Version

4

User Guide

CAFE VARIOME

User Guide

Department of Genetics
University Road
Leicester
LE1 7RH

Table of Contents

Installation

Server requirements

Cafe Variome requires the following components to be installed on the host server:

- Apache webserver
- PHP
- MySQL
- Elasticsearch (comes bundled with the Cafe Variome install package)

If you are a Windows user and wish to install Cafe Variome then we recommend you try using one of the following all-in-one WAMP solutions:

http://bitnami.com/stack/wamp

http://www.apachefriends.org/en/xampp.html

Unpack Installation Package

Download the Cafe Variome installation tarball and copy it to your webserver root directory (or anywhere in the web directory you prefer). Use the following command to unpack the tarball:

```
tar -pzxvf cafevariome.tar.gz
```

Once you have unpacked the tarball you should navigate to the install directory on the target machine in your web browser e.g.

`http://localhost/cafevariome/install`

Install Wizard

When you first navigate to the install page you will either be presented with the installation wizard or be directed to a page informing you of any errors that need to be resolved before installation can take place. Suggested ways to resolve these problems are given on this page and are also covered in the "Common Install Problems" section below.

If the install wizard appears then you should complete the fields in each step, interactive help is available for each field by hovering over the question mark button with your mouse. If all fields were filled in correctly then you will be presented with the install button on the final page of the wizard.

Sample Data

If this is your first installation of Cafe Variome, we recommend that you keep the boxed checked that installs sample data. The sample data helps to give new users an idea of how a typical Cafe Variome installation could be setup. It includes mock variants from the BRCA1 and BRCA2 genes that are being shared between three example diagnostic labs.

Starting Elasticsearch

After you have successfully installed Cafe Variome you will need to start Elasticsearch. Elasticsearch is required by Cafe Variome as it is used as the search engine for the discovery interface. N.B. you must use the version of Elasticsearch supplied with the installation package which is located in the elasticsearch directory. The server should be started from the command line using:

```
./elasticsearch/bin/elasticsearch
```

Common Install Problems

1. The most common issues that need to be addressed are permissions to directories that need to be accessed by the apache webserver. If you are presented with such errors then you should modify the permissions of each of the listed directories so that they are writable by the apache user (ideally at group level), e.g.

   ```
   chmod 775 /var/www/htdocs/cafevariome/application/config/database.php
   ```

2. Another common problem is with apache not being able to read the .htaccess file. To enable this you should add the following line to your apache config file and restart apache:

   ```
   AllowOverride All
   ```

   If the errors are still showing after this line is added then you should also check that mod_rewrite is enabled in apache. In Ubuntu mod_rewrite can be enable by the following command (requires apache to be restarted):

   ```
   sudo a2enmod rewrite

   sudo /etc/init.d/apache2 restart
   ```

3. If after clicking the install button the install progress popup takes longer than expected (more than a minute) then you check the apache error logs. If there are errors such as "Execute failed: (1364) Field doesn't have a default value" then you need to disable strict mode in MySQL. This can be done by editing the my.cnf file (often located at /etc/my.cnf or /usr/local/mysql/my.cnf or /etc/mysql/my.cnf or ~/.my.cnf) and remove the "STRICT_TRANS_TRABLES" option from the following line so it looks as follows:

   ```
   sql_mode=
   ```

You should then restart MySQL. Please note that we are working to resolve this issue and having to remove this option will not be necessary in future versions of Cafe Variome.

Remove Install Directory

If installation was successful you will be presented with a success page and it is recommended that you delete the installation directory before proceeding (especially if this is a live site), e.g.

```
rm –Rf /var/www/htdocs/cafevariome/install
```

Post Install Issues

If you see an error similar to the following: "*Warning: date() [function.date]: It is not safe to rely on the system's timezone settings…*" in the header of every page you should edit your php.ini file (usually found in /etc/php.ini) so that there is a value present for the date.timezone parameter, e.g.

```
date.timezone = "Europe/London"
```

You will need to restart your apache webserver after adding in this value.

