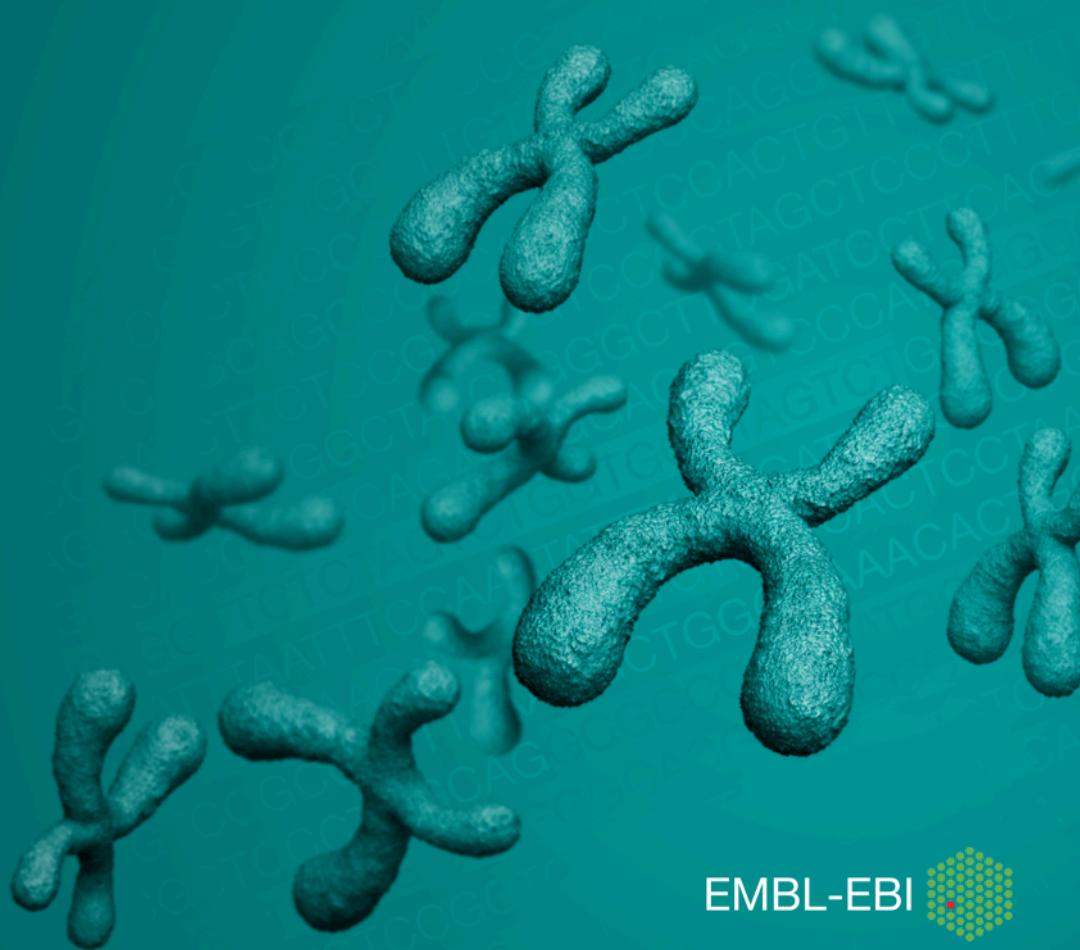


# The European Variation Archive at EMBL-EBI: a home for plant variation data

Gary Saunders, PhD

[www.ebi.ac.uk/eva](http://www.ebi.ac.uk/eva)

[eva-helpdesk@ebi.ac.uk](mailto:eva-helpdesk@ebi.ac.uk)

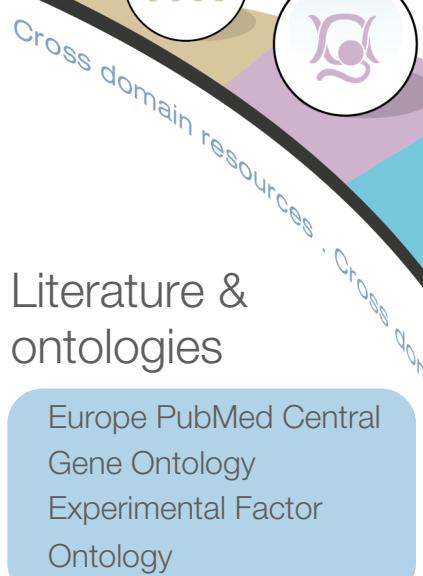


# Genes, genomes

European Nucleotide Archive  
1000 Genomes

Ensembl  
Ensembl Genomes

European Genome-phenome Archive  
Metagenomics portal



## Literature & ontologies

Europe PubMed Central  
Gene Ontology  
Experimental Factor Ontology

## Reactions, interactions & pathways

IntAct

Reactome

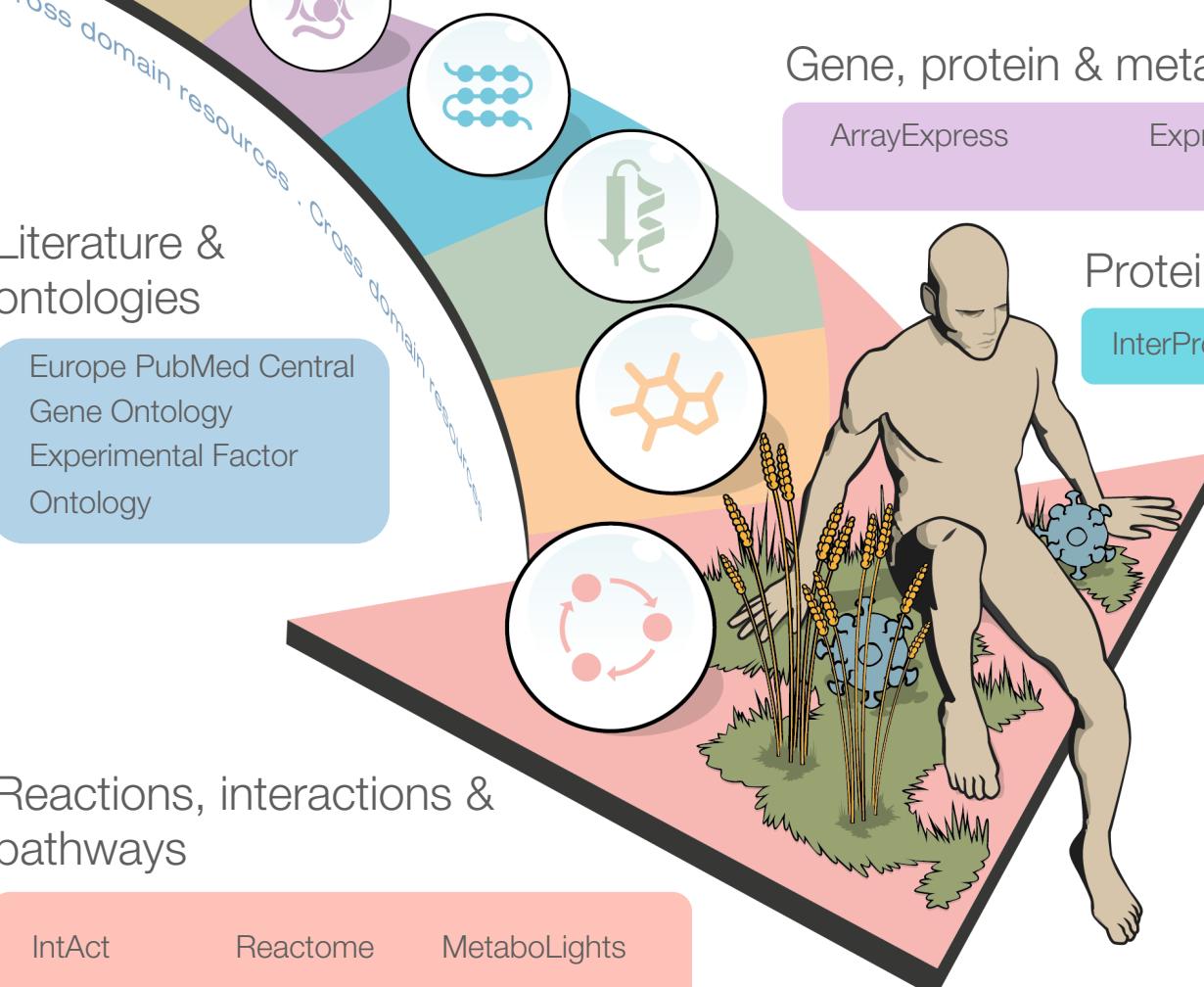
MetaboLights

## Gene, protein & metabolite expression

ArrayExpress

Expression Atlas

Metabolights  
PRIDE



## Protein sequences, families & motifs

InterPro

Pfam

UniProt

## Molecular structures

Protein Data Bank in Europe  
Electron Microscopy Data Bank

## Chemical biology

ChEMBL

ChEBI

## Systems

BioModels  
Enzyme Portal

BioSamples

# Genes, genomes & variation

European Nucleotide Archive  
1000 Genomes

Ensembl  
Ensembl Genomes

European Genome-phenome Archive  
Metagenomics portal

**European Variation Archive**



## Literature & ontologies

Europe PubMed Central  
Gene Ontology  
Experimental Factor  
Ontology

## Reactions, interactions & pathways

IntAct  
Reactome

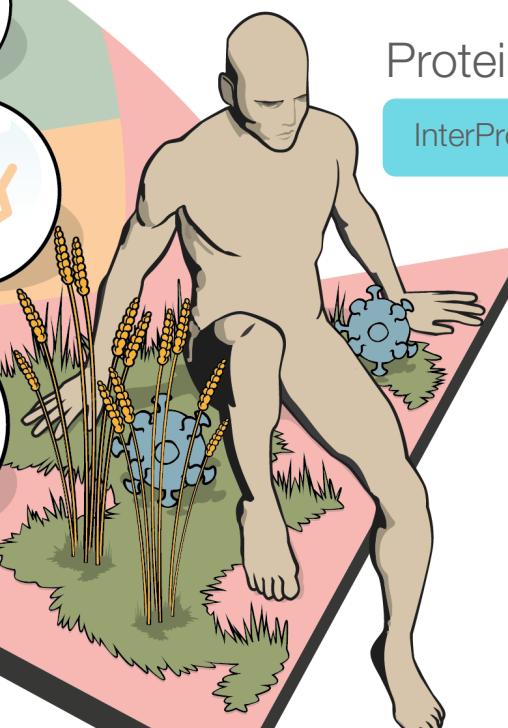
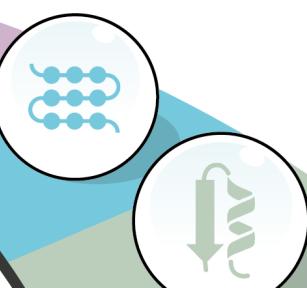
MetaboLights

