

FungiDB: SNPs and Population Genetics

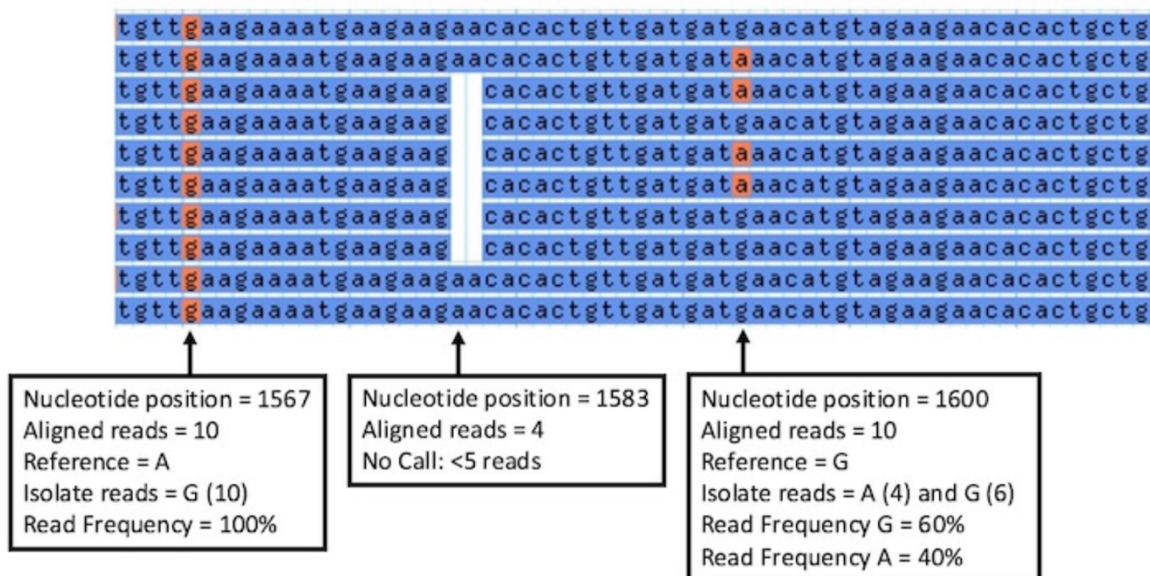
Single Nucleotide Polymorphisms (SNPs) can be used to characterize similarities and differences within a group of isolates or between two groups of isolates. They can also be used to identify genes that may be under evolutionary pressure, either to stay the same (purifying selection) or to change (diversifying or balancing selection).

Isolates are assayed for SNPs in EuPathDB by two basic methods: re-sequencing and the alignment of sequence reads to a reference genome or DNA hybridization to a SNP-chip array.

Read Frequency Threshold: Calling SNPs for each isolate in your group.

Each isolate's sequencing reads are aligned to a reference genome (Organism) and then each nucleotide position with 5 or more aligned reads is examined. A base call is made if the aligned reads meet your Read Frequency Threshold. For example, Isolate X has 10 aligned reads at nucleotide position 1600. If 6 reads are G and 4 reads are A, the read frequency is 60% for the G call and 40% for A. Running this search with the Read Frequency Threshold set to 80% will prevent a base call and consequently exclude Isolate X when returning SNPs for nucleotide position 1600. Running the search with the Read Frequency Threshold set to 60% will bring back a G for this isolate and a 40% threshold will return two calls (both G and A) at this position. The parameter lets you control the quality of the sequencing data and the confidence of the SNP calls. Read Frequency Threshold is a particularly important parameter when dealing with diploid (or aneuploid) organisms since a read frequency of ~50% is expected for heterozygous SNPs.

Isolate X aligned sequencing reads



Minor allele frequency: Parameter for calling SNPs across your isolate group.

The minor allele frequency refers to the least common base call for a single nucleotide position across all isolates. The default setting for this parameter is 0% and returns all SNPs - instances where at least one isolate has a base call that differs from reference. Increase the

Minor allele frequency to ensure that SNPs returned by the search are shared by a larger percentage of isolates in your group.

