YAQAAT webserver User Guide

YAQAAT automates the entire process of assembling Sanger reads into contigs from pairing forward and reverse segments of the AB1/FASTQ files (based on the filename), trimming (optional), de-novo/template guided assembly and alignment to the template sequence in order to scan for mutant bases (if a template is given. A video tutorial illustrating the entire process can also be accessed at (**Link to youtube**).

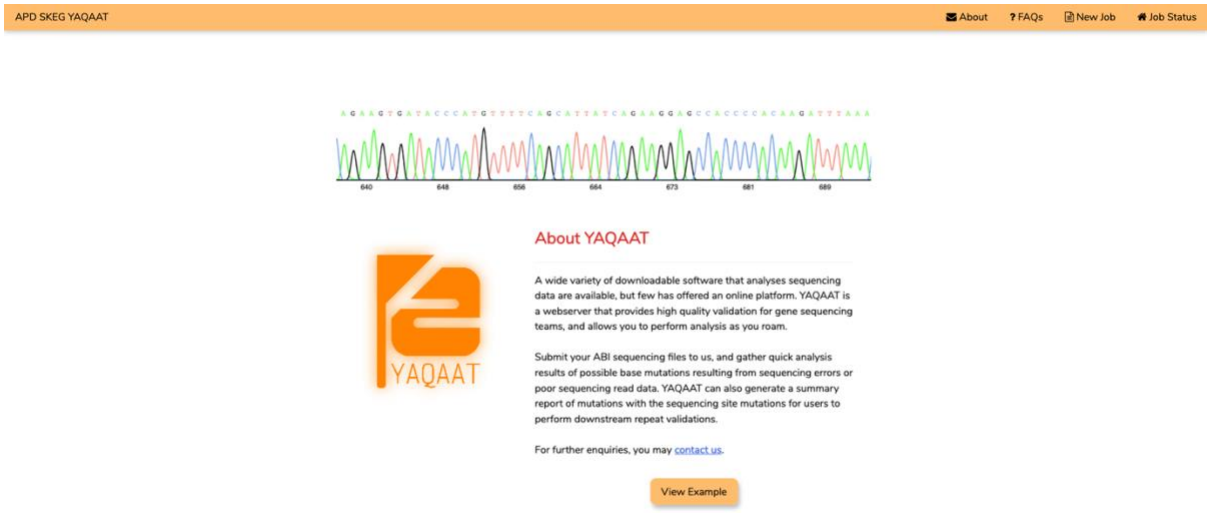

Figure 1: Main interface of YAQAAT

When the user enters the webserver, he / she will be greeted at the man landing page (Figure 1). “New Jobs” can be submitted at the top right and any ongoing / completed jobs are viewed under “Job Status”. To view an example of analysis by YAQAAT, click on “View Example”

Job Submission

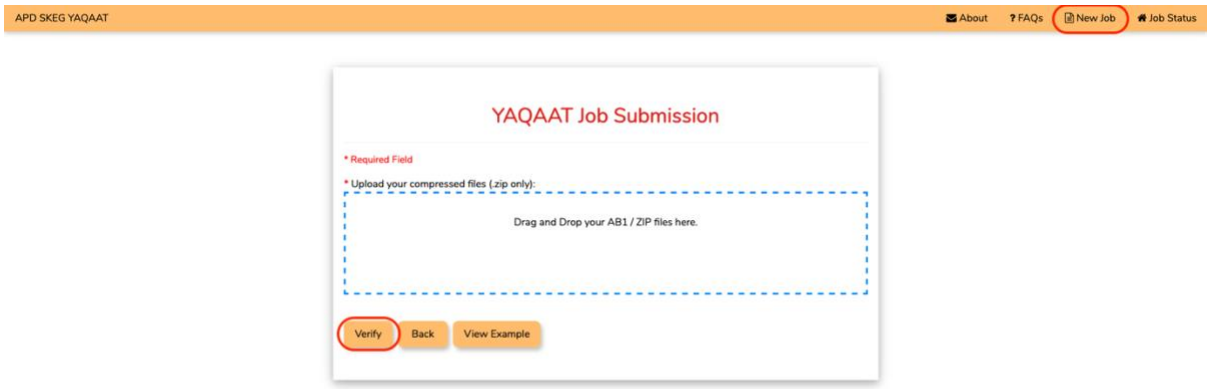

Figure 2: New job submission

To submit a job, click on “New Job” (Figure 2). The user will then be requested to upload the AB1 / FASTQ files to be analyzed. Click “Verify” to proceed.

