VDJ SERVER

Quickstart Guide

VDJServer Release 1.0

Prepared by: The VDJServer Team

Date: Last updated March 2017

Quickstart Guide

GENERAL INFORMATION

VDJServer (www.vdjserver.org) is a free, scalable resource for performing immune repertoire analysis and for sharing data, analysis pipelines, and results. It is comprised of a suite of repertoire analysis tools that are accessible via a web interface and run on the high-performance computing resources at the Texas Advanced Computing Center (www.tacc.utexas.edu). VDJServer has two goals: (1) to facilitate repertoire analysis by providing access to sophisticated analysis tools and high-performance computing through an intuitive user interface, and (2) to make repertoire analysis easily and reliably reproducible by providing standardized analysis pipelines and mechanisms for publishing custom pipelines, data, and results.

VDJServer currently provides the following analysis components:

- Read pre-processing, such as quality filtering and demultiplexing, using the VDJPipe and/or pRESTO applications
- V(D)J assignment using the IgBlast application
- Repertoire analysis using the RepCalc application

These components can be configured to run as seamless pipelines through the VDJServer web interface, or they can be downloaded for local installation and use. All software is available under an open source license.
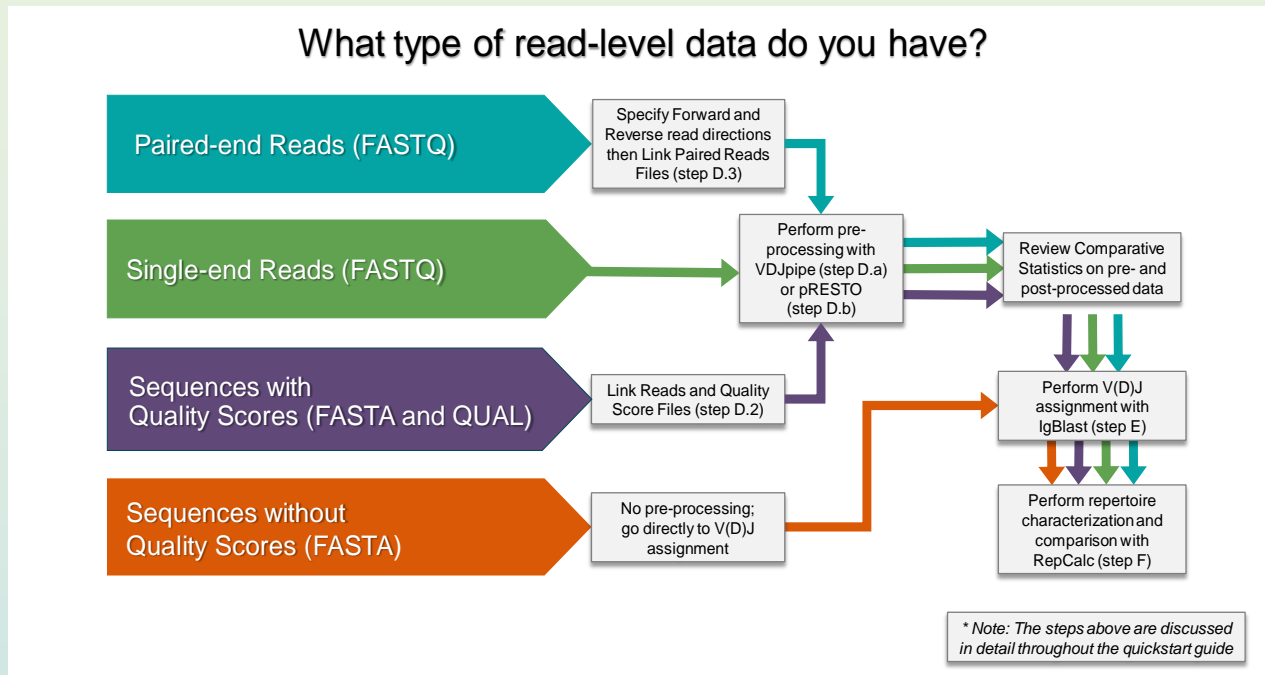
VDJServer has been developed with funding from the National Institute of Allergy and Infectious Diseases (AI097403).