Isolate consensus sequences aligned to reference genome.

reference	TGGTGATACT	GGTTTTTGTA	CTCCACTTCT	CGGTGCTTCA	TTTTCTACTG
303.1	TGGTGATACT	GGTTTTTGTA	CTCCACTTCT	CGGTGCTTTA	TTTTCTACTG
309.1	TGATAATNCT	GGTTTTTGTA	CTCCACTTCC	CAGTGCTTCA	TTTTCTACTG
RV_3600	TGGTGATACT	GGTTTTTGTA	CTCCACTTCT	CGGTGCTTCA	TTTTCTACTG
RV_3606	TGATAATNCT	GGTTTTTGTA	CTCCACTTCC	CAGTGCTTCA	TTTTCTACTG
RV_3610	TGATGATTCT	GGTTTTTGTA	CTCCACTTCC	CAGTGCTTCA	TTTTCTACTG
SenT119.09	TGGTGATACT	GGTTTTTGTA	CTCCACTTCT	CGGTGCTTCA	TTTTCTACTG
SenT123.09	TGATRAATTCT	GGTTTTTGTA	CTCCACTTCC	CAGTGCTTCA	TTTTCTACTG
SenT140.08	TGGTGATACT	GGTTTTTGTA	CTCCACTTCC	CGGTGCTTCA	TTTTCTACTG
SenT142.09	TGGTGATACT	GGTTTTTGTA	CTCCACTTCC	CAGTGCTTCA	TTTTCTACTG
SenT175.08	TGGTGATACT	GGTTTTTGTA	CTCCACTTCT	CGGTGCTTTA	TTTTCTACTG

Reference = G 6 isolate seq = G 4 isolate seq = A % with base call = 100 Minor allele = A Minor allele freq = 40% (4/10)	Reference = A 6 isolate seq = A 2 isolate seq = T 2 isolate seq = N (no call) % with base call = 80 Minor allele = T Minor allele freq = 25% (2/8)	Reference = G 5 isolate seq = G 5 isolate seq = A % with base call = 100 Minor allele = G or A Minor allele freq = 50% (5/10)
---	--	--

Percent isolates with a base call: Parameter for calling SNPs across your isolate group. Sometimes an isolate does not have a base call at a certain nucleotide position because the Read Frequency Threshold was not met or because there were less than 5 aligned sequencing reads for that nucleotide position. In this case, a SNP can be returned by the search based on a subset of your isolate group. The 'Percent isolates with a base call' parameter defines the fraction of isolates that must have a base call before a SNP is returned for that nucleotide position. The default setting for this parameter is 80% or 8 out of 10 isolates in your group must have a base call for a SNP to be returned by the search. The higher this parameter, the more likely the SNP is to be high quality as regions difficult to align or difficult to sequence will tend to have a lower percentage of calls since the coverage and/or quality will be lower in that region.

1. Identifying SNPs between fungal isolates collected in various geographical areas

The example described below identifies SNPs in *Coccidioides posadasii* (*C. posadasii*) str. Silveira isolates collected from patients with Coccidioidomycosis in the US and Latin America. Coccidioidomycosis, also known as Valley fever, is a fungal disease caused by two closely related species – *C. immitis* and *C. posadasii*. The disease is associated with high morbidity and mortality rates that affects tens of thousands of people each year. The two fungal species are endemic to several regions in the Western Hemisphere, but recent epidemiological and population studies suggest that the geographic range of these fungal species is becoming wider.

a) Identify SNPs based on differences between isolates collected in Guatemala and the US.

- From the *Search for Other Data Types* panel, navigate to the *Identify SNPs based on Differences Between Two Groups of Isolates*.

- In the resulting window first select the target organism '*C. posadasii* str. Silveira' then scroll through the metadata options on the left and make appropriate *Geographic Location* selections from the *Host* section of Characteristic separately for set A and set B isolates. *Set A isolates* should be set to *Guatemala* and *Set B* to the *United States of America*. All other parameters for both sets should be left as *default* (read frequency threshold – 80%, major allele frequency – 80, percent isolates with base call – 50).

Identify SNPs based on Differences Between Two Groups of Isolates

Organism
Coccidioides posadasii str. Silveira

Set A Isolates
68 Set A Isolates Total
5 of 68 Set A Isolates selected
Country x

Country
Keep checked values at top
1 of 68 Set A Isolates have no data provided for this filter

Country	Remaining Set A Isolates	Set A Isolates	Distribution	%
Argentina	1 (1%)	1 (1%)		(100%)
Brazil	1 (1%)	1 (1%)		(100%)
Guatemala	5 (7%)	5 (7%)		(100%)
Mexico	9 (13%)	9 (13%)		(100%)
Paraguay	1 (1%)	1 (1%)		(100%)
United States of America	50 (73%)	50 (73%)		(100%)

Set B Isolates
68 Set B Isolates Total
50 of 68 Set B Isolates selected
Country x

Country
Keep checked values at top
1 of 68 Set B Isolates have no data provided for this filter

