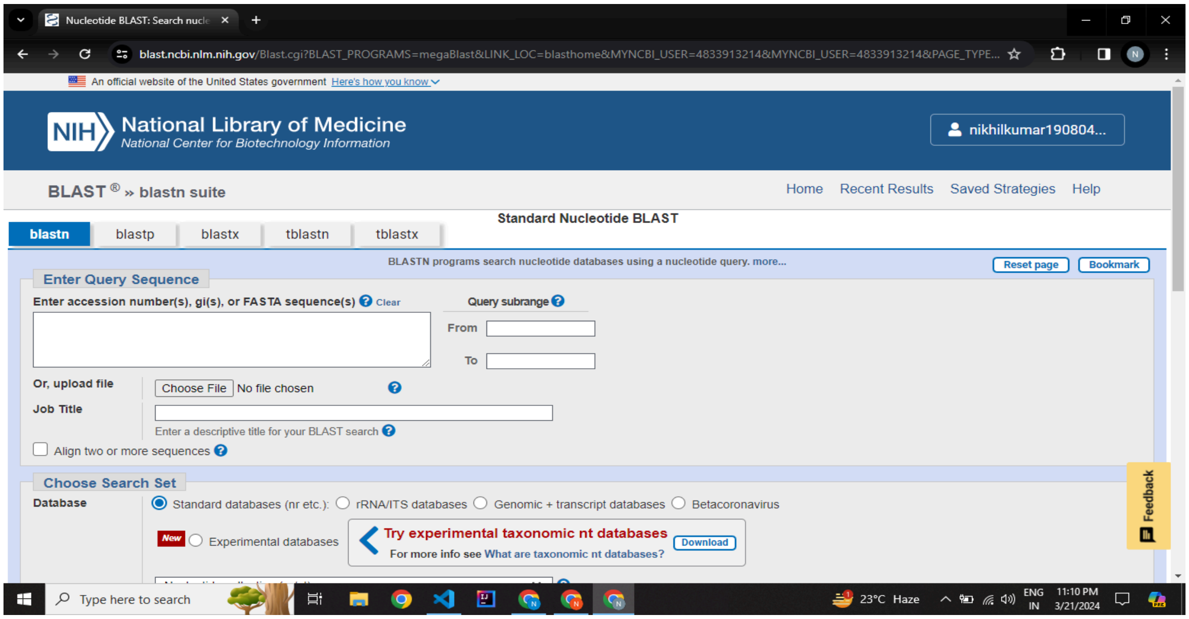
Using BLAST To Compare the PDC gene sequence across different Saccharomyces cerevisiae strains



Nucleotide BLAST: Search nucle

blast.ncbi.nlm.nih.gov/Blast.cgi?BLAST_PROGRAMS=megaBlast&LINK_LOC=blasthome&MYNCBI_USER=4833913214&MYNCBI_USER=4833913214&PAGE_TYPE...

An official website of the United States government Here's how you know

NIH National Library of Medicine
National Center for Biotechnology Information

nikhilkumar190804...

BLAST ® » blastn suite

Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. more...