## Gene, protein & metabolite expression

ArrayExpress

Expression Atlas

Metabolights  
PRIDE



## Protein sequences, families & motifs

InterPro  
Pfam  
UniProt

## Molecular structures

Protein Data Bank in Europe  
Electron Microscopy Data Bank

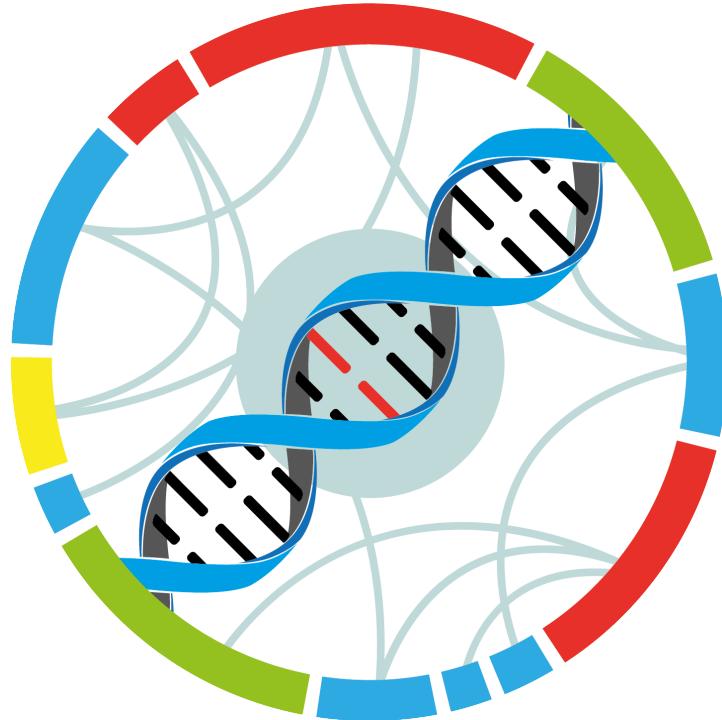
## Chemical biology

ChEMBL  
ChEBI

## Systems

BioModels  
Enzyme Portal  
BioSamples

# European Variation Archive – EVA (Eva)



- Submission based data sharing & analysis platform
- All types of variation:
- SNVs, MNVs, small indels and structural variation
- Germ line, somatic, within / cross population, potentially between species

Any variation represented as a change against a reference genome sequence

# European Variation Archive – EVA (Eva)



- Why archive VCF files?
- Administer file accessions
  - papers
  - share between researchers / labs
  - stable
- Administer variant accessions
  - TransPlant accessions
  - cluster variants based on position
  - stable

Sharing VCF files, and variant data, from a single resource

# European Variation Archive

Accept and Validate Submissions

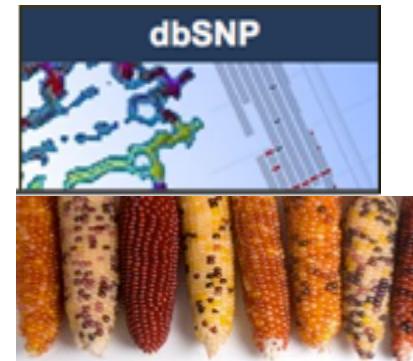
Support dataflow from and to other EBI resources

Provide stable accessions



Provide direct data access

Data exchange with peer archives



(dbVar)

# European Variation Archive

Accept and Validate Submissions

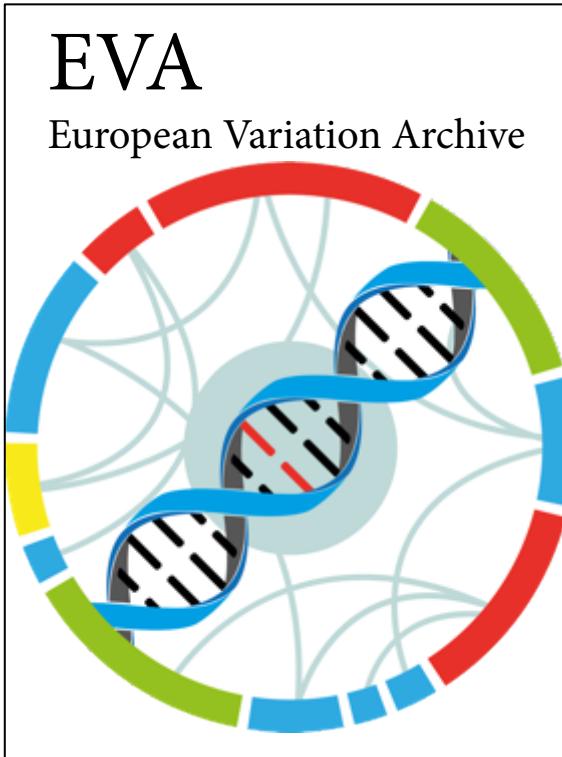
Data mining

Visualization

Support dataflow from and to other EBI resources

Databases

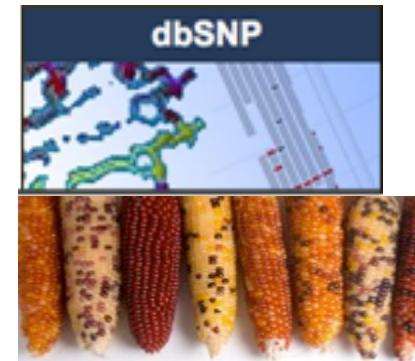
Provide stable accessions



Curation  
Provide direct data access

Ontologies

Data exchange with peer archives



(dbVar)

# EVA Is A Collaborative Archive

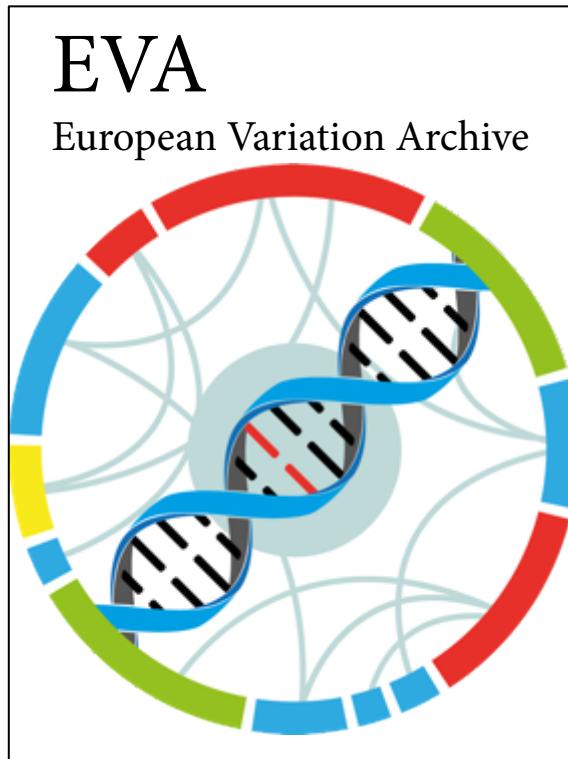
SEUVADIS  
CONSORTIUM

DGVa<sup>r</sup>chive

Direct  
Submissions



Europe PubMed  
Central



European  
ga  
enome-phenome  
archive

EMBL

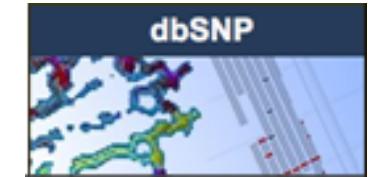
e!

