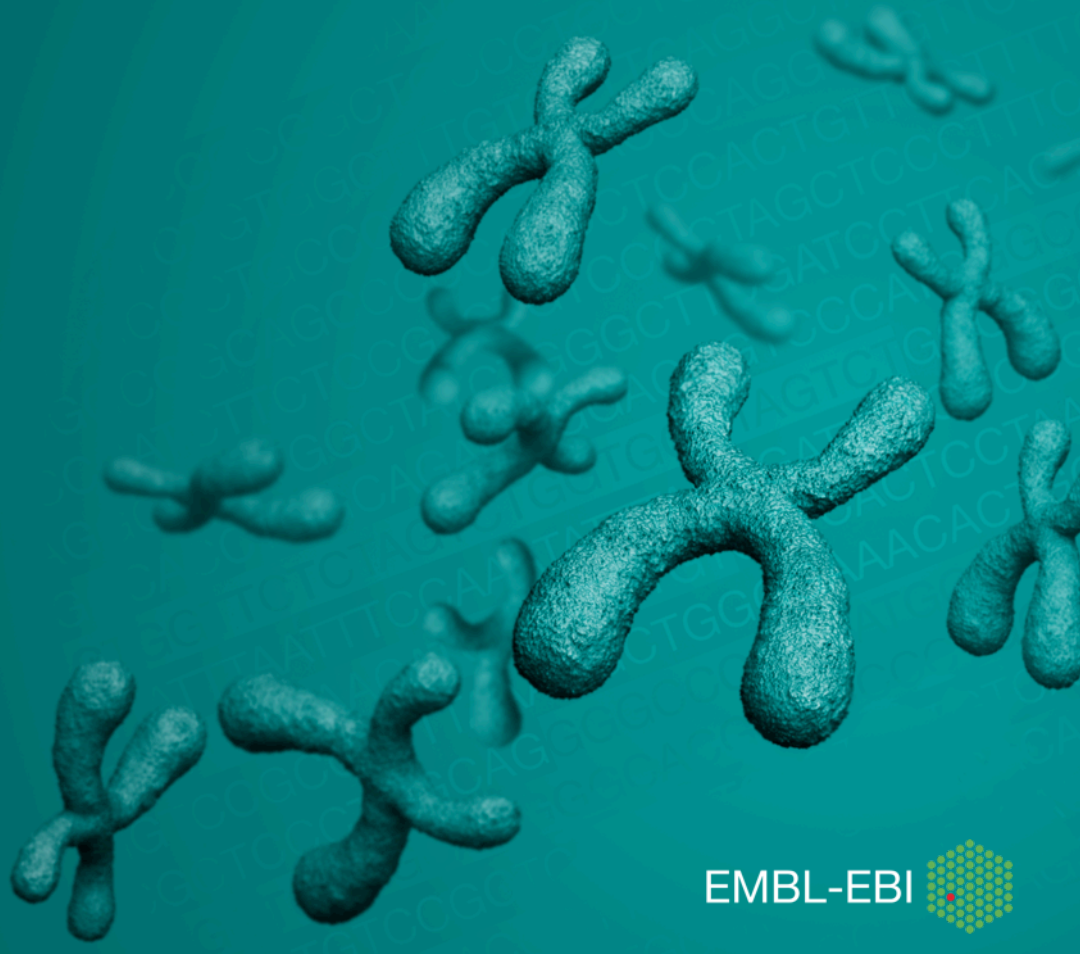
# The Transition of Non-Human Genetic Variation Data from dbSNP to EVA

Gary Saunders, PhD

AgBio, 7th February, 2018

www.ebi.ac.uk/eva

eva-helpdesk@ebi.ac.uk



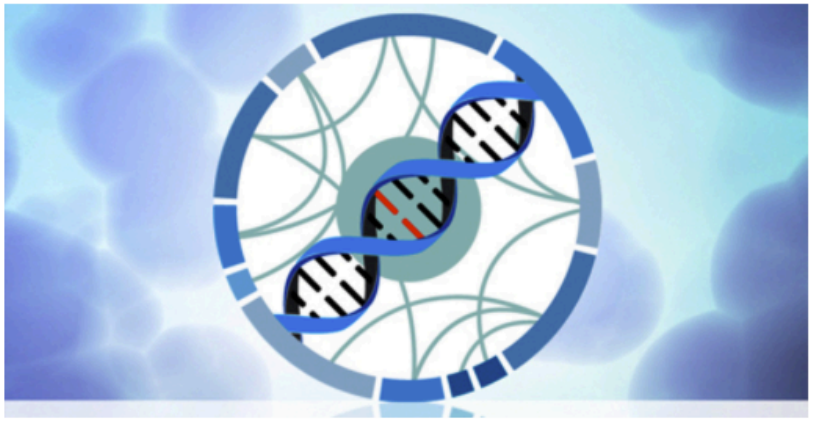
EMBL-EBI

# Outline

- Summary of dbSNP > EVA transition
- Why?
- Working timelines at EVA and dbSNP
- Submission of data to EVA
- Access to data at EVA
- EVA training materials

# Since September 2017 dbSNP no longer accepts non-human variants, data to be submitted to EVA

News overview News archive Photos of EMBL-EBI Blogs



## EVA issues long-term IDs for non-human variants

9 May 2017 - 15:18

### Summary

- New agreement between the NCBI and EMBL-EBI shares responsibility for managing data from

NCBI

About PubMed Labs What's New Quick Tips & Tricks Science Features

Posted on May 9, 2017 ← Previous Next →

## Phasing out support for non-human genome organism data in dbSNP and dbVar

15 Votes

*This blog post is directed toward people who use dbSNP and dbVar, particularly those who submit non-human data to the two databases.*

dbSNP and dbVar archive, process, display and report information related to germline and somatic variations from multiple species. These two databases have grown rapidly as sequencing and other discovery technologies have evolved, and now contain nearly two billion variants from over 360 species.

Based on projected growth and the resources required to archive and distribute the data, continued support for all organisms will become unsustainable for NCBI in the near future. Therefore, NCBI will phase out support for all non-human organisms in dbSNP and dbVar, and will support only human variation.

# Since September 2017 dbSNP no longer accepts non-human variants, data to be submitted to EVA

Why this change?

- Increased volume of data to handle
- Formal split of overlapping functionalities
- Quicker turnaround for community
  - Who may have been losing faith in system
- Change in focus of host institutions (NCBI / EBI)

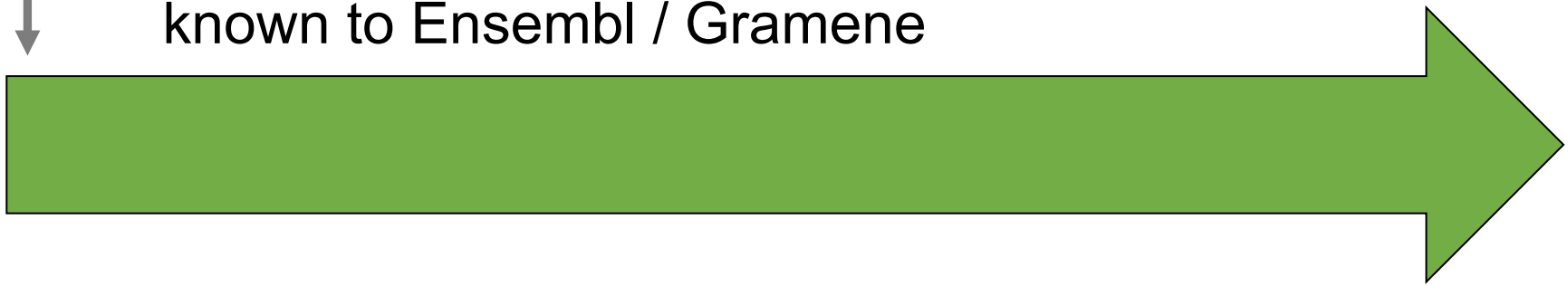
# Since September 2017 dbSNP no longer accepts non-human variants, data to be submitted to EVA

- EVA is committed to the continuity of existing dbSNP non-human ‘ss’ & ‘rs’ identifiers
- New non-human variants shall be accessioned with ‘ss’ and ‘rs’ identifiers
- Variants identifiers shall be searchable at the EVA website and programmatically
- EVA shall continue to roll-out information via the EVA website (www.ebi.ac.uk/eva) and Twitter @evarchive

# dbSNP > EVA transition working timeline

Sept ‘17

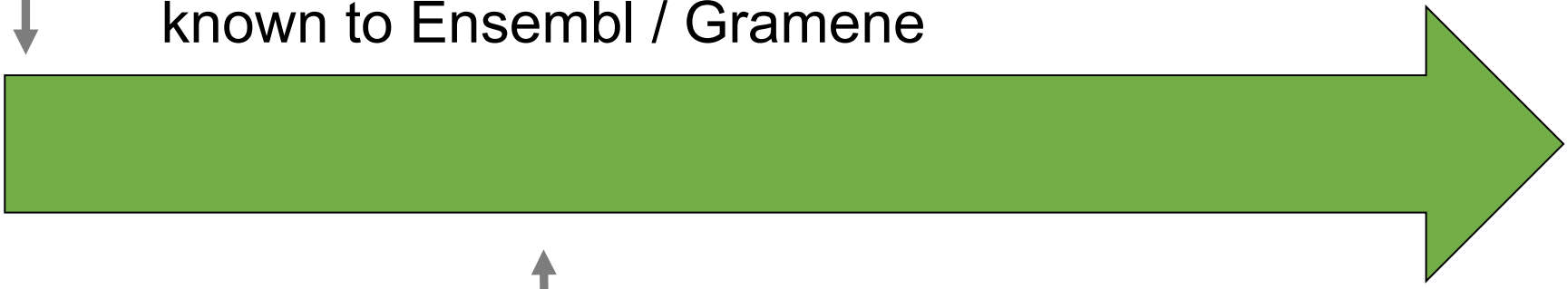
- Importing dbSNP variants
- Focus on species assemblies known to Ensembl / Gramene



