

Figure 4. 1: A HTS raw sequence read showing the location of a key sequence, Multiple Identifier (MID), forward or reverses primer and a sequence of targeted genomic region. A key sequence is a sequence of four base in any order (TGCA) that is used by Roche/454 high throughput sequencing platform for calibrating the number of base call while sequencing. The MID sequence identifies the sample from where the sequences derived, the primers (forward or reverse) identifies the genomic amplicon region that was resequenced. The sequence is the actual sequence of the amplicon region.

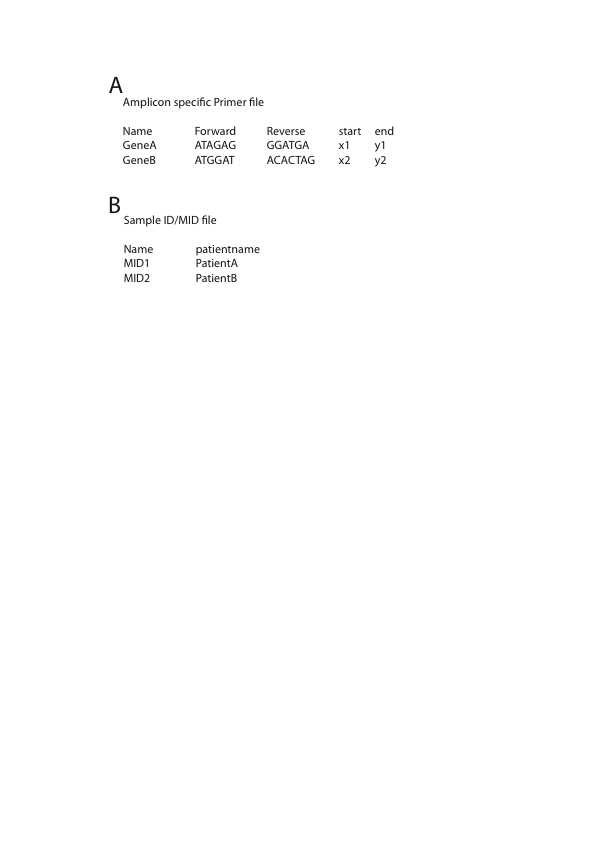


Figure 4. 2: A) Tab delimited primer file containing five columns – name of amplicon, forward and primers used for resequencing the amplicon and start and end positions of the amplicons set by first nucleotide position of forward primer and last nucleotide position of reverse primer relative to a standard HIV *pol* reference sequence. B) Tab delimited two column MID file containing the name of Roche/454 standard MID in first column and the patient name or the sample name in the second column. In an instance, when no Roche/454 MID is used, the actual nucleotide sequence used as MID can be supplied in the first column.

Note: x1, x2, y1, y2 can be replaced with the actual numbers corresponding to reference sequence

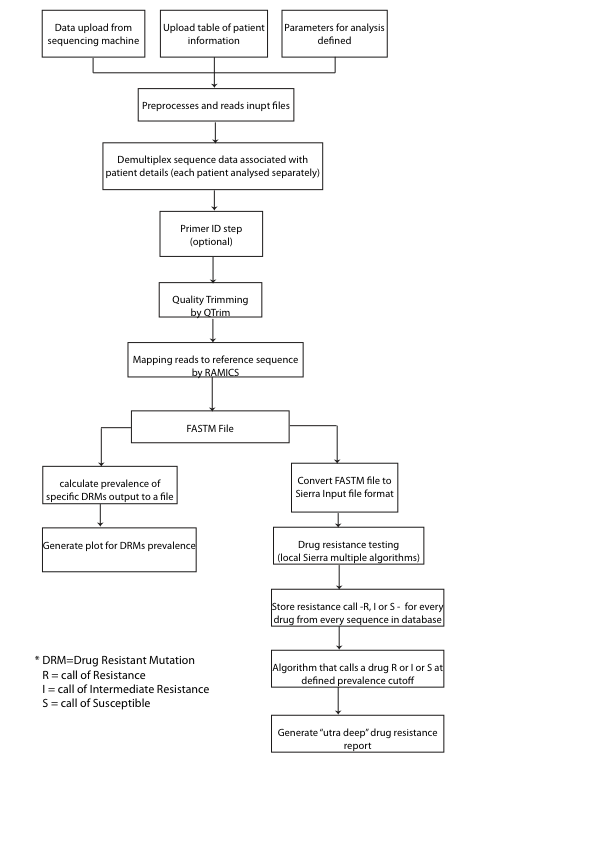


Figure 4. 3: Seq2Res pipeline workflow

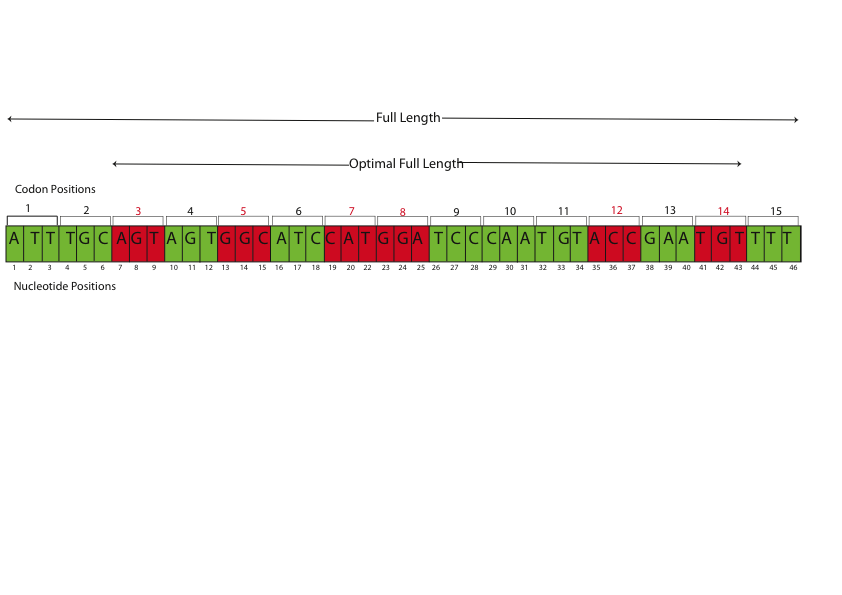


Figure 4. 4: An example of an amplicon showing optimal full length and full length. The amplicon is 46 nucleotides in length starting from 1 to 46 and has 15 codon positions (each three nucleotide compose one codon). The green codons are non drug resistant codon positions and the red codon positions are drug resistant codon positions. A Seq2Res user has to input start and end position in nucleotide numbering as 1 and 46 respectively for the amplicon. Seq2Res processes that the start and end nucleotide positions fall on codon position 1 and codon position 15 respectively, which is the full length. For the known drug resistant codon positions (in this example: 3, 5, 7, 8, 12, 14), Seq2Res processes the start and end codon positions to get the first and last drug resistant codon positions that is 3 and 14 as the new start and end positions. Positions 3 to 14 define the optimal full length.

