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

# ELB19F

## Entry Level Bioinformatics

04-08 February 2019

(First 2019 run of this Course)

## Basic Bioinformatics Sessions

## Practical 4: Primer Design

Wednesday 30 January 2019

# Primer Design

The prime intention of this exercise is to design a way to amplify a DNA fragment of reasonable size that includes a specific portion of the **PAX6** gene. The target region is that which includes the mutation you glanced at earlier, that is a major cause of Aniridia. That is, the substitution that mutates the **33rd** amino acid position of the **PAX6** human protein. I remind that the details you discovered earlier are as follows:

| | | |
|---|---|---|
| **Affected Patient:** | **33rd amino acid position** | **mRNA Base position** |
| | **Proline** (**P**) | **459 (CCT)** |
| **Wild Type:** | **33rd amino acid position** | **Genomic DNA Base position** |
| | **Alanine** (**A**) | **15915 (GCT)** |

The isolation, amplification and analysis of the target region of the genome could be affected by using restriction enzymes. In this case, there is more than one restriction enzyme whose cut site is dependant upon the mutation and so would produce a differing set of restriction fragments when used with the DNA of **Aniridia** affected patients to that normally expected. As long as those differences were course enough to be detected by a Restriction Fragment Length Polymorphism (**RFLP**) experiment. Software exists to select enzymes to isolate a chosen region of genomic DNA and to fragment that isolate in such a way it is possible to determine whether it includes the unfortunate mutation or not from the pattern of fragments generated.

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 **P**olymerase **C**hain **R**eaction (**PCR**) to isolate the region around the mutation and then to sequence samples from 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 Technology (**MIT**)[1]:

`http://bioinfo.ut.ee/primer3/`

This site is popular with many users wanting the very latest version of the software, complete control over the various options offered by **primer3** and are not too concerned with using a database search to check the uniqueness of the products they will produce.

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:

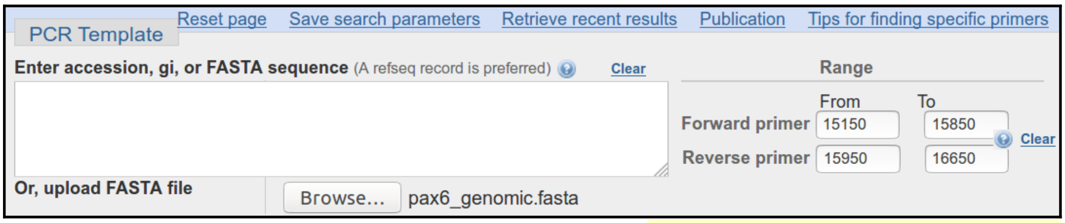
`http://www.ncbi.nlm.nih.gov`

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

[1] The **MIT** link here is to the latest version of **primer3** (**version 4.0.0**, soon **primer4** maybe?).

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

You have established that the mutation of greatest interest is the **G/C** substitution at position **15915** 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 **15150** and continuing **To** base pair **15850**. Set the range for the **Reverse primer** to be **From 15950** and **To 16650**.



PCR Template

Reset page | Save search parameters | Retrieve recent results | Publication | Tips for finding specific primers

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

Or, upload FASTA file Browse... pax6_genomic.fasta

Range

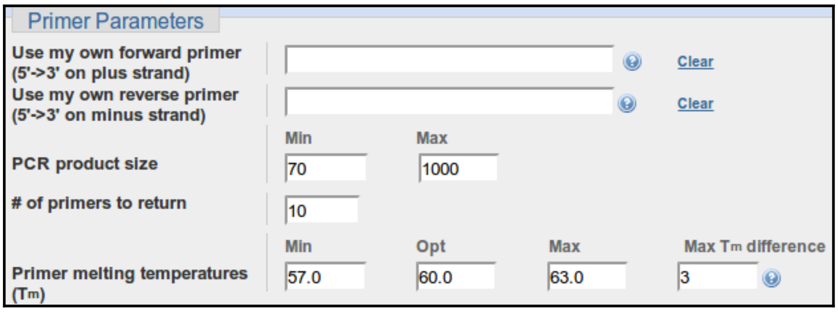
| | From | To |
|---|---|---|
| Forward primer | 15150 | 15850 |
| Reverse primer | 15950 | 16650 |

Clear

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.