Administrator Dashboard

Once installation is complete you may log into your instance using the user details you entered during the install process. This user is automatically set to be a member of the admin user group and so has full access the main administrator dashboard, which can be accessed through the link in the header bar once you are logged in.

The administrator dashboard is a central part of a Cafe Variome installation and the sections that follow aim to give a brief overview of the functionality of this interface and how access to variants can be controlled. It is important that as an administrator that you understand how each of the sections of the dashboard work.

Sharing Policies

A key component of the "data discovery" concept is the use of sharing policies. Variants that are imported into a Cafe Variome instance can be released under one of the following three sharing policies:

- **openAccess**: Like a journal where variant is fully available for public download
- **restrictedAccess**: In order to access the variant the user must belong to a pre-approved group
- **linkedAccess**: Only reports the existence of a variant and the user is linked to the source database in order to access the full record

Users and Groups

Users and groups are controlled through the admin panel and are used for access control – that is, the control of access to sources and variants. Once a user has been created they can be assigned to groups which permits them pre-approved access any restrictedAccess variants in sources (controlled through the sources and variants admin panels). There are two groups available by default, the admin group and the general group. Any number of additional groups may be created by the administrator, for example, different groups to represent different diagnostic laboratories.

Admin User Group

If a user is added to the admin user group then they will have full access to the administrator dashboard.

Sources

Creating Sources and Source Groups

Sources are created and modified through the "Sources" admin panel. Additionally sources can be taken offline so that they are completely hidden from the discovery interface and will therefore be hidden from all users.

If you chose to install the sample data then there will be four example sources to show how a typical installation could be configured. In addition to dbSNP there are three mock diagnostic laboratories. In the sources section of the administrator dashboard specific groups have been assigned to these diagnostic laboratories and so only users belonging to these groups are permitted to access any restrictedAccess variants in these sources.

Variants

Adding Variants and Variant Sharing Policies

Variants can either be entered manually one at a time or imported in bulk. In the variants section of the administrator dashboard the method of import can be decided by clicking on the "+" action icon for a source. After clicking the add icon a popup will appear which gives the user the option to either import in bulk or to alternative use a form to enter variants one at a time. These two import options are discussed below. Other action icons allow the user to modify/curate variants in a source (including deleting individual variants), or delete variants in bulk.

Sharing policies can be set for variants in a source through the variants panel. This can either be done in bulk (for every single variant in a source), or individually through the curate action button in the variants panel. If variants are set as restrictedAccess – only users belong to the required group will be able to automatically access these variants through the main discovery interface.

Bulk Import

Variants can be imported in bulk in a variety of formats. Import templates for Excel and tab-delimited formats can be generated through the "Settings" section of the admin interface. **N.B. The import template system is explained in more detail in the "Import Templates" section the Settings description below.**

Manual Entry (per variant)

Selecting the "Manually Enter Variants" option directs the user to a form that can be used to enter variants one by one. Autocomplete functionality in fields guides the user through the import process. Variant HGVS nomenclatures are required to be validated by Mutalyzer. If the variant is successfully validated, the genomic coordinate information is automatically populated in the respective fields. If you are using phenotype ontologies then a field for entering a phenotype term will appear for each ontology (see next section "Phenotypes" for more details).

Other Import Formats

Additional custom (or common) import formats can easily be supported, please contact us if you would like to discuss adding an import format.

Phenotypes

Each variant in the system can be annotated with any number of terms from any of the NCBO BioPortal phenotype ontologies. This flexibility allows a variant to be associated with a single disease term, or a complex combination of Entity and Quality (EQ) phenotype descriptions.

Adding Ontologies

New ontologies can be added to your installation through the ontologies section in the phenotypes administration dashboard page. Please note that an ontology cannot be removed if you have already annotated phenotype terms to variants from this ontology.

