ECFG15
ROME • ITALY 2020

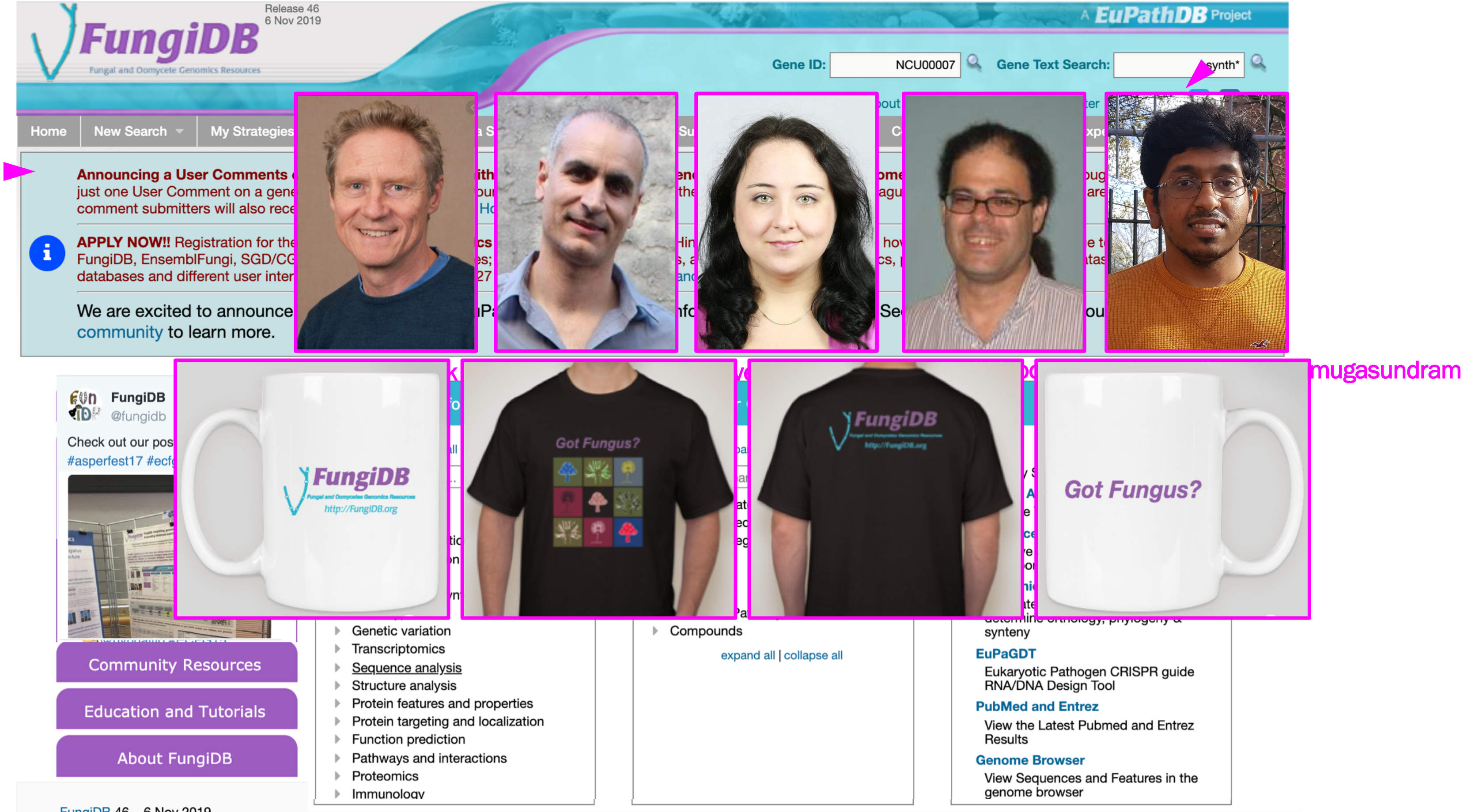
FungiDB
Fungal and Oomycete Genomics Resources

Genomics data, advanced search capabilities, large-scale data analysis & more.

Evelina Basenko
FungiDB Outreach
On behalf of VEuPathDB Team

News and Updates

ECFG15
ROME • ITALY 2020



Release 46
6 Nov 2019

FungiDB
Fungal and Oomycete Genomics Resources

A EuPathDB Project

Gene ID: NCU00007
Gene Text Search: synth*

Home | New Search | My Strategies

Announcing a User Comments
just one User Comment on a gene
comment submitters will also rece

APPLY NOW!! Registration for the
FungiDB, EnsemblFungi, SGD/CG
databases and different user inter

We are excited to announce
community to learn more.

mugasundram

FungiDB
@fungidb
Check out our pos
#asperfest17 #ecf

- Genetic variation
- Transcriptomics
- Sequence analysis
- Structure analysis
- Protein features and properties
- Protein targeting and localization
- Function prediction
- Pathways and interactions
- Proteomics
- Immunology

- Compounds

expand all | collapse all

synteny

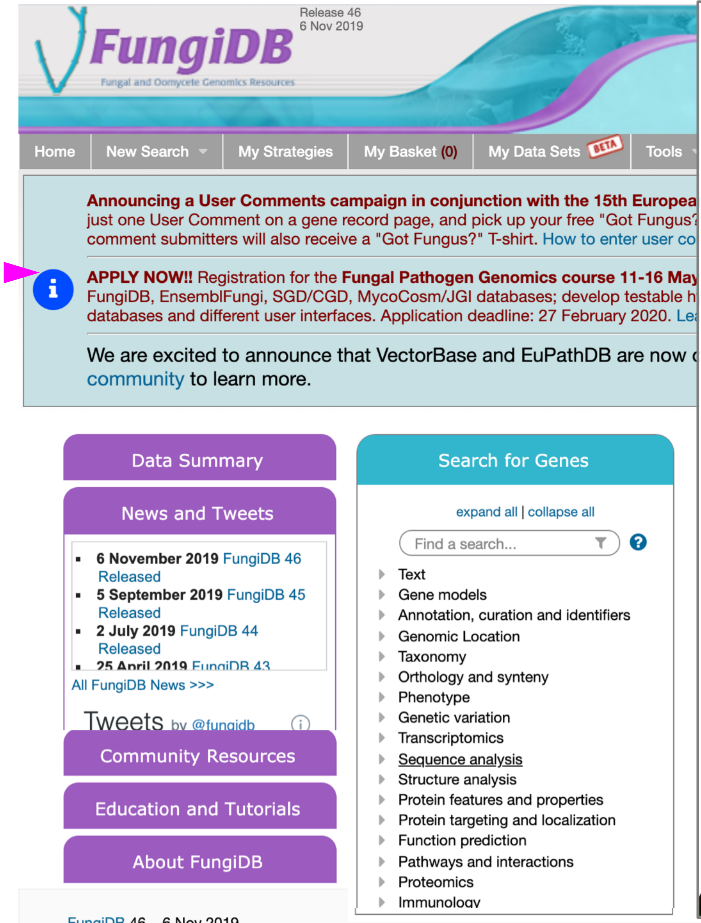
EuPaGDT
Eukaryotic Pathogen CRISPR guide RNA/DNA Design Tool

PubMed and Entrez
View the Latest Pubmed and Entrez Results

Genome Browser
View Sequences and Features in the genome browser

FungiDB 46 6 Nov 2019

News and Updates



Release 46
6 Nov 2019

Announcing a User Comments campaign in conjunction with the 15th Europea just one User Comment on a gene record page, and pick up your free "Got Fungus? comment submitters will also receive a "Got Fungus?" T-shirt. How to enter user co

APPLY NOW!! Registration for the **Fungal Pathogen Genomics course 11-16 May** FungiDB, EnsemblFungi, SGD/CGD, MycoCosm/JGI databases; develop testable h databases and different user interfaces. Application deadline: 27 February 2020. Le

We are excited to announce that VectorBase and EuPathDB are now c community to learn more.