Country	Remaining Set B Isolates	Set B Isolates	Distribution	%
Argentina	1 (1%)	1 (1%)		(100%)
Brazil	1 (1%)	1 (1%)		(100%)
Guatemala	5 (7%)	5 (7%)		(100%)
Mexico	9 (13%)	9 (13%)		(100%)
Paraguay	1 (1%)	1 (1%)		(100%)
United States of America	50 (73%)	50 (73%)		(100%)

The search strategy returns SNPs rather than genes, which are classified by genomic location within the results table. When individual SNPs fall within a gene, its corresponding Gene ID

(SNPs) Strategy: Two Groups(2) * [Rename](#) [Duplicate](#) [Save As](#) [Share](#) [Delete](#)

Two Groups 13982 SNPs Step 1 [Add Step](#)

13982 SNPs from Step 1 [Revise](#) Strategy: Two Groups(2)

SNP Results [First](#) [1](#) [2](#) [3](#) [4](#) [5](#) [Next](#) [Last](#) [Advanced Paging](#) [Download](#) [Add to Basket](#) [Add Columns](#)

SNP Id	Location	Gene ID	Position in protein	Coding
NGS_SNP.GL636486.1022206	GL636486: 1,022,206	CPSG_00348	1616	coding
NGS_SNP.GL636486.1125536	GL636486: 1,125,536	CPSG_00368	98	coding
NGS_SNP.GL636486.1140082	GL636486: 1,140,082	CPSG_00376	1	coding
NGS_SNP.GL636486.1144476	GL636486: 1,144,476	CPSG_00379	68	coding
NGS_SNP.GL636486.1159591	GL636486: 1,159,591	CPSG_00387	215	coding

is listed next to the SNP record.

- To examine a SNP record page, click on the *SNP.GL636486.1125536* in the CPSG_00368 gene. *Note, you might have to scroll down to find the SNP or you can follow the next step.*
 - If your results table looks somewhat different and you cannot easily locate the SNP mentioned above – can you think of other ways to locate this SNP within your results?

Hint: Click Add Step and look up the SNP by its ID: SNP.GL636486.1125536

The screenshot shows the 'Add Step' interface. On the left, there are three options: 'Run a new Search for', 'Add contents of Basket', and 'Add existing Strategy'. In the center, a list of categories is shown: Genes, Genomic Segments, SNPs, and ORFs. The 'SNPs' category is selected, and a sub-menu is open showing options: 'SNP ID(s)', 'Genomic Location', 'Differences Within a Group of Isolates', 'Differences Between Two Groups of Isolates', and 'Gene IDs'. An orange arrow points from the 'SNP ID(s)' option to the 'Add Step 2 : SNP ID(s)' section below.

Add Step 2 : SNP ID(s)

SNP ID input set

Enter a list of IDs or text:

☐ Copy from My Basket: 0 SNPs will be copied from your Basket.

☐ Copy from My Strategy: Choose a SNP strategy: JoF SNPs between Two Groups (Guatemala & US) (19280 SNPs)

Combine SNPs in Step 1 with SNPs in Step 2:

☒ 1 Intersect 2 ☐ 1 Minus 2
☐ 1 Union 2 ☐ 2 Minus 1
☐ 1 Relative to 2, using genomic colocation

(SNP)

Step 1: Two Groups 13987 SNPs

Step 2: SNP ID(s) 1 SNP

Add Step

Run Step

SNP location, allele summary, associated GeneID, major and minor allele records can be found at the top of the page, followed by DNA polymorphism summary and SNP records table that is searchable by isolates.

SNP: NGS_SNP.GL636486.1125536

Organism: *Coccidioides posadasii* str. Silveira

Location: GL636486: 1,125,536

Type: coding

Number of Strains: 68

Gene ID: CPSG_00368

Gene Strand: reverse

Major Allele: A (0.84)

Minor Allele: G (0.16)

Distinct Allele Count: 2

Reference Allele: A

Reference Product: L 98

Allele (gene strand): T

SNP context: CGTCCATCCTCTCACTCCCTGTCCCTGCCAATCGGTGTCGAAGTGTGTGGCTGAGATCTC

SNP context (gene strand): GAGATCTCAGCCACACACTTCGACACCGATTGGCAGGGACAGGGAGTGAGAGGATGGACG

Genomic location, SNP type and aligned reads can be displayed in GBrowse by clicking on the *View in genome browser* button. SNP tracks can be activated from the *Select Tracks* tab by selecting *SNPs by coding potential* under *DNA polymorphism* in the *Genetic variation* section. Hover over SNPs labeled as red diamonds (nonsense SNPs) to get more information.

