Instruction of SuperDecode toolkit (v1.0)

Yao-Guang Liu's Lab
South China Agricultural University, College of Agriculture

SuperDecode (v1.0) is an integrated toolkit of a standalone software package and web-base tool for mutation analysis. It comprises three modules, **DSDecodeMS, HiDecode, and LaDecode, which were designed to automatically decode mutations from Sanger, high-throughput short-read (next- generation sequencing), and long-read sequencing data (third-generation sequencing), respectively, from target-site-containing amplicons.** DSDecodeMS is an updated version of our previously developed web-based tools DSDecode and DSDecodeM in CRISPR-GE (http://skl.scau.edu.cn/), which can identify mutations directly from Sanger sequencing chromatograms of target-site-containing PCR amplicons (Liu et al., 2015; Xie et al., 2017). DSDecodeMS is used to identify mutant alleles from superimposed sequencing peaks derived from biallelic and heterozygous mutations in diploid organisms. HiDecode is used for large- scale mutation identification from one or two target-site-containing amplicons in a large number of samples using NGS. Various small mutations, including chimeric mutations, can be analyzed by HiDecode in diverse materials, such as diploid and polyploid organisms, culture cell lines, protoplasts, and calli. LaDecode is a dedicated tool for decoding multiple and complex mutations induced by multiplex target-site editing within a certain region by long- read TGS of long-range PCR amplicons, including PacBio and Nanopore sequencing. Leveraging the advantages of long sequencing reads, LaDecode is capable of identifying complex mutations, including complex SVs, and distinguishing all haplotypes in the samples. Considering the time and cost, it is recommended to use Sanger sequencing and DSDecodeMS analysis for fewer samples; NGS and HiDecode analysis are used for large numbers of samples (more than 30 or 50 samples).

This software package is publicly available for non-profit use by researchers; the use of this software package for profit and commercial purposes by the companies or institutes requires our authorization and permission.

- **Software download and installation**

1. Software download

We provide GUI version for Windows, MacOS systems, and a command-line version for Linux system. You can obtain the packages, including this document, from GitHub (https://github.com/xiexr/SuperDecode). Alternatively, you can directly download the packages from Cowtransfer link: https://tbtools.cowtransfer.com/s/d69cdec128f64e. The use of this software package for profit purposes by companies requires our authorization. SuperDecode can also be accessed from the plugin store of TBtools.

For convenience, **we also developed the web-based version of SuperDecode**

(http://www.crispr-ge.com/superdecode/), which provides the principal functionalities as the standalone version. You can also download the user-document (in Chinese an English) from this website.

2. Software installation

The packages can be directly used in your own PC without installation. After you download the bianary file, unzip it to a local directory on your disk (note: no space or Chinese characters allowed in the directory path). For GUI-versions, clicking the "SuperDecode.exe" in the unziped directory to open the program. For convenience, you can send a shortcut of "SuperDecode.exe" to Desktop. It will take a few minutes to open the program of MacOS for the first time.

For Linux version, Unzip the tar.gz file to your local disk (using command: tar -zxf superdecode_linux.tar.gz). Ensure no space in the path. An example script is provided in the directory for you to run the test example data of each module.

3. Download the example file

SuperDecode provides example file for each module, which can be accessed in the folder of SuperDecode package. For convenience, the example files can be loaded automatically by clicking the “Load example files” from the “Help” menu.

4. The hardware requirements of SuperDecode

For a better experience, we recommend running the software on computers with a memory greater than 4 GB and at least 4 cores in the CPU.

- **DSDecodeMS module**

DSDecodeMS decodes mutations directly from Sanger sequencing chromatograms of target-site(s) contained amplicons. It is majorly used for analyzing simple mutation types (such as homozygous, heterozygous and biallelic) of diploid organisms.

1. Amplification of target-site(s) contained fragments

Due to the limitation of read length of Sanger sequencing, the size of amplified fragment is suggested to be about 500~800 bp. Site-specific primers T#-F and T#-R are designed about 200~300 bp upsteam and downstream of the target site, repectively. The length of T#-F and T#-R is about 18~21 nt. Tm value of T#-F and T#-R is 58℃~60℃, which is calculated according to the formula Tm = (GC% × 41 + 69.3) - 650/L (L = primer length). We highly suggest to check the specificity of the primers using the “Primer-BLAST” tool of NCBI. The amplicon fragment can contain 1~2 target sites. If it has highly homologue sequences in the genome, the length of amplified fragment can be extended appropriately. The recommended PCR mixture is shown in Table 1-1.

Table 1-1 The PCR reaction mixture of the DSDecodeMS (25 μL)

2× Taq mix	12.5 μL
T#-F (10 μM)	0.5 μL (0.2 μM final concentration)
T#-R (10 μM)	0.5 μL
gDNA	20~30 ng
ddH_2O	up to 25 μL

The PCR program is shown in Table 1-2.

Table 1-2 PCR program for amplifying target-site contained fragment

Initial denaturing	94℃, 3 min	
Denaturing	96℃ 15 s	30~32 cycles
Annealing	58℃ 15 s	
Extension	65℃ 10 s, 68℃ 10 s, 72℃ 10 s, 65℃ 10 s (If the target fragment is >1 kb, increase the time by 3~5 s per temperature section)	
Final extension	72℃, 5 min	

(Note: use of variable temperatures in extension can improve the amplification efficiency of sequences with uneven GC distribution)

Take 3~4 μL of several PCR products for agarose gel (1%) electrophoresis to confirm the size and specificity. If the concentration of products is too low, 2~3 cycles can be appropriately added to increase the concentration.

The PCR products (20~25 μL) are sent to the company for Sanger sequencing (most companies provide the service of PCR product purification). Alternatively, user can purify the PCR products using a purification kit and send the purified products to the company for Sanger sequencing. To obtain high-quality sequencing chromatograms, we highly recommend to use an inner primer (17~20 nt with Tm=56℃~58℃) that is positioned at 150~250 bp upstream or downstream of the target site for sequencing, but not to employ the used PCR primers for sequencing, which usally produce low-quality sequencing files. If two target sites arepresent in the amplicon, a forwad primer upstream of the first target site and a reverse primer downstream the second target site can be designed for sequencing each site.

2. Decode Sanger sequencing files using DSDecodeMS

The DSDecodeMS module in SuperDecode was used to detect mutation on the ab1 file obtained by Sanger sequencing.

(1) The interface of DSDecodeMS

The interface of DSDecodeMS mainly includes a reference panel, a sequencing file input panel, a viewer for chromatogram traces, and a result panel. (Figure 1-1).

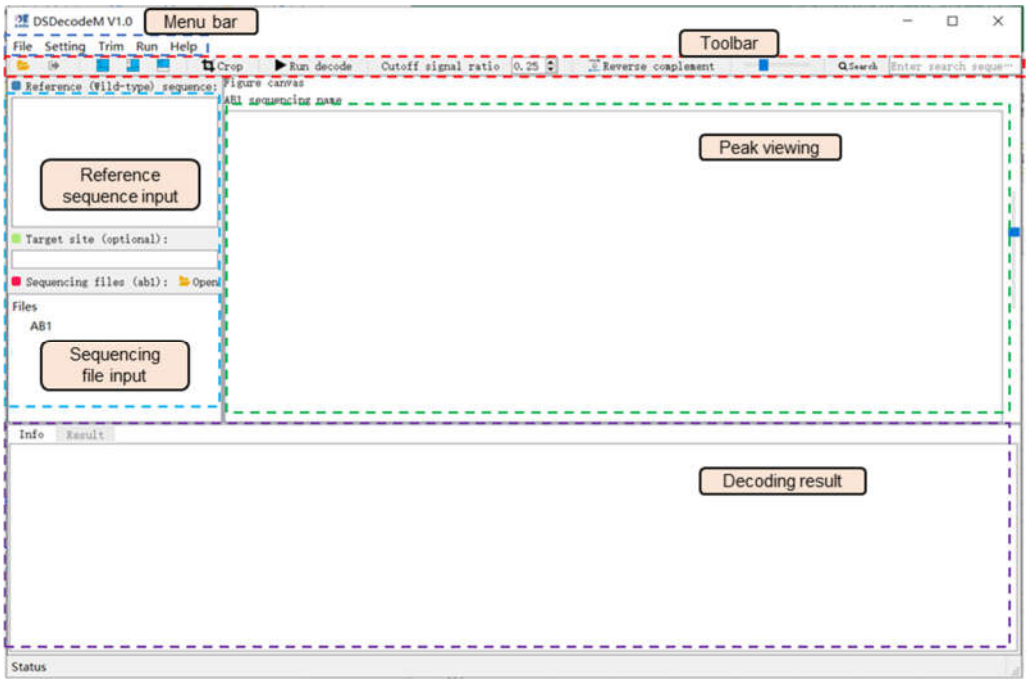

Figure 1-1 The interface of DSDecodeMS

(2) Decoding operation of DSDecodeMS

The use of DSDecodeMS is mainly includes following three steps:

i) Paste the wild-type reference sequence corresponding to the amplicon in the sequence input module box (the wild-type reference sequence is between the T#-F and T#-R primers). The input type is common text or Fasta format (same direction as the sequencing chain or reverse complementary). Next, enter the target sequence (with or without PAM), which will be serviced for locating mutations close to the target, but this is an optional steps not needing in most cases. Clicking the “Open” button to open the sequencing file input dialog. User can input the sequencing files (in ab1 format) corresponding to the same reference sequence and target in this dialog. By double-clicking each sequencing file name, user can view the peak on the right viewer track; checking the peak quality one by one is recommended before starting the analysis.

ii) Generally, the preset parameters of DSDecodeMS is good for analysis, and no parameter setting adjustment is required. If some sequencing files of slightly poor quality

cannot produce correct results, users can adjust the sequencing cutoff signal threshold (default value is 0.25). Other optional settings, including the length of anchor sequenc and length of degenerate sequence can be adjuted through the “Setting” from the menu.