- **6 November 2019** FungiDB 46 Released
- **5 September 2019** FungiDB 45 Released
- **2 July 2019** FungiDB 44 Released
- **25 April 2019** FungiDB 43

All FungiDB News >>>

FUNGAL PATHOGEN GENOMICS

11-16 May 2020
Wellcome Genome Campus, UK

This week-long hands-on course is a collaborative teaching effort between the web-based fungal data mining resources: FungiDB/VEuPathDB, EnsemblFungi, SGD/CGD, MycoCosm/JGI.

Daily activities will include individual and group training exercises, supplementary lectures on bioinformatics techniques and tools used by various databases, and presentations by distinguished guest speakers.

PROGRAMME:

- Comparative genomics, gene trees, whole-genome alignment
- Identification of orthologs and orthology-based inference
- Genome browsers and gene pages
- RNA-Seq analysis and visualization in VEuPathDB Galaxy
- Variant calling analysis and Ensembl Variant Effect Predictor (VEP) tool
- Development of advanced biologically relevant queries using FungiDB 'search strategies' and mining integrated datasets
- Genetic interactions, virulence genes, secondary metabolites
- Overview of ontology structure, evidence, available tools, slimming and enrichment
- Introduction to annotation and curation of fungal genomes

INSTRUCTORS:

David Roos University of Pennsylvania, USA
Nishadi De Silva EMBL-EBI, UK
Omar Harb University of Pennsylvania, USA
Evelina Basenko University of Liverpool, UK
Marek Skrzypek Stanford University, USA
Michal Szpak EMBL-EBI, UK
Kathryn Crouch University of Glasgow, UK
Stephen Mondo University of California, USA

GUEST SPEAKERS:

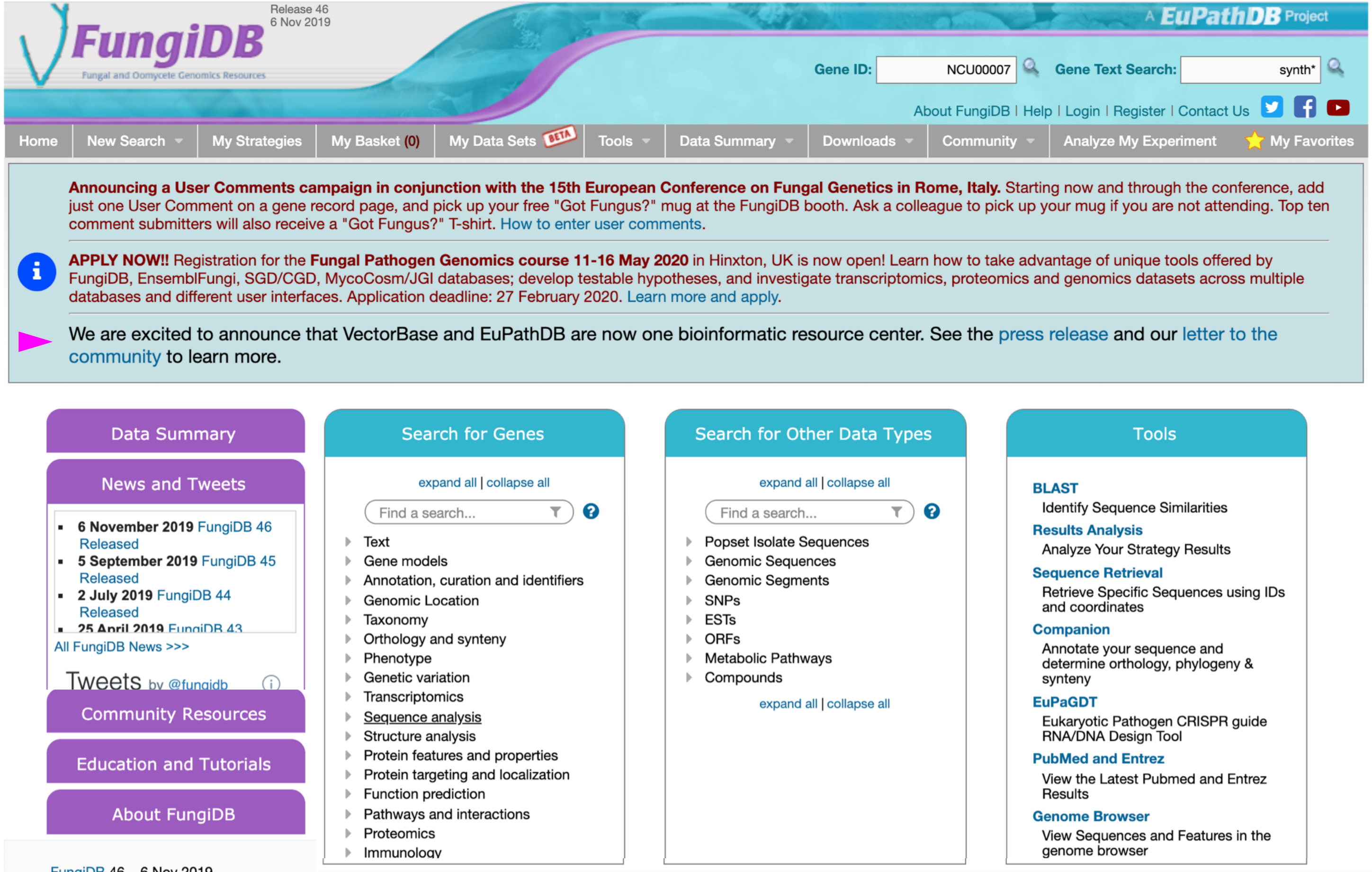
Lynne Boddy Cardiff University, UK
Jo Dicks Quadram Institute, UK
Johanna Rhodes Imperial College London, UK

The course fee is subsidised for non-commercial applicants and limited bursaries are available. Please see the course website for full details.

APPLICATION & BURSARY DEADLINE
27 February 2020

Full details at: bit.ly/FPG20_P

News and Updates



Announcing a User Comments campaign in conjunction with the 15th European Conference on Fungal Genetics in Rome, Italy. Starting now and through the conference, add just one User Comment on a gene record page, and pick up your free "Got Fungus?" mug at the FungiDB booth. Ask a colleague to pick up your mug if you are not attending. Top ten comment submitters will also receive a "Got Fungus?" T-shirt. How to enter user comments.

APPLY NOW!! Registration for the **Fungal Pathogen Genomics course 11-16 May 2020** in Hinxton, UK is now open! Learn how to take advantage of unique tools offered by FungiDB, EnsemblFungi, SGD/CGD, MycoCosm/JGI databases; develop testable hypotheses, and investigate transcriptomics, proteomics and genomics datasets across multiple databases and different user interfaces. Application deadline: 27 February 2020. Learn more and apply.

We are excited to announce that VectorBase and EuPathDB are now one bioinformatic resource center. See the press release and our letter to the community to learn more.

