	24%	0.0	100%	NM_001310160.1
□	Homo sapiens paired box 6 (PAX6), transcript variant 8, mRNA	9659	12929	20%	0.0	100%	NM_001310158.1
□	Homo sapiens paired box 6 (PAX6), transcript variant 7, mRNA	9659	12729	20%	0.0	100%	NM_001258465.1
□	Homo sapiens paired box 6 (PAX6), transcript variant 6, mRNA	9659	12761	20%	0.0	100%	NM_001258464.1
□	Homo sapiens paired box 6 (PAX6), transcript variant 5, mRNA	9659	12737	20%	0.0	100%	NM_001258463.1
□	Homo sapiens paired box 6 (PAX6), transcript variant 4, mRNA	9659	12862	20%	0.0	100%	NM_001258462.1
□	Homo sapiens paired box 6 (PAX6), transcript variant 2, mRNA	9659	12833	20%	0.0	100%	NM_001604.5
□	Homo sapiens paired box 6 (PAX6), transcript variant 1, mRNA	9659	12942	20%	0.0	100%	NM_000280.4
□	Homo sapiens paired box 6 (PAX6), transcript variant 3, mRNA	9659	12791	20%	0.0	100%	NM_001127612.1
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X13, mRNA	6613	10063	15%	0.0	100%	XM_005252958.3
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X12, mRNA	6613	9439	14%	0.0	100%	XM_011520153.1
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X11, mRNA	6613	9329	14%	0.0	100%	XM_006718246.2
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X10, mRNA	6613	9410	14%	0.0	100%	XM_011520152.1
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X9, mRNA	6613	10507	16%	0.0	100%	XM_005252956.3
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X8, mRNA	6613	9783	15%	0.0	100%	XM_005252955.3
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X7, mRNA	6613	9091	14%	0.0	100%	XM_011520151.1
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X6, mRNA	6613	9637	15%	0.0	100%	XM_011520150.1
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X5, mRNA	6613	11324	17%	0.0	100%	XM_011520149.1
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X4, mRNA	6613	9814	15%	0.0	100%	XM_005252954.3
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X3, mRNA	6613	9172	14%	0.0	100%	XM_011520148.1
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X2, mRNA	6613	9502	15%	0.0	100%	XM_011520147.1
□	PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X1, mRNA	6613	9576	15%	0.0	100%	XM_011520146.1
□	PREDICTED: Homo sapiens elongator acetyltransferase complex subunit 4 (ELP4), transc	1775	1775	2%	0.0	100%	XM_005252865.2
□	Homo sapiens paired box 6 (PAX6), transcript variant 9, mRNA	647	2630	4%	0.0	100%	NM_001310159.1

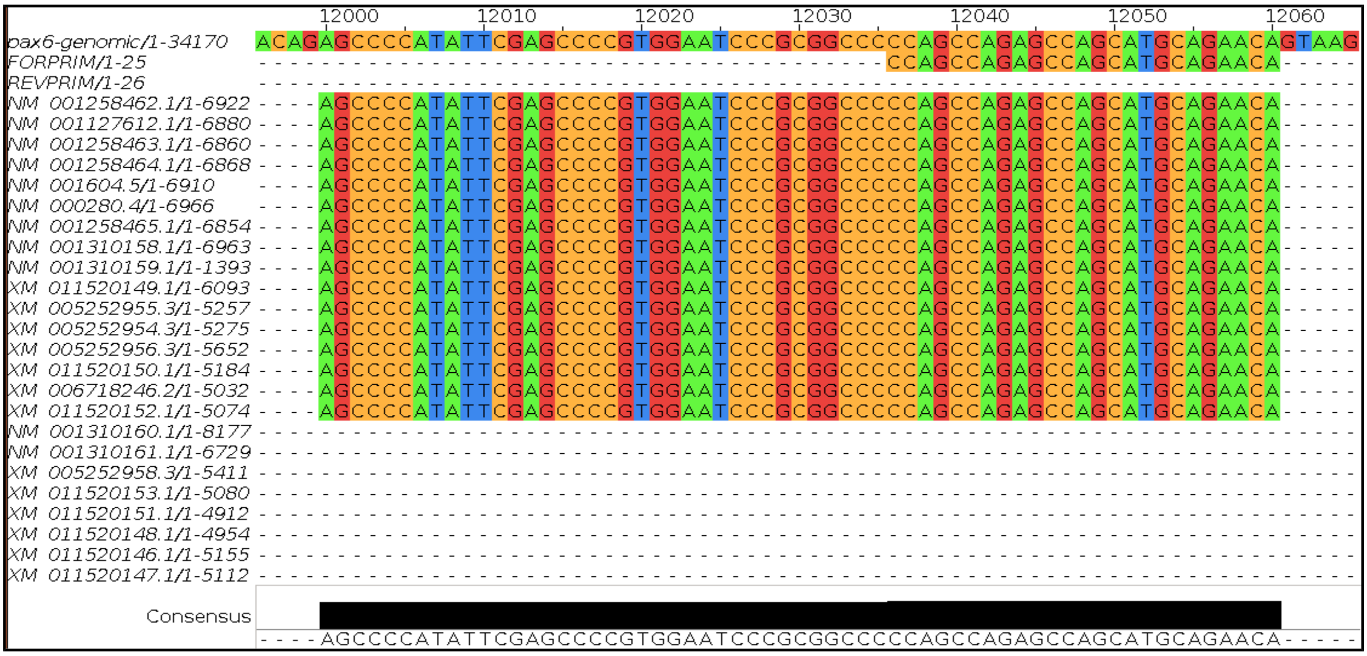
It might also be that unwanted PCR products are eliminated/introduced due to variations in the predicted transcripts. However, this can be ruled out as previous experiments, **blast** assures us that all **24** potential transcripts match the genomic sequence exactly.