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



Primer Parameters

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

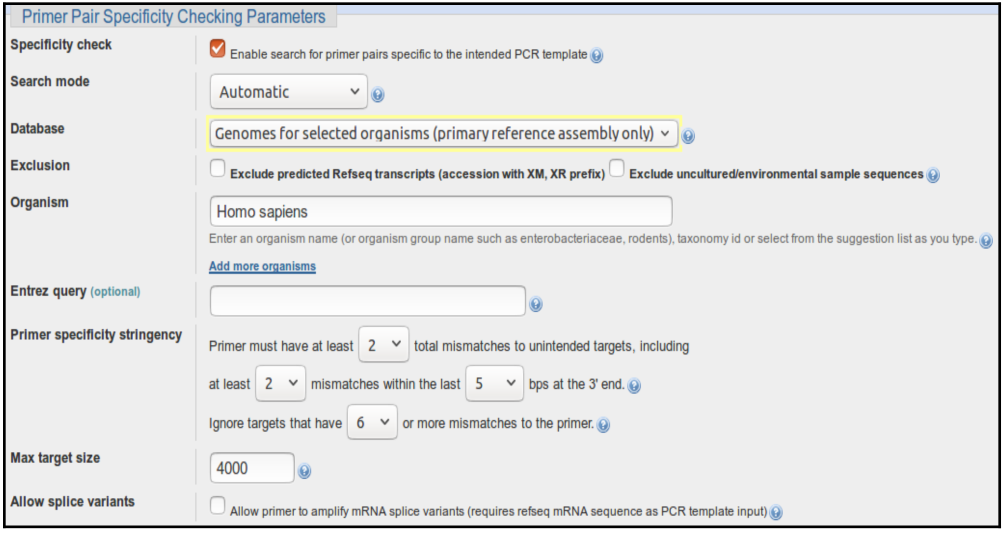
Use my own reverse primer (5'->3' on minus strand) Clear

| | Min | Opt | Max | Max Tm difference |
|---|---|---|---|---|
| PCR product size | 70 | | 1000 | |
| # of primers to return | 10 | | | |
| Primer melting temperatures (Tm) | 57.0 | 60.0 | 63.0 | 3 |

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 **Genomes for selected organisms (primary reference assembly only)**. Leave the **Organism** set as **Homo sapiens**.

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.



Primer Pair Specificity Checking Parameters

Specificity check: ☑ Enable search for primer pairs specific to the intended PCR template

Search mode: Automatic

Database: Genomes for selected organisms (primary reference assembly only)

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.

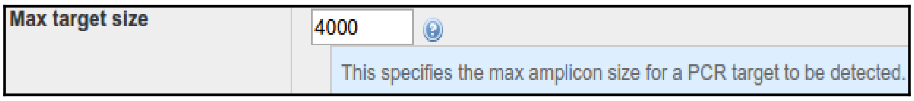
Max target size: 4000

Allow splice variants: □ Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input)

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.

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**.



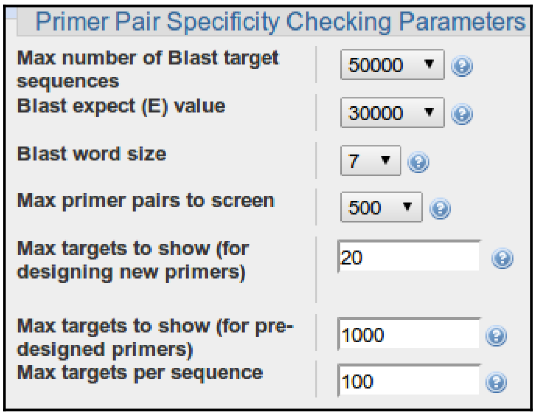
Max target size: 4000

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

I draw your attention to this parameter as, in the next part of this exercise, you will need to set it to a rather surprising value.

[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.

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.

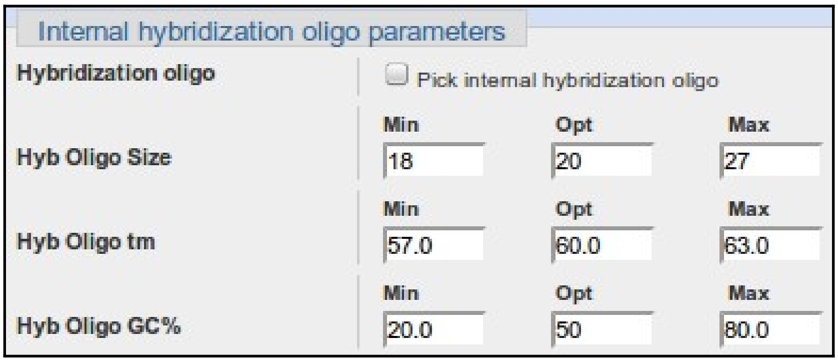


Primer Pair Specificity Checking Parameters

| | |
|---|---|
| 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. You are essentially requesting that exclusion of results with high **Expect Scores** be disabled.

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



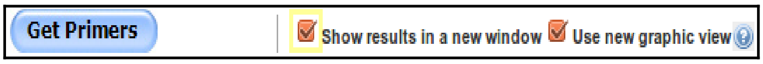
Internal hybridization oligo parameters

| | Min | Opt | Max |
|---|---|---|---|
| Hybridization oligo | □ Pick internal hybridization oligo | | |
| Hyb Oligo Size | 18 | 20 | 27 |
| Hyb Oligo tm | 57.0 | 60.0 | 63.0 |
| 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.