- Examine SNP record page further. Note that in addition to US and Guatemala SNP records it also contains information for other isolates collected elsewhere, where individual reads can be activated by clicking on the *view alignment* link from within the table. This action will re-direct you to the GBrowse where you can select either all or specific isolates listed under the *Aligned Genomic Sequence Reads for C. posadasii str. Silveira* to view specific tracks.

Strains / Samples Download Data sets								
Search this table... Q Showing 68 rows								
Geographic Location	Strain	Sample	Allele	Allele (gene strand)	Product	Coverage	Read Frequency	View Alignment
	Silveira (reference)		A	T	L			
Argentina	B0727_Argentina	EUSMPL0054-1-57	A	T	L	9	100	view alignment
Brazil	B5773_Brazil	EUSMPL0054-1-62	A	T	L	15	100	view alignment
Guatemala	730332_Guatemala	EUSMPL0054-1-65	G	C	L	7	100	view alignment
Guatemala	730333_Guatemala	EUSMPL0054-1-66	G	C	L	10	100	view alignment
Guatemala	730334_Guatemala	EUSMPL0054-1-67	A	T	L	13	100	view alignment
Guatemala	B0858_Guatemala	EUSMPL0054-1-58	G	C	L	21	100	view alignment

b) Determine genes that map to each of the SNPs identified in Step 1.

- Add Step, Run a new Search for, Genes, Taxonomy, and choose *C. posadasii str. Silveira*

Add Step 2 : Organism

1 selected

posad

☒ Fungi

☒ Eukaryotes

☒ Coccidioides

☒ Coccidioides posadasii

☐ Coccidioides posadasii C735 delta SOWgp

☐ Coccidioides posadasii RMSCC 3488

☒ Coccidioides posadasii str. Silveira

[add these](#) | [clear these](#) | [select only these](#) | [select all](#) | [clear all](#)

Combine SNPs in Step 1 with Genes in Step 2:

☒ 1 Intersect 2
 ☐ 1 Minus 2
 ☐ 1 Union 2
 ☐ 2 Minus 1
 ☒ 1 Relative to 2, using genomic colocation

[Continue....](#)

Add Step

Genomic Colocation

Combine Step 1 and Step 2 using relative locations in the genome

You had **13982 SNPs** in your Strategy (Step 1). Your new **Genes** search (Step 2) returned **10379 Genes**.

"Return each **Gene from Step 2** whose **exact region** overlaps the **exact region** of a SNP in Step 1 and is on **either strand**"

(10379 Genes in Step 2)

Region

Gene

☒ Exact

☐ Upstream: 1000 bp

☐ Downstream: 1000 bp

☐ Custom:

begin at: start + 0 bp

end at: stop + 0 bp

(13982 SNPs in Step 1)

Region

SNP

☒ Exact

☐ Upstream: 1000 bp

☐ Downstream: 1000 bp

☐ Custom:

begin at: start + 0 bp

end at: stop + 0 bp

[Submit](#)

[Close](#)

- Next window will bring up a colocation tool where you will be able to set parameters of your gene search.
- Choose to Return each *Gene from Step 2* whose exact region overlaps the exact region of a SNP in Step 1 and is on *either strand*
- Click *Submit*
- Examine gene list returned

(Genes)

Organism: 10379 Genes

Two Groups: 13982 SNPs

4071 Genes

[Add Step](#)

Strategy: Two Groups(2) *

[Rename](#)
[Duplicate](#)
[Save As](#)
[Share](#)
[Delete](#)

4071 Genes from Step 2

Strategy: Two Groups(2)

[Revise](#)

Click on a number in this table to limit/filter your results

Gene Results | [Genome View](#) | [Analyze Results](#)

Genes: 4071 Transcripts: 4072 ☐ Show Only One Transcript Per Gene

First 1 2 3 4 5 Next Last Advanced Paging

[Download](#) [Add to Basket](#) [Add Columns](#)

