

# Viewing mutations in sequencing data



## GOALS

1. Implement online genetics analysis tools.
2. Evaluate potential mutations in whole-genome sequencing data.
3. Identify different types of mutations and their effects.
4. Form specific hypotheses on how mutations may contribute to altered fitness.

## OVERVIEW

You or other high school students performed evolution experiments to select for yeast mutants that are more resistant to a specific stressor. These mutants will possess a few changes to their genetic code that may contribute to this resistance. We isolated yeast from each experiment and sequenced their genomes to identify their mutations.

You will receive the genome sequences of the ancestral yeast strain and an evolved yeast strain, and a list of mutations that a computer program has identified as differences between the ancestor and evolved yeast. Using an online genome viewer, you will first evaluate the mutations called by the program to determine if they are new changes in your evolved strain. Then you will identify the gene and type of mutation, and research its function in yeast. From this information, you will form hypotheses about how this mutation may contribute to the resistance of your evolved strain.

## MATERIALS AND EQUIPMENT

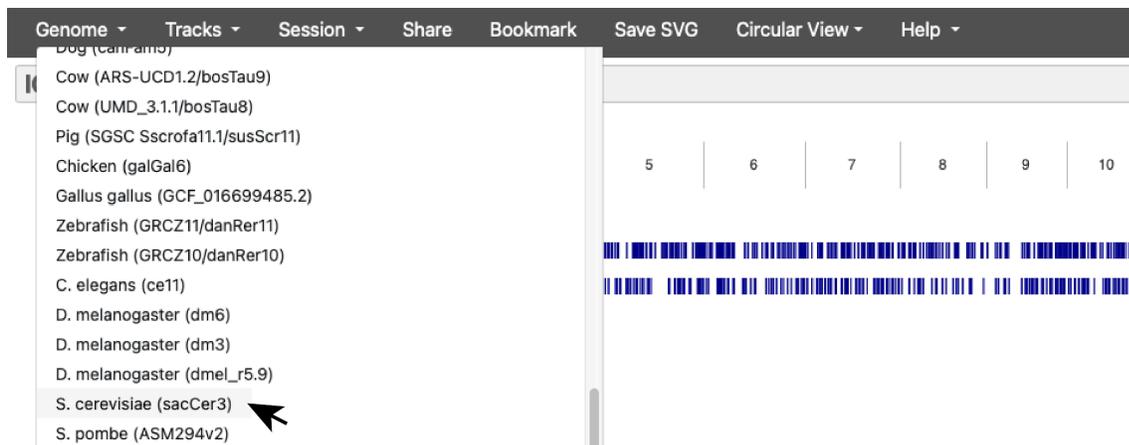
- Computer and internet access for each participant
- Example files:
  - Ancestor strain sequence files:  
example\_ancestor\_YMD4612\_comb\_R1R2.RG.MD.realign.sort.bam and .bam.bai
  - Evolved strain sequence files, from yeast evolved to have high caffeine tolerance:  
example\_evolved\_YMD4700\_comb\_R1R2.RG.MD.realign.sort.bam and .bam.bai
  - Mutations listed by computer program in Variant Call Format (VCF) file:  
example\_evolved\_YMD4700\_called\_mutations.txt
- Your own sequencing files:
  - Ancestor strain .bam and .bam.bai
  - Evolved strain .bam and .bam.bai
  - Mutations in your evolved strain listed by computer program .txt

## TASK 1: Evaluate potential mutations

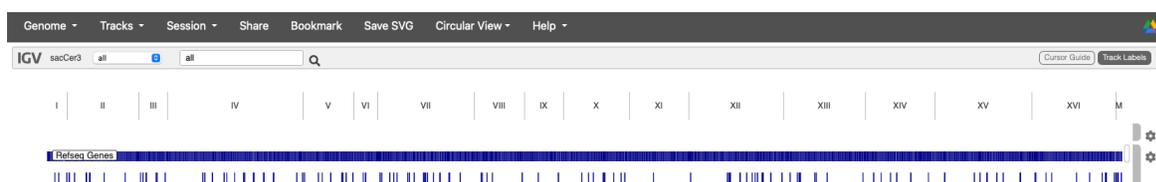
There are over 12 million DNA bases in the yeast genome! When we want to understand how the DNA of one yeast strain differs from another, we use a computer program to identify the sites where the two genomes differ so that it can analyze all 12 million sites. However, we want to confirm each of these mutations found by the computer, because sometimes noise in the sequencing data leads to the computer identifying potential sites that are not actual mutations. The error rate of the sequencing is very low at 0.6%, but with a genome with 12 million base pairs that has been sequenced to have about 50 reads at each site, that is still about 3.6 million errors! Also, some regions with a lot of repeats (e.g. hundreds of “AT” in a row) make assembling the sequence computationally challenging, or having only a few sequencing reads in the ancestor strain at a site lead to it being identified as “new” mutation. This is why your first task is to look at each potential mutated site in the evolved strain and compare it to the ancestor.

We will walk through some of the mutations in the example file, and then you can look at the sequencing data from your own evolved yeast! We will do this using a web app called the Integrative Genome Viewer (IGV).