Enough! Only because I want to, I will compute the alignments to prove the missing primer matches. Read no further unless you are truly in the mood. Much of the reason for recording the rest of this answer is that, apart from enjoying the pursuit of irrelevant detail, I also wanted to remember how I made the alignments and certainly feel I could have made both these, and my point much more simply? Suggestions welcome.

OK, I started by computing an alignment that was a mapping of all **24** transcripts onto the **PAX6** genomic regions as represented in the file **pax6_genomic.fasta**. I used a program called **gmap**, which like **spline** (used in the exercise) is designed to align cDNA/mRNA sequences with corresponding genomic sequences. The version of **gmap** I used runs under **linux** from the command line. It has the advantage over **spline** that is will align more than one cDNA/mRNA sequence against the genome in one run. Unfortunately, it does not generate an output format that can be easily displayed in the way I required here. I did try to persuade a couple of general multiple alignment programs (**clustalw** & **muscle**) to make me a usable alignment, but ran into the same difficulties we experienced in the exercise. I failed to find gap penalties that would get the programs to gap the larger introns. Even if I had succeeded to get the gaps in the right place, I would not have believed them to be placed with sufficient accuracy for the same reasons this was not possible when we tried the same trick with general alignment software for just one cDNA sequence against the genome in the exercise.

So, I made a rough alignment with **clustalw** and edited it to exactly what was suggested by **gmap** using **jalview**. This took **HOURS**. There has to be a better way!! You have already used all the software mentioned except **gmap** and **clustalw**. You will use **clustalw** and see how **jalview** can be use to edit, as well as just view, alignments a little later.

All that effort to show that the region around the forward primer looks like this:



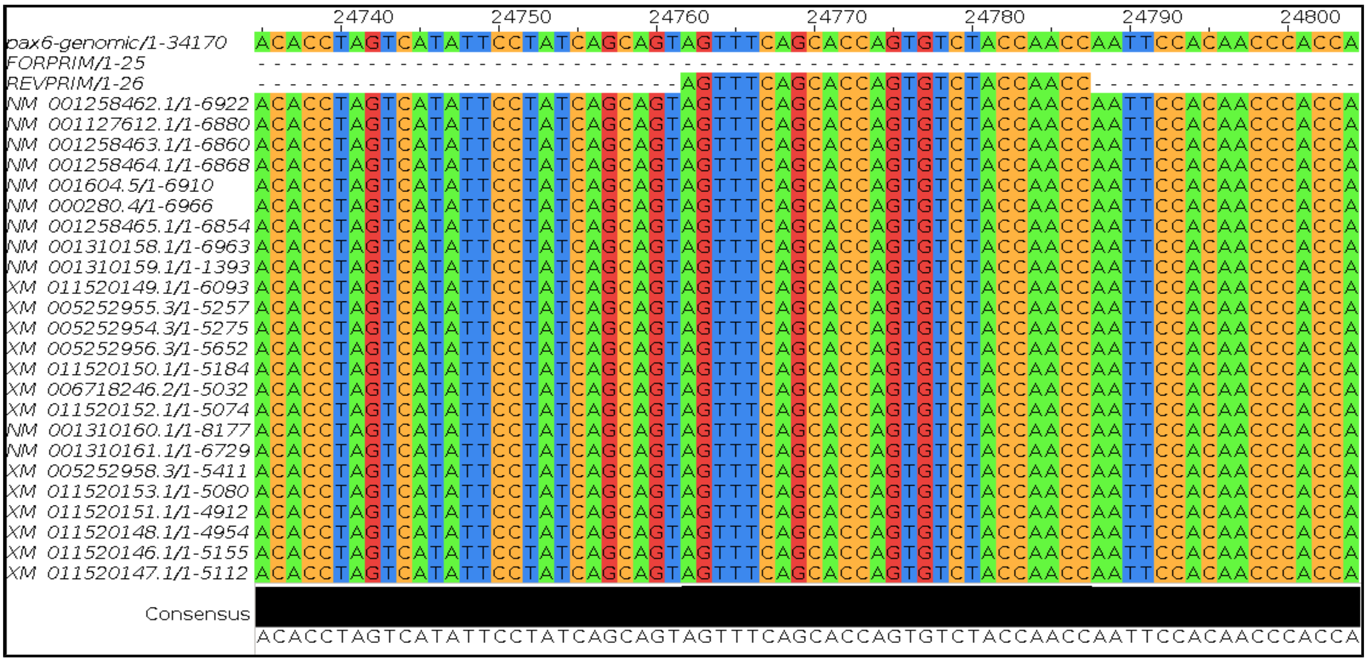
```
                          12000     12010     12020     12030     12040     12050     12060
pax6-genomic/1-34170     ACAGAGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACAGTAAG
FORPRIM/1-25             ----------------------------------------CCAGCCAGAGCCAGCATGCAGAACA-----
REVPRIM/1-26             ----------------------------------------------------------------------
NM 001258462.1/1-6922    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
NM 001127612.1/1-6880    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
NM 001258463.1/1-6860    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
NM 001258464.1/1-6868    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
NM 001604.5/1-6910       ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
NM 000280.4/1-6966       ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
NM 001258465.1/1-6854    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
NM 001310158.1/1-6963    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
NM 001310159.1/1-1393    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
XM 011520149.1/1-6093    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
XM 005252955.3/1-5257    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
XM 005252954.3/1-5275    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
XM 005252956.3/1-5652    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
XM 011520150.1/1-5184    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
XM 006718246.2/1-5032    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
XM 011520152.1/1-5074    ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
NM 001310160.1/1-8177    ----------------------------------------------------------------------
NM 001310161.1/1-6729    ----------------------------------------------------------------------
XM 005252958.3/1-5411    ----------------------------------------------------------------------
XM 011520153.1/1-5080    ----------------------------------------------------------------------
XM 011520151.1/1-4912    ----------------------------------------------------------------------
XM 011520148.1/1-4954    ----------------------------------------------------------------------
XM 011520146.1/1-5155    ----------------------------------------------------------------------
XM 011520147.1/1-5112    ----------------------------------------------------------------------

            Consensus
                         ----AGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA-----
```

Showing clearly that the **8** transcripts:

NM_001310160.1
NM_001310161.1
XM_005252958.3
XM_011520153.1
XM_011520151.1
XM_011520148.1
XM_011520146.1
XM_011520147.1

Have the exon that includes the forward primer spliced out and so will not produce any PCR product. Feel free to check this by comparing the textual results of your **blast** of the genomic sequence against the **RefSeq** mRNAs and the results of **PRIMER-BLAST**. I did, it was lots and lots of fun and I ended up content that all was logically consistent.

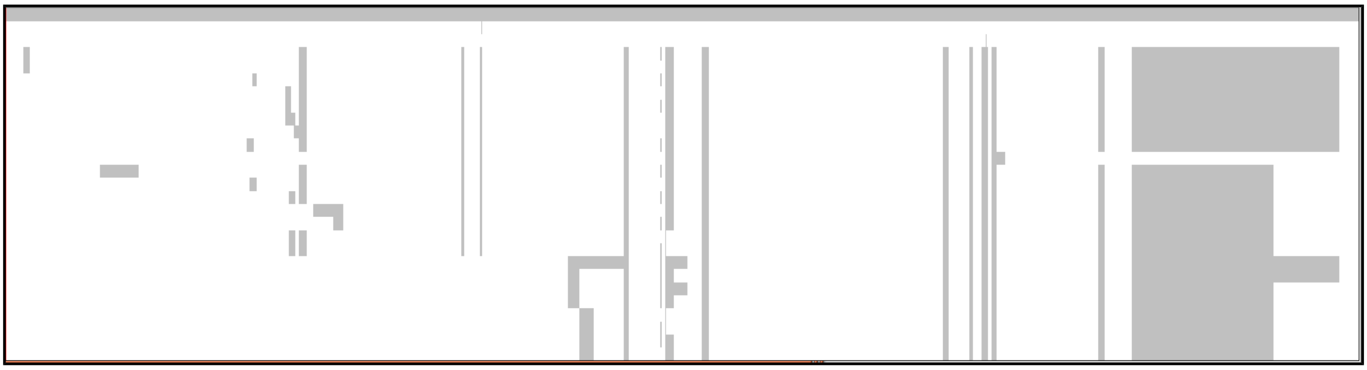
The alignment around the reverse primer looks like this:



```
                            24740     24750     24760     24770     24780     24790     24800
pax6-genomic/1-34170     ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
FORPRIM/1-25             ----------------------------------------------------------------------
REVPRIM/1-26             ---------------------------AGTTTCAGCACCAGTGTCTACCAACC-----------------
NM 001258462.1/1-6922    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 001127612.1/1-6880    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 001258463.1/1-6860    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 001258464.1/1-6868    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 001604.5/1-6910       ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 000280.4/1-6966       ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 001258465.1/1-6854    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 001310158.1/1-6963    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 001310159.1/1-1393    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 011520149.1/1-6093    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 005252955.3/1-5257    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 005252954.3/1-5275    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 005252956.3/1-5652    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 011520150.1/1-5184    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 006718246.2/1-5032    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 011520152.1/1-5074    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 001310160.1/1-8177    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
NM 001310161.1/1-6729    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 005252958.3/1-5411    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 011520153.1/1-5080    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 011520151.1/1-4912    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 011520148.1/1-4954    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 011520146.1/1-5155    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
XM 011520147.1/1-5112    ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA

            Consensus
                         ACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCA
```

