OOPS
USER GUIDE & TUTORIAL

Requirements

OOPS requires the Java Runtime Environment (version 1.8 or higher), which can be downloaded or updated from https://java.com/en/download .

Download & Installation

OOPS can be downloaded from https://github.com/alexcpa/antiSMASH-OOPS. Once downloaded, one should place the antiSMASH-OOPS.jar file within the folder he considers more appropriate.

To launch the OOPS application, open either the terminal (on Linux/MacOS) or the command prompt (on Windows: Start > All Programmes > Accessories > Command Prompt) and navigate to the folder containing the antiSMASH-OOPS.jar file (e.g. `cd C:\path\to\folder`), then type the following:

```
java -jar antiSMASH-OOPS.jar
```

and hit enter. The application will load as shown.

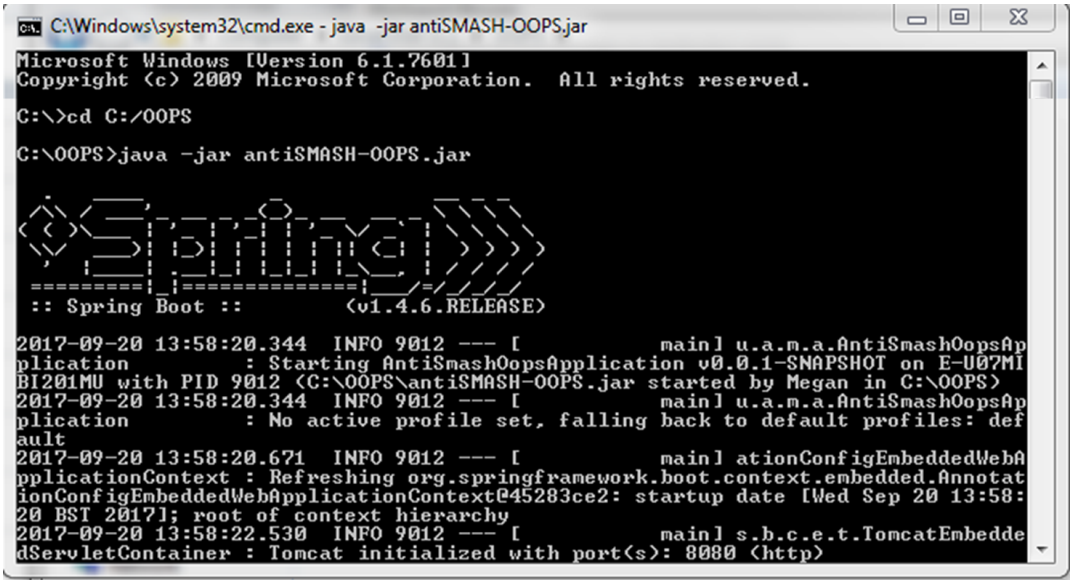

Figure 1: CommandPrompt screenshot

On Windows it is also possible to run the application with a double click. However, this is not advised, since using the terminal allows the user to better monitor the application processes. Once the application has loaded one can view the user interface by opening a web browser (Chrome or Safari recommended) and typing the following in the URL bar before pressing enter:

```
localhost:8080
```

The landing page of the system will show up immediately and will list any available workspace.

