USER MANUAL FOR

QTLIT

(Quantitative Trait Loci Identifcation Tool)

version 1.0

Abhijeet Roy, UB Angadi, Abhishek Mazumder, Subhadip Das, Tapan Kumar Mondal

Disclaimer: The tool has been tested on various datasets, at ICAR - NIPB. However, the results generated for any specific dataset are not gauranteed. We recommend to cross check with other available tools for reliability.

for standalone download: https://github.com/TKM-Lab/QTLIT

Packages utilized in this tool:

rtracklayer for import/export fasta files,
ShortRead for quality check of raw fastq files,
gmapR for reference genome indexing and short reads alignment,
Rsamtools for indexing binary alignment (bam) file,
vcfR for vcf plotting,
Biostrings for manipulation of large sequences,
stringr for pattern matching,
rvest for manipulation of HTML file,
ggplot2 for enhanced plotting,
rTASSEL for vcf filteration,
ASmap for linkage map construction,
LinkageMapView for linkage map generation,
qtl for QTL analysis ,
gridExtra for placing multiple plots in a grid,
parallel for parallelization of the process,
qtlhot for phenotype input and
ggrepel for labelling of the significant markers.

All the packages are tested on RStudio (v 2023.12.1+402).
Moreover, a custom R script was also incorporated for the conversion
of vcf file to R/qtl cross object
(https://github.com/RimGubaev/vcf_to_qtl/blob/master/vcf_to_qtl_converter.R) in the shiny app scripting.
Apart from the R packages, bcftools (v 1.13) was utilized for the variant
calling utilizing system ( ) command in the R environment for the Shiny server.

References

Broman,K.W., Wu,H., Sen,Ś. and Churchill,G.A. (2003) R/qtl: QTL mapping in
experimental crosses. Bioinformatics, 19, 889–890.
Monier,B., Casstevens,T.M., Bradbury,P.J. and Buckler,E.S. (2022) rTASSEL: An R interface to TASSEL for analyzing genomic diversity. J. Open Source Softw., 7, 4530.
Lawrence,M., Gentleman,R. and Carey,V. (2009) rtracklayer: an R package for interfacing with genome browsers. Bioinformatics, 25, 1841–1842.

Morgan,M., Anders,S., Lawrence,M., Aboyoun,P., Pagès,H. and Gentleman,R. (2009) ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics, 25, 2607–2608.
Wu,T.D., Reeder,J., Lawrence,M., Becker,G. and Brauer,M.J. (2016) GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. In Mathé,E., Davis,S. (eds), Statistical Genomics: Methods and Protocols. Springer, New York, NY, pp. 283–334.
Morgan,M., Pagès,H., Obenchain,V. and Hayden,N. (2024) Rsamtools: Binary
alignment (BAM), FASTA, variant call (BCF), and tabix file import.
Knaus,B.J. and Grünwald,N.J. (2017) vcfr: a package to manipulate and visualize variant call format data in R. Mol. Ecol. Resour., 17, 44–53.
Pagès,H., Aboyoun,P., Gentleman,R. and DebRoy,S. (2024) Biostrings: Efficient manipulation of biological strings.
Taylor,J. and Butler,D. (2017) R Package ASMap: Efficient Genetic Linkage Map Construction and Diagnosis. J. Stat. Softw., 79, 1–29.
Ouellette,L.A., Reid,R.W., Blanchard,S.G. and Brouwer,C.R. (2018) LinkageMapView—rendering high-resolution linkage and QTL maps. Bioinformatics,34, 306–307.
Auguie,B. (2017) gridExtra: Miscellaneous Functions for ‘Grid’ Graphics.
Neto,E.C. and Yandell,B.S. (2018) qtlhot: Interface for QTL Hotspots.
Slowikowski,K. (2024) ggrepel: Automatically Position Non-Overlapping Text Labels with ‘ggplot2’.
Danecek,P., Bonfield,J.K., Liddle,J., Marshall,J., Ohan,V., Pollard,M.O., Whitwham,A.,Keane,T., McCarthy,S.A., Davies,R.M., et al. (2021) Twelve years of SAMtools and BCFtools. GigaScience, 10, giab008.

INTRODUCTION

1.1 About QTLIT

Quantitative Trait Loci Identification Tool (QTLIT) , is a tool which takes raw fastq format files, either in-house or commercially generated and proceed to generate alignment files in binary alignment map (BAM) format files, variant called (VCF) files.The VCF file is further utilized for linkage mapping and QTL analysis. All at once, with the user having a choice of proceeding at any step or to utilize any of the files generated for the downstream analysis. Pipeline folows in the figure 1.

