Tutorial on how to estimate best PCR conditions and uniqueness of primers.

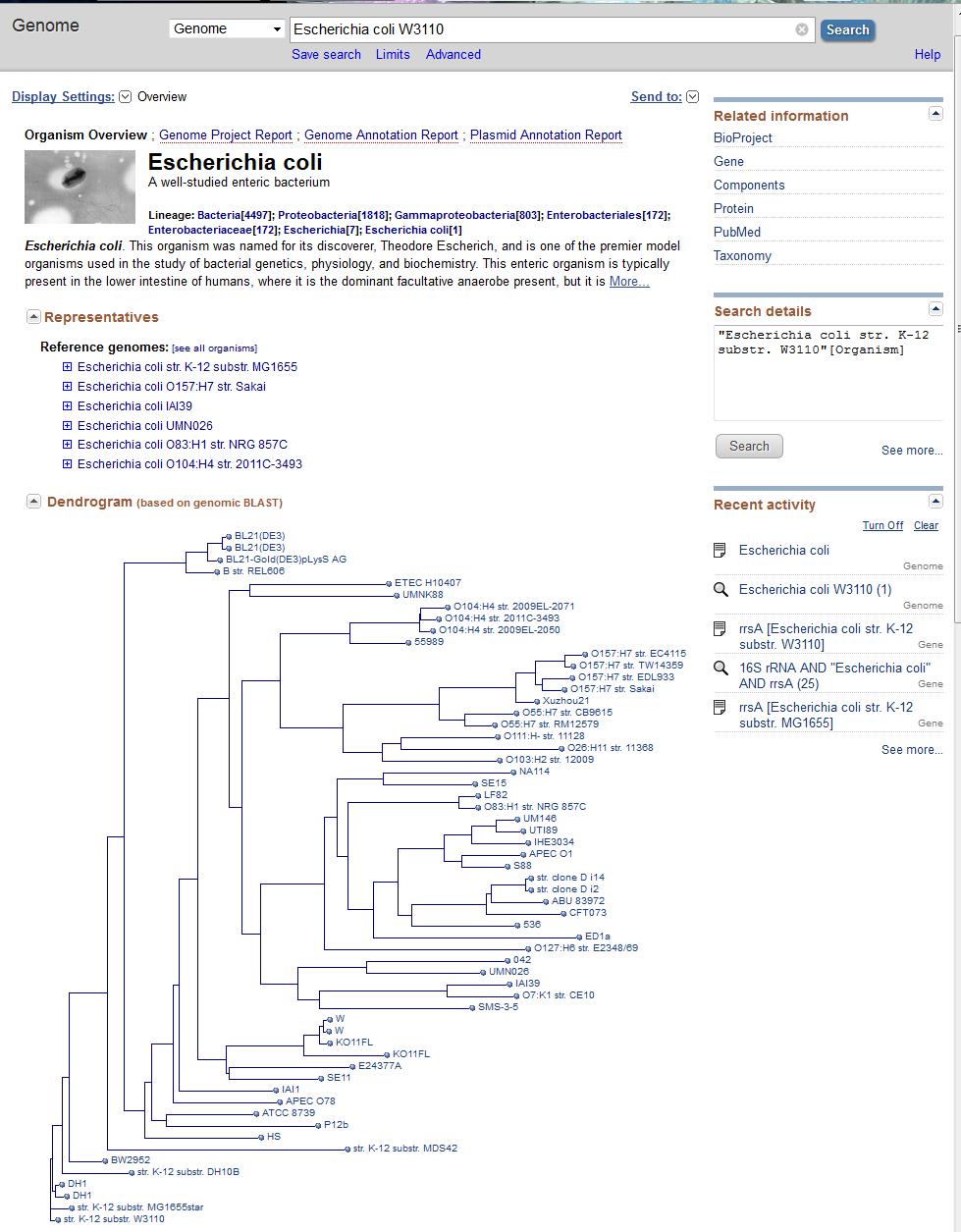
This case focuses on primers used to amplify the bacterial 16S rDNA genes, which are shared in prokaryotes and archaea.

We have two goals:

1. Make sure the primers work in E coli (our optimization genome)
2. Use the primers in human oral samples, so estimate the amount of bacterial DNA contributing to the total. Because of this we need primers that work across species, which means we are using some mixtures. This can make some of the tools we use choke – they don’t all accept mixed bases as something they can handle – you may have to try each of the individual sequences separately with some of the tools.

How did I find this? I looked at the description of the genomic DNA we bought, which says this is E coli K-12, substrain W3110.

In the NCBI Genome section I enter this information and get the following:



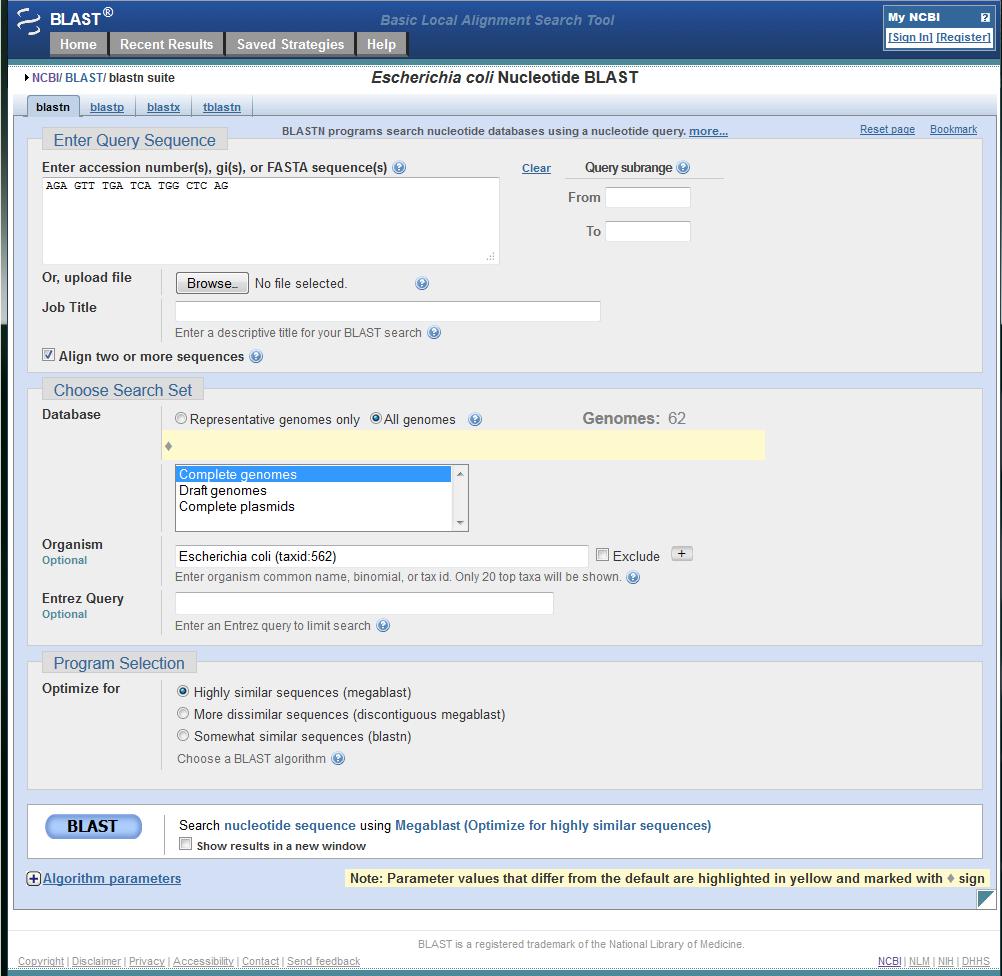
Click on the branch of the tree that has my substrain and you get the following:



If I click on the GenBank link I will get the ncbi Reference Sequence identifier, NC\_0007779.1, which is a useful bit of information.

