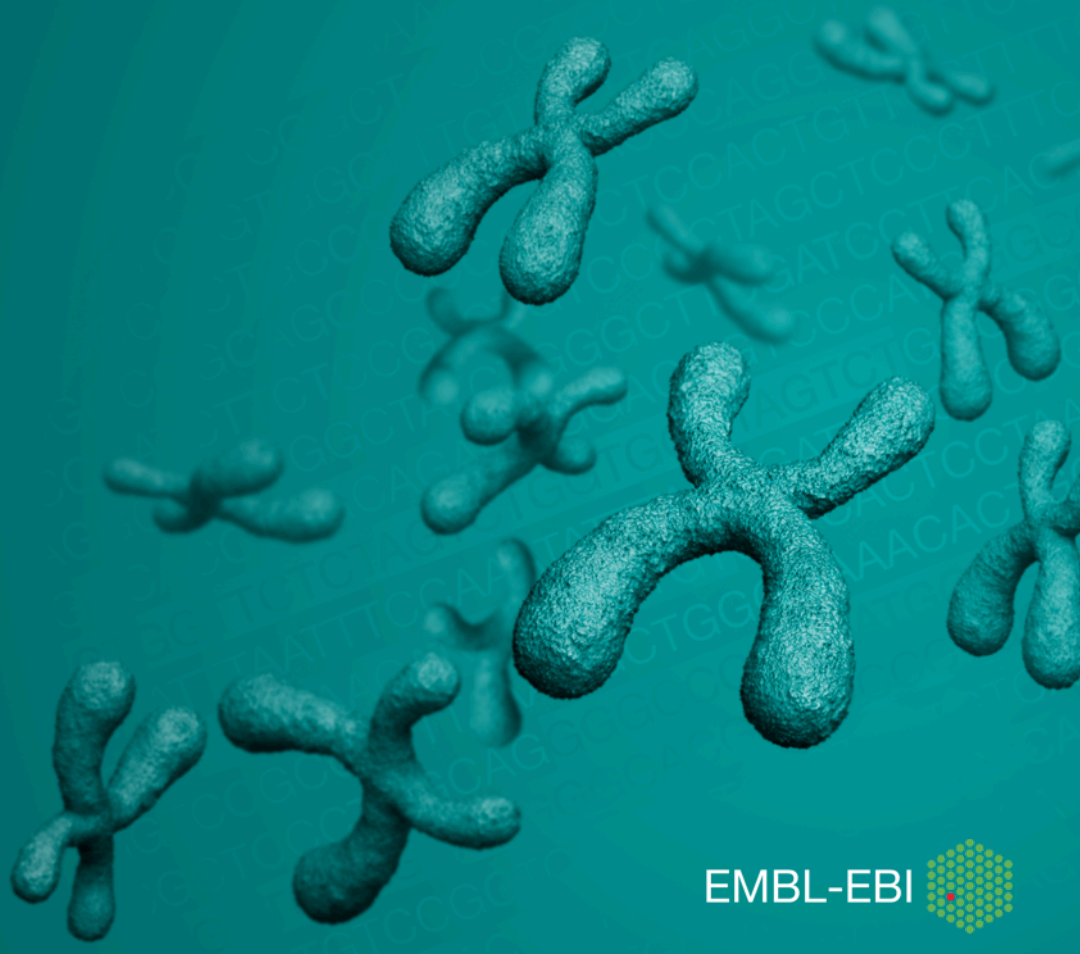
The European Variation Archive at EMBL-EBI: a home for plant variation data

Gary Saunders, PhD

www.ebi.ac.uk/eva

eva-helpdesk@ebi.ac.uk



EMBL-EBI

Genes, genomes

European Nucleotide Archive
1000 Genomes
Ensembl
Ensembl Genomes
European Genome-phenome Archive
Metagenomics portal

Gene, protein & metabolite expression

ArrayExpress
Expression Atlas
Metabolights
PRIDE

Literature & ontologies

Europe PubMed Central
Gene Ontology
Experimental Factor Ontology

Protein sequences, families & motifs

InterPro
Pfam
UniProt

Molecular structures

Protein Data Bank in Europe
Electron Microscopy Data Bank

Chemical biology

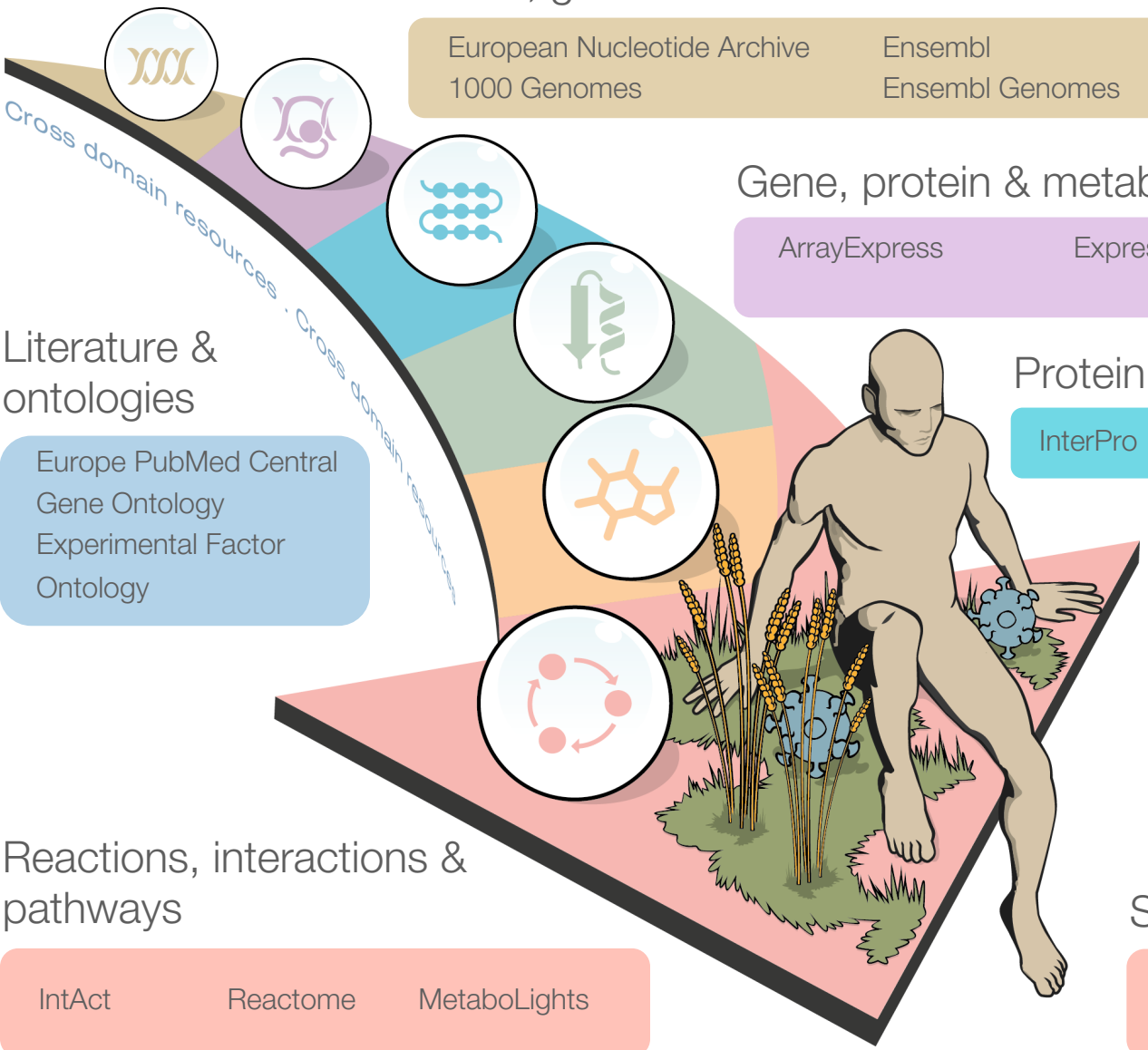
ChEMBL
ChEBI

Reactions, interactions & pathways

IntAct
Reactome
MetaboLights

Systems

BioModels
BioSamples
Enzyme Portal



EMBL-EBI

Genes, genomes & variation

European Nucleotide Archive
1000 Genomes
Ensembl
Ensembl Genomes
European Genome-phenome Archive
Metagenomics portal
European Variation Archive

Gene, protein & metabolite expression

ArrayExpress
Expression Atlas
Metabolights
PRIDE

Literature & ontologies

Europe PubMed Central
Gene Ontology
Experimental Factor Ontology

Protein sequences, families & motifs

InterPro
Pfam
UniProt

Molecular structures

Protein Data Bank in Europe
Electron Microscopy Data Bank

Chemical biology

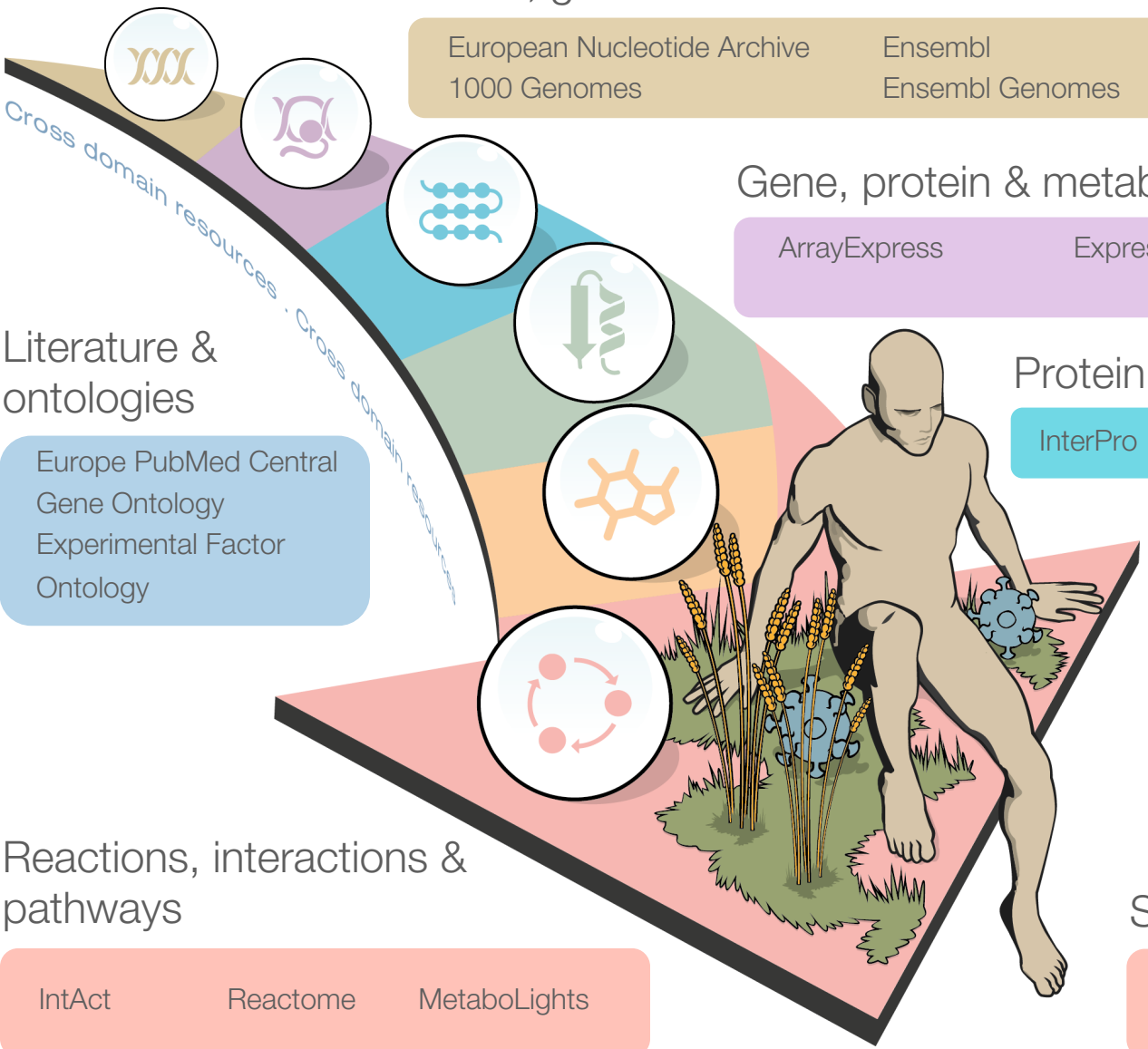
ChEMBL
ChEBI

Reactions, interactions & pathways

IntAct
Reactome
MetaboLights

Systems

BioModels
BioSamples
Enzyme Portal



EMBL-EBI

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)

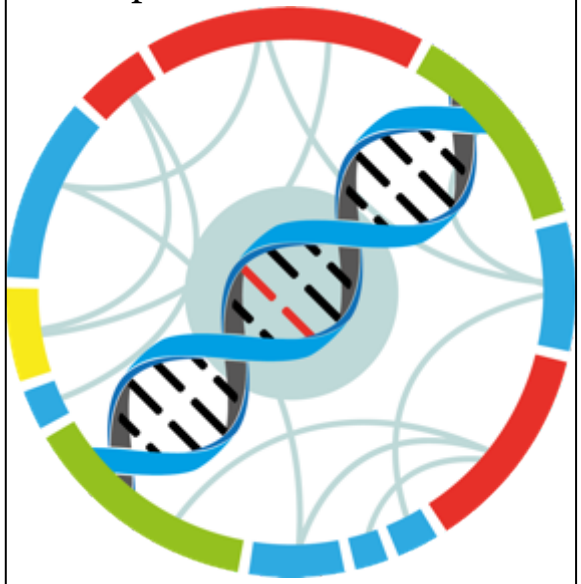
- **Why archive VCF files?**
- Administer file accessions
  - papers
  - share between researchers / labs
  - stable
- Administer variant accessions
  - TransPlant accessions
  - cluster variants based on position
  - stable

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

European Variation Archive

Accept and Validate Submissions

EVA
European Variation Archive



Provide direct data access

Support dataflow from and to other EBI resources

Data exchange with peer archives

(dbVar)

Provide stable accessions

European Variation Archive

Accept and Validate Submissions

Data mining

Visualization

Support dataflow from and to other EBI resources

Databases

EVA

European Variation Archive

Curation

Provide direct data access

Ontologies

Data exchange with peer archives

(dbVar)

Provide stable accessions

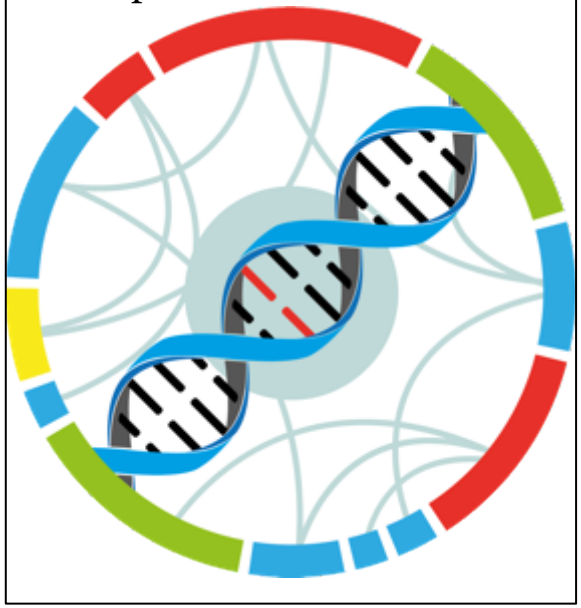
EVA Is A Collaborative Archive

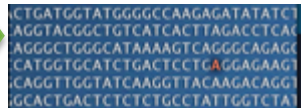
SEUVADIS CONSORTIUM

DGVarchive

Direct Submissions

EVA
European Variation Archive





Europe PubMed Central

LRG

ENA
European Nucleotide Archive

e!