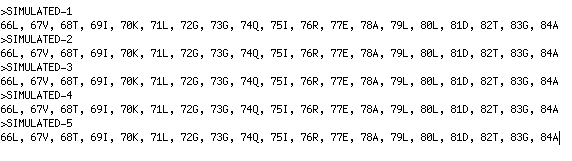


Figure 4. 5: A FASTA format like FASTM file. The FASTM format begins with sequence ID followed by list of codon positions that it covers with respected to the reference sequence and the single letter denotation of an amino acid at the codon positions.

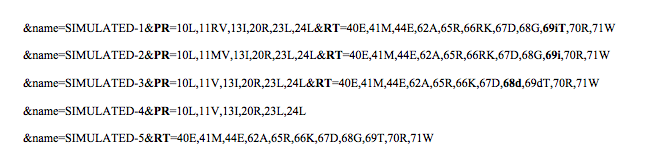


Figure 4. 6: Seq2Res generated file after translation codon positions of sequence reads from FASTM file. Each line in the file contains all information of a sequence like the complete sequence ID as the name, codon positions and their amino acids including insertion and deletions from PR and RT amplicons.

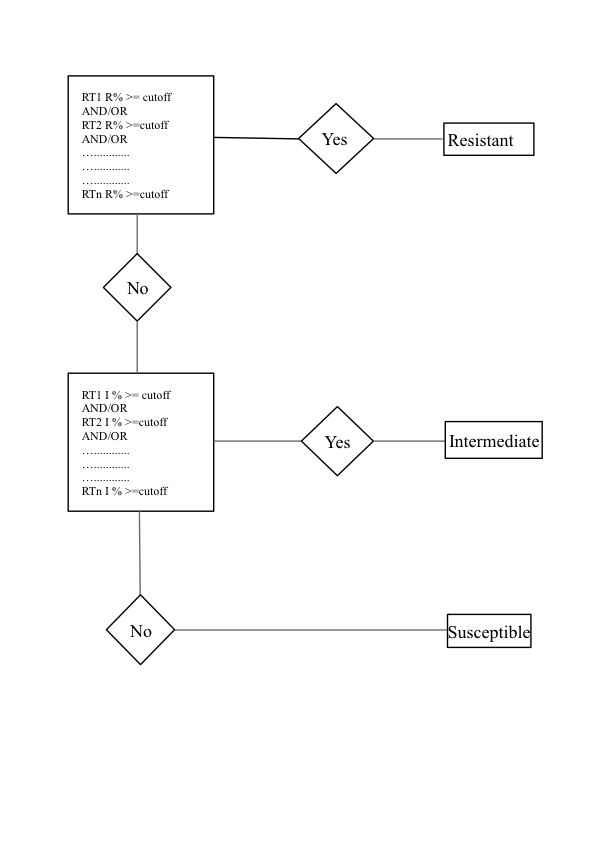


Figure 4. 7: Conditions applied in Seq2Res for drug susceptibility calls for a viral population when multiple amplicons are sequenced from a gene usually Reverse Transcriptase (RT). RT1, RT2, …. RTn are the amplicons from the gene RT. R% indicates the percentage of sequences predicted as resistant for a drug. I% indicates the percentage of sequence predicted as Intermediate Resistant for the same drug. The viral population is called as susceptible, if it is predicted neither resistant nor intermediate resistant.

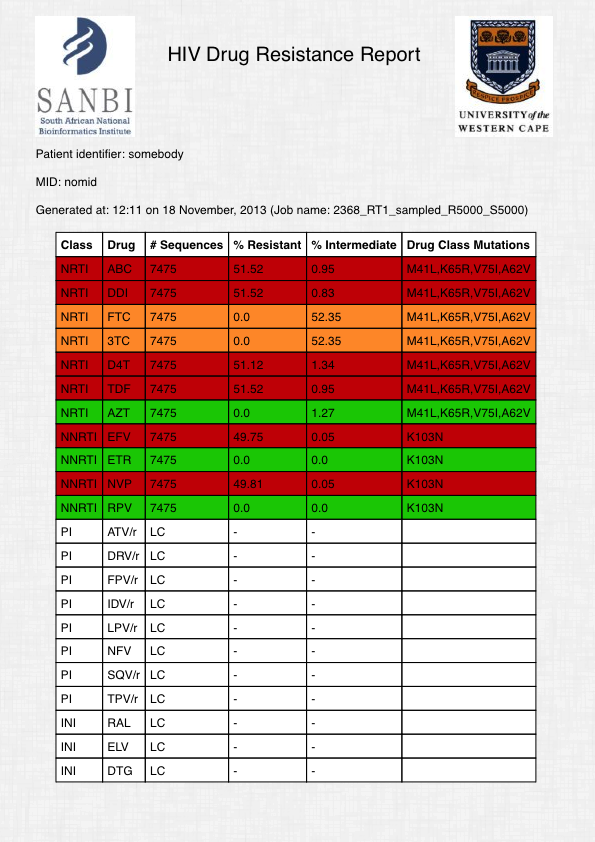


Figure 4. 8: An example of a drug resistance report. The columns from left to right in the report shows the drug class, the drug in the drug class, the number of sequence reads showing resistance to the associated drug, the percentage of sequence reads showing high level resistance to the associated drug, the percentage of sequence reads showing intermediate resistance to the associated drug and the drug resistant mutations in the observed sequence reads shows the resistance to the associated drug.



Figure 4. 9: QTrim analysis on number of sequence reads and read mean in both untrimmed and trimmed data in Seq2Res.

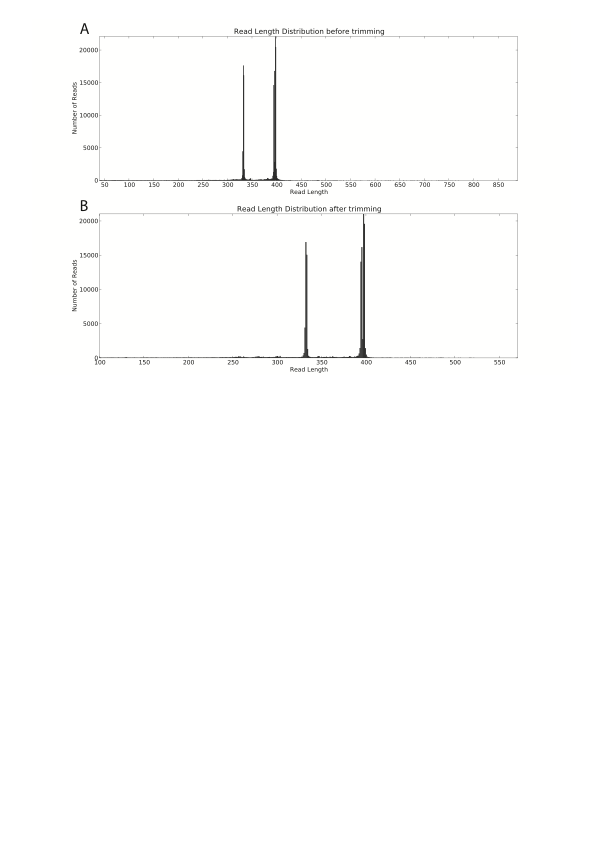


Figure 4. 10: QTrim analysis showing number of sequence reads with same read length in both untrimmed and trimmed data in Seq2Res.