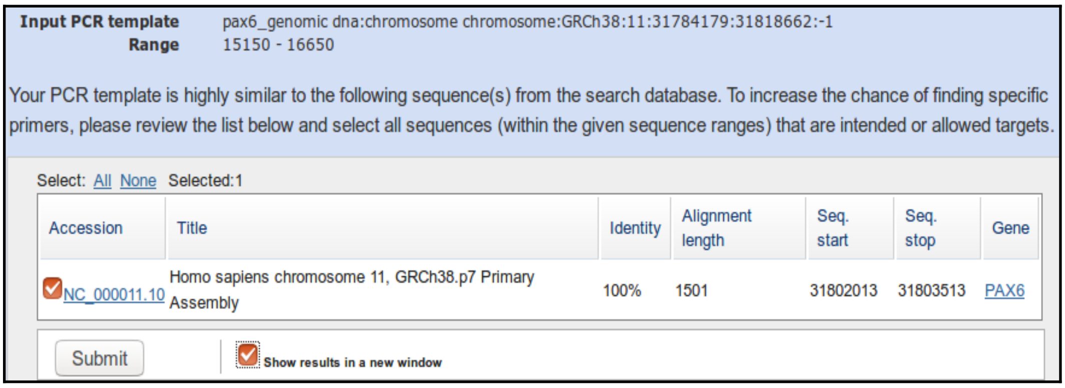
Accept all the **Advanced parameters** as they are. Ask **primer-BLAST** to **Show results in a new window**.

Click on the **Get Primers** button.



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. The **RefSeq** entry identified is the **PAX6 RefSeqGene** sequence you examined in a previous exercise.

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.



**Input PCR template** pax6_genomic dna:chromosome chromosome:GRCh38:11:31784179:31818662:-1
**Range** 15150 - 16650

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.

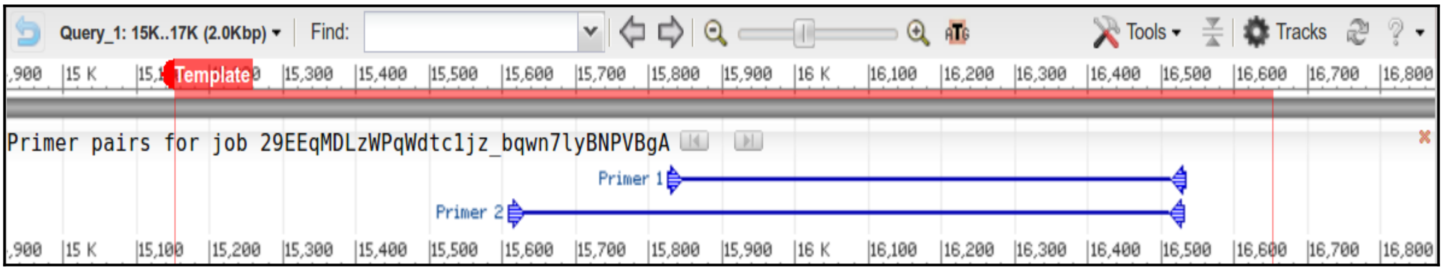
Select: All None Selected:1

| Accession | Title | Identity | Alignment length | Seq. start | Seq. stop | Gene |
|---|---|---|---|---|---|---|
| ☑ NC_000011.10 | Homo sapiens chromosome 11, GRCh38.p7 Primary Assembly | 100% | 1501 | 31802013 | 31803513 | PAX6 |

Finally, all is ready, so ask to **Show results in a new window** once more and then click on the **Submit** button.

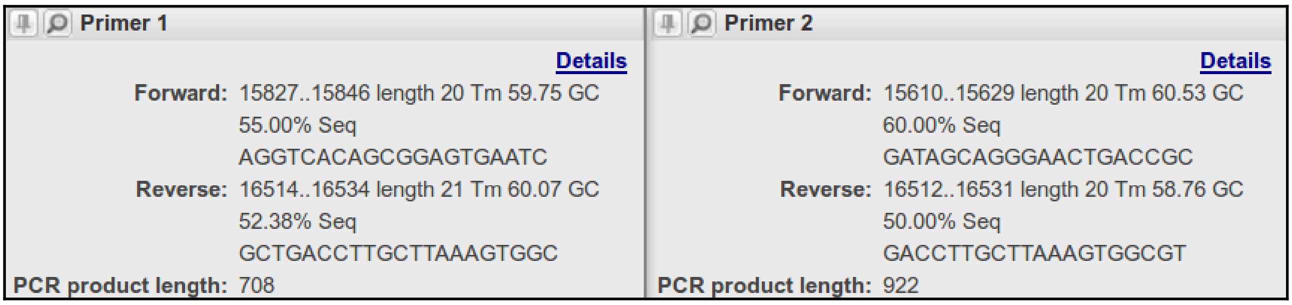
Once you have revelled in the opportunity to twiddle the fingers and scratch the ear(s) whilst **primers3** and **blast** go merrily about their appointed tasks, you will receive your results. These should look disarmingly like mine if all has gone well.

The summary **Graphic view** suggest just **2** solutions met the default criteria for success used by **primer3**. Up to **10** were permitted[3].