Gene ID	Transcript ID	Organism	Genomic Location (Transcript)	Product Description	Match Count	Region	Matched Regions	Gene Name or Symbol	Entrez Gene ID	Pre ID
CPSG_00001	CPSG_00001-i26_1	C. posadasii str. Silveira	GL636486:37354..39412(-)	dihydrofolate S-succinyltransferase	1	37246 - 39474 (-)	NGS_SNP.GL636486.37446: 37,446 - 37,446 (+)	N/A		N/A
CPSG_00002	CPSG_00002-i26_1	C. posadasii str. Silveira	GL636486:40033..40587(+)	LYR family protein	3	39629 - 40814 (+)	NGS_SNP.GL636486.39980: 39,980 - 39,980 (+); NGS_SNP.GL636486.39797: 39,797 - 39,797 (+); NGS_SNP.GL636486.40392: 40,392 - 40,392 (+)	N/A		N/A
CPSG_00003	CPSG_00003-i26_1	C. posadasii str. Silveira	GL636486:41089..41652(-)	histone H2A.2	2	40679 - 41780 (-)	NGS_SNP.GL636486.40844: 40,844 - 40,844 (+); NGS_SNP.GL636486.41181: 41,181 - 41,181 (+)	N/A		N/A
CPSG_00004	CPSG_00004-i26_1	C. posadasii str. Silveira	GL636486:42275..42857(+)	histone H2b	2	42138 - 43101 (+)	NGS_SNP.GL636486.42241: 42,241 - 42,241 (+); NGS_SNP.GL636486.42606: 42,606 - 42,606 (+)	N/A		N/A
CPSG_00005	CPSG_00005-i26_1	C. posadasii str. Silveira	GL636486:43189..44647(-)	cysteine protease atp4	2	43022 - 44859 (-)	NGS_SNP.GL636486.43272: 43,272 - 43,272 (+); NGS_SNP.GL636486.43225: 43,225 - 43,225 (+)	N/A		N/A
CPSG_00007	CPSG_00007-i26_1	C. posadasii str. Silveira	GL636486:46300..48242(-)	hypothetical protein	1	45938 - 48389 (-)	NGS_SNP.GL636486.47106: 47,106 - 47,106 (+)	N/A		N/A
CPSG_00009	CPSG_00009-i26_1	C. posadasii str. Silveira	GL636486:60804..62170(-)	conserved hypothetical protein	3	60531 - 62359 (-)	NGS_SNP.GL636486.61445: 61,445 - 61,445 (+); NGS_SNP.GL636486.61439: 61,439 - 61,439 (+); NGS_SNP.GL636486.61444: 61,444 - 61,444 (+)	N/A		N/A

- Think about how can you analyze this data further?

*Hint: you can extract genes that have *hypothetical* in the product description via the Text search. You can also perform GO enrichment or identify orthologs in other species, or map to metabolic pathways etc., or you can take use other resources as shown previously to cross reference the integrated data.*

2. Identify SNPs within a group of isolates

- Navigate to the *Identify other Data Types* panel and select the *Differences Within a Group of Isolates* search from *SNPs* menu
- Let's look at the SNPs from *Coccidioides* species.
- Select *Guatemala, Texas, Phoenix, and Nevada* isolates.

Revise Step 1 : Differences Within a Group of Isolates

Organism

Samples

Select Samples

expand all | collapse all

Find a quality

Name

select all | clear all

Name	Total Samples	Matching Samples	Distribution
<input checked="" type="checkbox"/> 730332_Guatemala	1	1	<div></div>
<input checked="" type="checkbox"/> 730333_Guatemala	1	1	<div></div>
<input checked="" type="checkbox"/> 730334_Guatemala	1	1	<div></div>
<input type="checkbox"/> B0727_Argentina	1	1	<div></div>
<input type="checkbox"/> B0858_Guatemala	1	1	<div></div>
<input checked="" type="checkbox"/> B10757_Nevada	1	1	<div></div>
<input checked="" type="checkbox"/> B10813_Texas	1	1	<div></div>
<input checked="" type="checkbox"/> B1249_Guatemala	1	1	<div></div>
<input type="checkbox"/> B5773_Brazil	1	1	<div></div>
<input type="checkbox"/> Coahuila_1	1	1	<div></div>
<input type="checkbox"/> Coahuila_2	1	1	<div></div>
<input type="checkbox"/> Colorado_Springs_1	1	1	<div></div>
<input checked="" type="checkbox"/> GT002_Texas	1	1	<div></div>
<input type="checkbox"/> GT017_Paraguay	1	1	<div></div>
<input type="checkbox"/> Guerrero_1	1	1	<div></div>
<input type="checkbox"/> Michoacan_1	1	1	<div></div>
<input type="checkbox"/> Michoacan_2	1	1	<div></div>
<input type="checkbox"/> Nuevo_Leon_1	1	1	<div></div>
<input type="checkbox"/> Nuevo_Leon_2	1	1	<div></div>
<input checked="" type="checkbox"/> Phoenix_1	1	1	<div></div>
<input checked="" type="checkbox"/> Phoenix_2	1	1	<div></div>
<input checked="" type="checkbox"/> Phoenix_3	1	1	<div></div>
<input checked="" type="checkbox"/> Phoenix_4	1	1	<div></div>

☐ All Samples
☒ Samples remaining when other criteria have been applied.