Adding Ontology Terms (per variant)

Once an ontology has been added all the terms from this ontology will be available when creating or editing a variant (see the variants/manual entry section of this documentation). Terms will autocomplete using the BioPortal widget after 3 characters have been entered.

Adding Ontology Terms (bulk import)

If you wish to use phenotype ontologies for the bulk import interface you must use the following format for the phenotype field for each variant in your bulk import template:

```
Phenotype_source_ID|Phenotype_term_ID|Phenotype_term_name

e.g.

HP|http://purl.obolibrary.org/obo/HP_0001093|Optic nerve dysplasia
```

Multiple phenotypes can be added per variant by separating each phenotype with "||"

```
HP|http://purl.obolibrary.org/obo/HP_0001093|Optic nerve dysplasia||
HP|http://purl.obolibrary.org/obo/HP_0001254|Lethargy
```

Phenotype Local List

To add…

Searchable Ontology Tree

An up-to-date searchable term tree for all ontologies used in the annotations can be generated through the administrator interface (settings/maintenance). This tree needs to be updated after any new phenotypes have been added. Only terms related to both an annotation term and the ontology root are included in the trees. This functionality makes use of the BioPortal API to ensure that the latest version of all ontologies, and associated terms, are available to the user. If there are searchable phenotype terms present then an icon to display the tree will be shown in the variant discovery interface.

Variant Discovery Interface

Variants present in your installation are made discoverable through the discovery interface. This interface allows users to search the contents of an installation and subsequently discover whether a variant is present is a source. The main discovery interface is accessed via the "Discover" tab in the sites header bar. The details of the discovery interface are now discussed.

Select Sources

The user can either select a specific source to search or alternatively search through all sources. All sources are selected by default.

Search Field

The search interface is powered by ElasticSearch and allows for sophisticated "Google-like" queries. See here for an example of the types of queries that can be performed:

http://www.elasticsearch.org/guide/en/elasticsearch/reference/current/query-dsl-query-string-query.html

Example queries are present below the discovery search bar in order to give users an idea of what types of queries are possible. After the user clicks the discover button they are presented with counts of variants present in each source along with the sharing policy under which they can be accessed i.e.

openAccess variants can be viewed without restrictions. Clicking the access icon presents the user with a popup that allows them to view the results of the search term in a variety of formats.

restrictedAccess variants require the user to be logged in and to be a member of the required group in order to have pre-approved access to the variants.

linkedAccess variants are presented as core pointer links back to the source of the variant in the host database. Clicking a link directs the user to the full record of the source variant.

Settings

Administrators have an extensive

Display Fields

Administrators are able to control exactly what fields can be searched on and what is subsequently displayed to users in the discovery interface (both for an individual variant and also for lists of variant "hits"). This can be controlled through the display fields tab in the settings section of the administrator interface (instructions on how to control the display fields are also given on this page).

Autocomplete

A non-redundant list of all genes and reference sequences is used for the autocomplete functionality in the discovery interface. If a variant with a new gene or reference sequence is imported then this list needs to be updated and can be done by clicking on the "Regenerate Autocomplete" button in the maintenance tab of the settings section of the administrator interface. This list can be automatically regenerated daily via a cron job. The setting to enable this is in the automated sub-tab of the maintenance tab in the settings section of the administrator interface.

Import Templates

As discussed above, Excel or tab-delimited import templates can be generated through the "Settings" admin panel and gives user a flexible way for importing their variants in bulk. The core fields to be imported are selected using the "Set Core Fields" button, once this has been done you should click the generate button for whichever import template format you want (Excel or tab-delimited) and the template can be downloaded. You should then populate this template with your variant data (DO NOT change any of the headings) and this template can then be subsequently imported via the bulk import interface discussed above.

Other Settings

There are a number of other settings that can be modified through the core settings tab of the settings section of the admin interface. Help is available for each of these core settings by hovering over the corresponding question mark icon.

