European Variation Archive: Quick tour

This quick tour provides a brief introduction to the European Variation Archive - an open-access repository for genetic variation datasets that are submitted directly from the community, or are loaded to the resource by internal staff members.

This quick tour was published in August 2017.

Learning objectives:

- A basic understanding of the European Variation Archive resource and how to use it to explore variation data
- Know where to get help and find out more about the European Variation Archive resource

What is the European Variation Archive?

In 2014, we at EMBL-EBI decided to launch a portal to store genetic variation data in a regimented manner. We called it the European Variation Archive (EVA). The objective of the EVA is to serve as a 'one-stop-shop' of open-access genetic variation datasets; to negate the need for researchers, pre-doctoral students, reviewers (anyone, really) to search various locations to access genetic variation data. Instead, we load all datasets (those submitted by the community and those loaded manually by internal staff members) to a single repository at EMBL-EBI.

What can I do with the EVA?

At the EVA we have tried to ensure that you can achieve three key objectives easily (Figure 1):
1. Submit data to the archive
2. Use browsers on the EVA website to view the data
3. Pull data that is of interest to your local infrastructure, either by downloading flat files [2] or accessing the repository computationally via our API [3]

Figure 1 The EVA user experience.

Browsing datasets archived at the European Variation Archive

The EVA study browser [4] lists all datasets that have been archived at the resource (Figure 2). You can filter this list by variant length, genome assembly, and/or type of study (i.e. whole genome sequencing, exome sequencing, etc.).

Figure 2 The EVA Study Browser.
Each dataset is given its own study page (Figure 3), where you can see additional information such as a description of the study, how many samples were analysed, publications that are associated with the data, and the there are also links where you can download the VCF files as they were submitted to the EVA.

Figure 3 Example EVA Study Page, for the Genome of the Netherlands release 5 data.

Accessing variant data at the European Variation Archive

The variation data housed at the EVA has been described and annotated in different ways. Importantly, we normalise all variant data and annotate this homogenous variant population with only one variant consequence predictor: Ensembl’s Variant Effect Predictor [5]. Additionally, we calculate allele frequencies in a standardized manner - and also group variants from samples that are from a particular population together, in order to calculate population allele frequency values.

You can read more about our variant normalisation and processing steps here [6].

We provide access to these normalised and annotated variant data in two ways:

1. The EVA Variant Browser [7]

Filtering options are available on the left-hand side of the web browser (Figure 4). Once you have selected your
species/assembly combination the filters allow you to refine the variant population on any combination of variant consequence, allele frequency, and protein substitution score. Results can be shown from all studies that relate to a species/assembly combination, or limited to only a subset.

All results passing your filtering options are displayed in the top panel of the browser.

The bottom panel of the browser displays detailed information for a particular variant including overlapping genes and transcripts, all variant consequence annotations, datasets at EVA where the variant is present, sample level genotypes, population allele frequency statistics, and any clinically relevant assertions that are linked.

Figure 4 The EVA Variant Browser.

2. The EVA API [8]

You can computationally access the EVA normalised and annotated variant data via the API (Figure 5). We provide endpoints for species, studies, files and variants, and our API is also integrated with the GA4GH Beacon and Variants API. You can read more about our API [8] and on GitHub [9].
Submitting data to the European Variation Archive

What are the minimum requirements for submission to the European Variation Archive?

EVA accepts submission of genetic variation data based on three criteria:

1. The genome assembly used is International Nucleotide Sequence Database Collaboration [10] (INSDC) registered, or will be at point of submission

2. The variation data is described in valid VCF file(s) - this can be tested prior to submission using the EVA VCF Validation Suite [11]

3. Submitted variants are associated with allele frequency values and/or genotypes and/or the raw materials to calculate allele frequency values internally (i.e. Allele Count AND Allele Number values)

What are the key stages of the EVA submission process?

1. Contact

Contact eva-helpdesk [at] ebi.ac.uk to provide details of your submission. You will receive a custom private FTP [12] location for you to upload files.
2. Prepare

Submissions to the EVA consist of VCF file(s), any associated data file(s), and metadata [13] that describe sample(s), experiment(s), and analysis that produced the variant and/or genotype call(s). This metadata is described in an Excel template that can be found here [14]. You can also find a mocked up version [15] of this template that has been completed for a fictional study on our website.