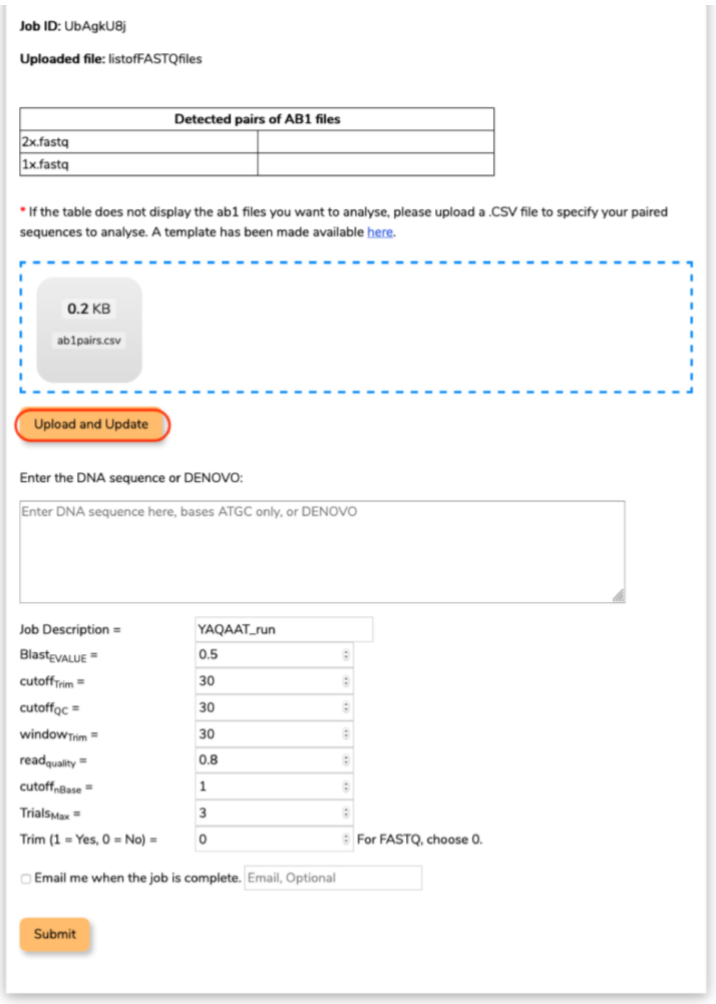

Figure 3: New Job submission

The program will check for AB1 pairs included in the uploaded files. In cases when no pairs are detected, the user is able to upload specify paired sequences by filling in the CSV template, followed by clicking on "Upload and Update" (Figure 3).

The Basic information fields for 'Job Description', '$Blast_{EVALUE}$', '$cutoff_{Trim}$', '$cutoff_{QC}$', '$window_{Trim}$', '$read_{quality}$', '$cutoff_{nBase}$', '$Trials_{Max}$' and '$Trim$' are available for adjustment, else values are set to default value as specified. To get notification when the jobs are completed, users may check the "Email me when job is complete" option. The job will be submitted upon clicking on "Submit" (Figure 3). To give an estimate of the analysis time, a pair of sequence takes less than one minute. The analysis time required is dependent on the number of sequence pairs to be analyzed and sequence length.

Viewing of results

Job status of the submission will be available after the user has returned to the home page (Figure 4).

Number of jobs in queue: 6

JobID	Status	Valid Till	Public IP ADDRESS
UbAgkU8j	Completed	5/4/2020	103.37.196.xxx
hxH5ADYM	Completed	5/4/2020	123.136.68.xxx
EBmpW2MI	Completed	5/4/2020	123.136.68.xxx
czelRYsG	Completed	5/4/2020	123.136.68.xxx
R9y359xN	Completed	2/4/2020	123.136.68.xxx
Yf27Ik6A	Completed	2/4/2020	103.37.196.xxx

Figure 4: Job status

To view results, click on "JobID" which has the status as "Completed". Due to constraints in computing resources, the completed job results will be valid for a set duration of time, after which the data would be removed and viewing of results will be unavailable.

When the user clicks on the JobID, they will be brought to the results page where they can select on the ab1 files for analyze (Figure 5). For pairs where reads are too poor in quality and skipped, the reason for skipping will be displayed. Downloading of the summary or full details of the analysis are available by clicking on “Download Summary” or “Download Full details” respectively.

Figure 5: Selection of ab1 files

By clicking “Analyze”, The results of the selected ab1 files are shown (Figure 6).