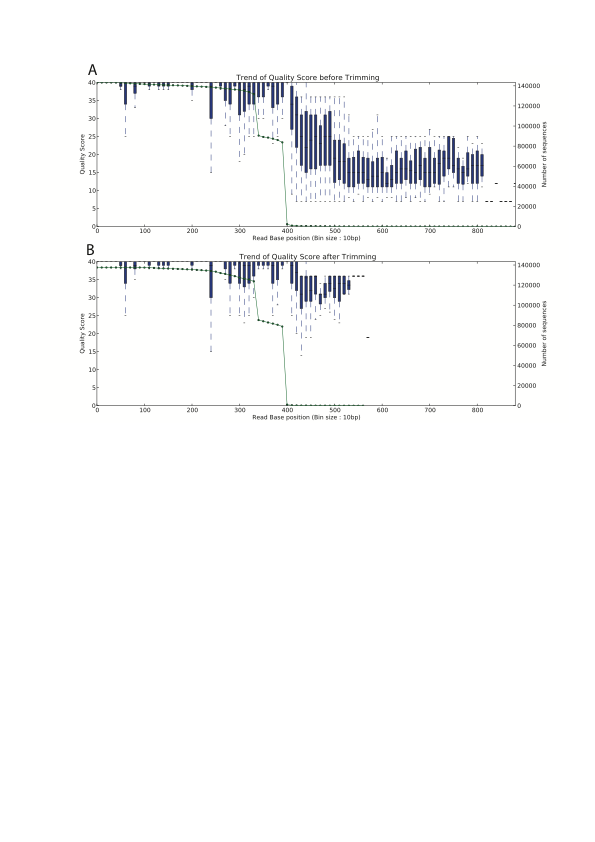


Figure 4. 11: QTrim analysis showing the trend of average quality score with increase in the read length of sequence reads at every 10 base pairs. The green dotted line shows the number of sequences that contribute in the average quality score.

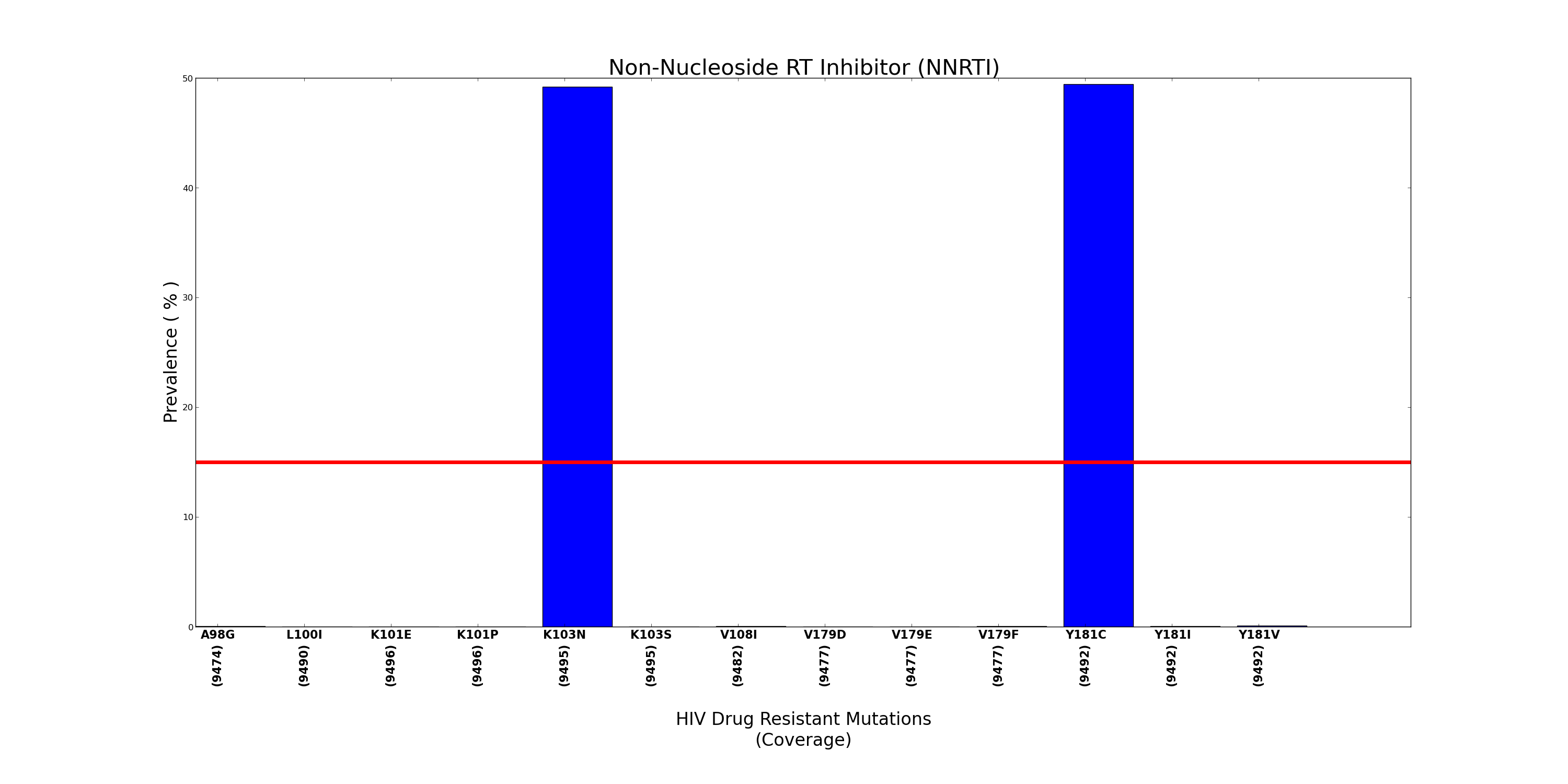


Figure 4. 12: An example of DRM prevalence plot in Seq2Res. The plot shows the prevalence of K103N and Y181C DRMs with prevalence of 49.50% and 49.85% respectively. The red line represents the user defined prevalence cutoff.

