

WebCircos User Manual

A. Szymanek, E. B. Sowka, S.C. Weill, R. J. Munro, Y. H. Lo

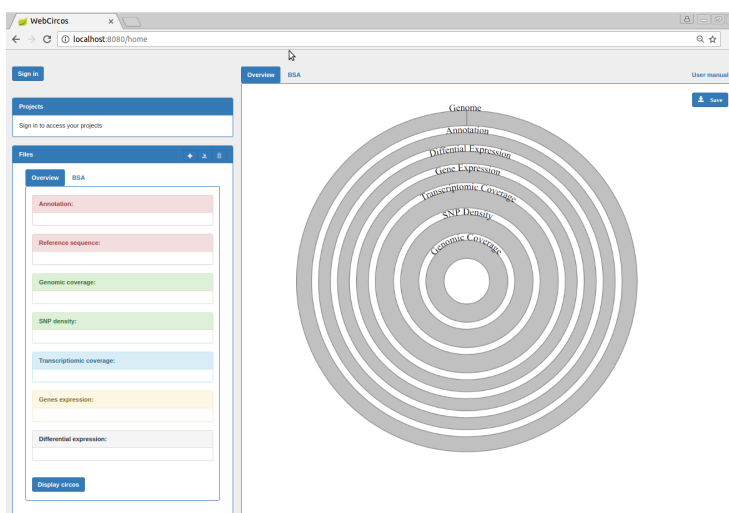
April 28, 2017

Abstract

WebCircos is a Web-based application for visualising both raw and processed Next Generation Sequencing (NGS) data, integrating the BioCircos.js **reference** library and a React.js **reference** based interactive interface. The program is capable of displaying different information tracks for a given genome and Bulk Segregant Analysis can also be performed to identify and highlight areas of probable causative mutations. As NGS experiments generate big datasets, a personalised user space is provided to any registered user, allowing to store information prepared for future visualisations.

1 Running WebCircos

WebCircos is a Web-based application, therefore it is accessible through any web browser - currently supported browsers are Mozilla Firefox, Google Chrome, Opera and Microsoft Internet Explorer. The application is currently run locally and it can be accessed under the address `localhost:8080`. After accessing the localhost address, a Graphical User Interface will be loaded onto the page.



2 User account

As files from Next Generation Sequencing experiments contain enormous amounts of data, WebCircos provides the possibility of creating a user account, which gives access to personal space for storing and managing uploaded files.

2.1 Creating an account and logging in

To login into the account, click the **Sign in** button, enter your user name and password in the pop-up window's form and click **Sign in** (Fig. 1). To create a new account, there is a join us prompt on the same pop-up window with a **Join us!** link, which will open the registration window **NEED A FIGURE**. All the fields need to be filled in order to create a new user account. To finalise registration click the **Register** button (Fig. 2).

Figure 1: Sign in pop-up window

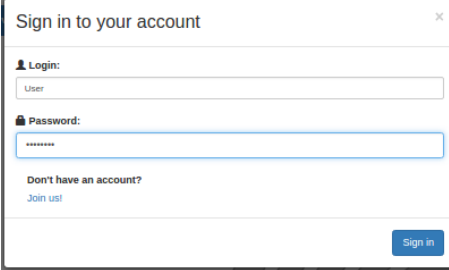
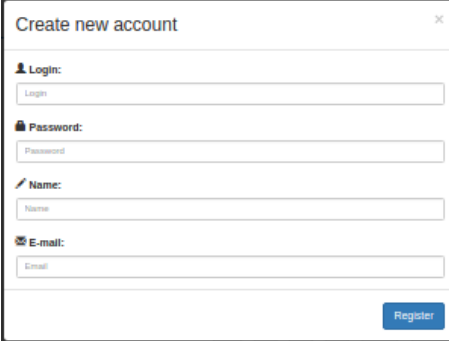
A screenshot of a 'Sign in to your account' pop-up window. The window has a title bar with a close button (X). It contains two input fields: 'Login:' with a placeholder 'User' and 'Password:' with a placeholder '*****'. Below the password field is a link that says 'Don't have an account? Join us!'. At the bottom right is a blue button labeled 'Sign in'.

Figure 2: User registration pop-up window

A screenshot of a 'Create new account' pop-up window. The window has a title bar with a close button (X). It contains four input fields: 'Login:' with a placeholder 'Login', 'Password:' with a placeholder 'Password', 'Name:' with a placeholder 'Name', and 'E-mail:' with a placeholder 'Email'. At the bottom right is a blue button labeled 'Register'.