# dbSNP > EVA transition working timeline

Sept ‘17

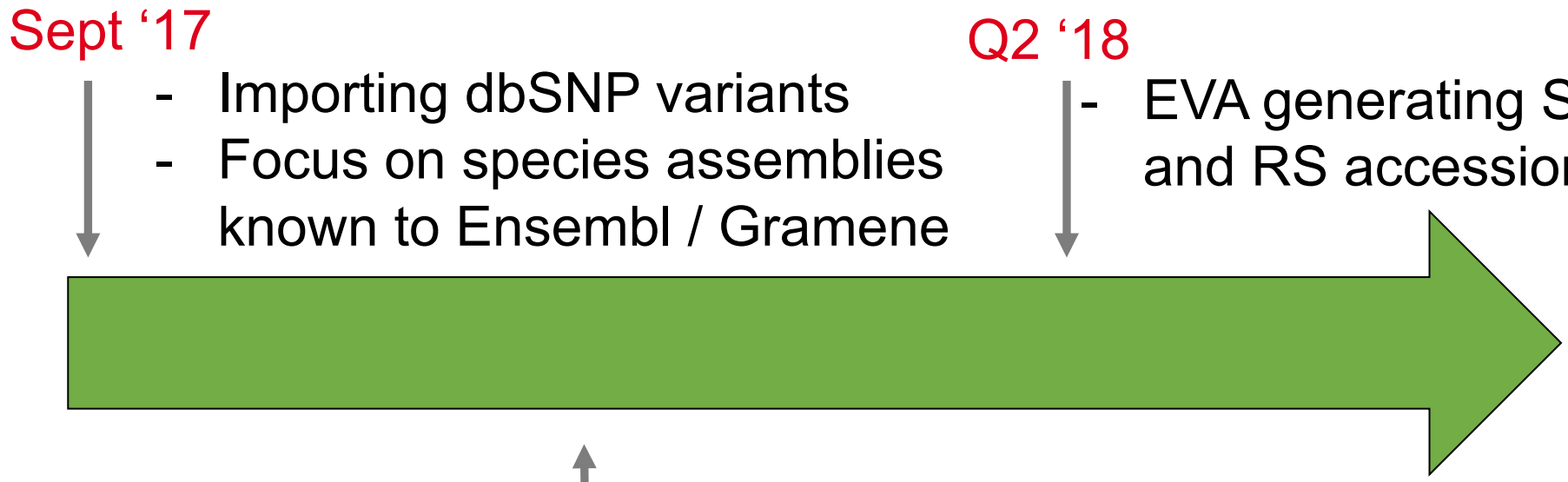
- Importing dbSNP variants
- Focus on species assemblies known to Ensembl / Gramene



- dbSNP turning off non-human data
- Remain available via FTP dumps

Q1 ‘18

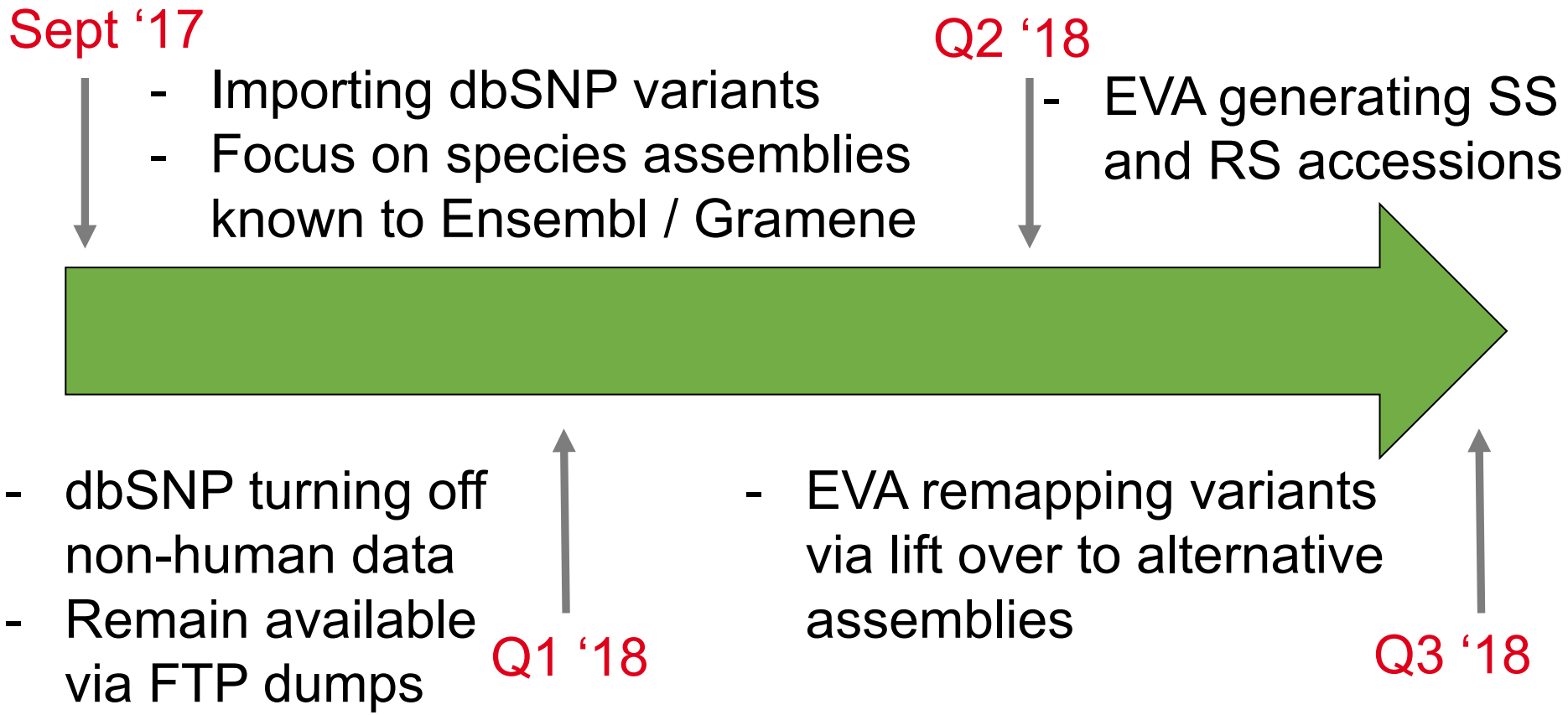
# dbSNP > EVA transition working timeline




- dbSNP turning off non-human data
- Remain available via FTP dumps

Q1 ‘18

# dbSNP > EVA transition working timeline



# Current status of dbSNP > EVA transition

- Importing dbSNP variant data to EVA
- Import completed for a number of species and can be checked at the EVA website summary table:
  - https://www.ebi.ac.uk/eva/?dbSNP-Import-Progress
- All variants imported; only variants that pass the EVA data requirements are shown
  - e.g. legacy data that does not match an assembly not shown
- Up-to-date dbSNP import status presented on the EVA website

# Current status of dbSNP > EVA transition

| Common name | Scientific name | Taxonomy ID | INSDC assembly accession | dbSNP build | All variants match INSDC assembly | Suitable for Variant Browser | Current dbSNP accessions searchable | Previous dbSNP accessions searchable | Supported by Ensembl |
|---|---|---|---|---|---|---|---|---|---|
| Cow | Bos taurus | 9913 | GCA_000003055.5 | 150 | ? | ✔ | In progress | | ✔ |
| Mouse | Mus musculus | 10090 | GCA_000001635.6 | 150 | ? | ✔ | In progress | | ✔ |
| Rat | Rattus norvegicus | 10116 | GCA_000001895.4 | 149 | ? | ✔ | In progress | | ✔ |
| Sorghum | Sorghum bicolor | 4558 | GCA_000003195.1 | 148 | ? | ✔ | In progress | | ✔ |
| Chicken | Gallus gallus | 9031 | GCA_000002315.3 | 151 | | ✔ | ✔ 10/1/2018 | | ✔ |
| Rice | Oryza sativa | 4530 | GCA_001433935.1 | 151 | ✔ | ✔ | ✔ 11/12/2017 | | ✔ |
| Pig | Sus scrofa | 9823 | GCA_000003025.6 | 150 | ✔ | ✔ | ✔ 12/1/2018 | | ✔ |
| Thale cress | Arabidopsis thaliana | 3702 | GCA_000001735.1 | 150 | | ✔ | ✔ 19/12/2017 | | ✔ |
| Goat | Capra hircus | 9925 | GCA_000317765.1 | 143 | ✔ | ✔ | ✔ 21/12/2017 | | |

http://www.ebi.ac.uk/eva/?dbSNP-Import-Progress

# Submission to EVA

- EVA accepts variant data described in Variant Call Format (VCF) files only
- Response time of 48 hrs
  - Accession number suitable for publication
- Data requirements:
  - The genome assembly used is International Nucleotide Sequence Database Collaboration (INSDC) registered
  - The variation data is described in valid VCF file(s)
  - We require that it be possible to compute allele frequencies for all submitted variants
    - (Contains genotypes or allele frequency (AF) values)

# EVA Archives Variant Call Format (VCF) Files

- Variant Call Format (VCF): the community standard way to describe genetic variants
- Currently on specification 4.3, EVA accepts all VCF versions