transPLANT

Gramene

LRG

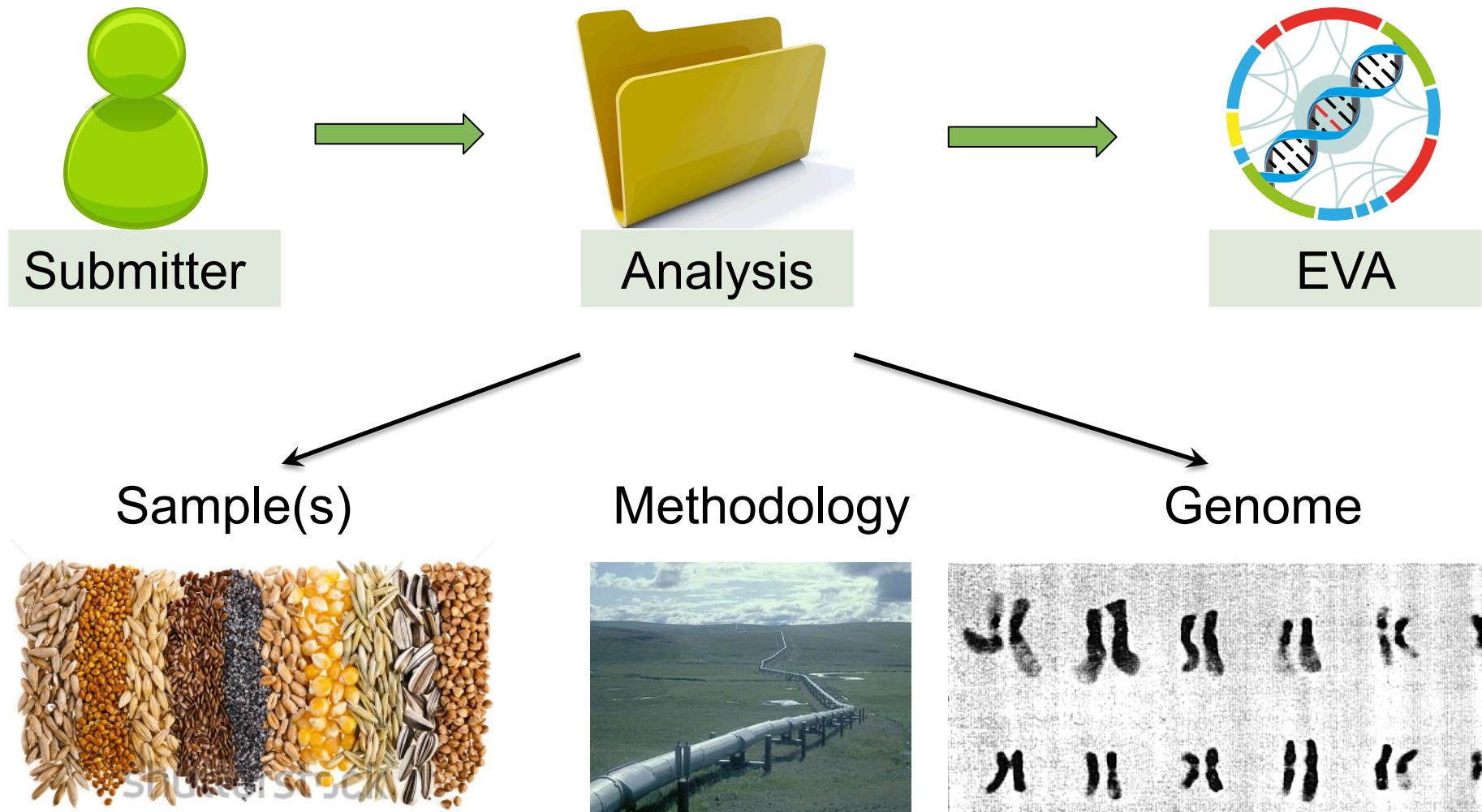
ENA



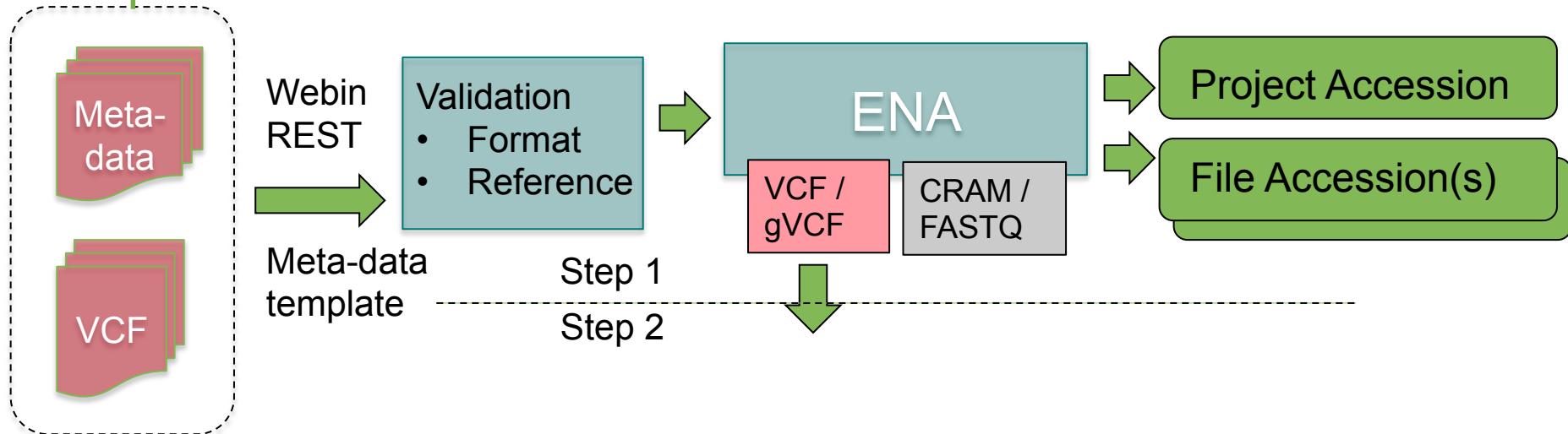
CTGATGGTATGGGCCAAGAGATAATCT  
AGGTACGGCTGTCACTCACTTAGCCTCAC  
AGGGCTGGCATAAAAGTCAGGGAGAGC  
CATGGTCATCTGACTCTGAGGAGAAGT  
CAGGTTGATCAAGGTTACAAGAGCAGG  
GCACTGACTCTCTGCTATTGGCTA