Click the “Run decode” button on the tool bar or form “Run” menu to start the mutation analysis.

DSDecodeMS will quickly outputs the decoding results in the result panel. Users can copy the results to a Word or PPT file; or save the results by clicking the the “Save” button at the upper right corner

(3) Other handy functions

i) Triming low-quality ends in sequencing chromatograms

The quality of the Sanger sequencing chromatograms is usually poor at the beginning and the end (about 30-60 bp). Low-quality ends in the sequencing chromatograms may also hamper decoding. DSDecodeMS provides the function to remove low-quality ends of the chromatograms using the Richard Mott algorithm or user-defined parameters from the “Trim” menu (Figure 1-2).

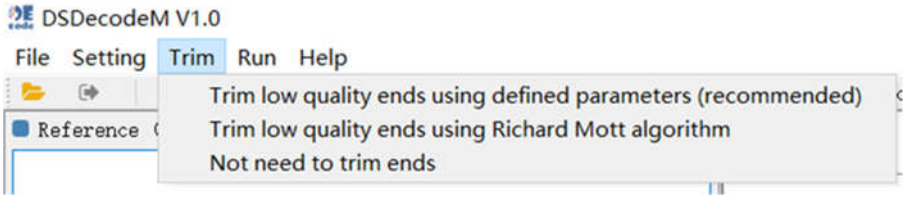

Figure 1-2 Trim the low-quality ends using DSDecodeMS

ii) Capturing figures using the “Crop” function

Each module of SuperDecode provides a screenshot function. Users can open the screenshot function by clicking the “Crop” button on the toolbar. After selecting the screenshot area by dragging the left mouse button, double-clicking to copy the screenshot or right-clicking to save the screenshot.

iii) Subsequence search

Enter a subsequence (4 nt or more) in the “Search” box on the toolbar for searching it in the opened chromatograms. After pressing the “Enter” key, the program will automatically searche for the subsequence. This function supports the search for degenerate sequences.

iv) Export fasta format sequence files from ab1 files

By clicking “Export fasta” in the “Run” menu, the sequence of each ab1 file can be exported into a local file in fasta format.

3. Decode Sanger sequencing files using web-based software tool DSDecodeM

(1) The interface of web-based software tool DSDecodeMS

The interface of web-based software tool DSDecodeM mainly includes a reference sequence input panel, a sequencing file input panel, and a parameter setting panel. (Figure 1-3).

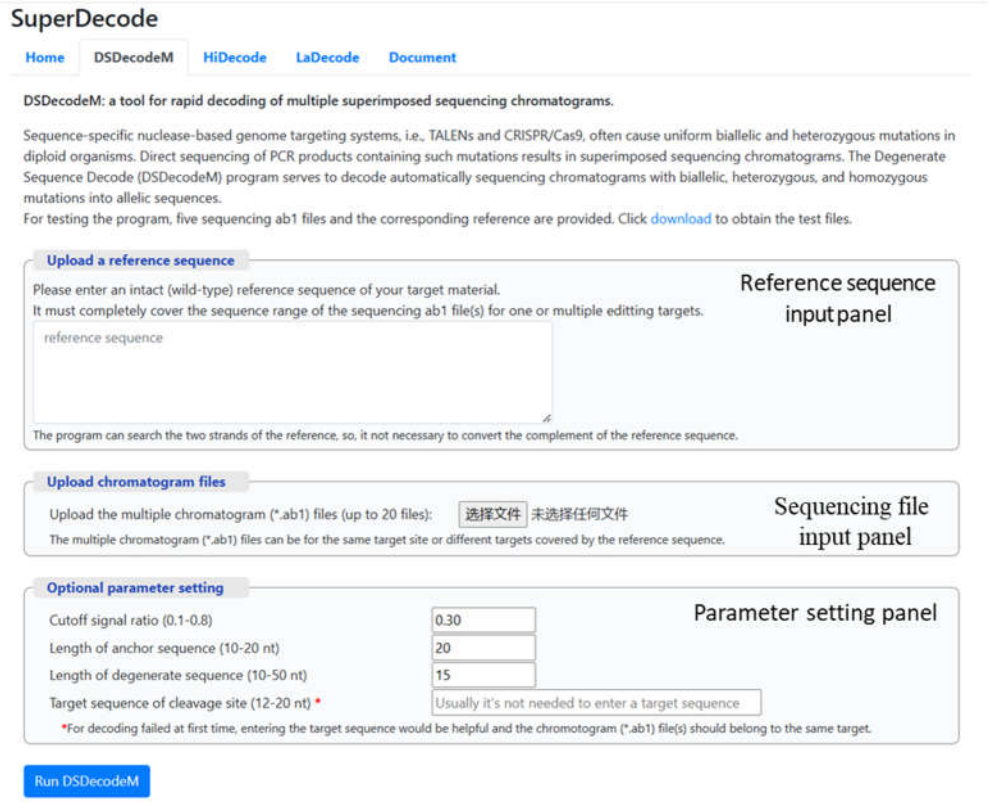

Figure 1-3 The interface of web-based software tool DSDecodeM

(2) Decoding operation of web-based software tool DSDecodeM

The use of web-based software tool DSDecodeMS is mainly includes following four steps:

i) Paste the wild-type reference sequence corresponding to the amplicon in the reference sequence input panel (the wild-type reference sequence is between the T#-F and T#-R primers). The input type is common text or Fasta format (same direction as the sequencing chain or reverse complementary).

ii) Upload the sequencing files (in ab1 format) in the sequencing file input panel. In web-based software tool DSDecodeMS, the size limitation for uploading sequencing files is up to 20.

iii) Setting the parameter for decoding, include target site, cutoff signal ratio and length of anchor and degenerate sequences in the parameter setting panel.

iv) Click the "Run DSDecodeM" button to start the mutation analysis.

After decoding, the results will output in a new webpage (Figure 1-4), user can select the corresponding results to copy and save.

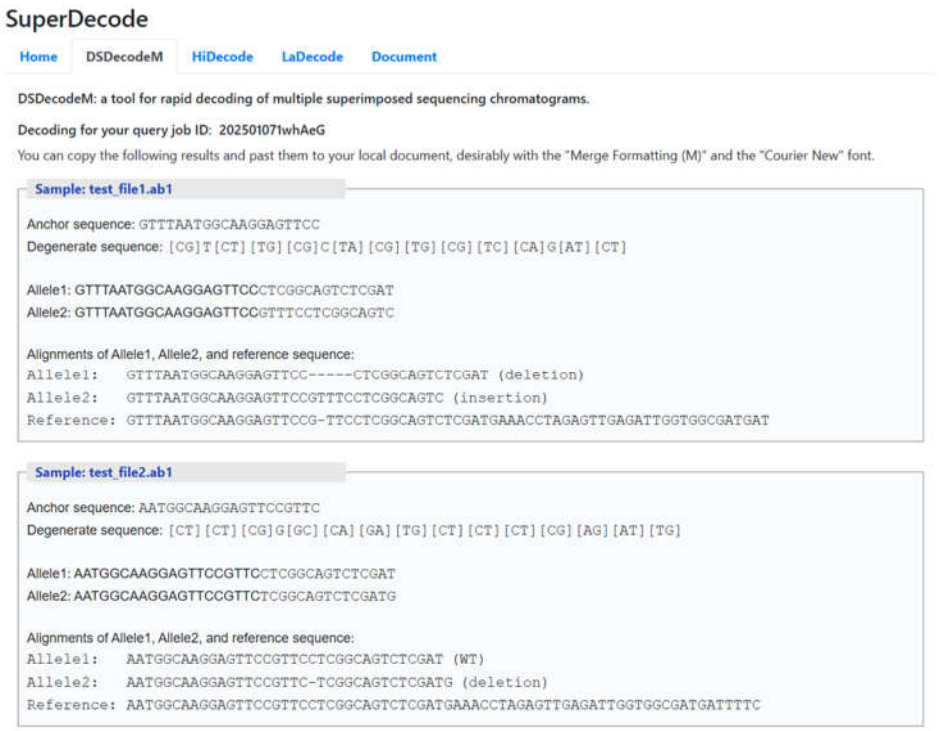

Figure 1-4 The output of web-based software tool DSDecodeMS

4. Causes and solutions for decoding failure

i) The PCR amplicons may contain non-specific products. Repeated use of amplification primers for sequencing may result in poor-quality sequencing chromatograms. It is highly recommended to use an inner primer for Sanger sequencing.

ii) The amplified sequence has homologous sequence(s) in the genome. Reset the site-specific primers for amplifying the target sequence by extending a certain distance to the outside of the target to design the specific primers. By utilizing the single or multi base differential sites between the target sequence and its homologous sequence(s), the primer 3 'end is designed to the differential site to enhance the specificity of the product. Users can also design two sets of specific primers for two rounds of nested PCR to improve the specificity of the product.

iii) The mutation of the sample is a complex mutation type such as chimeric mutations, or there are multiple targets in the target region that induce structural variations such as fragment deletion or inversion among the targets. For chimeric mutations, it is recommended to use HiDecode. For multiplex editing, it is recommended to use LaDecode.

iv) The input wild-type reference sequence does not match the sequencing file. Check whether the input reference sequence corresponds to the target amplification sequence.

v) Unenve GC or polyA(T) in the sequence region between the primer and target site causes low-quality chromatograms. Try to use another sequencing primer on the other side of

the target site that do not contain such complicated regions.