TYPES OF INPUT
VDJServer can currently accept (1) single-end sequencing reads in FASTQ format, (2) paired-end sequencing reads in FASTQ format, (3) sequences in FASTA format with quality score in separate files in QUAL format, and (4) sequences in FASTA format without quality scores. The workflow schematic on page 2 shows the basic steps for an analysis workflow according to the type of input.

Raw reads obtained directly from the sequencer (read-level data types #1, #2 or #3 above) should be pre-processed, such as by merging paired-end reads, quality filtering, barcode demultiplexing, and trimming primer sequences, before the sequences are used for V(D)J assignment. Files containing primer and barcode sequences can be uploaded for use during pre-processing. Sequences in FASTA format without quality scores (read-level data type #4 above) can be directly used for V(D)J assignment without pre-processing.

Workflow Schematic:

What type of read-level data do you have?

Paired-end Reads (FASTQ)	Specify Forward and Reverse read directions then Link Paired Reads Files (step D.3)
Single-end Reads (FASTQ)	
Sequences with Quality Scores (FASTA and QUAL)	Link Reads and Quality Score Files (step D.2)
Sequences without Quality Scores (FASTA)	No pre-processing; go directly to V(D)J assignment

Perform pre-processing with VDJpipe (step D.a) or pRESTO (step D.b)

Review Comparative Statistics on pre- and post-processed data

Perform V(D)J assignment with IgBlast (step E)

Perform repertoire characterization and comparison with RepCalc (step F)

* Note: The steps above are discussed in detail throughout the quickstart guide

Key Steps

A. Creating an account:

1. Select "Create Account" under the "Login" button on the homepage and fill in as required. The user will be required to enter an email address, username, and password.
2. You will receive a confirmation email to the address you provided. Confirm, and your account is now active.

B. Setting up a new project:

1. Select "Add Project" on the left side of the page under the VDJServer logo, type in a name, and then confirm by selecting the green "Create a new VDJServer Project" button.
2. The main page now displays a project tab in blue on the far left. When you create more projects, you will be able to quickly cycle between them using these tabs. For the active project, the main page shows three buttons: a green "Upload" button, a "Run Job" button, and an inactive "File Actions" button. Text in the "Let's Get Started!" field below these indicates the types of files that can be uploaded.

B. Setting up a new project (*cont.*):

3. To upload files, drag them into the "Let's Get Started" area or select the "Upload" button and choose to upload files from your computer or from Dropbox. Make sure to select the appropriate file type in the dropdown menu, as this will determine the ways in which different files can be used by VDJServer analysis components.

C. Entering study metadata:

The "Metadata Entry" tab for the project is used to define study, subject and sample metadata. User-specified sample groups (e.g. treatment/control) can also be defined. This can be done in two ways: 1) import a pre-existing, tab delimited file that contains this information via the "Import From File" option of the "Metadata Actions" button. The file needs to be uploaded into Project Data (step B.3) before importing. Or 2) manually enter this information directly on the Metadata Entry screen. Metadata is used by RepCalc (Step F below) for repertoire comparison between samples and sample groups.

D. Preprocessing of read data for a new project:

1. If your reads have barcodes or primers, you should upload files in FASTA format that contain the sequences for them. Make sure to select the appropriate file type (such as barcode or primer).
2. If your read sequences and quality scores are in separate files (e.g. FASTA and QUAL files), you will need to associate them using the "Link .fasta/.qual Files" tab on the left.
3. If your read data is paired-end sequences in separate FASTQ files, set the appropriate read direction (forward, reverse) on each file and associate the files using the "Link Paired Reads Files" tab on the left.
4. In the "Upload and Browse Project Data" tab, select the read-level data you wish to process by selecting the box to the left of the file name. Do not select any primer or barcode files at this time, as they will be specified during job submission.
5. Select "Run Job" and then "VDJPipe" or "pRESTO" from the dropdown menu.