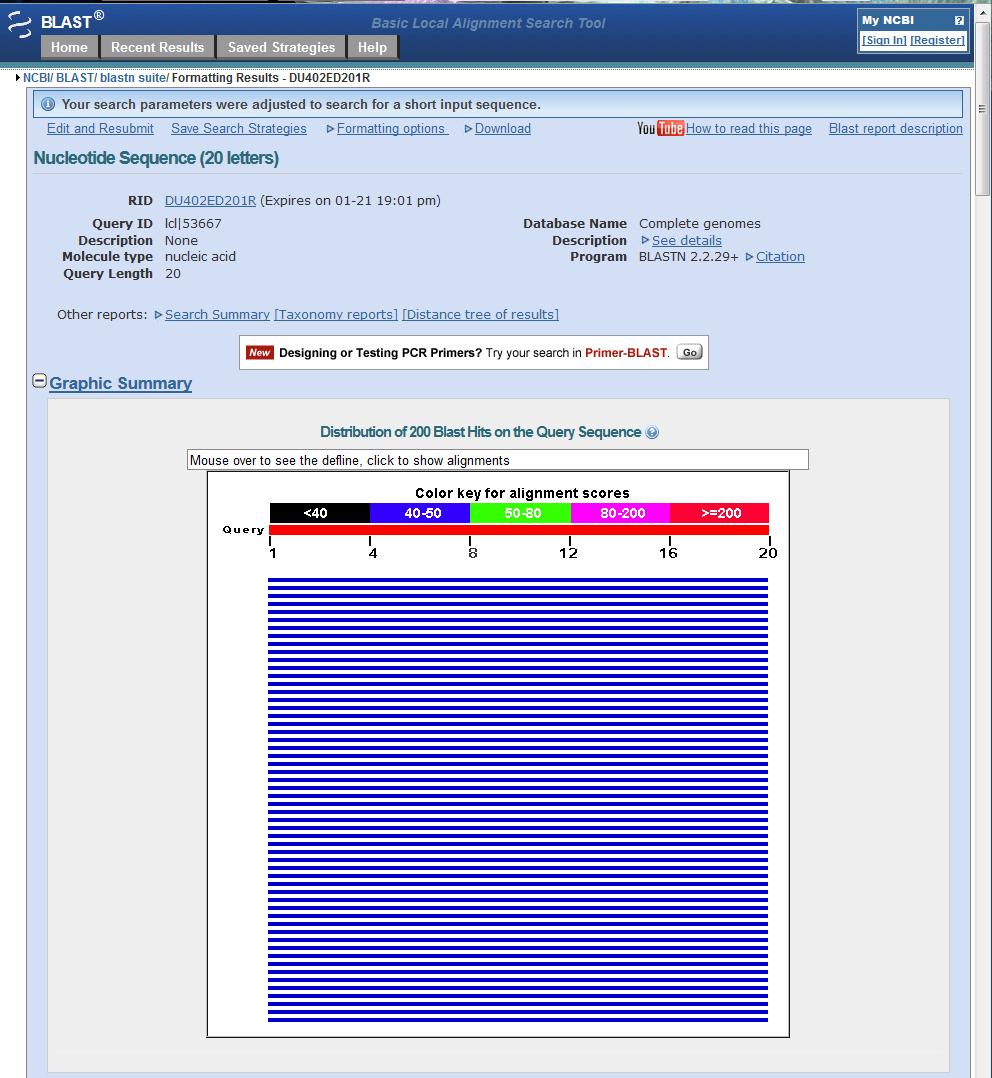
This tells me that the whole genome is indeed sequenced – I could BLAST my primers against this genome to find what region of the genome I am in and if my primers match.

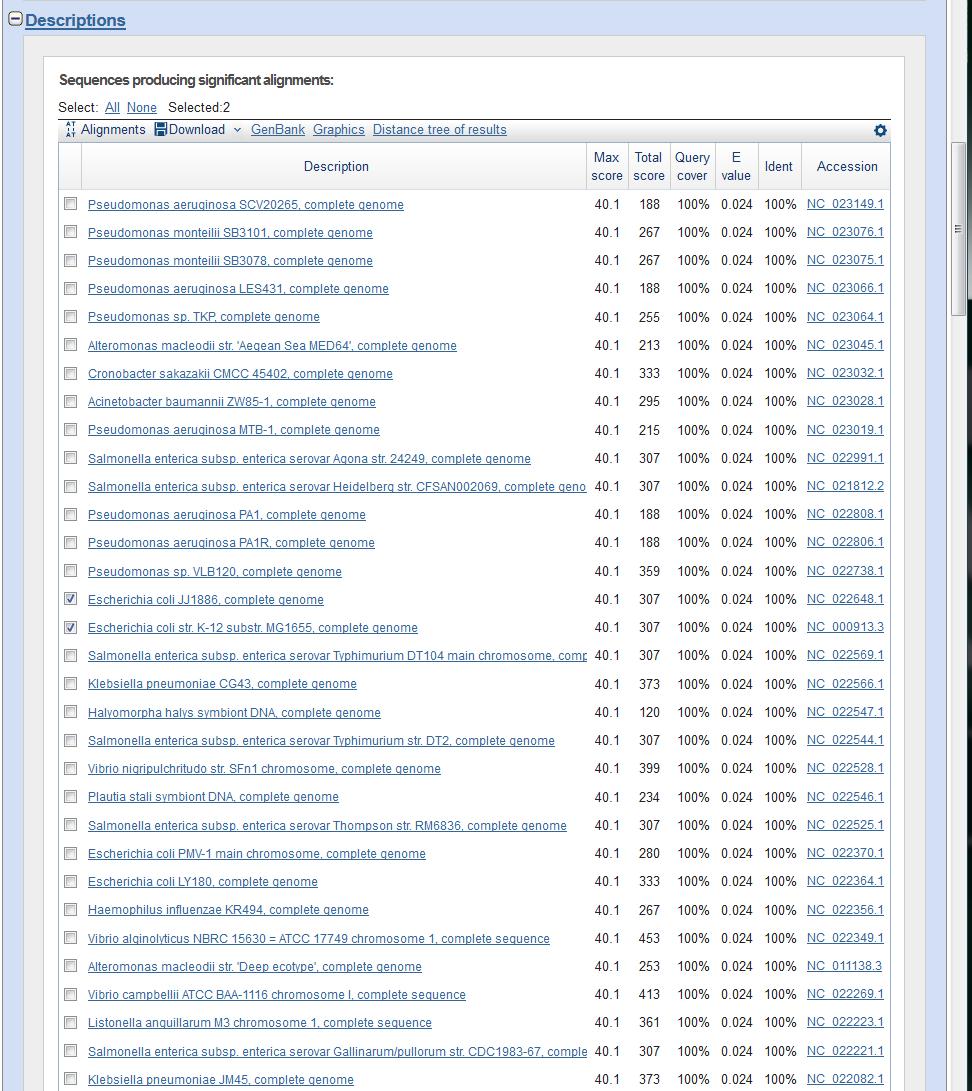
If I take my first primer and do a BLAST search against 62 complete bacterial genomes, I get this output:



You can see that I put in one of my 27F primers, and this screen lets me compare it to the set of E coli genomics (Organism, with the taxon id).

The output is shown below, I had to split it into 3 screens.