All **24** putative transcripts match the reverse primer perfectly. So blast should indeed find **16** of the **24** transcripts sequences in **RefSeq**, and it does.

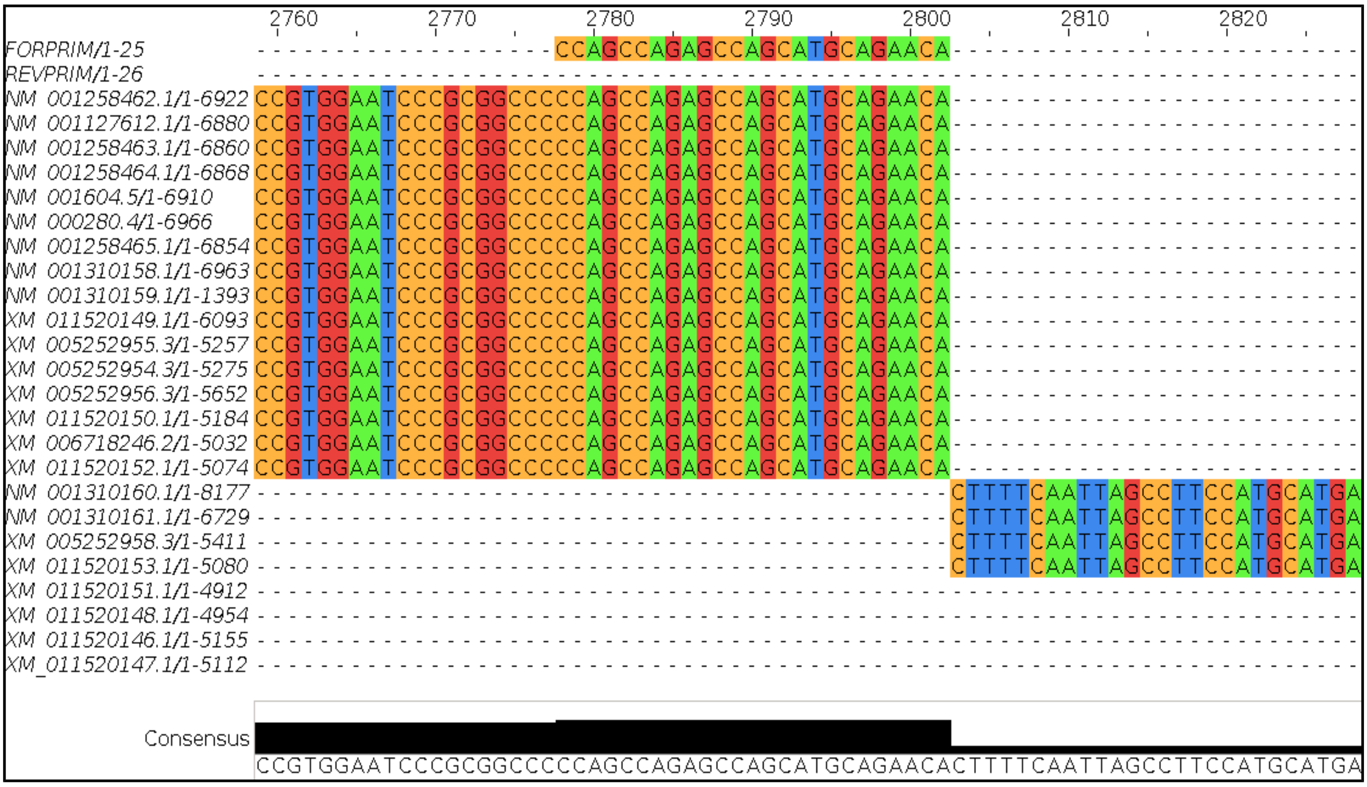
Jalview offers an overview of the entire alignment. The top row shows the genomic sequence. The second row shows the position of the forward primer. The third row shows the position of the reverse primer.