- **6 November 2019** FungiDB 46 Released
- **5 September 2019** FungiDB 45 Released
- **2 July 2019** FungiDB 44 Released
- **25 April 2019** FungiDB 43

BLAST
Identify Sequence Similarities

Results Analysis
Analyze Your Strategy Results

Sequence Retrieval
Retrieve Specific Sequences using IDs and coordinates

Companion
Annotate your sequence and determine orthology, phylogeny & synteny

EuPaGDT
Eukaryotic Pathogen CRISPR guide RNA/DNA Design Tool

PubMed and Entrez
View the Latest Pubmed and Entrez Results

Genome Browser
View Sequences and Features in the genome browser

FungiDB 46   6 Nov 2019

VEuPathDB.org
Eukaryotic Pathogen, Vector & Host Informatics Resources

HostDB

ClinEpiDB
Clinical Epidemiology Resources

MicrobiomeDB

NIH National Institute of Allergy and Infectious Diseases

BILL & MELINDA GATES *foundation*

wellcome

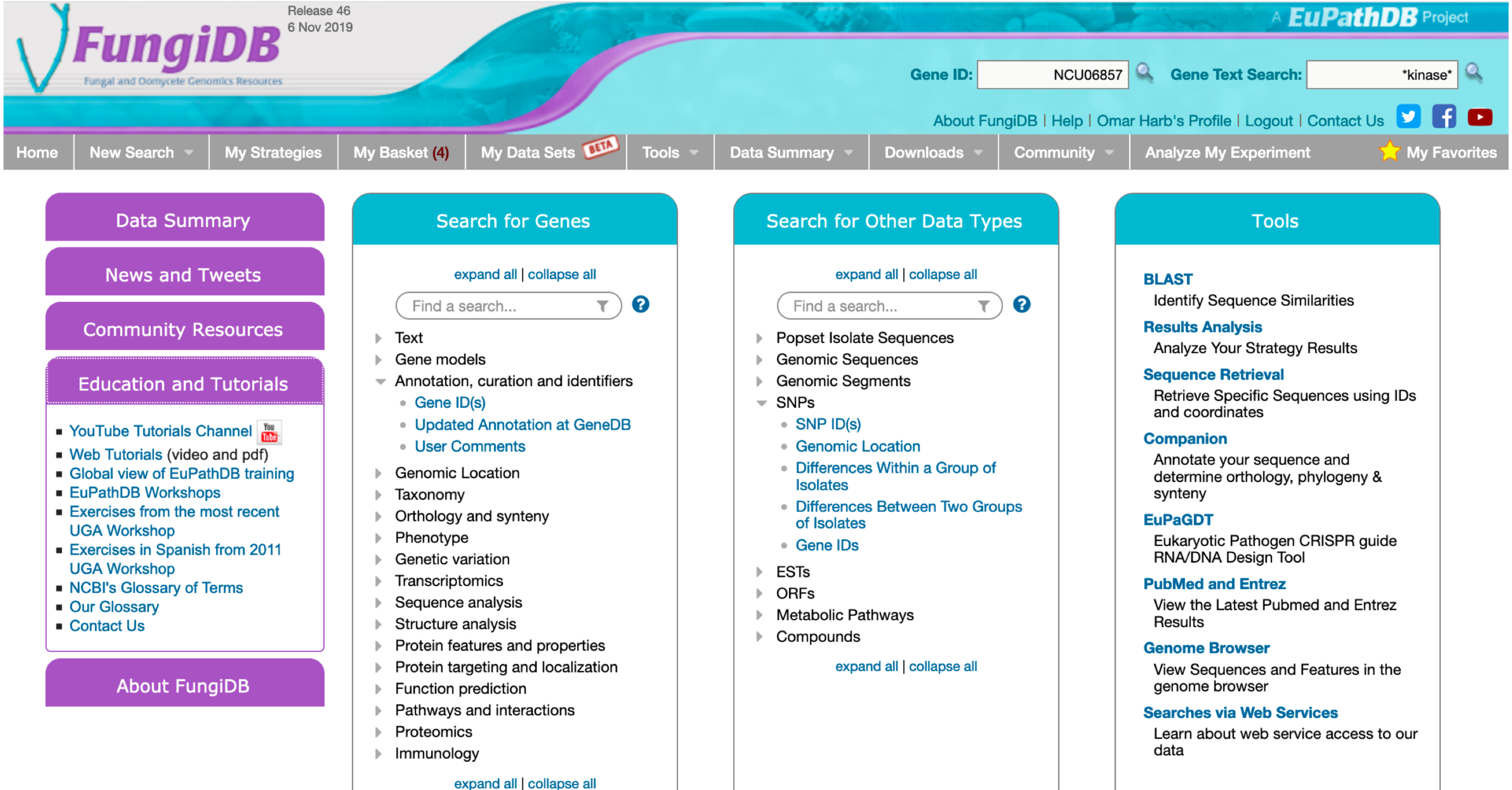
Our name has changed but the philosophy behind databases remains the same!

- ***Curated annotation, integrated with automated analyses.*** Predicted genes, proteins, motifs, structures, functions, pathways, *etc* … for both microbes and their host cells
- ***Incorporation & integration of diverse data types, including:*** genome sequences, variation data, field & clinical isolates (w/ extensive metadata on phenotypes & hosts/pathogen interactions), chromatin marks, transcriptomic & proteomic profiling (multiple platforms), interactomes, structural genomics, metabolomics, comparative genomics & orthology-based inference, *etc.*
- ***Encyclopedic gene & other record pages, Genome Browser, data visualization tools, etc.:*** data mining, result analysis (e.g. GO term enrichment), and private workspaces in Galaxy & FungiDB.
- ***Sustainable resources, universal global access.*** 7M hits/mo, from 63K unique users/mo (100+ countries), returning > twice monthly; >2Tb/yr downloads, >14K citations; extensive scientific outreach & education activities

❖ ***Enable researchers to ask their own questions!***

FungiDB key features & tools

ECFG15
ROME • ITALY 2020



Release 46
6 Nov 2019

Gene ID: NCU06857

Gene Text Search: *kinase*

Data Summary

News and Tweets

Community Resources

Education and Tutorials

- YouTube Tutorials Channel
- Web Tutorials (video and pdf)
- Global view of EuPathDB training
- EuPathDB Workshops
- Exercises from the most recent UGA Workshop
- Exercises in Spanish from 2011 UGA Workshop
- NCBI's Glossary of Terms
- Our Glossary
- Contact Us