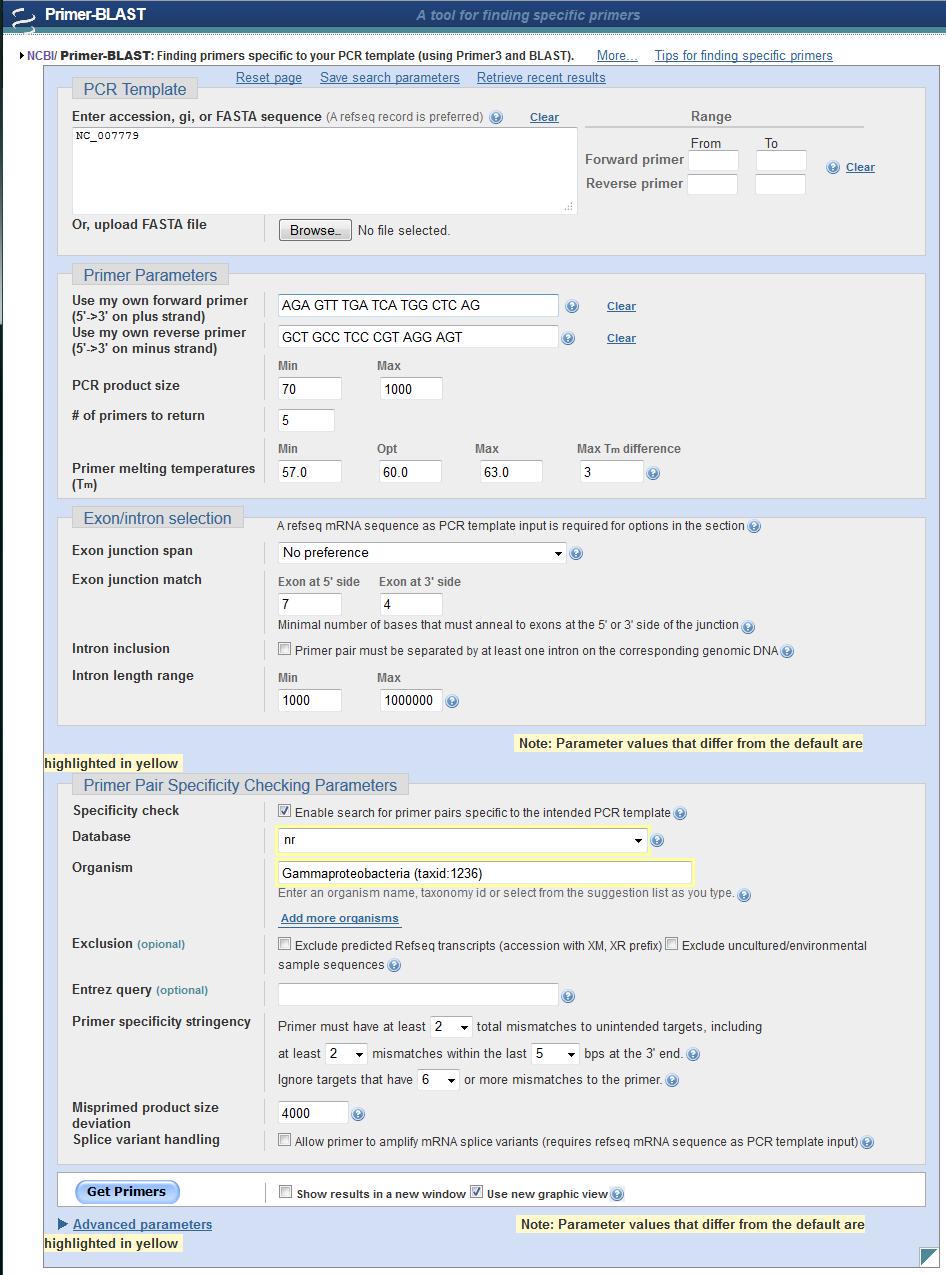


The above shows how conserved the primer sequence is across these genomes.

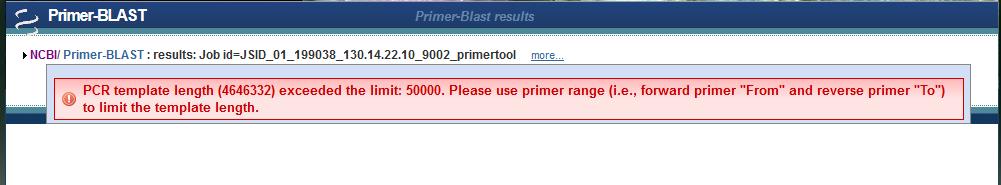
If you look at the alignment in one of the genomes, you will see there are several per genome – this is because most bacteria have multiple copies of the ribosomal RNA genes (usually identical).



I can also try the PrimerBLAST option, which lets me align two short sequences instead of one. In this case I need the accession number for this organism (noted during the Genome search). I decided to see what hits I get in the larger group of bacteria that E coli belongs to, which is the Gammaproteobacter (I can get this by looking at a taxonomic tree in NCBI).



Here is the output:

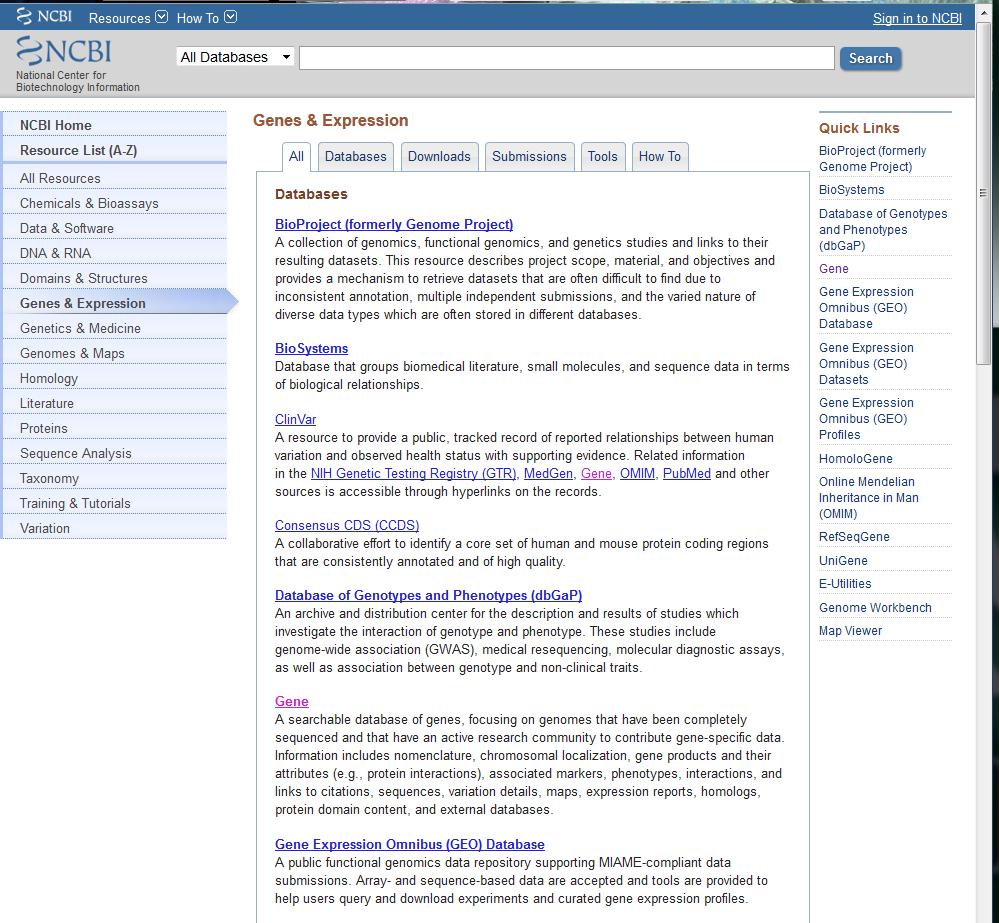


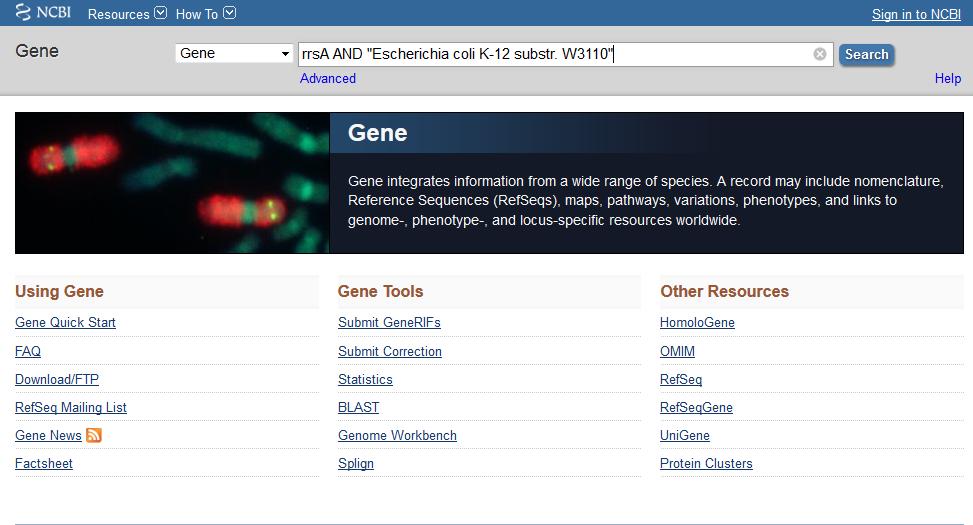
Well, rats. Even though I was invited to try this, it turns out that I cannot start with the whole genome.

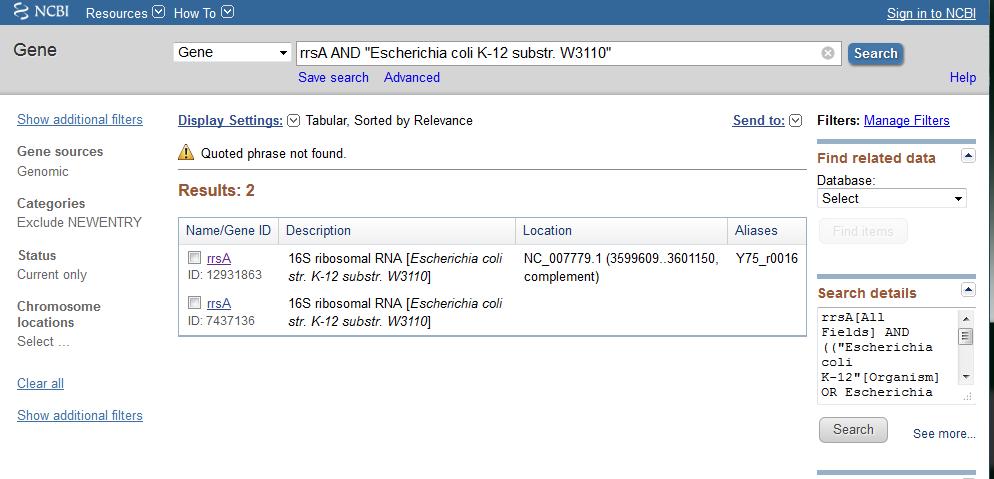
So, I need to find a different starting point. In this case, I do know what gene I am looking for, the ribosomal RNA gene for 16S. The code for these operons is rrs[a,b,c,d..] etc..