Read frequency threshold

Minor allele frequency >=

Percent isolates with a base call >=

- To set stringent control for quality and confidence of the SNP calls select 80% Read frequency threshold, leave minor allele frequency at default and percent isolates parameter at 80.
- How many SNPs were returned?
- How would you identify heterozygous SNPs?

Note: Create a new search or revise an existing search strategy. Modify a read frequency threshold of 40% and revise this search and increase the minor allele frequency threshold (try 20 and 40 and compare results).

My Strategies: New Opened (2) All (177) Basket Public Strategies (39) Help

(SNPs)

One Group 55815 SNPs Step 1 Add Step

STEP 1: One Group

Organism : Coccidioides posadasii str. Silveira

Samples : Name is "730332_Guatemala", "730333_Guatemala", "730334_Guatemala", "B0858_Guatemala", "Phoenix_1", "Phoenix_2", "Phoenix_3", "Phoenix_4", "Phoenix_5", "Phoenix_6", "Phoenix_7", "Phoenix_8", "Phoenix_9"

Read frequency threshold : 40%

Minor allele frequency >= : 40

Percent isolates with a base call >= : 40

Results: 55815 SNPs

Give this search a weight

SNP Results

First 1 2 3 4 5 Next Last Advanced Paging

SNP ID	Location	Gene ID	Position in protein	% Minor Alleles	% Calls	Phenotype	Major Allele	Minor Allele
NGS_SNP.GL636486.1000245	GL636486:1,000,245	CPSG_00339	155	46.2	100	syn	A	G
NGS_SNP.GL636486.100036	GL636486:100,036		N/A	44.4	69.2	non-coding	G	C
NGS_SNP.GL636486.100070	GL636486:100,070		N/A	44.4	69.2	non-coding	C	T
NGS_SNP.GL636486.1002006	GL636486:1,002,006		N/A	46.2	100	non-coding	T	A
NGS_SNP.GL636486.100242	GL636486:100,242		N/A	44.4	69.2	non-coding	G	A
NGS_SNP.GL636486.100370	GL636486:100,370		N/A	44.4	69.2	non-coding	C	T
NGS_SNP.GL636486.1003800	GL636486:1,003,800	CPSG_00341	354	46.2	100	non-syn	A	G
NGS_SNP.GL636486.100561	GL636486:100,561		N/A	44.4	69.2	non-coding	C	T
NGS_SNP.GL636486.1005788	GL636486:1,005,788	CPSG_00342	254	46.2	100	syn	A	G
NGS_SNP.GL636486.100583	GL636486:100,583		N/A	44.4	69.2	non-coding	C	T
NGS_SNP.GL636486.1006954	GL636486:1,006,954		N/A	46.2	100	non-coding	C	A
NGS_SNP.GL636486.100727	GL636486:100,727		N/A	44.4	69.2	non-coding	C	T
NGS_SNP.GL636486.1007469	GL636486:1,007,469	CPSG_00343	N/A	46.2	100	non-coding	A	G
NGS_SNP.GL636486.1008372	GL636486:1,008,372	CPSG_00343	N/A	46.2	100	non-coding	G	A
NGS_SNP.GL636486.100998	GL636486:100,998		N/A	40	76.9	non-coding	C	T

Note: Read frequency threshold applies to the sequencing reads of individual isolates and defines a stringency for data supporting a SNP call between an isolate and the reference genome (Organism). Each nucleotide position of each isolate is compared to the reference genome and a SNP call is made if the portion of the isolate's aligned reads that support the SNP is above the Read Frequency Threshold (RFT). Find high quality haploid SNPs with 80% RFT or heterozygous diploid/aneuploid SNPs with 40%.

- How many SNPs did you identify?
- Why might you want to increase the minor allele threshold when you run SNP searches?

Note: Minor Allele Frequency parameter applies to your group of isolates. A SNP can occur in any number of isolates in your group and the least frequent SNP call across all isolates is the Minor Allele Frequency. A SNP will be returned by the search if the frequency of the minor allele is equal to or greater than your Minor Allele Frequency.

1. Identify genes with nonsense SNPs

- Navigate to the Genes by SNP characteristics search, which can be found under the Genetic Variation category in the gene searches section.
- Select *Aspergillus fumigatus* from the list of organisms and configure the search to identify SNPs in isolates originating from 'environmental' sample types.

- Examine your results. How many genes were identified in your search?