About FungiDB

Search for Genes

expand all | collapse all

- Text
- Gene models
- Annotation, curation and identifiers
  - Gene ID(s)
  - Updated Annotation at GeneDB
  - User Comments
- Genomic Location
- Taxonomy
- Orthology and synteny
- Phenotype
- Genetic variation
- Transcriptomics
- Sequence analysis
- Structure analysis
- Protein features and properties
- Protein targeting and localization
- Function prediction
- Pathways and interactions
- Proteomics
- Immunology

expand all | collapse all

Search for Other Data Types

expand all | collapse all

- Popset Isolate Sequences
- Genomic Sequences
- Genomic Segments
- SNPs
  - SNP ID(s)
  - Genomic Location
  - Differences Within a Group of Isolates
  - Differences Between Two Groups of Isolates
  - Gene IDs
- ESTs
- ORFs
- Metabolic Pathways
- Compounds

expand all | collapse all

Tools

BLAST
Identify Sequence Similarities

Results Analysis
Analyze Your Strategy Results

Sequence Retrieval
Retrieve Specific Sequences using IDs and coordinates

Companion
Annotate your sequence and determine orthology, phylogeny & synteny

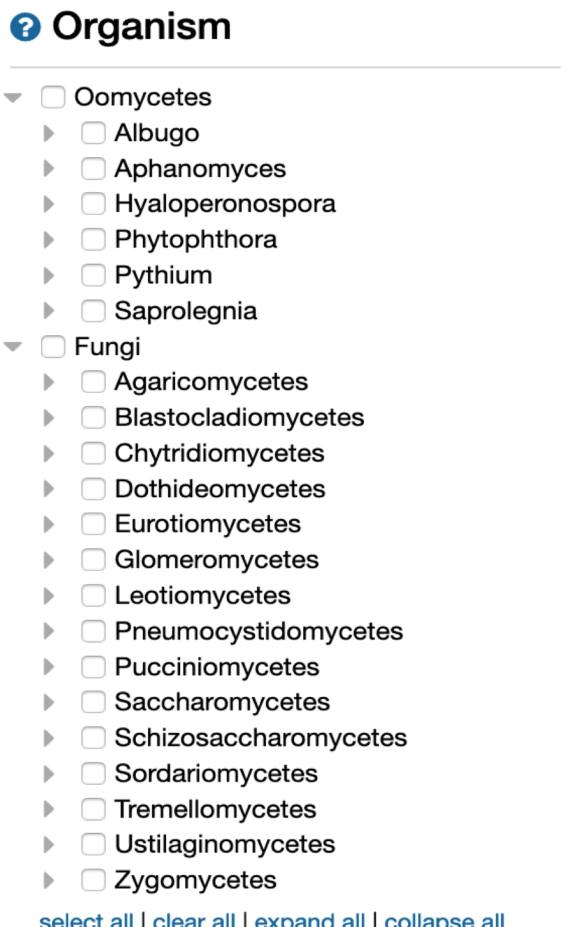
EuPaGDT
Eukaryotic Pathogen CRISPR guide RNA/DNA Design Tool

PubMed and Entrez
View the Latest Pubmed and Entrez Results

Genome Browser
View Sequences and Features in the genome browser

Searches via Web Services
Learn about web service access to our data

Supporting 148 organisms, 164 genomes & integrating omics scale datasets & more



Organism

- ☐ Oomycetes
  - ☐ Albugo
  - ☐ Aphanomyces
  - ☐ Hyaloperonospora
  - ☐ Phytophthora
  - ☐ Pythium
  - ☐ Saprolegnia
- ☐ Fungi
  - ☐ Agaricomycetes
  - ☐ Blastocladiomycetes
  - ☐ Chytridiomycetes
  - ☐ Dothideomycetes
  - ☐ Eurotiomycetes
  - ☐ Glomeromycetes
  - ☐ Leotiomycetes
  - ☐ Pneumocystidomycetes
  - ☐ Pucciniomycetes
  - ☐ Saccharomycetes
  - ☐ Schizosaccharomycetes
  - ☐ Sordariomycetes
  - ☐ Tremellomycetes
  - ☐ Ustilaginomycetes
  - ☐ Zygomycetes

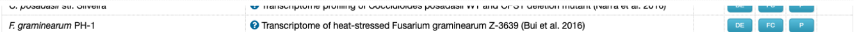
select all | clear all | expand all | collapse all

Identify Genes based on Phenotype Evidence

Filter Data Sets: Type keyword(s) to filter

Legend: CP Curated Phenotype PT Phenotype Text

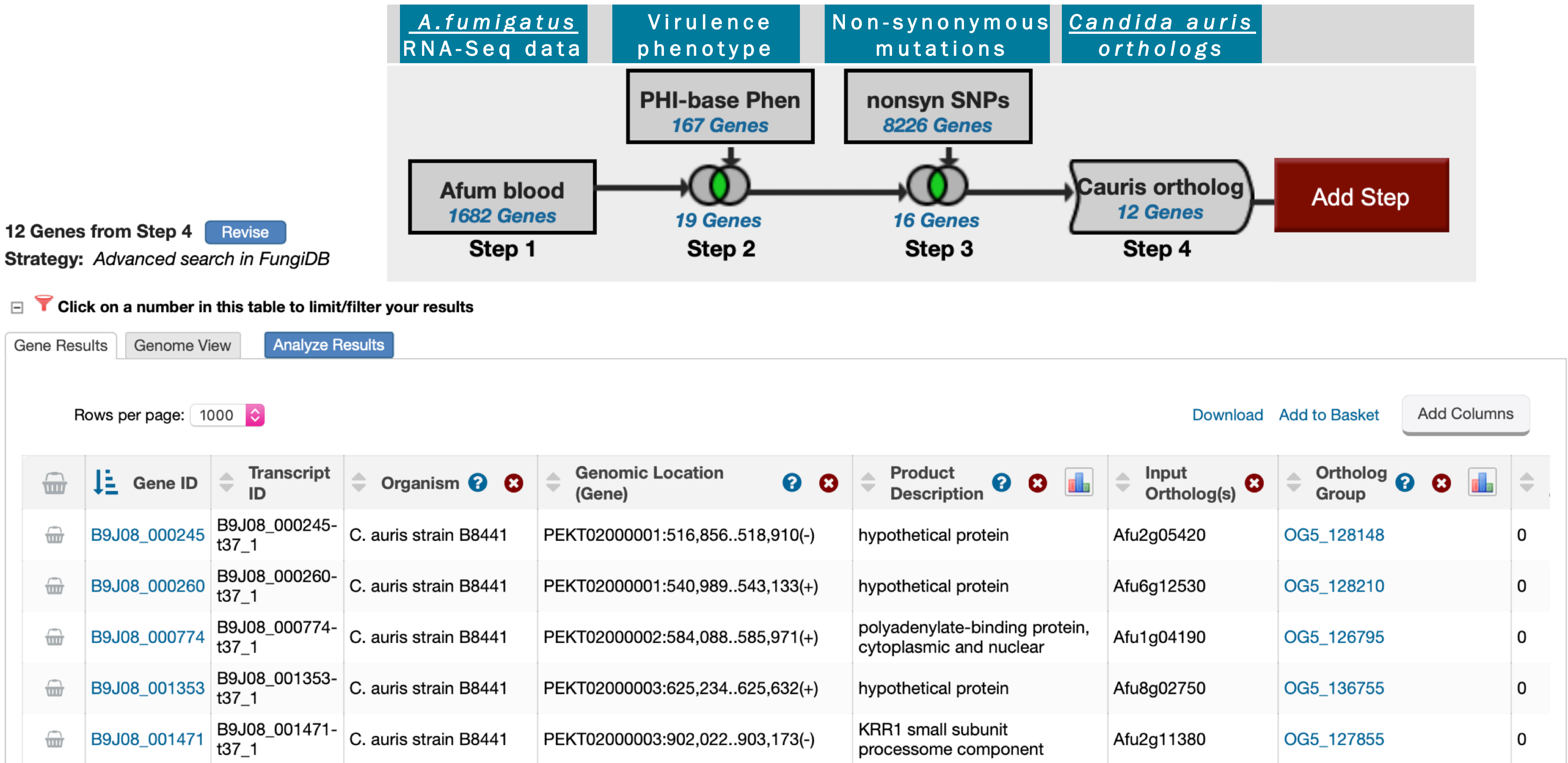
Organism	Data Set	Choose a search	
A. clavatus NRRL 1 *A. flavus* NRRL3357 *A. fumigatus* A1163 *A. fumigatus* Af293 *A. nidulans* FGSC A4 *A. oryzae* RIB40 *B. cinerea* B05.10 *C. albicans* SC5314 *C. albicans* WO-1 *C. deuterogattii* R265 *C. glabrata* CBS 138 *C. lusitaniae* ATCC 42720 *C. neoformans* var. grubii H99 *C. neoformans* var. neoformans B-3501A *C. neoformans* var. neoformans JEC21 *C. parapsilosis* CDC317 *C. posadasii* C735 delta SOWgp *C. tropicalis* MYA-3404 *F. fujikuroi* IMI 58289 *F. graminearum* PH-1 *F. oxysporum* f. sp. cubense tropical race 4 54006 *F. oxysporum* f. sp. lycopersici 4287 *F. oxysporum* f. sp. melonis 26406 *F. verticillioides* 7600 *H. arabidopsidis* Emoy2 *H. capsulatum* G186AR *M. larici-populina* 98AG31 *M. oryzae* 70-15 *P. graminis* f. sp. tritici CRL 75-36-700-3 *P. infestans* T30-4 *P. sojae* strain P6497 *R. delemar* RA 99-880 *S. cerevisiae* S288c *S. sclerotiorum* 1980 UF-70 *T. virens* Gv29-8 *U. maydis* 521	? PHI-base curated phenotypes (PHI-base database) NEW	CP	
A. fumigatus Af293 *A. nidulans* FGSC A4 *A. niger* CBS 513.88 *A. oryzae* RIB40	? Manually Curated *Aspergillus* Phenotypes (Eukaryotic Pathogen Database Resources (EuPathDB)) NEW	CP	PT
C. gattii WM276 *C. neoformans* var. grubii H99 *C. neoformans* var. neoformans JEC21	? Manually Curated *Cryptococcus* Phenotypes (Eukaryotic Pathogen Database Resources (EuPathDB))	CP	PT
F. graminearum PH-1	? Manually Curated *Fusarium* Phenotypes (Eukaryotic Pathogen Database Resources (EuPathDB)) NEW	CP	PT
N. crassa OR74A	? Phenotypic analysis of Neurospora crassa knockout mutants (Borkovich et al.)	CP	
N. crassa OR74A	? Neurospora Genome Project Phenotype Image Collection (Dunlap et al.)	CP	