Except for the order of the transcripts, this view is very similar to the overview graphic generated by **blast**. The transcripts missing the forward primer, which isoform each transcripts represents and the fact that all transcripts match the reverse primer should be very clear.

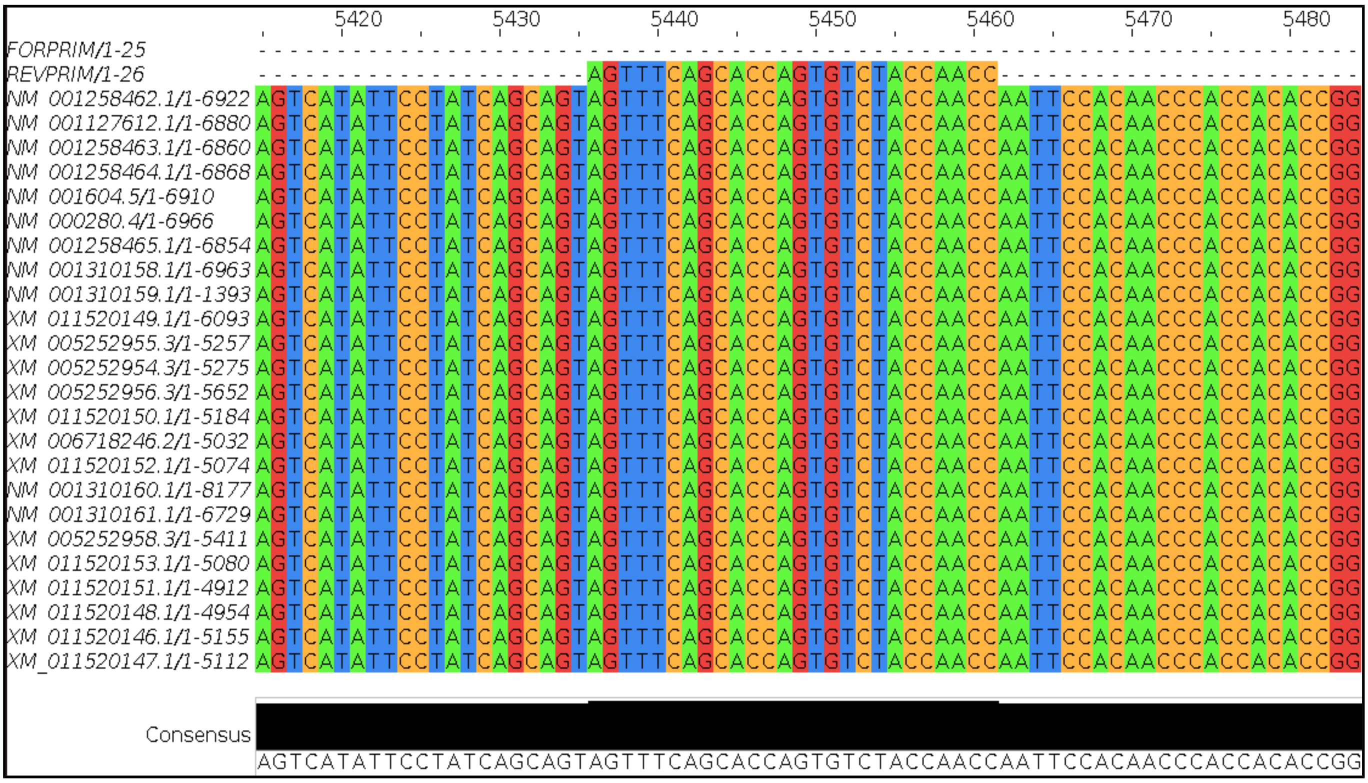
Finished Dave? Well no, not quite. I wondered why I had included the genomic sequence in my alignment. Finding no answer to that question, I tried to make an alignment of just the primer sequences and the mRNAs. I thought this would be easy. I was wrong. The general programs are still going to get the gaps wrong whatever penalties are used. Some transcripts have exons entirely missing in all other transcripts leaving no clues as to which way round they should be aligned. The scaffold provided by the genomic sequence was essential. So, I made an mRNA only alignment by editing the alignment discussed above with **jalview**. This was easy (although you would not think so given the time it took me to work out how to do it!). I loaded the alignment into **jalview**, deleted the genomic sequence and then removed all empty columns (that is, all columns with no bases in them due the the removal of the genomic sequence). Clever eh? Just because it is there, here are the pictures.