- **HiDecode module**

NGS library preparation and analysis of HiDecode

When large numbers (>50) of samples are needed to be analyzed, HiDecode is recommended to reduce the cost. This method is particularly suitable for detecting one- or two-targets contained sequences within a certain region (<150 bp) in samples such as chimeric mutation, polyploids, cell lines, protoplasts, or callus tissue. Here, we suggest a two-round PCR-based strategy to construct the NGS library of pooled amplicons. In addition to the default barcodes, users can also defined their own sample barcodes for mutation analysis using HiDecode. Moreover, HiDecode also support mutation analysis of simple targeted PCR amplicons of individual samples.

1. NGS library construction for HiDecode

The NGS library for HiDecode is constructed mainly through two rounds of PCR. A pair of unique barcodes are added to both ends of the amplicons and then mixed into a sequencing library for NGS (Figure 2-1).

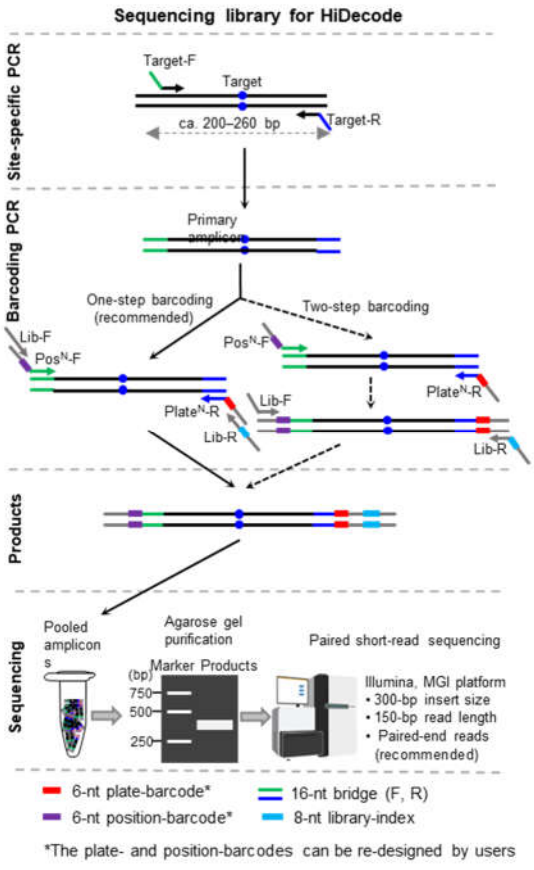

Figure 2-1 PCR-based construction of sequencing libraries for HiDecode

(1) Primer design for HiDecode

HiDecode library construction involves three pairs of primers, including site-specific primers T#-F and T#-R for amplifying target sequences, universal primers pos^N-F/col^N-F and plate^N-R/row^N-R (pos^N-F and plate^N-R labeling plate-well positions (from A1 to H12) and plate numbers for 96×n barcoding when the number of sample more than 96, col^N-F and row^N-R labeling column and row position in a 96-well plate for 12×8 barcoding strategy), and primers Lib-F and Lib- R for adding library adapters. The sequence structure of the NGS library is shown in Figure 2- 2.

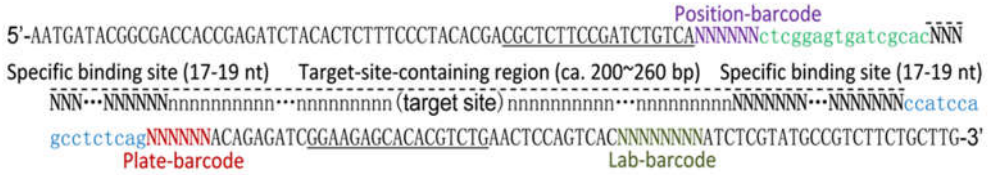

Figure 2-2 Sequence structure of NGS library

i) Site-specific primers T#-F and T#-R

Considering that the read length of current paired-end NGS tools is ca. 150 nt for each side, the fragment size of first-round PCR is designed within ca. 200–260 bp (including the two bridge sequences). After checking the specificity, the primers T#-F and T#-R are added with different bridge sequences (16 nt) at their 5′ ends.

Site-specific forward primer (T#-F): 5′-ctcggagtgatcgcacNNNNNN…NNNNN-3′ (17 to 19 N are the genomic site-specific region, Tm = 56℃~58℃; the 16 nt at the 5′ end is the bridge sequence that matchs pos^N-F);

Site-specific reverse primer (T#-R): 5′-ctgagaggctggatggNNNNNN…NNNNN-3′ (17 to 19 N are the genomic site-specific region, T_m = 56℃~58℃; the 16 nt at the 5′ end is the bridge sequence that matchs plate^N-R).

ii) Universal primers pos^N-F/col^N-F and plate^N-R/row^N-R

HiDecode support the analysis of sequencing data generated by 96×n barcoding strategy and 12×8 barcoding strategy. When the sample more than 96, it is recommended to construct the library with 96×n barcoding strategy (using pos^N-F and plate^N-R for barcoding). pos^N-F contains sequences (from 5′ to 3′) matching the library adapter, position barcode, and sequences (16 nt) binding to the 5′ sequence of the first round of products. plate^N-R contains sequences (from 5′ to 3′) matching the library adapter, plate barcode, and sequences (16 nt) binding to the 3′ sequence of the first round of products. A total of 96 position barcodes (6 nt) and 96 plate barcodes (6 nt) for labeling the positions (A01 to H12) of samples in 96- well plates and the corresponding plate number, respectively, thus creating uniquely barcoded amplicons for up to 9,216 samples (96 × 96). Users can synthesize a certain number of plate-barcoded primers based on the number of samples.

pos^{N}-F: 5′-CGCTCTTCCGATCTGTCANNNNNNctcggagtgatcgcac-3′ (the underlined sequence is the adapter sequence matching the library primer Lib-F, the 6 N are the position barcode, and the lowercase letters are the sequence binding to the 5′ sequence of the first round of products).

plate^{N}-R: 5′-CAGACGTGTGCTCTTCCGATCTCTGTNNNNNNctgagaggctggatgg-3′(the underlined sequence is the adapter sequence matching the library primer Lib-R, the 6 N are the plate barcode, and the lowercase letters are the sequence binding to the 3′ sequence of the first round of products).

For detailed sequences, see Appendix 1-HiDecode all primer sequences.xlsx

For convenience, the 96 position-barcoded primers can be prepared in a 96-well PCR plate and stored at -20℃ for future use.

Users can choose 12×8 barcoding strategy to construct the library while the number of sample less than 96 (using col^{N}-F and row^{N}-R for barcoding), col^{N}-F contains sequences (from 5′ to 3′) matching the library adapter, column barcode, and sequences (16 nt) binding to the 5′ sequence of the first round of products. row^{N}-R contains sequences (from 5′ to 3′) matching the library adapter, row barcode, and sequences (16 nt) binding to the 3′ sequence of the first round of products. A total of 12 column barcodes (6 nt) and 8 row barcodes (6 nt) for labeling the column positions (1 to 12) and row positions (A to H) of samples in a 96- well plates, respectively, thus creating uniquely barcoded amplicons for 96 samples. Users can synthesize a certain number of column barcoded primers and row barcoded primers based on the number of samples.

col^{N}-F: 5′-CGCTCTTCCGATCTGTCANNNNNNctcggagtgatcgcac-3′ (the underlined sequence is the adapter sequence matching the library primer Lib-F, the 6 N are the column barcode, and the lowercase letters are the sequence binding to the 5′ sequence of the first round of products).

row^{N}-R: 5′-CAGACGTGTGCTCTTCCGATCTCTGTNNNNNNctgagaggctggatgg-3′(the underlined sequence is the adapter sequence matching the library primer Lib-R, the 6 N are the row barcode, and the lowercase letters are the sequence binding to the 3′ sequence of the first round of products).