Figure 6: Result analysis output

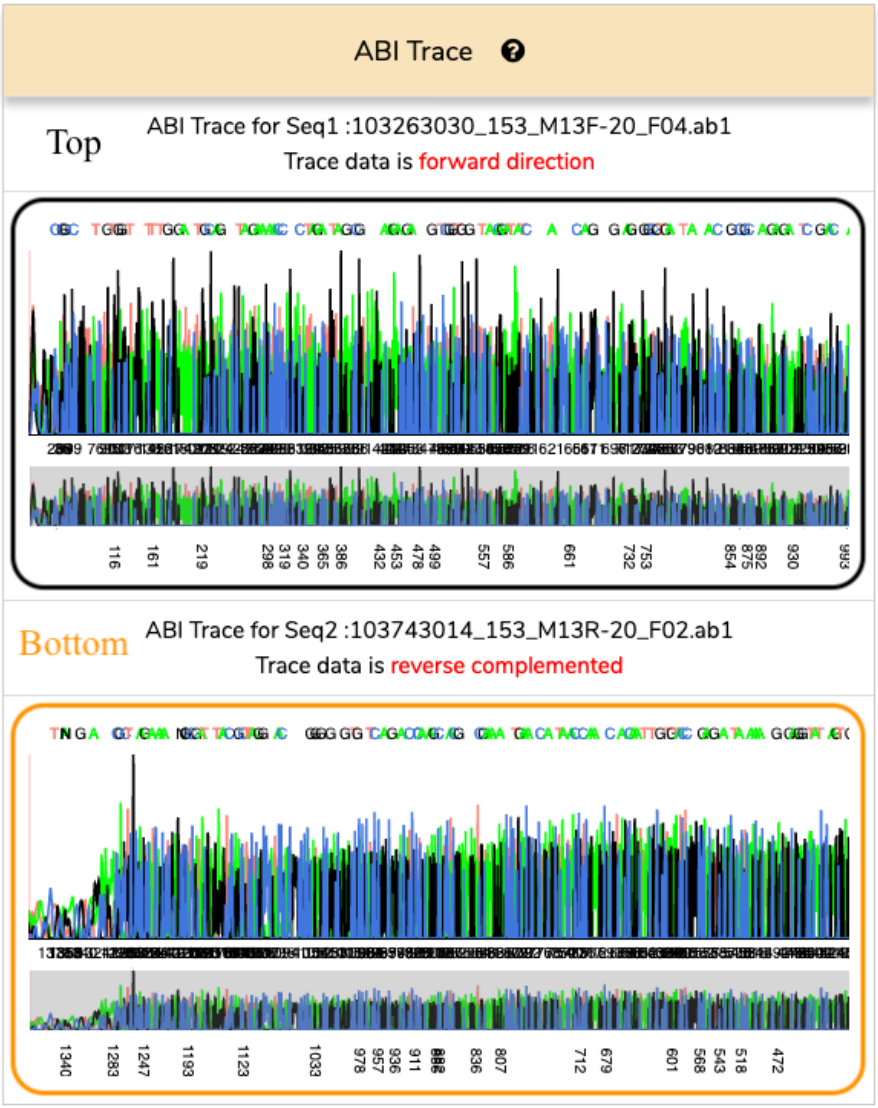

Figure 7: ABI Trace Panel, Top: Graph of ABI Trace, Bottom: Graph of ABI Trace in opposite direction

The ABI Trace Panel (Figure 7) includes the AB1 trace for the forward and reverse sequence pair chromatograms. To zoom in, click outside the shorter graph selection window (grey rectangle) to reset view. To view a specific region, drag the crosshair mouse on the smaller graph outside the current selection. Alternatively, scroll on the larger graph plot to zoom in and out.