So I can use the Gene option from NCBI .

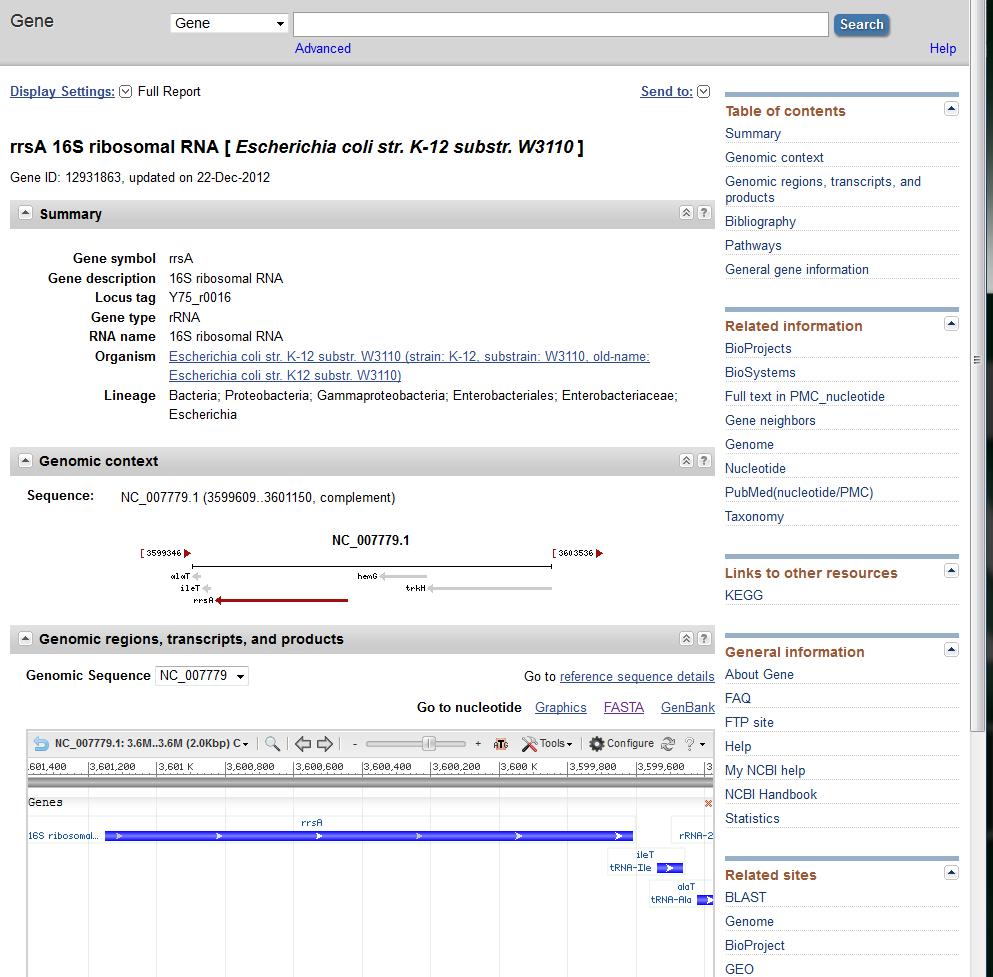
Here I am looking for the rrsA gene in E coli K-12, substrain W3110.



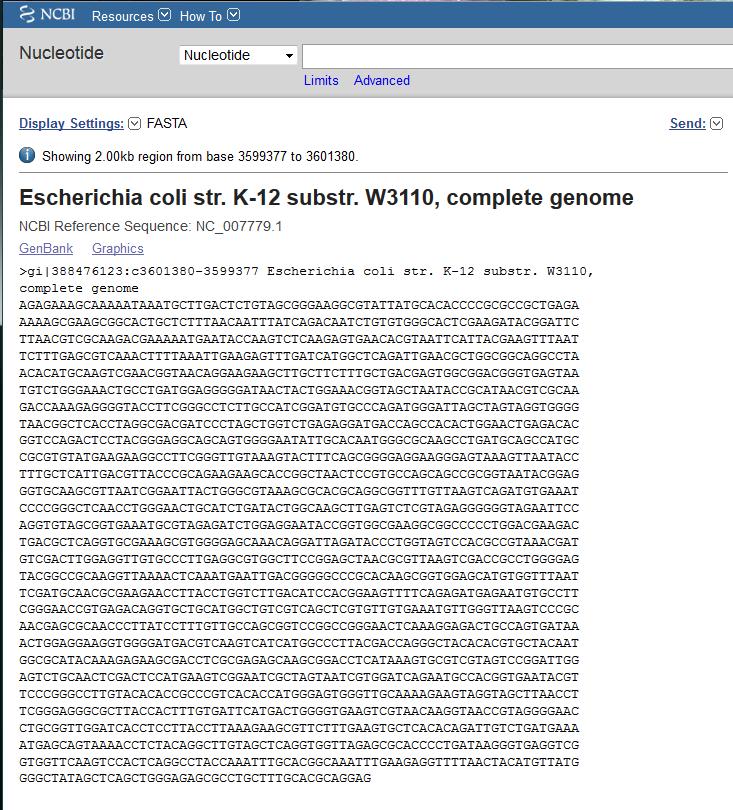




If you click on the top return you will get the following



If you click on FASTA you will get the sequence you need.



You can double-check the information: W3110 has accession NC\_007779 and the ribosomal operon has gi:388476123

The sequence itself is

AGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAGGCGTATTATGCACACCCCGCGCCGCTGAGA

AAAAGCGAAGCGGCACTGCTCTTTAACAATTTATCAGACAATCTGTGTGGGCACTCGAAGATACGGATTC

TTAACGTCGCAAGACGAAAAATGAATACCAAGTCTCAAGAGTGAACACGTAATTCATTACGAAGTTTAAT

TCTTTGAGCGTCAAACTTTTAAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTA

ACACATGCAAGTCGAACGGTAACAGGAAGAAGCTTGCTTCTTTGCTGACGAGTGGCGGACGGGTGAGTAA

TGTCTGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAA

GACCAAAGAGGGGTACCTTCGGGCCTCTTGCCATCGGATGTGCCCAGATGGGATTAGCTAGTAGGTGGGG

TAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGACAC

GGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGC

CGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACC

TTTGCTCATTGACGTTACCCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAG

GGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAAAT

CCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCC

AGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGAC

TGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGAT

GTCGACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAG

TACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAAT

TCGATGCAACGCGAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTCAGAGATGAGAATGTGCCTT

CGGGAACCGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGC

AACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTCCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAA

ACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAAT

GGCGCATACAAAGAGAAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGG

AGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGT

TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAACCT

TCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAAC

CTGCGGTTGGATCACCTCCTTACCTTAAAGAAGCGTTCTTTGAAGTGCTCACACAGATTGTCTGATGAAA

ATGAGCAGTAAAACCTCTACAGGCTTGTAGCTCAGGTGGTTAGAGCGCACCCCTGATAAGGGTGAGGTCG

GTGGTTCAAGTCCACTCAGGCCTACCAAATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTATG

GGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTGCACGCAGGAG

This screen also lets me pick primers (now I have a short enough sequence that it should work).

