

tbVar: A comprehensive genome variation resource for *Mycobacterium tuberculosis*

This is a general manual to help users search, navigate and even annotate their *Mycobacterium tuberculosis* (Mtb) genomic variations.

Homepage

Homepage of the database introduces the database and its purpose to the users. It also gives a brief overview of the data that the database holds.

Users can investigate the database using search box provided on the homepage using multiple parameters as shown in figures below.

The screenshot shows the tbVar homepage with several key features highlighted by red arrows and text boxes:

- Search box:** A search input field with the placeholder text "Search: Variant Location, Range, Gene Name, etc." and a "Eureka" button.
- Browse variations by position:** A text box pointing to the search input field.
- Browse variations by:** A text box pointing to the "Database can be accessed by:" section.
- Browse variations over a range of genomic position:** A text box pointing to the "Browse by Genome Position Range" section.
- Browse the genome:** A text box pointing to the "Browse the database" section.

The "Database can be accessed by:" section lists four methods:

- Browsing by Variant Location: Example: 1817019 | 3073507 | 4222225
- Browsing by Gene Name: Example: *katG* | *pncA* | *gyrA*
- Browsing RvID: Example: Rv1089 | Rv1090 | Rv1091
- Browsing Genome Position Range: Example: 1500-15000 | 30000-35000 | 80000-85000

The "Statistics" section features two charts:

- Venn Diagram:** Shows the overlap of variations between TBvar, TBDB, and MTCDB. The counts are: TBvar only (21632), TBDB only (1913), MTCDB only (208), TBvar & TBDB (1813), TBvar & MTCDB (162), TBDB & MTCDB (102), and all three (0).
- Pie Chart:** Shows the distribution of variation types: Stop gain (39), Stop loss (38), Upstream (77), Downstream (39), Others (228), Non-synonymous (4284), Synonymous (1495), and Upstream (39).

Citation: [Joshi et. al (2013): tbVar: A comprehensive genome variation resource for *Mycobacterium tuberculosis*]

Copyright ©2013 CSIR-Institute of Genomics and Integrative Biology

For the convenience of users database is made searchable through multiple features such as variant position, gene name/id or a range of genomic position as mentioned in a number of examples on the homepage.

Search term in any of the above mentioned form takes user to a separate result page which is further divided into different sections and easily.

tbvar


The output page for database query is divided into different sections which can be explored using different tabs at the top of result table. Different sections show different biological and other information related to query.

Genomic variations: This section gives information on the genomic position of the variations along with their genomic loci (e.g. whether they lie in genic part of the inter-genic region of the genome). This section also mentions the count and frequency of occurrence of the particular variation within the population of samples chosen for building database. Finally, this section also links out to other various databases for the variants found in other similar databases.

The screenshot displays the tbvar web application interface. At the top, there is a navigation bar with the tbvar logo and the text "Mycobacterium tuberculosis variome resource". Below the navigation bar are tabs for "HOME", "annoTB", "MANUAL", and "CONTACT". A search bar is located in the center, with the text "Search: Variant Location, Range, Gene or Rvid" and a "Eureka" button. Below the search bar are several tabs: "Genomic Variations", "Gene Annotation", "Functional Effects", "Regulatory Variations", "Strain Information", "Drug Resistance", and "Genome Browser". The "Genomic Variations" tab is selected, and a table of genomic variations is displayed below it.

Gene Id	Position	Ref Allele	Alt Allele	Location	Type	Variant Count	Frequency Percentage	External Link
Rv1908c	2154532	G	A	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2154724	C	A	exonic	nonsynonymous SNV	175	37.31	TBDB, MTCID
Rv1908c	2154730	T	G	exonic	nonsynonymous SNV	2	0.43	
Rv1908c	2154871	T	C	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2154902	A	G	exonic	synonymous SNV	2	0.43	
Rv1908c	2154980	G	T	exonic	nonsynonymous SNV	1	0.21	dbSNP
Rv1908c	2155140	G	A	exonic	synonymous SNV	3	0.64	
Rv1908c	2155167	G	T	exonic	nonsynonymous SNV	4	0.85	MTCID
Rv1908c	2155168	C	G	exonic	nonsynonymous SNV	52	11.09	TBDB, MTCID
Rv1908c	2155168	C	T	exonic	nonsynonymous SNV	6	1.28	TBDB, MTCID
Rv1908c	2155301	T	G	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155343	T	C	exonic	nonsynonymous SNV	2	0.43	
Rv1908c	2155412	C	T	exonic	nonsynonymous SNV	5	1.07	
Rv1908c	2155503	G	A	exonic	synonymous SNV	3	0.64	TBDB
Rv1908c	2155726	A	G	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155732	T	G	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155751	C	T	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155783	G	A	exonic	nonsynonymous SNV	1	0.21	MTCID
Rv1908c	2155832	C	T	exonic	nonsynonymous SNV	1	0.21	
Rv1908c	2155945	T	G	exonic	nonsynonymous SNV	2	0.43	
Rv1908c	2155964	G	A	exonic	stopgain SNV	1	0.21	