Forward primer region (the primer is right at the end of an exon):



```
                      2760      2770      2780      2790      2800      2810      2820
FORPRIM/1-25          -------------------CCAGCCAGAGCCAGCATGCAGAACA--------------------------
REVPRIM/1-26          ----------------------------------------------------------------------
NM 001258462.1/1-6922 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
NM 001127612.1/1-6880 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
NM 001258463.1/1-6860 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
NM 001258464.1/1-6868 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
NM 001604.5/1-6910    CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
NM 000280.4/1-6966    CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
NM 001258465.1/1-6854 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
NM 001310158.1/1-6963 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
NM 001310159.1/1-1393 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
XM 011520149.1/1-6093 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
XM 005252955.3/1-5257 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
XM 005252954.3/1-5275 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
XM 005252956.3/1-5652 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
XM 011520150.1/1-5184 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
XM 006718246.2/1-5032 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
XM 011520152.1/1-5074 CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACA--------------------------
NM 001310160.1/1-8177 --------------------------------------------CTTTTCAATTAGCCTTCCATGCATGA
NM 001310161.1/1-6729 --------------------------------------------CTTTTCAATTAGCCTTCCATGCATGA
XM 005252958.3/1-5411 --------------------------------------------CTTTTCAATTAGCCTTCCATGCATGA
XM 011520153.1/1-5080 --------------------------------------------CTTTTCAATTAGCCTTCCATGCATGA
XM 011520151.1/1-4912 ----------------------------------------------------------------------
XM 011520148.1/1-4954 ----------------------------------------------------------------------
XM 011520146.1/1-5155 ----------------------------------------------------------------------
XM_011520147.1/1-5112 ----------------------------------------------------------------------

            Consensus
                      CCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAACACTTTTCAATTAGCCTTCCATGCATGA
```

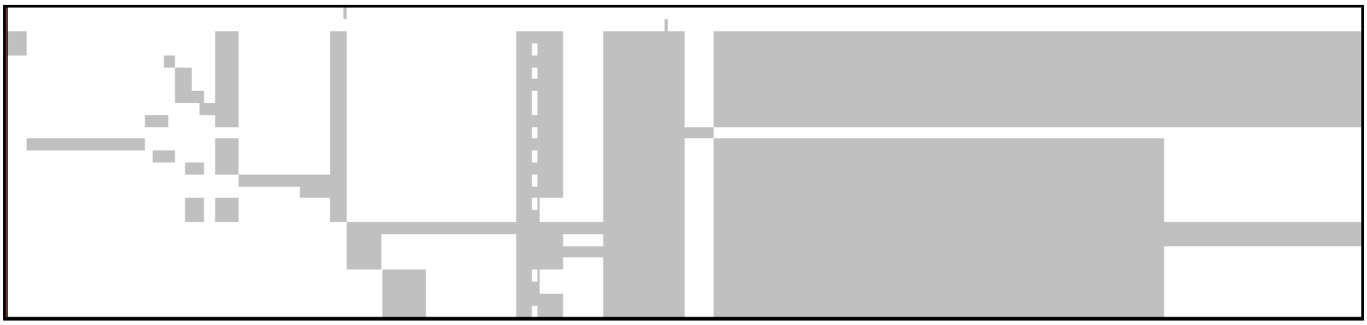
Reverse primer region:



```
                               5420      5430      5440      5450      5460      5470      5480
FORPRIM/1-25          ----------------------------------------------------------------------
REVPRIM/1-26          ---------------------AGTTTCAGCACCAGTGTCTACCAACC-----------------------
NM 001258462.1/1-6922 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 001127612.1/1-6880 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 001258463.1/1-6860 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 001258464.1/1-6868 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 001604.5/1-6910    AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 000280.4/1-6966    AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 001258465.1/1-6854 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 001310158.1/1-6963 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 001310159.1/1-1393 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 011520149.1/1-6093 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 005252955.3/1-5257 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 005252954.3/1-5275 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 005252956.3/1-5652 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 011520150.1/1-5184 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 006718246.2/1-5032 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 011520152.1/1-5074 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 001310160.1/1-8177 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
NM 001310161.1/1-6729 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 005252958.3/1-5411 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 011520153.1/1-5080 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 011520151.1/1-4912 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 011520148.1/1-4954 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM 011520146.1/1-5155 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
XM_011520147.1/1-5112 AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG

            Consensus
                      AGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTCCACAACCCACCACACCGG
```

Overview:

Without the evidence of the genomic sequence, the two leftmost exons could logically swap position. There is no transcript that includes both these exons and no overlap between either and any other exon in any transcript (most clearly verified from the previous **Overview** plot). Thus, there is no exon evidence of the order in which the two should appear.