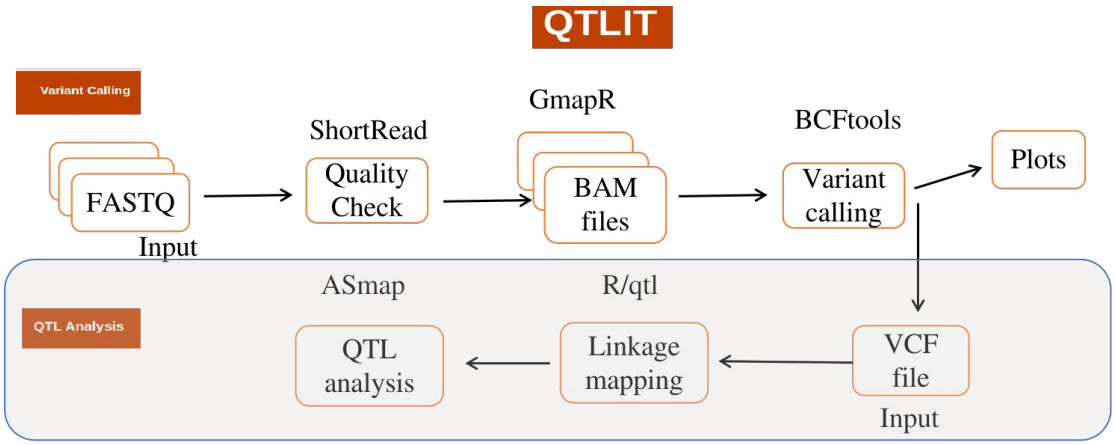


Figure 1. Pipeline of the QTLIT tool from variant calling to QTL analysis and the packages used in delveloping this tool.

1.2 Datasets

1.2.1 Fastq files

Raw Genotype By Sequencing (GBS) or Whole Genome Sequencing (WGS) data with their quality score is needed to proceed for sequence alignment. The subsequent files will get generated thereafter i.e., BAM and VCF files for further linkage mapping and QTL analysis.

Note: The file renaming should be not start with numeric, for eg., 1-23.fastq but could be alpha numeric or alaphabetic.

1.3 User Interface

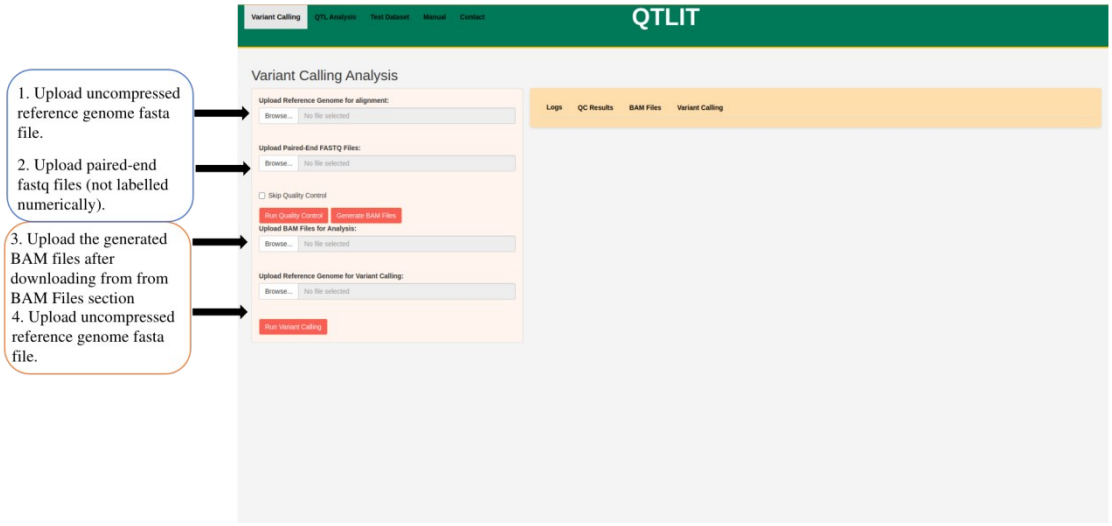


Figure 2. User interface for short reads alignment and variant calling.