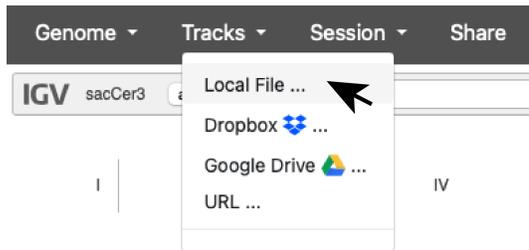
1. Make sure that you have the .bam and .bai files downloaded for both the evolved and ancestor strain. We will start with the example files.
2. Open the IGV web app <https://igv.org/app/>
3. Navigate to the Genome drop-down menu in the upper left corner and scroll down to select “S. cerevisiae (sacCer3)”. This is the reference genome for baker’s yeast.



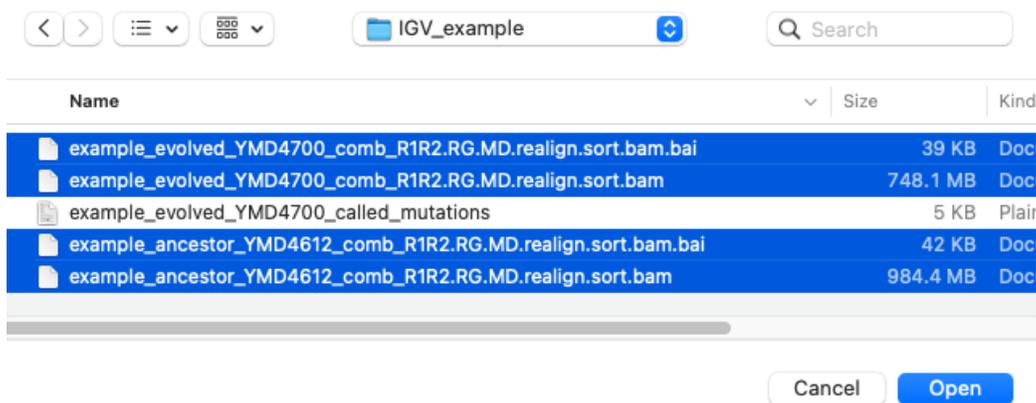
4. You can now see a diagram of all 16 yeast chromosomes and the mitochondrial DNA at the top of the window. It is labeled “Refseq Genes” because every blue bar is one gene. You will be able to see them more clearly when zoomed in after you upload data.



- Upload data by navigating to the Tracks drop-down menu and select “Local File ...”



- Select the .bam and .bai files for the evolved and ancestor strains, then click “Open.” *If you only upload the .bam you will receive an error message and need to upload the .bai at the same time. If you only upload the .bam and .bai file for one strain, you can repeat steps 5-6 to add another strain - it will not replace what you have already uploaded.*



- Now your files have been loaded, but there’s a lot of data so the viewer can’t show it all at once! You can see the file names, so for this example the upper file is the evolved strain and the lower is the ancestor.



8. To look at specific mutations, we need to see which were identified by the computer program. Open the Variant Call Format (VCF) file in a spreadsheet viewer like Excel, Numbers, or Google Sheets. The example file is `example_evolved_YMD4700_called_mutations.txt`
  - a. The first ~50 rows are from the program - you can ignore them!
  - b. The row that begins with `#CHROM` is the header. Each row below it is a mutation identified by the computer.
  - c. We will use the information on location (`#CHROM` and `POS`) to look at mutations in the viewer, and information about mutation type (`ANNOTATION`) and what it affects (`REGION` and `PROTEIN`) to form hypotheses about how it could contribute to resistance in later sections.

ignore {  
header →  
each row is a potential mutation {

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	RBcaff_Spinl	ANNOTATIO	REGION	PROTEIN
chrVII	471392	.	C	T	225	PASS	DP=89;VDB=	GT:PL	1/1:255,208,	coding-nonsy	YGL013C	E303K
chrIV	1167808	.	GAAAAA	GAAAA	135	PASS	INDEL;IDV=7	GT:PL	1/1:162,72,0	coding-nonsy	YDR346C	F284indel
chrVIII	526453	.	T	C	112	PASS	DP=143;VDB=	GT:PL	0/1:146,0,25	coding-synor	YHR211W	V354V
chrXIII	32149	.	C	G	225	PASS	DP=100;VDB=	GT:PL	1/1:255,241,	5'-upstream	YML118W	NA
chrXIII	699232	.	A	T	225	PASS	DP=111;VDB=	GT:PL	1/1:255,255,	coding-nonsy	YMR216C	L603H
chrXIV	98305	.	T	C	97	PASS	DP=140;VDB=	GT:PL	0/1:130,0,25	coding-synor	YNL284C-B	Q1309Q