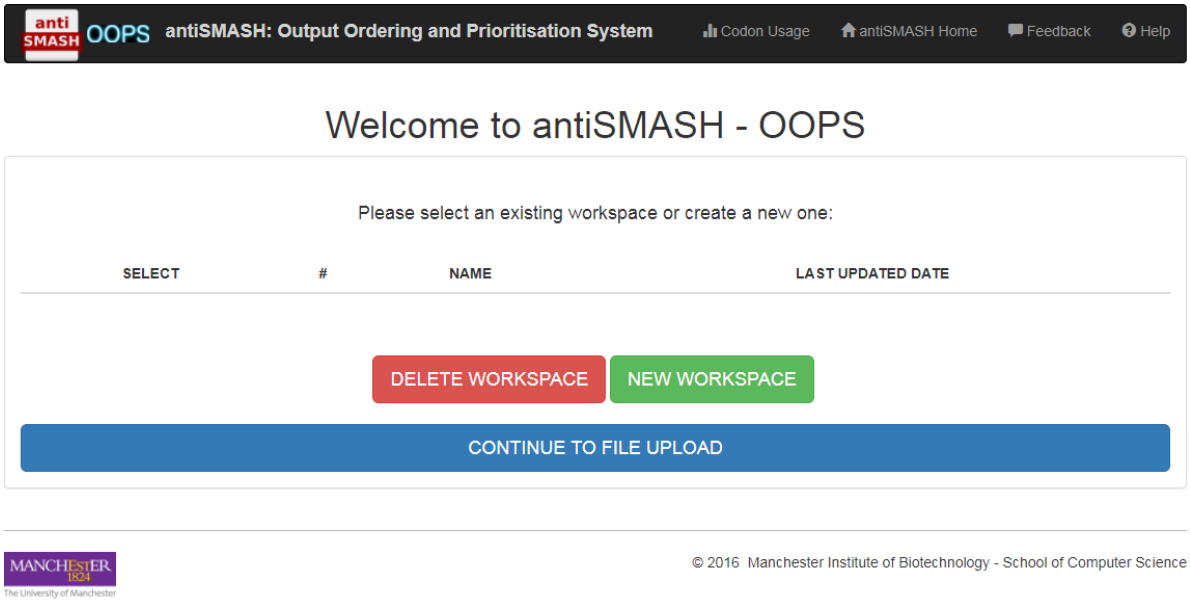

Figure 2: landing page

Workspace creation/load

Workspaces comprise input data previously loaded into OOPS. If the user has not previously used OOPS, no workspaces will be available and the user will need to create a new workspace by clicking on the 'new workspace' button. Here the user can give the workspace an appropriate name, e.g. 'My_first_OOPS_Workspace'. An empty workspace must be created and named prior to uploading any antiSMASH output data to OOPS.

The OOPS landing page allows you to manage workspaces, i.e. create a new workspace, delete an existing workspace or load workspaces previously built.

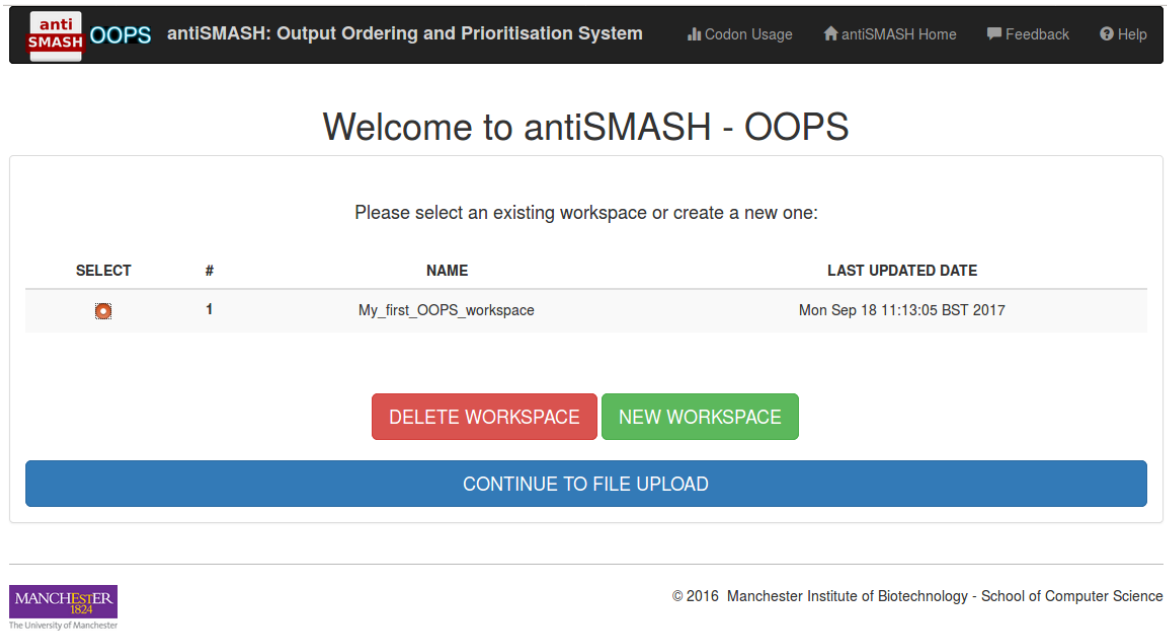

Figure 3: Create/load workspace initial page

antiSMASH outputs loading

OOPS works specifically with prokaryotic and fungal antiSMASH outputs, and it has been tested with the outputs of both antiSMASH 3 and 4. In order for OOPS to work properly, the antiSMASH analyses should be performed including the *Gene Cluster BLAST* and the *Known Gene Cluster BLAST* parameters.