F. graminearum PH-1	? Transcriptome of heat-stressed Fusarium graminearum Z-3639 (Bui et al. 2016)	DE	FC	P

Easy-to-use search strategy system can be used to create advanced searches across species

ECFG15
ROME • ITALY 2020



12 Genes from Step 4

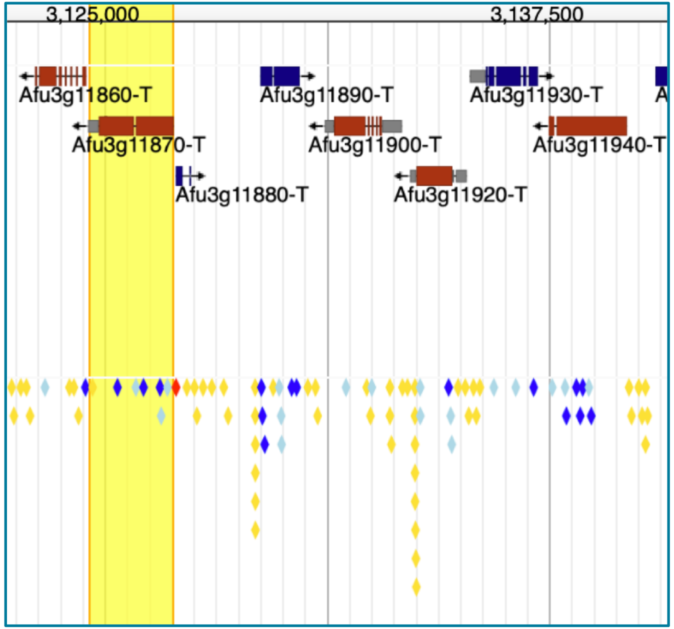
Strategy: *Advanced search in FungiDB*

Rows per page: 1000

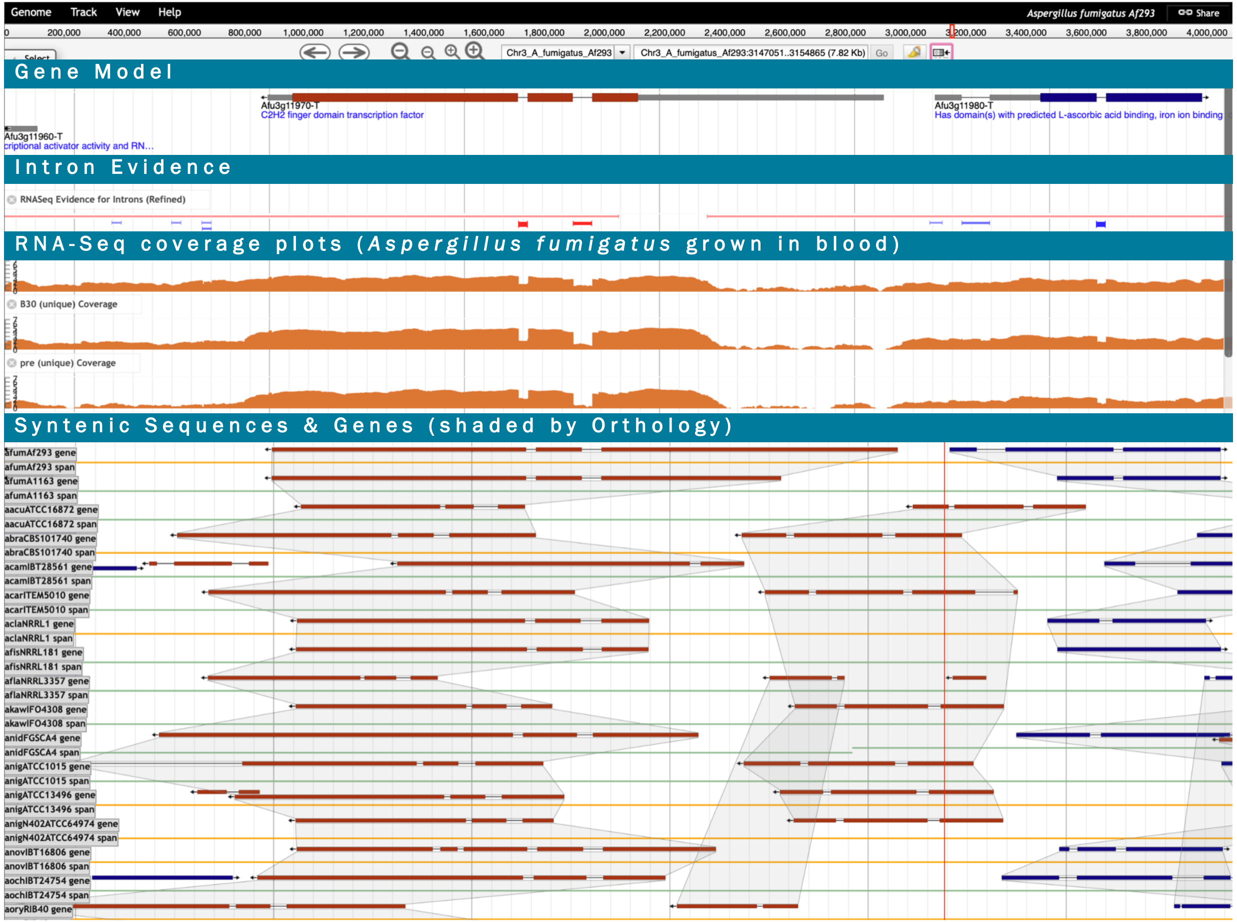
	Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description	Input Ortholog(s)	Ortholog Group	
	B9J08_000245	B9J08_000245-t37_1	C. auris strain B8441	PEKT02000001:516,856..518,910(-)	hypothetical protein	Afu2g05420	OG5_128148	0
	B9J08_000260	B9J08_000260-t37_1	C. auris strain B8441	PEKT02000001:540,989..543,133(+)	hypothetical protein	Afu6g12530	OG5_128210	0
	B9J08_000774	B9J08_000774-t37_1	C. auris strain B8441	PEKT02000002:584,088..585,971(+)	polyadenylate-binding protein, cytoplasmic and nuclear	Afu1g04190	OG5_126795	0
	B9J08_001353	B9J08_001353-t37_1	C. auris strain B8441	PEKT02000003:625,234..625,632(+)	hypothetical protein	Afu8g02750	OG5_136755	0
	B9J08_001471	B9J08_001471-t37_1	C. auris strain B8441	PEKT02000003:902,022..903,173(-)	KRR1 small subunit processome component	Afu2g11380	OG5_127855	0