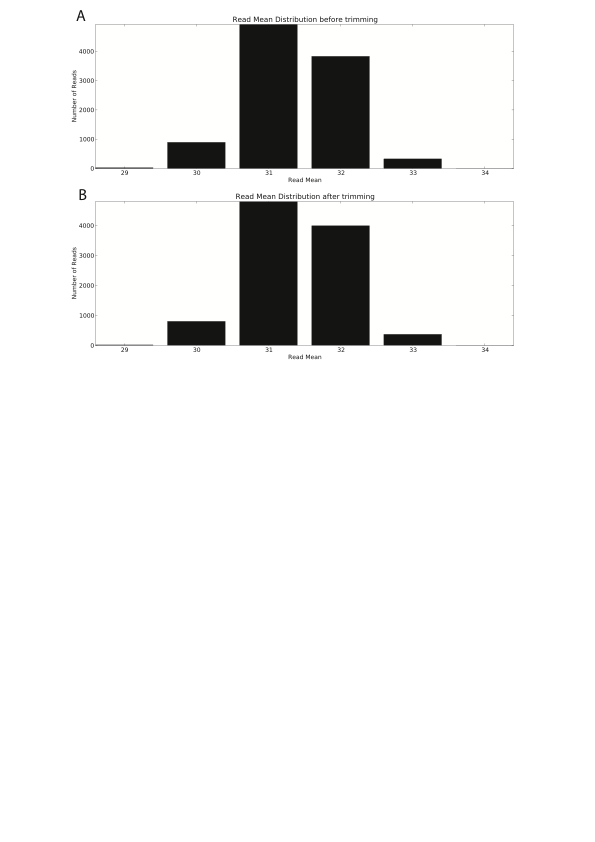


Figure 4. 13: Distribution of sequence reads by mean quality score in (A) untrimmed and (B) trimmed simulated data

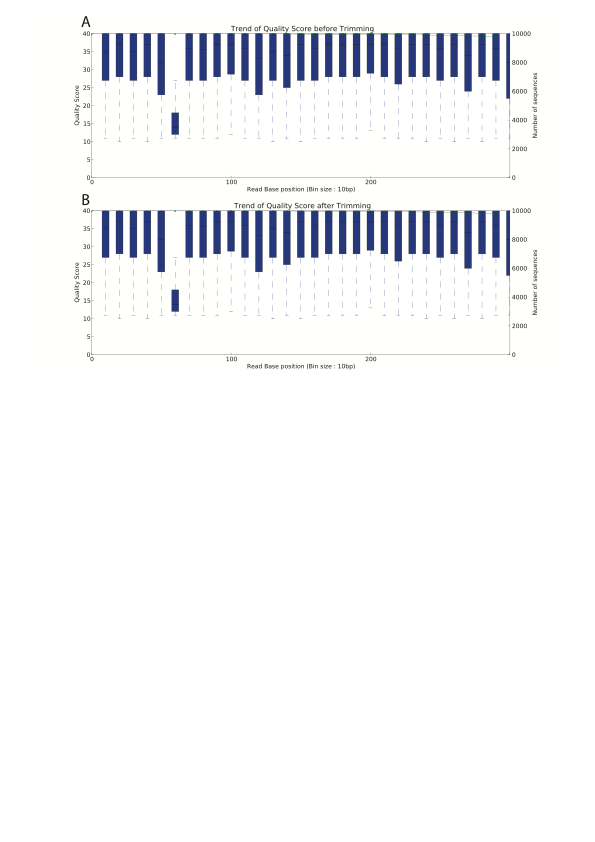


Figure 4. 14: Trend of average quality score at every 10th base pair across sequence reads in (A) an untrimmed and (B) a trimmed simulated data. The median quality score was increasing and decreasing across the sequence reads in both A and B.

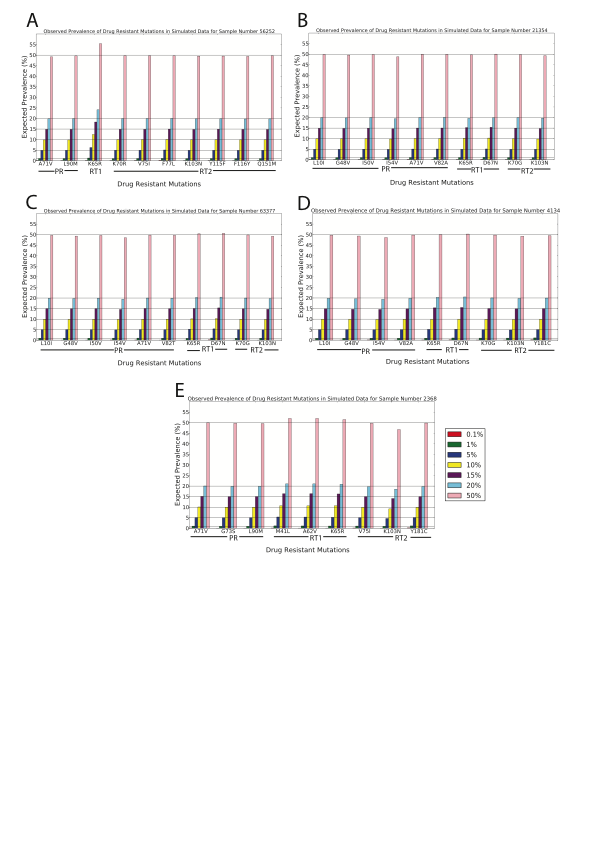


Figure 4. 15: Observed prevalence of the drug resistance mutations (DRMs) in the simulated data of samples. The horizontal lines show the expected prevalence of the DRMs while the colored bars show the observed prevalence.