Appearance Preferences

Appearance preferences for the installation can be modified through the "Appearance" panel in the administrator dashboard. Background, logo, header colour and fonts can

all be customized. Once you have customized the interface to your liking you may save the preferences as a theme.

News Feed

The main news feed appears on the front page of your installation. New articles can be added through the administrators interface. If you wish to turn off the news feed on the front page then there is a setting for this in the settings section of the administrator dashboard.

Content Management System

The content management system can be accessed through the content management icon in the administrator dashboard and allows menus items and their associated pages to be edited by the administrator. It aims to offer basic functionality similar to popular content management systems such as Wordpress, Joomla and Drupal.

Menus

New menu items can be added to the header bar for your installation. The order of the menus can also be re-arranged by clicking and holding the change order icon for a menu and then dragging it up or down. When you release the mouse button the page order will refresh and the order will be saved automatically. Please note that the "Home", "Discover" and "Contact" menus cannot be deleted, however it is possible edit the page content for the "Home" and "Contact" menus.

Pages

After creating a menu item you must link that menu to page content. This can be done when you edit or add a page. A WYSISYG editor allows you to format the content, HTML source code can also be used in this editor by clicking the source button in the editor.

DAS Server

Each Cafe Variome installation comes with a built-in DAS server. All variants that are set to openAccess or linkedAccess and have genomic coordinates available will be automatically available via DAS. The DAS Server can be switched off completely in the settings section of the administrator dashboard. The main page for the DAS server can accessed at the following URL (substitute CV_BASE_URL with the base path to your installation):

http://CV_BASE_URL/das

API Documentation

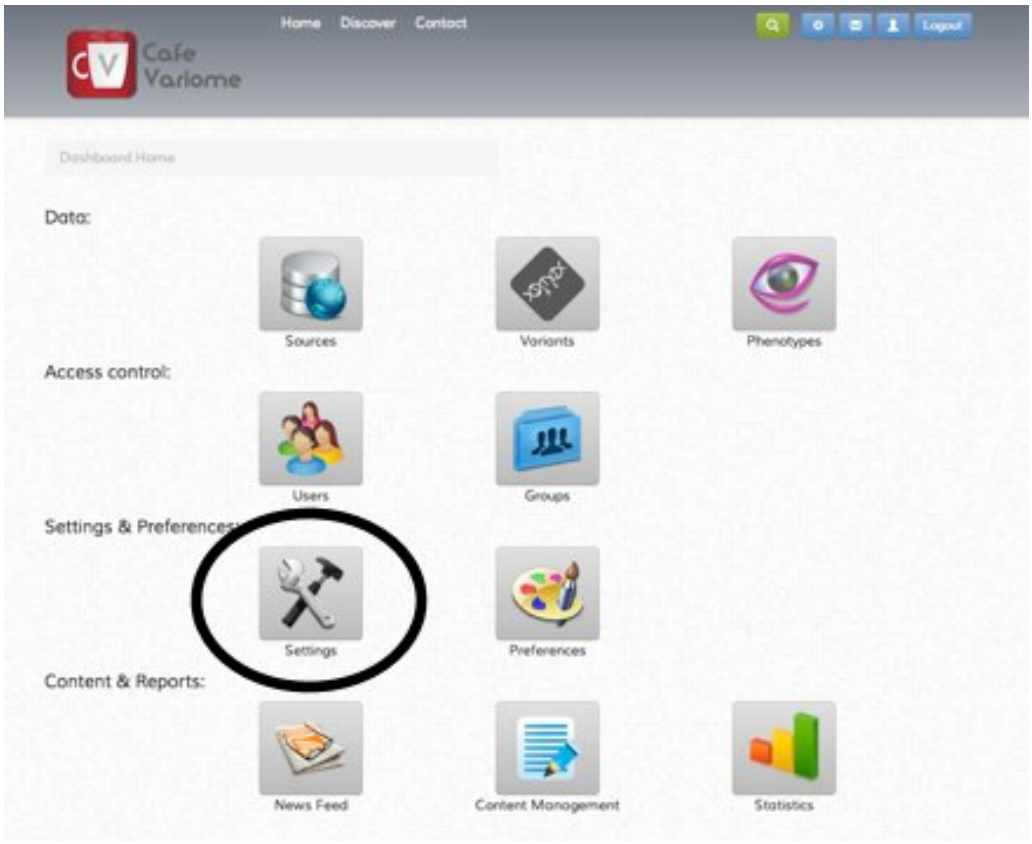
Full, interactive API documentation can be accessed at the following URL (substitute YOUR_CV_BASE_URL with the base path to your installation):