Query_1: 15K..17K (2.0Kbp)
Find:
Tools
Tracks
Template
,900 15 K 15,100 15,200 15,300 15,400 15,500 15,600 15,700 15,800 15,900 16 K 16,100 16,200 16,300 16,400 16,500 16,600 16,700 16,800
Primer 1
Primer 2

Primer pairs for job 29EEqMDLzWPqWdtc1jz_bqwn7lyBNPVBgA

Hover your mouse over one or more and further details will pop up in separate windows.



**Primer 1**

[Details](#)

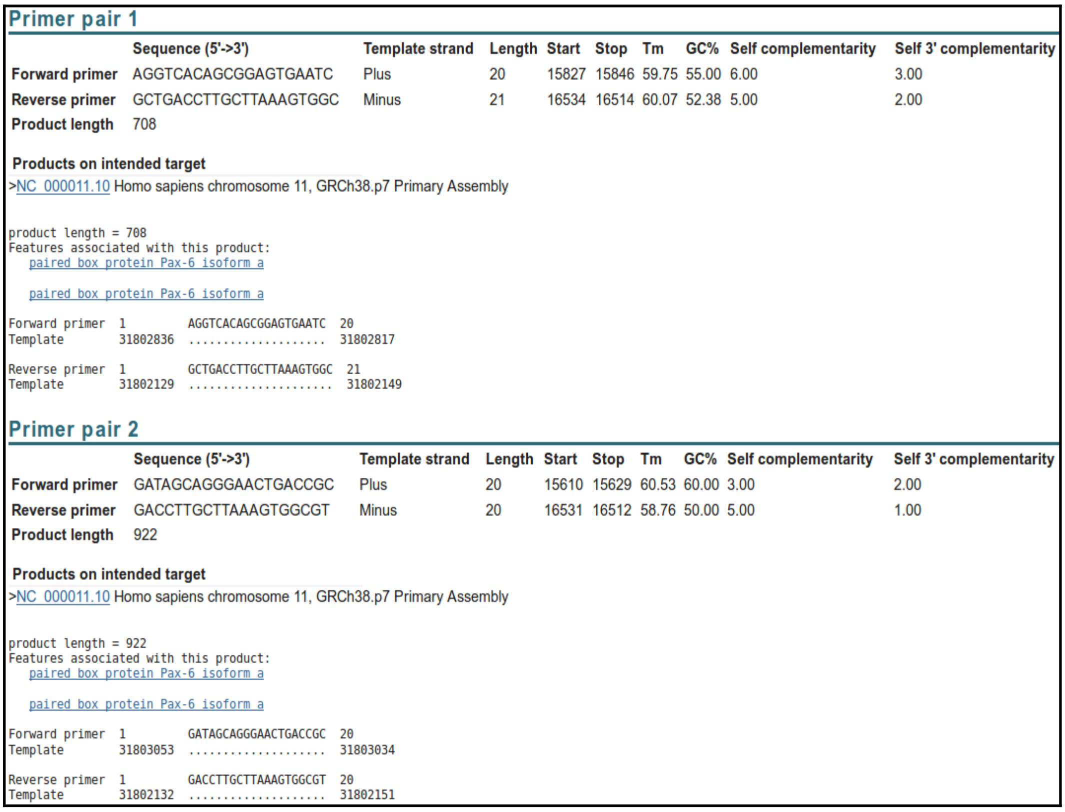
**Forward:** 15827..15846 length 20 Tm 59.75 GC 55.00% Seq AGGTCACAGCGGAGTGAATC
**Reverse:** 16514..16534 length 21 Tm 60.07 GC 52.38% Seq GCTGACCTTGCTTAAAGTGGC
**PCR product length:** 708

**Primer 2**

[Details](#)

**Forward:** 15610..15629 length 20 Tm 60.53 GC 60.00% Seq GATAGCAGGGAACTGACCGC
**Reverse:** 16512..16531 length 20 Tm 58.76 GC 50.00% Seq GACCTTGCTTAAAGTGGCGT
**PCR product length:** 922

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



Primer pair 1

| | Sequence (5'->3') | Template strand | Length | Start | Stop | Tm | GC% | Self complementarity | Self 3' complementarity |
|---|---|---|---|---|---|---|---|---|---|
| **Forward primer** | AGGTCACAGCGGAGTGAATC | Plus | 20 | 15827 | 15846 | 59.75 | 55.00 | 6.00 | 3.00 |
| **Reverse primer** | GCTGACCTTGCTTAAAGTGGC | Minus | 21 | 16534 | 16514 | 60.07 | 52.38 | 5.00 | 2.00 |
| **Product length** | 708 | | | | | | | | |

**Products on intended target**

>NC_000011.10 Homo sapiens chromosome 11, GRCh38.p7 Primary Assembly

```
product length = 708
Features associated with this product:
   paired box protein Pax-6 isoform a

   paired box protein Pax-6 isoform a

Forward primer  1         AGGTCACAGCGGAGTGAATC  20
Template        31802836  ....................  31802817

Reverse primer  1         GCTGACCTTGCTTAAAGTGGC  21
Template        31802129  .....................  31802149
```