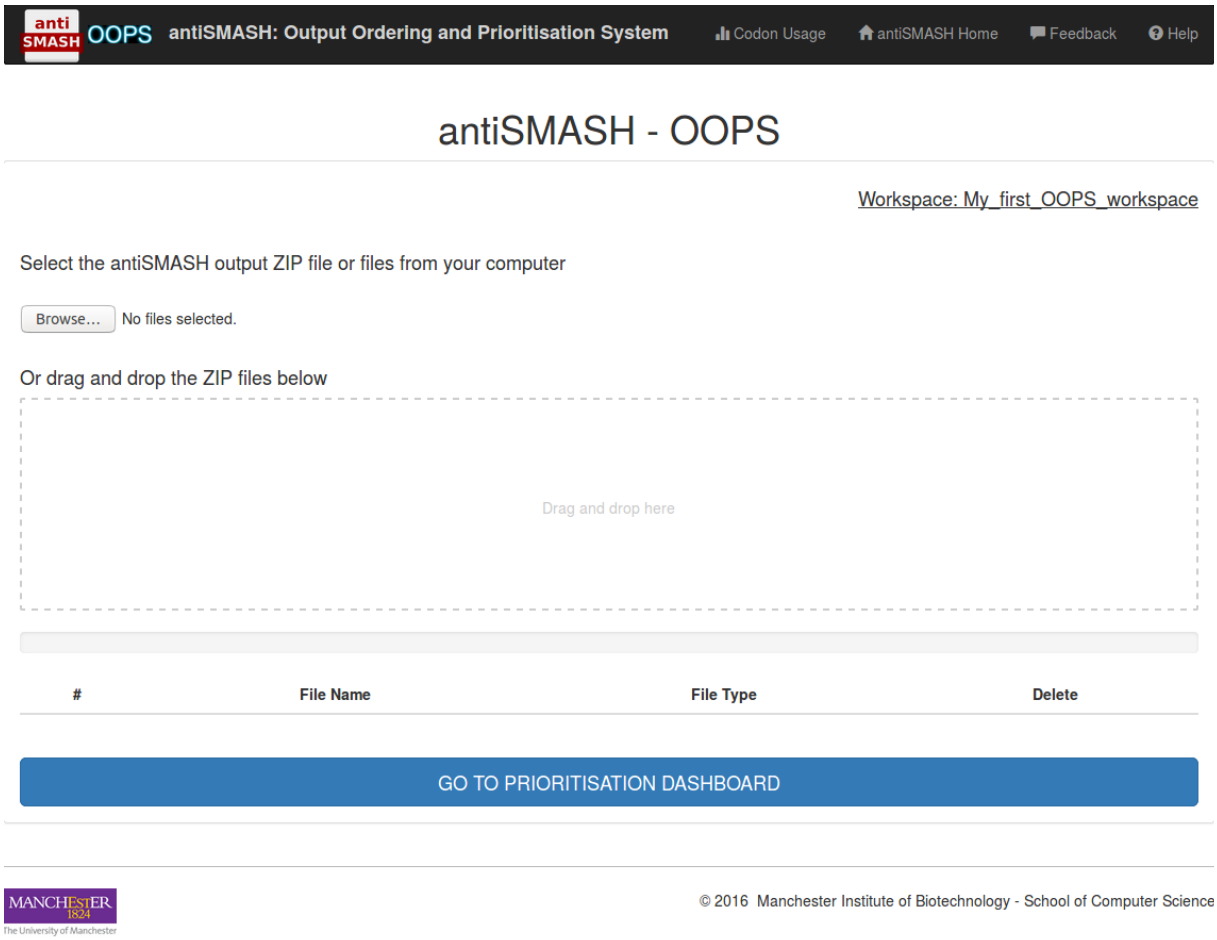

Figure 4: data upload page

It is possible to load as many antiSMASH outputs (as .zip files) as one likes into the OOPS workspace. The number of antiSMASH outputs is only limited by the computation power available. However, for rapid prioritisation we recommend limiting each prioritisation to fewer than 1000 BGCs

AntiSMASH outputs can be loaded into OOPS workspaces with a drag-and-drop function, or by browsing folders accessible via the 'browse' button. After selecting all the antiSMASH outputs of interest, the user could hit the *GO TO PRIORITIZATION DASHBOARD* button. At this point OOPS extracts all the information needed from the inputs. This process may take a while, depending on the size of the dataset. For example, it takes <5 seconds to load *Streptomyces coelicolor* A3(2) antiSMASH output, which contains 27 predicted BGCs.