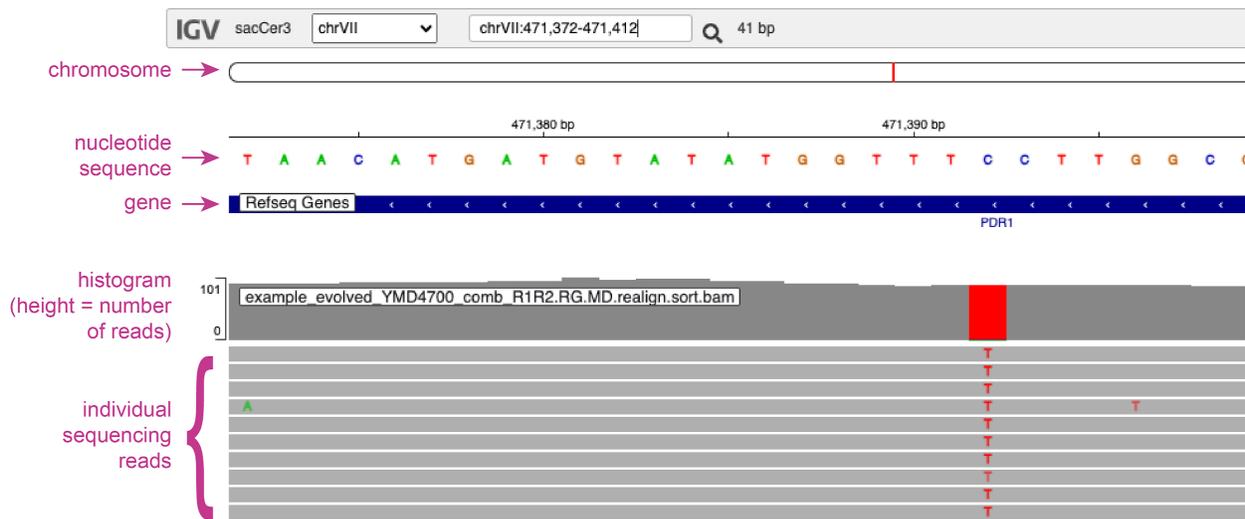
9. To look for a mutation in the IGV app, copy the entries for `#CHROM` and `POS` from one row in the VCF spreadsheet.
  - a. Also take a look at the `REF` and `ALT` columns for an idea of what to expect. `REF` is the reference sequence, and `ALT` is the alternate - what is seen in your evolved yeast. Here, `REF` is `GAAAAA` and `ALT` is `GAAAA`, so we expect to see a deletion of one `A` in the evolved strain.

#CHROM	POS	ID	REF	ALT
chrVII	471392	.	C	T
chrIV	1167808	.	GAAAAA	GAAAA
chrVIII	526453	.	T	C
chrXIII	32149	.	C	G

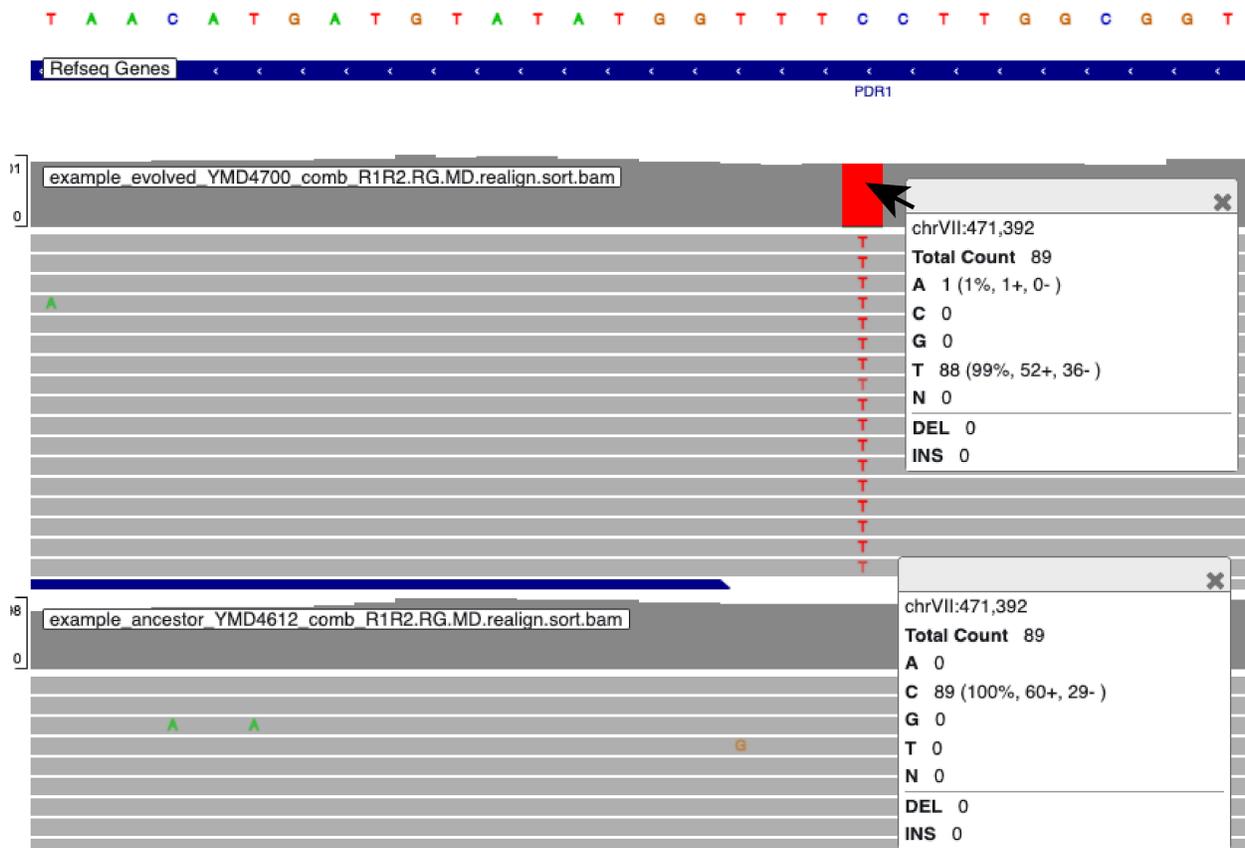
10. Paste them into the IGV search bar and type a colon between them so the format is `CHROM:POS`, e.g. `chrVII 471392` becomes `chrVII:471392`. Then hit `Enter` or click the search button (magnifying glass).