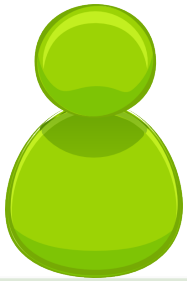
transPLANT

european genome-phenome archive

EMBL

Gramene

EVA Data Model



Submitter

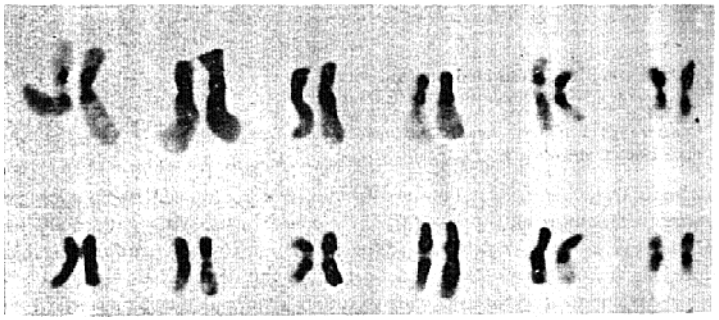
Analysis

EVA

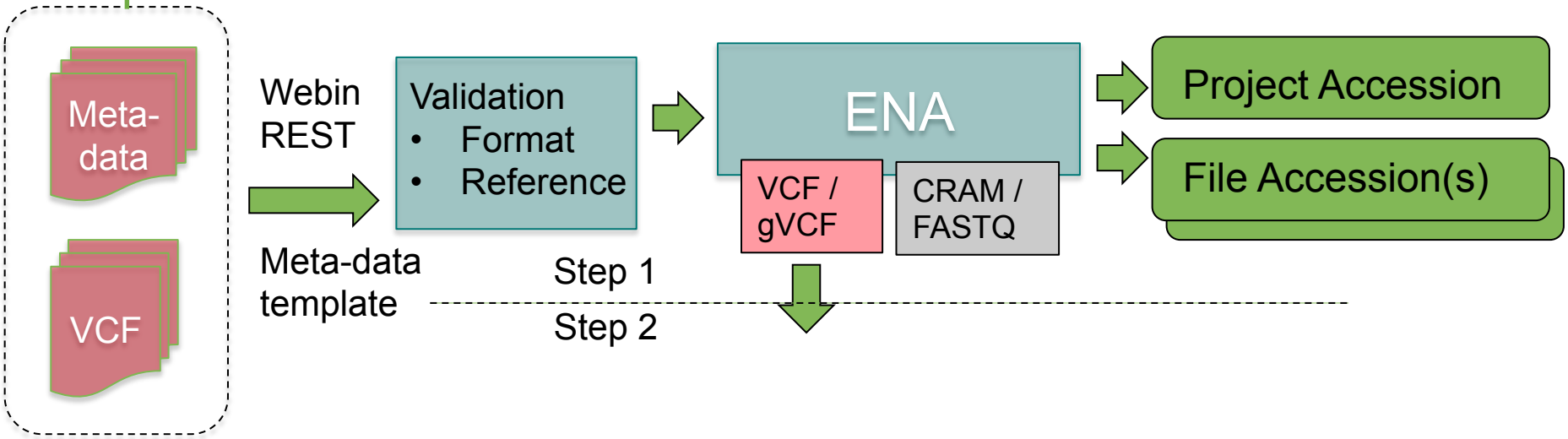
Sample(s)