Now I am done! This has to be the most over the top answer yet, but at least it kept me out of trouble for a while.

How would you tell quickly which isoform was represented by each mRNA listed here?

Summary:

All the mRNAs reported were of length **908**, **950**, **707** or **749**.

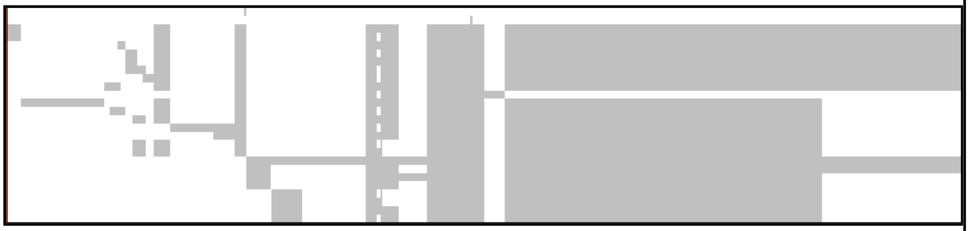
A reasonable guess might be based on the length of the products? All those that are **908** bases might be assume to produce the **422** amino acid **canonical isoform**. All those that are **950** (i.e. **42** base pairs longer) might be assumed to **436** produce amino acid **isoform 5a** proteins (i.e. **14** amino acids longer).

Analogous reasoning might be applied to the mRNAs that are either **707** or **749** base pairs in length.

Just a guess of course, but one I would be happy to have faith in. To be certain, one would need to read the annotations of each listed **RefSeq** entry!

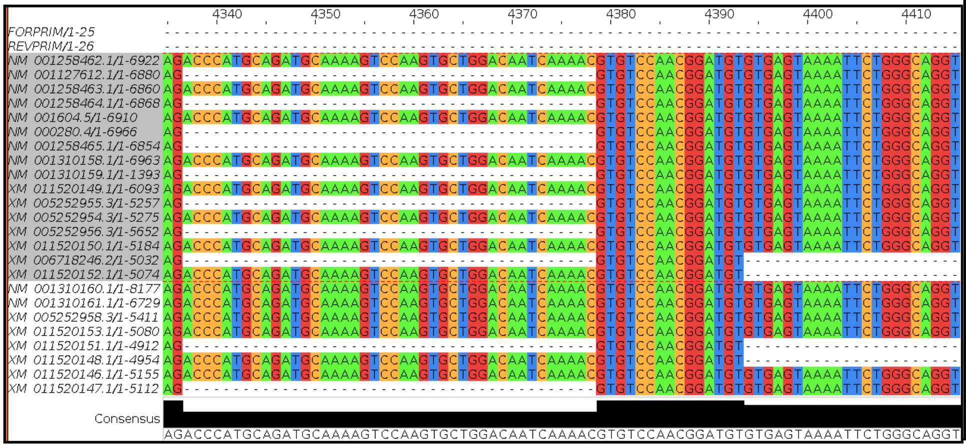
Full Answer:

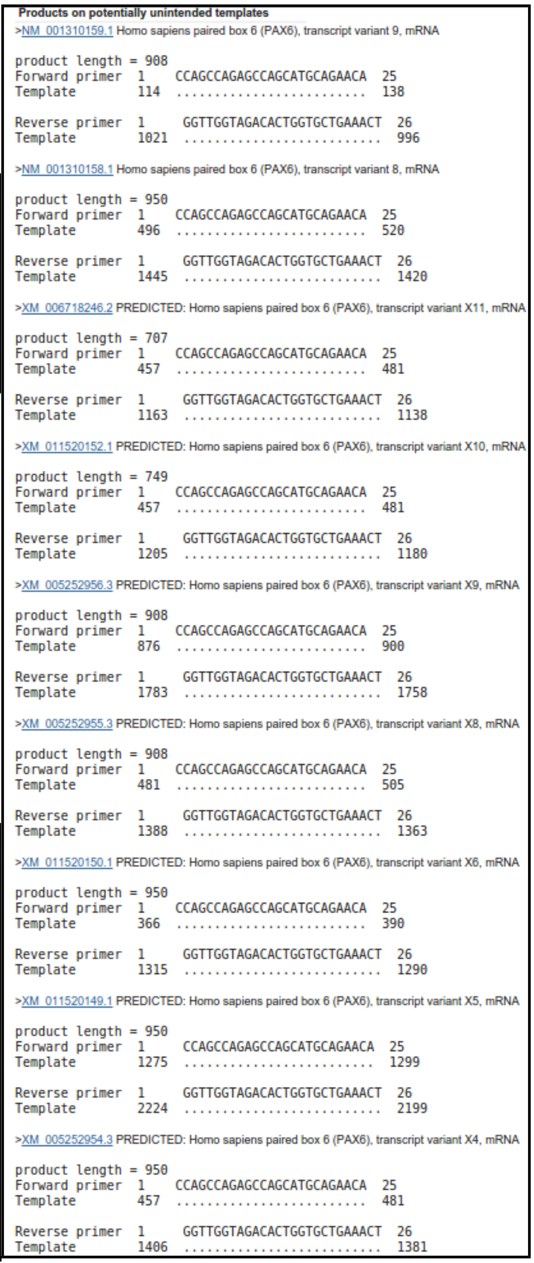
From the illustrations of the last "**Full Answer**" (in particular, the **jalview** overviews), it is clear that all the mRNAs that produce a product include the region that determines which isoform is represented. That is, all are one isoform or the other.