11. Now IGV shows the position you searched for in the middle of the window. In the example, the evolved strain is on top and the ancestor on the bottom.
- The chromosome position, DNA sequence and gene are shown at the top.
  - Each sequence track is topped by a histogram. Each bin represents one nucleotide. When the majority of the sequencing reads at a site are the same as the reference sacCer3 genome, the bin is gray. If many differ from the reference, it is colored by each different nucleotide at the site.
  - The lines below each histogram each represent one read from the sequencer. This is your raw data, which you can now see! Let's look at it more closely.



12. We want to know if there is a mutation at the site, and what it is. The colored bar of the histogram shows that there is a difference between our evolved strain and the reference, but we want to know if there is a difference from the ancestor strain!
- When we look at the ancestor strain's histogram below, we don't see the red bin representing the change to T. If we click on that bin, it shows the summary of all of the reads. In the ancestor, there are 89 reads ("Total Count"), and all 89 are C, which is the reference.
  - Now click on the red bin from the evolved strain at that site. It also has 89 reads, and 88 are T and 1 is A. The A is probably a sequencing error, since we know those happen at a low frequency, but the majority are certainly T.



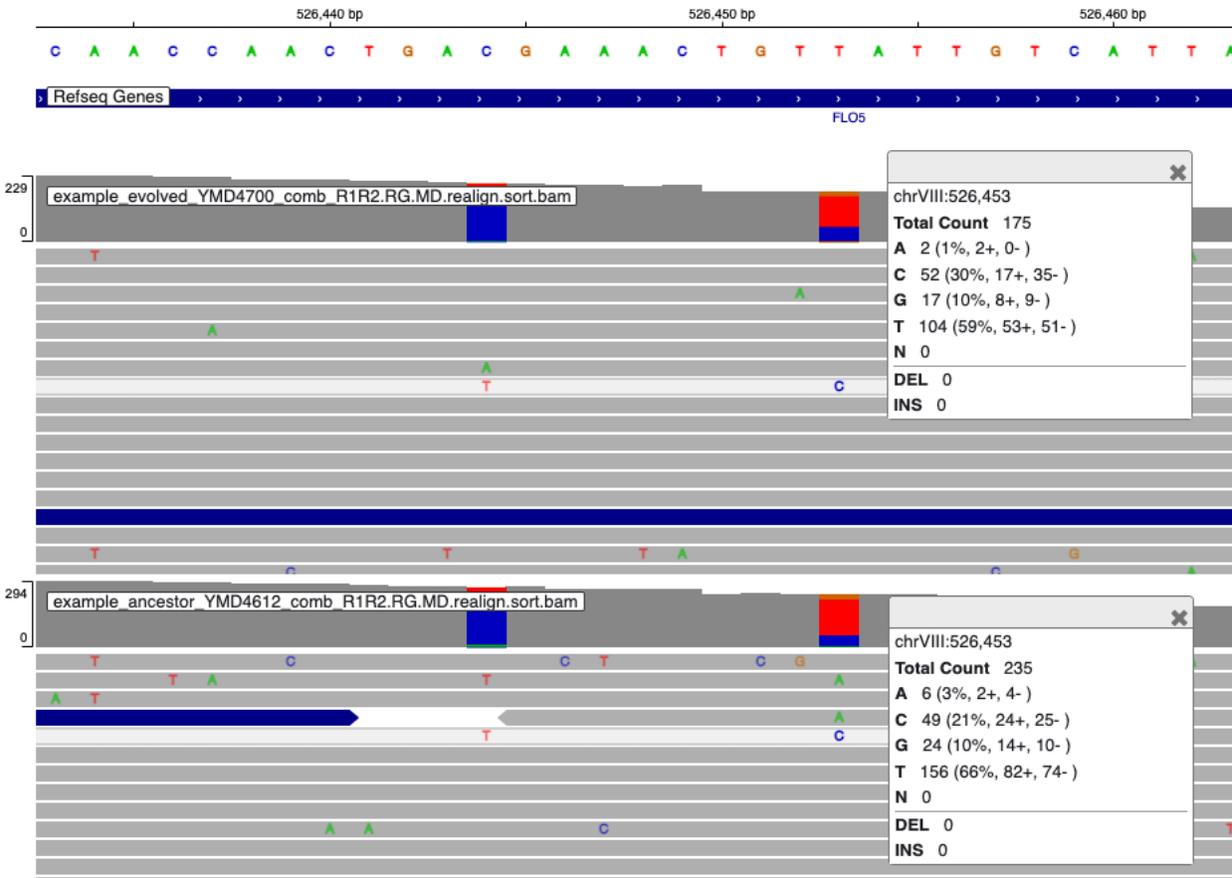
- Since the ancestor strain has a C at this position and the evolved has a T, we confirmed that this is a mutation where the reference is C and the alternate is T, or C>T. This is what was written in your VCF mutation file, so we confirmed this mutation! Make a note of this - you can do this directly in the VCF file by adding a column for your own notes.

```
##bcftools_filterVersion=1.9+htslib-1.9
##bcftools_filterCommand=filter -O v -o RBcaff_Spink2_S19_samtools_filtered.vcf -i 'MQ>30 & QUAL>75 & DP>10 & (DP4[2]+DP4[3])>4 & (DP4[2]+DP4[3])/DP>0.3 & (DP4[0]+DP4[2])/((DP4[0]+DP4[1]+DP4[2]+DP4[3]))'
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT RBcaff_Spinl ANNOTATIO REGION PROTEIN NOTES
chrVII 471392 . C T 225 PASS DP=89;VDB= GT:PL 1/1:255,208, coding-nonsy YGL013C E303K confirmed this mutation in IGV
chrIV 1167808 . GAAAA GAAAA 135 PASS INDEL;IDV=7 GT:PL 1/1:162,72,0 coding-nonsy YDR346C F284indel
chrVIII 526453 . T C 112 PASS DP=143;VDR:GT:PL 0/1:146,0,75 coding-nonsy YHR211W V354V
```