Primer pair 2

| | Sequence (5'->3') | Template strand | Length | Start | Stop | Tm | GC% | Self complementarity | Self 3' complementarity |
|---|---|---|---|---|---|---|---|---|---|
| **Forward primer** | GATAGCAGGGAACTGACCGC | Plus | 20 | 15610 | 15629 | 60.53 | 60.00 | 3.00 | 2.00 |
| **Reverse primer** | GACCTTGCTTAAAGTGGCGT | Minus | 20 | 16531 | 16512 | 58.76 | 50.00 | 5.00 | 1.00 |
| **Product length** | 922 | | | | | | | | |

**Products on intended target**

>NC_000011.10 Homo sapiens chromosome 11, GRCh38.p7 Primary Assembly

```
product length = 922
Features associated with this product:
   paired box protein Pax-6 isoform a

   paired box protein Pax-6 isoform a

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

Reverse primer  1         GACCTTGCTTAAAGTGGCGT  20
Template        31802132  ....................  31802151
```

[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 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 substantially more generous.

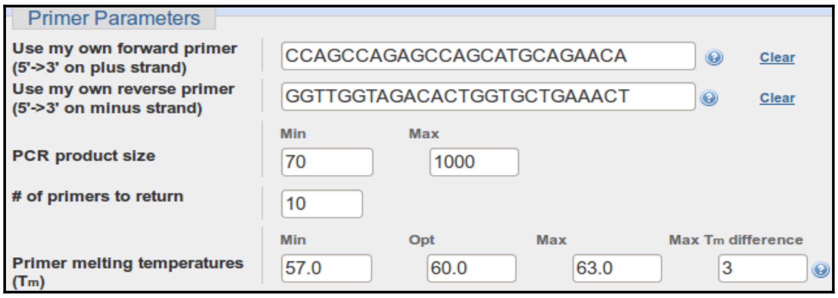
As well as suggesting primers for PCR (or other purposes) and (optionally) suggesting hybridisation oligos, **primer-BLAST** can be used to evaluate user-selected primers. Earlier, you saved a pair of primer sequences associated with **PAX6** when searching the nucleotide databases at the **NCBI**. It 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 from the **Specialized BLAST** section.

Primer-BLAST
Design primers specific to your PCR template

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

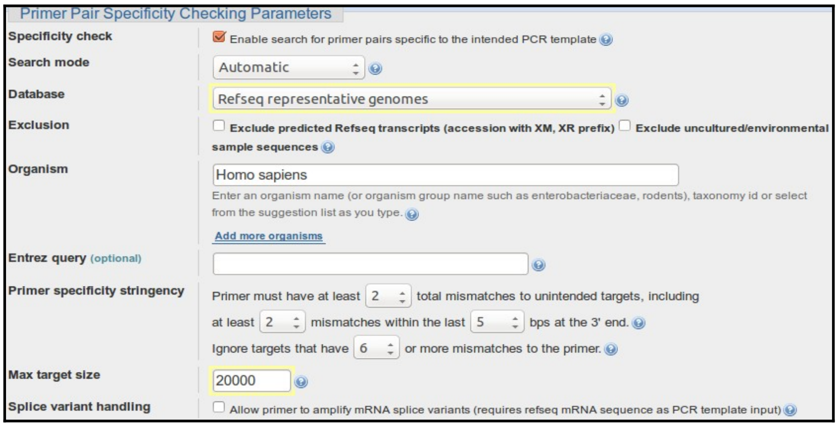


Primer Parameters

| | | | | | |
|---|---|---|---|---|---|
| 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

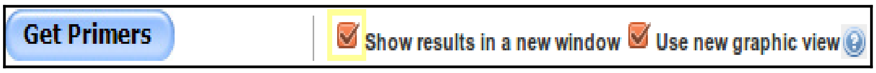
Splice variant handling: □ Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input)

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

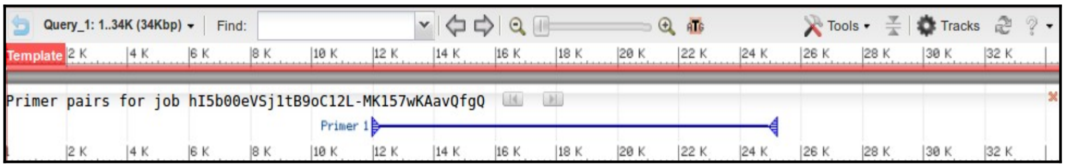
Leave the **Organism** as **Homo sapiens**.

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.