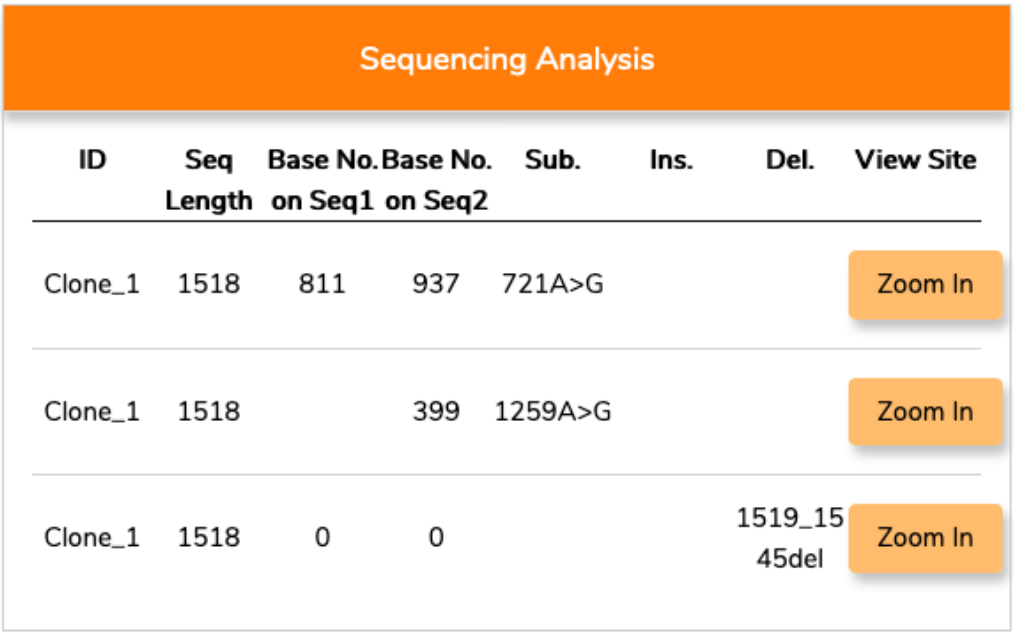

Figure 8: Sequencing Analysis Panel

The Sequencing Analysis Panel shows the SNPs / INDELS detected by YAQAAT. To view specific clones, click on "Zoom In".

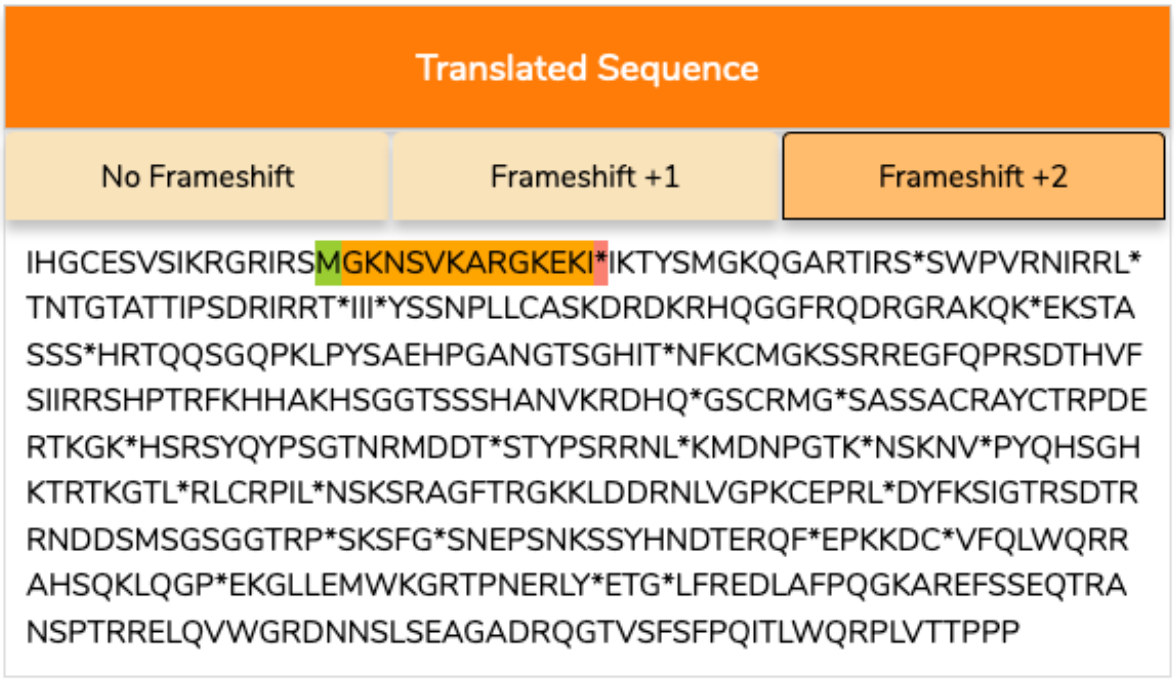

Figure 9: Translated Sequence Panel

In the Translated Sequence Panel, the trimmed consensus sequence produced by YAQAAT is translated based on human start and stop codons, "ATG" and "TAG, TAA, TGA" respectively.

“M” is highlighted in green for the start amino acid while red “ * ” is for the stop signal. By clicking on the frameshifts button, users are able to see the different sequences viewed in under different frame windows.