For detailed sequences, see Appendix 1-HiDecode all primer sequences.xlsx

For convenience, the column barcoded primers and row barcoded primers can be prepared in a 96-well PCR plate and stored at -20℃ for future use.

iii) Library adapter primers Lib-F and Lib-R

NGS sequencing adapters can be introduced into the amplicons by PCR. The adapter contains a laboratory-specific barcode (Lab barcode) to distinguish libraries from different customers. Sequencing companies demultiplex raw sequencing reads by this barcode. A

single- or dual-barcodes are supported. Following primers are designed with single Lab barcode.

Lib-F: 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCA-3′ (the underlined sequence is the adapter sequence bingding pos^N-F/col^N-F)

Lib-R: 5′-CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCC-3′ (8 N sequence is the Lab barcode that must be written as the sequence as in the + strand; the underlined sequence is the adapter sequence binding to $plate^N$-R/row^N-R primer). The Lab barcode needs to be reverse complement in the library information form when send it to a sequencing company. Lab barcodes are used to seperate different sequencing libraries from different users (laboratories) by sequencing compary, because the company may mix these libraries into the same batch for sequencing. For 96×n barcoding strategy, a laboratory can use different Lab barcodes or the same Lab barcode for different sublibraries for the same batch of sequencing, because they can be identified by using different $plate^N$-R primers. In addition, if dual-Lab barcodes are used, the sequences of Lib-F and Lib-R should be modified according to the recommendations of company.

Note: This is designed based on the Illumina platform. If other sequencing platforms are used, pos^N-F/col^N-F and $plate^N$-R/row^N-R, Lib-F and Lib-R are needed to be redesigned based on the platform adapter sequence.

(2) Steps of amplification

i) Preparation of genomic DNA and various primers

All primers and genomic DNA working solutions are prepared with 0.3× TE (do not use 1× TE, to aviod too much EDTA that inhibits enzyme activity).

Site-specific primers T#-F, T#-R, and $plate^N$-R: diluting to 10 μM as working stocks.

pos^N-F working solution: synthesize 96 pos^N-F (from pos^{A1}-F to pos^{H12}-F) with 5 OD each, 1 OD/tube. Dissolve one tube of each primer with 1× TE into 10 μM storage solution, and freeze the rest as dry powder at -20℃ for storage. Take 2.0 μL of each pos^N-F (10 μM) and add it to the corresponding wells (A1~H12) of a 96-well PCR plate. Add 18 μL of 0.3× TE to each well, which is **1.0 μM** pos^N-F working solution. Store at -20℃ (after multiple uses discard when the remaining is less than 4~5 μL).

col^N-F/row^N-R working solution: synthesize 12 col^N-F (from col^1-F to col^{12}-F) and 8 row^N-R (from row^A-R to row^H-R) with 5 OD each, 1 OD/tube. Dissolve one tube of each primer with 1× TE into 10 μM storage solution, and freeze the rest as dry powder at -20℃ for storage. Take 2.0 μL of each col^N-F (10 μM) and row^N-R (10 μM), and add it to the

corresponding wells (A1~H12) of a 96-well PCR plate. Add 16 μL of 0.3× TE to each well, which is **1.0 μM** col^{N}-F/row^{N}-R working solution. Store at -20℃ (after multiple uses discard when the remaining is less than 4~5 μL).

Genomic DNA working solution: take about 200 ng of genomic DNA from each sample and add it to the 96-well PCR plate (A1~H12). Add 0.3× TE to each well to 10 μL (20 ng/μL). Store at -20℃.

ii) First round of specific amplification

The HiDecode library is constructed via two rounds of PCR. In the first-round PCR, site-specific primers T#-F and T#-R are used to amplify the target-site contained fragment. The total volumn of PCR mixture is prepared according to the number of samples and then dispensed into a 96-well PCR plate. The PCR reaction mixture is shown in Table 2-1 (the number of samples is *n*). 15 μL of the reaction mixture is added to each well.

Table 2-1 Total reaction mixture of first-round PCR (15 μL per reaction)

2× Taq mix	7.5 μL × *n*
T#-F (10 μM)	0.3 μL × *n* (final concentration 0.2 μM)
T#-R (10 μM)	0.3 μL × *n* (final concentration 0.2 μM)
ddH_2O	7.5 μL × *n*

Use a 96-pin replicator (or multichannel pipette) to add ~1μL of template DNA into the wells that contain the PCR reaction mixture (After adding the template DNA, the 96-pin replicator is washed three times with pure water and is then washed once with ethanol, following subjected for flaming at high temperatures to remove the residual DNA). The program of PCR reaction is shown in Table 2-2.

Table 2-2 PCR reaction program of first-round PCR

Initial denaturing	94℃ 3 min	
Denaturing	96℃ 10 s	10 cycles
Annealing	58℃ 20 s	
Extension	68℃ 5 s	
Extension	72℃ 5 s	
Extension	68℃ 5 s	
Denaturing	96℃ 10 s	18~20 cycles
Annealing/ Extension	68℃ 5 s	

Extension	72℃ 5 s	
Extension	68℃ 5 s	
Final extension	68℃, 1 min	1 cycle

(Note: use of variable temperatures for extension can improve amplification efficiency of sequences with different GC contents)

If the target fragment has homologous sequence(s) in genome, causing T#-F and T#-R unable to specifically amplify the target fragment, new specific primers exT#-F and exT#-R (T_m = ~60℃) can be designed to extended regions. Perform the first round of nested PCR to obtain high-specific amplification products. The PCR reaction mixture is shown in Table 2-3.

Table 2-3 First-round nested PCR reaction mixture (15 μL per reaction)

2 × Taq mix	7.5 μL × *n*
exT#-F (10 μM)	0.15 μL × *n* (final concentration 0.1 μM)
exT#-R (10 μM)	0.15 μL × *n* (final concentration 0.1 μM)
ddH_2O	7.5 μL × *n*

The program of PCR reaction is shown in Table 2-4:

Table 2-4 The reaction program of first-round nested PCR

Initial denaturing	93℃ 3 min	
Denaturing	96℃ 10 s	15 cycles
Annealing	60℃ 20 s	
Extension	68℃ 10 s	
Extension	72℃ 30 s	
Extension	68℃ 10 s	

Mix T#-F and T#-R primers (10 μM) in equal amounts, and dilute with ddH_2O to 3 μM each (total > *n* μL). After completing the first-round PCR, add 1 μL T#-F/T#-R mixture (3 μM) to each well (final concentration is 0.2 μM). Continue the PCR amplification reaction to obtain the specific amplification products of the target fragment. The PCR reaction program is shown in Table 2-5.

Table 2-5 The reaction program of second-round nested PCR

Initial denaturing	96℃ 10 s	

Annealing / Extension	58℃ 20 s	18-20 cycles
Extension	68℃ 5 s	
Extension	72℃ 10 s	
Extension	68℃ 5 s	
Final extension	68℃ 1 min	

The fragment sizes of the first-round PCR products are 200~260 bp. The amplified products of several samples can be randomly selected from each plate for agarose gel (1.5%) electrophoresis to check the specificity and product size.

ii) Second-round PCR for barcoding

The purpose of the second-round of PCR reaction is to add the barcodes and sequencing adapters. Prepare the second-round of PCR mixture according to the number of samples and then dispensed into a 96-well PCR plate. The PCR mixture is shown in Table 2-6 (the number of samples is *n*):

Table 2-6 Second-round PCR reaction mixture (15 μL per reaction)

2× Taq mix	7.5 μL × *n*
$plate^N$-R (10 μM) (Note: $plate^N$-R does not need to added in mixture for library constructed with 12×8 barcoding.)	0.08 μL × *n* (final concentration 0.05 μM)
Lib-F (10 μM)	0.3 μL × *n* (final concentration 0.2 μM)
Lib-R (10 μM)	0.3 μL × *n* (final concentration 0.2 μM)
ddH_2O	7.5 μL × *n*

For 96×n barcoding, using a 96-pin replicator (or multichannel pipette) to take ~1μL of the pos^N-F (1.0 μM) into the reaction mixture in each PCR well (the final concentration of each pos^N-F is about 0.06 μM).

For 12×8 barcoding, using a 96-pin replicator (or multichannel pipette) to take ~1μL of the col^N-F/row^N-R working solution (1.0 μM) into the reaction mixture in each PCR well (the final concentration of each col^N-F and row^N-R is about 0.06 μM).

Add 30 μL ddH_2O to dilute the first-round of PCR products for 3 times. And then use a 96-pin replicator (or multichannel pipette) to take ~1μL of the diluted products into the reaction mixture in each PCR well (the total dilution of the products is about 45×).

The program of second-round PCR is shown in Table 2-7.

Table 2-7 Program of second-round PCR

Denaturing	96℃ 10 s	8 cycles
Annealing	57℃ 15 s	
Extension	68℃ 5 s	
Extension	72℃ 10 s	
Extension	68℃ 5 s	
Denaturing	96℃ 10 s	8~10 cycles
Annealing / Extension	68℃ 10 s	
Extension	72℃ 10 s	
Extension	68℃ 5 s	
Final extension	72℃ 1 min	1 cycle

The length of the final products after the second-round PCR is about 340~400 bp (including adapters and barcodes). Several products can be randomly selected from each plate for agarose gel (1.3%) electrophoresis to check the specificity and product size. If the concentration of products is too low, 2~3 cycles can be increased of extension to increase the concentration.