The last two of the mRNAs that produce a PCR product, have a bit chewed out just after the isoform defining region (an exon spliced out, if you prefer). It is logical to suppose these would be the mRNAs from which the two shorter products were generated.

Indeed, looking at the relevant part of the mRNA only alignment shows them to be **XM_006718246** (product length **707**, excluding the **isoform 5a** exon that suggest it codes for a **canonical** protein) and **XM_011520152** (product length **749**, including the extra **42** base pairs suggesting it codes for an **isoform 5a** protein).





```
Products on potentially unintended templates
>NM_001310159.1 Homo sapiens paired box 6 (PAX6), transcript variant 9, mRNA

product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        114   .........................  138

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1021  ..........................  996

>NM_001310158.1 Homo sapiens paired box 6 (PAX6), transcript variant 8, mRNA

product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        496   .........................  520

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1445  ..........................  1420

>XM_006718246.2 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X11, mRNA

product length = 707
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        457   .........................  481

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1163  ..........................  1138

>XM_011520152.1 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X10, mRNA

product length = 749
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        457   .........................  481

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1205  ..........................  1180

>XM_005252956.3 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X9, mRNA

product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        876   .........................  900

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1783  ..........................  1758

>XM_005252955.3 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X8, mRNA

product length = 908
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        481   .........................  505

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1388  ..........................  1363

>XM_011520150.1 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X6, mRNA

product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        366   .........................  390

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1315  ..........................  1290

>XM_011520149.1 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X5, mRNA

product length = 950
Forward primer  1      CCAGCCAGAGCCAGCATGCAGAACA  25
Template        1275   .........................  1299

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        2224  ..........................  2199

>XM_005252954.3 PREDICTED: Homo sapiens paired box 6 (PAX6), transcript variant X4, mRNA

product length = 950
Forward primer  1     CCAGCCAGAGCCAGCATGCAGAACA  25
Template        457   .........................  481

Reverse primer  1     GGTTGGTAGACACTGGTGCTGAAACT  26
Template        1406  ..........................  1381
```

All other transcripts that generate PCR products generate products of length either **908** or **950**. Given the difference (**42** base pairs) is exactly the size of the **isoform 5a** exon, it is reasonable to assume the transcripts generating PCR products of length **908** represent **canonical** proteins, whereas the transcripts generating PCR products of length **950** represent **isoform 5a** proteins.

Prettier? True, and I submit including sufficient evidence to be more that just a guess now.

Is the number of "**potentially unintended products**" as you would you expect, given the evidence from **GeneCards**, **Ensembl** and **blast**?

Yes, I think so, given you accept my investigation (see above) as to why there were only **16** "**potentially unintended products**" when you might have expected **24**, given your **blast** results. **GeneCards** now encourages an initial expectation of **24** "**potentially unintended products**". **Ensembl** only uses the higher quality **RefSeq** mRNAs. Currently, **Ensembl** uses **10** of the **11** good quality **RefSeq** mRNAs to make its transcripts predictions. Close enough?

For all the "**potentially unintended products**", the selected primers match exactly. Can you explain this?

Well, of course they do??? All the transcripts found are generated from the same region of genomic DNA and therefore will be identical in all shared regions, including the primer regions. I suppose, in other instances, it would be possible to have transcripts with variation in the regions matching the primers insufficient to stop the primers working? But not in this case.

One might conclude there are no genuinely "unintended" products? All are real **PAX6** transcripts of varying certainty. A genuine unintended product would come from an entirely different part of the genome and would not necessarily match exactly with respect to the primers. They would just need to be "good enough to work".

The "**potentially unintended products**" are of different sizes. Can you explain the difference between the possible product lengths?

Are the numbers of "**potentially unintended products**" of each possible length consistent with your **blast** results?

Yes yes yes! I think both these questions made a bit more sense a few generations of these notes ago. We have already answered them sufficiently I suggest. I refer you to the answers above.

DPJ – 2016.12.05