Reset page Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

Query subrange
From
To

Or, upload file: Choose File No file chosen

Job Title: Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

Database: ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

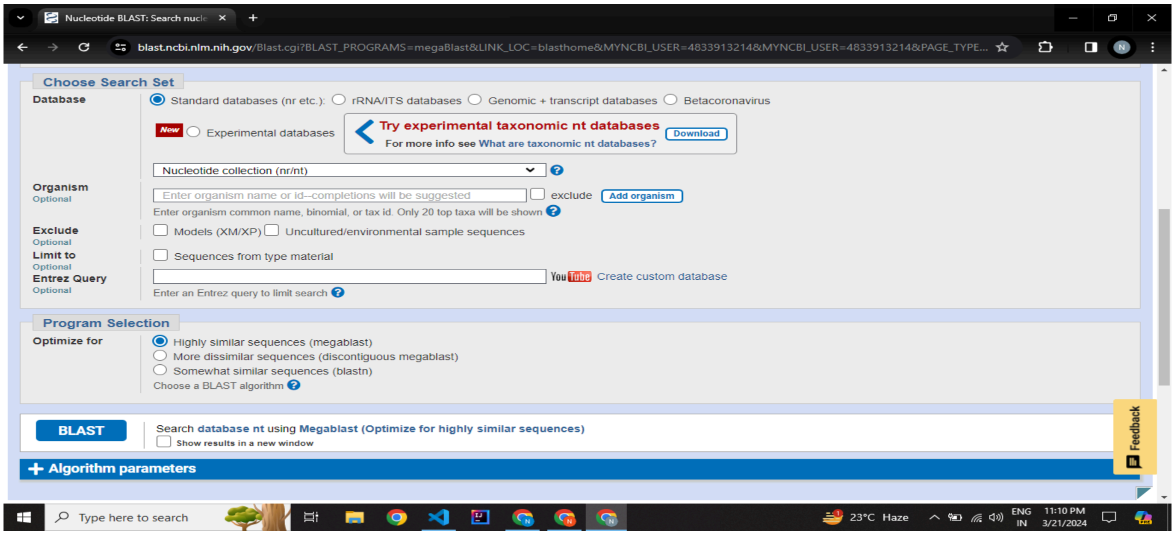
New ○ Experimental databases

Try experimental taxonomic nt databases Download
For more info see What are taxonomic nt databases?

Feedback

Type here to search 23°C Haze ENG IN 11:10 PM 3/21/2024

This is the standard view of the BLASTn. It is divided into three sections, namely: “Enter Query Sequence,” “Choose Search Set”, and” Program Selection.”
In the section “Enter Query Sequence,” we have to Enter the input query i.e.- the sequence for which we want to perform the BLAST. Here we can either paste the sequence directly or give the accession number of that sequence and provide a subrange for that sequence.



Nucleotide BLAST: Search nucle

blast.ncbi.nlm.nih.gov/Blast.cgi?BLAST_PROGRAMS=megaBlast&LINK_LOC=blasthome&MYNCBI_USER=4833913214&MYNCBI_USER=4833913214&PAGE_TYPE...

Choose Search Set

Database: ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

New ○ Experimental databases

Try experimental taxonomic nt databases Download
For more info see What are taxonomic nt databases?

Nucleotide collection (nr/nt)

Organism (Optional): Enter organism name or id--completions will be suggested □ exclude Add organism
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude (Optional): □ Models (XM/XP) □ Uncultured/environmental sample sequences

Limit to (Optional): □ Sequences from type material

Entrez Query (Optional): YouTube Create custom database
Enter an Entrez query to limit search

Program Selection

Optimize for: ● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)
Choose a BLAST algorithm

BLAST Search database nt using Megablast (Optimize for highly similar sequences)
□ Show results in a new window

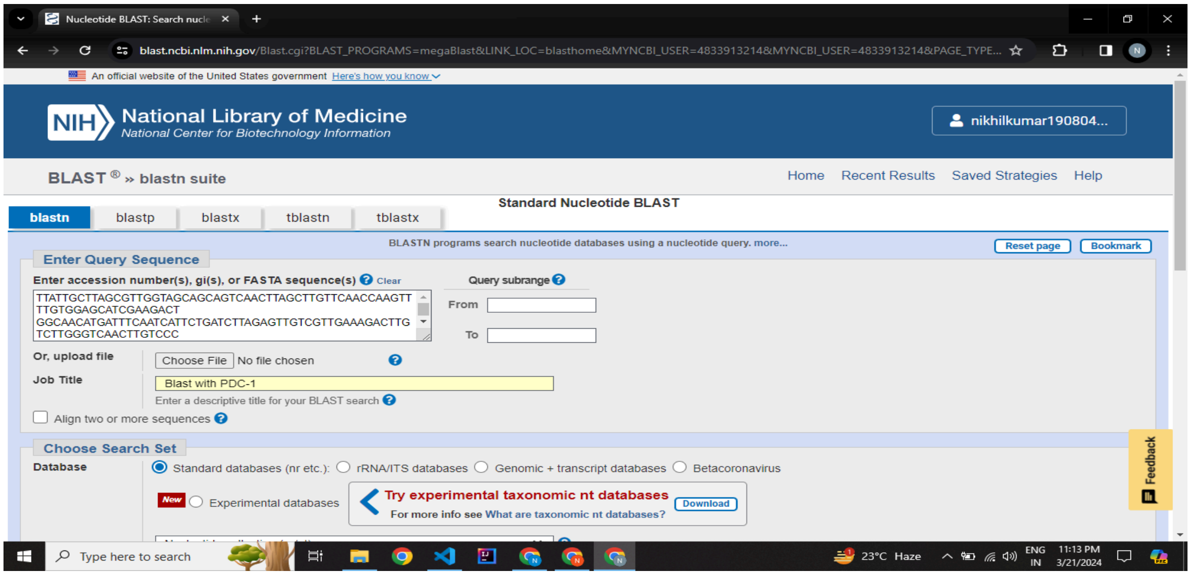
+ Algorithm parameters

Feedback

Type here to search 23°C Haze ENG IN 11:10 PM 3/21/2024

Now, in the section “Choose Search Set” we can select the database we want to use for BLAST. By default, it is selected as “Nucleotide” and standard database (non-redundant). Also here, we have the option to add an organism or to exclude an organism from the database. And we also have the option to filter our database using an Entrez Query.
And in the last section, “Program Selection” we have some optimization options to enhance our searches.
And lastly we can change some algorithm parameters according to our needs.