# EVA Archives Variant Call Format (VCF) Files

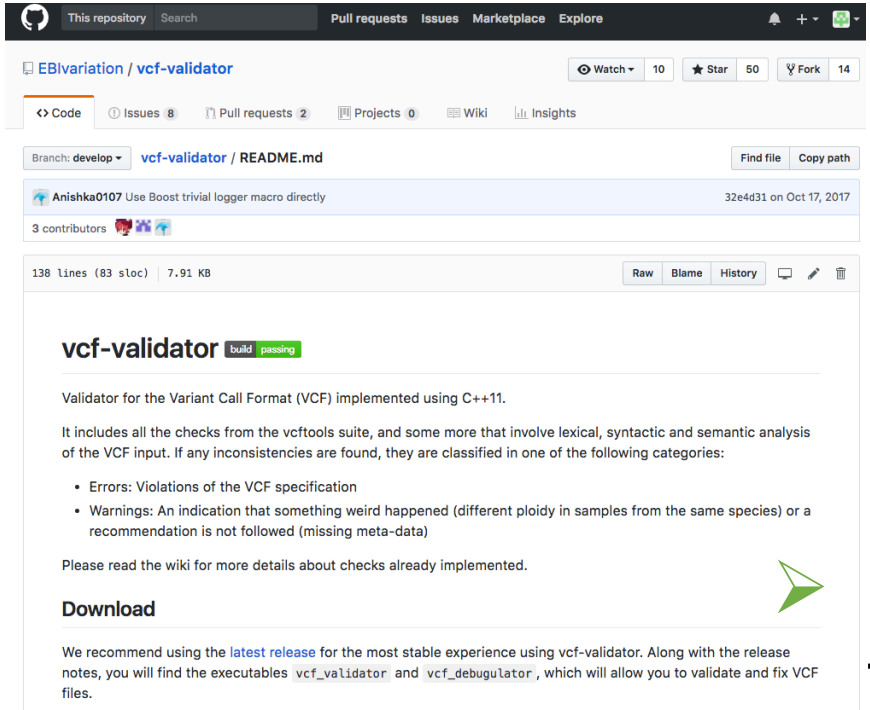
- Variant Call Format (VCF): the community standard way to describe genetic variants
- Currently on specification 4.3, EVA accepts all VCF versions

- Of all ca.4000 VCFs loaded to EVA <10% truly valid on first pass

**Most VCFs publically available are not truly valid**

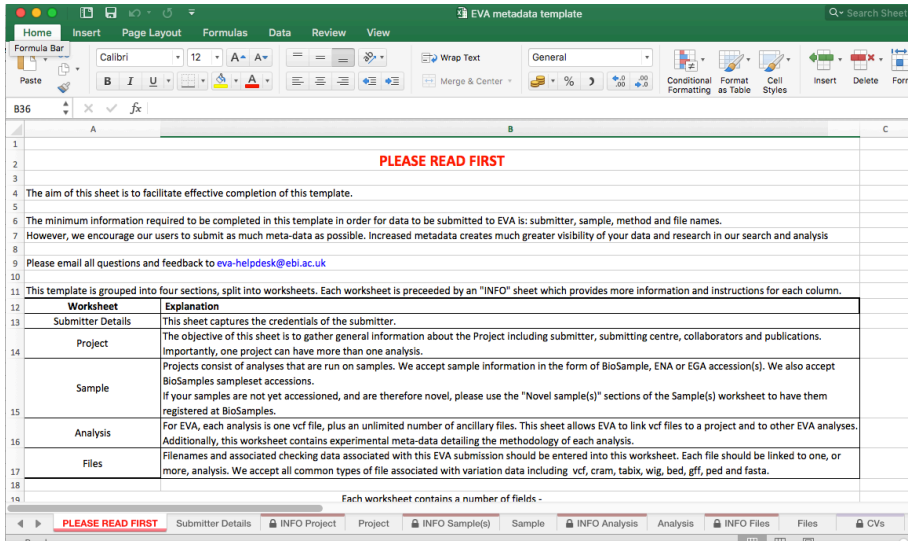
- EVA VCF validator tool used in submission processing to ensure all EVA VCF files are truly valid to specification
  - https://github.com/EBIvariation/vcf-validator

# EVA Archives Variant Call Format (VCF) Files




- Submission of pre-validated files improves processing time
- Works best on Linux machine
- Available as Docker container
- Plans to make available on Windows and MacOS X

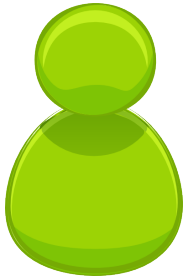
# VCF(s) associated with an EVA metadata template

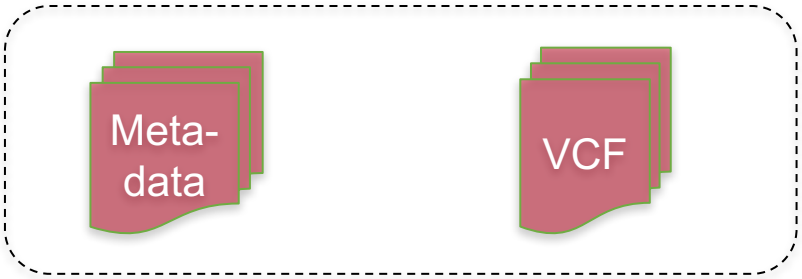



- Excel template
- Submitter
- Project
- Samples
- VCFs
- CVs, guidelines

- Can be filled manually or automatically
- Example template for a fictional study available at EVA website
  - https://www.ebi.ac.uk/eva/?Submit-Data

# Submission Pipeline to EVA

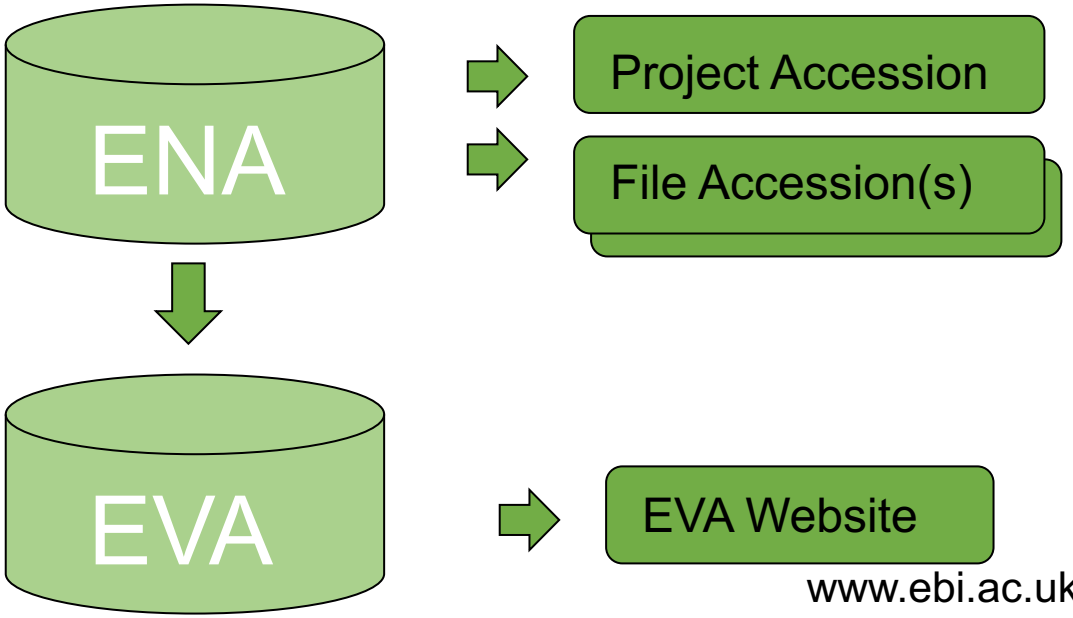





Custom Meta-data template:
- Data owner and Project summary
- Sample attributes
- Variant calling pipeline
- Description of each VCF

Validation
- Format
- Reference

EVA acts as a broker of valid VCF files to ENA. Permitting links to other projects (e.g. core read data, array data)




File level archive

www.ebi.ac.uk/eva

# EVA Data Access: EVA website

European Variation Archive

Home | Submit Data | Study Browser | Variant Browser | Clinical Browser | GA4GH | API | dbSNP Import Progress | Help | Feedback

EVA / HOME

Support for non-human variant data archival and accessioning is transitioning from dbSNP to EVA from September 2017. [Read more]

## Overview

The European Variation Archive is an open-access database of all types of genetic variation data from all species.

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.