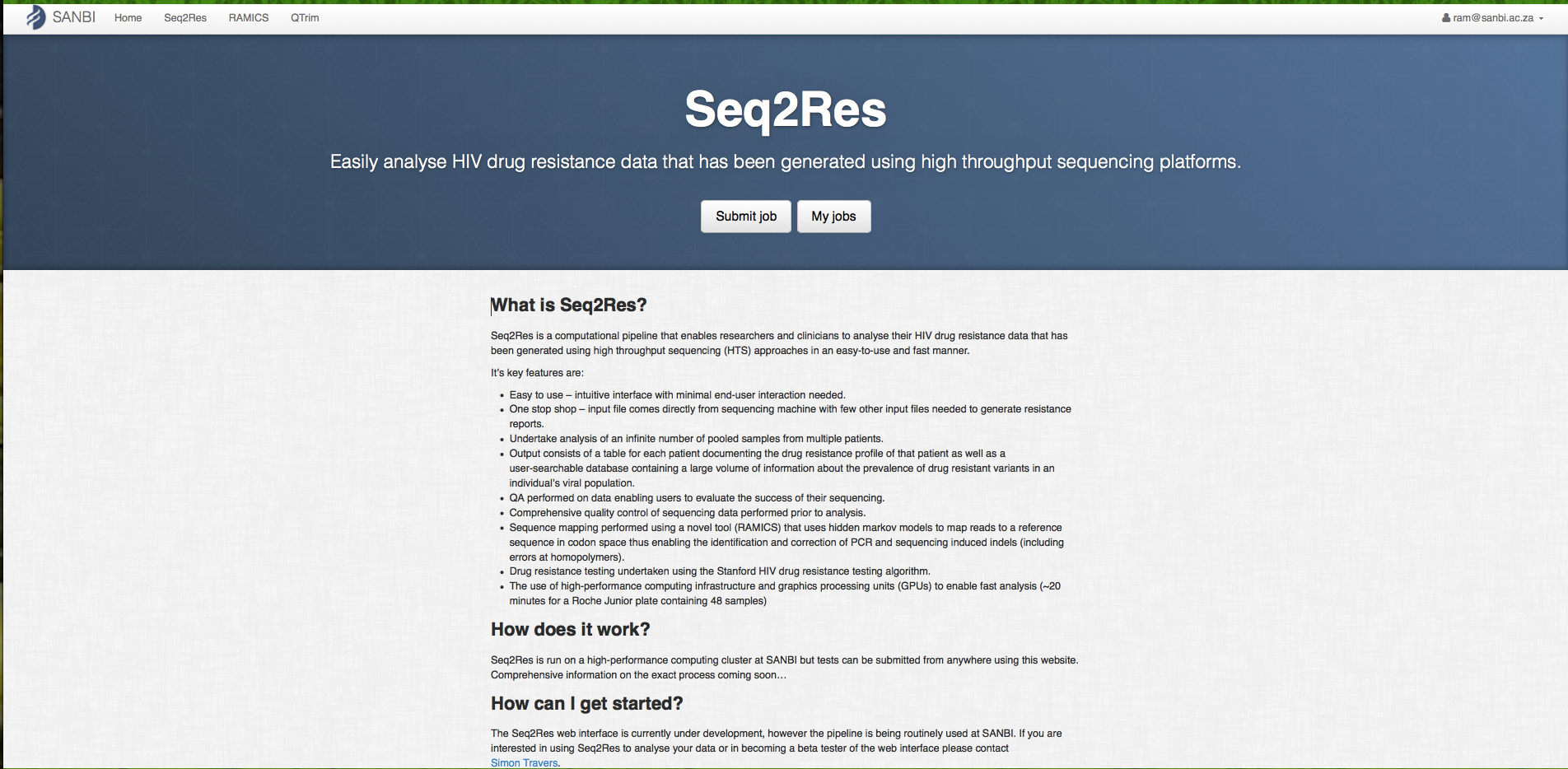


Figure 4. 16: Seq2Res homepage. Users are required to register before Seq2Res access. New users have to contact Prof Simon Travers at [simon@sanbi.ac.za](mailto:simon@sanbi.ac.za) for registration. Academic users are registered for free whereas business users have to purchase license. Once registered, users get a login ID and password, which can be used to log in to get Seq2Res access. With “Submit Job” button, users can submit new job and with “My Jobs” button, users can view previously submitted jobs that are both completed or in process.

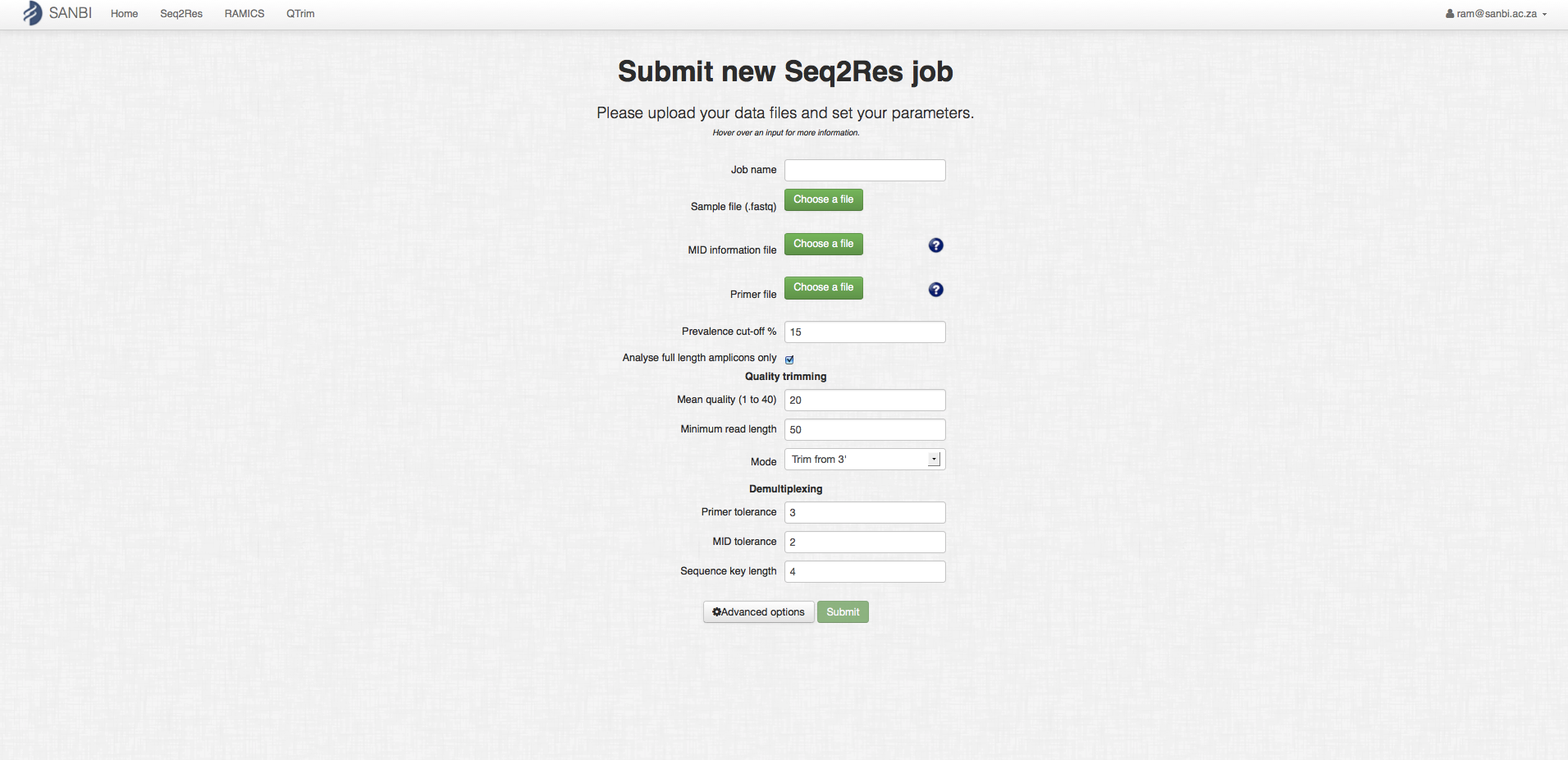


Figure 4. 17: Seq2Res job submission homepage. Users can input a raw sequence file, primer file containing primers for the raw sequence file and sample specific MID file by exploring the files in their computer file system. Users can click on help button (with symbol ?) for any confusion in the file format of the primer and MID files. In advance options, users can change the analysis parameters. Users can hover the mouse cursor point on the fill up area to get help on the parameters. The “submit” button will light up after entering all the required fields. Users can then click the “submit” button to submit their job. Users will automatically get an email if Seq2Res outputs error while processing or the job is processed successfully.