14. Let's look at one more example together, chrVIII:526453. The VCF mutation file says we should expect to see a T>C mutation.

- a. This time the mutation shows a mix of different nucleotides, and also appears in the ancestor strain. That's enough to tell us that it isn't a mutation in the evolved strain, since it was already in the ancestor. *This can happen for many reasons, but is often because some regions of the genome have almost identical sequence so those reads can get mapped to multiple places, resulting in a mix of nucleotides at the positions that differ.*



15. Repeat steps 9-12 to look at the remaining three mutations in the example file, and determine if they are differences between the evolved strain and the ancestor. Conclusions are provided on the next page to check your understanding.

16. Remaining mutations in evolved yeast YMD4700
- chrXIII:32149 is a C>G point mutation. It is a new mutation in the evolved strain since 99/100 reads show a G, and in the ancestor there is only one read with a G and the other 114 reads have the reference nucleotide, C.
  - chrXIII:699232 is an A>T point mutation. It is a new mutation in the evolved strain since all 111 reads have a T, and in the ancestor there all 82 reads have the reference nucleotide, A.
  - chrXIV:98305 is not a new mutation in the evolved strain because a mix of C and T are present in both the evolved and ancestor strains.
17. Now you should be more comfortable with viewing mutations in IGV, and are ready to investigate the mutations in your own samples. Repeat steps 1-12 to investigate each mutation listed in your evolved strain's VCF mutation file. Use the Resource 1 summary sheet on the next page to help you; it also shows one type of mutation we did not explore in the example - insertions. At the end, you should have made notes about which mutations you confirmed.

### **TASK 1 QUESTIONS**

- How many mutations did you identify in your evolved yeast strain's DNA?
- What types of mutations did you identify (point mutations, deletions, insertions)?
- How many mutations were predicted by the computer in the VCF mutation file, which after viewing in IGV you concluded are not differences between your evolved strain and the ancestor? What evidence do you have to support your conclusion?
- If you are unsure about any of the mutations in your evolved strain, talk to your instructor! You need to have your final list of mutations for the next tasks.

## RESOURCE 1: Identifying mutation types in IGV

Point Mutations	Deletions
<p>One nucleotide is <b>changed</b> in the evolved strain and not in the ancestor. Represented by a color change in the histogram.</p>	<p>One or more nucleotides are <b>missing</b> in the evolved strain and not in the ancestor. Represented by a gap spanned by a black line in the read.</p>
<p><b>Evolved</b></p> <p><b>Ancestor</b></p>	<p><b>Evolved</b></p> <p><b>Ancestor</b></p>
Insertions	Not a new mutation
<p>One or more nucleotides are <b>added</b> in the evolved strain and not in the ancestor. Represented by a purple bracket in each read. Click on a bracket to see which nucleotides are inserted.</p>	<p>Point mutations, deletions, or insertions are present in a large number of reads for both the evolved and ancestor strains.</p>
<p><b>Evolved</b></p> <p><b>Ancestor</b></p>	<p><b>Evolved</b></p> <p><b>Ancestor</b></p>

## TASK 2: Explore mutation types

In order to form hypotheses about how mutations could be affecting the fitness of the yeast, we need to understand what type of mutations they are, and what gene they affect. The “ANNOTATION” column of the VCF mutation file lists one of the following types of consequences for each mutation:

1. **coding-synonymous mutations** are DNA changes that do not change an amino acid. You will not be able to see these in the protein alignment. They are sometimes referred to as silent mutations because they do not impact a gene’s protein.
2. **coding-nonsynonymous mutations** change one or more amino acids in a gene’s protein. They may change the function of that protein. There are different types of nonsynonymous mutations:
  - a. **Missense mutations** change one amino acid to another amino acid.
  - b. **Nonsense mutations** result in an early stop in the gene. You will notice the \* indicating which amino acid’s codon has been changed to a stop codon. None of the amino acids that come after this stop will be added to the protein, so the resulting protein will be shorter than the original version.
  - c. **Frameshift mutations** are when one or a few nucleotides are inserted or deleted - noted as “indel.” When the number of nucleotides inserted or deleted is not a multiple of 3, this results in a change in the reading frame of a protein, shifting how the code of the gene is read. Nearly all the amino acids after the mutation will be changed, and this frequently leads to early stops.
3. **5’-upstream mutations** occur before the coding sequence of the gene starts, such as in the promoter. They can affect expression of the gene.
4. **intergenic mutations** are between genes. They could affect gene regulation, but many will have no effect.
5. **ARS** stands for autonomously replicating sequence, which is a part of each chromosome where DNA replication is initiated.