The primers I want to check include

27f

We have 27f-2\_1, which is AGA GTT TGA TCC TGG CTC AG \*

And 27f-YM which is a combination of

AGA GTT TGA TYM TGG CTC AG

AGA GTT TGA TCA TGG CTC AG \*\*

AGA GTT TGA TCC TGG CTC AG \*

AGA GTT TGA TTA TGG CTC AG

AGA GTT TGA TTC TGG CTC AG

338R\_V2\_BMC

GCT GCC TCC CGT AGG AGT

536R\_BMC

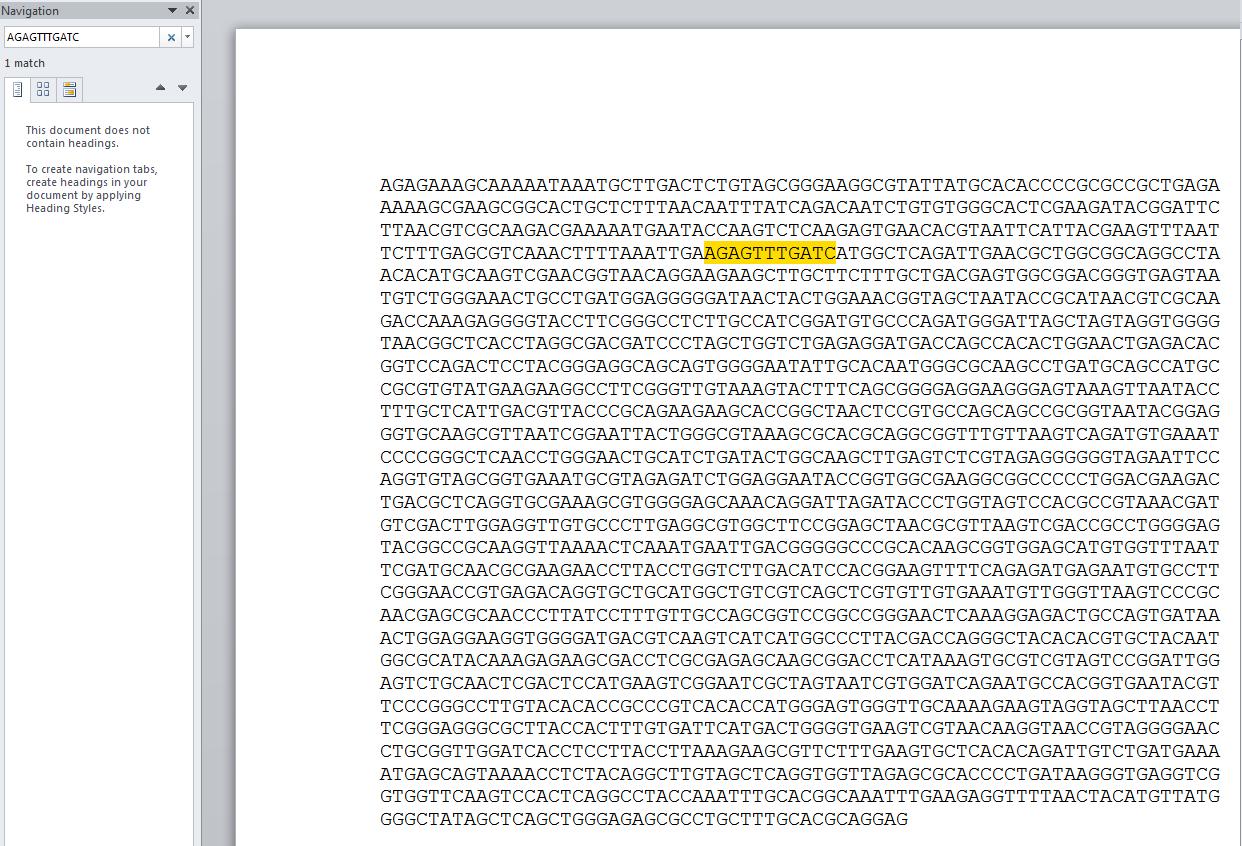
GTA TTA CCG CGG CKG CTG

GTA TTA CCG CGG CGG CTG

GTA TTA CCG CGG CTG CTG

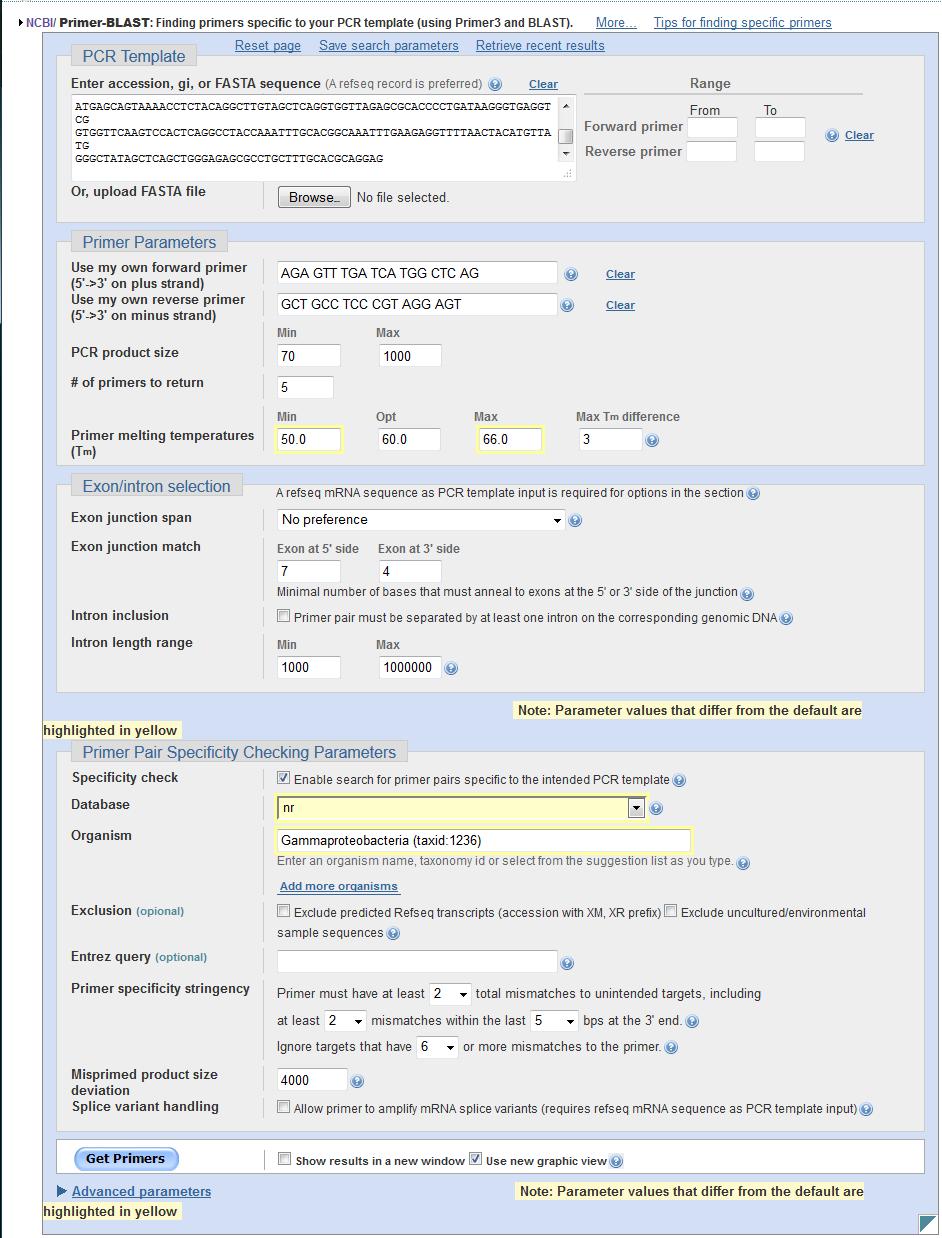
Other pairs that should give us products, but they will be larger, include 27F with 1492R, and 519F\_BMC with 1492R.

Note that unlike using BLAST, when you pick primers, you have to have an exact match (that is the point of the exercise – to have the best possible primers). Since our primers are mixtures that aim to amplify a number of species, and we don’t know beforehand which ones are to our genome, I like to do a quick screen for the primers that work. You can simply use FIND in Word. I usually enter just the first 6 bases, then keep adding more until I get a perfect match, or figure out where the mismatch is and correct it. Note that spaces are also considered, so if my query has no spaces I need to leave them out of the primer sequence). Then I can put in the sequences that match in PrimerBLAST and see what other information I get.