Pool 3 μL of each sample from each 96-well plate to generate a polled group (about 300 μL per group) for purification using 1.3% agarose gel electrophoresis followed by a DNA purification kit. After quantifying the concentration of each group of purified products, mix the purified products of the groups with equal amounts (the group with less than 96 samples should reduce the mixing amount accordingly). The final concentration of the total mixed product should be > 50 ng/μL, and the product volume should be > 60 μL.

2. NGS sequencing

The pooled library is sent to a company for sequencing through Illumina, X-Ten, or Nova-seq platforms (if other platforms are used, pos^N-F and $plate^N$-R, Lib-F and Lib-R are needed to be redesigned according to the adapter sequences provided by the platform). Approximately 1 Gb (a minimum output for each sequencing task by a company) of clean NGS data (with an average of ca. 300 bp of adapter-deleted reads) should be obtained per 1~ 10 × 96 samples, which will produce an average sequencing depth of >30,000× (or >3,000×) in theory.

Some companies in China (such as Wuhan Kinstar Global, HaploX, and Geneplus) can complete the sequencing within 3~5 workdays after receiving a customer's sequencing library samples.

3. Mutation analysis using HiDecode

(1) The interface of HiDecode

The main interface of HiDecode includes the menu, toolbar, file input panel, plate canvas, and decoding results panel (Figure 2-3). The menu bar provides commonly used setting functions and plugin. The toolbar provides some setting options such as setting mutation threshold and CPU thread number. The file input panel is used to input wild-type reference sequence file, target sequence file, and paired-end sequencing files. In the barcode settings, user can select the used position- and plate-barcodes. The plate canvas is used to interactively display the analysis results of each sample.

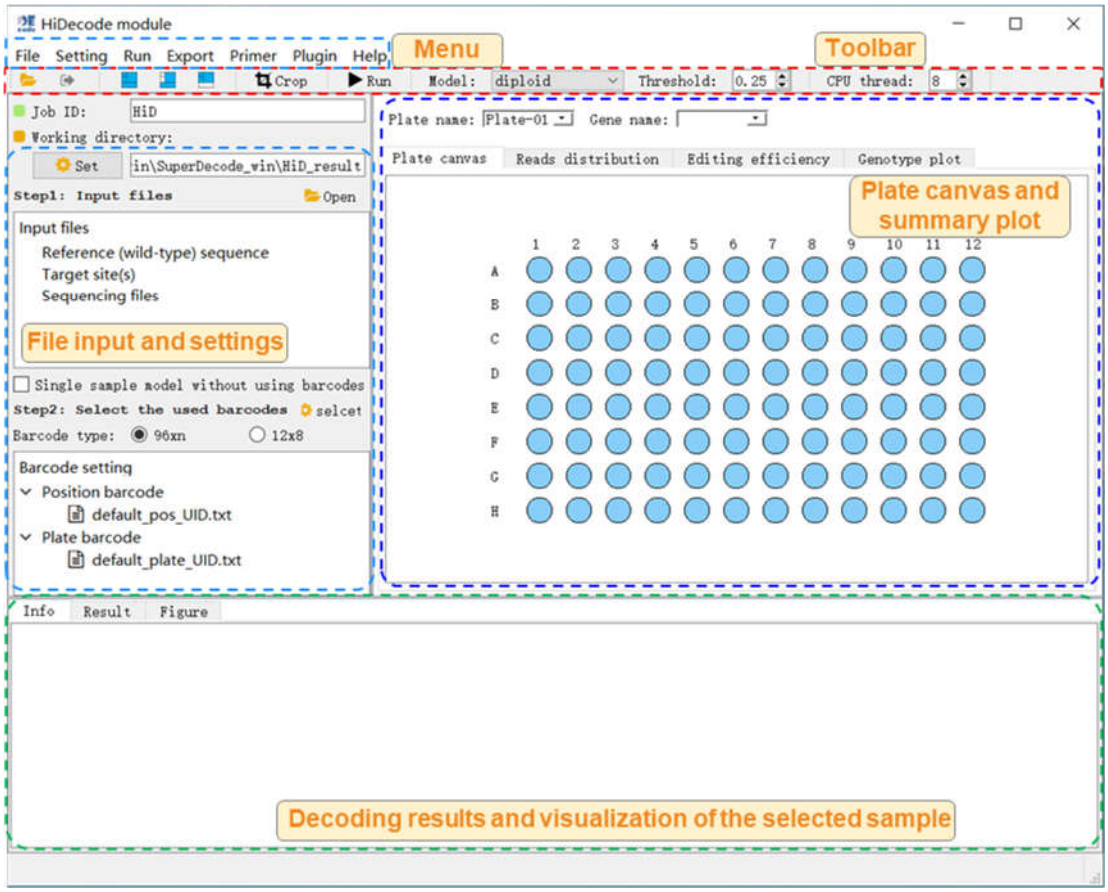

Figure 2-3 The interface of HiDecode

(2) Decoding operation of HiDecode

The use of HiDecode for decoding mutations from NGS reads of pooled amplicons is mainly includes following three steps:

i) File iniput: Click the “File” button on the menu bar or the “Open” button from the file input panel to open the file input dialog (Figure 2-4). Next, input the files, including: the

reference file in Fasta format (covering the wild-type genome sequence between T#-F and T#-R), optional target sequence file (with or without PAM sequence), and sequencing file (paired-end data, which requires clean data after filtering and removing adapters). The format of reference and target site files can be viewed by clicking the provided example files.

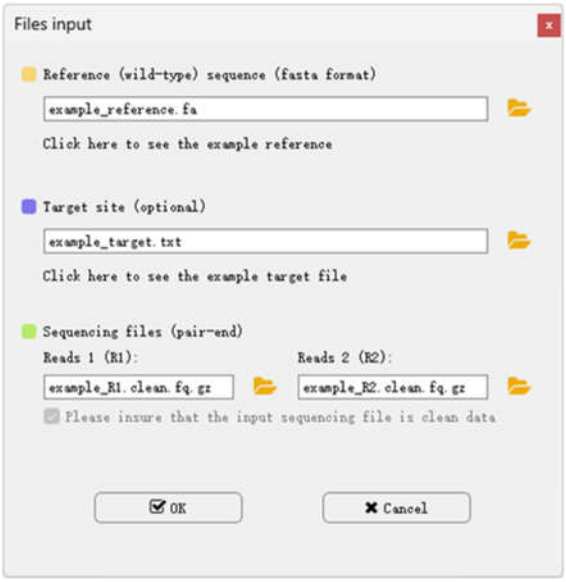

Figure 2-4 The interface of file input of HiDecode

ii) Accoding to the strategy of library construction, select the barcode type in the barcode setting panel. For 96 × n barcoding, double-click “default_pos_UID.txt” and “default_plate_UID.txt” in the barcode setting panel to select the used position barcode and plate barcode in amplification. For 12×8 barcoding, double-click “default_col_UID.txt” and “default_row_UID.txt” in the barcode setting panel to select the used column barcode and row barcode in amplification. After opening the setting box, check the barcode to be used.

iii) Set the parameters for data analysis. HiDecode provides some built-in threshold values for different analysis models, including the diploid, polyploid, and low-frequency models, based on our laboratory tests. The low-frequency model is mainly used to identify mutations in cell lines, calli, or plant protoplasts; the output is all possible variations at the threshold. In addition to the preset threshold, users can adjust the threshold as needed. When the proportion of reads of a certain variant site in the sample is greater than the set threshold, the program considers the genotype as a realiable mutation and outputs the mutation site as well the frequency.

iv) Click the “Run” button to start the analysis task.

(3) Analysis results

The analysis results of each well will be displayed in the decoding result panel. Yellowed position(s) indicate no result was produced for the sample(s) (Figure 2-5). When a sample is

selected, the mutation result will be displayed in the decoding result panel at the bottom, including the mutation type, total number of reads covered, sequence alignment, and supported read count of each allele. By draging the slider above, users can adjust the sequence range to be displayed.