Gene annotation: This section shows all the information pertaining to a gene harboring variations. It shows gene name and its ID. This tab also gives annotation for the respective gene. Genomic coordinates of genes and its orientation is also presented in this section.

tbvar Mycobacterium tuberculosis genome resource 

HOME annoTB MANUAL CONTACT


Search: Variant Location, Range, Gene or Rvid

Genomic Variations Gene Annotation Functional Effects Regulatory Variations Strain Information Drug Resistance Genome Browser

Gene Id	Gene Name	Gene Annotation	Start	Stop	Orientation
Rv1908c	katG	Catalase-peroxidase-peroxynitritase T KatG	2153888	2156111	-

Copyright©2013 CSIR

Functional effects: This section shows SIFT prediction and SIFT score for non-synonymous variations. SIFT predicts whether a genomic variation has any functional consequence on the corresponding protein. Here we consider a change in protein structure as predicted by SIFT to be DELETERIOUS, while no change in the protein structure is considered as TOLERATED. SIFT score is a major parameter in prediction and is provided in this tab.

tbvar Mycobacterium tuberculosis genome resource 

HOME annoTB MANUAL CONTACT

Search: Variant Location, Range, Gene or Rvid

Genomic Variations Gene Annotation Functional Effects Regulatory Variations Strain Information Drug Resistance Genome Browser

Position	Amino Acid Mutation	Functional Effect	SIFT Score
2154532	S527L	DELETERIOUS	0.02
2154724	R463L	TOLERATED	1
2154730	Q461P	DELETERIOUS	0
2154871	K414R	DELETERIOUS	0
2154902	L404L	Not Scored	Not Scored
2154980	L378M	DELETERIOUS	0.03
2155140	T324T	Not Scored	Not Scored
2155167	S315R	DELETERIOUS	0
2155168	S315T	DELETERIOUS	0
2155168	S315N	DELETERIOUS	0
2155301	T271P	DELETERIOUS	0
2155343	M257V	DELETERIOUS	0
2155412	G234R	DELETERIOUS	0
2155503	T203T	Not Scored	Not Scored
2155726	F129S	DELETERIOUS	0
2155732	Q127P	DELETERIOUS	0
2155751	G121S	DELETERIOUS	0
2155783	A110V	DELETERIOUS	0
2155832	D94N	DELETERIOUS	0
2155945	D56A	DELETERIOUS	0.01
2155984	Q50X	Not Scored	Not Scored

Regulatory variations: Variations lying in the regulatory elements on Mtb genome as found through Transcription factor (TF) ChIP-seq of 50 transcription factors are reported in this section. The targets predicted against these TFs are also reported in this section.

The screenshot shows the tbvar website interface. At the top, there is a navigation bar with the tbvar logo and the text 'Mycobacterium tuberculosis variome resource'. To the right is the IGB logo. Below the navigation bar are tabs for 'HOME', 'annoTB', 'MANUAL', and 'CONTACT'. A search bar is present with the text 'Search: Variant Location, Range, Gene or Rvid' and a 'Eureka' button. Below the search bar are several tabs: 'Genomic Variations', 'Gene Annotation', 'Functional Effects', 'Regulatory Variations' (which is highlighted), 'Strain Information', 'Drug Resistance', and 'Genome Browser'. The main content area displays a table with the following data:

Position	Regulator	Target	Start	Stop
2154532	Rv3133c	Rv1908c	2154420	2154639
2154532	Rv3133c	Rv1908c	2154421	2154645
2154724	Rv3133c	Rv1908c	2154711	2154947
2154724	Rv2989	Rv1908c	2154549	2154888
2154730	Rv3133c	Rv1908c	2154711	2154947
2154730	Rv2989	Rv1908c	2154549	2154888
2154871	Rv3133c	Rv1908c	2154711	2154947
2154871	Rv2989	Rv1908c	2154549	2154888
2154871	Rv3133c	Rv1908c	2154758	2154926
2154902	Rv3133c	Rv1908c	2154711	2154947
2154902	Rv3133c	Rv1908c	2154758	2154926
2154980	No Regulatory Effect			
2155140	No Regulatory Effect			
2155167	No Regulatory Effect			
2155168	No Regulatory Effect			
2155301	No Regulatory Effect			
2155343	No Regulatory Effect			
2155412	No Regulatory Effect			
2155503	No Regulatory Effect			
2155726	No Regulatory Effect			
2155732	No Regulatory Effect			
2155751	No Regulatory Effect			