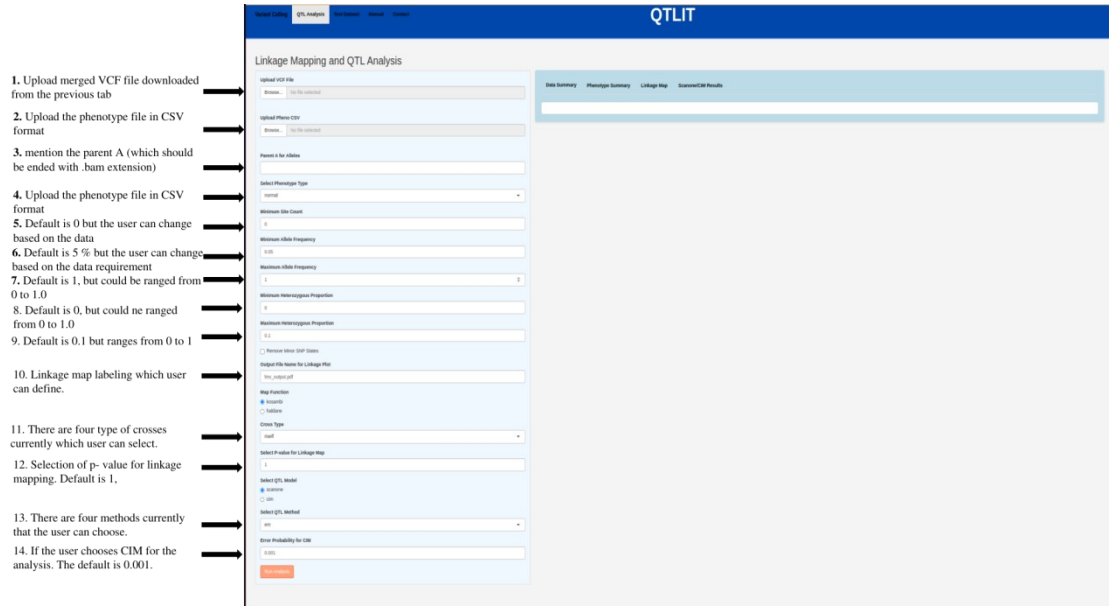
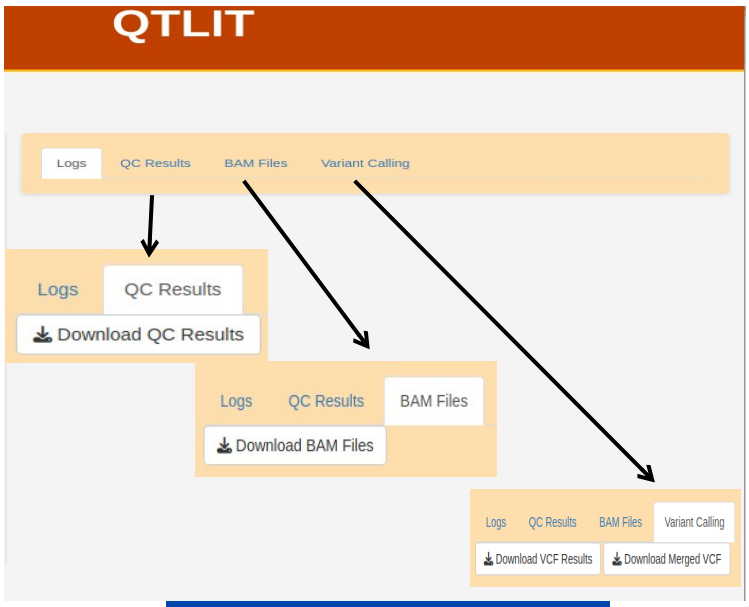


Figure 3. User interface for linkage mapping and QTL analysis page.

2. Results

Figure 4 depicts the overview of the results which can be downloaded by the users.

a.



b.

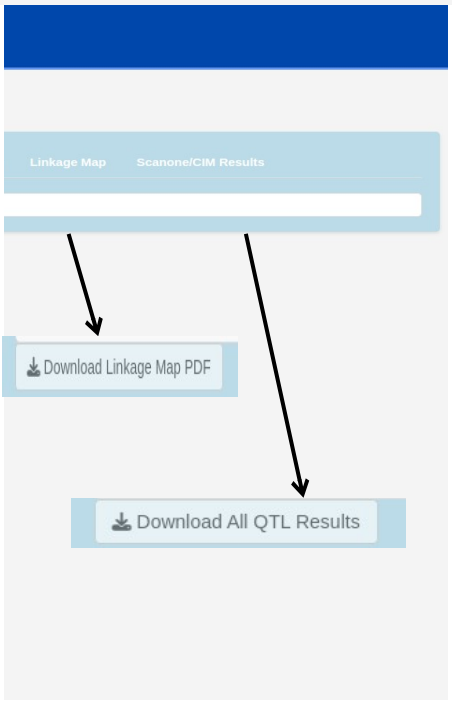


Figure 4. User interface of QTLIT tool (a) Variant calling analysis with side bar for uploading reference genome, bam files and analysis running button and each of the analysis having a dedicated tabs for downloading results (b) QTL analysis console with side bar included in inlay depicting different selection parameters.