Figure 4. 18: Seq2Res page for viewing user submitted list of jobs. The page shows job specific details like name of the job, the date when the job was submitted, the status of the job either completed process or in pending and an option to delete the job. Users can sort the jobs by name and by date, clicking at “name” and “date” respectively

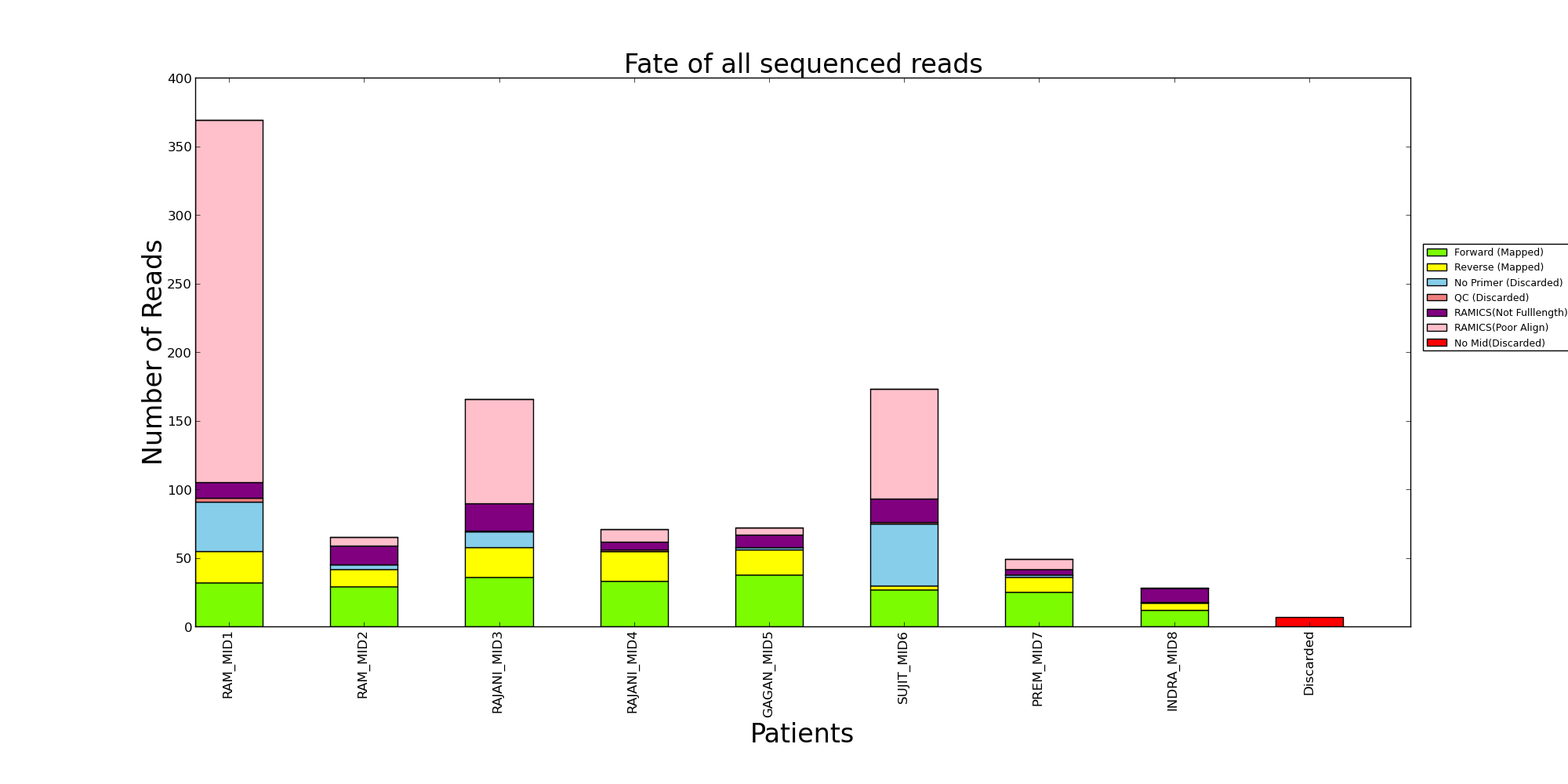


Figure 4. 19: Number of sequences mapped that went to final result (grass green and bright yellow color) and the number of sequences that are discarded (other colors) per patient at different processing steps of Seq2Res.