Strain information: This section reports information on the samples and corresponding experiment and study from which the variations were derived.

tbvar | Mycobacterium tuberculosis
variant resource

HOME | annoTB | MANUAL | CONTACT

Search: Variant Location, Range, Gene or Rvid

Genomic Variations | Gene Annotation | Functional Effects | Regulatory Variations | **Strain Information** | Drug Resistance | Genome browser

Position	Sample	Experiment	Reference
2154532	ERS019751	ERX013857	ERP000436
2154724	ERS019735	ERX013859	ERP000436
2154724	ERS019739	ERX013859	ERP000436
2154724	ERS019741	ERX013859	ERP000436
2154724	ERS019754	ERX013857	ERP000436
2154724	ERS019755	ERX013857	ERP000436
2154724	ERS019757	ERX013857	ERP000436
2154724	ERS019760	ERX013860	ERP000436
2154724	ERS019761	ERX013860	ERP000436
2154724	ERS019762	ERX013860	ERP000436
2154724	ERS019766	ERX013860	ERP000436
2154724	ERS019767	ERX013860	ERP000436
2154724	ERS019768	ERX013860	ERP000436
2154724	ERS019773	ERX013856	ERP000436
2154724	ERS019778	ERX013856	ERP000436
2154724	ERS019779	ERX013856	ERP000436
2154724	ERS019785	ERX013858	ERP000436
2154724	ERS020002	ERX014930	ERP000436
2154724	ERS020005	ERX014930	ERP000436
2154724	ERS020012	ERX014928	ERP000436
2154724	ERS020018	ERX014928	ERP000436
2154724	ERS020019	ERX014928	ERP000436
2154724	ERS020026	ERX014930	ERP000436

Drug resistance: Variations known to be annotated as drug resistant are reported in this section of the database. Variations having resistance to known anti-tb drug along with the antibiotic and corresponding resistant gene information is shown in this tab. Peer reviewed reference from which the information was derived is also reported. This tab appears only if the query has a drug resistant mutant associated with it.

tbvar Mycobacterium tuberculosis variantome resource

HOME annoTB MANUAL CONTACT

Search: Variant Location, Range, Gene or Rvid Eureka

Genomic Variations Gene Annotation Functional Effects Regulatory Variations Strain Information **Drug Resistance** Genome Browser

Position	Resistant Drug	Resistant Gene	Reference
2155167	INH	katG	Lipin MY CMI 2007
2155168	INH	katG	Musser JM JID 1996
2155412	INH	katG	Chan RCY JAC 2007
2155732	INH	katG	Chan RCY JAC 2007
2155751	INH	katG	Gagneux S PLOS Path 2006
2155783	INH	katG	Wei CJ AAC 2003

Copyright©2013 CSIR

ncRNA: Variations falling in the non-coding RNA regions of the genome are reported in this tab. This is also an optional tab which appears only if any variation in the query lies in the nc region of the genome.

Genome Browser: A genome browser hosting tracks from TBrowse and other relevant variation information and region surrounding the region of interest is shown in this section. Users can navigate through the genome by clicking and dragging the browser area. New tracks of interest can be added to the browser by a simple drag and drop from the vertical panel on the left. Users can also upload their own tracks of interest to show up in the browser by choosing File -> Open -> Select Files... / Paste remote URLs. Users can upload files from one of these file formats (GFF, BigWig, BAM & BAI). Configuration of the information to show in browser is customizable. Users have an option to either open the file or directly add it as a track.

tbvar Mycobacterium tuberculosis variome resource

IGB

HOME annoTB MANUAL CONTACT

Search: Variant Location, Range, Gene or Rvid Eureka

Genomic Variations Gene Annotation Functional Effects Regulatory Variations Strain Information Drug Resistance Genome Browser

Available Tracks

- 16S_rRNA
- 23S_28S_rRNA
- 591_434_FragA
- 5S_rRNA
- APFY
- Antigen
- cobal
- COG
- CsrB_RsmB_rRNA
- ctRNA
- deg
- DIV_MGDD_133
- DIV_MGDD_BCG

Copyright©2013 CSIR

annoTB

This is novel feature of tbvar, where doctors can simply upload Mtb genomes in simple SNP file format which is used by a parser to map those variations onto variations housed in 'tbvar'. Since tbvar is a comprehensive compendium of known genomic variations in Mtb, most of the variations in clinical sample genome are annotated.

tbvar Mycobacterium tuberculosis variome resource

HOME annoTB MANUAL CONTACT

Batch query:

30944 C T
30954 A T
31468 T C
31468 T C
34063 C T
35063 C A
7585 G C
81549 C G
8428 G T
9557 G C
9597 T C
1417019 C T
3037367 C T
4222628 C T
2154532 G A
2164532 C A

Add/Delete
Submit Query Reset

Simple text input for variation

Example input

Reset button

Upload SNPs

Report summary

A report showing variation annotation and corresponding information is available for doctors to analyze. The report gives a summary of annotated variations and also shows the variations known to confer drug resistance.