VCF file(s) submitted to the EVA must be truly valid a 4.X version of the file format specification [16]. Files can be validated prior to submission using our validation suite that is available on GitHub [17].

3. Submit

Upload your VCF file(s), associated data file(s) and EVA metadata template to your private EVA FTP location.

4. Receive

The EVA aims to process submissions within two business days. Accession numbers will be sent via email to the submitter upon successful archival of the deposited data.

Get help and support on the European Variation Archive

Support

- There are separate sections for general information, data access, submissions, and variant accessions at our EVA help pages, found here [6]

- If you would like to be notified of changes and improvements to the European Variation Archive, you can subscribe to our low-traffic announcement mailing list [18]

- For general enquires, or to start a submission to the EVA please contact eva-helpdesk [at] ebi.ac.uk

- For bug reports, or to suggest a new feature please start a ticket at our GitHub page here [19]

Related courses

An introductory webinar to the European Variation Archive can be found here [20].

Collaborators

The ??EVA & GE?UVADIS European Exome Variant Server ?(GEEVS) work in collaboration to coordinate common data formats for data exchange. As part of this collaboration, we fully endorse the variant calling protocol detailed on the GEEVS website [21], as adherence to this protocol for variant calling permits direct comparison and/or aggregation of results from different datasets.

Some of the technical and analytical features of the EVA were developed in collaboration with the department of Computational Genomics led by Joaquin Dopazo at the Principe Felipe Research Centre, Computational Genomics Department (CIPF) [22].
Your feedback

Please tell us what you thought about this Quick tour. Your feedback is invaluable and helps us to improve our courses and enhance your learning experience.

Contributors

Gary Saunders [1]

EMBL-EBI
Genetic Variation Scientific Curator - Paschall team: Variation

Gary Saunders is an EBI curator of the European Variation Archive and related resources: the Database of Genomics Variants archive and the European Genome-phenome Archive. It is Gary's responsibility to manage the data within these resource(s) to ensure accuracy, clarity and discoverability. Previous to this position, Gary was a curator of the GENCODE project, which provides the gene set for the Ensembl genome browser.

Gary moved into curation following the completion of his PhD at the University of Glasgow, where he employed a variety of phylogenomic and bioinformatic methods to investigate drug resistance in nematode parasites of human and livestock importance. ORCID iD: 0000-0002-7468-0008

Source URL: https://www.ebi.ac.uk/training/online/course/european-variation-archive-quick-tour

Links
[1] https://www.ebi.ac.uk/training/online/trainers/garys
[2] https://www.ebi.ac.uk/training/online/glossary/flat-files
[3] https://www.ebi.ac.uk/training/online/glossary/api
[4] http://www.ebi.ac.uk/eva/?Study%20Browser&amp;browserType=sgv
[7] http://www.ebi.ac.uk/eva/?Variant%20Browser&amp;species=hsapiens_grch37&amp;selectFilter=region&amp;studies=PRJEB4019%2CPRJEB6930%2CPRJEB5439%2CPRJEB5829%2CPRJEB8652%2CPRJEB8650%2CPRJEB8639%2CPRJEB6042%2CPRJX00001%2CPRJEB15385%2CPRJEB8705%2CPRJEB17529%2CPRJEB19524%2CPRJEB19523%2CPRJNA289433%2CPRJEB8661%2CPRJEB7895%2CPRJEB7217%2CPRJEB7218%2CPRJEB6041&amp;region=2%3A48000000-49000000
[8] http://www.ebi.ac.uk/eva/?API
[12] https://www.ebi.ac.uk/training/online/glossary/ftp
[13] https://www.ebi.ac.uk/training/online/glossary/metadata
[14] https://www.ebi.ac.uk/eva/files/EVA_Submission_template.V1.0.5.xlsx
[17] https://github.com/EBIvariation/vcf-validator
[19] https://github.com/EBIvariation/
[20] https://www.ebi.ac.uk/training/online/course/european-variation-archive-embl-ebi-webinar
[22] http://bioinfo.cipf.es/