Prioritization dashboard

After uploading the files, the user is brought to the prioritization dashboard. There are seven prioritization parameters: number of genes, CDS length, GC content, codon bias (against reference

species), similarity to known clusters, self-similarity and phylogenetic diversity. There also is an additional parameter that allows the user to filter BGCs according to the predicted class of the end compound.

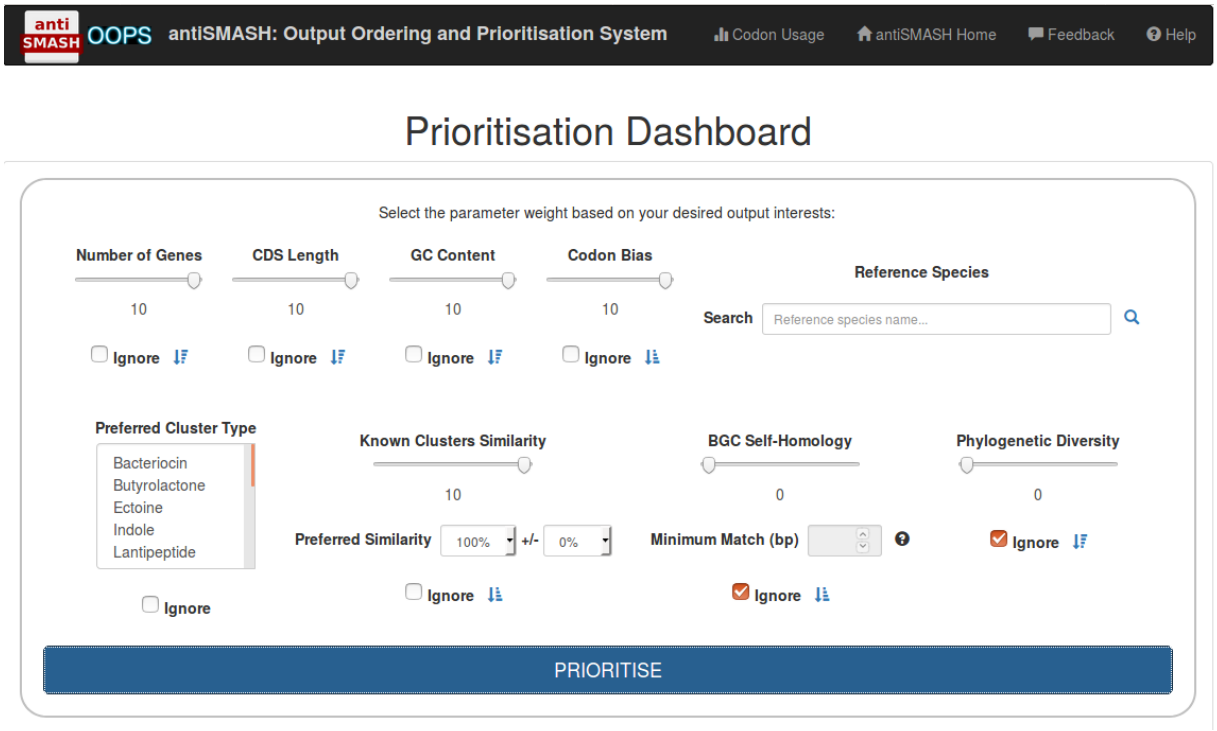

Figure 5: Prioritisation dashboard

Each parameter can be weighted (using a slider) or ignored (using the checkbox), allowing the user to build her own prioritization rules. Moreover, the user can choose in which direction (decreasing or increasing) each parameter should be considered.

Pre-computed dataset

For evaluation and training purposes, it is possible to download a large pre-computed dataset containing the antiSMASH outputs of all the actinobacterial genomes considered in the antiSMASH database [1]. The dataset can be downloaded from https://doi.org/10.5281/zenodo.1000774

After the first start, OOPS automatically creates the folder '`/appData`' where all datasets are stored. After unzipping the downloaded file, the user should place the precomputed dataset in this folder and restart the application (if running). It will now be possible to choose the *antiSMASH_Actinobacterial_BGCs* dataset from the initial workspace selection page. If the user selects this workspace and presses *CONTINUE TO FILE UPLOAD*, the precomputed data is loaded (it might take a couple of minutes) and the user is directly brought to the prioritization pipeline, skipping the upload page.