## News

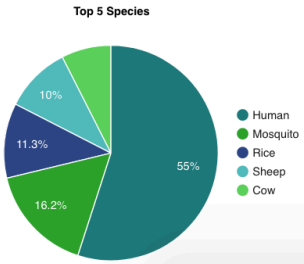
EVA Retweeted

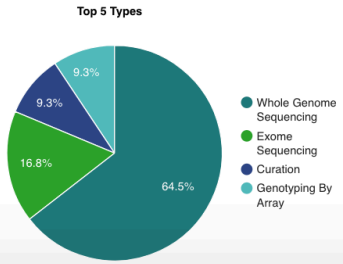
**Thomas Keane** @drtkeane

We are looking for an enthusiastic bioinformatician to be the front face of the European Variation Archive. Great opportunity to interact with many species communities, and gain practical bioinf experience. embl.de/jobs/searchjob…

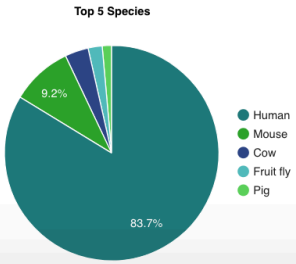
## Statistics

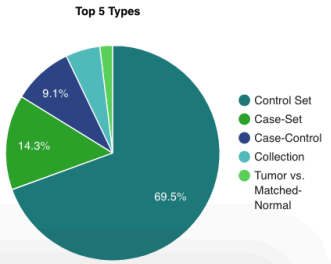
### Short genetic variants studies (<50bp)




### Structural variants studies (>50bp)

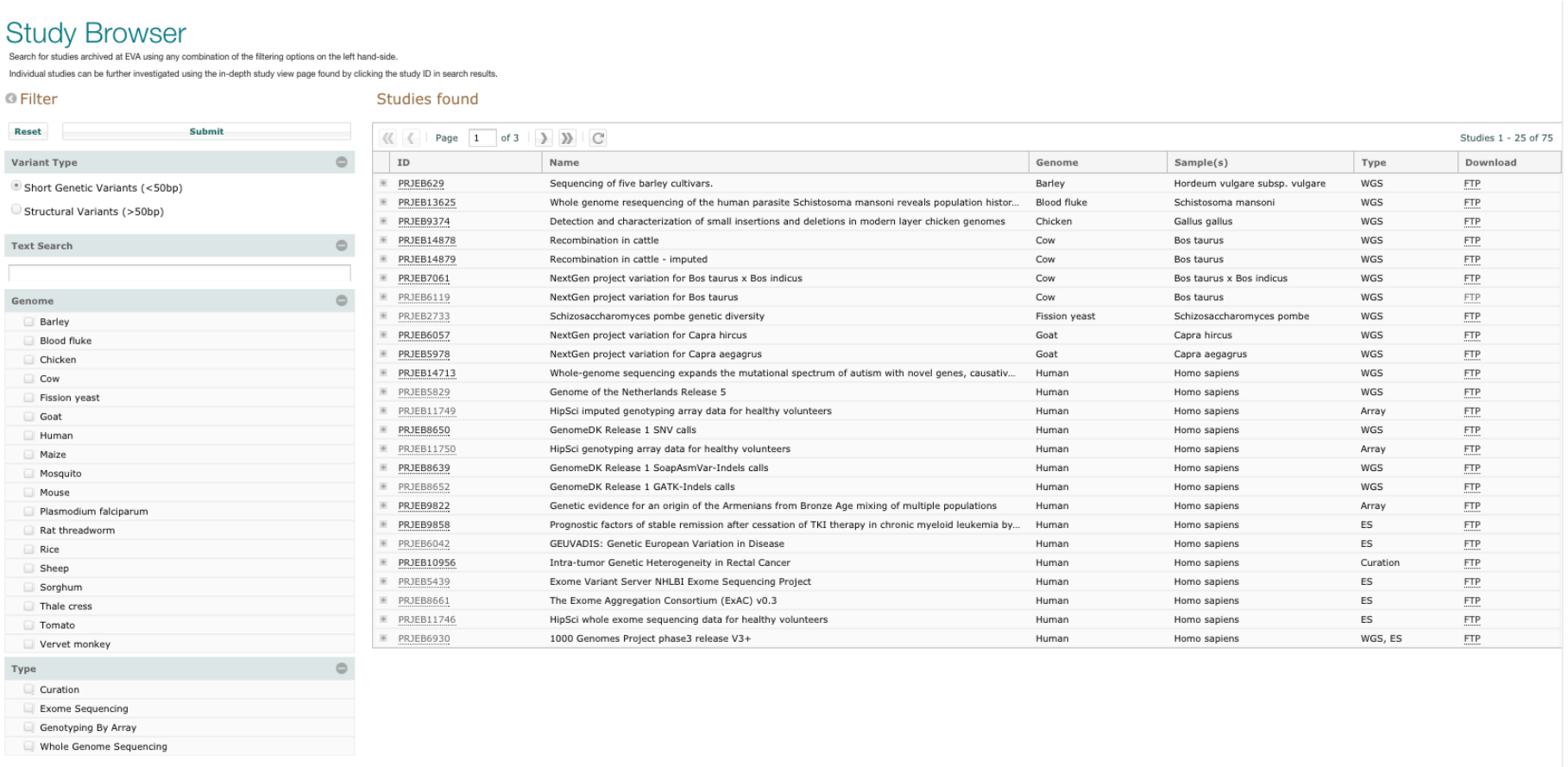



EMBL-EBI

# EVA Data Access: Study Browser

- Core functionality: portal to open-access genetic variation data submitted to EVA:



## Study Browser

Search for studies archived at EVA using any combination of the filtering options on the left hand-side.

Individual studies can be further investigated using the in-depth study view page found by clicking the study ID in search results.

### Filter

Reset | Submit

**Variant Type**

- Short Genetic Variants (<50bp)
- Structural Variants (>50bp)

**Text Search**

**Genome**

- Barley
- Blood fluke
- Chicken
- Cow
- Fission yeast
- Goat
- Human
- Maize
- Mosquito
- Mouse
- Plasmodium falciparum
- Rat threadworm
- Rice
- Sheep
- Sorghum
- Thale cress
- Tomato
- Vervet monkey

**Type**

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

### Studies found

Page 1 of 3 — Studies 1 - 25 of 75

| ID | Name | Genome | Sample(s) | Type | Download |
|---|---|---|---|---|---|
| PRJEB629 | Sequencing of five barley cultivars. | Barley | Hordeum vulgare subsp. vulgare | WGS | FTP |
| PRJEB13625 | Whole genome resequencing of the human parasite Schistosoma mansoni reveals population histor... | Blood fluke | Schistosoma mansoni | WGS | FTP |
| PRJEB9374 | Detection and characterization of small insertions and deletions in modern layer chicken genomes | Chicken | Gallus gallus | WGS | FTP |
| PRJEB14878 | Recombination in cattle | Cow | Bos taurus | WGS | FTP |
| PRJEB14879 | Recombination in cattle - imputed | Cow | Bos taurus | WGS | FTP |
| PRJEB7061 | NextGen project variation for Bos taurus x Bos indicus | Cow | Bos taurus x Bos indicus | WGS | FTP |
| PRJEB6119 | NextGen project variation for Bos taurus | Cow | Bos taurus | WGS | FTP |
| PRJEB2733 | Schizosaccharomyces pombe genetic diversity | Fission yeast | Schizosaccharomyces pombe | WGS | FTP |
| PRJEB6057 | NextGen project variation for Capra hircus | Goat | Capra hircus | WGS | FTP |
| PRJEB5978 | NextGen project variation for Capra aegagrus | Goat | Capra aegagrus | WGS | FTP |
| PRJEB14713 | Whole-genome sequencing expands the mutational spectrum of autism with novel genes, causativ... | Human | Homo sapiens | WGS | FTP |
| PRJEB5829 | Genome of the Netherlands Release 5 | Human | Homo sapiens | WGS | FTP |
| PRJEB11749 | HipSci imputed genotyping array data for healthy volunteers | Human | Homo sapiens | Array | FTP |
| PRJEB8650 | GenomeDK Release 1 SNV calls | Human | Homo sapiens | WGS | FTP |
| PRJEB11750 | HipSci genotyping array data for healthy volunteers | Human | Homo sapiens | Array | FTP |
| PRJEB8639 | GenomeDK Release 1 SoapAsmVar-Indels calls | Human | Homo sapiens | WGS | FTP |
| PRJEB8652 | GenomeDK Release 1 GATK-Indels calls | Human | Homo sapiens | WGS | FTP |
| PRJEB9822 | Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations | Human | Homo sapiens | Array | FTP |
| PRJEB9858 | Prognostic factors of stable remission after cessation of TKI therapy in chronic myeloid leukemia by... | Human | Homo sapiens | ES | FTP |
| PRJEB6042 | GEUVADIS: Genetic European Variation in Disease | Human | Homo sapiens | ES | FTP |
| PRJEB10956 | Intra-tumor Genetic Heterogeneity in Rectal Cancer | Human | Homo sapiens | Curation | FTP |
| PRJEB5439 | Exome Variant Server NHLBI Exome Sequencing Project | Human | Homo sapiens | ES | FTP |
| PRJEB8661 | The Exome Aggregation Consortium (ExAC) v0.3 | Human | Homo sapiens | ES | FTP |
| PRJEB11746 | HipSci whole exome sequencing data for healthy volunteers | Human | Homo sapiens | ES | FTP |
| PRJEB6930 | 1000 Genomes Project phase3 release V3+ | Human | Homo sapiens | WGS, ES | FTP |

# EVA Study View

- One-page summary of each project: data submitter, publication, number of samples, brief description, etc.