## TASK 2 QUESTIONS

1. Which of the mutation types above did you identify in your evolved yeast strain in Task 1? For coding-nonsynonymous mutations, are they missense, nonsense, or frameshift? Look at the PROTEIN column of your VCF mutation file to see if the amino acid is changed to another amino acid, indel, or stop codon \*.
2. Which of your identified mutations do you think will have the largest impact on the gene product? Explain your reasoning based on the mutation type descriptions above, or the properties of amino acids listed in Resource 2.

## RESOURCE 2

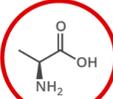
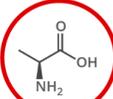
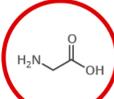
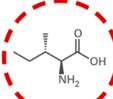
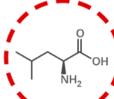
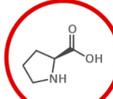
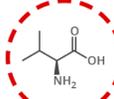
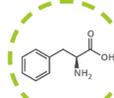
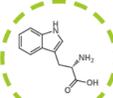
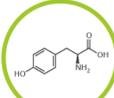
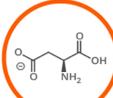
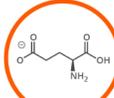
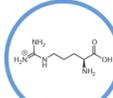
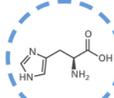
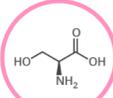
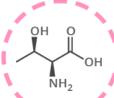
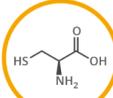
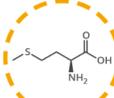
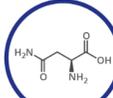
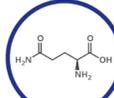
View full size resource:

<https://i0.wp.com/www.compoundchem.com/wp-content/uploads/2014/09/20-Common-Amino-Acids-v3.png>

# A GUIDE TO THE TWENTY COMMON AMINO ACIDS

AMINO ACIDS ARE THE BUILDING BLOCKS OF PROTEINS IN LIVING ORGANISMS. THERE ARE OVER 500 AMINO ACIDS FOUND IN NATURE - HOWEVER, THE HUMAN GENETIC CODE ONLY DIRECTLY ENCODES 20. 'ESSENTIAL' AMINO ACIDS MUST BE OBTAINED FROM THE DIET, WHILST NON-ESSENTIAL AMINO ACIDS CAN BE SYNTHESISED IN THE BODY.

**Chart Key:** ● ALIPHATIC ● AROMATIC ● ACIDIC ● BASIC ● HYDROXYLIC ● SULFUR-CONTAINING ● AMIDIC ○ NON-ESSENTIAL ○ ESSENTIAL

 <p><b>Chemical Structure</b> single letter code <b>A</b> three letter code DNA codons</p>	 <p><b>ALANINE (A)</b> <i>Ala</i> GCT, GCC, GCA, GCG</p>	 <p><b>GLYCINE (G)</b> <i>Gly</i> GGT, GGC, GGA, GGG</p>	 <p><b>ISOLEUCINE (I)</b> <i>Ile</i> ATT, ATC, ATA</p>	 <p><b>LEUCINE (L)</b> <i>Leu</i> CTT, CTC, CTA, CTG, TTA, TTG</p>	 <p><b>PROLINE (P)</b> <i>Pro</i> CCT, CCC, CCG, CCG</p>	 <p><b>VALINE (V)</b> <i>Val</i> GTT, GTC, GTA, GTG</p>
 <p><b>PHENYLALANINE (F)</b> <i>Phe</i> TTT, TTC</p>	 <p><b>TRYPTOPHAN (W)</b> <i>Trp</i> TGG</p>	 <p><b>TYROSINE (Y)</b> <i>Tyr</i> TAT, TAC</p>	 <p><b>ASPARTIC ACID (D)</b> <i>Asp</i> GAT, GAC</p>	 <p><b>GLUTAMIC ACID (E)</b> <i>Glu</i> GAA, GAG</p>	 <p><b>ARGININE (R)</b> <i>Arg</i> CGT, CGC, CGA, CCG, AGA, AGG</p>	 <p><b>HISTIDINE (H)</b> <i>His</i> CAT, CAC</p>
 <p><b>LYSINE (K)</b> <i>Lys</i> AAA, AAG</p>	 <p><b>SERINE (S)</b> <i>Ser</i> TCT, TCC, TCA, TCG, AGT, AGC</p>	 <p><b>THREONINE (T)</b> <i>Thr</i> ACT, ACC, ACA, ACG</p>	 <p><b>CYSTEINE (C)</b> <i>Cys</i> TGT, TGC</p>	 <p><b>METHIONINE (M)</b> <i>Met</i> ATG</p>	 <p><b>ASPARAGINE (N)</b> <i>Asn</i> AAT, AAC</p>	 <p><b>GLUTAMINE (Q)</b> <i>Gln</i> CAA, CAG</p>

**Note:** This chart only shows those amino acids for which the human genetic code directly codes for. Selenocysteine is often referred to as the 21st amino acid, but is encoded in a special manner. In some cases, distinguishing between asparagine/aspartic acid and glutamine/glutamic acid is difficult. In these cases, the codes asx (B) and glx (Z) are respectively used.