Methodology

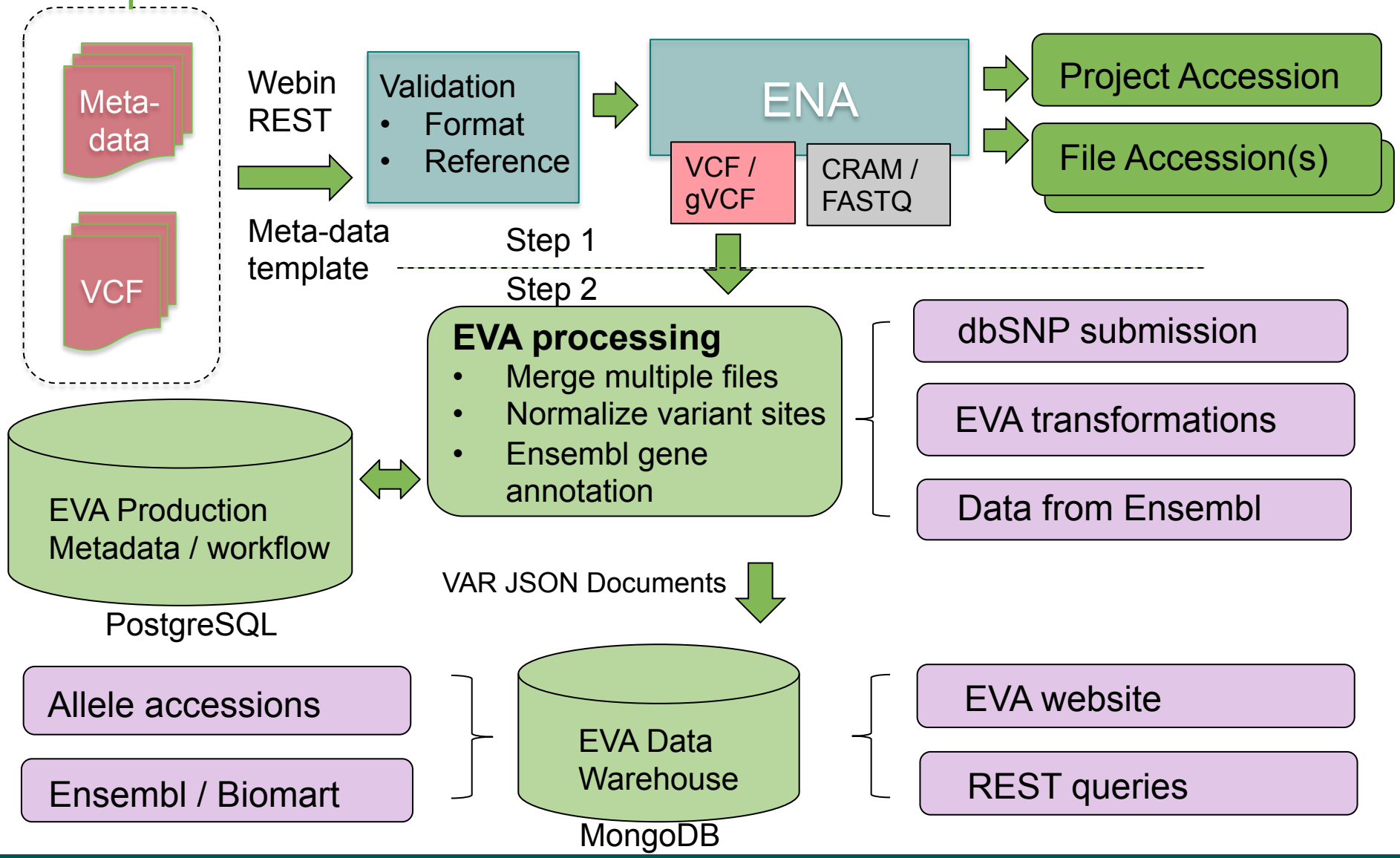
Genome



Pipeline Structure of EVA

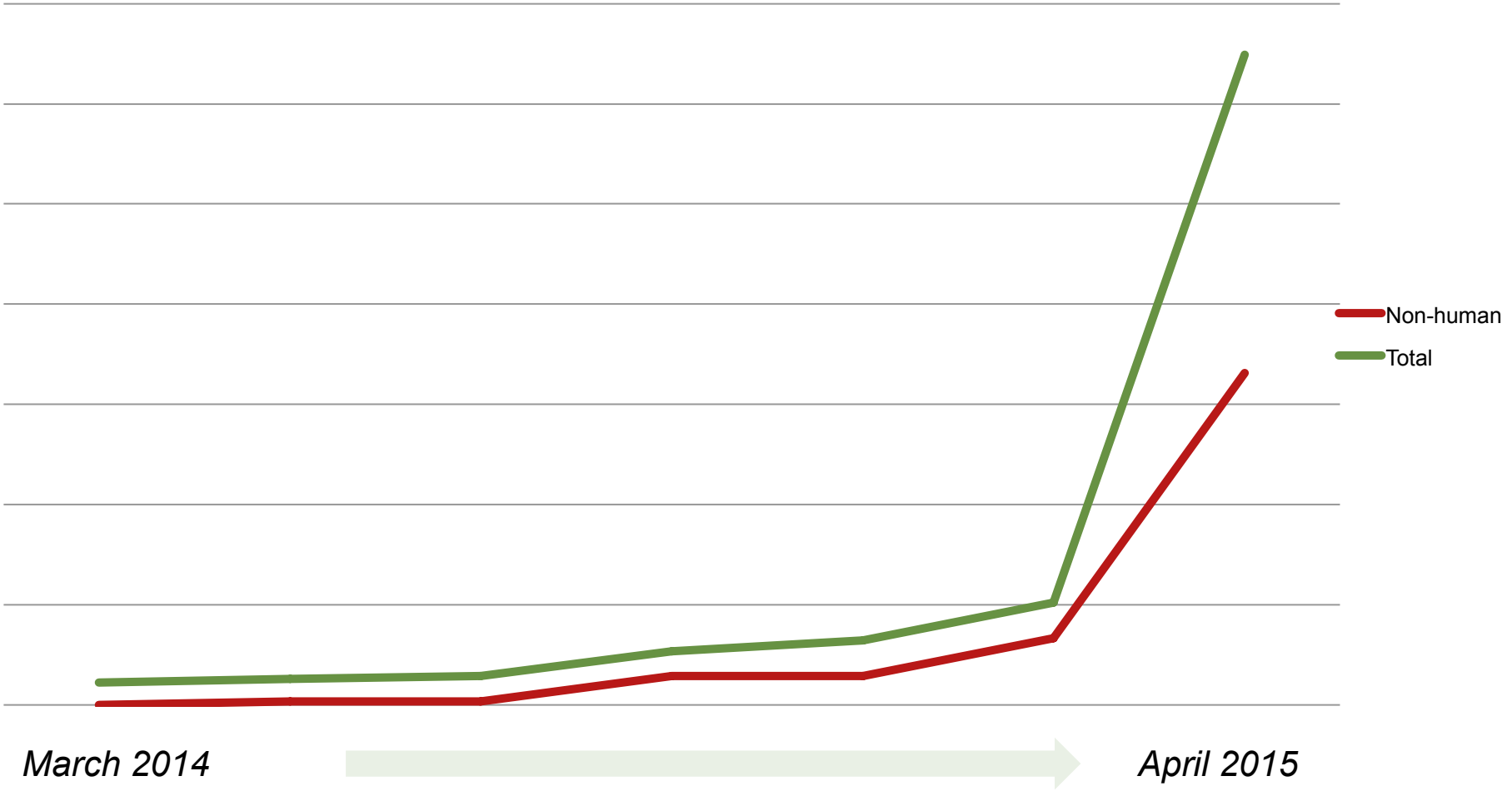


Pipeline Structure of EVA



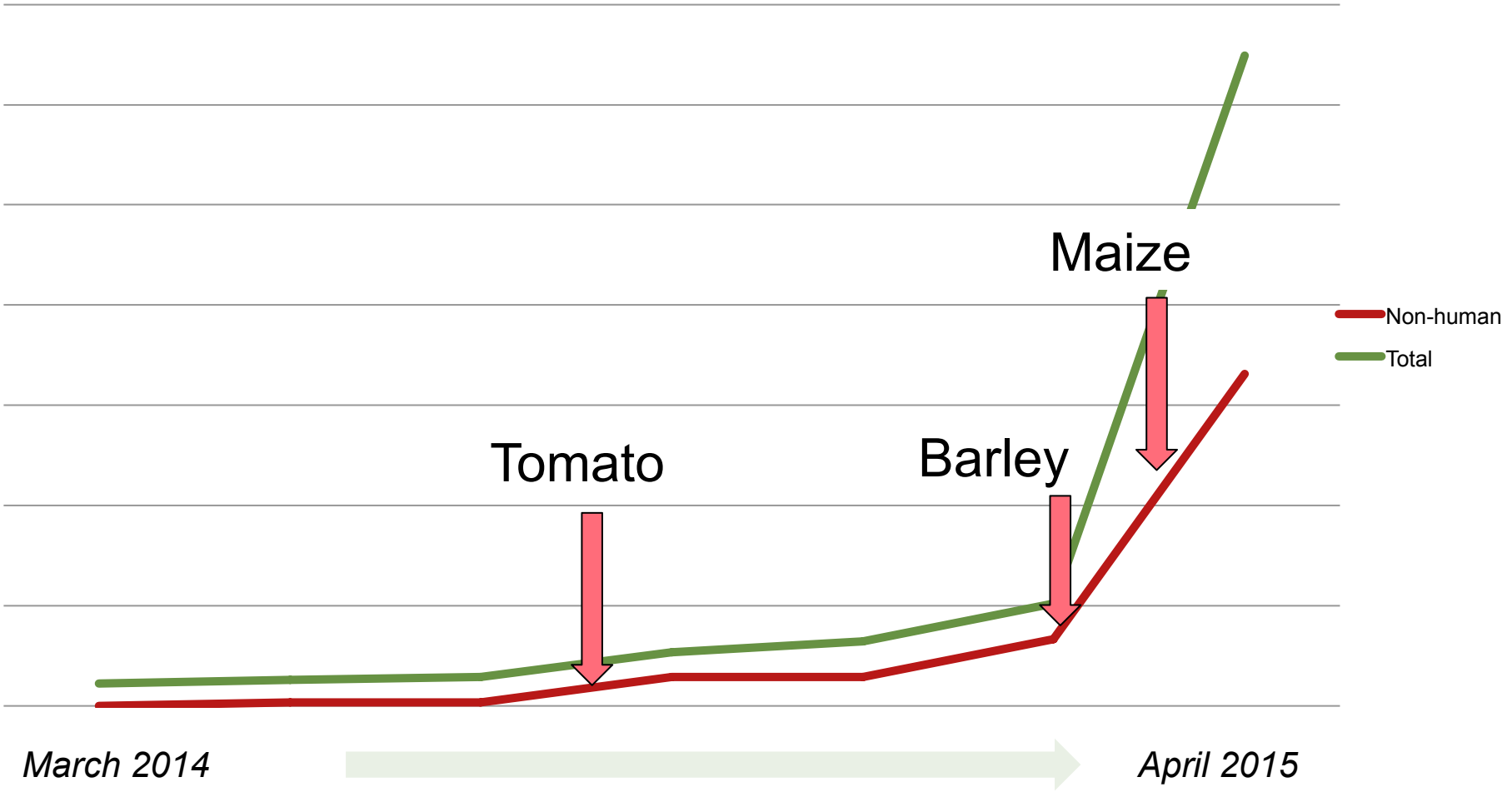
EVA Content

- Submitted variants



EVA Content

- Submitted variants



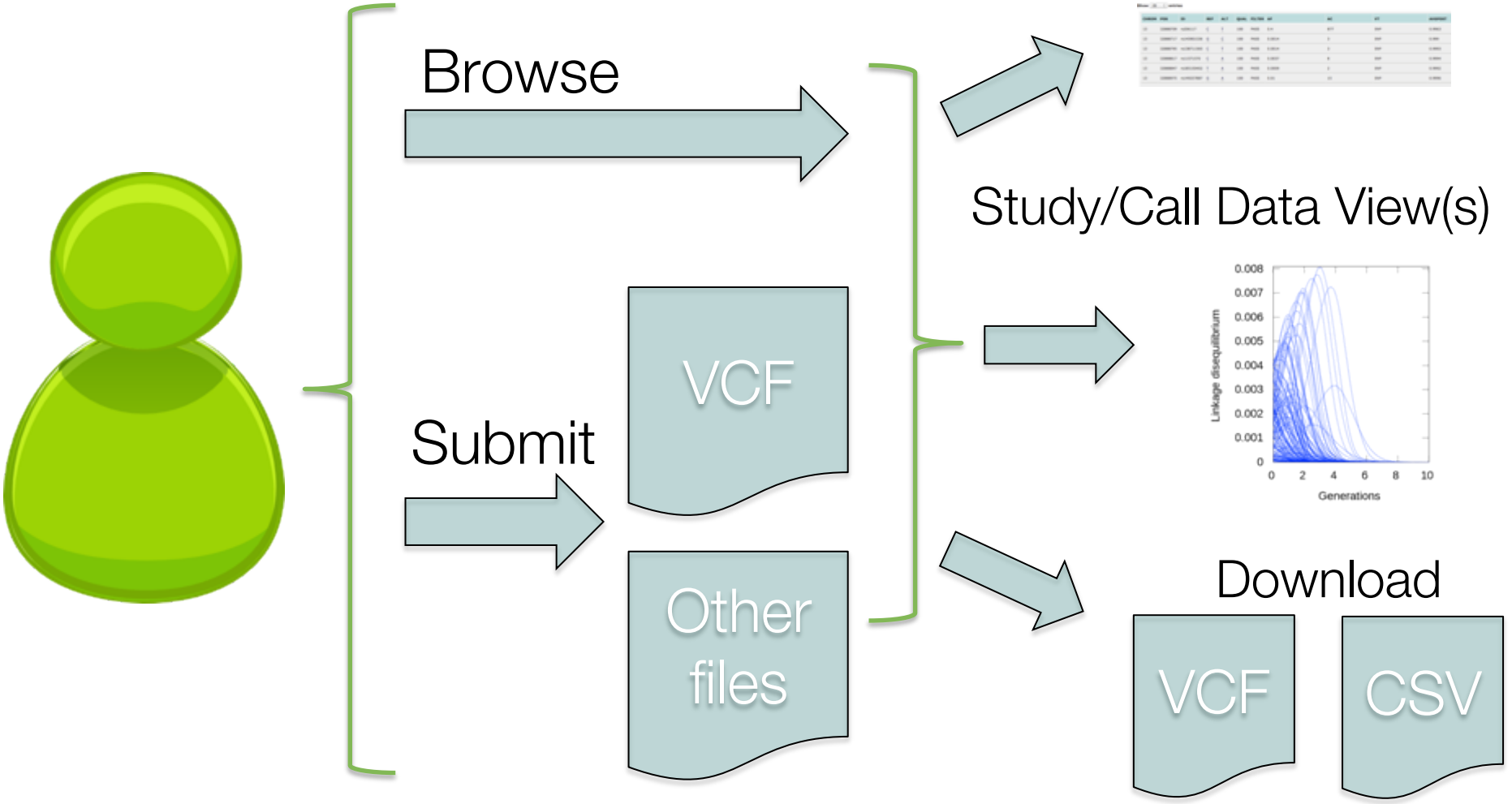
EVA Content

- 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

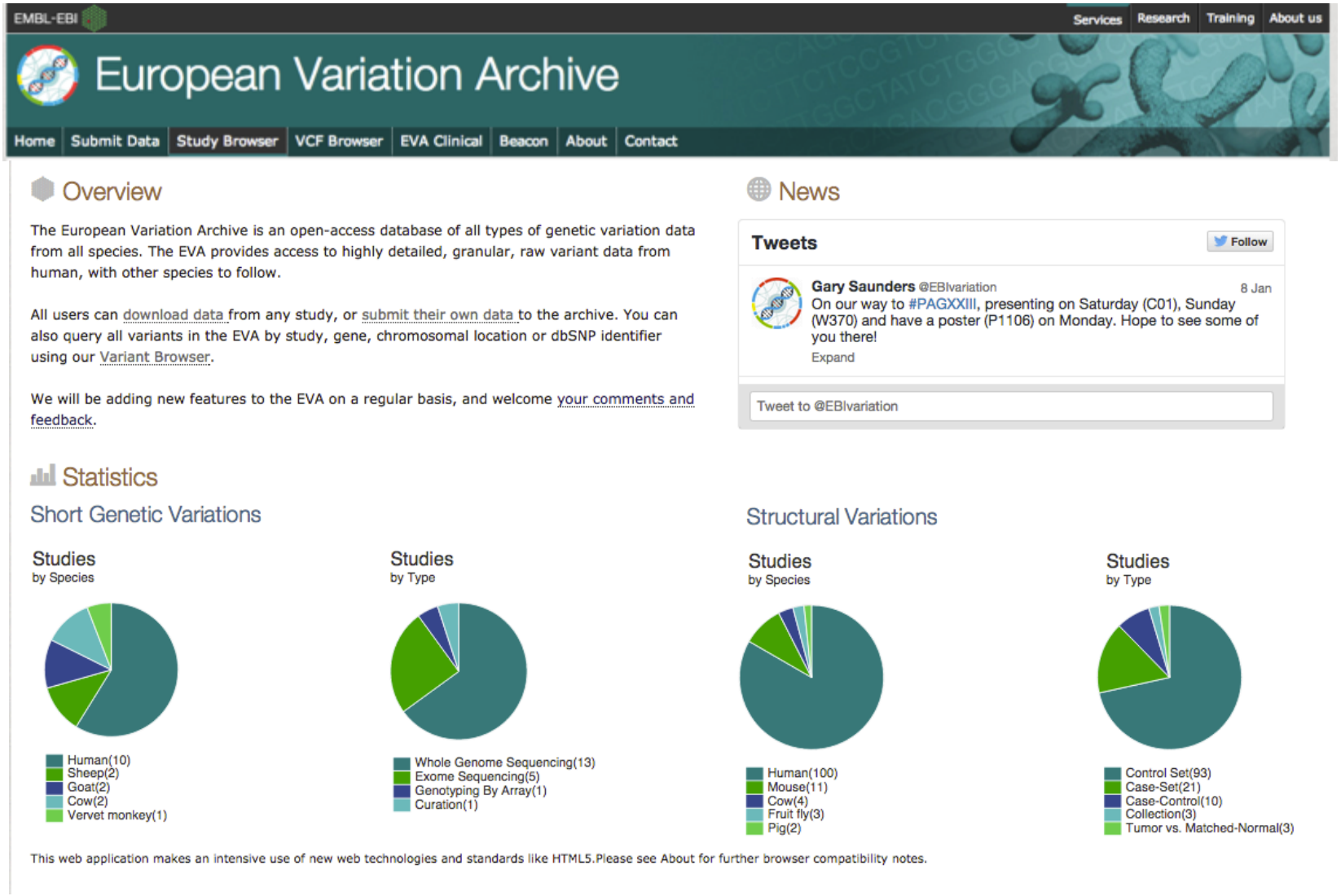
- Shrinking percentage of the dbSNP legacy database will not be loaded as individual studies but available via dbSNP legacy track only e.g. older (array-based) technologies, very small studies