EVA / STUDY BROWSER / STUDY VIEW

1001 Genomes: A Catalog of Arabidopsis thaliana Genetic Variation

General Information

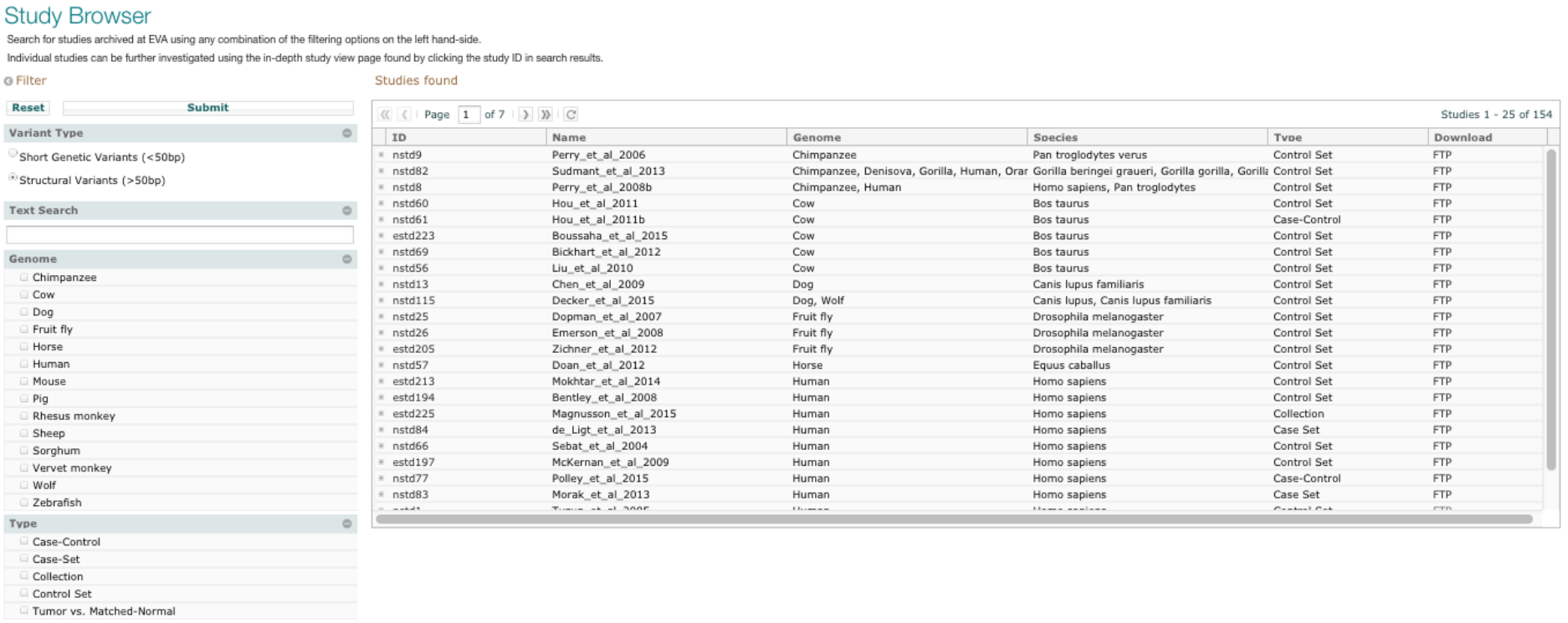
| | |
|---|---|
| **Organism** | Thale cress |
| **Scientific Name** | Arabidopsis thaliana |
| **Taxonomy ID** | 3702 |
| **Center** | The 1001 Genomes Project Consortium |
| **Material** | DNA |
| **Scope** | multi-isolate |
| **Type** | Whole Genome Sequencing |
| **Genome Assembly** | GCA_000001735.1 |
| **Source Type** | Germline |
| **Platform** | Illumina HiSeq 2000 |
| **Samples** | 1135 |
| **Description** | 1001 Genomes: A Catalog of Arabidopsis thaliana Genetic Variation |
| **Resource** | http://1001genomes.org/about.html<br>ENA:PRJNA273563 |
| **Download** | FTP |
| **Publications** | 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana.<br>1001 Genomes Consortium. Electronic address: magnus.nordborg@gmi.oeaw.ac.at, 1001 Genomes Consortium.<br>*Cell* **166**:2016 481-491 |

Files

| File Name | Samples with Genotypes | Variants Count | SNP Count | Indel Count | Pass Count | Transitions/Transversions Ratio | Mean Quality |
|---|---|---|---|---|---|---|---|
| 1001genomes_snp-short-indel_only_ACGTN_v3.1.snpeff.garys.final.vcf.gz | 1135 | 14144402 | 12545539 | 1598863 | 14144402 | 1.09 (6823574/6272040) | 33.89 |

# EVA Study Browser – Structural Variants

- SVs from EMBL-EBI resource Database of Genomic Variants archive (DGVa)



# EVA Study View – Structural Variants

- One-page summary of each project: data submitter, publication, number of samples, brief description, etc.

Zheng_et_al_2011

General Information

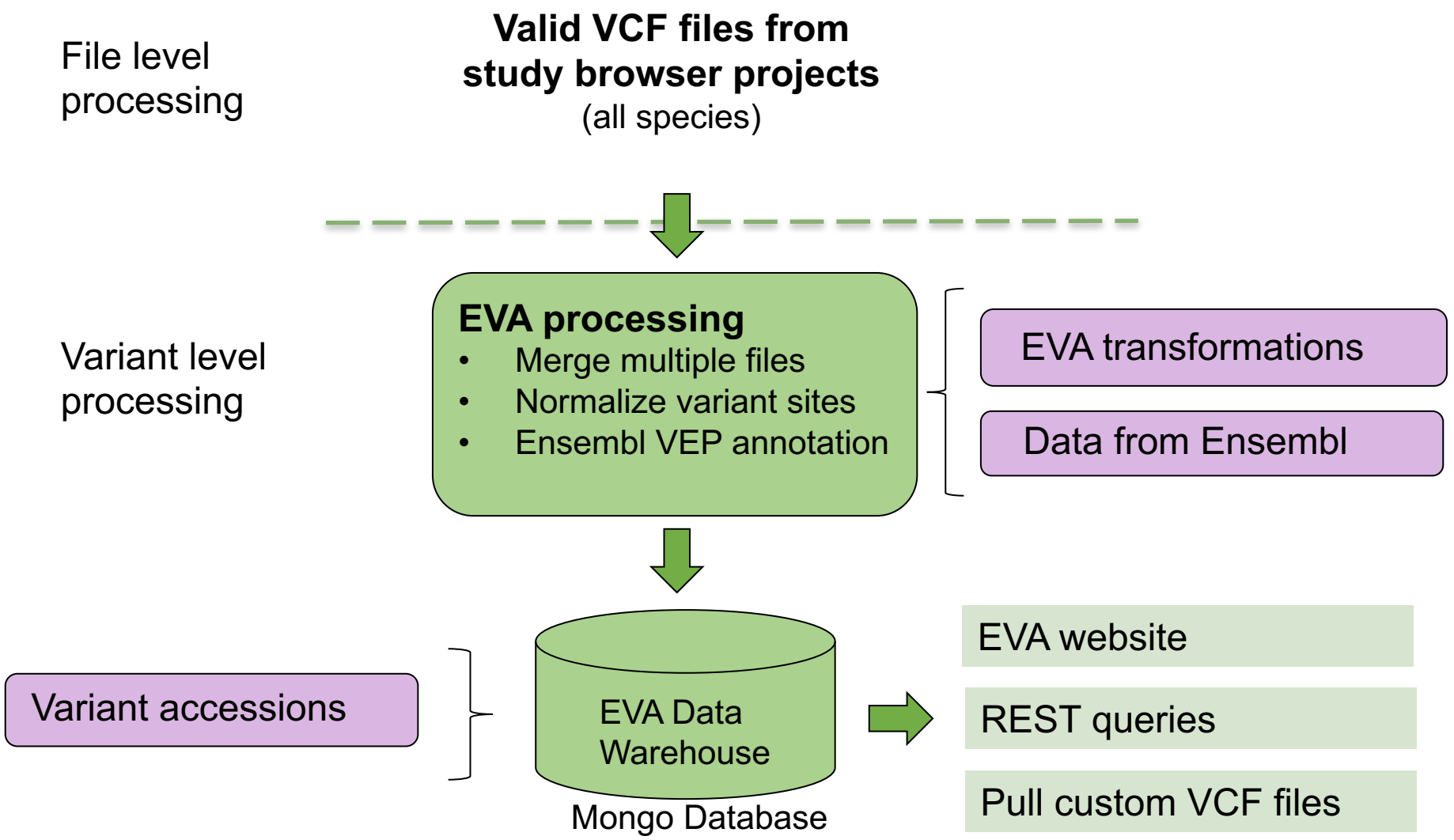
| | |
|---|---|
| **Genome** | Sorghum |
| **Sample(s)** | Sorghum bicolor |
| **Taxonomy ID** | 4558 |
| **Study Type** | Control Set |
| **Experiment Type** | Paired-end mapping, Read depth |
| **Platform** | Illumina Solexa Genome Analyzer |
| **Genome Assembly** | Sorbi1 |
| **Number of Variants** | 64507 |
| **Description** | Sorghum (Sorghum bicolor) is globally produced as a source of food, feed, fibre and fuel. Grain and sweet sorghums differ in a number of important traits including stem sugar and juice accumulation, plant height and production of grain and biomass. The first whole genome sequence of a grain sorghum is available, but additional genome sequences are required to study genome-wide and intraspecies variation for dissecting the genetic basis of these important traits and for tailor-designed breeding of this important C4 crop. We resequenced two sweet and one grain sorghum inbred lines, and identified a set of nearly 1,500 genes differentiating sweet and grain sorghum. In addition, we uncovered 1,057,018 SNPs, 99,948 indels of 1-10bp in length and 16,487 presence/absence variations. In addition, 17,111 CNVs were detected. This is a first report on the identification of genome-wide patterns of genetic variation in sorghum. |
| **Download** | FTP |
| **Publications** | Genome-wide patterns of genetic variation in sweet and grain sorghum (Sorghum bicolor).<br>Zheng LY, Guo XS, He B, Sun LJ, Peng Y, Dong SS, Liu TF, Jiang S, Ramachandran S, Liu CM, Jing HC.<br>*Genome Biol* **12**:2011 R114 |

# EVA Added Value – Variant processing pipeline

File level processing

**Valid VCF files from study browser projects**
(all species)

# EVA Added Value – Variant processing pipeline



# EVA Processing Steps Documented

- EVA FAQs

## FAQ

**What is the European Variation Archive (EVA)?**

The European Variation Archive (EVA) is EMBL-EBI's open-access genetic variation archive. The EVA accepts submission of all types of genetic variants, ranging from single nucleotide polymorphisms to large structural variants, observed in germline or somatic sources, from any eukaryotic organism. The EVA permits access to these data at two distinct levels:

i) The raw variant data as was submitted to the EVA, via the EVA Study Browser

ii) The normalised and processed variant data, via the EVA Variant Browser and EVA API

Is my data suitable for submission to the EVA?