http://YOUR_CV_BASE_URL/docs/api

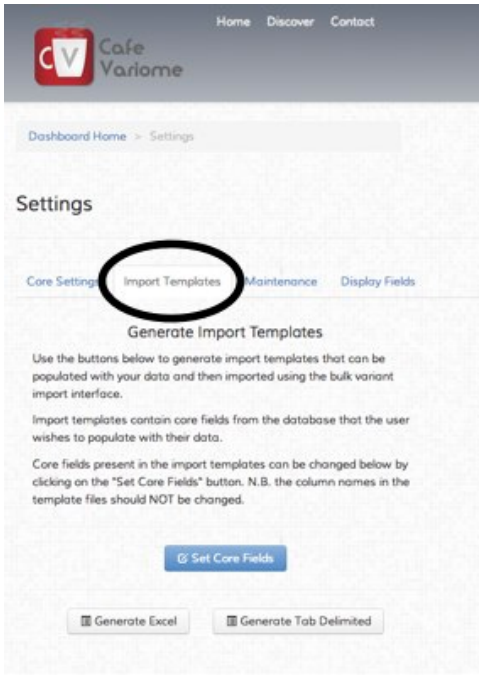
The API allows other tools and resources to programmatically interact with a Cafe Variome installation.

HOWTOs

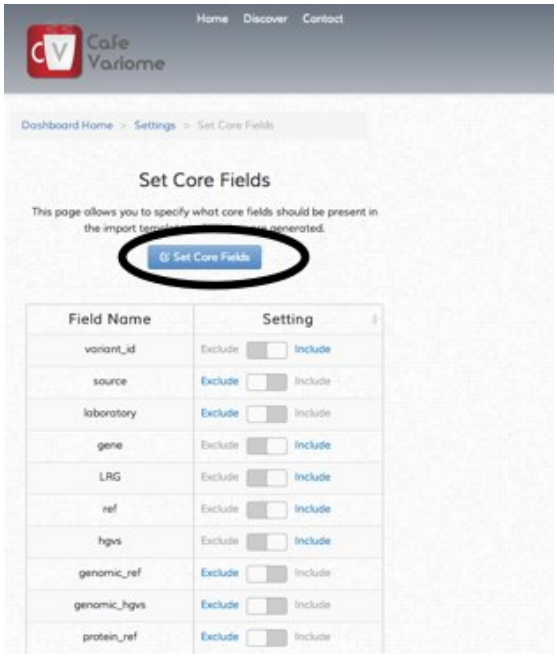
Import Variants in Bulk



1. First you need to generate an Excel import template, to do this click the settings icon in the administrator dashboard. N.B. If you already have a Excel template that you have populated with your data you may skip to step number #6.