© COMPOUND INTEREST 2014 - WWW.COMPOUNDCHEM.COM | Twitter: @compoundchem | Facebook: www.facebook.com/compoundchem  
Shared under a Creative Commons Attribution-NonCommercial-NoDerivatives licence.



### TASK 3: Learn about the genes affected by your mutations

We will now move onto the *Saccharomyces* Genome Database (SGD) website to learn more about the genes affected by the mutations in your evolved yeast strain. SGD contains information from thousands of published studies on yeast, which has been curated to make it easier to find information about a specific gene.

<https://www.yeastgenome.org/>

For this example, I'm going to show you the page for the first mutation we confirmed in the caffeine-tolerant evolved yeast strain YMD4700 used in the Task 1 example.

1. Copy the REGION from the VCF mutation file and paste it into the SGD search bar

2. It takes us to a page for *PDR1*, the common name of the gene encoded by the locus YGL013C. This page contains a ton of info organized into sections that are listed in light blue along the left side of the page. The sections I like to use are *Locus Overview* and *Summary Paragraph*. The top menu also has a helpful tab, *Phenotype*. The others are great, but for now you can ignore them. We will walk through these three sections.

PDR1 / YGL013C Overview	
Standard Name:	PDR1 <sup>1</sup>
Systematic Name:	YGL013C
SGD ID:	SGD:S000002981
Aliases:	AMY1 <sup>19</sup> , ANT1 <sup>20</sup> , BOR2 <sup>21</sup> , CYH3 <sup>22</sup> , NRA2, SMR2, TIL1, TPE1 <sup>23</sup> , TPE3
Feature Type:	ORF, Verified
Description:	Transcription factor that regulates the pleiotropic drug response; zinc cluster protein that is a master regulator involved in recruiting other zinc cluster proteins to pleiotropic drug response elements (PDREs) to fine tune the regulation of multidrug resistance genes; relocates to the cytosol in response to hypoxia; PDR1 has a paralog, PDR3, that arose from the whole genome duplication <sup>2,3,4</sup>
Name Description:	Pleiotropic Drug Resistance <sup>1</sup>
Paralog:	PDR3 <sup>3</sup>
Comparative Info:	

- Locus Overview:** When you first enter the page you should see the Locus Overview, displayed as “PDR1 / YGL013C Overview”. The Description section will give you a short description of what is known about the gene. It’s extremely dense, and you will encounter a lot of unfamiliar terminology. The trick is to look through for terms that are related to the selection condition, which for YMD4700 was caffeine.

## PDR1 / YGL013C Overview

---

<b>Standard Name:</b>	PDR1 <sup>1</sup>
<b>Systematic Name:</b>	YGL013C
<b>SGD ID:</b>	SGD:S000002981
<b>Aliases:</b>	AMY1 <sup>19</sup> , ANT1 <sup>20</sup> , BOR2 <sup>21</sup> , CYH3 <sup>22</sup> , NRA2, SMR2, TIL1, TPE1 <sup>23</sup> , TPE3
<b>Feature Type:</b>	ORF, Verified
<b>Description:</b>	Transcription factor that regulates the pleiotropic drug response; zinc cluster protein that is a master regulator involved in recruiting other zinc cluster proteins to pleiotropic drug response elements (PDREs) to fine tune the regulation of multidrug resistance genes; relocalizes to the cytosol in response to hypoxia; PDR1 has a paralog, PDR3, that arose from the whole genome duplication <sup>2 3 4</sup>

In this case, “pleiotropic drug response” and “multidrug resistance genes” are good clues. Caffeine is a drug, and it seems like PDR1 is important for how yeast cells respond to drugs. It sounds like this is a good candidate to explore further, so let’s look at the summary paragraph for more information.

- Summary Paragraph:** The summary paragraph is a longer description of what is known about this gene. It’s usually in more of a narrative format and can be easier to read (but not always). This one contains a lot of very specific information about what genes are regulated by Pdr1 (written in this paragraph as “Pdr1p” which is sometimes used so the final “p” indicates that it is the protein encoded by the *PDR1* gene). Further into the paragraph though there is a very helpful sentence: “Hyperactive mutants of Pdr1p and Pdr3p often lead to enhanced drug resistance due to an increase in drug transporters (reviewed in 7).” This will be helpful information to form hypotheses about what our mutation could be doing! If you wanted more information, you could click on the “7” link to be taken to the cited paper.