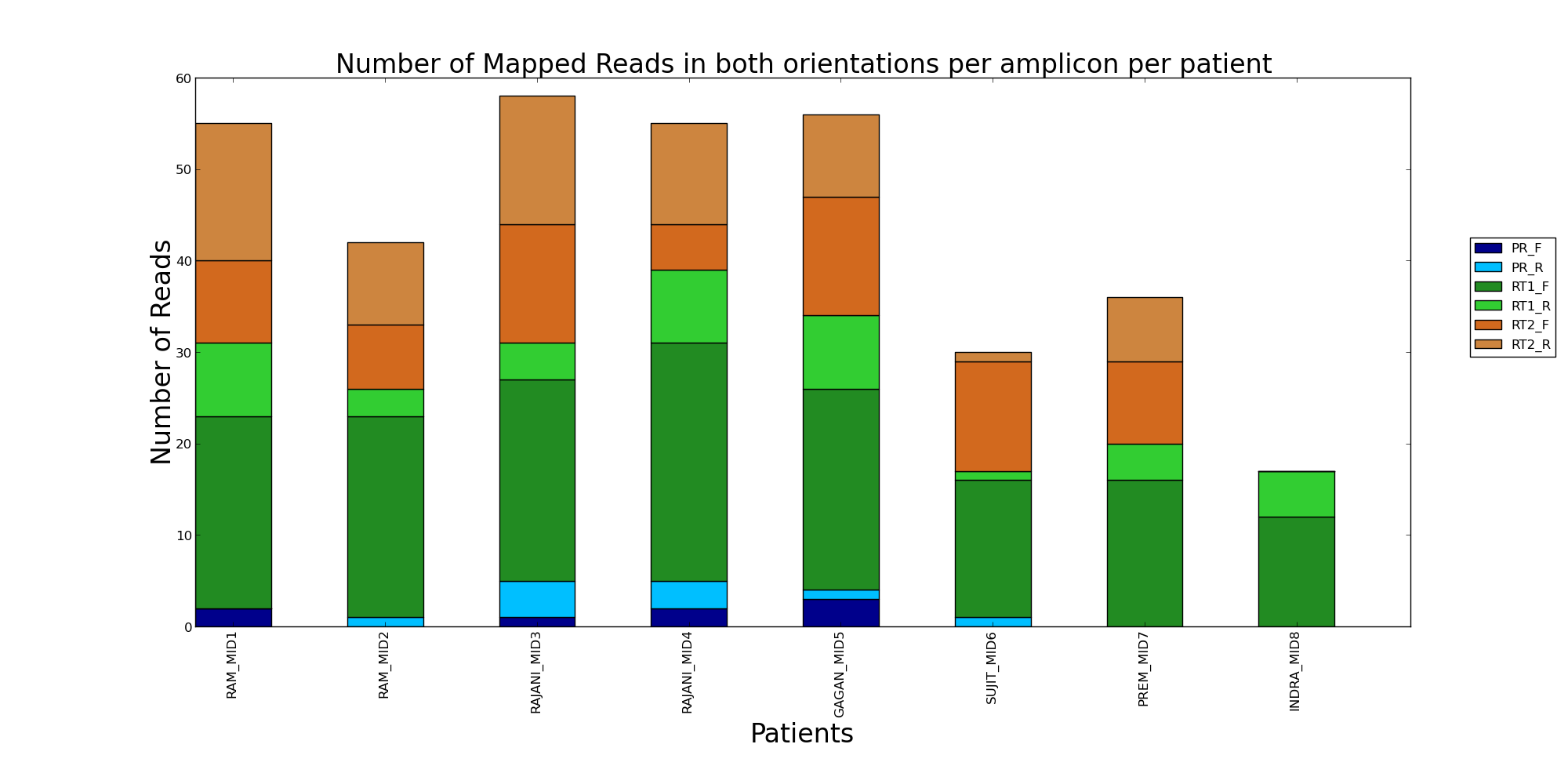


Figure 4. 20: Number of sequences in forward (with “\_F” in legend) and reverse (with “\_R” in legend) strands per amplicon per sample that were mapped to reference sequences and went to final result.

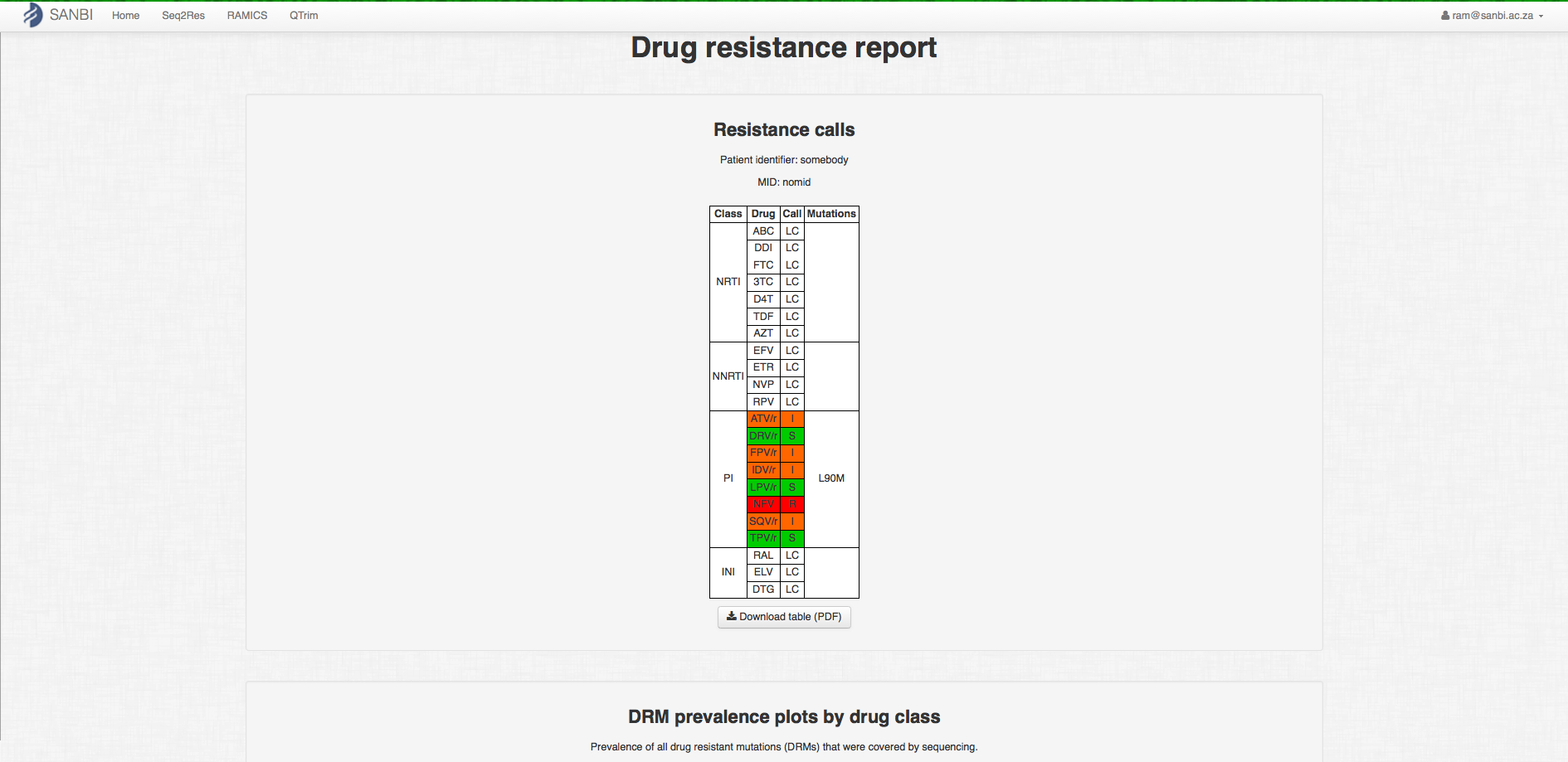


Figure 4. 21: Patient specific drug resistant result page showing the drug resistant report of the sample at the top in the page.

Color code: Red – highly resistant, Orange – intermediate resistant, Green - Susceptible