For FASTQ results, raw FASTQ sequences are displayed instead of the chromatograms. (Figure 10)

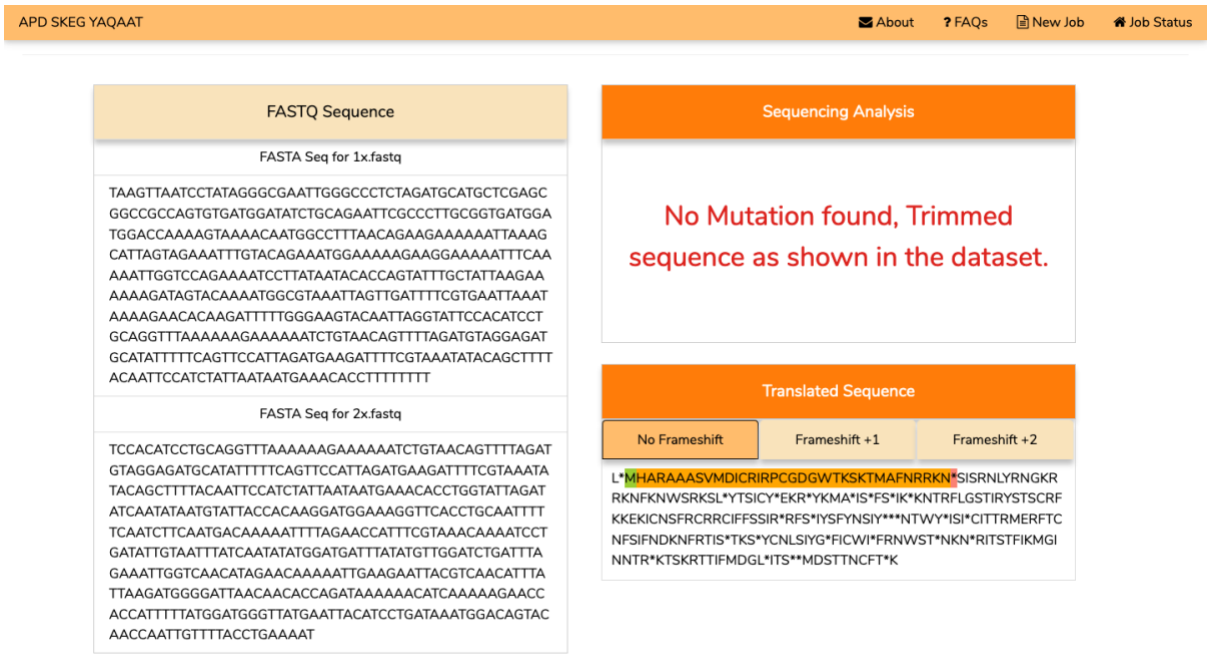

Figure 10: FASTQ result analysis output

YAQAAT List of Parameters

Supplementary Table 1 list of tweak-able parameters for trimming, evaluation of read quality and alignment of reads to the template.

Parameter	Description	Adjustment
$Trim_{on/off}$	Trim sequences based on Phred quality.	Default = Off, to avoid unnecessarily truncation of assembled sequences. May be toggled on when sequencing files have noisy reads at the ends.
$sliding - window_{phred}$	length of sliding window used to calculate instantaneous gradient of Phred values in order to determine window of regions at the ends of the reads which are of poor quality to be trimmed out from evaluation of the overall sequence quality of assemble reads.	Default = 30, To increase for additional trimming or accuracy of the quality evaluation procedure.
$cutoff_{Trim}$, $cutoff_{QC}$	Cutoff used to evaluate the quality and to trim AB1/FASTQ reads. A base Phred value of 20 and 40 corresponds to a base call accuracy of 99% and 99.99%, respectively.	Default = 30, This may be increased to increase the stringency of the quality evaluation procedure or trimming or reduced to allow poor quality reads to be assembled or

		over-trimming of relevant sequence regions which may consist of bases with low Phred quality values
$read_{quality}$	A high read_quality_score indicates good Sanger sequencing reads with little noise in the middle of reads where quality is not typically expected to be poor.	Default = 0.8 , increase to permit assembly of sequences with noisier reads and vice versa.
$cutoff_{nBase}$	Bases below this cutoff are assigned the base "n"	Default = 1 , increase to assign bases below the cutoff as "n"
$Blast_{EVALUE}$	Used to align contigs to the template or to each other	Default = 0.5 this may be increased if contigs cannot be aligned or if they share too little homologous regions.
$Trials_{Max}$	Number of times the sliding window, Trimming_Phred_window, Quality and Quality_evaluation_Phred_window are extended by a multiple of 1.	Default = 3, reduce to increase processing speed and increase to possibly increase the degree of trimming low quality bases.