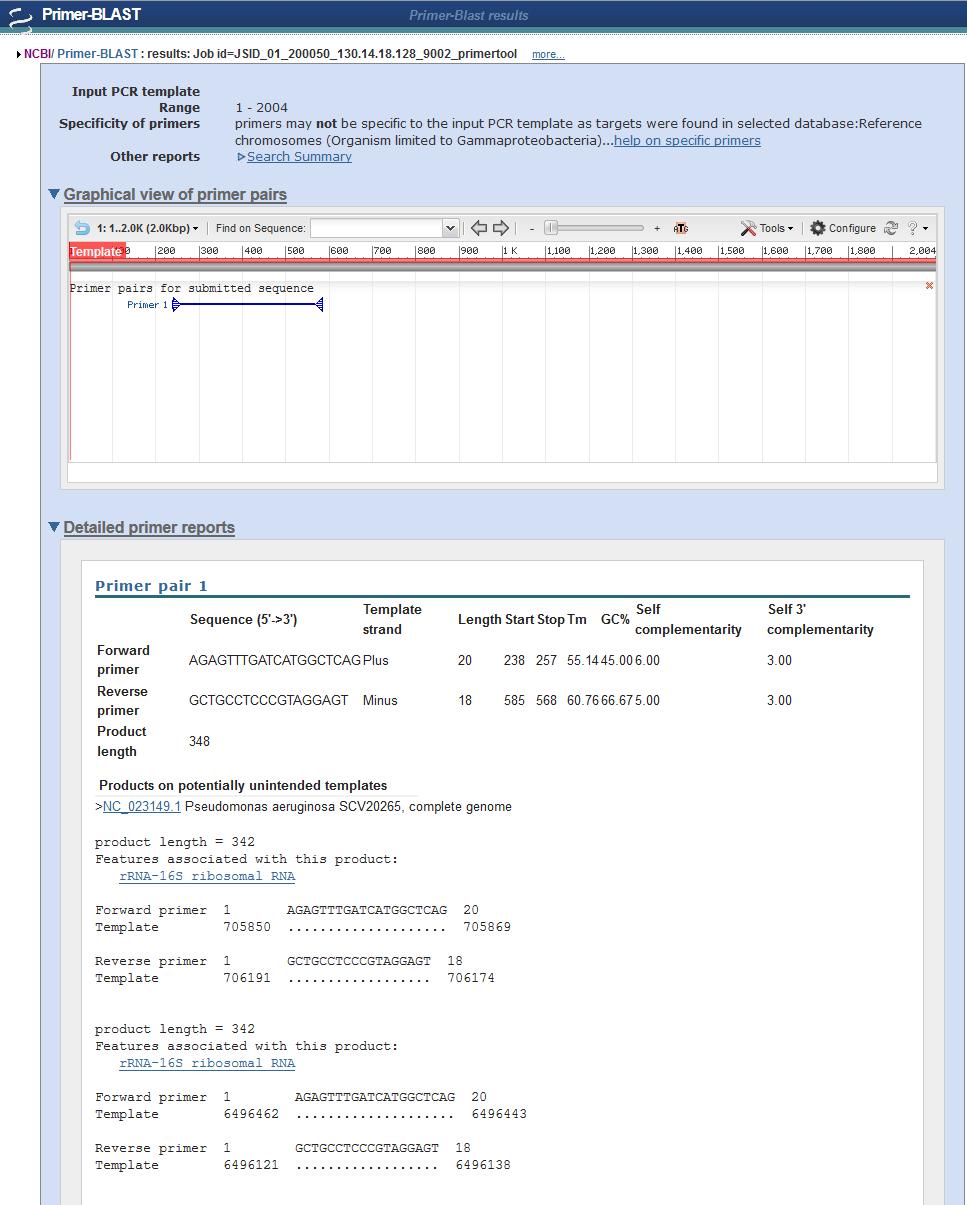


This is how I figured out that in this genome I could not use 27f-2\_1.

Using the two I did find, I can go back to the Pick Primers at the ncbi web site.



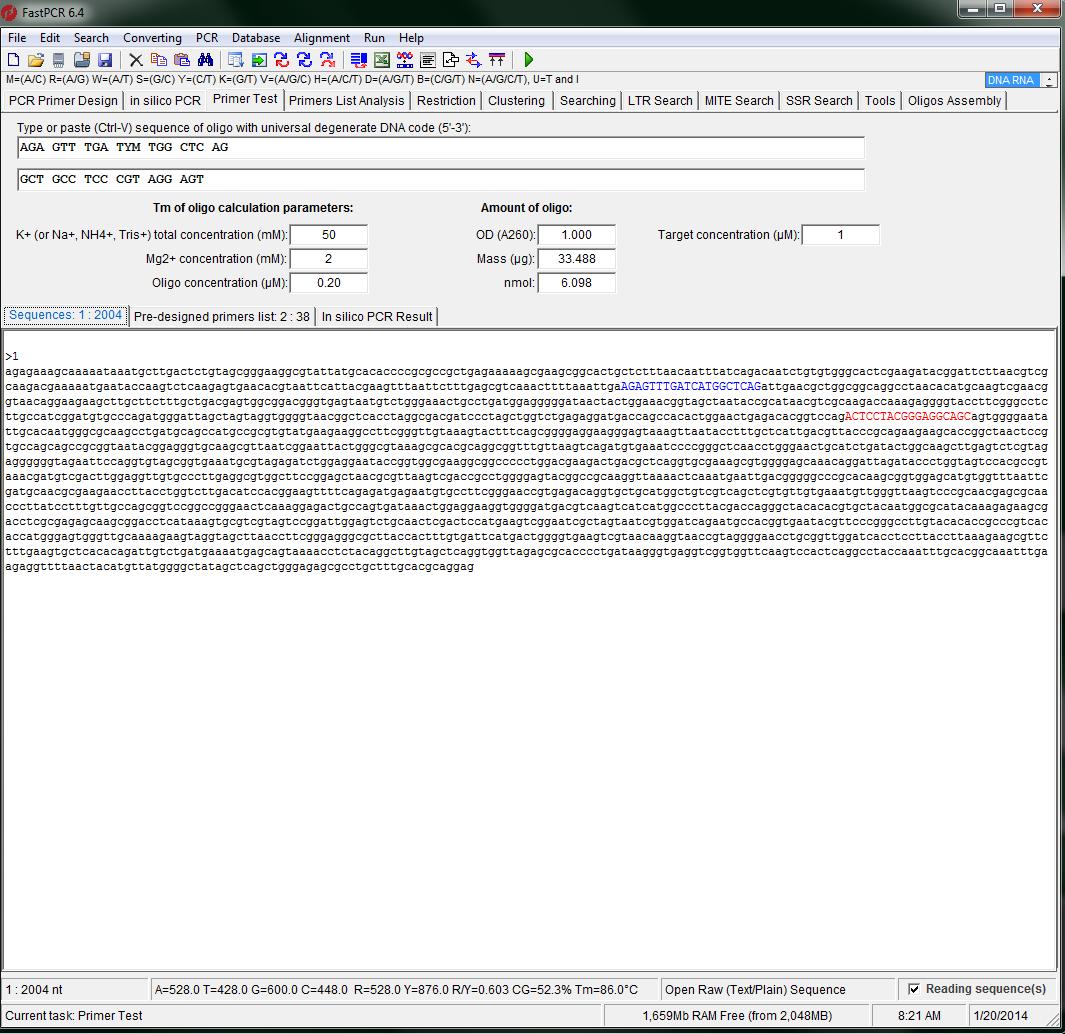
Results are:

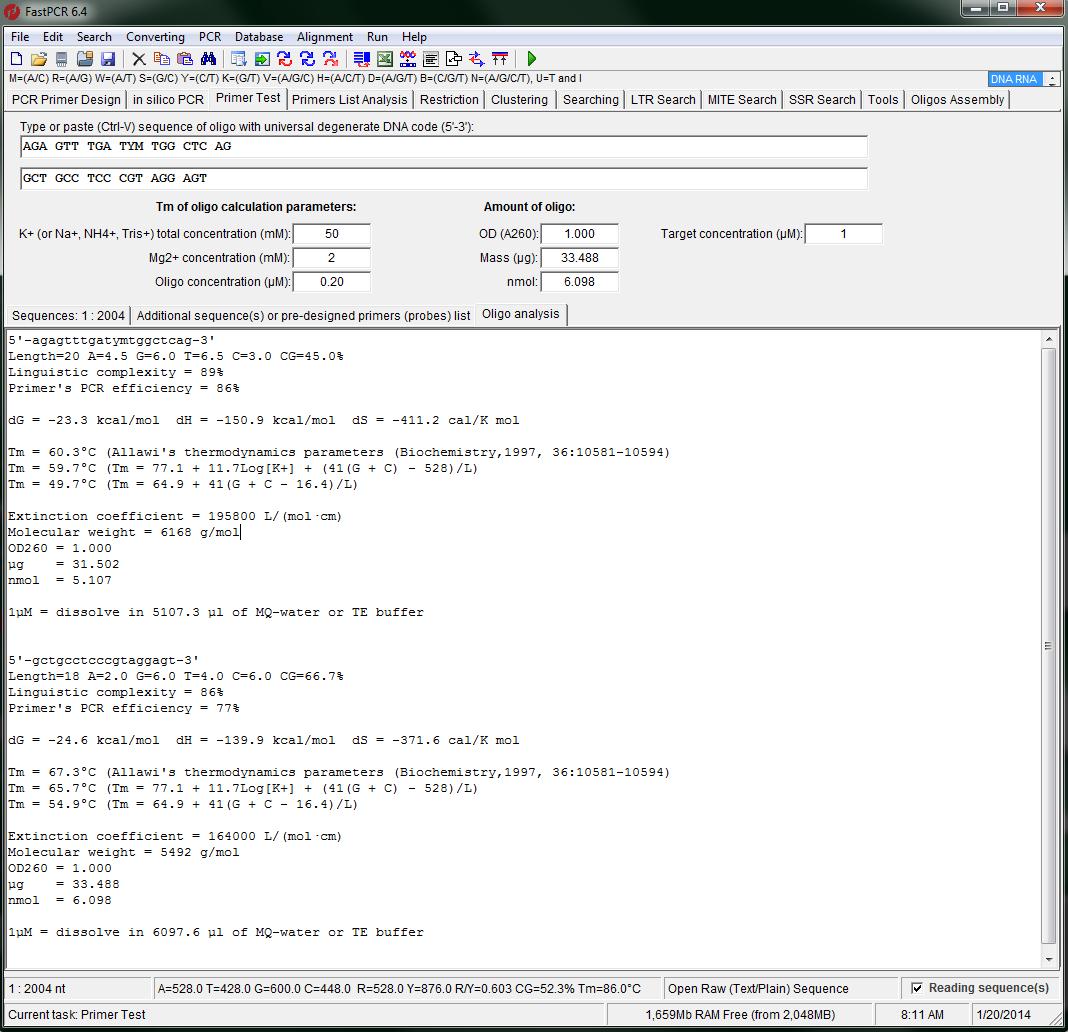


The primary information shows that the product length is 348bp, that the Tm of the first is 55C and the second is 60.7 C (marginal since ideally they would be less than 5C apart). You can also look to see what other organisms will amplify – such as the Pseudomonas. If we were trying to amplify a unique gene out of a mixture this would be bad, but since we want all bacteria, this is fine, it will work in our pure DNA and also in our oral stuff.

This Pick Primer software is not very sophisticated, so you can use other programs to determine how well the primers might work. For example FastPCR (which you can download) shows the following in a Primer Test (note that you are allowed to put in mixed bases, so you don’t have to do them one at a time).

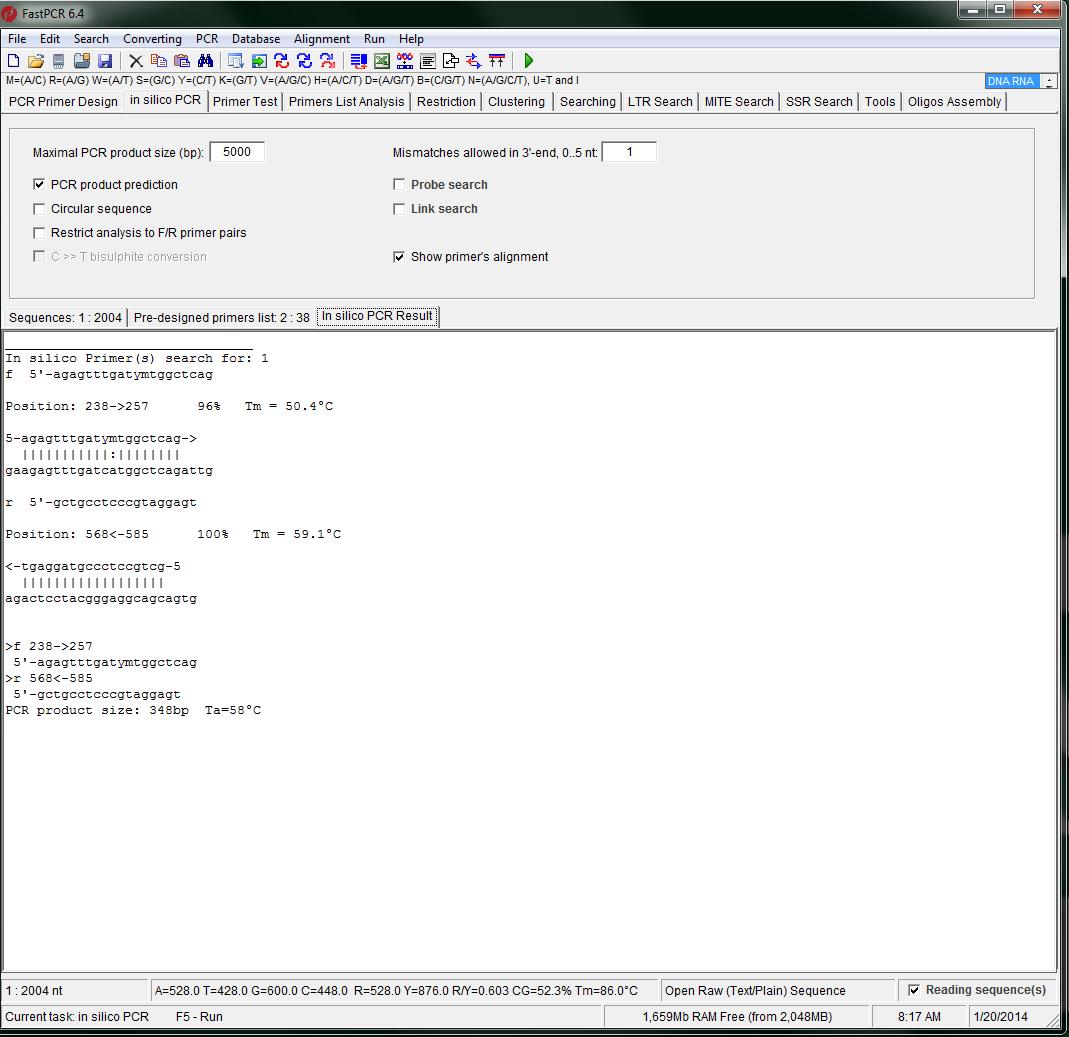
I put in the template sequence here along with PCR conditions.





You can see that the primers are still about 7C apart, and that depending on which equation you use you get somewhat different results for the Tm than ncbi – this takes buffer conditions into account. I like the fact that this gives you an estimate of the efficiency of the PCR primers.

I can also use FastPCR to do in silico PCR (you have to hit Run, or F5) and check the product:



This predicts a 348bp product and has a combined annealing temperature of 58C.

There are a lot of other programs you can run – many will give you slightly different values because they use somewhat different models or parameters in those models.

The next bit is my results using FastPCR and the primers I have listed above.

1. 27f\_YM with 536R\_BMC

The forward primer stays the same, the reverse primer shows a Tm of 66.7 and an efficiency of 60%. The product is 529bp with a Ta of 59C.

1. 27f\_YM with 1492R (these are our possibilities)

1492R TAC CTT GTT ACG ACT T

1492R\_Wiki AGA GTT TGG ATC MTG GCT CAG

U1492R GGT TAC CTT GTT ACG ACT T

1492R\_BMC\_11Y TAC CTT GTT AYG ACT T

The Forward primer has a Tm of 60.5 and an efficiency of 80%.

For 1492R, we have a Tm of 51 for the reverse primer and a product size of 1500bp. The reverse primer has an efficiency of 75%.

For 1492R\_Wiki there are a lot of mismatches – this is not a good reverse primer for our test gDNA.

For U1492R the Tm is 58C and the product size is 1503bp. The efficiency is 73%.

For 1492R\_BMC\_11Y the product is 1500 bp. The Tm of the reverse primer is 50C and the efficiency is 70%.

So the best-matched primers are 27f\_YM and U1492R. Since this makes a very long product, we will not pursue this set.

1. 519F\_BMC with 1492R

519F\_BMC CAG CMG CCG CGG TAA TAC

The forward primer has a Tm of 66.7 and an efficiency of 58%. It is not going to be a particularly good pair of primers. The product is 989bp, which is also pretty long – we will not pursue this. The best reverse primer would be U1492R again.

For other primers that I just ordered, you can see the spreadsheet for this information.