Why submit data to the EVA?

What happens to my data once submitted?

What are the EVA normalisation and variant processing steps?

What data is shown in the EVA Clinical Browser?

With whom does the EVA collaborate?

Which browsers does the EVA website support?

## Normalisation

Variants submitted to the EVA have been determined by a number of different algorithms and software packages. As a result, the VCF files generated by these differing methodologies describe variants in a number of different ways. The primary processing step of the EVA is to normalise variant representation following two basic rules:

1. Each variant is shifted to be left-aligned
2. The Start and End positions represent exactly the range where the variation occurs (which could, in the case of insertions, result in the reference allele being recorded as 'empty')

Examples of our variant normalisation process can be seen here

## Annotation

Once variants have been normalised, the EVA uses the Variant Effect Predictor (VEP) of Ensembl to annotate variant consequences. The variant consequences are described using Sequence Ontology terms and both the VEP version and Ensembl gene build used are described via the "i" help bubbles on the EVA Variant Browser.

N.B. Variants that have been mapped to a reference genome sequence that is not supported by Ensembl are not annotated.

## Statistical calculations

The EVA adopts the classical definition of allele frequency (AF): 'a measure of the relative frequency of an allele at a genetic locus in a given population'. The AF value(s) stored by the EVA for each variant is (are) study specific - i.e. the same variant reported in two distinct studies shall be given two allele frequencies, one for each study. There are two methodologies by which the EVA is able to determine allele frequency values, dependent on the datatype of the study in question:

**Variants associated with genotypes:**

For variants associated with genotypes, the EVA determines the AF values via the calculation:

AF = (number of alternate allele observations (AC)) / (number of observations (AN))

The result of this calculation allows the EVA to also store the minor allele frequency (MAF) for each variant (defined as the minimum of the reference or alternative allele frequency) and the MAF allele (the allele associated with the MAF).

**Variants not associated with genotypes:**

For variants that are not associated with genotypes, the EVA is dependent on the AF value(s) estimated from the primary data and provided in the submitted VCF files(s). AF values that are specifically provided in the submitted aggregated VCF file(s) are directly stored. In cases where no AF is provided then the EVA uses the AC and AN values in the submitted aggregated VCF file(s) to calculate AF value(s) via the calculation:

AF = AC / AN

**Population / sample cohort allele frequency values:**

The EVA accepts submission of pedigree files, or structured samples (using "derived_from" and/or "subject" layers), to define populations and cohorts within studies. In cases where such information is associated with variants that have genotypes then the EVA calculates intra-study population/cohort specific AF values via the method described above, with the caveat that the (total number of populations/cohorts):(total number of samples) ratio must be less than 1:10. For studies that do not contain genotypes but instead provide intra-study population/cohort AF values in the submitted aggregated VCF file(s), or AC and AN values, then these are directly stored, or calculated by the EVA using the method described above, again with the caveat that a ratio of 1:10 (total number of populations/cohorts):(total number of samples) must not be exceeded.

# EVA Variant Browser

- Build view(s) of the data within EVA variant warehouse
- 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 Variant Browser

## Variant Browser

Search the EVA variant warehouse using any combination of the filtering options on the left hand-side.

Search results can be exported in CSV format and individual variants can be further investigated using the in-depth Variant Data tabs found below the main results table.

### Filter

Reset | Submit

**Genome Assembly**

Organism / Assembly:

Cow / Bos_taurus_UMD_3.1

**Position**

Filter By:

Chromosomal Location

1:3000000-3100000

**Consequence Type**

**Minor Allele Frequency**

MAF: ex: >= | ex: 0.3

**Protein Substitution Score**

PolyPhen2 >: ex:0.5

Sift <: ex:0.1

**Studies Mapped To Assembly**

search

| Name |
|---|
| NextGen project variation for Bos taurus (PRJEB6119) |
| NextGen project variation for Bos taurus x Bos indicus (PRJEB7061) |
| Recombination in cattle (PRJEB14878) |

### Variants found

Page 1 of 1

| Chr | Position | Variant ID | Alleles | Class | Most Severe Consequence Type |
|---|---|---|---|---|---|
| 1 | 3075045 | ss1083280832 | C/T | SNV | synonymous_variant |
| 1 | 3082169 | ss1083280845 | A/G | SNV | missense_variant |
| 1 | 3082172 | - | G/A | SNV | missense_variant |
| 1 | 3082603 | ss1083280847 | A/G | SNV | synonymous_variant |
| 1 | 3095575 | ss1083280903 | A/C | SNV | synonymous_variant |

Results per Page: 10 | Export as CSV

### Variant Data

Annotation | Files | Genotypes | Population Statistics

#### Annotations

Page 1 of 1

| Ensembl Gene ID | Ensembl Gene Symbol | Ensembl Transcript ID | Ensembl Transcript Biotype | SO Term(s) |
|---|---|---|---|---|
| ENSBTAG00000018855 | SCAF4 | ENSBTAT00000025104 | protein_coding | synonymous_variant |

EMBL-EBI

# EVA Variant Browser

## Variant Browser

Search the EVA variant warehouse using any combination of the filtering options on the left hand-side.

Search results can be exported in CSV format and individual variants can be further investigated using the in-depth Variant Data tabs found below the main results table.

### Filter

Reset | Submit

**Genome Assembly**

Organism / Assembly:

Cow / Bos_taurus_UMD_3.1

**Position**

Filter By:

Chromosomal Location

1:3000000-3100000

**Consequence Type**

**Minor Allele Frequency**

MAF: ex: >= | ex: 0.3

**Protein Substitution Score**

PolyPhen2 >: ex:0.5

Sift <: ex:0.1

**Studies Mapped To Assembly**

search

Name

- NextGen project variation for Bos taurus (PRJEB6119)
- NextGen project variation for Bos taurus x Bos indicus (PRJEB7061)
- Recombination in cattle (PRJEB14878)

### Variants found

Page 1 of 1

| Chr | Position | Variant ID | Alleles | Class | Most Severe Consequence Type |
|---|---|---|---|---|---|
| 1 | 3075045 | ss1083280832 | C/T | SNV | synonymous_variant |
| 1 | 3082169 | ss1083280845 | A/G | SNV | missense_variant |
| 1 | 3082172 | - | G/A | SNV | missense_variant |
| 1 | 3082603 | ss1083280847 | A/G | SNV | synonymous_variant |
| 1 | 3095575 | ss1083280903 | A/C | SNV | synonymous_variant |

Results per Page: 10 | Export as CSV

### Variant Data

Annotation | Files | Genotypes | Population Statistics

#### Files

NextGen project variation for Bos taurus (PRJEB6119 - **ERZ019405**)

| AC | NS | AN | FILTER | AF |
|---|---|---|---|---|
| 1 | 8 | 16 | PASS | 0.0625 |

Show Full Header

EMBL-EBI

# EVA Variant Browser

## Variant Browser

Search the EVA variant warehouse using any combination of the filtering options on the left hand-side.

Search results can be exported in CSV format and individual variants can be further investigated using the in-depth Variant Data tabs found below the main results table.

**NextGen project variation for Bos taurus (PRJEB6119 - ERZ019405)**

| AC | NS | AN | FILTER | AF |
|---|---|---|---|---|
| 1 | 8 | 16 | PASS | 0.0625 |

Hide Full Header

```
##fileformat=VCFv4.1
##EVA_ProjectName="NextGen project variation for Bos (cattle)"
##EVA_ProjectDescription="Variation site discovery and genotype calls in the genus Bos (cattle) by the NEXTGEN project (Next generation methods to preserve farm
##EVA_AnalysisTitle="Iranian Bos taurus (cattle) filtered variation calls from whole genome sequencing"
##EVA_PipelineDescription="This data set represents all variation discovered in the Iranian Bos taurus (cattle) population passing the NextGen project filters.
##EVA_Scope=multi-isolate
##EVA_Selection=genome
##EVA_Material=DNA
##EVA_TaxID=9913
##EVA_Platform="Illumina"
##EVA_Sample=<ID=IRBT-H9-0008,ACCESSION=ERS403311,TAXID=9913,SCIENTIFIC_NAME="Bos taurus",COMMON_NAME="cattle",COUNTRY="Iran",CLOSEST_CITY="Sanandaj",CLOSEST_LO
```

EVA Variant Browser

Name

- NextGen project variation for Bos taurus (PRJEB6119)
- NextGen project variation for Bos taurus x Bos indicus (PRJEB7061)
- Recombination in cattle (PRJEB14878)

| AC | NS | AN | FILTER | AF |
|---|---|---|---|---|
| 1 | 8 | 16 | PASS | 0.0625 |

Show Full Header

EMBL-EBI

# EVA Variant Browser

## Variant Browser

Search the EVA variant warehouse using any combination of the filtering options on the left hand-side.

Search results can be exported in CSV format and individual variants can be further investigated using the in-depth Variant Data tabs found below the main results table.