User Experience

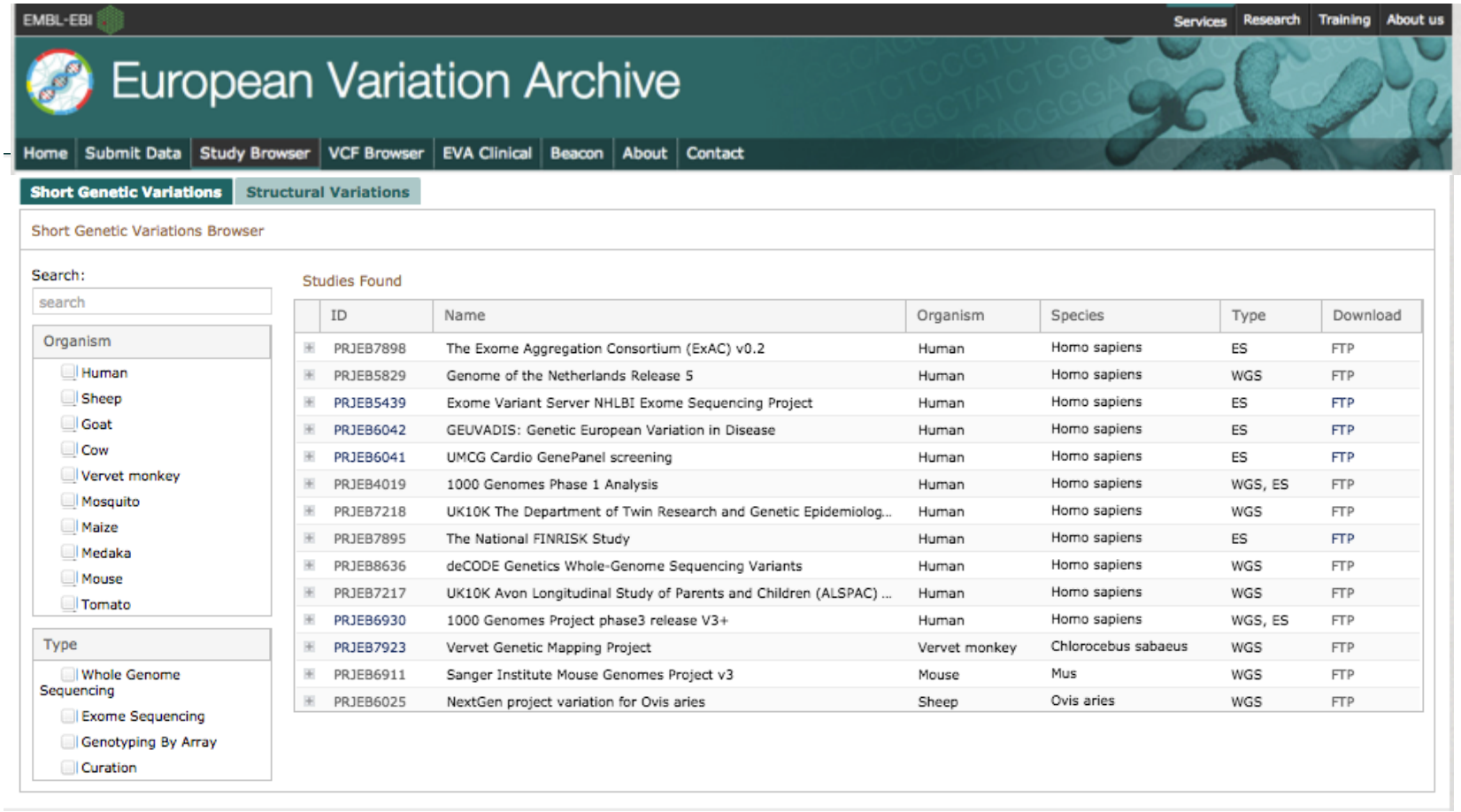


EVA Website

www.ebi.ac.uk/eva/ eva-helpdesk@ebi.ac.uk

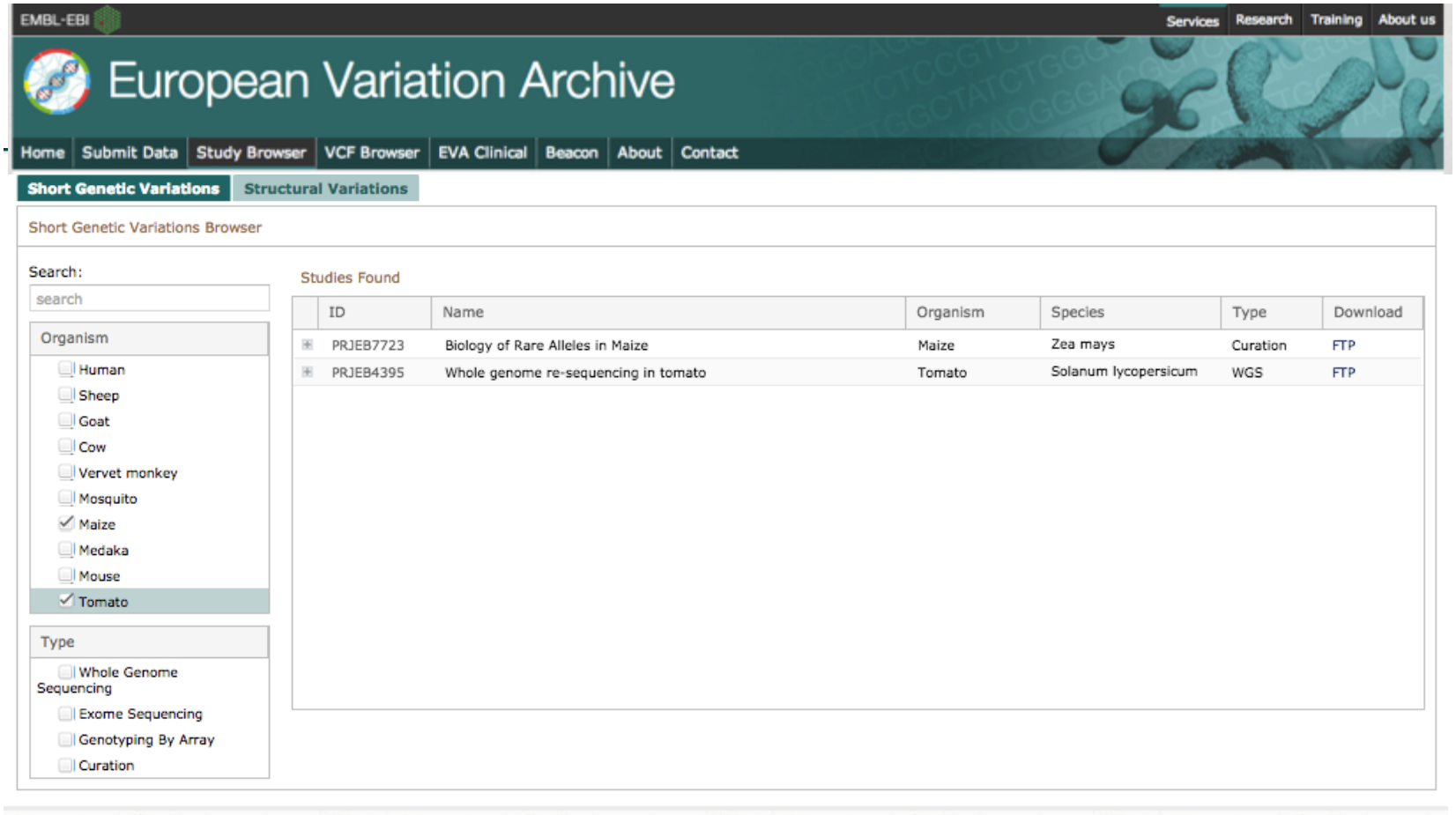


EVA Study Browser



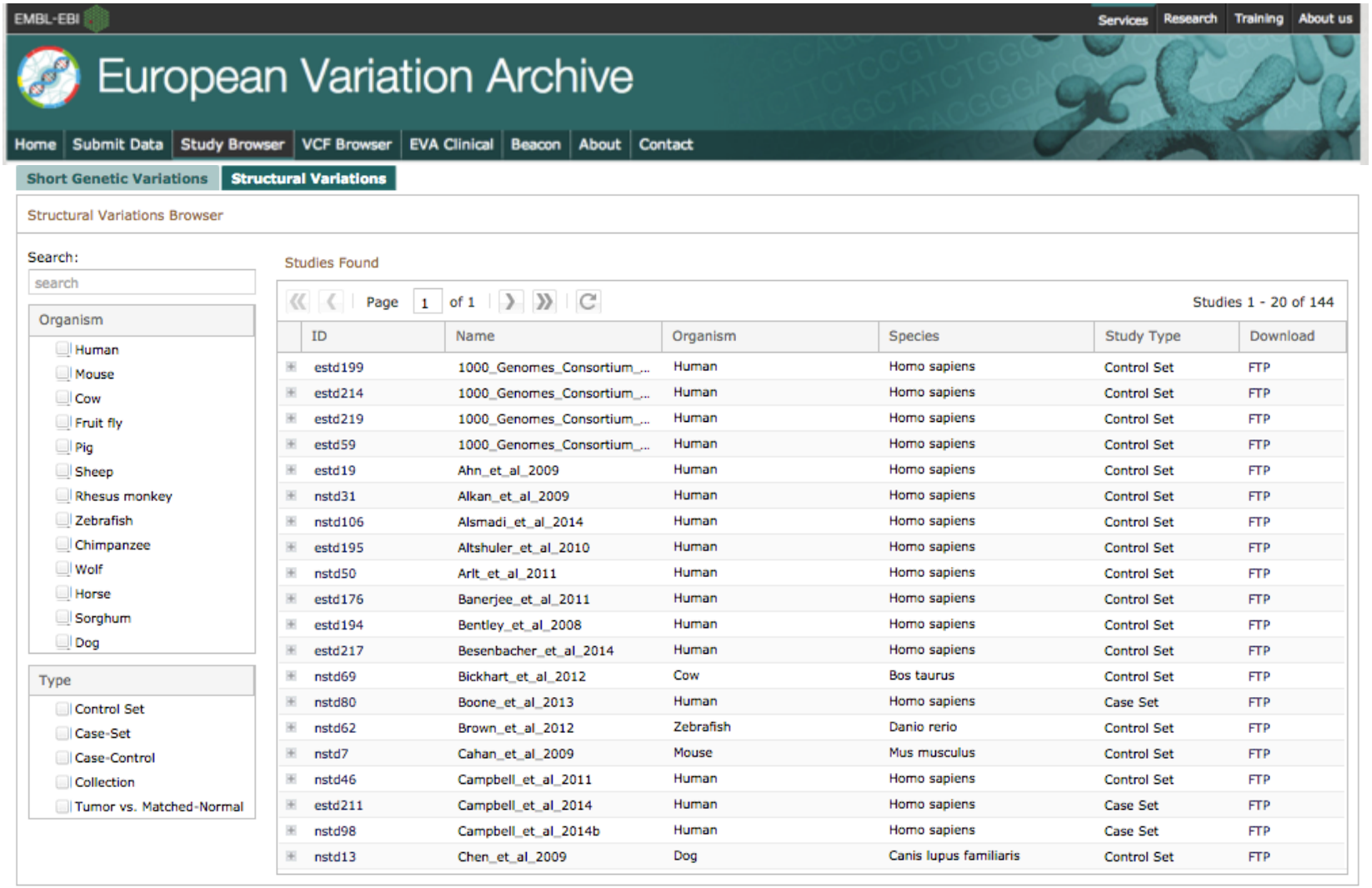
www.ebi.ac.uk/eva/ eva-helpdesk@ebi.ac.uk