If you want to check the oligos for hairpin and dimer formation, the IDT tools are fine:

<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>

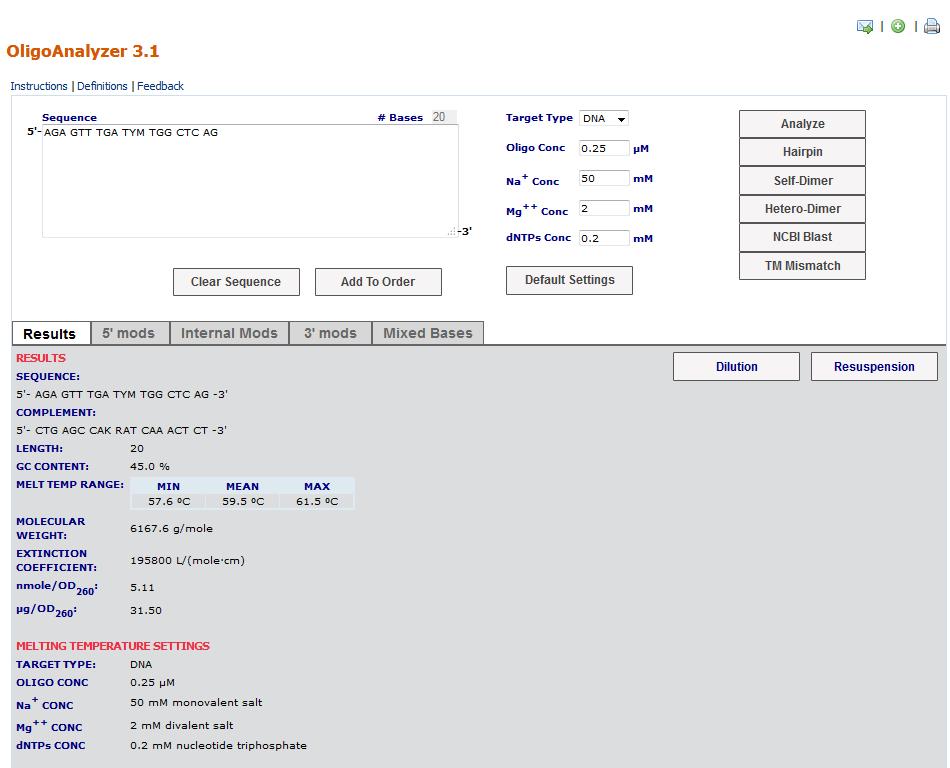
Here I just show the 27f\_YM and 338R\_V2\_BMC and 536R since they are the ones that appear to be useful (348 and 529bp products).

27f\_YM AGA GTT TGA TYM TGG CTC AG

338R\_v2\_BMC GCT GCC TCC CGT AGG AGT

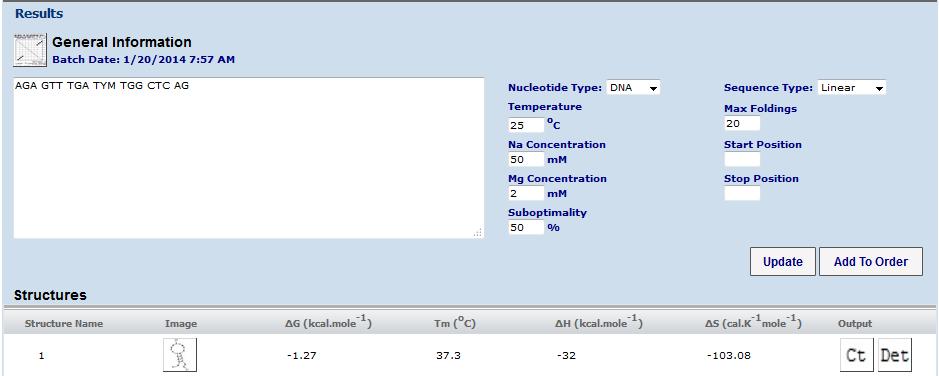
538R\_BMC GTA TTA CCG CGG CKG CTG

For 27f\_YM

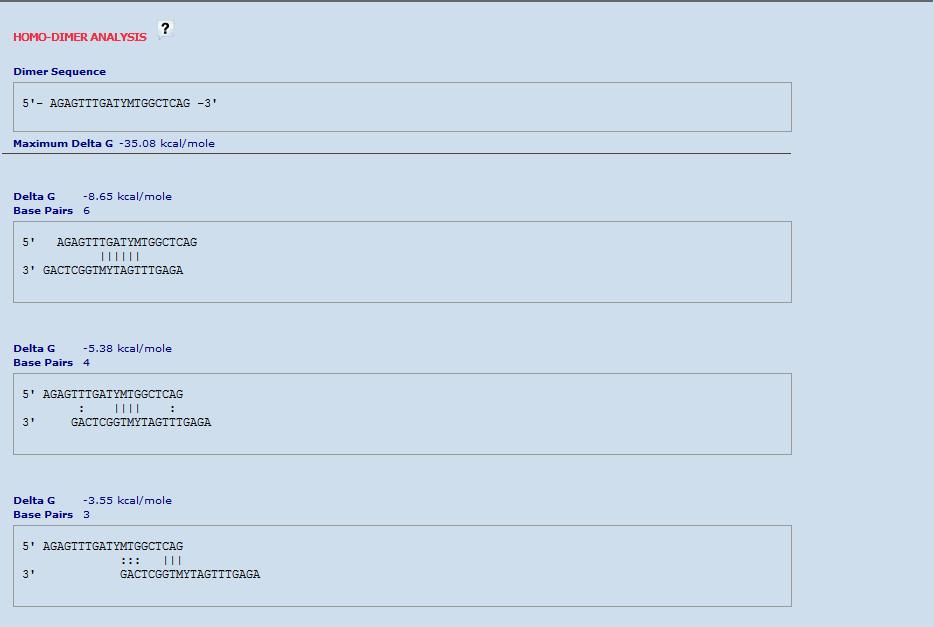


The mean Tm is consistent with the Tm FastPCR predicted.

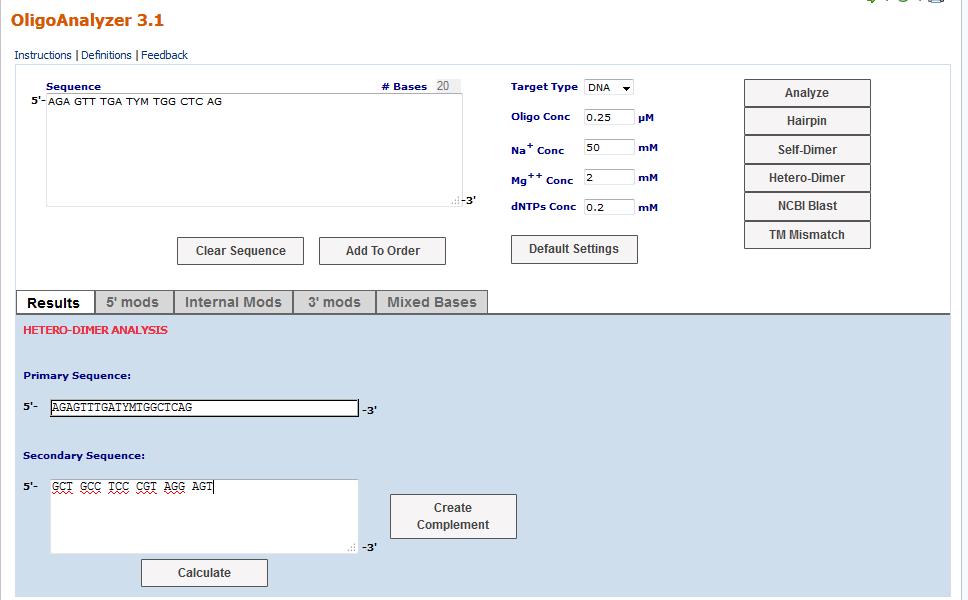
One hairpin is predicted with a Tm of 37C:



For self-dimer we get one with 6bp (which is normally over my limit)



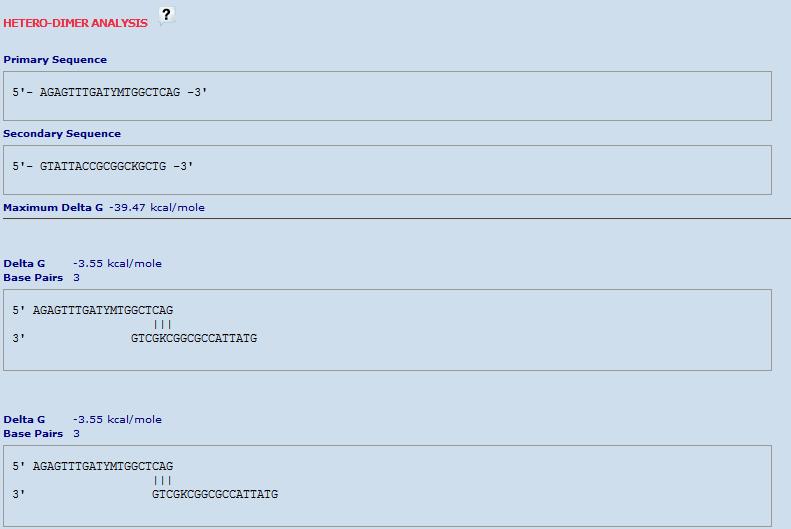
To check for heterodimers, you put in the reverse primer (338R in this case)





This part looks fine.

When I make a heterodimer with 538R I also don’t see a mjor problem, although with 3bp they do line up end to end, which can dispose towards making concatamers.



Here ends the tutorial. As you can see, it can take a lot of tools, but once you have the workflow figured out it will go fairly quickly (for example, once you have a target sequence from a genome, you are better off not using NCBI Primer BLAST on an organism, go to the Gene level, or just use a tool like FastPCR).