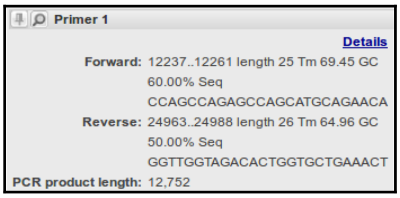


Get Primers ☑ Show results in a new window ☑ Use new graphic view



Query_1: 1..34K (34Kbp)
Find:
Tools
Tracks
Template
2 K 4 K 6 K 8 K 10 K 12 K 14 K 16 K 18 K 20 K 22 K 24 K 26 K 28 K 30 K 32 K
Primer pairs for job hI5b00eVSj1tB9oC12L-MK157wKAAvQfgQ
Primer 1

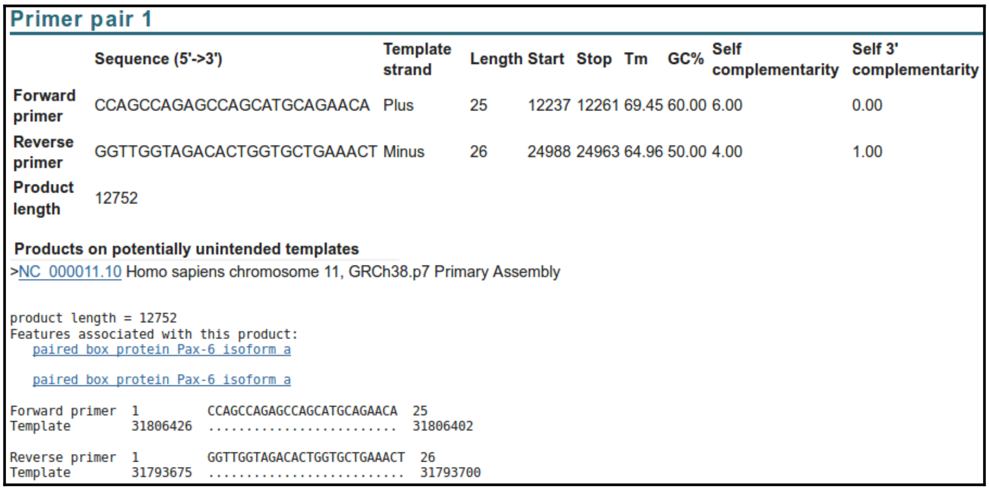
After a short thrill filled pause, you will receive a result that should again looks more that a trifle like mine.



Primer 1

Details

Forward: 12237..12261 length 25 Tm 69.45 GC 60.00% Seq CCAGCCAGAGCCAGCATGCAGAACA
Reverse: 24963..24988 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 | 12237 | 12261 | 69.45 | 60.00 | 6.00 | 0.00 |
| Reverse primer | GGTTGGTAGACACTGGTGCTGAAACT | Minus | 26 | 24988 | 24963 | 64.96 | 50.00 | 4.00 | 1.00 |
| Product length | 12752 | | | | | | | | |

Products on potentially unintended templates

>NC_000011.10 Homo sapiens chromosome 11, GRCh38.p7 Primary Assembly

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

   paired box protein Pax-6 isoform a

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

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

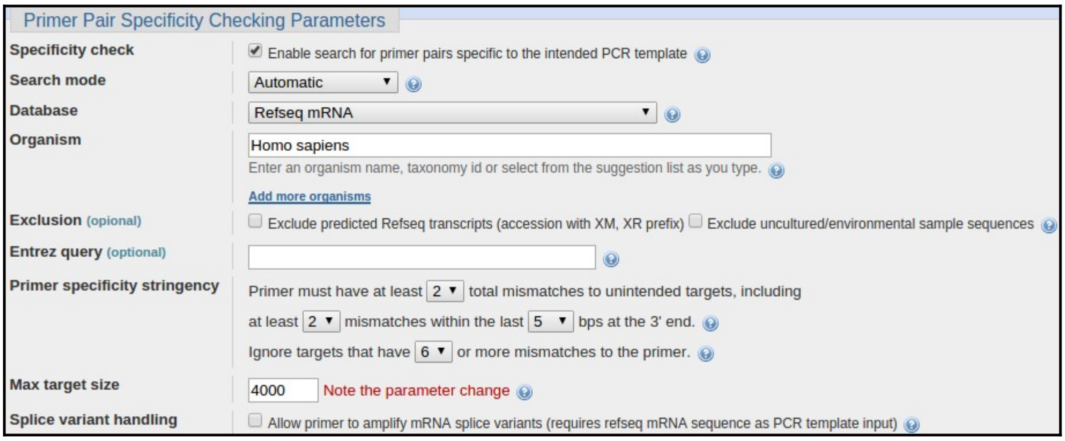
Seemingly a fine match. Even the single **potentially unintended product** reported is actually the **intended product**. For some reason, **Primer-BLAST** does not eliminate predictable intended products when investigating user specified primers[5]?

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.

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

[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.

Reading the only paper referenced seems a little like hard work! Better by far to investigate the only sensible reason for the prediction of such an outrageously large PCR product, by experiment. A sensible conjecture is 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_mrna.fasta` as the **PCR Template**.



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)

Move back to your last **primer-BLAST** launch page. This time, load `pax6_mrna.fasta` as the **PCR Template**.

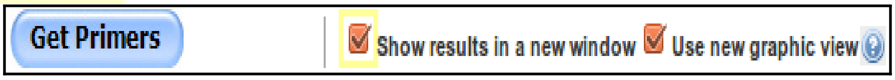
In the **Primer Pair Specificity Checking Parameters** section, set the **Database** selection set to **Refseq mRNA** and leave the organism set to **Homo sapiens**.

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**.