EVA Study Browser

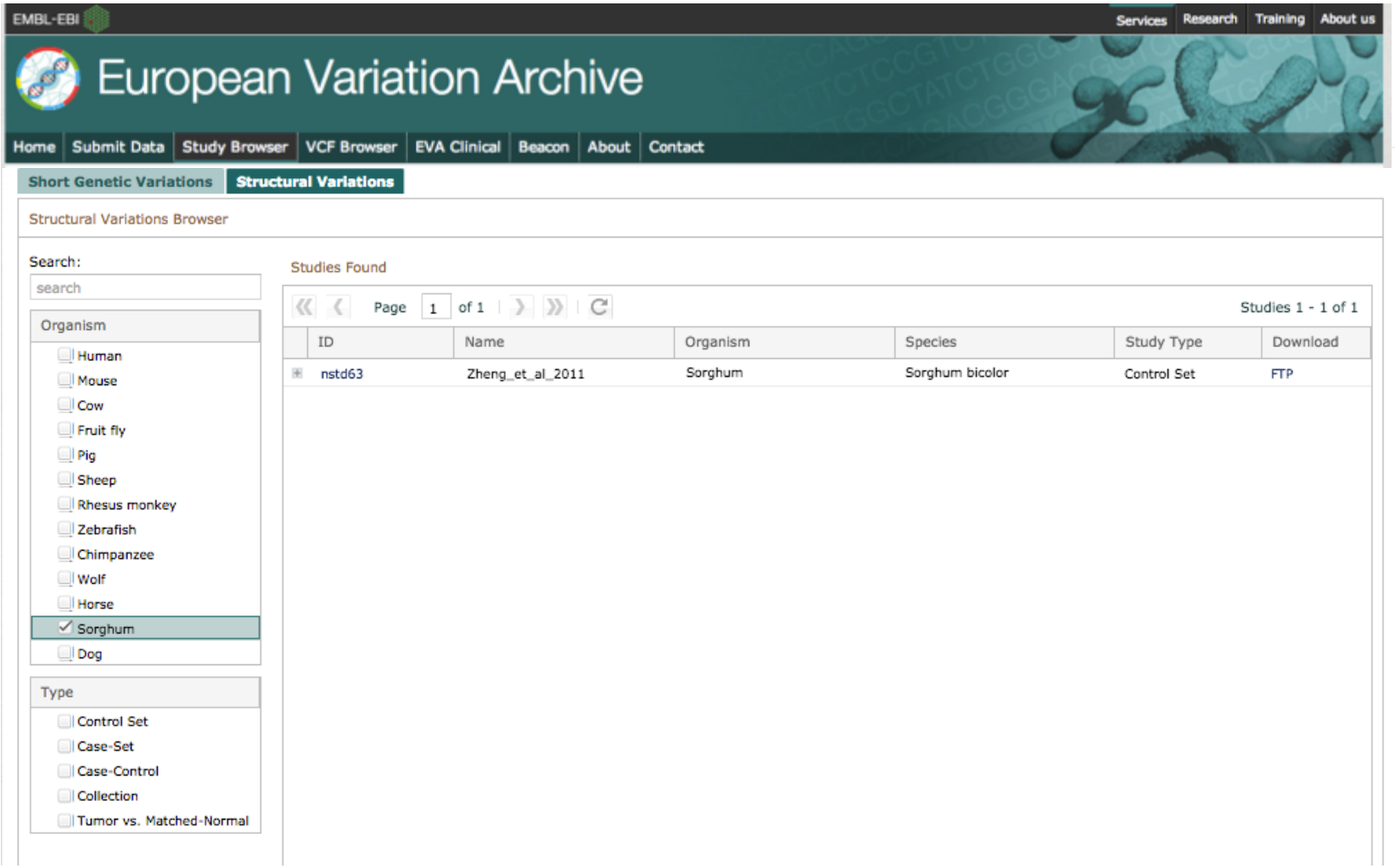


www.ebi.ac.uk/eva/ eva-helpdesk@ebi.ac.uk

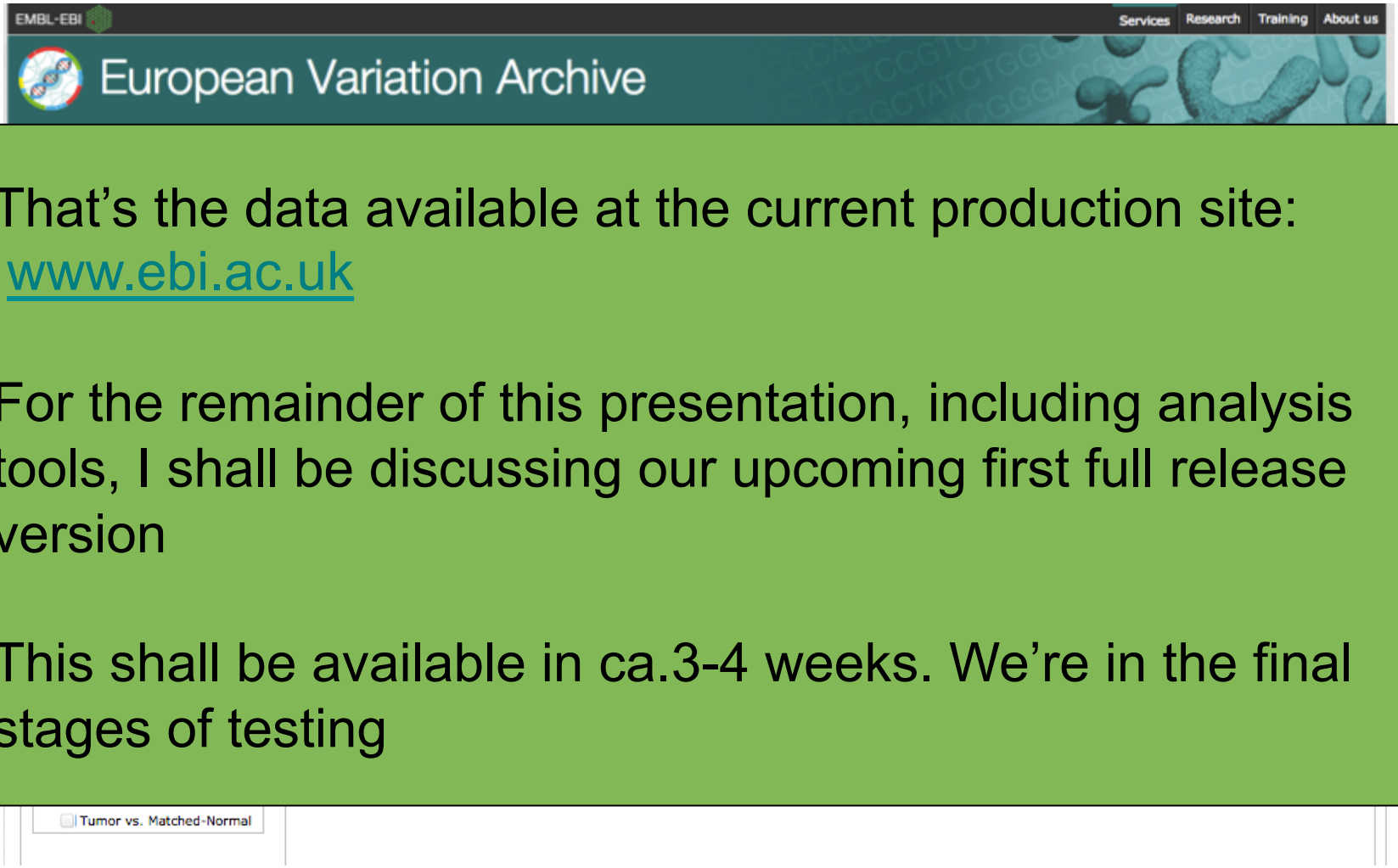
EVA Study Browser – Structural Variants



EVA Study Browser – Structural Variants



EVA 1.0

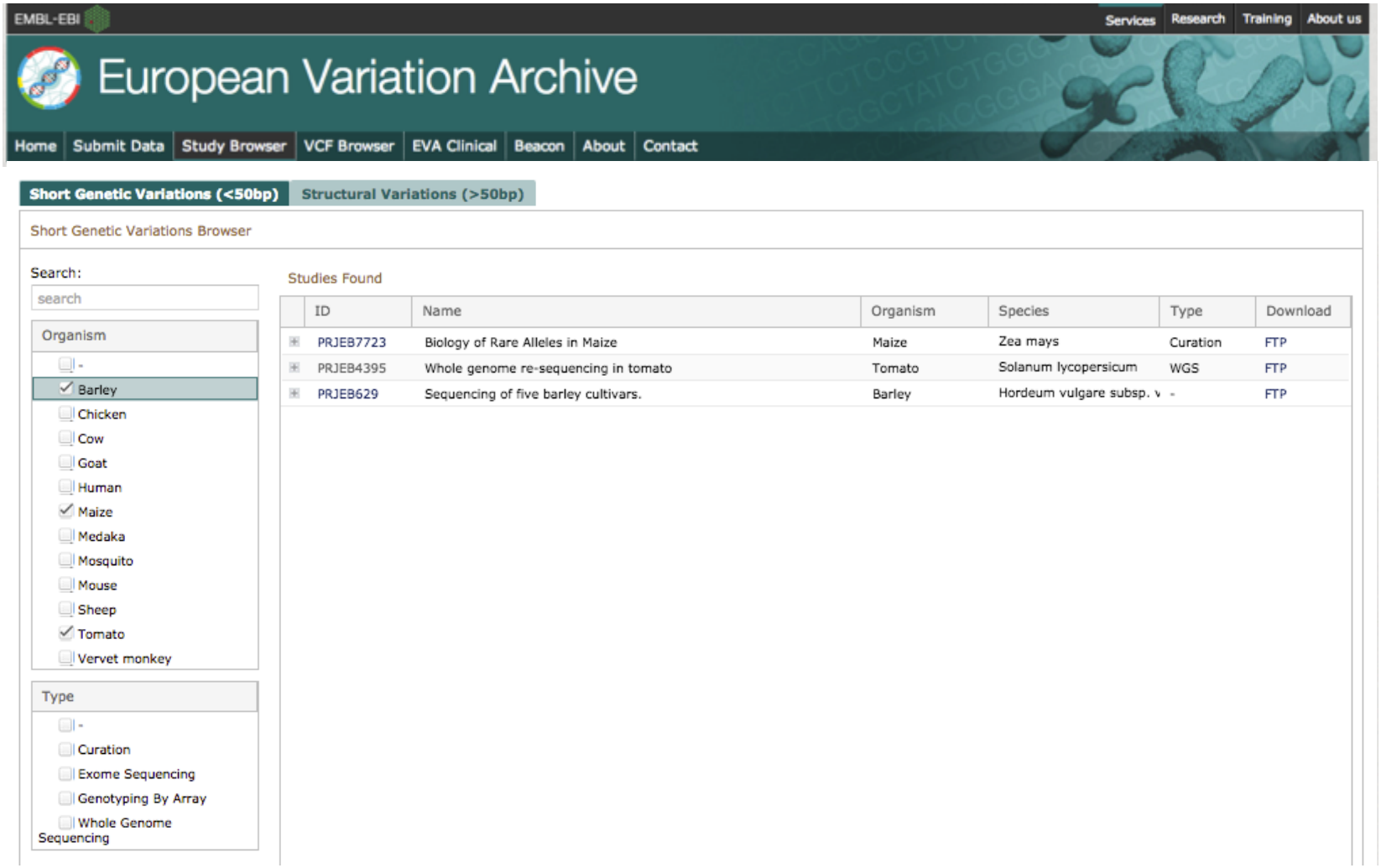


That's the data available at the current production site: www.ebi.ac.uk

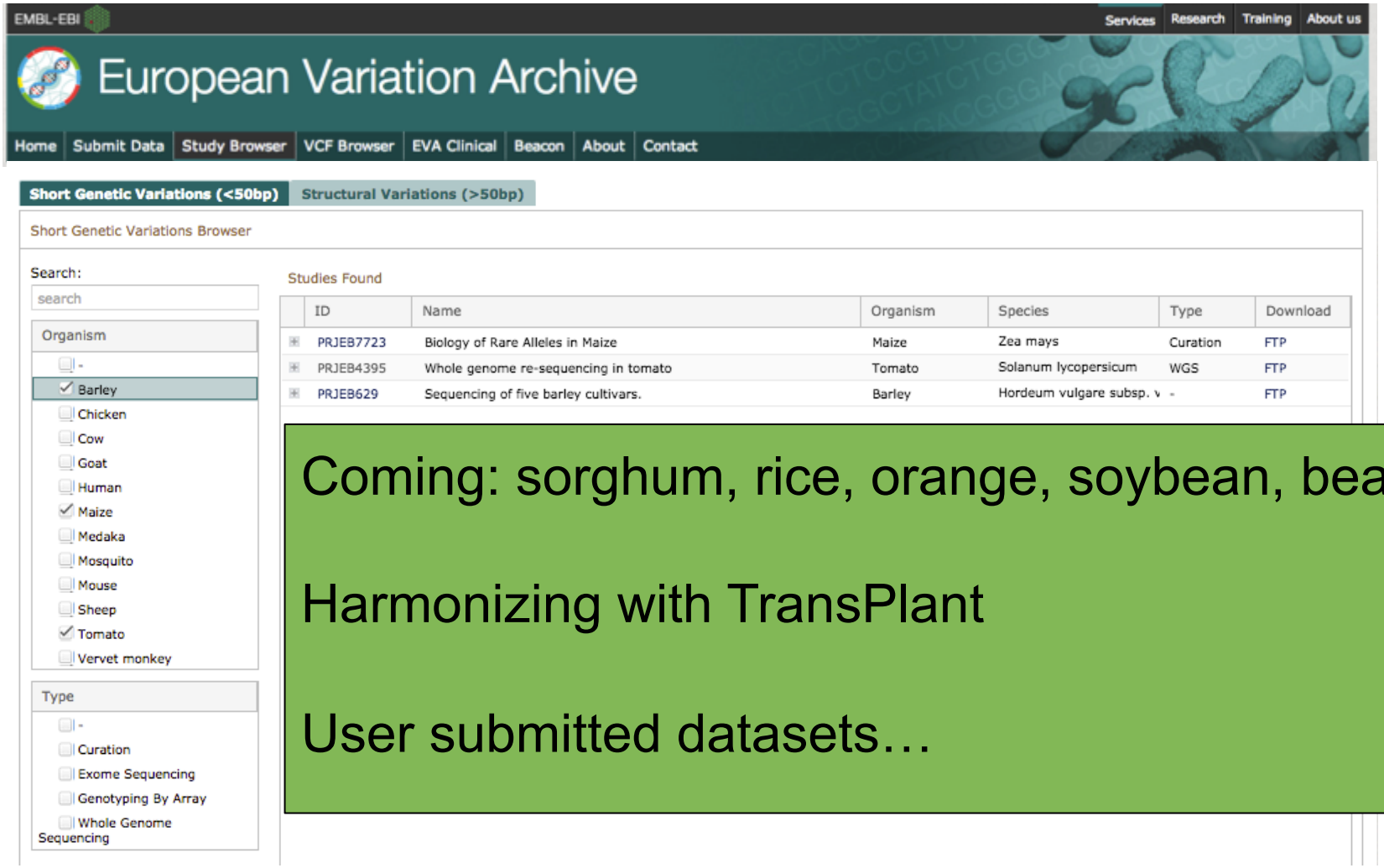
For the remainder of this presentation, including analysis tools, I shall be discussing our upcoming first full release version