BLAST With PDC-1


Nucleotide BLAST: Search nucle
blast.ncbi.nlm.nih.gov/Blast.cgi?BLAST_PROGRAMS=megaBlast&LINK_LOC=blasthome&MYNCBI_USER=4833913214&MYNCBI_USER=4833913214&PAGE_TYPE...
An official website of the United States government Here's how you know
NIH National Library of Medicine
National Center for Biotechnology Information
nikhilkumar190804...
BLAST ® » blastn suite
Home
Recent Results
Saved Strategies
Help
Standard Nucleotide BLAST
blastn
blastp
blastx
tblastn
tblastx
BLASTN programs search nucleotide databases using a nucleotide query. more...
Reset page
Bookmark
Feedback
Type here to search
23°C Haze
ENG IN
11:13 PM 3/21/2024

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ? Clear

```
TTATTGCTTAGCGTTGGTAGCAGCAGTCAACTTAGCTTGTTCAACCAAGTT
TTGTGGAGCATCGAAGACT
GGCAACATGATTTCAATCATTCTGATCTTAGAGTTGTCGTTGAAAGACTTG
TCTTGGGTCAACTTGTCCC
```

Query subrange ? From To

Or, upload file: Choose File No file chosen

Job Title: Blast with PDC-1

Enter a descriptive title for your BLAST search

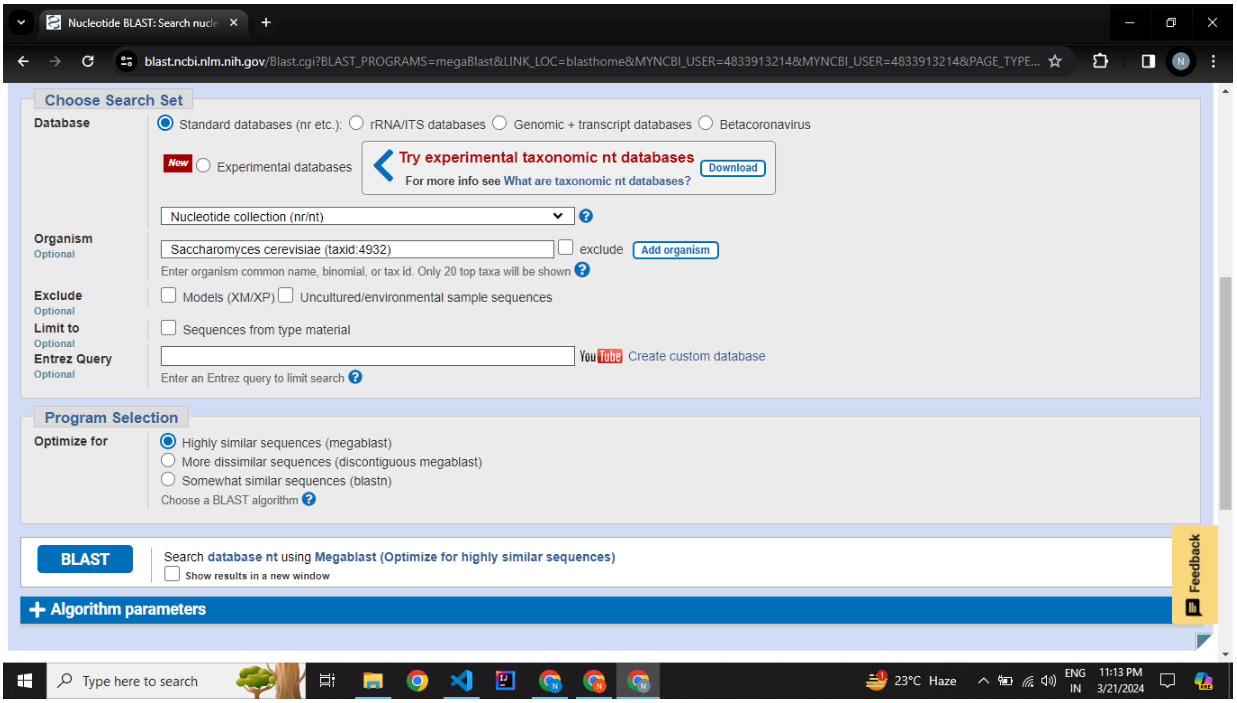
□ Align two or more sequences

Choose Search Set

Database: ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

New ○ Experimental databases — Try experimental taxonomic nt databases Download — For more info see What are taxonomic nt databases?