SNPs
261 Genes
Step 1

Add Step

261 Genes from Step 1 [Revise](#)
Strategy: SNPs(4)

Click on a number in this table to limit/filter your results

Gene Results [Genome View](#) [Analyze Results](#)

First 1 2 3 Next Last [Advanced Paging](#) [Download](#) [Add to Basket](#) [Add Columns](#)

Gene ID	Transcript ID	Product Description	Total SNPs	Nonsynonymous SNPs	Synonymous SNPs	Nonsense SNPs	Non-coding SNPs	Nonsyn/syn SNP ratio	SNPs per Kb (CDS)
Afu5g00340	Afu5g00340-T	Has domain(s) with predicted ATP binding, ATPase activity, ATPase activity, coupled to transmembrane...	172	60	32	3	77	1.88	36.58
Afu8g00342	Afu8g00342-T	Ortholog of A. fumigatus Af293 : Afu5g01010, A. niger CBS 513.88 : An11g08170, Neosartorya fischeri ...	132	5	0	1	126	5	6.74
Afu8g06132	Afu8g06132-T	Ortholog of Aspergillus brasiliensis : Aspb1_0060310, Aspergillus glaucus : Aspg1_1496032, Neosart...	129	45	37	11	36	1.22	53.44
Afu6g14630	Afu6g14630-T	Ortholog of A. nidulans FGSC A4 : AN5945, Neosartorya fischeri NRRL 181 : NFIA_060670 and Aspergillu...	128	82	30	16	0	2.73	96.39

- How do these results change if you modify the stringency of the selection criteria?
- Afu5g00340 is one of the genes with several nonsense SNPs. Navigate to its gene record page in FungiDB and click on the **SNPs** data shortcut to be redirected to the Genetic variation section of the page.

Afu5g00340 Has domain(s) with predicted ATP binding, ATPase activity, ATPase activity, coupled to tran...

Type: protein coding
Chromosome: 5
Location: Chr5_A_fumigatus_Af293:93,541..98,243(-)
Species: *Aspergillus fumigatus*
Strain: Af293
Status: Curated Reference Strain

Shortcuts

Synergy BLAT Alignments **SNPs** Transcriptsomics Protein Features Proteomics

Also see Afu5g00340 in the JBrowse Genome Browser [View](#) or Protein Browser [View](#)

Add the first user comment

Contents

expand all | collapse all

Search section names...

- 1 Gene models
- 2 Annotation, curation and identifiers
- 3 Link outs
- 4 Genomic Location
- 5 Literature
- 6 Taxonomy
- 7 Orthology and synteny
- 8 Phenotype
- 9 Genetic variation
- 10 Transcriptsomics
- 11 Sequence analysis
- 12 Sequences
- 13 Structure analysis
- 14 Protein features and properties
- 15 Function prediction
- 16 Pathways and interactions
- 17 Proteomics
- 18 Immunology

expand all | collapse all

SNPs

View in genome browser

Chr5_A_fumigatus_Af293

Reannotated Transcripts (UTRs in gray when available)

afu5g00340-T

afu5g00340-T

afu5g00340-T

SNPs by coding potential

View in genome browser

Non-Coding SNPs All Strains 81

NonSyn/Syn SNP Ratio All Strains 1.94

NonSynonymous SNPs All Strains 70

SNPs with Stop Codons All Strains 4

Synonymous SNPs All Strains 36

Total SNPs All Strains 191

- Navigate to GBrowse by clicking on the *View in genome browser* button. Once in GBrowse, activate *SNPs by coding potential* track (Hint: navigate to the *Select Tracks* tab to select)

Browser **Select Tracks** Snapshots Custom Tracks Preferences

<< Back to Browser Show Active Tracks Only Show Favorites Only

Search for Specific Tracks

Search: SNPs by coding potential ☒ Stick to top when scrolled

Clear Search

Tracks

- ☒ %0/var/www/w2.fungidb.org/cgi-bin/gbrowse Genetic variation
- ☒ DNA polymorphism ☒ All on ☐ All off
- ☒ SNPs by coding potential [7]
- ☒ %0/var/www/w2.fungidb.org/cgi-bin/gbrowse Sequence analysis
- ☒ Sequence sites, features and motifs
- ☒ DNA binding
- ☐ Restriction Sites ☐ All on ☐ All off
- ☐ Restriction Sites