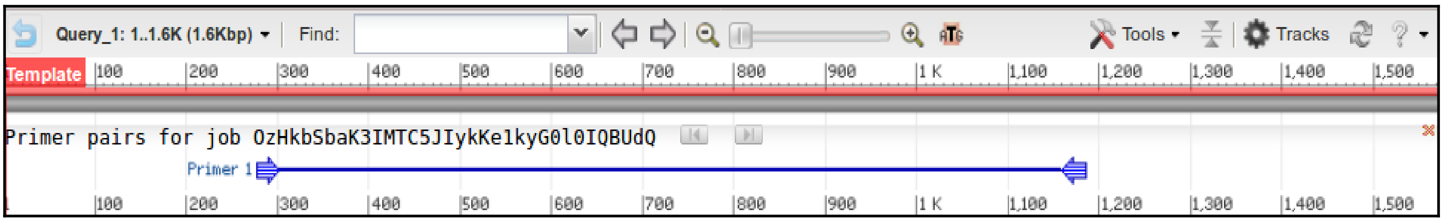
These selections suppose that the design of **PCR** product was for selection from a library of all human cDNAs.

Ask **primer-BLAST** to **Show results in a new window**.

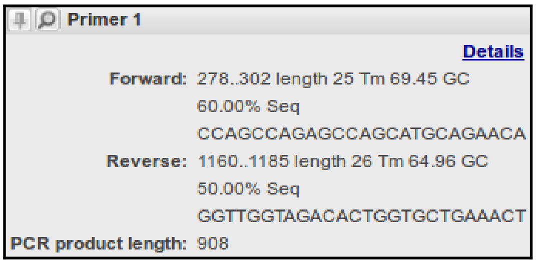
Click on the **Get Primers** button.



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.