D (a). Preprocessing of read data for a new project using VDJPipe:

1. The default job name will have a timestamp and can be edited to be more descriptive by typing in the "Job Name" box.
2. Select the desired workflow from the dropdown menu. For VDJPipe, VDJServer offers a pre-configured complete workflow that includes all pre-processing steps. However, individual pre-processing steps can also be performed.
3. Change values for each pre-processing step, select barcode and primer files or remove those steps (by selecting the "X") if not needed. When you are satisfied with the workflow configuration, select the green "Launch Job" button.
4. The job progress and output can be viewed from the "View Analyses and Results" tab for your project. When the "View Analyses" column on the far right displays a blue "View Output" button, the job has finished running. "View Output" provides a number of interactive charts to visualize comparative statistics on your data pre- and post-processing.
5. When you are satisfied with the post-processed output data, select "Make Available in Project Data Area" for your job output files, and continue to V(D)J assignment.

D (b). Preprocessing of read data for a new project using pRESTO:

1. The default job name will have a timestamp and can be edited to be more descriptive by typing in the "Job Name" box.
2. Currently, VDJServer offers a pre-configured complete workflow that includes all pre-processing steps for pRESTO.
3. Change values for each pre-processing step, select barcode and primer files or remove those steps (by selecting the "X") if not needed. When you are satisfied with the workflow configuration, select the green "Launch Job" button.
4. The job progress and output can be viewed from the "View Analyses and Results" tab for your project. When the "View Analyses" column on the far right displays a blue "View Output" button, the job has finished running. "View Output" provides a number of interactive charts to visualize comparative statistics on your data pre- and post-processing.
5. When you are satisfied with the post-processed output data, select "Make Available in Project Data Area" for your job output files, and continue to V(D)J assignment.

E. Running the IgBlast application for V(D)J assignment of read data:

1. Select the read level data you wish to analyze by selecting the box to the left of the file name on the main page.
2. Select "Run Job" and then "IgBlast" from the dropdown menu.
3. The default job name will have a timestamp but should be edited for a more descriptive name by replacing the text inside the "Job Name" box. The "Selected Files" tab should indicate which sequences files will be analyzed. If multiple files were selected and you wish to not consider those files, you can remove them by selecting the "x" to the right of the file name.
4. For the IgBlast application, you will need to specify the organism and sequence type. After this, select the green "Launch Job" button.
5. The job progress and output can be monitored from the "View Analyses and Results" tab for the designated project. When the "View Analysis" column on the left side displays a blue "View Output" button, the job has finished running. Selecting this button will display the output files.

F. Running the RepCalc application for repertoire analysis of read data:

1. Select "Run Job" and then "RepCalc" from the dropdown menu. Do not select any files from Project Data when running RepCalc.
2. The Job Submission screen for RepCalc will check that you have Study Metadata defined which it will use for repertoire calculations and comparison.
3. The Job Submission screen for RepCalc will ask you to pick a job, which contains V(D)J assignment data, to perform analysis upon. By default, the most recent job will be automatically selected for you, but you may pick a different job from the list.
4. After the repertoire analysis is complete you may view the results by selecting the "View Analysis and Results" option from the project menu.
5. Select the "View Output" option for the job. You will see a list of job output files and analysis charts. Each of the analysis charts allows customization for the comparison and visualization of samples, sample groups, and associated files.

ADDITIONAL VDJSERVER RESOURCES

A brief introductory video for VDJServer can be viewed at:
https://www.youtube.com/watch?v=DgqKP-L75CM

For more information about pRESTO (pre-processing application) and ChangeO (an external tool available via the RepCalc application), please refer to:

https://presto.readthedocs.io/en/latest/

http://changeo.readthedocs.io/en/latest/

For queries and correspondence, email us at vdjserver@utsouthwestern.edu

For online discussions with the VDJServer community join us at:

http://forums.vdjserver.org/

Check out our wiki page at:

http://wiki.vdjserver.org

Like us on Facebook:

VDJServer | Facebook