2.1 Variant Calling

The files getting generated in this part are i. QC report ii. Variant calling plots (Figure 5). iii. BAM files in compressed format which can be downloaded from the respective section and iv. vcf file.

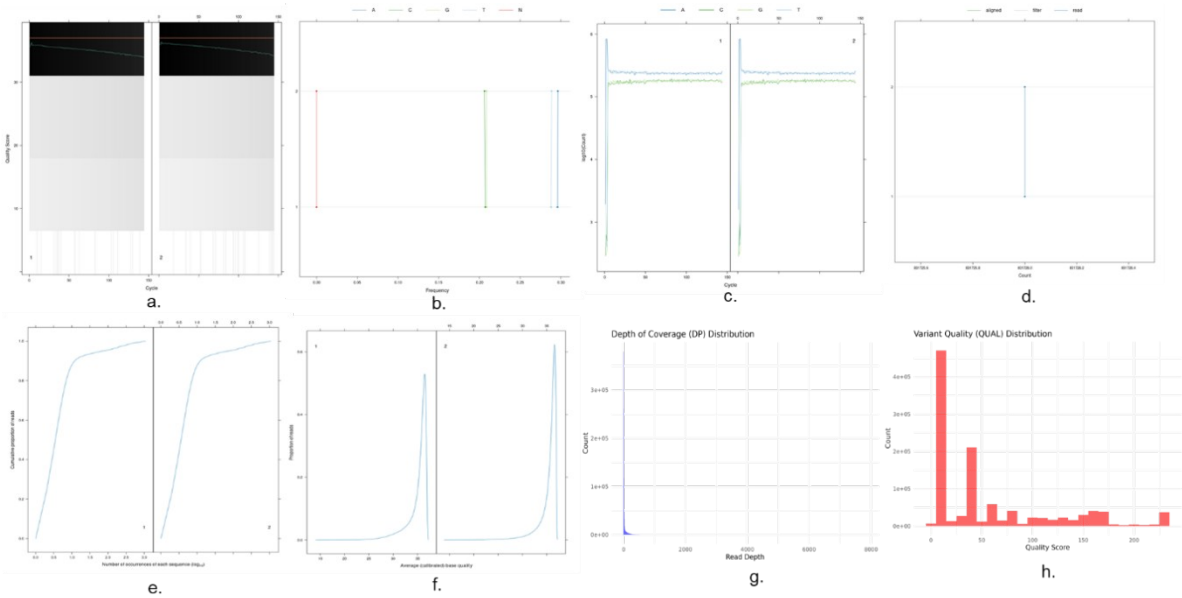


Figure 5. Various plots generated in the VCF calling pipeline (a) Quality score per cycle (b) Frequency of bases (c) Base counts per cycle (d) Average number of reads (e) Cumulative proportion of reads and the number of sequences (f) Proportion of reads with average base quality (g) Depth of coverage in the vcf file (h) Count wise variant quality score.

2.2 QTL analysis

In this section, the user will find a real time plots for the phenotypic data (figure 6) and the linkage map with QTL analysis results (figure 7). Also the user will have the the csv files of the QTL data as mentioned in the figure 4.These results can be viewed in real time or can be saved as pdf while downloading.