JBrowse: Access gene model, transcriptomics, population diversity data, orthology & more

ECFG15
ROME • ITALY 2020



- ◆ Non-coding
- ◆ Non-synonymous
- ◆ Synonymous
- ◆ Nonsense



Gene Model

Intron Evidence

RNA-Seq coverage plots (*Aspergillus fumigatus* grown in blood)

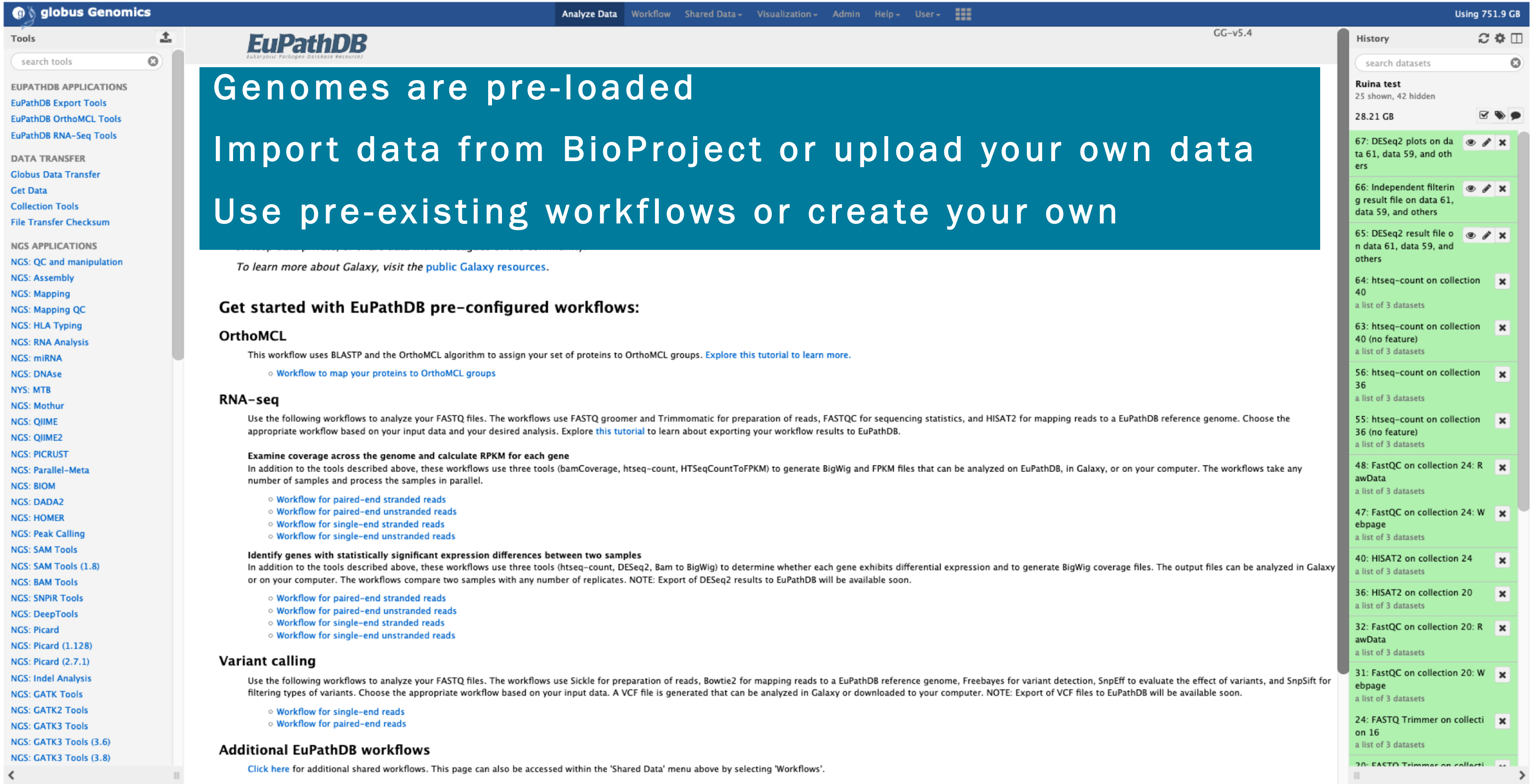
Syntenic Sequences & Genes (shaded by Orthology)

VEuPathDB Galaxy offers private workspaces

Genomes are pre-loaded

Import data from BioProject or upload your own data

Use pre-existing workflows or create your own



To learn more about Galaxy, visit the public Galaxy resources.

Get started with EuPathDB pre-configured workflows:

OrthoMCL

This workflow uses BLASTP and the OrthoMCL algorithm to assign your set of proteins to OrthoMCL groups. Explore this tutorial to learn more.

- Workflow to map your proteins to OrthoMCL groups

RNA-seq

Use the following workflows to analyze your FASTQ files. The workflows use FASTQ groomer and Trimmomatic for preparation of reads, FASTQC for sequencing statistics, and HISAT2 for mapping reads to a EuPathDB reference genome. Choose the appropriate workflow based on your input data and your desired analysis. Explore this tutorial to learn about exporting your workflow results to EuPathDB.

Examine coverage across the genome and calculate RPKM for each gene
In addition to the tools described above, these workflows use three tools (bamCoverage, htseq-count, HTSeqCountToFPKM) to generate BigWig and FPKM files that can be analyzed on EuPathDB, in Galaxy, or on your computer. The workflows take any number of samples and process the samples in parallel.