2.2 User functionalities

Creating a user account enables increased functionalities to help support visualisation management and storage. Anonymous users are able to upload and visualise their data, but this information is only kept temporarily. A registered users account provides storage space and a project management system to order files.

2.2.1 Creating project

After creating a new account, an empty project panel is displayed. To upload files, users must create a new project by clicking the *plus* symbol in the upper-left of the projects panel. Specify a project name in the pop-up window and confirm by clicking the **Create project** button.

2.2.2 Uploading files

Unauthenticated user

To upload a file click the *plus* symbol displayed in the files panel. In the newly opened pop-up, choose the file to upload. WebCircos will try to recognise the data type by the file extension by comparing it to a list of extensions:

- fasta (along with: fa, frn, fn, fas, fna and faa),
- gff (along with: gff2, gff3),
- gtf,
- bedcov,
- results,
- results.sorted,
- vcf,

If the application is not able to recognise an appropriate extension it takes the first line of text found in the file into consideration. A list of recognised text contents is available in Table 1. This allows the application to accept custom .txt files, however the first line must be as stated in the Table for file recognition to work. The application has a few error trapping features shown in the figures below. Errors will show up in pop-up windows if a file with the same name already exists in the selected project and if the file is not recognised. WebCircos also recognises if a file is zipped or compressed and asks for an unpacked version. It is impossible at this time to upload gene expression, differential expression and transcriptomic coverage data without previously uploading an annotation file. If one annotation file has been added to the project and another annotation is about to be uploaded, the previous file is deleted along with any expression and transcriptomic coverage related files.

To change the recognised data type to a different type, tick a different checkbox in the pop-up window and click **upload**. The file is now uploaded and added to a temporary file and the content is being parsed. When these processes are finished the pop-up window will close and the newly added file will be shown in the appropriate dropdown menu in the files panel.

Registered users

For a logged in user, files are stored under projects. To upload a file, first choose a created project, or if none exist create one as outlined earlier, then click the *plus* symbol displayed in the files panel. In the newly opened pop-up, choose the file to upload. WebCircos will again try to recognise the data type by the extension by comparing it to listed extensions:

Table 1: A table of the various tracks WebCircos can display, along with the file types associated with those tracks, the Application recognised first line of the file, and the essential content the Application needs to display the data.

Track:	Recognised file extension:	Recognised first line of file:	Essential content of file:
Sequence	fasta, fa, frn, fn, fas, fna, faa	>	chromosome sequence and name
Annotation	gff, gff2, gff3, gtf	##gff	gene position and name
Expression	results	gene_id	gene expression level
Differential expression	results.sorted	"PPEE"	comparative expression levels
Transcriptomic coverage	bedcov	mRNA:	average reads depth
Variants	vcf	##fileformat=VCF	SNP number, SNP read depth and allelic frequency

- fasta (along with: fa, frn, fn, fas, fna and faa),
- gff (along with: gff2, gff3),
- gtf,
- bedcov,
- results,
- results.sorted,
- vcf,

If the application is not able to recognise an appropriate extension, it takes the first line of text into consideration. A list of recognized text contents is available in Table 1 below. To change the recognised data type to an alternate type, tick a different check-box and click the **Upload** button. The file is then uploaded to the project folder and the content is being parsed. When these processes are finished, the pop-up will close, and the newly added file will be shown in the appropriate dropdown menu in the files panel.

2.2.3 Removing files

To remove a file from the presently open project, click the *bin* symbol in the files panel. In the newly opened pop-up window choose the file to delete. There is no possibility to recover the file.

2.2.4 Importing files

To import files, choose the project to which you want the files to be uploaded. Next click the *import* symbol located between the *plus* symbol and *bin* symbol. In the newly opened pop-up window choose the project from which you want the file to be taken and the file itself. The file will then be copied to your new project.

3 Data visualisation

Data visualisation in WebCircos is split into two cases: integrative visualisation of information stored in provided files and Bulk Segregant Analysis visualisation, as they both require slightly different approaches and cannot be visualised on the same Circos.