Now, here i am performing BLAST with input as the PDC-1 gene. I pasted the sequence of the PDC-1 gene, which I downloaded using Python, and gave a suitable Job Title to this.
The accession number for PDC-1 is NC_001144.5 and the sequence starts from 234081 to 232390. It is in reverse complement form.


Nucleotide BLAST: Search nucle
blast.ncbi.nlm.nih.gov/Blast.cgi?BLAST_PROGRAMS=megaBlast&LINK_LOC=blasthome&MYNCBI_USER=4833913214&MYNCBI_USER=4833913214&PAGE_TYPE...
Feedback
Type here to search
23°C Haze
ENG IN
11:13 PM 3/21/2024

Choose Search Set

Database: ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

New ○ Experimental databases — Try experimental taxonomic nt databases Download — For more info see What are taxonomic nt databases?

Nucleotide collection (nr/nt)

Organism (Optional): Saccharomyces cerevisiae (taxid:4932) □ exclude Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude (Optional): □ Models (XM/XP) □ Uncultured/environmental sample sequences

Limit to (Optional): □ Sequences from type material

Entrez Query (Optional): YouTube Create custom database

Enter an Entrez query to limit search

Program Selection

Optimize for: ● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

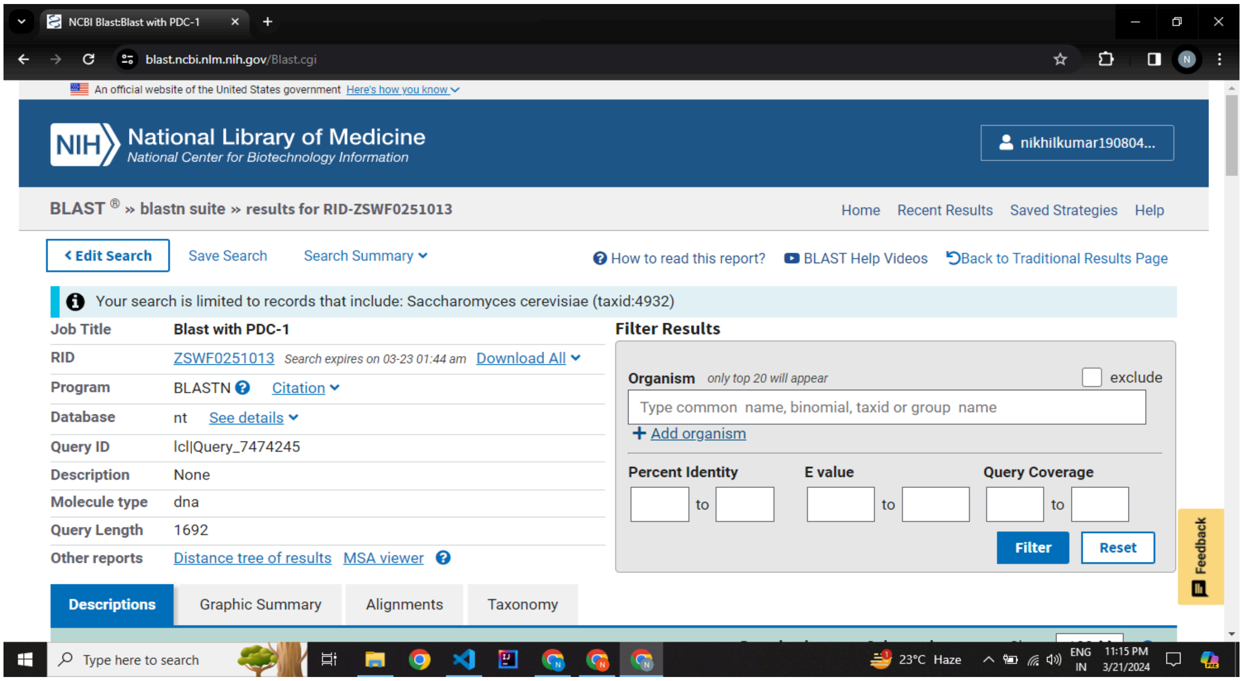
BLAST — Search database nt using Megablast (Optimize for highly similar sequences)