EMBL-EBI

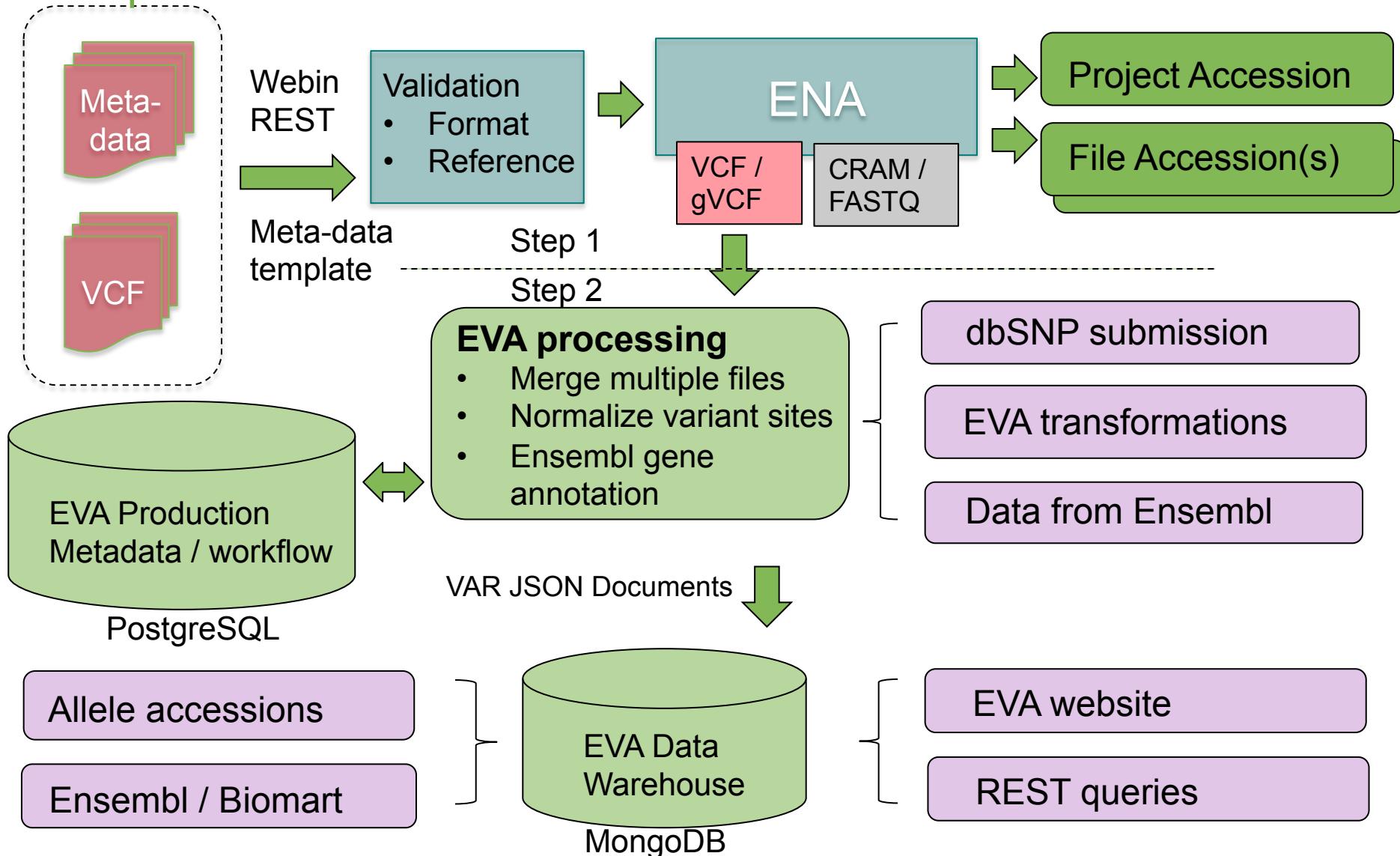
# EVA Data Model



# Pipeline Structure of EVA

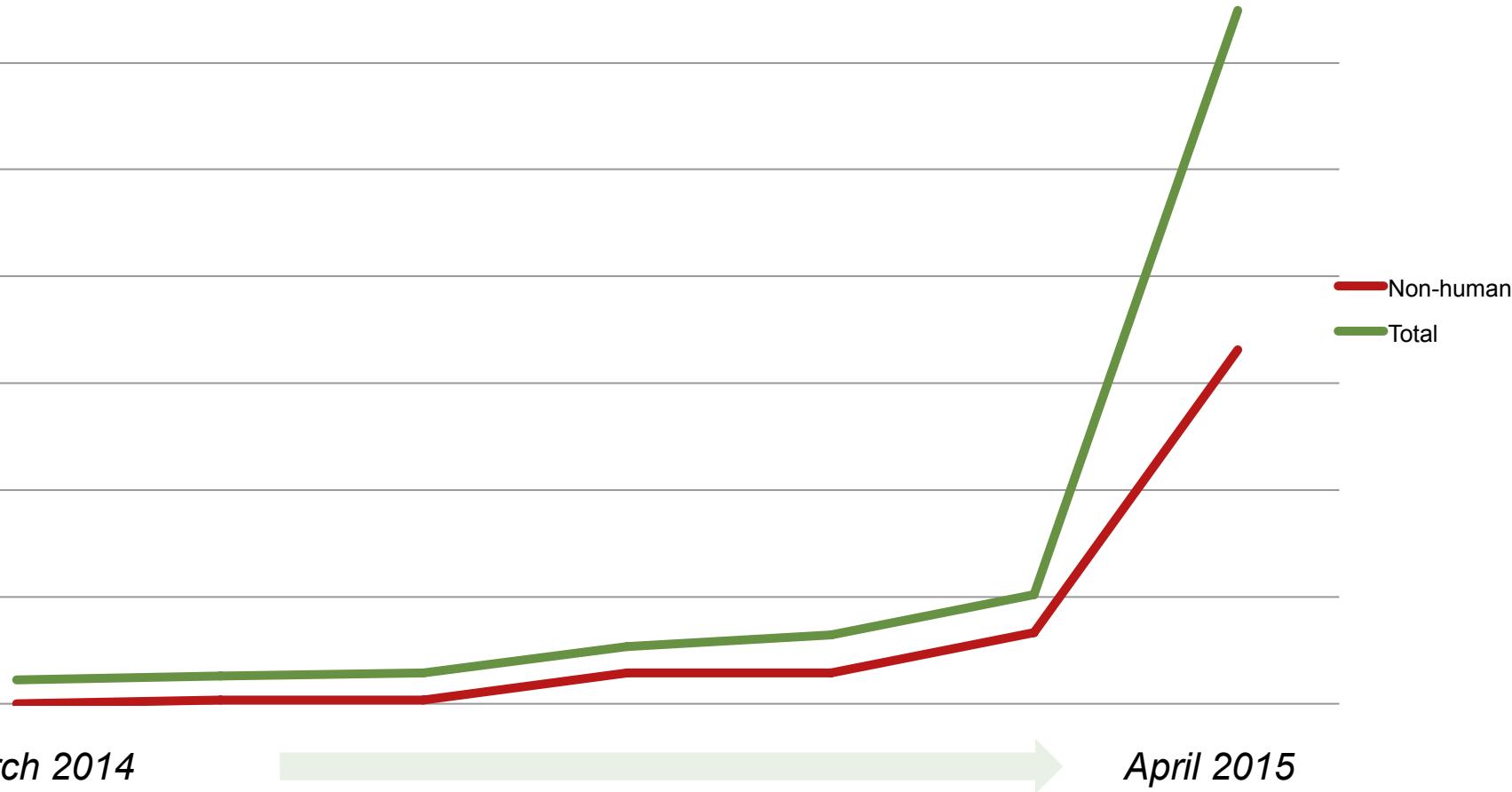


# Pipeline Structure of EVA



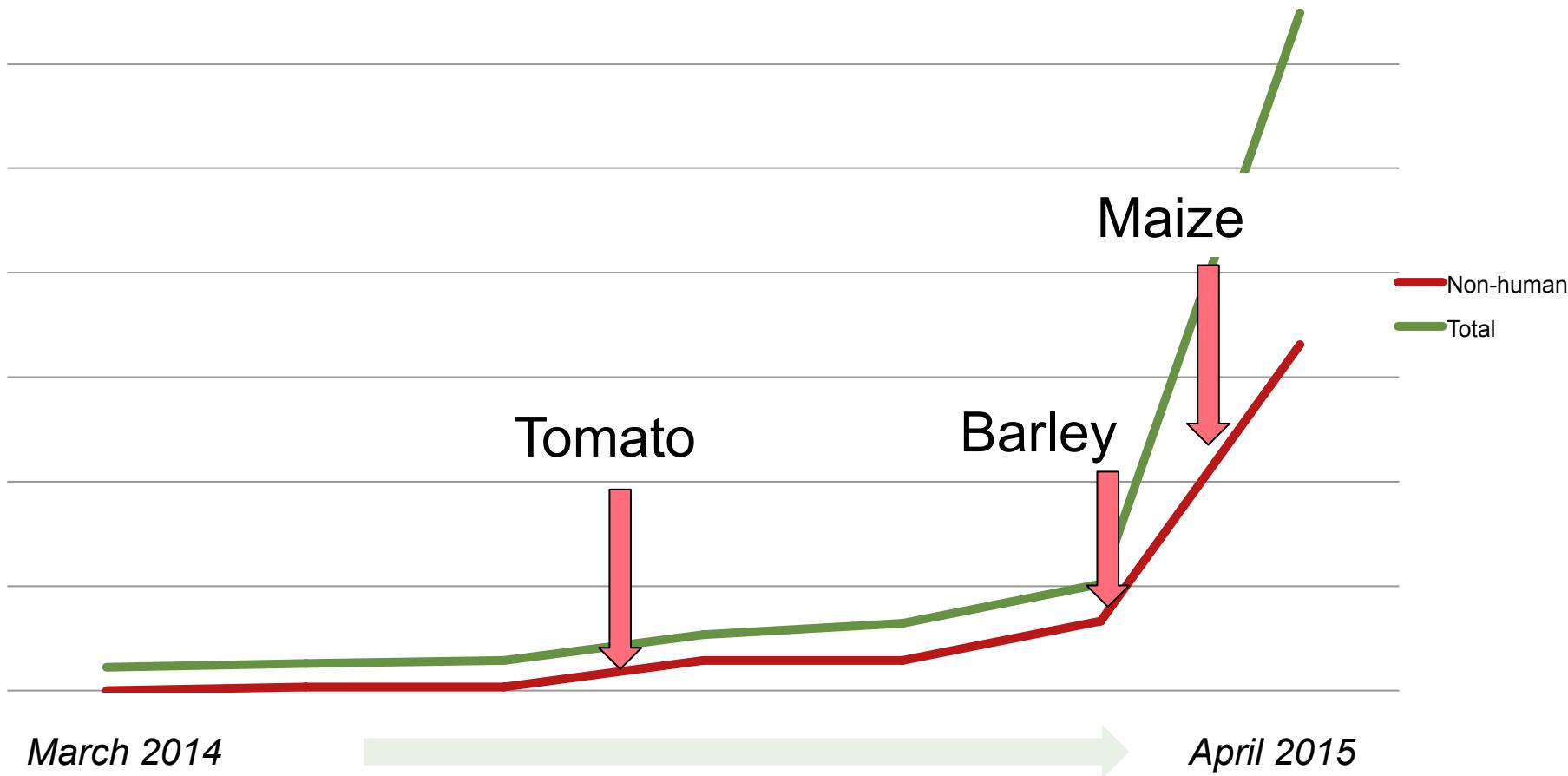
# EVA Content

- Submitted variants



# EVA Content

- Submitted variants

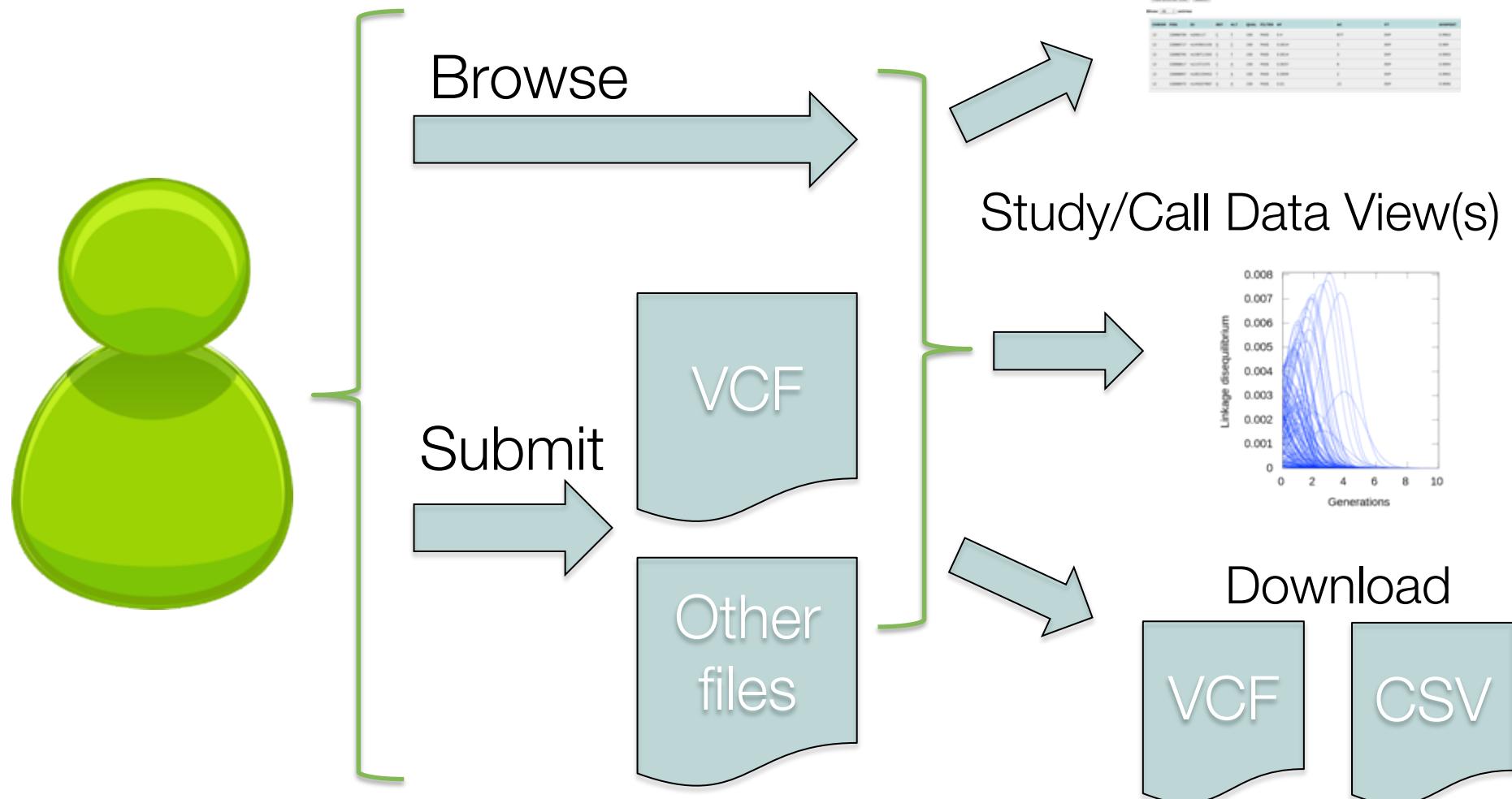


# EVA Content

- 3 main methods of increasing the data in the EVA archive:
  - 1) Large next generation sequence datasets are archived as individual studies
  - 2) User submitted datasets
  - 3) Smaller legacy datasets from dbSNP
- Shrinking percentage of the dbSNP legacy database will not be loaded as individual studies but available via dbSNP legacy track only e.g. older (array-based) technologies, very small studies

# User Experience

Search



## Overview

The European Variation Archive is an open-access database of all types of genetic variation data from all species. The EVA provides access to highly detailed, granular, raw variant data from human, with other species to follow.

All users can [download data](#) from any study, or [submit their own data](#) to the archive. You can also query all variants in the EVA by study, gene, chromosomal location or dbSNP identifier using our [Variant Browser](#).

We will be adding new features to the EVA on a regular basis, and welcome [your comments and feedback](#).

## Statistics

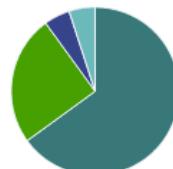
### Short Genetic Variations

Studies  
by Species



Human(10)  
Sheep(2)  
Goat(2)  
Cow(2)  
Vervet monkey(1)

Studies  
by Type



Whole Genome Sequencing(13)  
Exome Sequencing(5)  
Genotyping By Array(1)  
Curation(1)

## News

### Tweets

[Follow](#)



Gary Saunders @EBIvariation

8 Jan

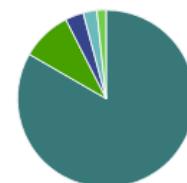
On our way to #PAGXXIII, presenting on Saturday (C01), Sunday (W370) and have a poster (P1106) on Monday. Hope to see some of you there!

[Expand](#)

[Tweet to @EBIvariation](#)

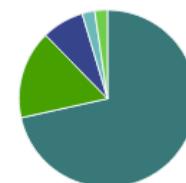
### Structural Variations

Studies  
by Species



Human(100)  
Mouse(11)  
Cow(4)  
Fruit fly(3)  
Pig(2)

Studies  
by Type



Control Set(93)  
Case-Set(21)  
Case-Control(10)  
Collection(3)  
Tumor vs. Matched-Normal(3)

This web application makes an intensive use of new web technologies and standards like HTML5. Please see [About](#) for further browser compatibility notes.

# EVA Study Browser

The screenshot shows the European Variation Archive (EVA) Study Browser. At the top, there's a navigation bar with links for Home, Submit Data, Study Browser, VCF Browser, EVA Clinical, Beacon, About, and Contact. Below the navigation is a search bar with the placeholder "search". To the right of the search bar is a sidebar titled "Studies Found" containing two sections: "Organism" and "Type". The "Organism" section lists Human, Sheep, Goat, Cow, Vervet monkey, Mosquito, Maize, Medaka, Mouse, and Tomato. The "Type" section lists Whole Genome Sequencing, Exome Sequencing, Genotyping By Array, and Curation. The main content area displays a table of study results with columns for ID, Name, Organism, Species, Type, and Download. The table contains 15 rows of study information.