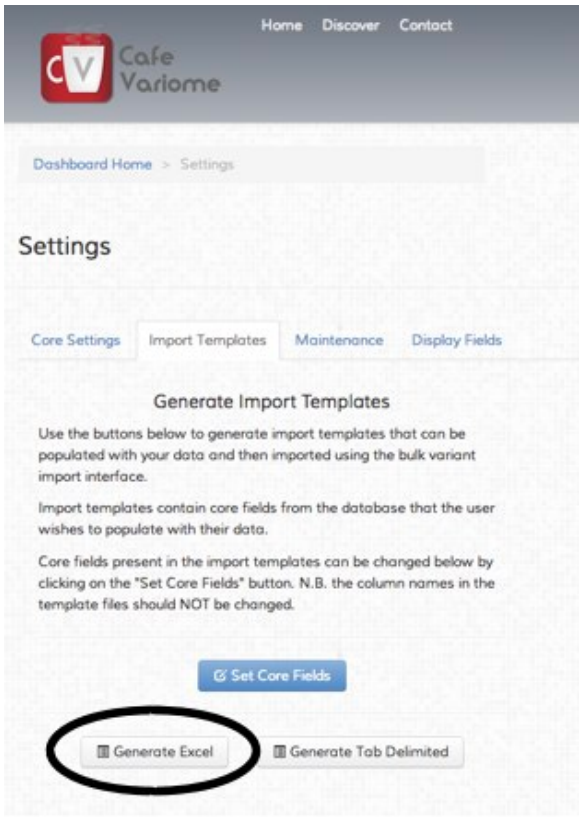


2. Select the "Import Templates" tab and click the "Set Core Fields" button.

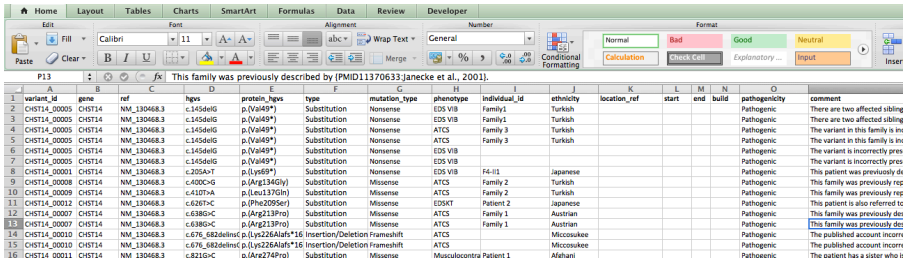


3. In the "Set Core Fields" page you can select which fields you would like to be in your Excel template i.e. the data fields you would like to import. N.B. even after creating your template you do not have to populate all the fields in the template for any

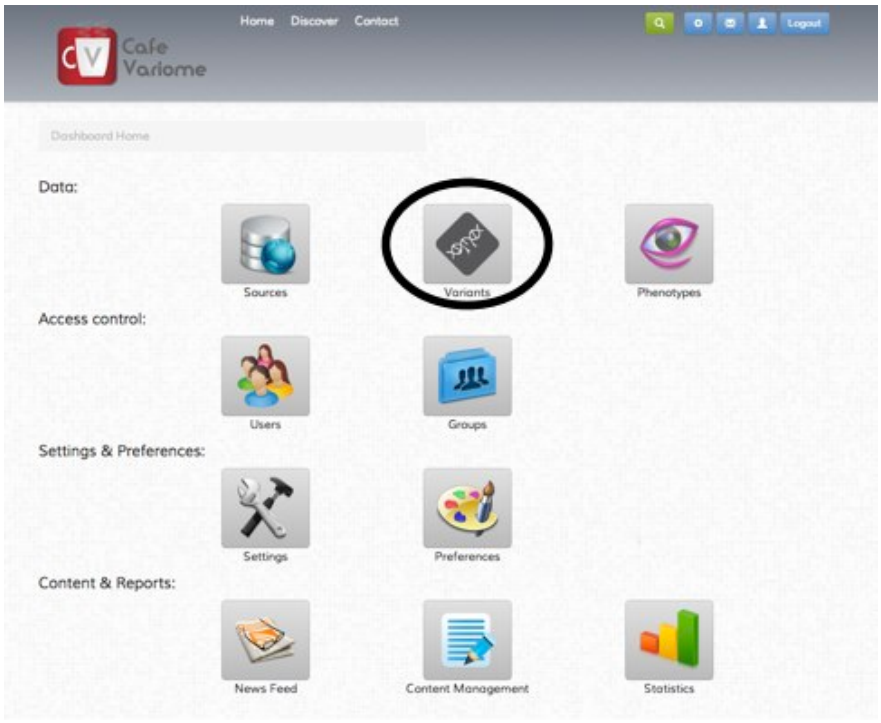
variant. Once you have selected your fields then save by clicking "Set Core Fields" then go back to the "Settings" page.