□ Show results in a new window

+ Algorithm parameters

Here I selected the database as Standard database, which is a by default option, and selected the Nucleotide collection (nr/nt) which again was by default. After that i added the Organism “Saccharomyces cerevisiae,” and I selected the top among them with taxid as 4932.
After that, in section “Program Selection” I kept the options by default, and in algorithm parameters I kept everything by default only.
And finally, i clicked on the button for BLAST.
It took some time to complete this request, and after some time, the result page was shown.

Result of BLAST



BLAST ® » blastn suite » results for RID-ZSWF0251013

Your search is limited to records that include: Saccharomyces cerevisiae (taxid:4932)

Job Title	Blast with PDC-1
RID	ZSWF0251013 Search expires on 03-23 01:44 am Download All
Program	BLASTN Citation
Database	nt See details
Query ID	lcl|Query_7474245
Description	None
Molecule type	dna
Query Length	1692
Other reports	Distance tree of results MSA viewer

Filter Results

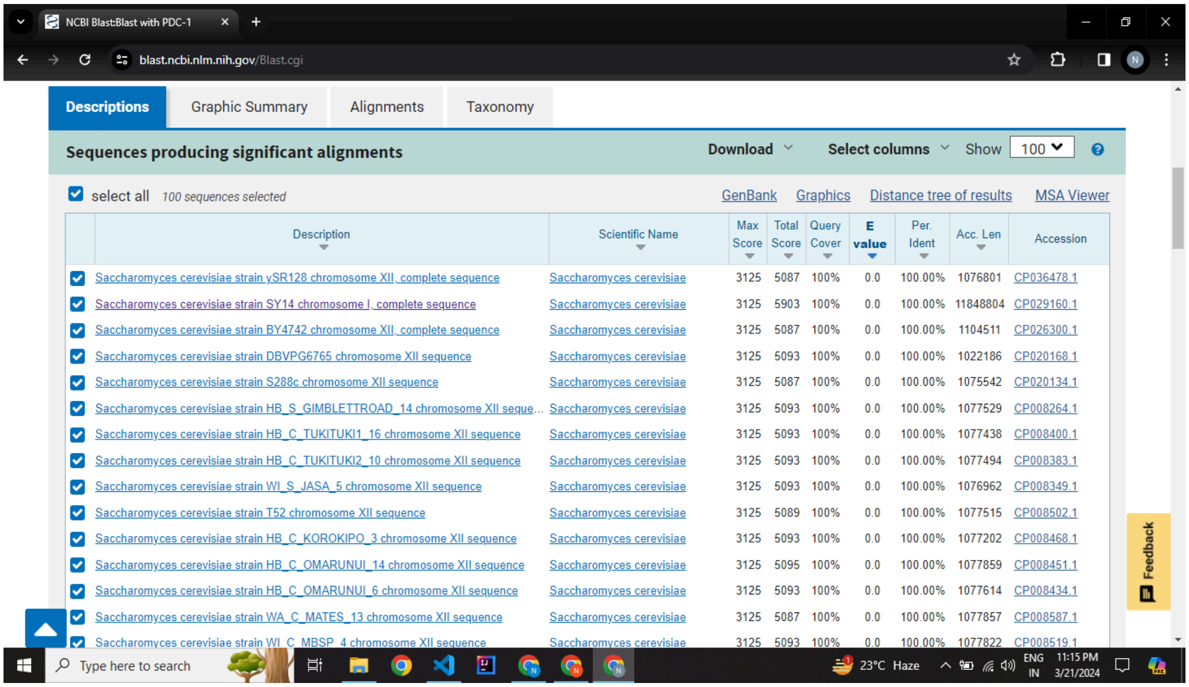
Organism *only top 20 will appear* □ exclude

Percent Identity ___ to ___ E value ___ to ___ Query Coverage ___ to ___

NCBI Blast:Blast with PDC-1
blast.ncbi.nlm.nih.gov/Blast.cgi
An official website of the United States government Here's how you know
NIH National Library of Medicine
National Center for Biotechnology Information
nikhilkumar190804...
Home
Recent Results
Saved Strategies
Help
Edit Search
Save Search
Search Summary
How to read this report?
BLAST Help Videos
Back to Traditional Results Page
Type common name, binomial, taxid or group name
Add organism
Filter
Reset
Feedback
Descriptions
Graphic Summary
Alignments
Taxonomy
Type here to search
23°C Haze
ENG IN
11:15 PM
3/21/2024

This is the result page of the BLAST. On the left side, it provides the details related to the search like Job Title, a unique ID for every search, the BLAST program that was selected (blastn.blastp,tblastx etc.), the database which was selected, and the query length which was 1692 bases long for the PDC-1.