Primer pairs for job OzHkbSbaK3IMTC5JIykKe1kyG0l0IQBUdQ



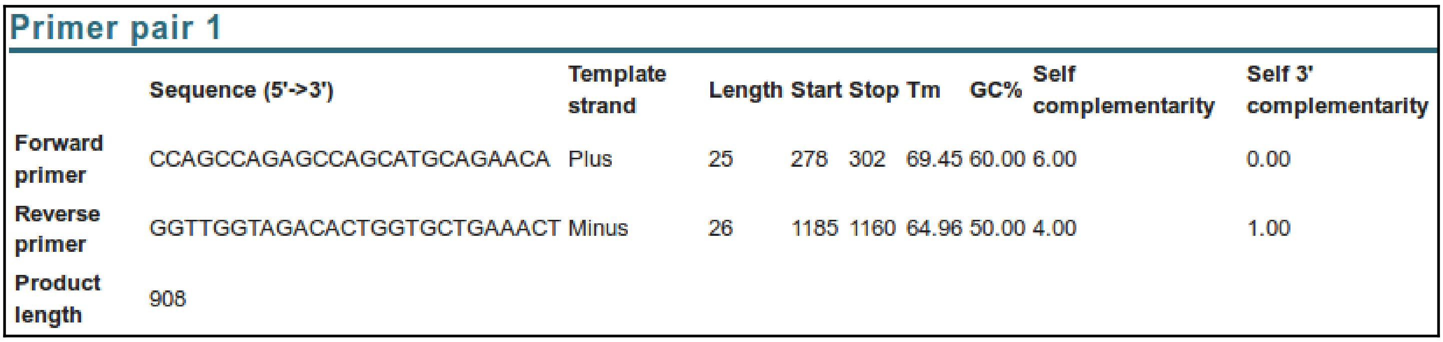
Primer 1

Details

Forward: 278..302 length 25 Tm 69.45 GC 60.00% Seq CCAGCCAGAGCCAGCATGCAGAACA

Reverse: 1160..1185 length 26 Tm 64.96 GC 50.00% Seq GGTTGGTAGACACTGGTGCTGAAACT

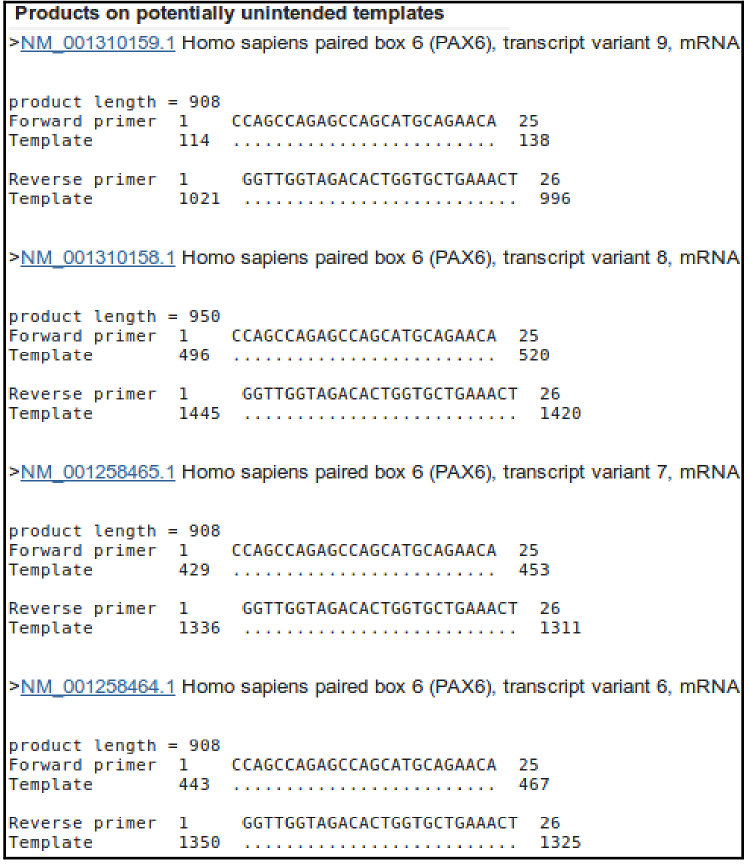
PCR product length: 908



Primer pair 1

| | Sequence (5'->3') | Template strand | Length | Start | Stop | Tm | GC% | Self complementarity | Self 3' complementarity |
|---|---|---|---|---|---|---|---|---|---|
| Forward primer | CCAGCCAGAGCCAGCATGCAGAACA | Plus | 25 | 278 | 302 | 69.45 | 60.00 | 6.00 | 0.00 |
| Reverse primer | GGTTGGTAGACACTGGTGCTGAAACT | Minus | 26 | 1185 | 1160 | 64.96 | 50.00 | 4.00 | 1.00 |
| Product length | 908 | | | | | | | | |

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.



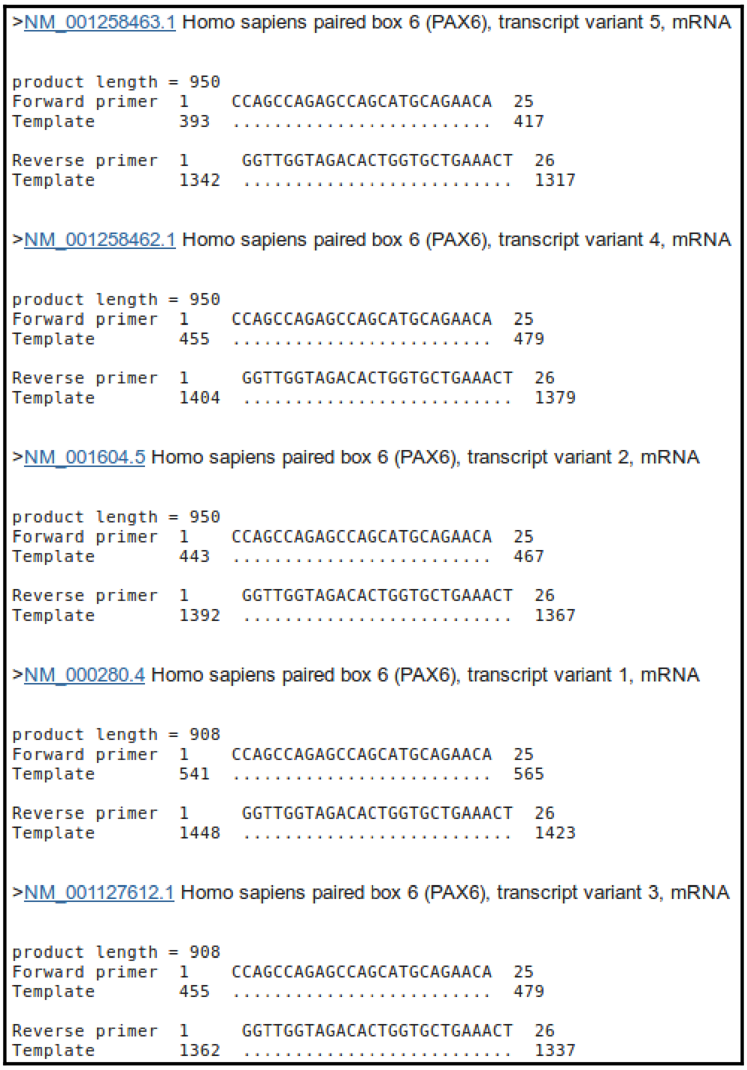
**Products on potentially unintended templates**

```
>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

>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
```

The first thing to note is that the template (the mRNA sequence in the file `pax6_mrna.fasta`) is not a **RefSeq** mRNA. It comes from the **GenBank** database and so was included in the "non-redundant" union of databases you searched earlier.

**Genbank** sequences are generally generated directly from a specific sequencing project. **RefSeq** mRNAs are generally consensus sequences computed from the evidence represented by **Genbank** sequences. Consequently, there is no unintended product that we can ignore because it relates to the original template sequence.

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 **9** unintended products come from **9** of the **11 RefSeq PAX6** transcripts found in the databases by test search and later detected by **blast**.



```
>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
```

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 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?

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

**DPJ – 2019.01.30**

# 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**

### 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.

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

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.

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

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

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).

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!

### 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. 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".

# DPJ – 2019.01.30