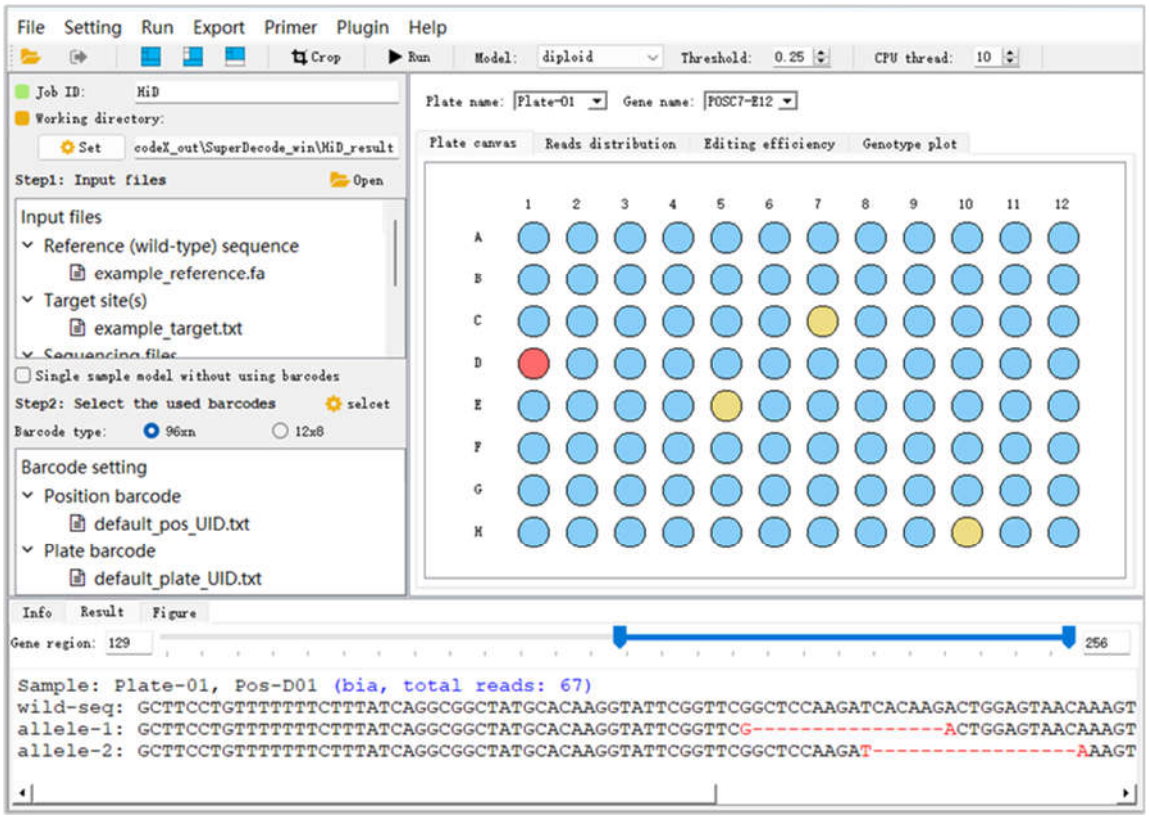

Figure 2-5 The decoding result of HiDecode

HiDecode modules offer comprehensive analysis results that can be viewed in an interactive mode. After the analysis, plate canvas generates plots summarizing the distribution of covered reads, editing efficiency, and mutation genotype for the samples in the selected plate (Figure 2-6).

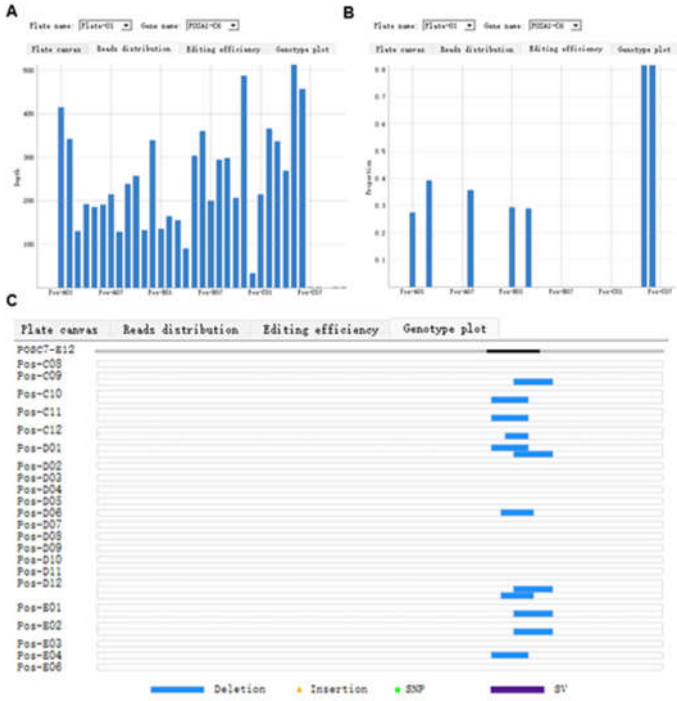

Figure 2-6 The reads distribution, editing efficiency and mutation genotype for the samples in the selected plate of HiDecode

Moreover, HiDecode modules generate figures depicting the mutation ratio, InDel size distribution, and mutation frequency at each position across the reference sequence for the selected sample (Figure 2-7).

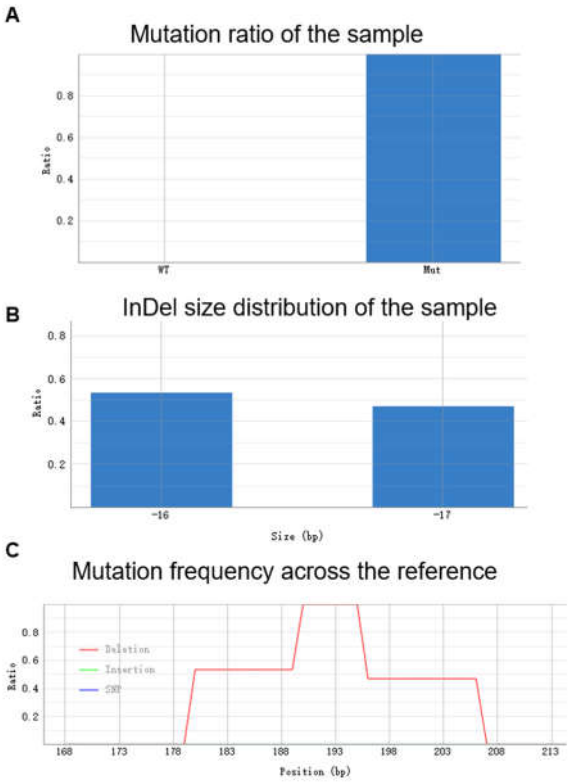

Figure 2-7 The mutation ratio, InDel size distribution, and mutation frequency at each position across the reference sequence for the selected sample of HiDecode

The results can be exported to an Excel file via the “Export” button in the menu bar. Two types of result files can be exported. The function of “Export all results to excel file” is to export the detailed decoding results of all wells and plates (including the number of reads, mutant sequences, and the sequence alignments) (Figure 2-8), and “Fetch mutations to a single file” is to export simplified information, including mutant sequences, the count of supported reads, and mutant types.

Figure 2-8 Detailed mutation result table of HiDecode

In addition, the “Export mutation genotypes of all samples” in “Export” can output the genotype plot of all samples, allowing users to intuitively view the mutation results of the samples (Figure 2- 7).

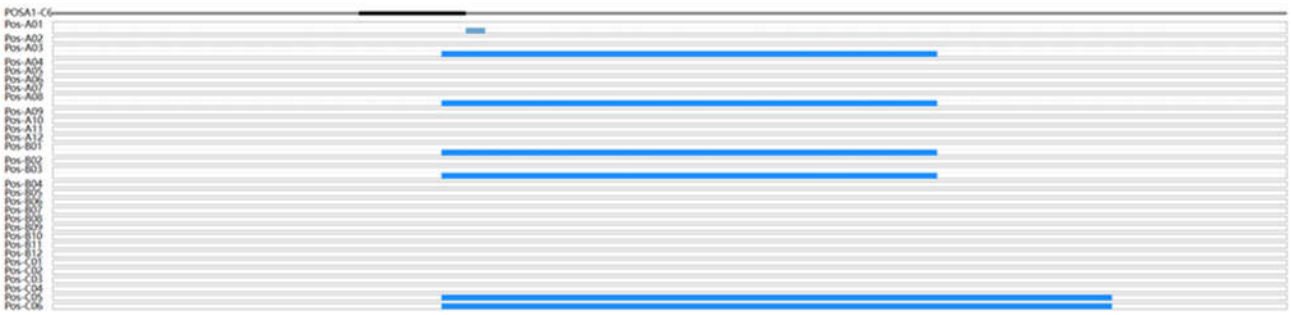

Figure 2-9 The mutation genotype plot of each sample in HiDecode

(4) Other handy functions

i) Generation of site-specific primers

HiDecode can automatically add the bridge sequences to the site-specific primers for first-round PCR. The program allows users to modify the bridge sequences by clicking the “edit” button (Figure 2-10). After entering the bridge sequence, the target name, and the

specific sequence (5′-3′), click the “Generate primers” button to generate site-specific primers containing bridge sequences for the first-round PCR reaction. (Note: this function does not have the function to detect the specificity of the primer. It is recommended that use other tools such as the “Primer-BLAST” of NCBI or “primerDesign” of CRISPR-GE (http://skl.scau.edu.cn/) to detect the specificity of the primer.)