ID	Name	Organism	Species	Type	Download
PRJEB7898	The Exome Aggregation Consortium (ExAC) v0.2	Human	Homo sapiens	ES	FTP
PRJEB5829	Genome of the Netherlands Release 5	Human	Homo sapiens	WGS	FTP
PRJEB5439	Exome Variant Server NHLBI Exome Sequencing Project	Human	Homo sapiens	ES	FTP
PRJEB6042	GEUVADIS: Genetic European Variation in Disease	Human	Homo sapiens	ES	FTP
PRJEB6041	UMCG Cardio GenePanel screening	Human	Homo sapiens	ES	FTP
PRJEB4019	1000 Genomes Phase 1 Analysis	Human	Homo sapiens	WGS, ES	FTP
PRJEB7218	UK10K The Department of Twin Research and Genetic Epidemiolog...	Human	Homo sapiens	WGS	FTP
PRJEB7895	The National FINRISK Study	Human	Homo sapiens	ES	FTP
PRJEB8636	deCODE Genetics Whole-Genome Sequencing Variants	Human	Homo sapiens	WGS	FTP
PRJEB7217	UK10K Avon Longitudinal Study of Parents and Children (ALSPAC) ...	Human	Homo sapiens	WGS	FTP
PRJEB6930	1000 Genomes Project phase3 release V3+	Human	Homo sapiens	WGS, ES	FTP
PRJEB7923	Vervet Genetic Mapping Project	Vervet monkey	Chlorocebus sabaeus	WGS	FTP
PRJEB6911	Sanger Institute Mouse Genomes Project v3	Mouse	Mus	WGS	FTP
PRJEB6025	NextGen project variation for Ovis aries	Sheep	Ovis aries	WGS	FTP

[www.ebi.ac.uk/eva/](http://www.ebi.ac.uk/eva/)    [eva-helpdesk@ebi.ac.uk](mailto:eva-helpdesk@ebi.ac.uk)

# EVA Study Browser

EMBL-EBI

## European Variation Archive

Home | Submit Data | Study Browser | VCF Browser | EVA Clinical | Beacon | About | Contact

Short Genetic Variations Structural Variations

Short Genetic Variations Browser

Search:

Organism

- Human
- Sheep
- Goat
- Cow
- Vervet monkey
- Mosquito
- Maize
- Medaka
- Mouse
- Tomato

Type

- Whole Genome Sequencing
- Exome Sequencing
- Genotyping By Array
- Curation

ID	Name	Organism	Species	Type	Download
PRJEB7723	Biology of Rare Alleles in Maize	Maize	Zea mays	Curation	FTP
PRJEB4395	Whole genome re-sequencing in tomato	Tomato	Solanum lycopersicum	WGS	FTP

[www.ebi.ac.uk/eva/](http://www.ebi.ac.uk/eva/)

[eva-helpdesk@ebi.ac.uk](mailto:eva-helpdesk@ebi.ac.uk)

# EVA Study Browser – Structural Variants

The screenshot shows the European Variation Archive (EVA) Study Browser interface. At the top, there's a navigation bar with links for Home, Submit Data, Study Browser, VCF Browser, EVA Clinical, Beacon, About, and Contact. The main title "European Variation Archive" is displayed above a search bar. Below the search bar, there are two tabs: "Short Genetic Variations" and "Structural Variations", with "Structural Variations" being the active tab. A sidebar on the left contains filters for "Organism" (Human, Mouse, Cow, Fruit fly, Pig, Sheep, Rhesus monkey, Zebrafish, Chimpanzee, Wolf, Horse, Sorghum, Dog) and "Type" (Control Set, Case-Set, Case-Control, Collection, Tumor vs. Matched-Normal). The main content area is titled "Studies Found" and shows a table of 144 studies. The table has columns for ID, Name, Organism, Species, Study Type, and Download. The first few rows of the table are as follows:

ID	Name	Organism	Species	Study Type	Download
estd199	1000_Genomes_Consortium...	Human	Homo sapiens	Control Set	FTP
estd214	1000_Genomes_Consortium...	Human	Homo sapiens	Control Set	FTP
estd219	1000_Genomes_Consortium...	Human	Homo sapiens	Control Set	FTP
estd59	1000_Genomes_Consortium...	Human	Homo sapiens	Control Set	FTP
estd19	Ahn_et_al_2009	Human	Homo sapiens	Control Set	FTP
nstd31	Alkan_et_al_2009	Human	Homo sapiens	Control Set	FTP
nstd106	Alsmadi_et_al_2014	Human	Homo sapiens	Control Set	FTP
estd195	Altshuler_et_al_2010	Human	Homo sapiens	Control Set	FTP
nstd50	Arlt_et_al_2011	Human	Homo sapiens	Control Set	FTP
estd176	Banerjee_et_al_2011	Human	Homo sapiens	Control Set	FTP
estd194	Bentley_et_al_2008	Human	Homo sapiens	Control Set	FTP
estd217	Besenbacher_et_al_2014	Human	Homo sapiens	Control Set	FTP
nstd69	Bickhart_et_al_2012	Cow	Bos taurus	Control Set	FTP
nstd80	Boone_et_al_2013	Human	Homo sapiens	Case Set	FTP
nstd62	Brown_et_al_2012	Zebrafish	Danio rerio	Control Set	FTP
nstd7	Cahan_et_al_2009	Mouse	Mus musculus	Control Set	FTP
nstd46	Campbell_et_al_2011	Human	Homo sapiens	Control Set	FTP
estd211	Campbell_et_al_2014	Human	Homo sapiens	Case Set	FTP
nstd98	Campbell_et_al_2014b	Human	Homo sapiens	Case Set	FTP
nstd13	Chen_et_al_2009	Dog	Canis lupus familiaris	Control Set	FTP

# EVA Study Browser – Structural Variants

The screenshot displays the European Variation Archive (EVA) Study Browser interface. At the top, there is a navigation bar with links to Home, Submit Data, Study Browser, VCF Browser, EVA Clinical, Beacon, About, and Contact. The main content area is titled "Structural Variations" and includes a search bar and a sidebar for filtering studies by organism and type.

**Search:**  
search

**Organism**

- Human
- Mouse
- Cow
- Fruit fly
- Pig
- Sheep
- Rhesus monkey
- Zebrafish
- Chimpanzee
- Wolf
- Horse
- Sorghum
- Dog

**Type**

- Control Set
- Case-Set
- Case-Control
- Collection
- Tumor vs. Matched-Normal

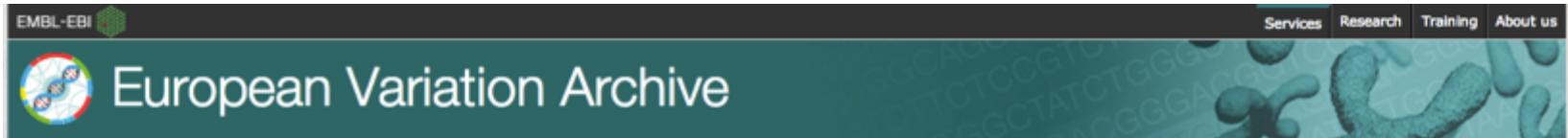
**Studies Found**

Page 1 of 1

ID	Name	Organism	Species	Study Type	Download
nstd63	Zheng_et_al_2011	Sorghum	Sorghum bicolor	Control Set	FTP

Studies 1 - 1 of 1

# EVA 1.0



That's the data available at the current production site:  
[www.ebi.ac.uk](http://www.ebi.ac.uk)

For the remainder of this presentation, including analysis tools, I shall be discussing our upcoming first full release version

This shall be available in ca.3-4 weeks. We're in the final stages of testing

Tumor vs. Matched-Normal

# EVA Study Browser – 1.0

The screenshot shows the European Variation Archive (EVA) Study Browser interface. At the top, there is a navigation bar with links for Home, Submit Data, Study Browser, VCF Browser, EVA Clinical, Beacon, About, and Contact. The main header features the EMBL-EBI logo and the text "European Variation Archive". Below the header, there are two tabs: "Short Genetic Variations (<50bp)" (selected) and "Structural Variations (>50bp)". The main content area is titled "Short Genetic Variations Browser" and contains a search form. The search form includes a text input field labeled "search" and a dropdown menu for "Organism" with options like Barley, Chicken, Cow, Goat, Human, Maize, Medaka, Mosquito, Mouse, Sheep, Tomato, and Vervet monkey. There is also a section for "Type" with options like "-", Curation, Exome Sequencing, Genotyping By Array, and Whole Genome Sequencing. To the right of the search form, a table titled "Studies Found" lists three entries:

ID	Name	Organism	Species	Type	Download
PRJEB7723	Biology of Rare Alleles in Maize	Maize	Zea mays	Curation	FTP
PRJEB4395	Whole genome re-sequencing in tomato	Tomato	Solanum lycopersicum	WGS	FTP
PRJEB629	Sequencing of five barley cultivars.	Barley	Hordeum vulgare subsp. v -		FTP

# EVA Study Browser – 1.0

The screenshot shows the European Variation Archive (EVA) Study Browser. At the top, there's a navigation bar with links for Home, Submit Data, Study Browser, VCF Browser, EVA Clinical, Beacon, About, and Contact. The main header features the EMBL-EBI logo and the text "European Variation Archive". Below the header, a banner displays a DNA sequence and some chromosomes.

The main content area is titled "Short Genetic Variations (<50bp)" and includes a search bar and filters for Organism (Barley selected) and Type (-, Curation, Exome Sequencing, Genotyping By Array, Whole Genome Sequencing). A table lists three studies:

ID	Name	Organism	Species	Type	Download
PRJEB7723	Biology of Rare Alleles in Maize	Maize	Zea mays	Curation	FTP
PRJEB4395	Whole genome re-sequencing in tomato	Tomato	Solanum lycopersicum	WGS	FTP
PRJEB629	Sequencing of five barley cultivars.	Barley	Hordeum vulgare subsp. v -		FTP

A large green box on the right side contains promotional text:

Coming: sorghum, rice, orange, soybean, bean  
Harmonizing with TransPlant  
User submitted datasets...