### Summary Paragraph ⓘ

---

PDR1 and PDR3 encode zinc finger transcription factors that are regulators of the pleiotropic drug response in *S. cerevisiae*. Pdr1p and Pdr3p are 36% identical in amino acid composition (5) and can form homodimers or heterodimers (6). Pdr1p and Pdr3p serve as both transcriptional activators and repressors by binding to pleiotropic drug response elements (PDREs) present in the promoters of target genes involved in multidrug resistance (reviewed in 7, 8, and 9). A key feature of the PDRE consensus sequence, 5'-TCCGCGGA-3', is the presence of CCG triplets in an everted repeat orientation (10) and both Pdr1p and Pdr3p constitutively occupy both perfect and degenerate PDREs (6). These two factors have overlapping but not identical sets of target genes and the individual effect on any given shared target gene can also differ (11, 12 and references therein). This variation in regulatory ability may be due either to differences in post-translational modification or heterodimer formation with other transcriptional factors such as Rdr1p and Stb5p (13, 2). Targets include the ABC transporters encoded by PDR5, PDR10, PDR15, SNQ2, and YOR1, the hexose transporter genes HXT9 and HXT11, and sphingolipid biosynthetic genes such as IPT1 (14, and reviewed in 9). Pdr3p also participates in other processes that do not involve Pdr1p, such as retrograde response signaling (15, 16), as well as regulating the DNA damage-inducible genes MAG1 and DDI1 (17). Loss of either PDR1 or PDR3 results in differential drug tolerance, and loss of both *pdr1* and *pdr3* results in severe drug hypersensitivity. Single *pdr1* null mutants are markedly decreased in their resistance to different drugs while the affect of a single *pdr3* null mutation is less severe (5 and 18). Hyperactive mutants of Pdr1p and Pdr3p often lead to enhanced drug resistance due to an increase in drug transporters (reviewed in 7), but only about 10% of the roughly 200 genes containing a PDRE-like element in their promoters respond transcriptionally to the hyperactive forms of Pdr1p and Pdr3p, indicating that factors beyond the presence of a PDRE may be necessary for transcriptional activation by Pdr1p and Pdr3p (reviewed in 8).

5. **Phenotype:** Navigate to the menu bar at the top and click on “Phenotype” to learn more about how the gene contributes to observable traits: the phenotype. This page begins with a summary of the function of the gene. Once again, this emphasizes the role of Pdr1 in how yeast respond to various drugs.

Protein Gene Ontology **Phenotype** Interactions Regulation Expression Literature Homology

## PDR1 / YGL013C Phenotype <sup>?</sup> [Phenotype Help <sup>?</sup>](#)

**Summary:** Non-essential gene; null and point mutations confer resistance or sensitivity to various drugs; in large-scale studies, null mutation confers increased competitive fitness in glycerol medium, decreased fitness in minimal medium, and altered resistance to a wide variety of drugs; overexpression causes slow growth

Scroll down to “Annotations.” This is a table of information collected from research papers, where each row represents a different experiment. Some genes might not have many entries in this table, but PDR1 has a lot.

### Annotations <sup>?</sup> 71 entries for 12 phenotypes

Phenotype	Experiment Type	Mutant Information	Strain Background	Chemical	Details	Reference
competitive fitness: decreased	competitive growth <sup>?</sup>	null Allele: pdr1-Δ	S288C		Media: minimal medium Details: Relative fitness score: 0.984	Breslow DK, et al. (2008) PMID:18622397
competitive fitness: increased	systematic mutation set	null Allele: pdr1-Δ	S288C		Media: glycerol medium, YPG	Qian W, et al. (2012) PMID:23103169

Since we’re interested in caffeine, we can type “caffeine” in the “Filter table” box at the upper right of the table. This returns one row, with “6mM caffeine” highlighted.

Phenotype	Experiment Type	Mutant Information	Strain Background	Chemical	Details	Reference
resistance to chemicals: decreased	homozygous diploid, competitive growth	null Allele: pdr1-Δ	S288C	6 mM caffeine		Brown JA, et al. (2006) PMID:16738548

The Phenotype column says “resistance to chemicals: decreased.” Here the chemical is 6mM caffeine, so the yeast are less resistant to 6mM caffeine.

To find out what change in the yeast made them less resistant to caffeine, look at Mutant Information. It says “null Allele: pdr1Δ.” Null means that the gene is not functional, and “Δ” indicates that is because the gene has been deleted. Putting this information together, that means when PDR1 is not functional, yeast are less resistant to caffeine.

6. Now it's time to synthesize this information to form a hypothesis.
  - We know that Pdr1 regulates drug transport, and that caffeine is a drug.
  - We know that when Pdr1 is not functioning yeast are less resistant to caffeine.
  - From Task 2, we know that the mutation in *PDR1* is a missense mutation, changing one amino acid.

So I can write my hypothesis: Since *PDR1* regulates how yeast respond to drugs, and losing the function of *PDR1* makes yeast less resistant to caffeine, I hypothesize that the missense mutation in YMD4700 makes *PDR1* more active so that the yeast are more resistant to caffeine.

### **TASK 3 QUESTIONS**

Go to the SGD page for the genes from your VCF mutation file from Task 1 where you confirmed the mutations, and answer the questions below for **each gene**. If you have more than 3, you may choose to research a minimum of 3 genes.

1. What is the common name of the gene?
2. What is the function of your gene? Write a 1-sentence summary of what the gene does.
3. Did you notice any keywords related to the selection condition used in your evolution experiment?
4. Are there any experiments from the Phenotype tab that have been done previously to study how yeast respond to the selection you studied?
5. Write a hypothesis about if you think that the mutation in the gene is affecting the ability of your yeast to be more fit in the selective condition. If you don't find a connection, it is ok to hypothesize that the mutation does not affect fitness! However, you should identify at least one mutation that could be contributing to your evolved yeast's increased fitness.