- Workflow for paired-end stranded reads
- Workflow for paired-end unstranded reads
- Workflow for single-end stranded reads
- Workflow for single-end unstranded reads

Identify genes with statistically significant expression differences between two samples
In addition to the tools described above, these workflows use three tools (htseq-count, DESeq2, Bam to BigWig) to determine whether each gene exhibits differential expression and to generate BigWig coverage files. The output files can be analyzed in Galaxy or on your computer. The workflows compare two samples with any number of replicates. NOTE: Export of DESeq2 results to EuPathDB will be available soon.

- Workflow for paired-end stranded reads
- Workflow for paired-end unstranded reads
- Workflow for single-end stranded reads
- Workflow for single-end unstranded reads

Variant calling

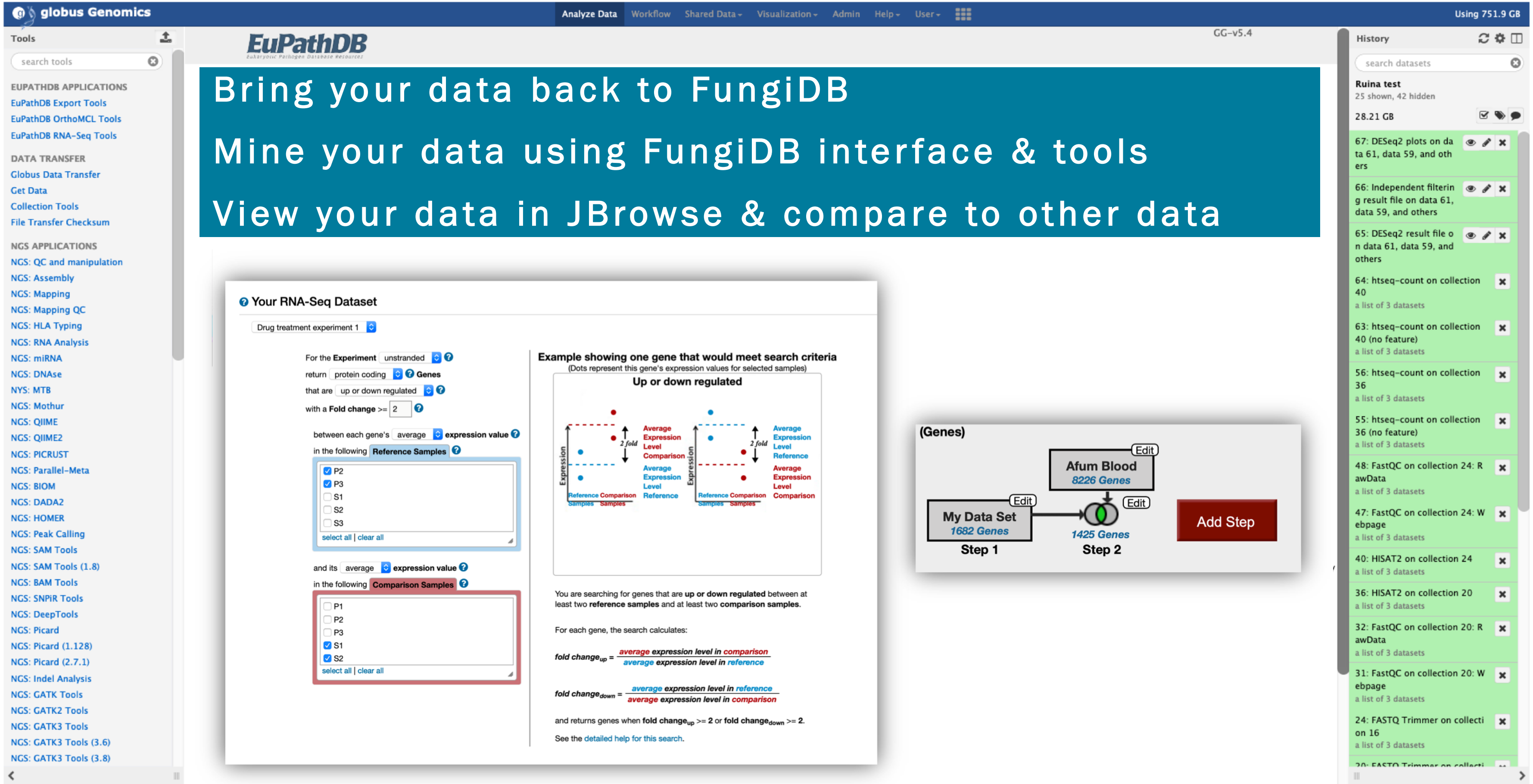
Use the following workflows to analyze your FASTQ files. The workflows use Sickle for preparation of reads, Bowtie2 for mapping reads to a EuPathDB reference genome, Freebayes for variant detection, SnpEff to evaluate the effect of variants, and SnpSift for filtering types of variants. Choose the appropriate workflow based on your input data. A VCF file is generated that can be analyzed in Galaxy or downloaded to your computer. NOTE: Export of VCF files to EuPathDB will be available soon.

- Workflow for single-end reads
- Workflow for paired-end reads

Additional EuPathDB workflows

Click here for additional shared workflows. This page can also be accessed within the 'Shared Data' menu above by selecting 'Workflows'.

Import Galaxy results into private My Data Sets in FungiDB



Bring your data back to FungiDB

Mine your data using FungiDB interface & tools

View your data in JBrowse & compare to other data

EUPATHDB APPLICATIONS
- EuPathDB Export Tools
- EuPathDB OrthoMCL Tools
- EuPathDB RNA-Seq Tools

DATA TRANSFER
- Globus Data Transfer
- Get Data
- Collection Tools
- File Transfer Checksum

NGS APPLICATIONS
- NGS: QC and manipulation
- NGS: Assembly
- NGS: Mapping
- NGS: Mapping QC
- NGS: HLA Typing
- NGS: RNA Analysis
- NGS: miRNA
- NGS: DNAse
- NYS: MTB
- NGS: Mothur
- NGS: QIIME
- NGS: QIIME2
- NGS: PICRUST
- NGS: Parallel-Meta
- NGS: BIOM
- NGS: DADA2
- NGS: HOMER
- NGS: Peak Calling
- NGS: SAM Tools
- NGS: SAM Tools (1.8)
- NGS: BAM Tools
- NGS: SNPiR Tools
- NGS: DeepTools
- NGS: Picard
- NGS: Picard (1.128)
- NGS: Picard (2.7.1)
- NGS: Indel Analysis
- NGS: GATK Tools
- NGS: GATK2 Tools
- NGS: GATK3 Tools
- NGS: GATK3 Tools (3.6)
- NGS: GATK3 Tools (3.8)

Your RNA-Seq Dataset

Drug treatment experiment 1

For the **Experiment** unstranded

return protein coding **Genes**

that are up or down regulated

with a **Fold change** >= 2

between each gene's average **expression value**

in the following **Reference Samples**

☑ P2
☑ P3
☐ S1
☐ S2
☐ S3

select all | clear all

and its average **expression value**

in the following **Comparison Samples**

☐ P1
☐ P2
☐ P3
☑ S1
☑ S2

select all | clear all

Example showing one gene that would meet search criteria
(Dots represent this gene's expression values for selected samples)

Up or down regulated

You are searching for genes that are **up or down regulated** between at least two **reference samples** and at least two **comparison samples.**

For each gene, the search calculates:

$$\text{fold change}_{up} = \frac{\text{average expression level in comparison}}{\text{average expression level in reference}}$$

$$\text{fold change}_{down} = \frac{\text{average expression level in reference}}{\text{average expression level in comparison}}$$

and returns genes when **fold change$_{up}$** >= 2 or **fold change$_{down}$** >= 2.