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

EVA Study Browser – 1.0



EVA Study Browser – 1.0



Coming: sorghum, rice, orange, soybean, bean

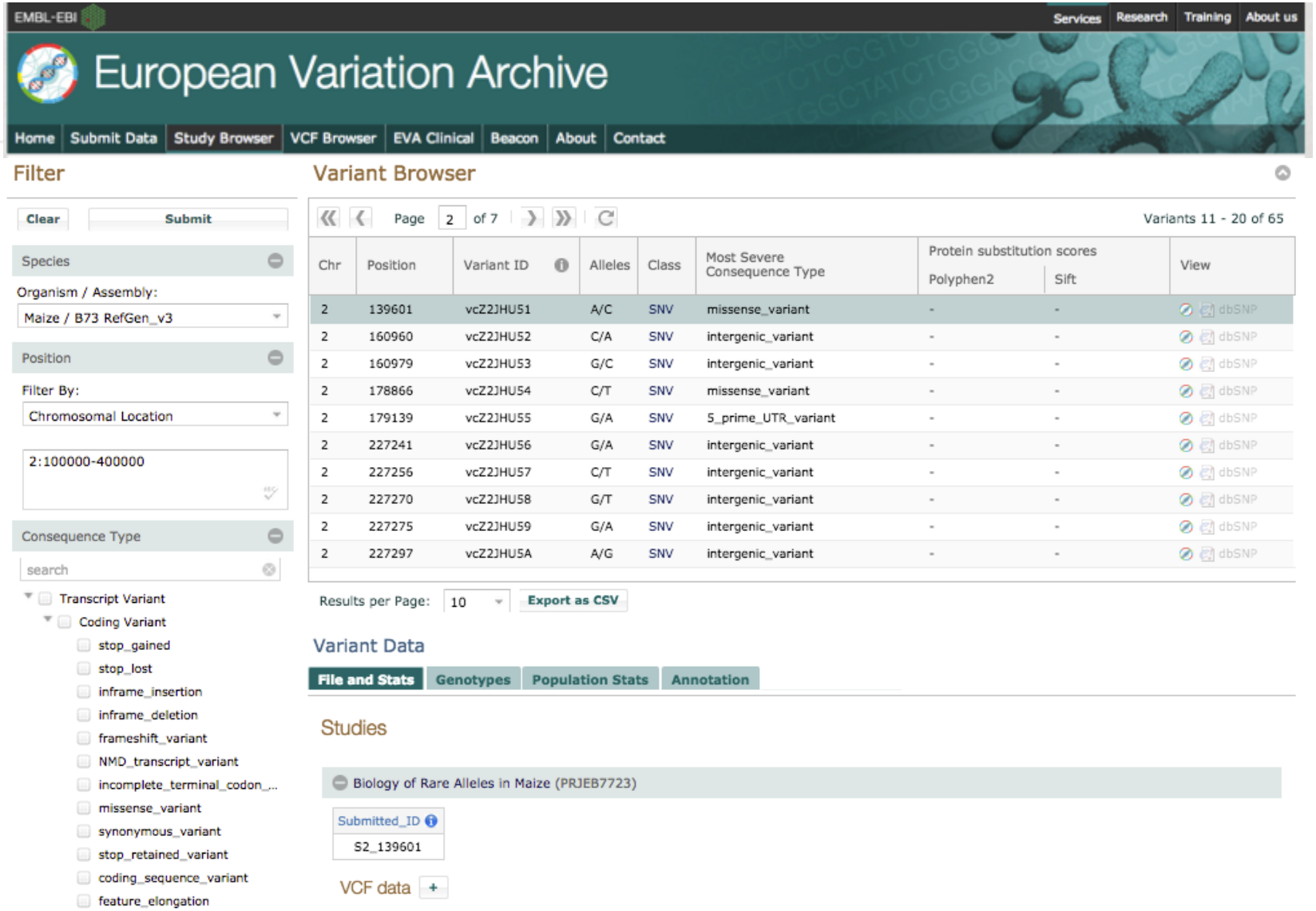
Harmonizing with TransPlant

User submitted datasets…

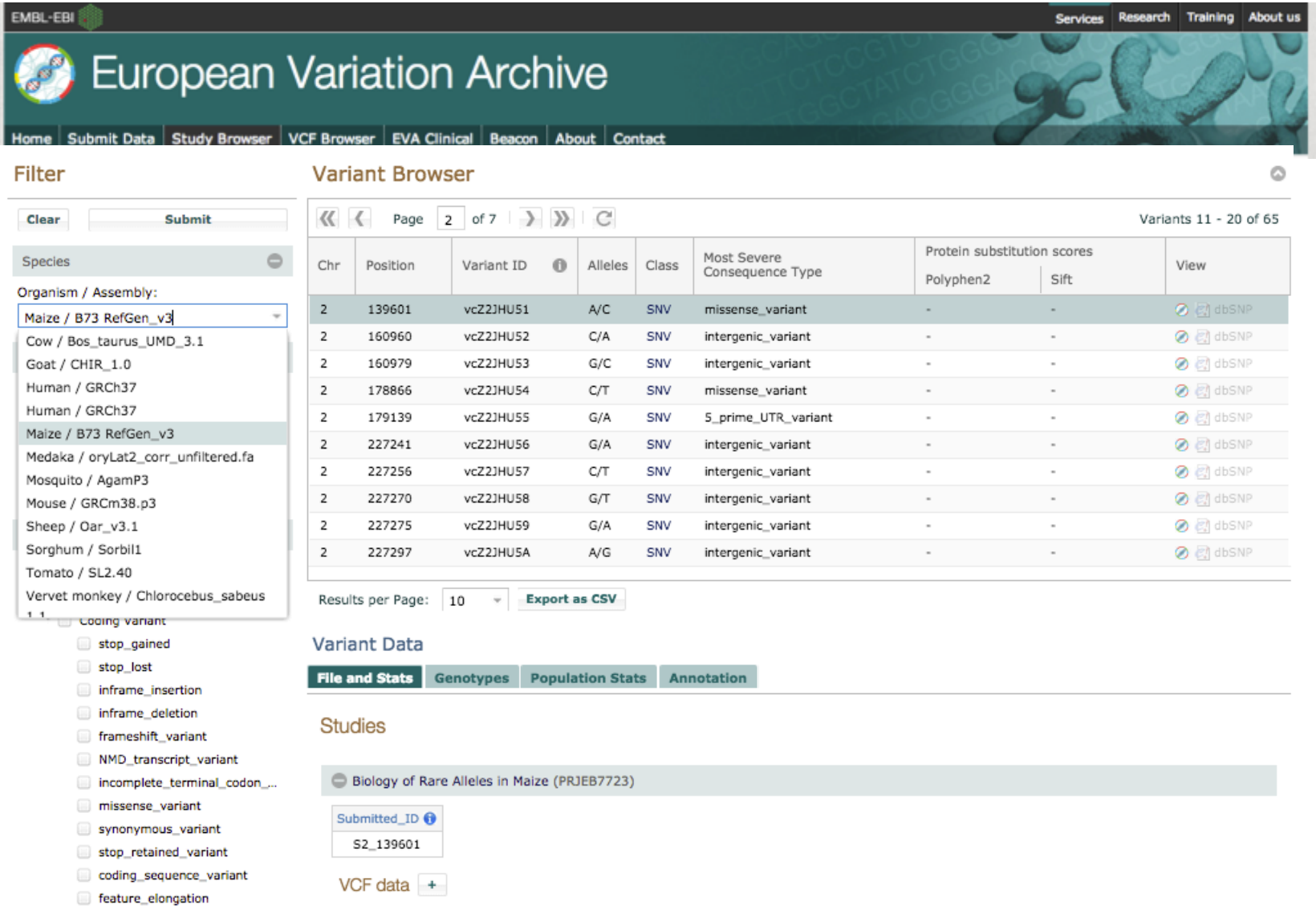
EVA VCF Browser

- 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

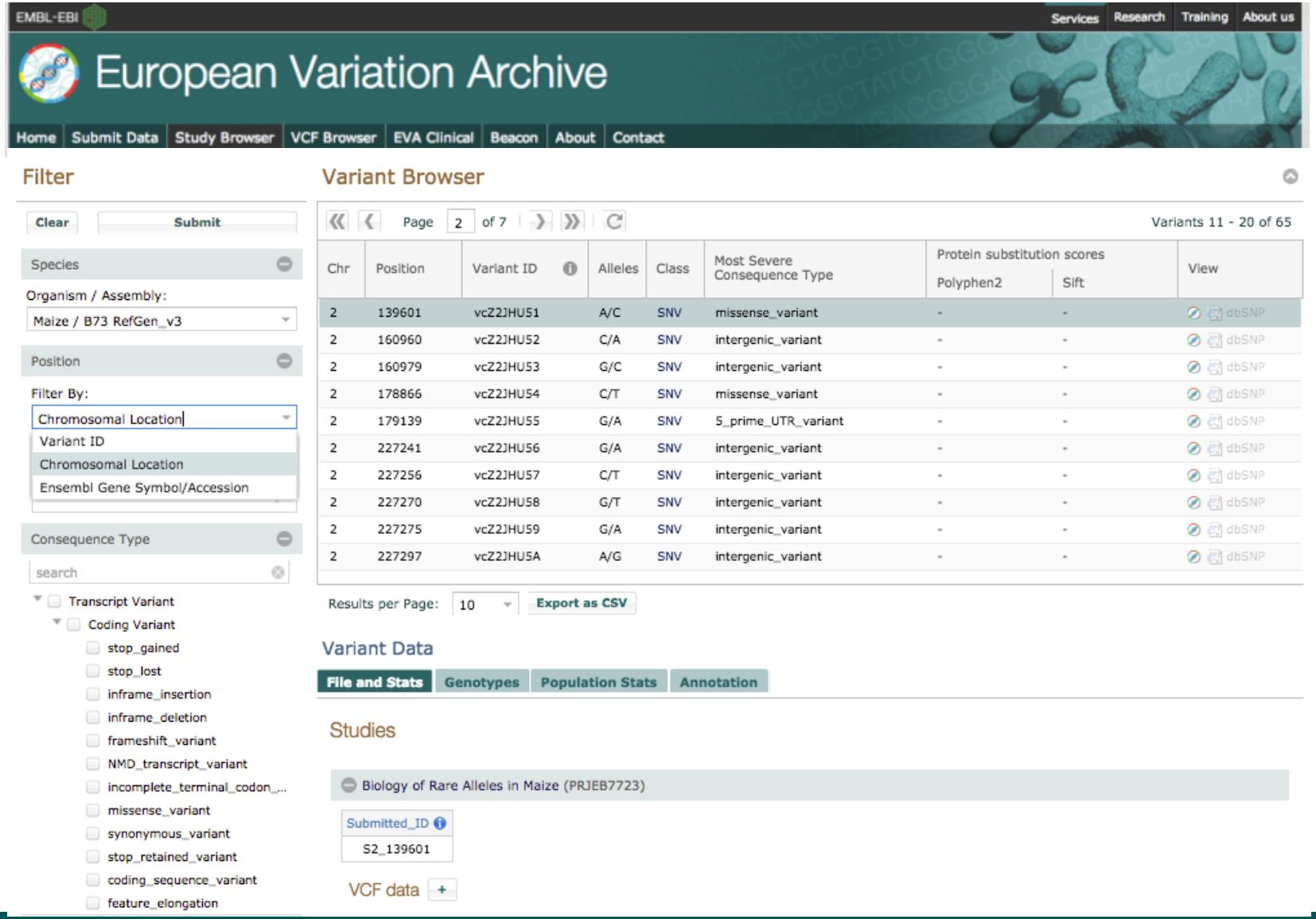
EVA VCF Browser – 1.0



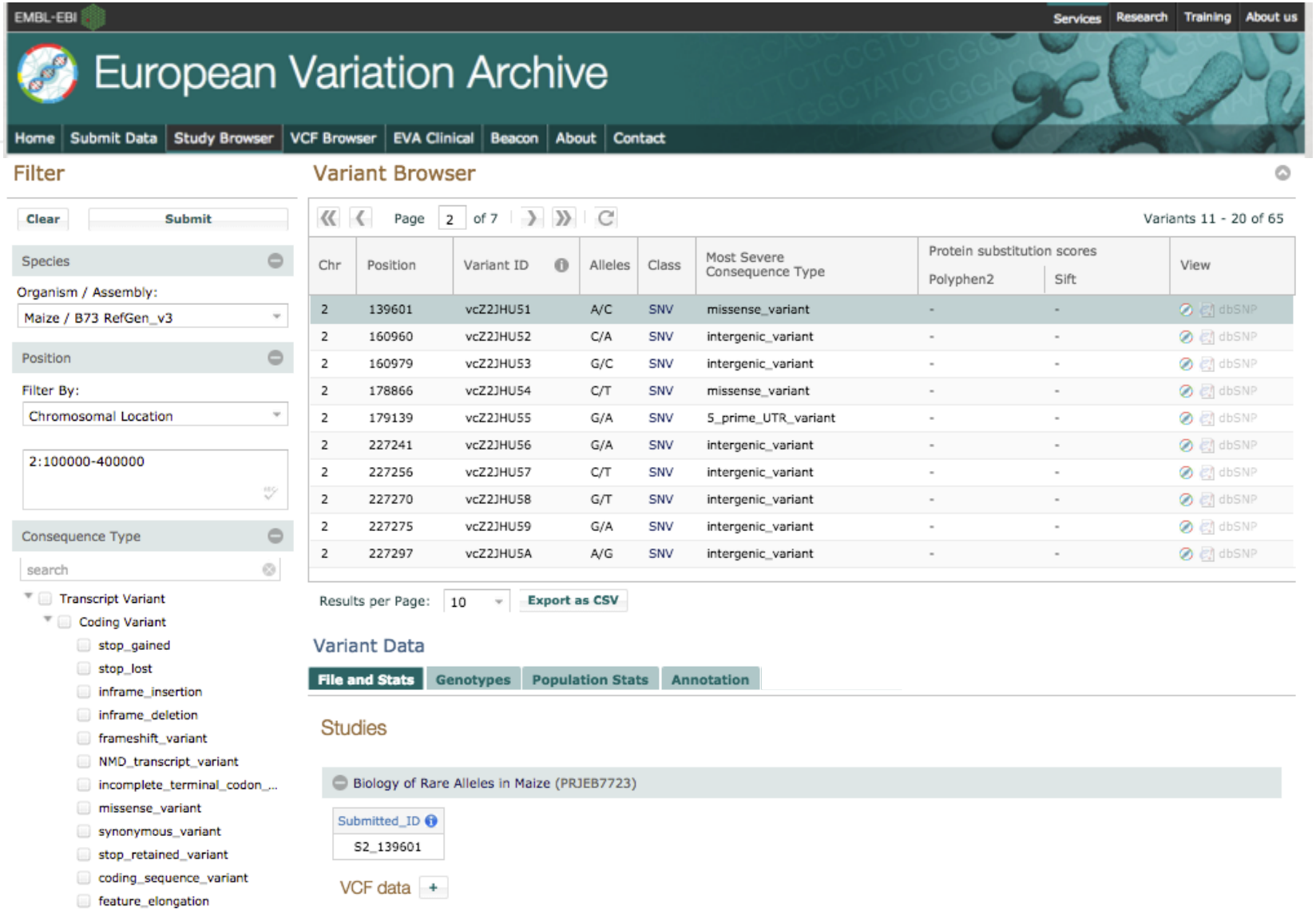
EVA VCF Browser – 1.0



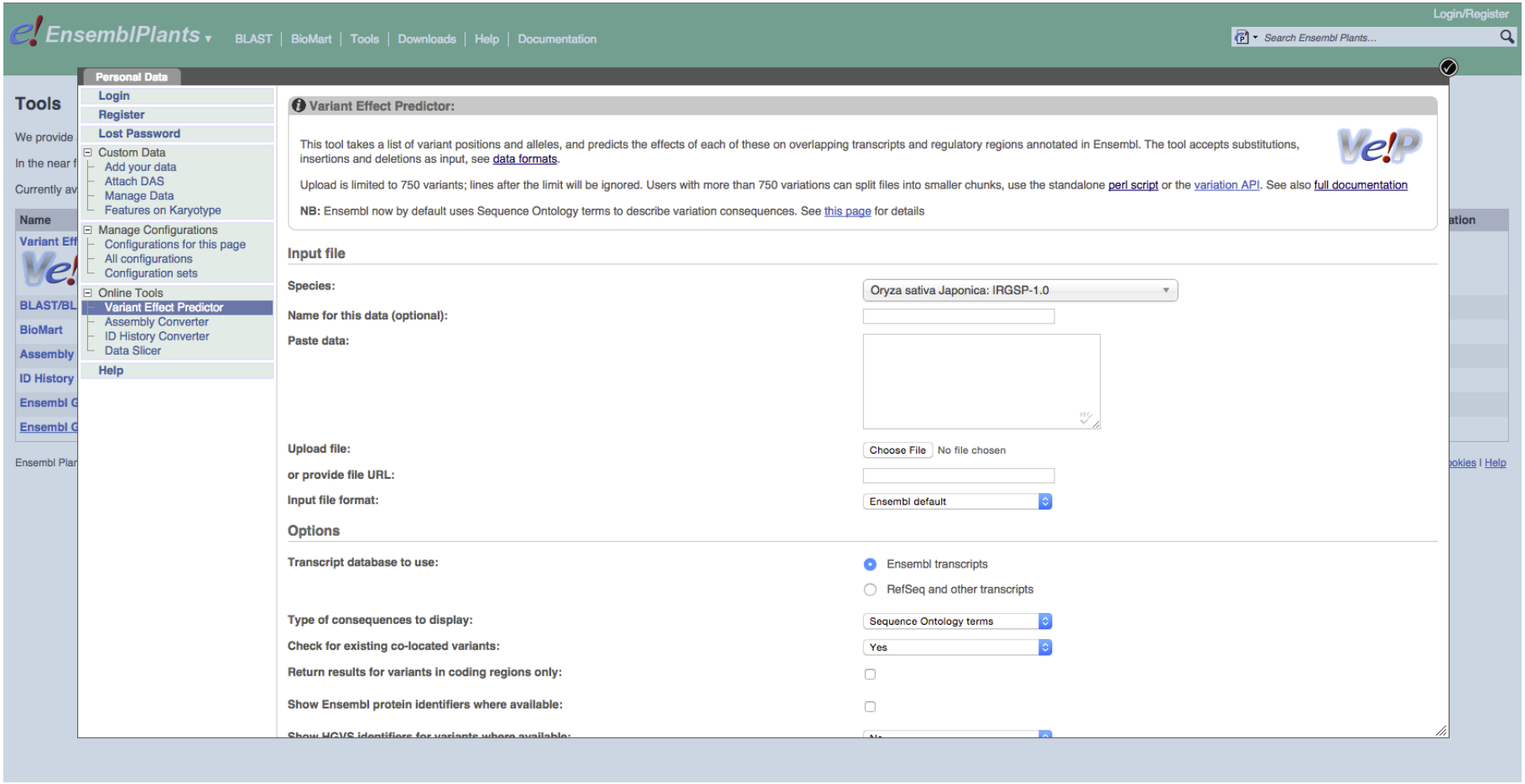
EVA VCF Browser – 1.0



EVA VCF Browser – 1.0

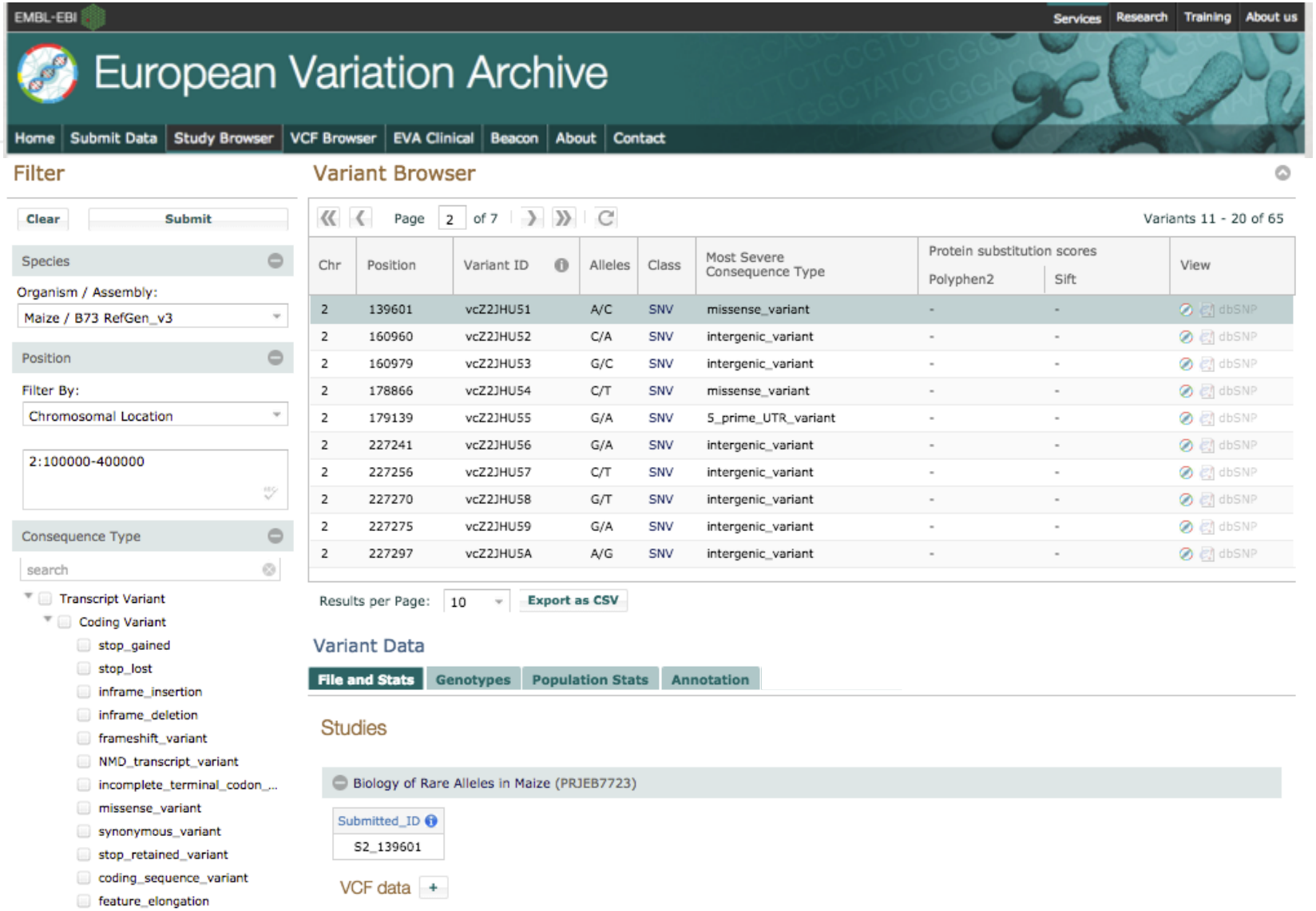


Ensembl Plants VEP

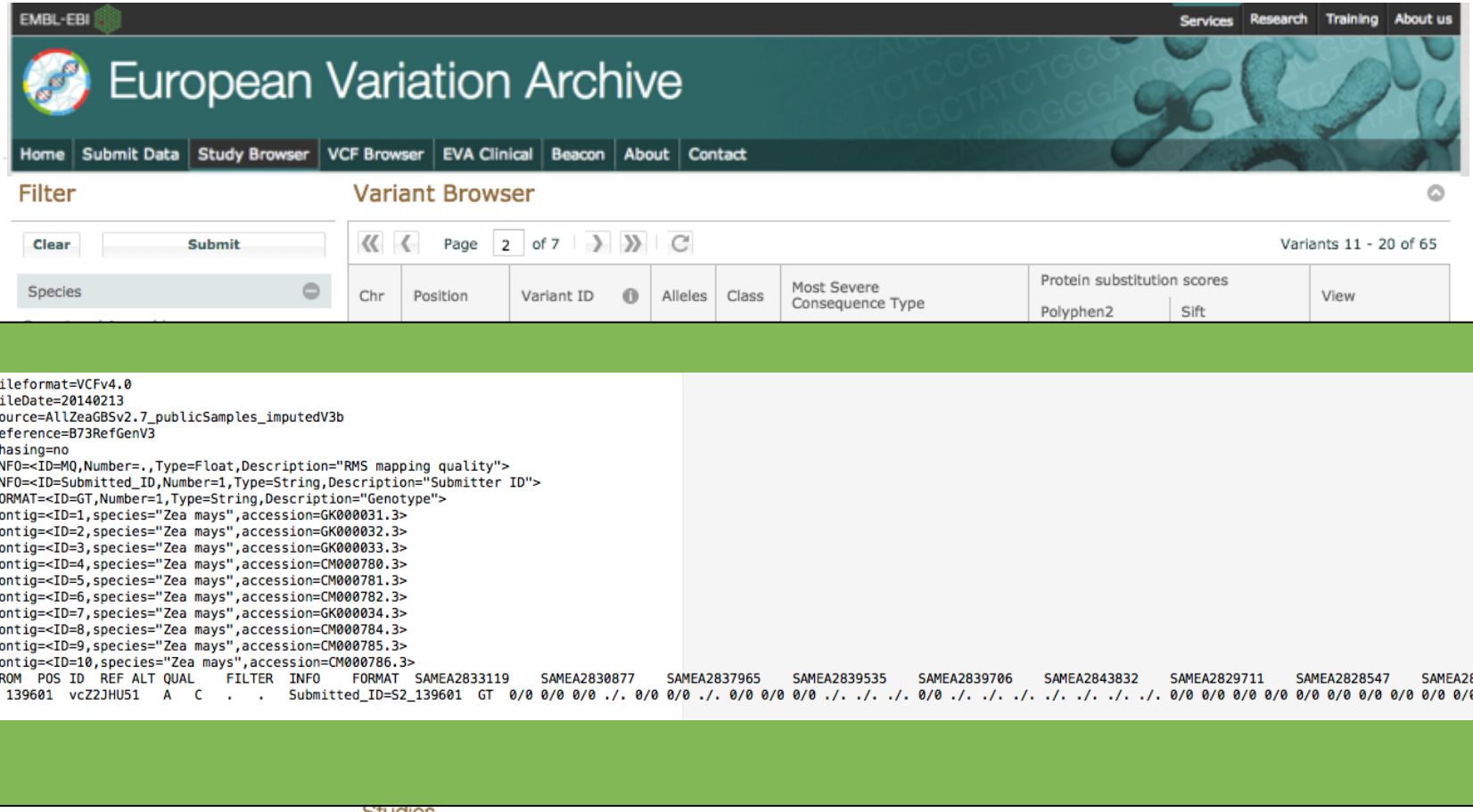


EMBL-EBI

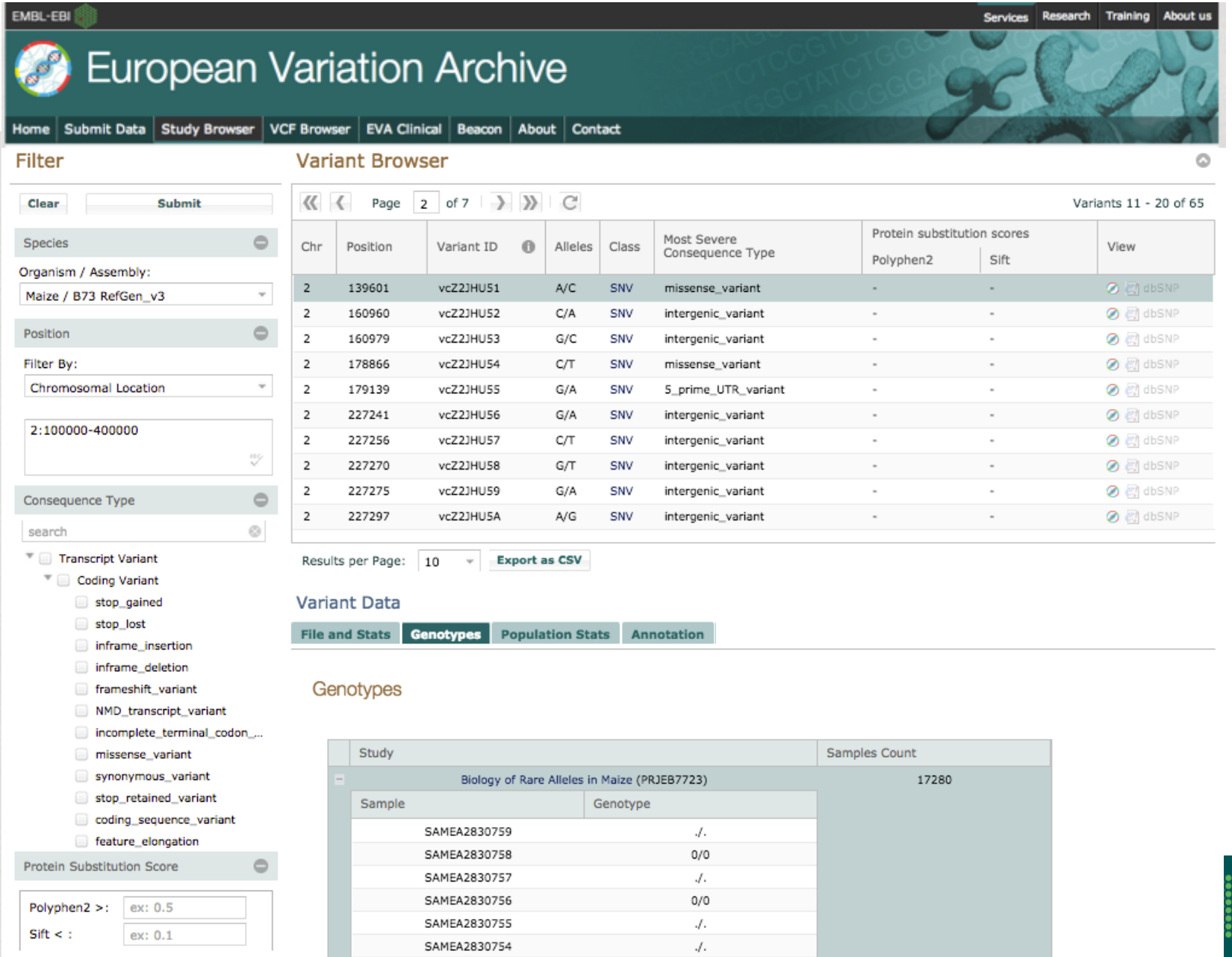