3.1 Integrative data visualisation

For each Circos track only one file can be visualised at a time. As previously stated, only one annotation file can be uploaded to the project and if the file has been uploaded, it will be included in the Circos automatically. In the **Overview** tab of the files panel, choose a file to display for each track from the corresponding dropdown list **NEED A FIGURE**. If not all tracks are to be displayed, select '—' **MAYBE A SCREENSHOT OF A BUTTON?** from the files dropdown list for any empty tracks. To render a Circos, click the **Display circos** button - visualisation will be displayed in the corresponding tab of visualisation panel **BEAUTIFUL SCREENSHOT OF A PANEL WITH CIRCOS**. A list of tracks that can be displayed and the files needed for that track can be found in Table 1.

3.2 Bulk Segregant Analysis

The second case in WebCircos is visualisation of Bulk Segregant Analysis (BSA) experiments, an analytical method designed to identify genetic links between two parents and their sibling crossed F2 generation. One of the parents is Homozygously wildtype, displaying a measurable phenotype, and the other is Homozygously mutant, with a disruption to the WT phenotype phenotype. In order to identify the causative mutation, the F2 generation is pooled by whether or not individuals display the measured phenotype. Individuals within a single bulk are characterised by identical phenotypic or genomic features, but may differ in terms of all unlinked features, therefore two bulks are genetically different in given region, but potentially consistent in others Michelson et al. (1991).

In order to visualise BSA results, files with information about variants must be provided for both parents and both pool samples in the Variant calling file (VCF) format, along with a reference sequence for the organisms genome. The pool VCFs are intersected against each other, to leave only unique SNPs between them. In order to link regions, the allelic frequency is used to identify the number of Homozygous SNPs and Heterozygous SNPs in that region. The homozygosity ratio is then calculated for all four files, and the homozygosity ratio for each pool is compared against both parents. When the homozygosity

ratio is high (currently more than 0.99) for both pool and parent, that region will be shown in a link.**PICTURE HERE PLOX** Therefore, in order to perform BSA visualisation five files are needed (one sequence and four VCF files) are required.

In order to display the BSA circos, select the BSA tab located at the top of the files panel. Upload the four VCFs and sequence file by clicking the *plus* symbol located at the top right of the files tab, and selecting the required file in the newly opened pop-up window. Once uploaded, select the correct files in the dropdown menus in the BSA tab. To get an accurate image, it is necessary to select the correct file for each dropdown menu. The dropdown menus are labelled as followed -

- Reference Sequence
- Parent 1 (WT)
- Parent 2 (MT)
- Pool 1 (WT)
- Pool 2 (MT)

The corresponding VCF file must be uploaded to the dropdown menu, i.e the wild type parent VCF must be added to the Parent 1 (WT) menu and so on. Once all five files are added, click the *Display BSA circos* button at the bottom of the panel. As data have to be processed when the *Display BSA circos* button is clicked, rendering will take some time. For 1GB of data it will take approximately 5 to 10 minutes.

3.3 Detailed view

Circos is an interactive plot, thus, by hovering over the different tracks tooltips will appear. Zooming in and out of the plot is also possible either by scrolling or double clicking. Panning around the plot is enabled on a click and drag basis.

Tooltips content

Tooltips are temporarily displayed text, enhancing the circos by displaying useful information about the part of the circos currently being hovered over. As information across tooltips changes based on the part of the track being hovered over, the summary table below details the information for each track. Tooltips don't exist for coverage tracks as coverage is displayed as a line and not as a fixed point or inside data bins, making a single point display impossible at this time.

Chromosome view

Chromosome view can be entered by clicking on a chromosome. This view contains the same information as that in the integrative overall view but only for the selected chromosome. Tooltips are also available in this view by hovering over tracks. Switching back to the overview is performed by clicking on the *back* dot in the middle of the plot.

Tracks	Information	Tool-tip information
GENOME	sequence	sequence length
ARC	genes	chromosome name start position stop position gene name
HEATMAP	gene expression	chromosome name and differential expression position track name log fold change value
LINE	genomic coverage and transcriptomic coverage	no information
HISTOGRAM	SNP density	chromosome name position SNPs count per million base

Table 2: Information summary for all the displayed tracks for the integrative plot

Tracks	Information	Tooltip information
GENOME	An overall representation of the genome	Sequence Length
LINK	Homozygosity between pool and parent	link name (chromosome, start and stop position)

Table 3: Information summary for all the displayed tracks for the BSA plot

3.4 Installation notes

References

Michelmore, R. W. / Paran, I. / Kesseli, R. V.(1991): *Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations.*9828-9832.