# EVA VCF Browser

- Build view(s) of the data within EVA
- Users can query and understand data before download
- Ease of use
- Gene annotation from Ensembl
- Direct download of query results
- HTML5 technologies
- Cross-browser compatibility

# EVA VCF Browser – 1.0

EMBL-EBI European Variation Archive

Home Submit Data Study Browser VCF Browser EVA Clinical Beacon About Contact

Filter Variant Browser

Clear Submit

Species: Maize / B73 RefGen\_v3

Organism / Assembly: Maize / B73 RefGen\_v3

Position: 2:100000-400000

Filter By: Chromosomal Location

Consequence Type: missense\_variant

search

Transcript Variant: Coding Variant

- stop\_gained
- stop\_lost
- inframe\_insertion
- inframe\_deletion
- frameshift\_variant
- NMD\_transcript\_variant
- incomplete\_terminal\_codon...
- missense\_variant
- synonymous\_variant
- stop\_retained\_variant
- coding\_sequence\_variant
- feature\_elongation

Page 2 of 7

Variants 11 - 20 of 65

Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Protein substitution scores	View
						Polyphen2 Sift	
2	139601	vcZ2JHU51	A/C	SNV	missense_variant	- -	dbSNP
2	160960	vcZ2JHU52	C/A	SNV	intergenic_variant	- -	dbSNP
2	160979	vcZ2JHU53	G/C	SNV	intergenic_variant	- -	dbSNP
2	178866	vcZ2JHU54	C/T	SNV	missense_variant	- -	dbSNP
2	179139	vcZ2JHU55	G/A	SNV	5_prime_UTR_variant	- -	dbSNP
2	227241	vcZ2JHU56	G/A	SNV	intergenic_variant	- -	dbSNP
2	227256	vcZ2JHU57	C/T	SNV	intergenic_variant	- -	dbSNP
2	227270	vcZ2JHU58	G/T	SNV	intergenic_variant	- -	dbSNP
2	227275	vcZ2JHU59	G/A	SNV	intergenic_variant	- -	dbSNP
2	227297	vcZ2JHU5A	A/G	SNV	intergenic_variant	- -	dbSNP

Results per Page: 10 Export as CSV

Variant Data

File and Stats Genotypes Population Stats Annotation

Studies

Biology of Rare Alleles in Maize (PRJEB7723)

Submitted\_ID: S2\_139601

VCF data +

# EVA VCF Browser – 1.0

EMBL-EBI

European Variation Archive

Home Submit Data Study Browser VCF Browser EVA Clinical Beacon About Contact

Filter

Species: Maize / B73 RefGen\_v3

Organism / Assembly:

- Maize / B73 RefGen\_v3
- Cow / Bos\_taurus\_UMD\_3.1
- Goat / CHIR\_1.0
- Human / GRCh37
- Human / GRCh37
- Maize / B73 RefGen\_v3
- Medaka / oryLat2\_corr\_unfiltered.fa
- Mosquito / AgamP3
- Mouse / GRCm38.p3
- Sheep / Oar\_v3.1
- Sorghum / Sorbil1
- Tomato / SL2.40
- Vervet monkey / Chlorocebus\_sabaeus

Coding variant

- stop\_gained
- stop\_lost
- inframe\_insertion
- inframe\_deletion
- frameshift\_variant
- NMD\_transcript\_variant
- incomplete\_terminal\_codon...
- missense\_variant
- synonymous\_variant
- stop\_retained\_variant
- coding\_sequence\_variant
- feature\_elongation

Variant Browser

Page 2 of 7

Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Protein substitution scores	View
						Polyphen2 Sift	
2	139601	vcZ2JHU51	A/C	SNV	missense_variant	- -	dbSNP
2	160960	vcZ2JHU52	C/A	SNV	intergenic_variant	- -	dbSNP
2	160979	vcZ2JHU53	G/C	SNV	intergenic_variant	- -	dbSNP
2	178866	vcZ2JHU54	C/T	SNV	missense_variant	- -	dbSNP
2	179139	vcZ2JHU55	G/A	SNV	5_prime_UTR_variant	- -	dbSNP
2	227241	vcZ2JHU56	G/A	SNV	intergenic_variant	- -	dbSNP
2	227256	vcZ2JHU57	C/T	SNV	intergenic_variant	- -	dbSNP
2	227270	vcZ2JHU58	G/T	SNV	intergenic_variant	- -	dbSNP
2	227275	vcZ2JHU59	G/A	SNV	intergenic_variant	- -	dbSNP
2	227297	vcZ2JHU5A	A/G	SNV	intergenic_variant	- -	dbSNP

Results per Page: 10 Export as CSV

Variant Data

File and Stats Genotypes Population Stats Annotation

Studies

Biology of Rare Alleles in Maize (PRJEB7723)

Submitted\_ID: S2\_139601

VCF data +

EMBL-EBI

# EVA VCF Browser – 1.0

EMBL-EBI European Variation Archive

Home Submit Data Study Browser VCF Browser EVA Clinical Beacon About Contact

## Filter

Clear Submit

Species: Maize / B73 RefGen\_v3

Organism / Assembly: Maize / B73 RefGen\_v3

Position

Filter By: Chromosomal Location

Consequence Type: search

Transcript Variant

Coding Variant

- stop\_gained
- stop\_lost
- inframe\_insertion
- inframe\_deletion
- frameshift\_variant
- NMD\_transcript\_variant
- incomplete\_terminal\_codon...
- missense\_variant
- synonymous\_variant
- stop\_retained\_variant
- coding\_sequence\_variant
- feature\_elongation

Variant Browser

Page 2 of 7

Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Protein substitution scores	View
						Polyphen2 Sift	
2	139601	vcZ2JHU51	A/C	SNV	missense_variant	- -	
2	160960	vcZ2JHU52	C/A	SNV	intergenic_variant	- -	
2	160979	vcZ2JHU53	G/C	SNV	intergenic_variant	- -	
2	178866	vcZ2JHU54	C/T	SNV	missense_variant	- -	
2	179139	vcZ2JHU55	G/A	SNV	5_prime_UTR_variant	- -	
2	227241	vcZ2JHU56	G/A	SNV	intergenic_variant	- -	
2	227256	vcZ2JHU57	C/T	SNV	intergenic_variant	- -	
2	227270	vcZ2JHU58	G/T	SNV	intergenic_variant	- -	
2	227275	vcZ2JHU59	G/A	SNV	intergenic_variant	- -	
2	227297	vcZ2JHU5A	A/G	SNV	intergenic_variant	- -	

Results per Page: 10 Export as CSV

## Variant Data

File and Stats Genotypes Population Stats Annotation

## Studies

Biology of Rare Alleles in Maize (PRJEB7723)

Submitted\_ID: S2\_139601

VCF data +

# EVA VCF Browser – 1.0

EMBL-EBI European Variation Archive

Home Submit Data Study Browser VCF Browser EVA Clinical Beacon About Contact

Filter Variant Browser

Clear Submit

Species: Maize / B73 RefGen\_v3

Organism / Assembly: Maize / B73 RefGen\_v3

Position: 2:100000-400000

Filter By: Chromosomal Location

Consequence Type: search

Transcript Variant:

- Coding Variant
  - stop\_gained
  - stop\_lost
  - inframe\_insertion
  - inframe\_deletion
  - frameshift\_variant
  - NMD\_transcript\_variant
  - incomplete\_terminal\_codon...
  - missense\_variant
  - synonymous\_variant
  - stop\_retained\_variant
  - coding\_sequence\_variant
  - feature\_elongation

Page 2 of 7

Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Protein substitution scores	View
						Polyphen2 Sift	
2	139601	vcZ2JHU51	A/C	SNV	missense_variant	- -	dbSNP
2	160960	vcZ2JHU52	C/A	SNV	intergenic_variant	- -	dbSNP
2	160979	vcZ2JHU53	G/C	SNV	intergenic_variant	- -	dbSNP
2	178866	vcZ2JHU54	C/T	SNV	missense_variant	- -	dbSNP
2	179139	vcZ2JHU55	G/A	SNV	5_prime_UTR_variant	- -	dbSNP
2	227241	vcZ2JHU56	G/A	SNV	intergenic_variant	- -	dbSNP
2	227256	vcZ2JHU57	C/T	SNV	intergenic_variant	- -	dbSNP
2	227270	vcZ2JHU58	G/T	SNV	intergenic_variant	- -	dbSNP
2	227275	vcZ2JHU59	G/A	SNV	intergenic_variant	- -	dbSNP
2	227297	vcZ2JHU5A	A/G	SNV	intergenic_variant	- -	dbSNP

Results per Page: 10 Export as CSV

Variant Data

File and Stats Genotypes Population Stats Annotation

Studies

Biology of Rare Alleles in Maize (PRJEB7723)

Submitted\_ID: S2\_139601

VCF data +

# Ensembl Plants VEP

Ensembl Plants ▾

BLAST | BioMart | Tools | Downloads | Help | Documentation

Login Register Lost Password

Custom Data Add your data Attach DAS Manage Data Features on Karyotype

Manage Configurations Configurations for this page All configurations Configuration sets

Online Tools Variant Effect Predictor

Assembly Converter ID History Converter Data Slicer

Help

Variant Effect Predictor:

This tool takes a list of variant positions and alleles, and predicts the effects of each of these on overlapping transcripts and regulatory regions annotated in Ensembl. The tool accepts substitutions, insertions and deletions as input, see [data formats](#).

Upload is limited to 750 variants; lines after the limit will be ignored. Users with more than 750 variations can split files into smaller chunks, use the standalone [perl script](#) or the [variation API](#). See also [full documentation](#)

**NB:** Ensembl now by default uses Sequence Ontology terms to describe variation consequences. See [this page](#) for details

**Input file**

Species:

Name for this data (optional):

Paste data:

Upload file:  No file chosen

or provide file URL:

Input file format:

**Options**

Transcript database to use:  Ensembl transcripts  RefSeq and other transcripts

Type of consequences to display:

Check for existing co-located variants:

Return results for variants in coding regions only:

Show Ensembl protein identifiers where available:

Show HGVS identifiers for variants where available:

Search Ensembl Plants...