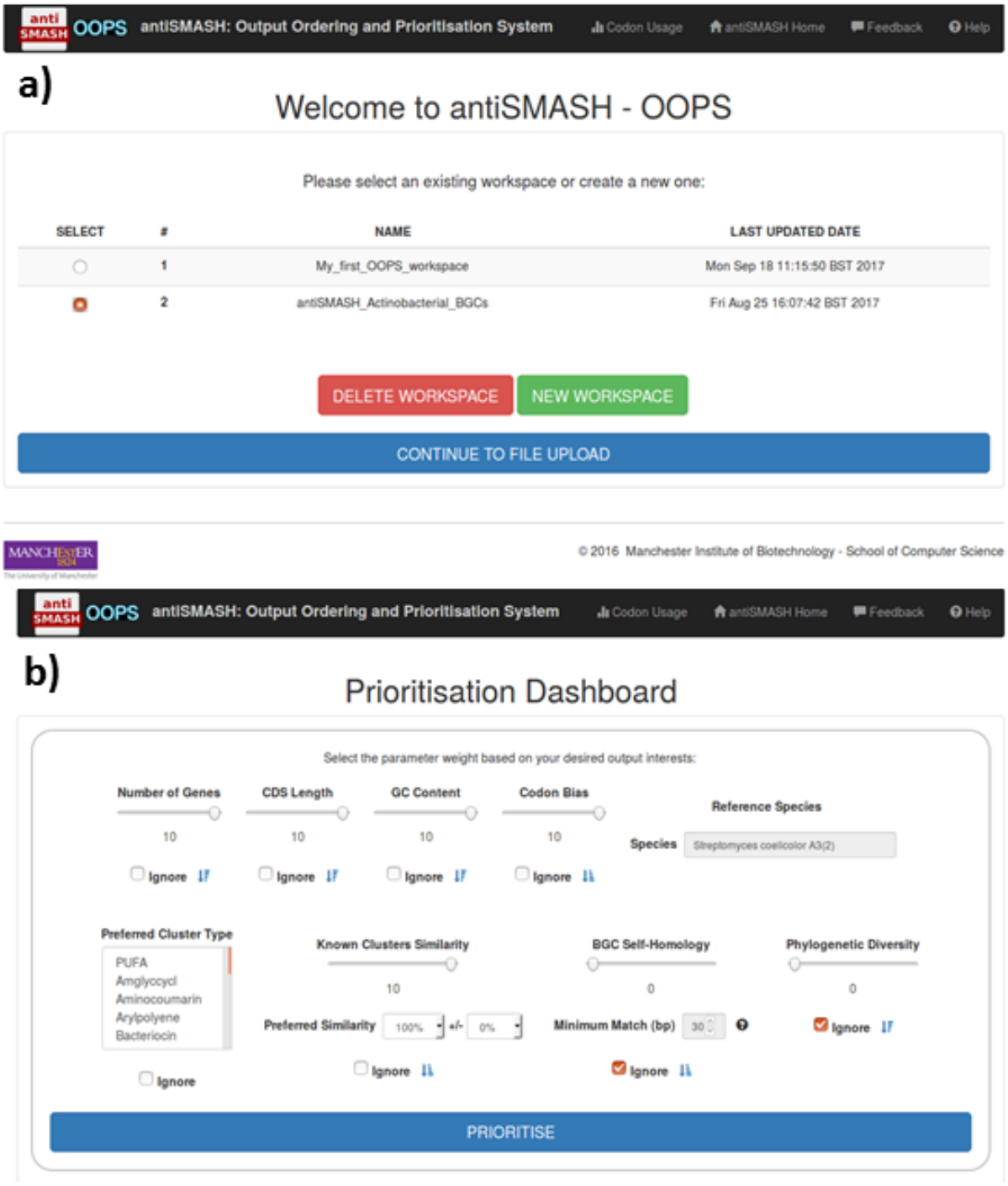

Figure 6: Prioritisation dashboard for the precomputed dataset. One should notice that changing the Reference Species or the Minimum Match parameter will result in a huge increase of the computational time needed, since all the pre-computed data has to be computed again

References

[1] Blin, K., Medema, M. H., Kottmann, R., Lee, S. Y., & Weber, T. (2017). The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Research* 45(Database issue): D555–D559.