EVA VCF Browser – 1.0



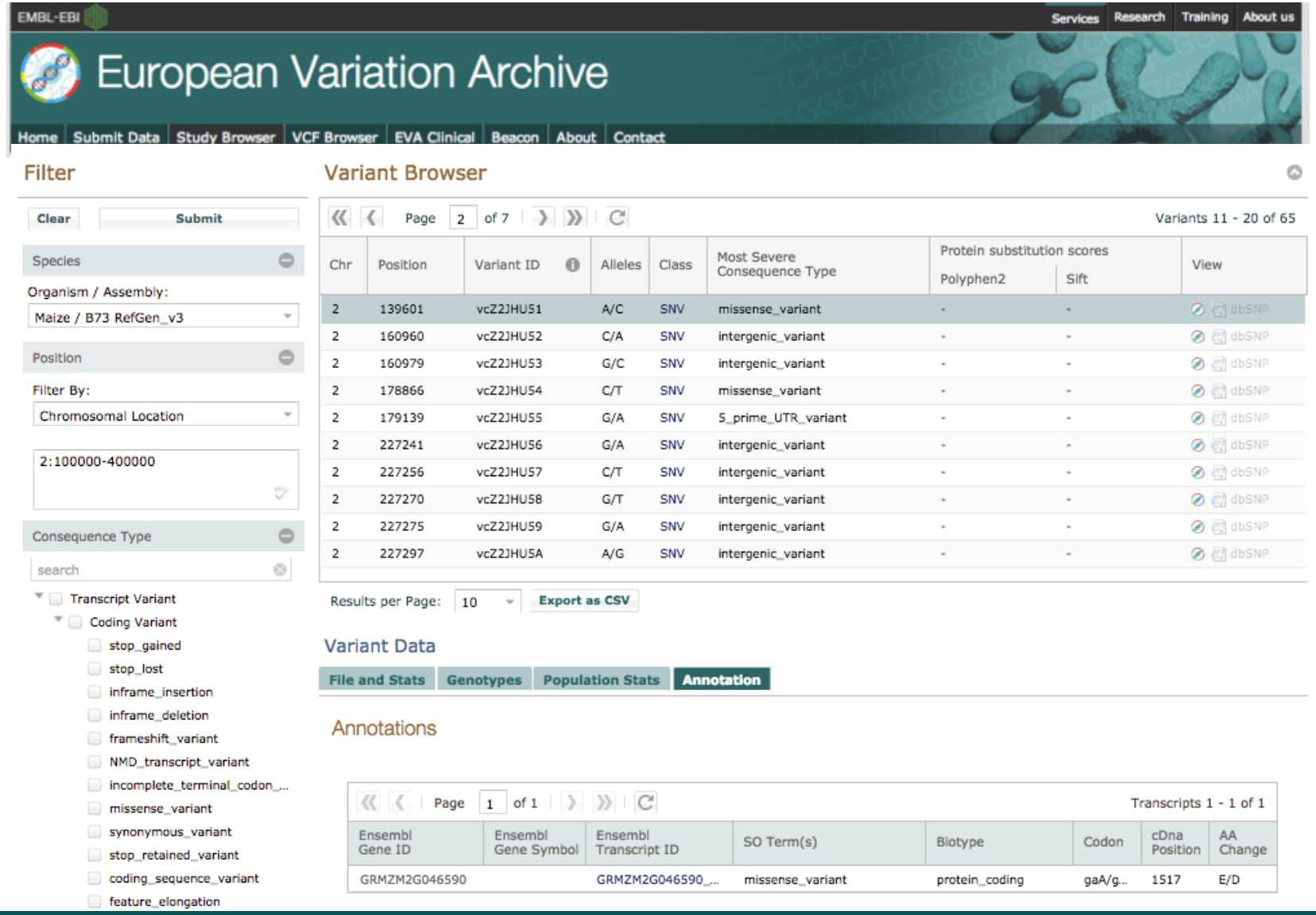
EVA VCF Browser – 1.0



EVA VCF Browser – 1.0

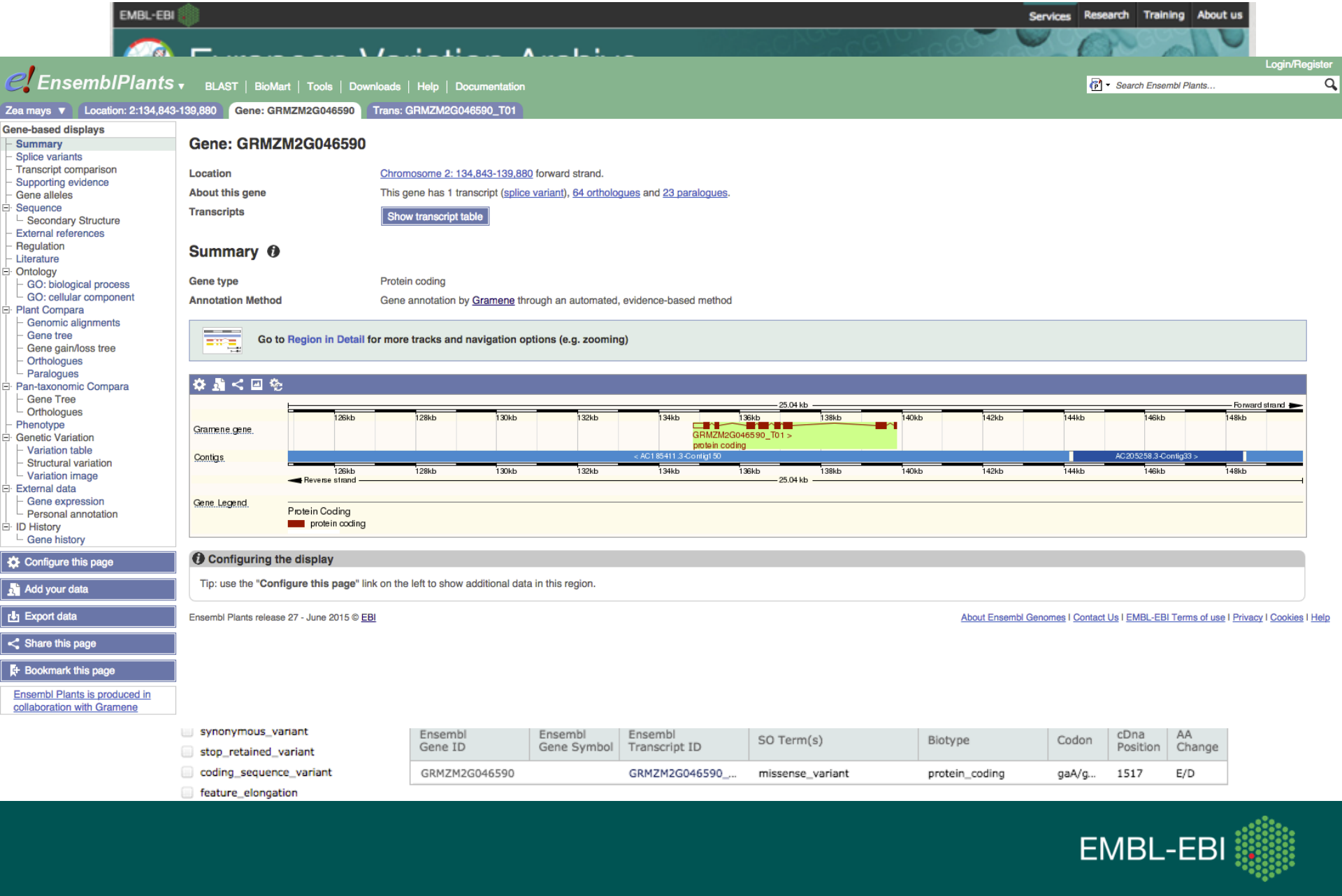


EVA VCF Browser – 1.0



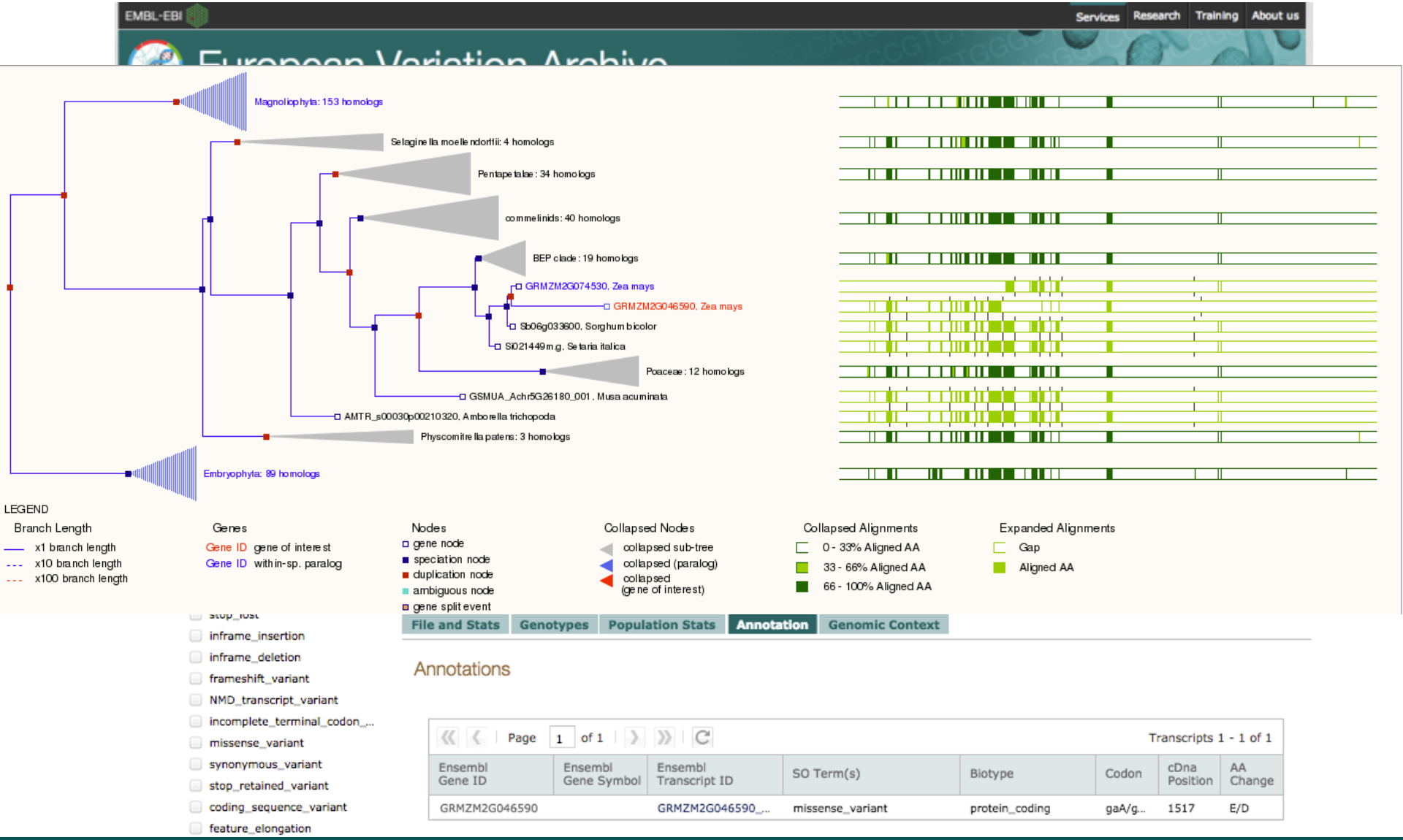
EMBL-EBI

EVA VCF Browser – 1.0



EMBL-EBI

EVA VCF Browser – 1.0



inframe_insertion
inframe_deletion
frameshift_variant
NMD_transcript_variant
incomplete_terminal_codon_...
missense_variant
synonymous_variant
stop_retained_variant
coding_sequence_variant
feature_elongation

File and Stats | Genotypes | Population Stats | **Annotation** | Genomic Context

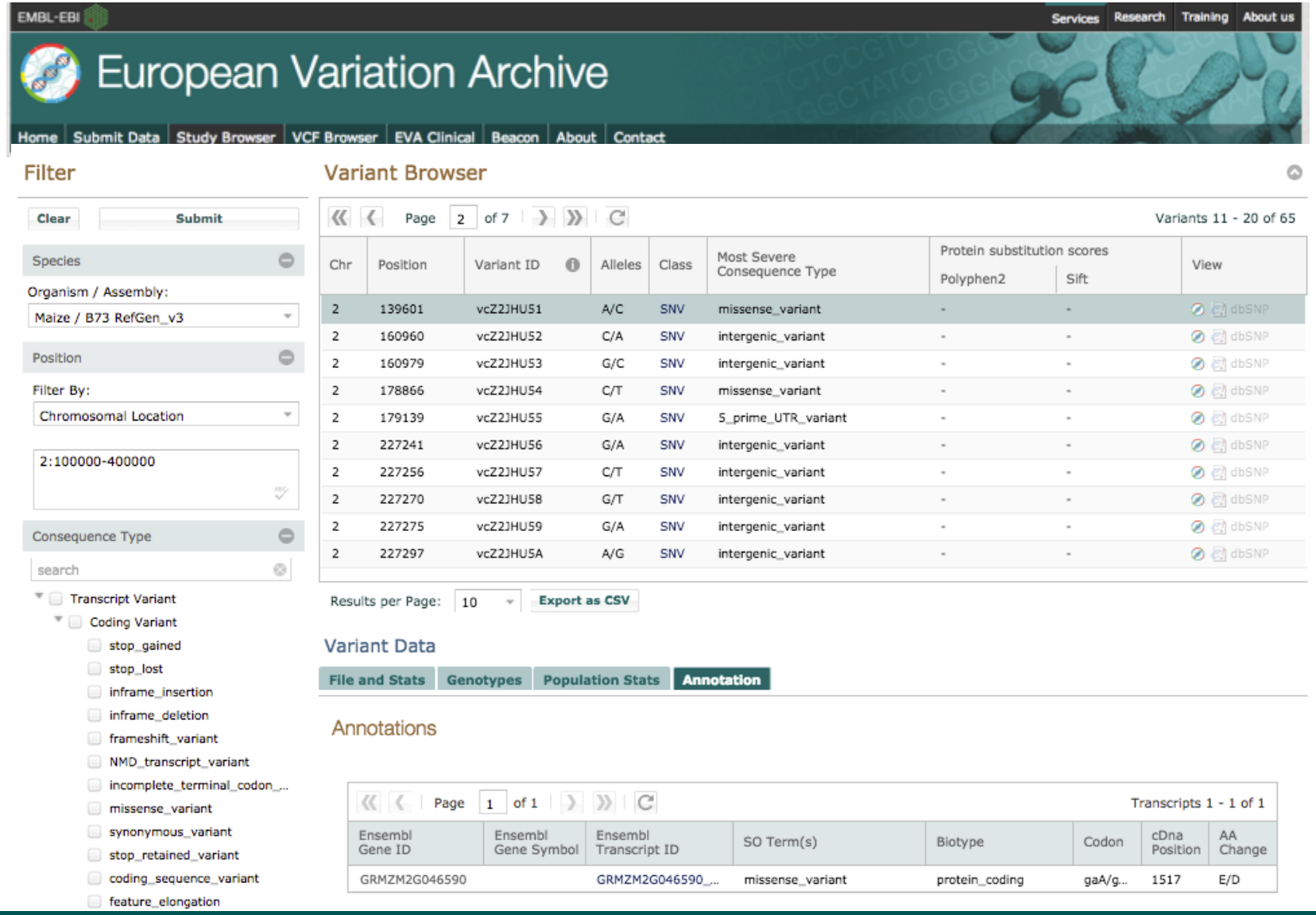
Annotations

Page 1 of 1 — Transcripts 1 - 1 of 1

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

EMBL-EBI

EVA VCF Browser – 1.0



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!

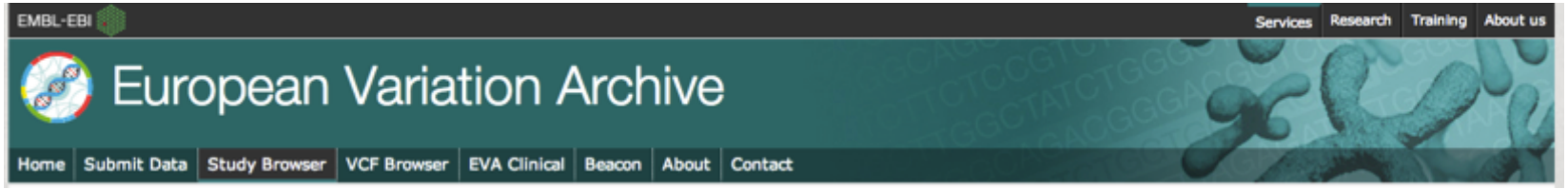
- EVA is now in beta-release, please suggest changes to make it more useful

- 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
  - I'm also around this week

Submit data to EVA



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 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. 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 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.

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

- 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

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

@ebivariation