Logout

# EVA VCF Browser – 1.0

EMBL-EBI European Variation Archive

Home Submit Data Study Browser VCF Browser EVA Clinical Beacon About Contact

## Variant Browser

Filter

Species: Maize / B73 RefGen\_v3

Organism / Assembly: Maize / B73 RefGen\_v3

Position: 2:100000-400000

Filter By: Chromosomal Location

Consequence Type: missense\_variant, intergenic\_variant

Alleles: A/C, C/A, G/C, C/T, G/A, G/T, A/G

Class: SNV

Most Severe Consequence Type: missense\_variant, 5\_prime\_UTR\_variant, intergenic\_variant

Protein substitution scores: Polyphen2, Sift

View: dbSNP

Results per Page: 10 Export as CSV

### Variant Data

File and Stats Genotypes Population Stats Annotation

### Studies

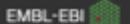
Biology of Rare Alleles in Maize (PRJEB7723)

Submitted\_ID: S2\_139601

VCF data +

Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Protein substitution scores	View
						Polyphen2 Sift	
2	139601	vcZ2JHU51	A/C	SNV	missense_variant	- -	dbSNP
2	160960	vcZ2JHU52	C/A	SNV	intergenic_variant	- -	dbSNP
2	160979	vcZ2JHU53	G/C	SNV	intergenic_variant	- -	dbSNP
2	178866	vcZ2JHU54	C/T	SNV	missense_variant	- -	dbSNP
2	179139	vcZ2JHU55	G/A	SNV	5_prime_UTR_variant	- -	dbSNP
2	227241	vcZ2JHU56	G/A	SNV	intergenic_variant	- -	dbSNP
2	227256	vcZ2JHU57	C/T	SNV	intergenic_variant	- -	dbSNP
2	227270	vcZ2JHU58	G/T	SNV	intergenic_variant	- -	dbSNP
2	227275	vcZ2JHU59	G/A	SNV	intergenic_variant	- -	dbSNP
2	227297	vcZ2JHU5A	A/G	SNV	intergenic_variant	- -	dbSNP

# EVA VCF Browser – 1.0



EMBL-EBI Services Research Training About us

# European Variation Archive

Home Submit Data Study Browser VCF Browser EVA Clinical Beacon About Contact

Filter Variant Browser

Clear Submit Page 2 of 7 Variants 11 - 20 of 65

Species	Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Protein substitution scores	View
Zea mays	1	123456789	rs123456789	A>C	SNP	missense	Polyphen2 Sift	<a href="#">View</a>
Zea mays	2	123456789	rs123456789	T>G	SNP	missense	Polyphen2 Sift	<a href="#">View</a>
Zea mays	3	123456789	rs123456789	C>T	SNP	missense	Polyphen2 Sift	<a href="#">View</a>
Zea mays	4	123456789	rs123456789	G>A	SNP	missense	Polyphen2 Sift	<a href="#">View</a>
Zea mays	5	123456789	rs123456789	T>C	SNP	missense	Polyphen2 Sift	<a href="#">View</a>
Zea mays	6	123456789	rs123456789	A>G	SNP	missense	Polyphen2 Sift	<a href="#">View</a>
Zea mays	7	123456789	rs123456789	C>G	SNP	missense	Polyphen2 Sift	<a href="#">View</a>
Zea mays	8	123456789	rs123456789	T>G	SNP	missense	Polyphen2 Sift	<a href="#">View</a>
Zea mays	9	123456789	rs123456789	C>T	SNP	missense	Polyphen2 Sift	<a href="#">View</a>
Zea mays	10	123456789	rs123456789	G>A	SNP	missense	Polyphen2 Sift	<a href="#">View</a>

```
##fileformat=VCFv4.0
##fileDate=20140213
##source=AllZeaGBSv2.7_publicSamples_imputedV3b
##reference=B73RefGenV3
##phasing=no
##INFO=<ID=MQ,Number=.,Type=Float,Description="RMS mapping quality">
##INFO=<ID=Submitted_ID,Number=1,Type=String,Description="Submitter ID">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##contig=<ID=1,species="Zea mays",accession=GK000031.3>
##contig=<ID=2,species="Zea mays",accession=GK000032.3>
##contig=<ID=3,species="Zea mays",accession=GK000033.3>
##contig=<ID=4,species="Zea mays",accession=CM000780.3>
##contig=<ID=5,species="Zea mays",accession=CM000781.3>
##contig=<ID=6,species="Zea mays",accession=CM000782.3>
##contig=<ID=7,species="Zea mays",accession=GK000034.3>
##contig=<ID=8,species="Zea mays",accession=CM000784.3>
##contig=<ID=9,species="Zea mays",accession=CM000785.3>
##contig=<ID=10,species="Zea mays",accession=CM000786.3>
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMEA2833119 SAMEA2830877 SAMEA2837965 SAMEA2839535 SAMEA2839706 SAMEA2843832 SAMEA2829711 SAMEA2828547 SAMEA2844507
2 139601 vcZ2JH51 A C . Submitted_ID=S2_139601 GT 0/0 0/0 0/0 ./ 0/0 0/0 ./ 0/0 0/0 0/0 ./ ./ ./ 0/0 ./ ./ ./ ./ ./ 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0 0/0 ./
```

## **Studies**

- ❑ frameshift\_variant
  - ❑ NMD\_transcript\_variant
  - ❑ incomplete\_terminal\_codon...
  - ❑ missense\_variant
  - ❑ synonymous\_variant
  - ❑ stop\_retained\_variant
  - ❑ coding\_sequence\_variant
  - ❑ feature\_elongation

Biology of Rare Alleles in Maize (PRJEB7723)

Submitted\_ID

## VCF data

# EVA VCF Browser – 1.0

EMBL-EBI

European Variation Archive

Home Submit Data Study Browser VCF Browser EVA Clinical Beacon About Contact

Filter Variant Browser

Clear Submit

Species: Maize / B73 RefGen\_v3

Organism / Assembly: Maize / B73 RefGen\_v3

Position: Chromosomal Location: 2:100000-400000

Filter By: Consequence Type: search

Transcript Variant

- Coding Variant
  - stop\_gained
  - stop\_lost
  - inframe\_insertion
  - inframe\_deletion
  - frameshift\_variant
  - NMD\_transcript\_variant
  - incomplete\_terminal\_codon...
  - missense\_variant
  - synonymous\_variant
  - stop\_retained\_variant
  - coding\_sequence\_variant
  - feature\_elongation

Protein Substitution Score

Polyphen2 >: ex: 0.5

Sift < : ex: 0.1

Page 2 of 7

Variants 11 - 20 of 65

Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Protein substitution scores	View
						Polyphen2 Sift	
2	139601	vc22JHU51	A/C	SNV	missense_variant	- -	dbSNP
2	160960	vc22JHU52	C/A	SNV	intergenic_variant	- -	dbSNP
2	160979	vc22JHU53	G/C	SNV	intergenic_variant	- -	dbSNP
2	178866	vc22JHU54	C/T	SNV	missense_variant	- -	dbSNP
2	179139	vc22JHU55	G/A	SNV	5_prime_UTR_variant	- -	dbSNP
2	227241	vc22JHU56	G/A	SNV	intergenic_variant	- -	dbSNP
2	227256	vc22JHU57	C/T	SNV	intergenic_variant	- -	dbSNP
2	227270	vc22JHU58	G/T	SNV	intergenic_variant	- -	dbSNP
2	227275	vc22JHU59	G/A	SNV	intergenic_variant	- -	dbSNP
2	227297	vc22JHU5A	A/G	SNV	intergenic_variant	- -	dbSNP

Results per Page: 10 Export as CSV

Variant Data

File and Stats Genotypes Population Stats Annotation

Genotypes

Study	Samples Count
Biology of Rare Alleles in Maize (PRJEB7723)	17280
Sample	Genotype
SAMEA2830759	./.
SAMEA2830758	0/0
SAMEA2830757	./.
SAMEA2830756	0/0
SAMEA2830755	./.
SAMEA2830754	./.

# EVA VCF Browser – 1.0

EMBL-EBI

## European Variation Archive

Home Submit Data Study Browser VCF Browser EVA Clinical Beacon About Contact

Filter Variant Browser

Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Protein substitution scores	View
						Polyphen2 Sift	
2	139601	vc22JHU51	A/C	SNV	missense_variant	- -	dbSNP
2	160960	vc22JHU52	C/A	SNV	intergenic_variant	- -	dbSNP
2	160979	vc22JHU53	G/C	SNV	intergenic_variant	- -	dbSNP
2	178866	vc22JHU54	C/T	SNV	missense_variant	- -	dbSNP
2	179139	vc22JHU55	G/A	SNV	5_prime_UTR_variant	- -	dbSNP
2	227241	vc22JHU56	G/A	SNV	intergenic_variant	- -	dbSNP
2	227256	vc22JHU57	C/T	SNV	intergenic_variant	- -	dbSNP
2	227270	vc22JHU58	G/T	SNV	intergenic_variant	- -	dbSNP
2	227275	vc22JHU59	G/A	SNV	intergenic_variant	- -	dbSNP
2	227297	vc22JHU5A	A/G	SNV	intergenic_variant	- -	dbSNP

Results per Page: 10 Export as CSV

### Variant Data

File and Stats Genotypes Population Stats Annotation

### Annotations

Ensembl Gene ID	Ensembl Gene Symbol	Ensembl Transcript ID	SO Term(s)	Biotype	Codon	cDNA Position	AA Change
GRMZM2G046590	GRMZM2G046590...	missense_variant	protein_coding	gaA/g...	1517	E/D	

Transcripts 1 - 1 of 1

# EVA VCF Browser – 1.0

EMBL-EBI

Services Research Training About us

Login/Register

**e! EnsemblPlants** ▾ BLAST | BioMart | Tools | Downloads | Help | Documentation

Location: 2:134,843-139,880 Gene: GRMZM2G046590 Trans: GRMZM2G046590\_T01

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Gene alleles
- Sequence
  - Secondary Structure
- External references
- Regulation
- Literature
- Ontology
  - GO: biological process
  - GO: cellular component
- Plant Compara
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
- Pan-taxonomic Compara
  - Gene Tree
  - Orthologues
- Phenotype
- Genetic Variation
  - Variation table
  - Structural variation
  - Variation image
- External data
  - Gene expression
  - Personal annotation
- ID History
  - Gene history

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Ensembl Plants is produced in collaboration with Gramene

synonymous\_variant  
 stop\_retained\_variant  
 coding\_sequence\_variant  
 feature\_elongation

Ensembl Gene ID: GRMZM2G046590 | Ensembl Gene Symbol: GRMZM2G046590... | Ensembl Transcript ID: GRMZM2G046590\_T01 | SO Term(s): missense\_variant | Biotype: protein\_coding | Codon: gaA/g... | cDNA Position: 1517 | AA Change: E/D

Chromosome 2: 134,843-139,880 forward strand.

This gene has 1 transcript ([splice variant](#)), [64 orthologues](#) and [23 paralogues](#).