See the detailed help for this search.

(Genes)

Ruina test
25 shown, 42 hidden

28.21 GB

67: DESeq2 plots on data 61, data 59, and others

66: Independent filtering result file on data 61, data 59, and others

65: DESeq2 result file on data 61, data 59, and others

64: htseq-count on collection 40
a list of 3 datasets

63: htseq-count on collection 40 (no feature)
a list of 3 datasets

56: htseq-count on collection 36
a list of 3 datasets

55: htseq-count on collection 36 (no feature)
a list of 3 datasets

48: FastQC on collection 24: RawData
a list of 3 datasets

47: FastQC on collection 24: Webpage
a list of 3 datasets

40: HISAT2 on collection 24
a list of 3 datasets

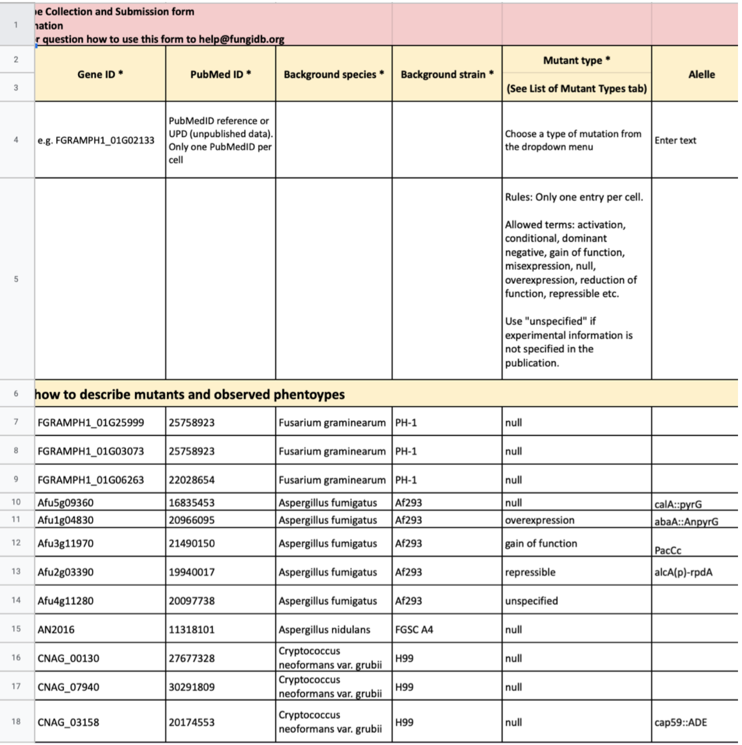
36: HISAT2 on collection 20
a list of 3 datasets

32: FastQC on collection 20: RawData
a list of 3 datasets

31: FastQC on collection 20: Webpage
a list of 3 datasets

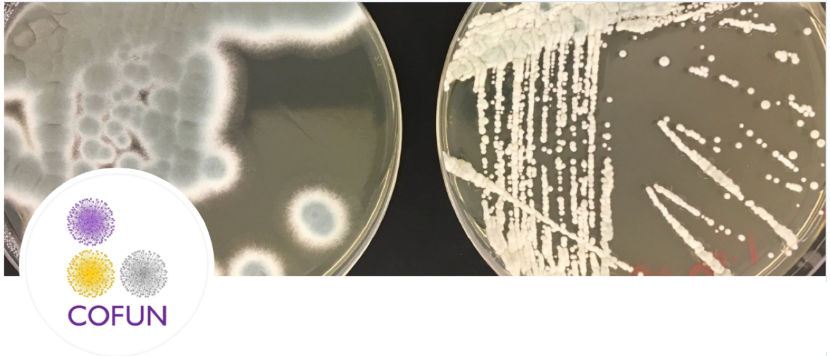
24: FASTQ Trimmer on collection 16
a list of 3 datasets

Phenotype data: using ontologies to create searchable datasets



1	ɪe Collection and Submission form ɪation r question how to use this form to help@fungidb.org					
2	Gene ID *	PubMed ID *	Background species *	Background strain *	Mutant type *	Alelle
3					(See List of Mutant Types tab)	
4	e.g. FGRAMPH1_01G02133	PubMedID reference or UPD (unpublished data). Only one PubMedID per cell			Choose a type of mutation from the dropdown menu	Enter text
5					Rules: Only one entry per cell. Allowed terms: activation, conditional, dominant negative, gain of function, misexpression, null, overexpression, reduction of function, repressible etc. Use "unspecified" if experimental information is not specified in the publication.	
6	how to describe mutants and observed phentoypes					
7	FGRAMPH1_01G25999	25758923	Fusarium graminearum	PH-1	null	
8	FGRAMPH1_01G03073	25758923	Fusarium graminearum	PH-1	null	
9	FGRAMPH1_01G06263	22028654	Fusarium graminearum	PH-1	null	
10	Afu5g09360	16835453	Aspergillus fumigatus	Af293	null	calA::pyrG
11	Afu1g04830	20966095	Aspergillus fumigatus	Af293	overexpression	abaA::AnpyrG
12	Afu3g11970	21490150	Aspergillus fumigatus	Af293	gain of function	PacCc
13	Afu2g03390	19940017	Aspergillus fumigatus	Af293	repressible	alcA(p)-rpdA
14	Afu4g11280	20097738	Aspergillus fumigatus	Af293	unspecified	
15	AN2016	11318101	Aspergillus nidulans	FGSC A4	null	
16	CNAG_00130	27677328	Cryptococcus neoformans var. grubii	H99	null	
17	CNAG_07940	30291809	Cryptococcus neoformans var. grubii	H99	null	
18	CNAG_03158	20174553	Cryptococcus neoformans var. grubii	H99	null	cap59::ADE

Coming soon..



Create or improve gene models with Apollo

Create new gene models

Import your own evidence

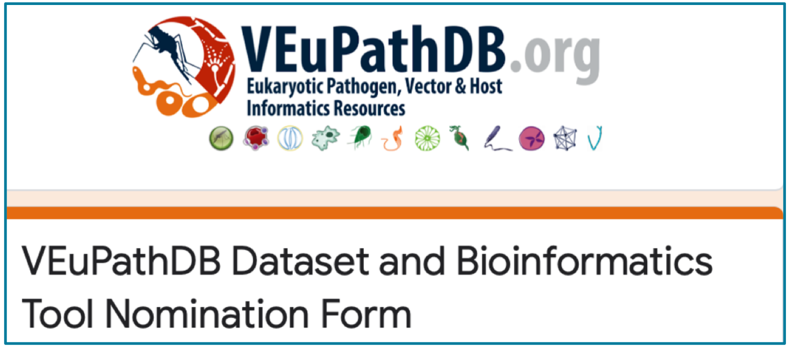
Engaging the community via forums, dataset & tools nomination forms

Would you like to join a community forum?

Email help@fungidb.org

Nominate datasets for integration (e.g. genomes, RNA-Seq etc.)

https://tinyurl.com/nominatedataset



VEuPathDB.org
Eukaryotic Pathogen, Vector & Host Informatics Resources

VEuPathDB Dataset and Bioinformatics Tool Nomination Form

Flyer at the booth

Thank You!

ECFG15
ROME • ITALY 2020

VEuPathDB team

VEuPathDB.org
Eukaryotic Pathogen, Vector & Host
Informatics Resources