Different panels below report summary give information on annotation of individual variation section-wise.

Drug resistance panel lists the variations annotated to be drug resistant.

Drug Resistant Variations

Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	Type	Resistant Drug
7588	G	C	75,14	rrhA	non-synonymous SNV	FLQ
761822	A	G	1,89	rrsB	non-synonymous SNV	SM
1417019	C	T	7,23	rrsA	non-synonymous SNV	EMB
2155168	C	G	12,13	katB	non-synonymous SNV	INH
2155168	C	T	12,13	katB	non-synonymous SNV	INH
2517810	G	A	5,53	katB	non-synonymous SNV	INH
4246871	C	T	8,72	rrsC	non-synonymous SNV	EMB
4241042	A	G	7,02	rrsC	non-synonymous SNV	EMB
4245969	C	T	7,23	rrsA	non-synonymous SNV	EMB
4247730	G	A	1,26	rrsB	non-synonymous SNV	EMB

Deleterious variations panel lists deleterious variations predicted by SIFT which exist in database.

Deleterious Variations						
Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	SIFT Score	
3446	C	T	3.19	refF	0.03	
8452	C	T	7.23	gprA	0.01	
11879	A	G	81.49	Rv0898c	0.04	
13288	G	C	7.23	Rv0814c	0.03	
20544	G	C	1.28	rsdA	0.03	
26347	C	G	7.23	Rv0821c	0	
85883	C	G	7.23	rskA1	0	
84528	T	G	5.53	Rv0875	0	
98968	G	C	7.23	Rv0899	0	
189850	A	G	3.19	PE4	0	
234268	G	T	1.28	Rv0187	0.04	

Non-synonymous and synonymous variations are listed in next two panels.

Non-synonymous Variations						
Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	Type	
7585	G	C	75.95	gprA	nonsynonymous SNV	
8428	G	T	0.21	gprA	nonsynonymous SNV	
30944	C	T	0.21	Rv0826	nonsynonymous SNV	
31468	T	C	0.21	Rv0827	nonsynonymous SNV	
1417019	C	T	8.81	embR	nonsynonymous SNV	
2154532	G	A	0.21	katG	nonsynonymous SNV	
2154730	T	G	0.43	katG	nonsynonymous SNV	
2155412	C	T	1.06	katG	nonsynonymous SNV	
2155783	G	A	0.21	katG	nonsynonymous SNV	
4222628	C	T	0.21	Rv3776	nonsynonymous SNV	

Synonymous Variations						
Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	Type	
9557	G	C	0.21	gprA	synonymous SNV	
81649	C	G	1.70	Rv0872	synonymous SNV	

Regulatory variations in uploaded SNP file matching to those present in database are shown in regulatory variation panel.

Regulatory Variations								
Variant Position	Ref Allele	Alt Allele	Variant Count	Gene	Type	Regulator	Target	
1977	A	G	76.17	-	-	Rv0081	Rv0002	
1977	A	G	76.17	-	-	Rv0081	Rv0001	
1977	A	G	76.17	-	-	Rv0081	Rv0002	
1977	A	G	76.17	-	-	Rv0081	Rv0001	
3446	C	T	3.19	refF	nonsynonymous SNV	Rv0324	Rv0002	
3446	C	T	3.19	refF	nonsynonymous SNV	Rv0324	Rv0003	
4013	T	C	80.85	refF	nonsynonymous SNV	Rv0574	Rv0004	
4013	T	C	80.85	refF	nonsynonymous SNV	Rv0574	Rv0002	

The last panel lists those variations not present in 'tbvar'. Users have an option to submit these novel variations to the database.

Novel Variations		
Variant Position	Ref Allele	Alt Allele
2269	A	G
12697	C	S
13304	G	C
17088	G	C
24692	C	A
24721	A	R
48426	G	G
55208	G	K
55540	A	R

When users press the submit button, a form asking for information on the submitter and submission. By submitting the form users submit the SNP file they uploaded to the server where manual curation is done and the data is included into the database.