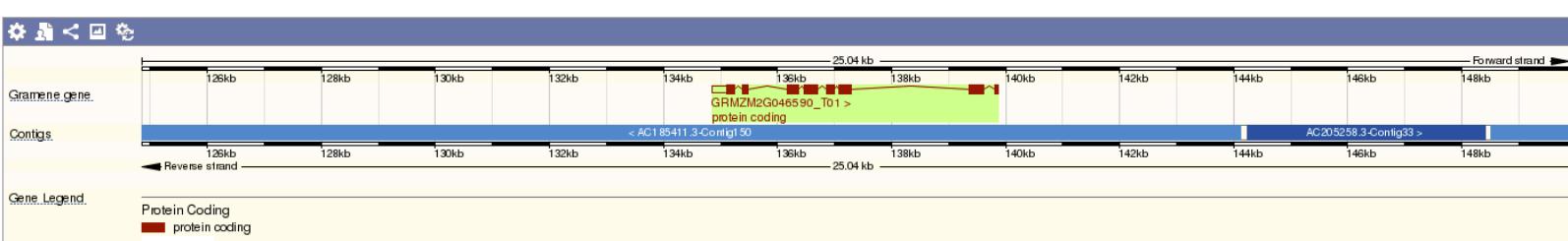
Show transcript table

## Summary

Gene type: Protein coding

Annotation Method: Gene annotation by [Gramene](#) through an automated, evidence-based method

Go to [Region in Detail](#) for more tracks and navigation options (e.g. zooming)



Gramene gene

Contigs

Gene Legend

Protein Coding

Forward strand

Reverse strand

< AC185411.3-Contig150 >

AC205258.3-Contig33 >

25.04 kb

126kb 128kb 130kb 132kb 134kb 136kb 138kb 140kb 142kb 144kb 146kb 148kb

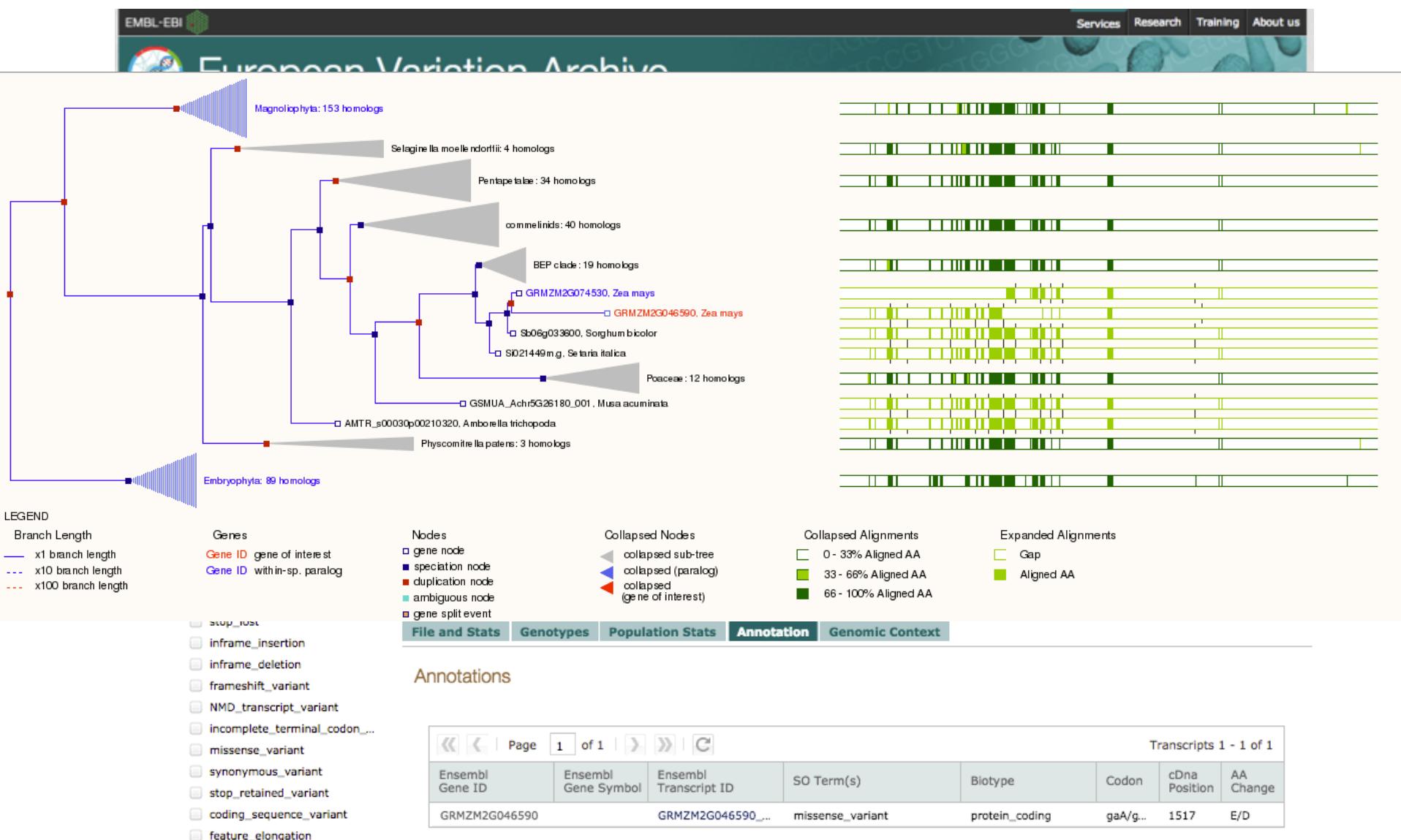
136kb 128kb 130kb 132kb 134kb 136kb 138kb 140kb 142kb 144kb 146kb 148kb

25.04 kb

Tip: use the "Configure this page" link on the left to show additional data in this region.

About Ensembl Genomes | Contact Us | EMBL-EBI Terms of use | Privacy | Cookies | Help

# EVA VCF Browser – 1.0



# EVA VCF Browser – 1.0

EMBL-EBI

## European Variation Archive

Home Submit Data Study Browser VCF Browser EVA Clinical Beacon About Contact

Filter Variant Browser

Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Protein substitution scores	View
						Polyphen2 Sift	
2	139601	vc22JHU51	A/C	SNV	missense_variant	- -	dbSNP
2	160960	vc22JHU52	C/A	SNV	intergenic_variant	- -	dbSNP
2	160979	vc22JHU53	G/C	SNV	intergenic_variant	- -	dbSNP
2	178866	vc22JHU54	C/T	SNV	missense_variant	- -	dbSNP
2	179139	vc22JHU55	G/A	SNV	5_prime_UTR_variant	- -	dbSNP
2	227241	vc22JHU56	G/A	SNV	intergenic_variant	- -	dbSNP
2	227256	vc22JHU57	C/T	SNV	intergenic_variant	- -	dbSNP
2	227270	vc22JHU58	G/T	SNV	intergenic_variant	- -	dbSNP
2	227275	vc22JHU59	G/A	SNV	intergenic_variant	- -	dbSNP
2	227297	vc22JHU5A	A/G	SNV	intergenic_variant	- -	dbSNP

Results per Page: 10 Export as CSV

### Variant Data

File and Stats Genotypes Population Stats Annotation

### Annotations

Ensembl Gene ID	Ensembl Gene Symbol	Ensembl Transcript ID	SO Term(s)	Biotype	Codon	cDNA Position	AA Change
GRMZM2G046590	GRMZM2G046590...	missense_variant	protein_coding	gaA/g...	1517	E/D	

Transcripts 1 - 1 of 1

# EVA API

- EVA VCF browser is our GUI view of web service results
- Efficient programmatic access through a RESTful web services API
- All EVA data available regardless of the programming language
- Results provided as JSON objects: easily parsed by Python, R, JAVA, for example
- Web services for:
  - files, segments, studies, variants
  - full documentation at EVA website (1.0)

# EVA needs your input!

- EVA is now in beta-release, please suggest changes to make it more useful
- EVA 1.0, first full release, July 2015, comments and suggestions are the foundations of changes
- Data submissions are key, ideally with genotypes
- Contact at [eva-helpdesk@ebi.ac.uk](mailto:eva-helpdesk@ebi.ac.uk)
  - I'm also around this week

# Submit data to EVA

The screenshot shows the European Variation Archive (EVA) homepage. At the top left is the EMBL-EBI logo. To its right is a horizontal navigation bar with links: Services, Research, Training, and About us. Below this is a main banner with the text "European Variation Archive" and a decorative background of a DNA double helix and some text. Underneath the banner is another horizontal navigation bar with links: Home, Submit Data, Study Browser, VCF Browser, EVA Clinical, Beacon, About, and Contact.

## European Variation Archive submissions

EVA follows the infrastructure of fellow EMBL-EBI resources European Nucleotide Archive (ENA) and European Genome-phenome Archive (EGA) to accept, archive, and accession VCF files. Submissions consist of VCF file(s) and metadata that describe sample(s), experiment(s), and analysis that produced the variant and/or genotype call(s).

EVA works in collaboration with the Database of Genomics Variants Archive (DGVa) to accession and archive structural variants. DGVa relies on a template based submission process that is explained in detail [here](#).

Data submitted to EVA is brokered to our collaborating databases at NCBI, dbSNP and dbVar. It is therefore unnecessary to submit data to multiple resources. Please contact [eva-helpdesk@ebi.ac.uk](mailto:eva-helpdesk@ebi.ac.uk) if you would like any further information on this brokering process or collaboration.

If you have data in a format other than VCF, which cannot be converted to VCF, please contact [eva-helpdesk@ebi.ac.uk](mailto:eva-helpdesk@ebi.ac.uk). Additional submission formats may be supported over time as required by the scientific community.

### Key stages of EVA submissions

#### Contact

Contact [eva-helpdesk@ebi.ac.uk](mailto:eva-helpdesk@ebi.ac.uk) in order to provide details of your submission.

#### Receive

Download your [submission pack](#), which will include:

- Details for your submission uploads
- [Templates](#) to capture your associated metadata
- Key stages for your submission

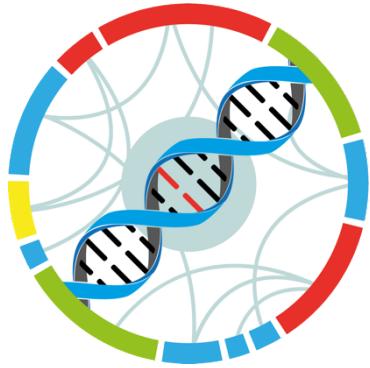
#### Submit

Upload your data files to your private submission upload account or directly to the [eva-helpdesk@ebi.ac.uk](mailto:eva-helpdesk@ebi.ac.uk).

# Submission to EVA

- Data submitted to EVA is shared with NCBI:
  - dbSNP, dbVar
  - No need to submit data twice
- Response time of 48 hrs
  - Accession number suitable for publication
- EVA dynamic study loading pipeline
  - No need to wait to release date to see study on website or data in variant browser
- Focus on growth of PAG relevant data

# Conclusion



European Variation Archive  
[www.ebi.ac.uk/eva](http://www.ebi.ac.uk/eva)

- Variant file archive
- All types of variants, all species
- Provides direct views of raw VCF files
- Full API
- Completely free to use

# Acknowledgments

## EVA / DGVA

Justin Paschall

Ignacio Medina Castello

Gary Saunders

Cristina Yenyxe Gonzalez

Jag Kandasamy

Ilkka Lappalainen

## EGA

Jeff Almeida-King

Vasudev Kumanduri

Saif Ur-Rehman

Tom Smith



@ebivariation

## Ensembl Variation

Fiona Cunningham

Sarah Hunt

William McLaren

Anja Thormann

Laurent Gil

## ENA team

Rasko Leinonen

Rajesh Radhakrishnan

Daniel Vaughan

## Ensembl Genomes

Paul Kersey

Dan Bolser

Christoph Grabmuller

## Funding