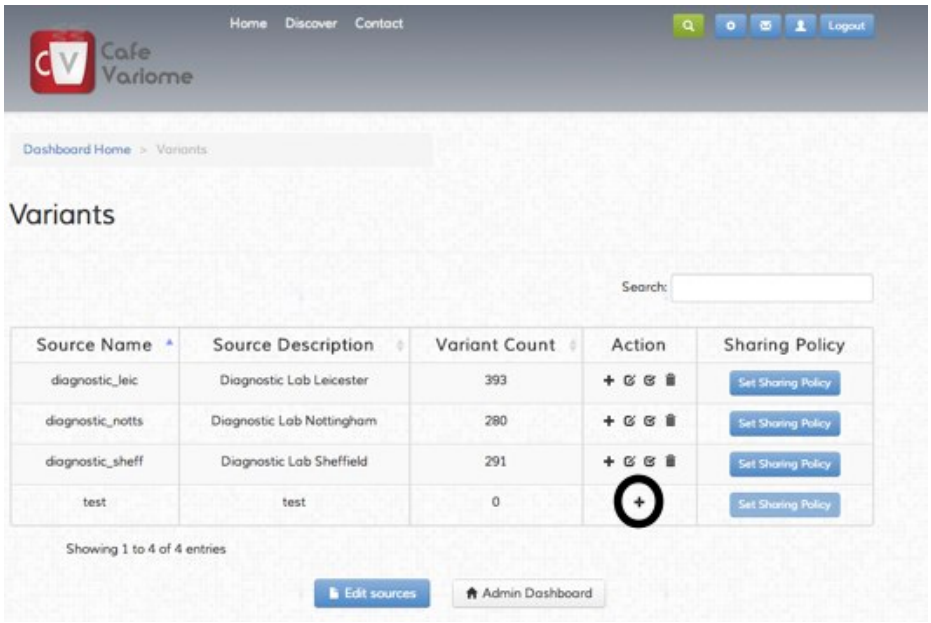
4. Click the "Generate Excel" button to create your Excel template, a popup will appear where you can choose the location and name of the file to save.



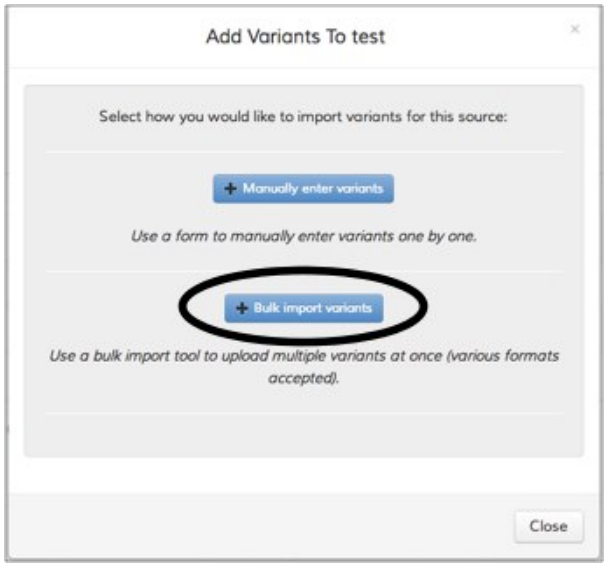
5. Open the Excel document and paste/enter the data you wish to import. N.B. You should not change the header name of any of the columns.