On the right side, it gives the option to filter out the results based on some parameters like percent identity and E value or by using an organism name.



Sequences producing significant alignments

☑ select all *100 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Saccharomyces cerevisiae strain ySR128 chromosome XII, complete sequence	Saccharomyces cerevisiae	3125	5087	100%	0.0	100.00%	1076801	CP036478.1
☑	Saccharomyces cerevisiae strain SY14 chromosome I, complete sequence	Saccharomyces cerevisiae	3125	5903	100%	0.0	100.00%	11848804	CP029160.1
☑	Saccharomyces cerevisiae strain BY4742 chromosome XII, complete sequence	Saccharomyces cerevisiae	3125	5087	100%	0.0	100.00%	1104511	CP026300.1
☑	Saccharomyces cerevisiae strain DBVPG6765 chromosome XII sequence	Saccharomyces cerevisiae	3125	5093	100%	0.0	100.00%	1022186	CP020168.1
☑	Saccharomyces cerevisiae strain S288c chromosome XII sequence	Saccharomyces cerevisiae	3125	5087	100%	0.0	100.00%	1075542	CP020134.1
☑	Saccharomyces cerevisiae strain HB_S_GIMBLETTROAD_14 chromosome XII seque...	Saccharomyces cerevisiae	3125	5093	100%	0.0	100.00%	1077529	CP008264.1
☑	Saccharomyces cerevisiae strain HB_C_TUKITUKI1_16 chromosome XII sequence	Saccharomyces cerevisiae	3125	5093	100%	0.0	100.00%	1077438	CP008400.1
☑	Saccharomyces cerevisiae strain HB_C_TUKITUKI2_10 chromosome XII sequence	Saccharomyces cerevisiae	3125	5093	100%	0.0	100.00%	1077494	CP008383.1
☑	Saccharomyces cerevisiae strain WI_S_JASA_5 chromosome XII sequence	Saccharomyces cerevisiae	3125	5093	100%	0.0	100.00%	1076962	CP008349.1
☑	Saccharomyces cerevisiae strain T52 chromosome XII sequence	Saccharomyces cerevisiae	3125	5089	100%	0.0	100.00%	1077515	CP008502.1
☑	Saccharomyces cerevisiae strain HB_C_KOROKIPO_3 chromosome XII sequence	Saccharomyces cerevisiae	3125	5093	100%	0.0	100.00%	1077202	CP008468.1
☑	Saccharomyces cerevisiae strain HB_C_OMARUNUI_14 chromosome XII sequence	Saccharomyces cerevisiae	3125	5095	100%	0.0	100.00%	1077859	CP008451.1
☑	Saccharomyces cerevisiae strain HB_C_OMARUNUI_6 chromosome XII sequence	Saccharomyces cerevisiae	3125	5093	100%	0.0	100.00%	1077614	CP008434.1
☑	Saccharomyces cerevisiae strain WA_C_MATES_13 chromosome XII sequence	Saccharomyces cerevisiae	3125	5087	100%	0.0	100.00%	1077857	CP008587.1
☑	Saccharomyces cerevisiae strain WI_C_MBSP_4 chromosome XII sequence	Saccharomyces cerevisiae	3125	5093	100%	0.0	100.00%	1077822	CP008519.1

Now, here we can clearly see the BLAST of PDC-1 against various Saccharomyces cerevisiae strains. Along with this we also have their alignments with PDC-1 and with percent identity, E value, scores, etc.

There were more than 600 strains displayed on keeping the number of records to a maximum allowed value in NCBI BLAST. Hence, this shows the BLAST of Saccharomyces cerevisiae strains with the PDC-1 gene.

A similar approach can be used to perform the BLAST with every PDC gene (I performed this in my code).

Now for further SNP analysis, I selected some of the Saccharomyces cerevisiae strains and wrote a Python code to find the SNPs between the PDC gene and strain sequence.
I selected these two strains: Saccharomyces cerevisiae YJM996 and Saccharomyces cerevisiae YJM1199.
The script I wrote prints the position of SNP change along with the base being changed.