### Filter

Reset | Submit

**Genome Assembly**

Organism / Assembly:

Cow / Bos_taurus_UMD_3.1

**Position**

Filter By:

Chromosomal Location

1:3000000-3100000

**Consequence Type**

**Minor Allele Frequency**

MAF: ex: >= | ex: 0.3

**Protein Substitution Score**

PolyPhen2 >: ex:0.5

Sift <: ex:0.1

**Studies Mapped To Assembly**

search

| Name |
|---|
| NextGen project variation for Bos taurus (PRJEB6119) |
| NextGen project variation for Bos taurus x Bos indicus (PRJEB7061) |
| Recombination in cattle (PRJEB14878) |

### Variants found

Page 1 of 1

| Chr | Position | Variant ID | Alleles | Class | Most Severe Consequence Type |
|---|---|---|---|---|---|
| 1 | 3075045 | ss1083280832 | C/T | SNV | synonymous_variant |
| 1 | 3082169 | ss1083280845 | A/G | SNV | missense_variant |
| 1 | 3082172 | - | G/A | SNV | missense_variant |
| 1 | 3082603 | ss1083280847 | A/G | SNV | synonymous_variant |
| 1 | 3095575 | ss1083280903 | A/C | SNV | synonymous_variant |

Results per Page: 10 | Export as CSV

### Variant Data

Annotation | Files | Genotypes | Population Statistics

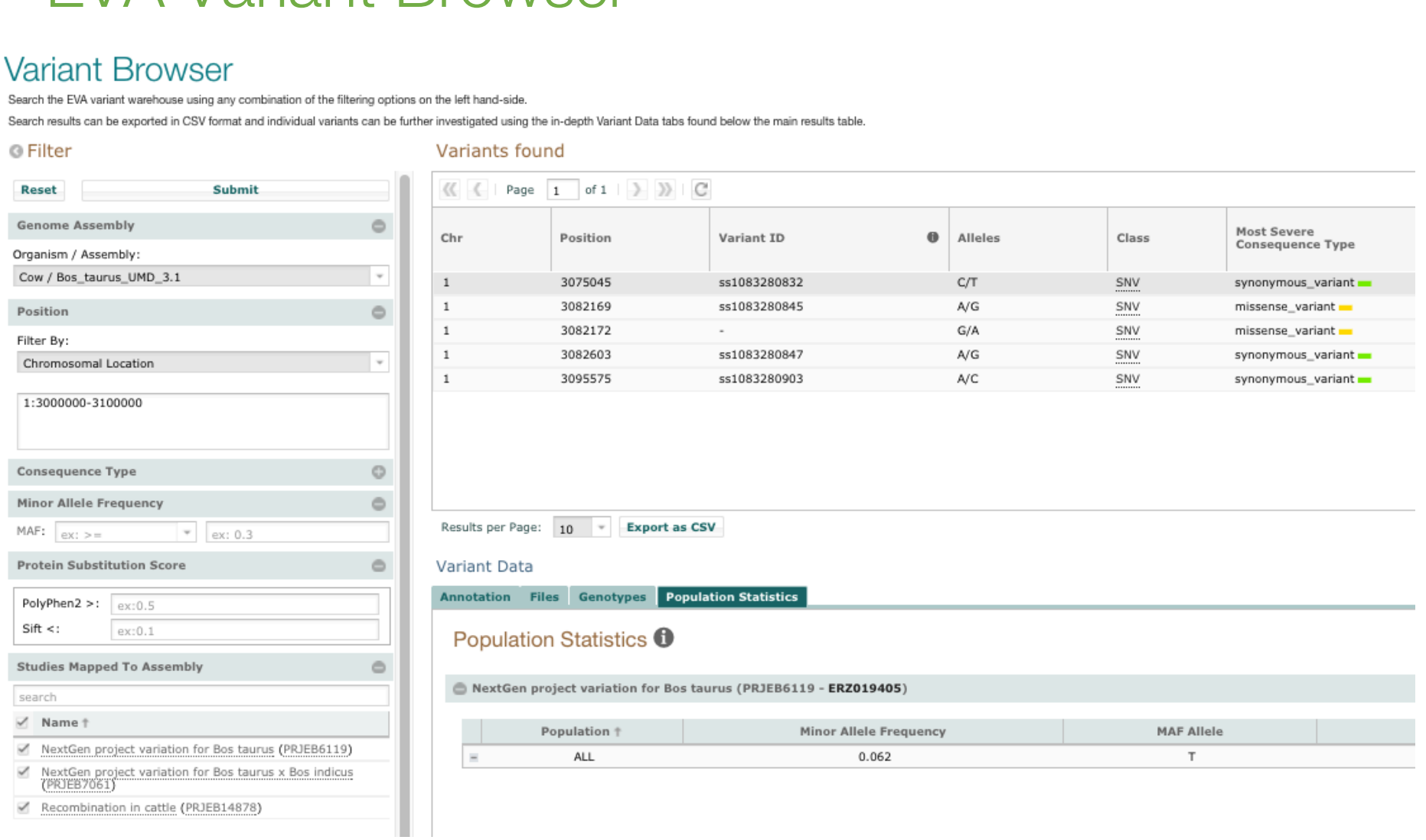
#### Genotypes

NextGen project variation for Bos taurus (PRJEB6119 - **ERZ019405**)

| Sample | Genotype |
|---|---|
| IRBT-E10-0001 | 0\|0 |
| IRBT-E10-0002 | 0\|0 |
| IRBT-F10-0003 | 0\|1 |
| IRBT-G8-0004 | 0\|0 |
| IRBT-G8-0005 | 0\|0 |
| IRBT-H9-0006 | 0\|0 |
| IRBT-H9-0007 | 0\|0 |


EMBL-EBI

# EVA Variant Browser




EMBL-EBI

# EVA API

- EVA Variant 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

# EVA API

## API

### The general structure of a EVA RESTful web service URL is:

http://www.ebi.ac.uk/eva/webservices/rest/{version}/{category}/IDs/{resource}?{filters}

### Where:

* *version*: indicates the version of the API, this defines the available filters and JSON schema to be returned. Currently there is only version 'v1'.
* *category*: this defines what objects we want to query. Currently there are five different categories: variants, segments, genes, files and studies.
* *resource*: specifies the resource to be returned, therefore the JSON data model.
* *filters*: each specific endpoint allows different filters.

RESTfull web services have been implemented using GET protocol since only queries are allowed so far. Several IDs can be concatenated using comma as separator. For more detailed information about the API and filters you can visit the project wiki and Swagger documentation.

### Some example of queries include:

* To fetch all the variants in a segment region:
http://www.ebi.ac.uk/eva/webservices/rest/v1/segments/11:128446-128446/variants?species=hsapiens_grch37

* To fetch all the info of a variant:
http://www.ebi.ac.uk/eva/webservices/rest/v1/variants/rs666/info?species=hsapiens_grch37

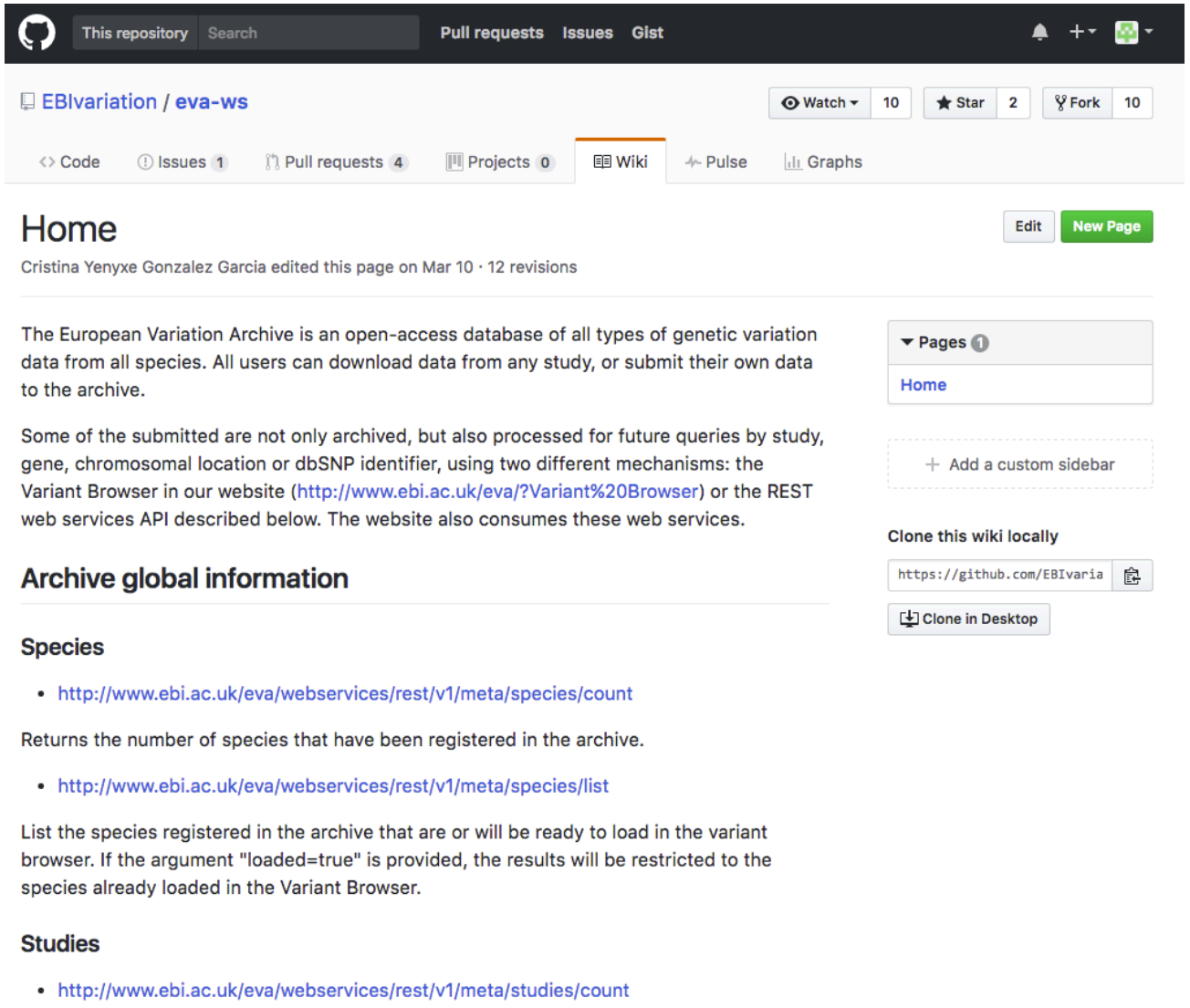
* To fetch all the Short Genetics Variations studies:
http://www.ebi.ac.uk/eva/webservices/rest/v1/meta/studies/all