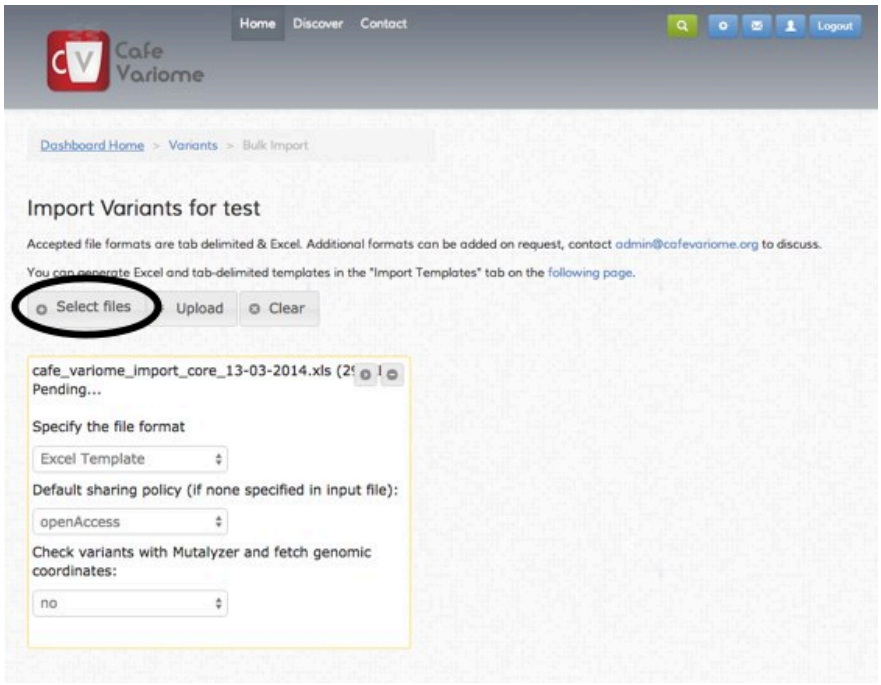
6. Begin the import of your data by going to the variants administrator page by clicking the "Variants" button.



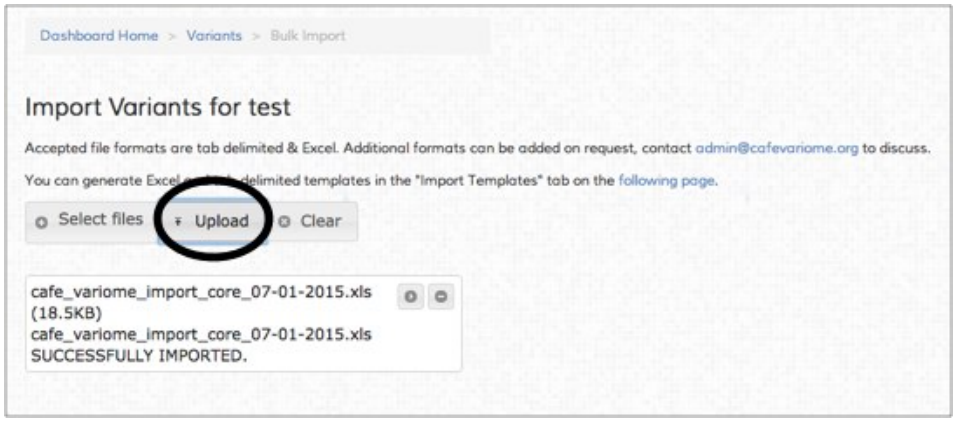
7. Click the "+" button for the source that you would like to import the variants into. If you have not created a source you should do this from the "Sources" page which can be accessed from the main administrator dashboard.



8. After clicking the "+" button the popup above should appear, click the "Bulk import variants" button.



9. Click the "Select files" button and choose the Excel template document you have just created and populated with your variants. Keep "Excel template" selected as the format. You may also choose the sharing policy that variants will be set to.



10. Click the "Upload" button and if all goes well your variants should be imported.

If you encounter any problems with importing data or wish to have your own custom format added to the import formats then please email Owen Lancaster at ol8@le.ac.uk