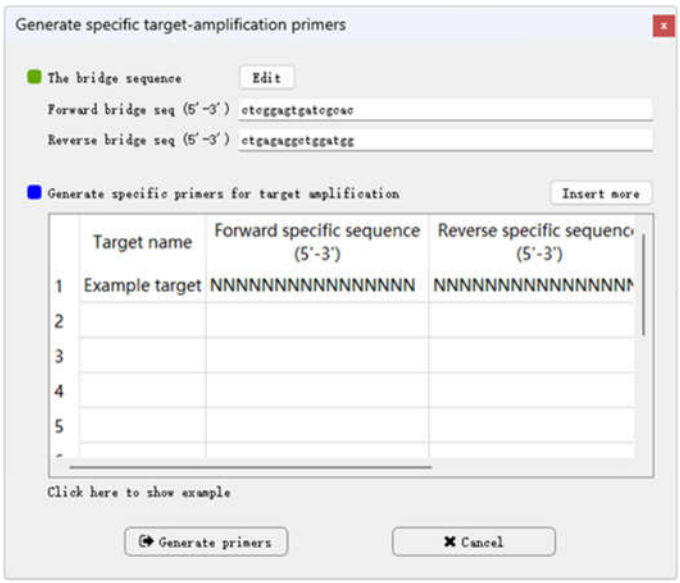

Figure 2-10 Primer generation dialog windows for HiDecode

ii) Barcode setting

To improve the decoding speed of the program, HiDecode extracts the barcode sequences according to their position in the amplicon for distinguishing different samples.

For 96×n barcoding, if the user modifies the position of the barcodes in pos^{N}-F and plate^{N}-R, the barcode position and length can be reset in the “Reset the default 96×n barcodes (define your own barcodes)” dialog from the “Setting” menu (Figure 2-11).

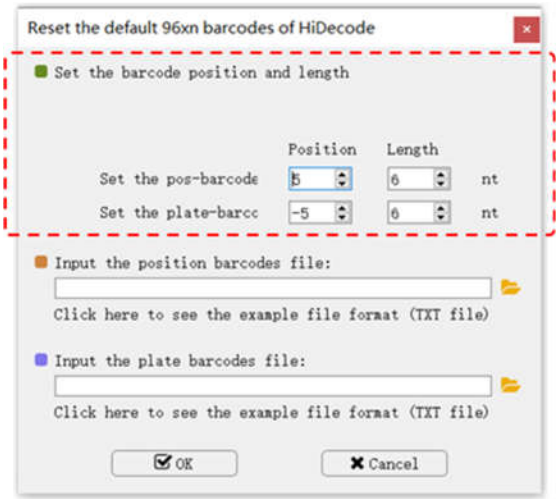

Figure 2-11 Barcode position and length setting in HiDecode for 96×n barcoding

Insteading using the default postion and plate barcodes, users can define their own barcodes according to actual use during PCR. Open the barcode setting dialog by clicking the “Reset the default 96×n barcodes (define your own barcodes)” from the “Setting” menu to define new barcodes in batch. Alternatively, one by one modification is allowed via the barcode selection dock (Figure 2-12).

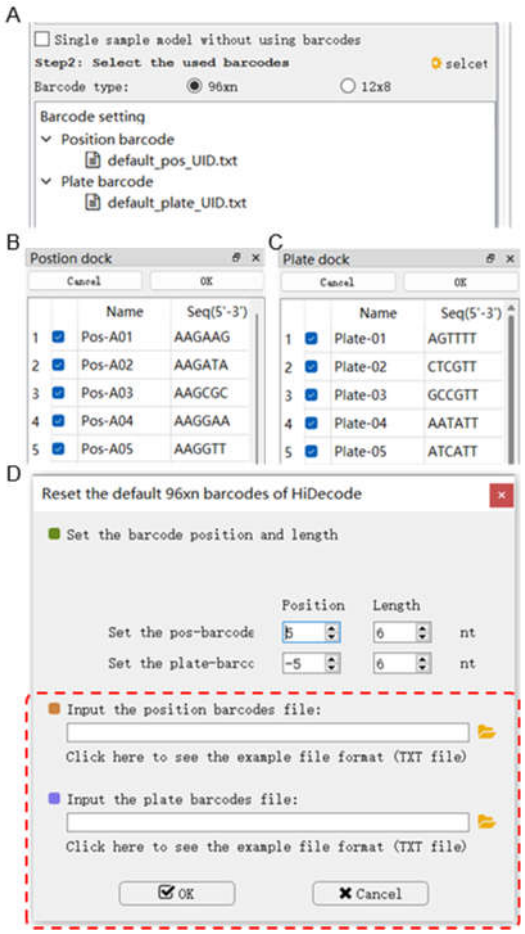

Figure 2-12 Barcode setting of Hidecode for 96×n barcoding

For 12×8 barcoding, the position and length of column barcode and row barcode can be reset in the “Reset the default 12×8 barcodes (define your own barcodes)” dialog from the “Setting” menu (Figure 2-13).

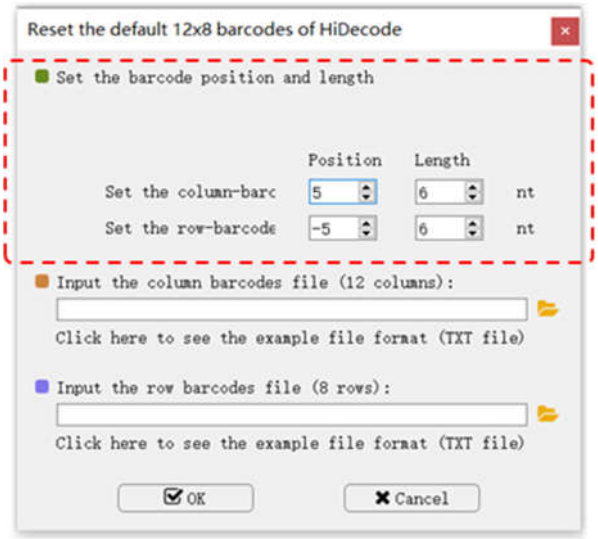

Figure 2-13 Barcode position and length setting in HiDecode for 12×8 barcoding

Users also can define column barcode and row barcode by clicking the “Reset the default 12×8 barcodes (define your own barcodes)” from the “Setting” menu. Alternatively, one by one modification is allowed via the barcode selection dock (Figure 2-14).

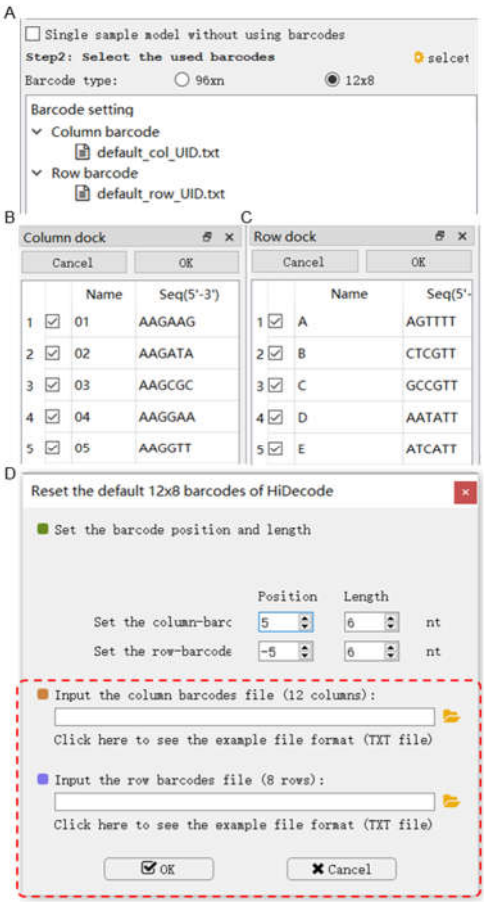

Figure 2-14 Barcode setting of Hidecode for 12×8 barcoding

iii) Removing sequencing adapters and filtering reads from raw sequencing files

Most of companies provide clean data that the sequencing adapters have been removed and low-quality reads have been filtered. In case some companies just provide raw data, HiDecode provides the function for removing adapter sequence and filtering raw quality reads via the “Quality control for sequencing files” from the “Plugin” menu (Figure 2-15). The program will automatically remove sequencing adapters and filter low-quality reads for the raw sequencing reads, and output a quality control reports for users.

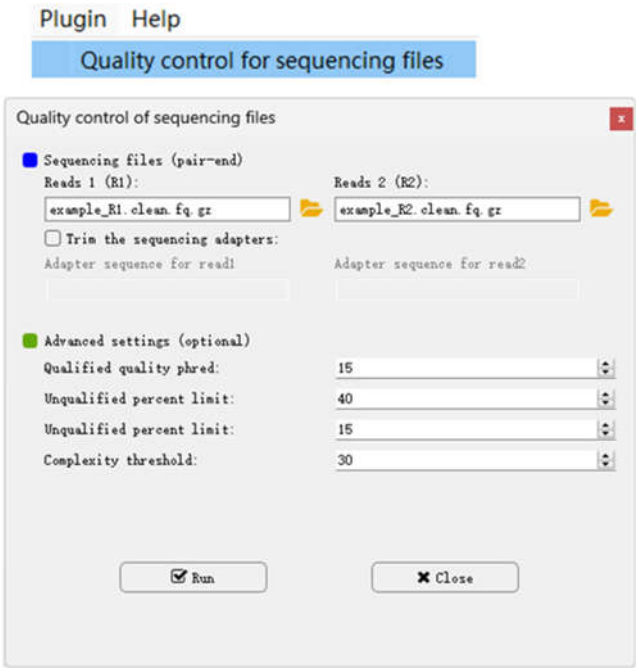

Figure 2-15 The quality control function of HiDecode for removing the sequencing adapters and low-quality reads of raw sequencing files

iv) Generation of site-specific primers

In addition to handling sequencing data from pooled amplicons, HiDecode is capable of analyzing NGS data from a single sample in which the target-site(s)-containing amplicons are directly subjected to sequencing without the barcoding step using the “Single sample model without using barcodes” function on the bottom of the file input and setting panel. In the dialog (Figure 2-16), input the reference sequence, target sequence (optional), and sequencing read files for starting the analysis.