* To fetch all the Structural Variations studies:
http://www.ebi.ac.uk/eva/webservices/rest/v1/meta/studies/all?structural=true

* To fetch all info of a study:
http://www.ebi.ac.uk/eva/webservices/rest/v1/studies/PRJEB4019/summary

* To fetch all file information of a study:
http://www.ebi.ac.uk/eva/webservices/rest/v1/studies/PRJEB4019/files?species=hsapiens_grch37

# EVA API


This repository
Search
Pull requests
Issues
Gist
EBIvariation / eva-ws
Watch 10
Star 2
Fork 10
Code
Issues 1
Pull requests 4
Projects 0
Wiki
Pulse
Graphs
Home
Edit
New Page
Cristina Yenyxe Gonzalez Garcia edited this page on Mar 10 · 12 revisions
The European Variation Archive is an open-access database of all types of genetic variation data from all species. All users can download data from any study, or submit their own data to the archive.
Some of the submitted are not only archived, but also processed for future queries by study, gene, chromosomal location or dbSNP identifier, using two different mechanisms: the Variant Browser in our website (http://www.ebi.ac.uk/eva/?Variant%20Browser) or the REST web services API described below. The website also consumes these web services.
Archive global information
Species
http://www.ebi.ac.uk/eva/webservices/rest/v1/meta/species/count
Returns the number of species that have been registered in the archive.
http://www.ebi.ac.uk/eva/webservices/rest/v1/meta/species/list
List the species registered in the archive that are or will be ready to load in the variant browser. If the argument "loaded=true" is provided, the results will be restricted to the species already loaded in the Variant Browser.
Studies
http://www.ebi.ac.uk/eva/webservices/rest/v1/meta/studies/count
Pages 1
Home
Add a custom sidebar
Clone this wiki locally
https://github.com/EBIvaria
Clone in Desktop

# Other input

- Bug reports / feature requests welcome
- Contact:
  - E-mail: eva-helpdesk@ebi.ac.uk
  - Open ticket at github repo:
    - https://github.com/EBIvariation/eva-web
    - https://github.com/EBIvariation/eva-ws

# Training Materials

- European Variation Archive at EMBL-EBI: webinar

Train online

Training | Train online | About Train online | Glossary | Support and feedback | Login/register

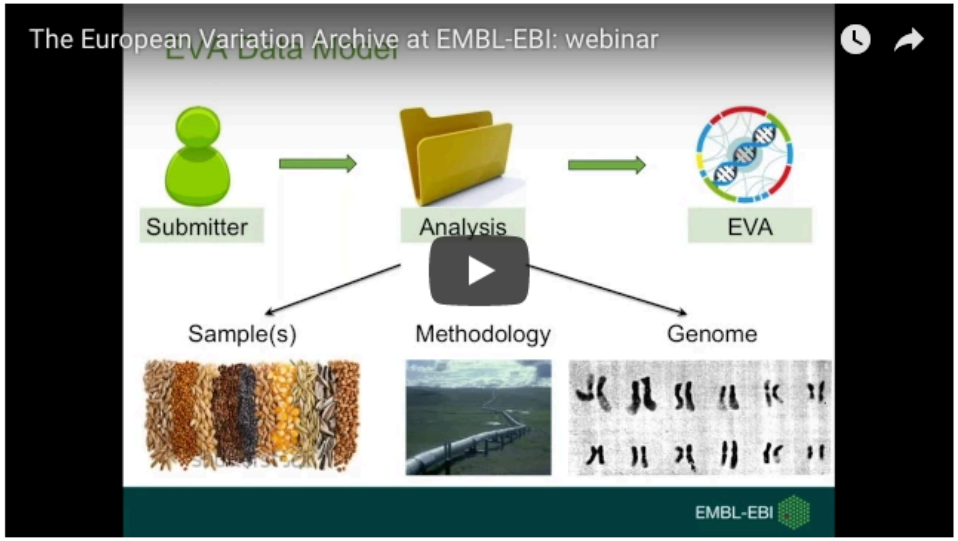
European Variation Archive at EMBL-EBI: webinar

Your feedback

Contributors



## European Variation Archive at EMBL-EBI: webinar



# Training Materials

- European Variation Archive at EMBL-EBI: webinar
- EVA, programmatically: webinar

Train online

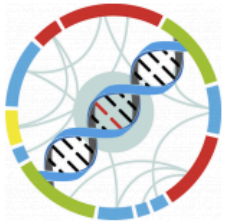
Training | Train online | About Train online | Glossary | Support and feedback | Login/register

EMBL-EBI, PROGRAMMATICALLY: TAKE A REST FROM MANUAL SEARCHES / EVA, PROGRAMMATICALLY

EMBL-EBI, programmatically: take a REST from manual searches

- Introduction to EMBL-EBI resources
- Introduction to programmatic access
- Europe PMC, programmatically
- Ensembl, programmatically
- UniProt, programmatically

## EVA, programmatically



The European Variation Archive (EVA) is an open-access database of all types of genetic variation data from all species.

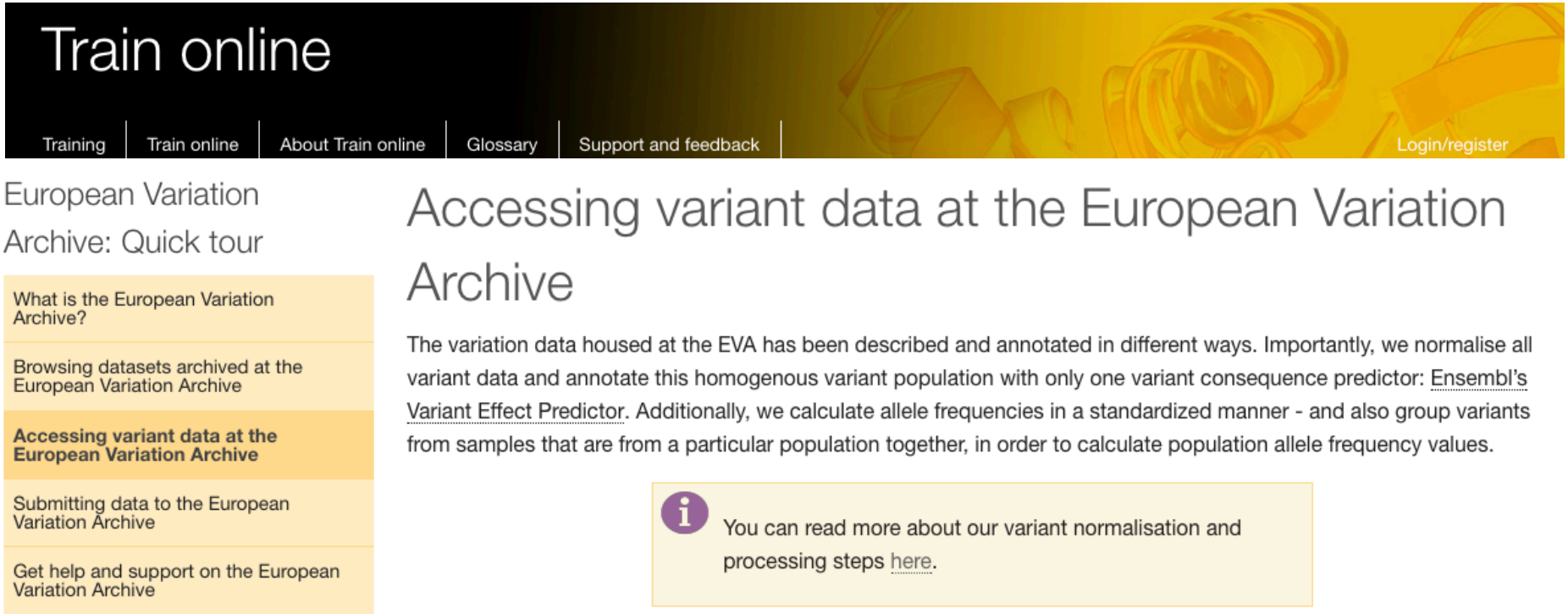
Using EVA, you can download data or submit your own data to the archive. You can also query variants in the EVA by study, gene, chromosomal location or variant consequence using our variant browser.

You can learn more about the EVA in our general introduction webinar, or via our Quick Tour.

EMBL-EBI

# Training Materials

- European Variation Archive at EMBL-EBI: webinar
- EVA, programmatically: webinar
- EVA Quick Tour: ca.30 mins, basics of using EVA



# Conclusion

European Variation Archive
www.ebi.ac.uk/eva

- All types of variants, all species
- Administers long-term variant IDs
- Provides direct views of variant data
- Full API
- Completely free to use

# Acknowledgments


## EVA / DGVA

Thomas Keane

Cristina Yenyxe Gonzalez

Pablo Arce Garcia

Jose Miguel Mut

Sundararaman Venkataraman

Tom Smith

Jag Kandasamy

## EGA

Dylan Spalding

Jeff Almeida-King

Saif Ur-Rehman

Jorge Izquierdo Ciges

## Ensembl Variation

Fiona Cunningham

Sarah Hunt

William McLaren

Anja Thormann

Laurent Gil

## ENA team

Rasko Leinonen

Marc Rosello

Daniel Vaughan

## Ensembl Genomes

Paul Kersey

Dan Bolser

Gareth Maslen

Kevin Howe

## Funding


@evarchive

# General Qs / feedback