Browser **Select Tracks** Snapshots Custom Tracks Preferences

Search

Landmark or Region: Chr5_A_fumigatus_Af293:91,190..100,5 Search

Annotate Restriction Sites Configure... Go

Save Snapshot Load Snapshot

Scroll/Zoom: Show 9.405 kbp Flip

Overview

Chr5_A_fumigatus_Af293

Region

Chr5_A_fumigatus_Af293: 9,405 kbp

Details

SNPs by coding potential

Annotated Transcripts (UTRs in gray when available)

afu5g00340-T

afu5g00340-T

afu5g00340-T

RNASeq evidence for introns (filtered) (Showing 2 of 2 subtracks)

afu5g00340-T

afu5g00340-T

afu5g00340-T

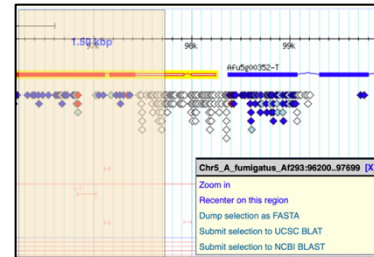
RNASeq evidence for introns (inclusive) (Showing 2 of 2 subtracks)

afu5g00340-T

afu5g00340-T

afu5g00340-T

- Zoom in to region with several nonsense SNPs.
- Identify isolates that contain nonsense mutations.
- Click on the link to be redirected to the SNP record page.
 - Examine other records on the isolate record page.
- How many alleles are reported for this locus?



SNP: NGS_SNP.Chr5_A_fumigatus_Af293.96655

Organism: *Aspergillus fumigatus* Af293
Location: Chr5_A_fumigatus_Af293: 96,655
Type: coding
Number of Strains: 48
Gene ID: Afu5g00340
Gene Strand: reverse
Major Allele: G (0.96)
Minor Allele: A (0.04)
Distinct Allele Count: 2
Reference Allele: G
Reference Product: R 238
Allele (gene strand): C
SNP context: CGGCTGCGATGATCATGATCCTGCGGTATCGGGCGCCAGCAGCGAGTTCGTCAACACGCAG
SNP context (gene strand): CTGCGTGTTGACGAACCTCGCTGCTGGCGCCGATACCGCAGGATCATGATCATCGCAGCCG

- Can you identify the specific isolates that contained a nonsense mutation?
Hint: Look in Strains/Sample table.

▼ Strains / Samples Download Data Sets

Search this table... Showing 48 rows

Geographic Location	Strain	Sample	Allele	Allele (gene strand)	Product	Coverage	Read Frequency	DNA-seq reads for strain
	Af293 (reference)		G	C	R			
India	Afu_1042-09	EUSMPL0067-1-16	G	C	R	60	100	view DNA-seq reads
India	Afu_124-E11	EUSMPL0067-1-19	G	C	R	66	100	view DNA-seq reads
India	Afu_166-E11	EUSMPL0067-1-20	G	C	R	54	100	view DNA-seq reads
India	Afu_218-E11	EUSMPL0067-1-22	G	C	R	55	100	view DNA-seq reads
India	Afu_257-E11	EUSMPL0067-1-21	G	C	R	49	100	view DNA-seq reads
India	Afu_343-P-11	EUSMPL0067-1-17	G	C	R	62	98.41	view DNA-seq reads
India	Afu_591-12	EUSMPL0067-1-18	G	C	R	44	100	view DNA-seq reads
India	Afu_942-09	EUSMPL0067-1-15	G	C	R	74	100	view DNA-seq reads
Kingdom of the Netherlands	08-12-12-13	EUSMPL0067-1-7	G	C	R	92	98.92	view DNA-seq reads
Kingdom of the Netherlands	08-19-02-10	EUSMPL0067-1-14	G	C	R	120	100	view DNA-seq reads
Kingdom of the Netherlands	08-19-02-30	EUSMPL0067-1-11	A	T	*	45	100	view DNA-seq reads

- Navigate to GBrowse by clicking on View in genome browser button and activate coverage tracks for 08-19-02-30 and 12-7504462 isolates from the Aligned genome sequence reads menu.

Browser Select Tracks Snapshots Custom Tracks Preferences

<< Back to Browser Show Active Tracks Only Show Favorites Only

Search for Specific Tracks

Search: 08-19-02-30 Stick to top when scrolled

Clear Search

Tracks

- %0/var/www/w2.fungidb.org/cgi-bin/gbrowse Genetic variation
 - A. fumigatus Af293
 - Genomic Context of Azole Resistance Mutations in *Aspergillus fumigatus*
 - Aligned Genomic Sequence Reads All on All off
 - 08-19-02-30 Coverage aligned to afumAf293 [Reads < 3k; Bases < 100bp] [?]
- %0/var/www/w2.fungidb.org/cgi-bin/gbrowse Sequence analysis
 - Sequence sites, features and motifs
 - DNA binding
 - Restriction Sites All on All off
 - Restriction Sites

- [illegible]

9 Genetic variation

▼ 9.1 DNA polymorphism

▼ SNPs