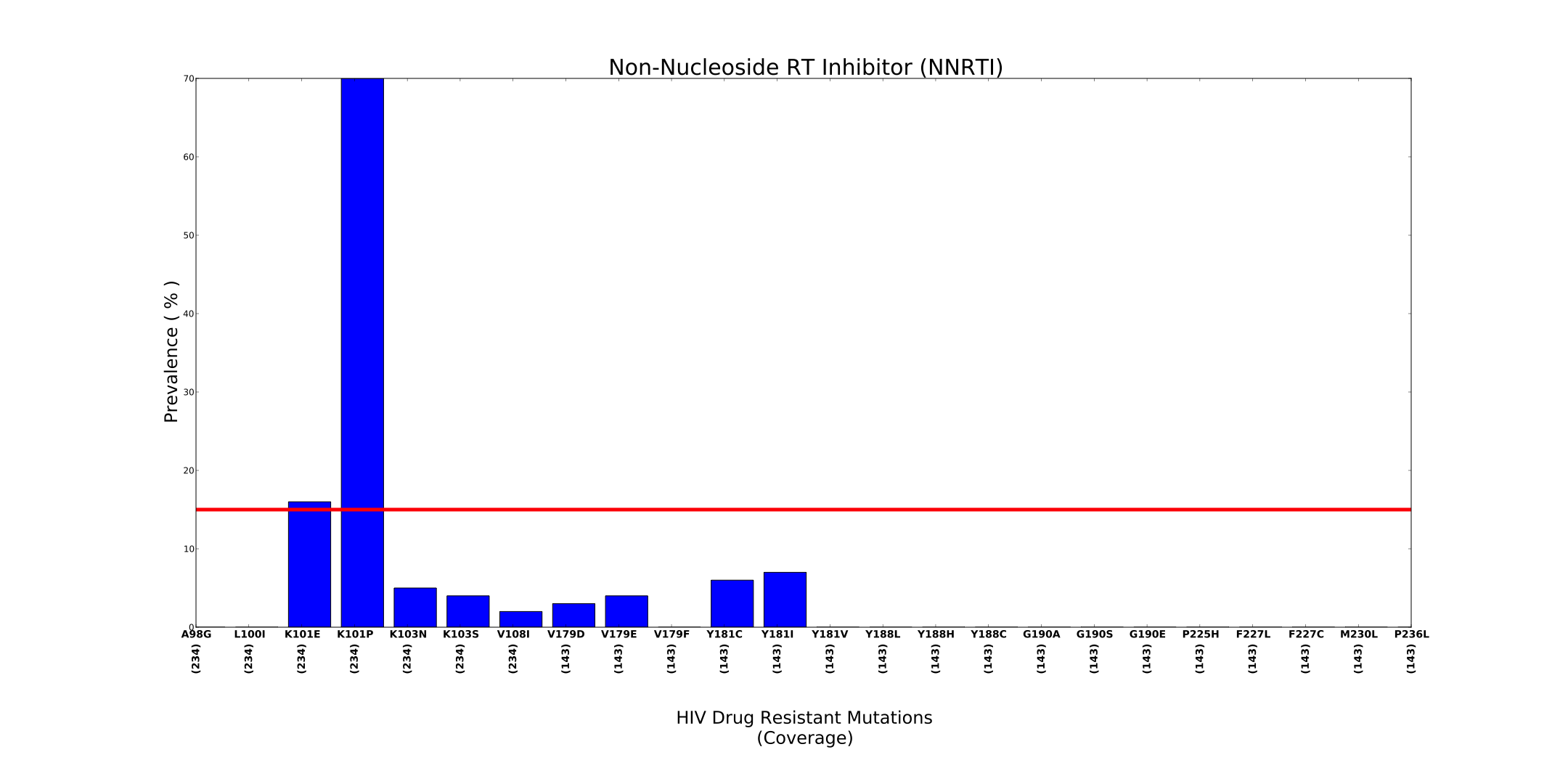


Figure 4. 22: A prevalence plot showing the observed prevalence of mutations that are resistant to NNRTI drugs. A number (in vertical orientation) shows the coverage of the drug resistant mutations. A horizontal red line shows the cutoff prevalence.



Figure 4. 23: Difference in nucleotide sequence alignment (A and C) and the corresponding amino acid sequence alignment (B and D) as obtained from RAMICS (A and B) and Sierra web service (C and D). The numbers indicate the codon position corresponding to the HIV *pol* reference sequence.

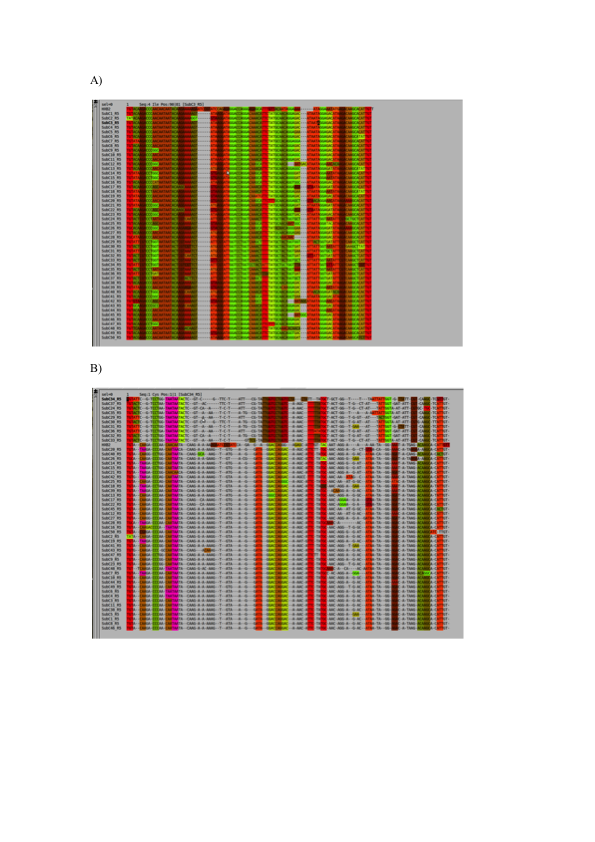


Figure 4. 24: Mapping of HIV subtype C V3 sequence reads to the HIV HXB2 reference sequence using A) RAMICS that maps at codon space and B) Muscle that maps at nucleotide space. Each color stripe represents an amino acid. Both the alignments are shown using alignment viewer called seaview (Gouy et al., 2010).

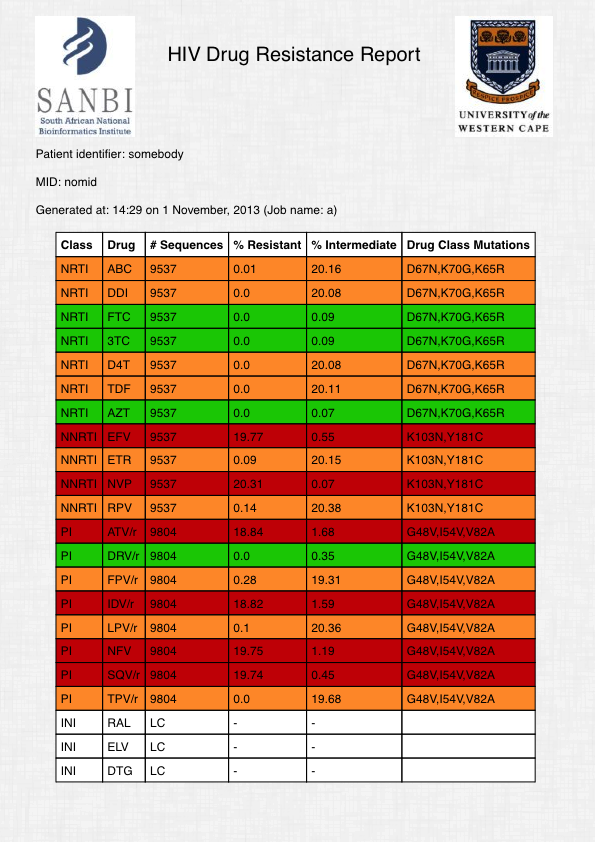


Figure 4. 25: Drug resistance test report of a sample in a table. The table shows number of sequences analyzed for the corresponding drug (third column), the percentage of sequences that are resistant (fourth column) and intermediate resistant (fifth column) to the corresponding drug and the drug resistant mutations (sixth column)

Color code: Red – highly resistant, Orange – intermediate resistant, Green - Susceptible