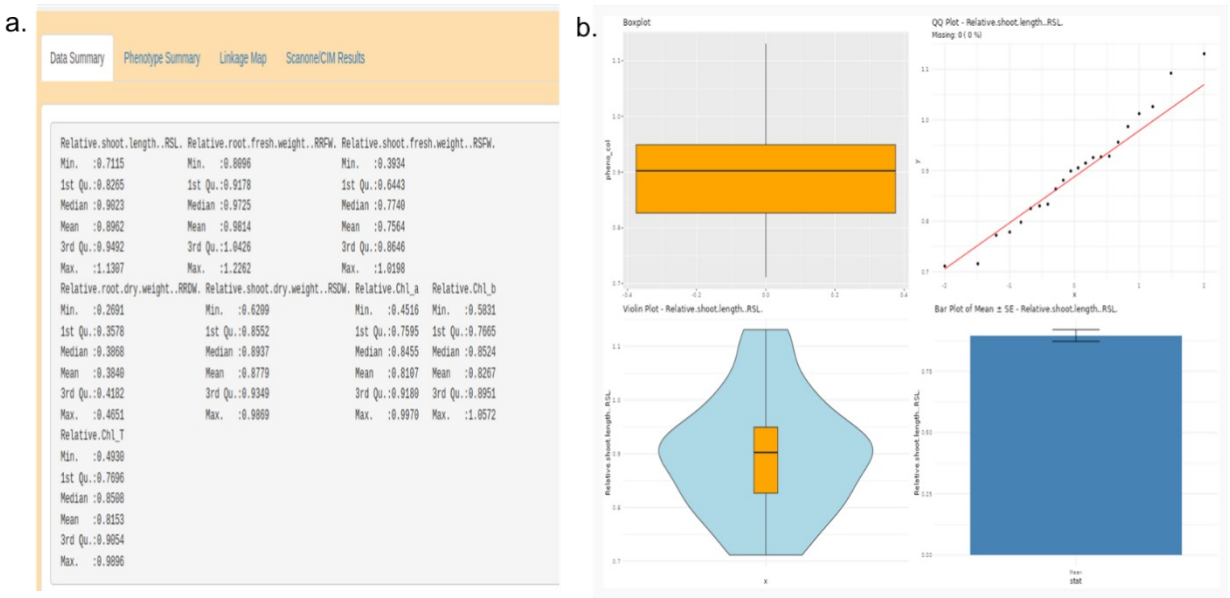


Figure 6. (a) Phenotypic data summary off all the phenotypes provided by the user (b) Box plot, Q-Q plot, Violin plot and mean with SD of the selected phenotype. (SD: Standard deviation, Q-Q: Quantile-Quantile)

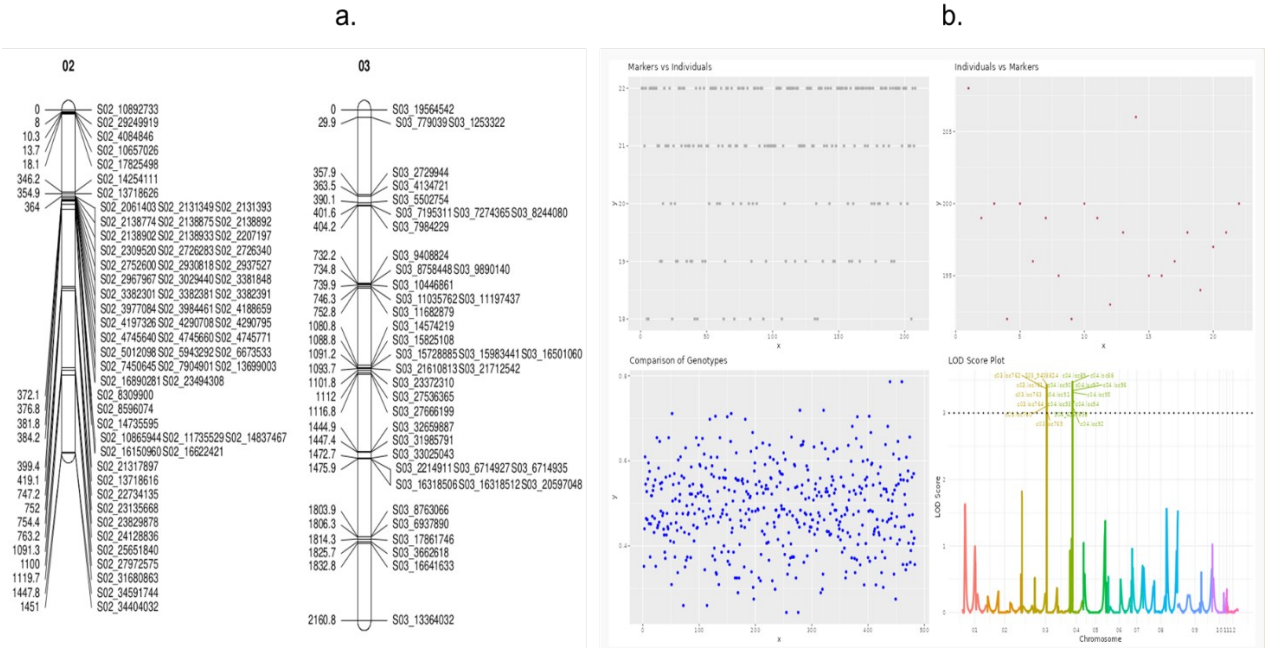


Figure 7. (a) A snapshot of the Linkage map PDF generated in the analysis (b) Various plot of QTL analysis including markers and individuals, genotype comparison and LOD plot with significant markers labelled.