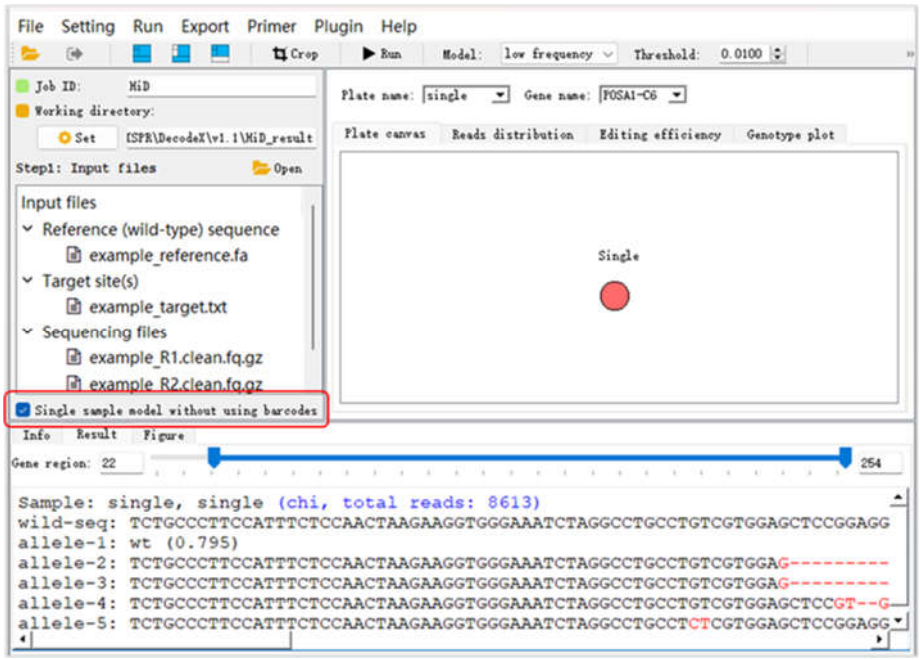

Figure 2-16 The mutation analysis from NGS data for a single sample

4. Mutation analysis using web-based software tool HiDecode

(1) The interface of web-based software tool HiDecode

The interface of web-based software tool HiDecode mainly includes a includes file input panel and barcode settings panel. (Figure 2-17).

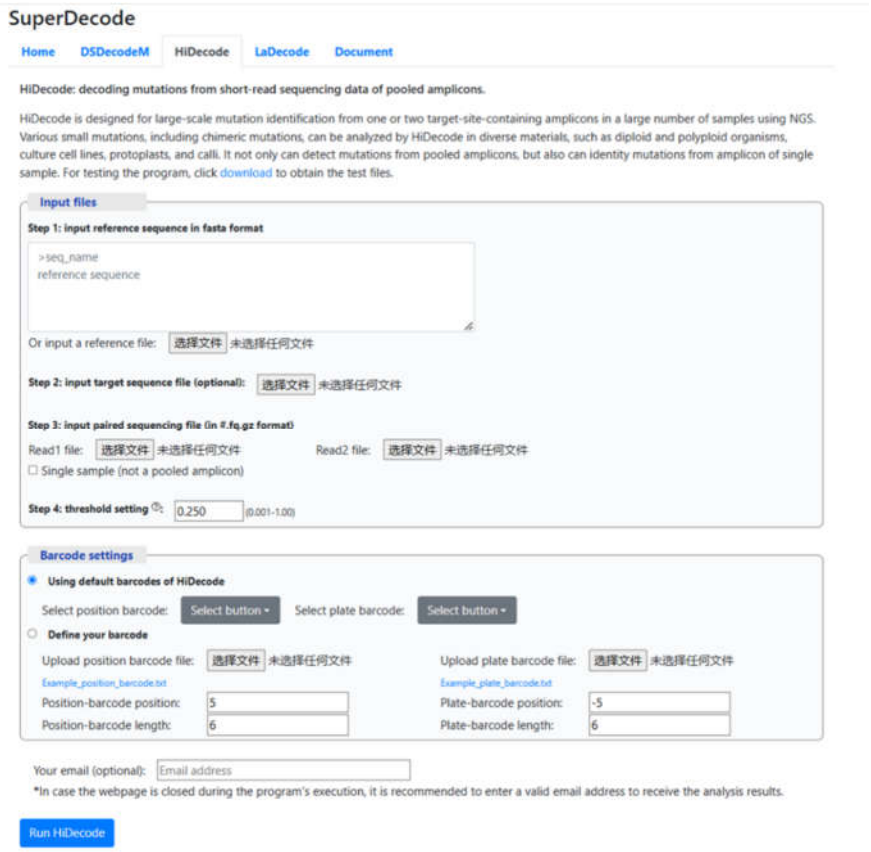

Figure 2-17 The interface of web-based software tool HiDecode

(2) Decoding operation of web-based software tool HiDecode

The use of web-based software tool HiDecode is mainly includes following four steps:

i) Input the files, including: the reference file in Fasta format (covering the wild-type genome sequence between T#-F and T#-R), optional target sequence file (with or without PAM sequence), and sequencing file (paired-end data, which requires clean data after filtering and removing adapters). The format of reference and target site files can be viewed by clicking the provided example files.

ii) Setting the threshold and model of the task, if analysis the mutation from NGS data for a single sample, click the "Single sample (not a pooled amplicon)" button in the file input panel.

iii) Setting the barcode sequence, the length of barcode and the position of barcode in the barcode settings panel.

iv) Click the "Run HiDecode" button to start the analysis task. (Note: before starting the analysis task, it is recommendes to enter the email to receive the decoding results in the email input box.)

After decoding, the results will output in a new webpage (Figure 2-18), user can select the sample to display the decoding result, including the the sequence alignments, mutation efficiency, InDel size distribution, and mutation frequency at each position across the reference sequence for the selected sample. Users also can save the decoding results via click the download button.

Figure 2-18 The decoding result of web-based software tool HiDecode

- **TGS module**

Library preparation and analysis of LaDecode

Multiplex editing within a certain region otfen induce complex structural variations (SVs), such as fragment deletions, inversions, and even duplications among target sites, making it difficult to design primers to amplify each target site. LaDecode is a dedicated tool for decoding multiple and complex mutations induced by multiplex target-site editing within a certain region by long-read TGS of long-range PCR amplicons, including PacBio and Nanopore sequencing.

1. TGS library construction of LaDecode

The library construction procedure of LaDecode is similar to HiDecode. Through two rounds of PCR reactions, a pair of unique long-barcodes are added to the ends of amplicons (Figure 3-1). Considering the higher native error rates of Nanopore and PacBio reads, a set of longer barcodes (12 nt or longer) is introduced into the second-round PCR primers to confidently assign the reads to their template molecules for each sample. In addition, the sequencing adapter is added by sequencing company.

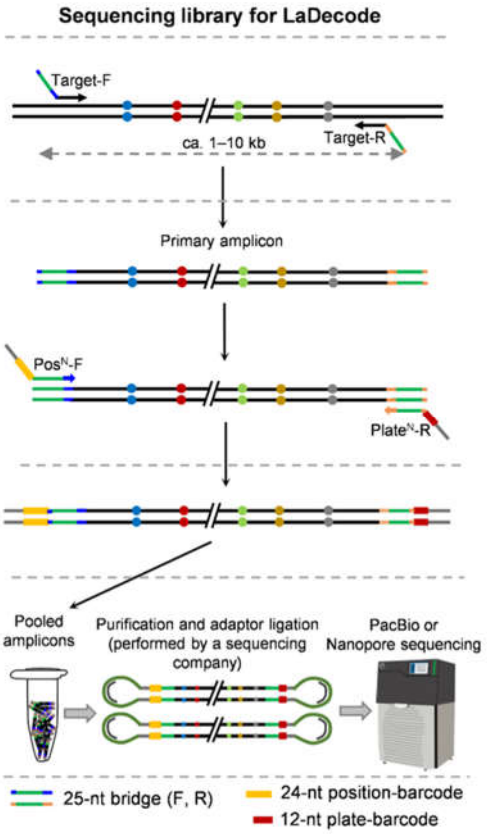

Figure 3-1 Construction of TGS library for LaDecode

(1) Primer design of LaDecode

Similar to HiDecode library construction, library construction of LaDecode involves two pairs of primers, including site-specific primers (Target-F and Target-R) for amplifying the

target sequence, and barcoding primers (pos^N-F and plate^N-R) for labeling samples.

i) Site-specific primers Target-F and Target-R

The site-specific forward primer (Target-F) is designed about 100~200 bp upstream of the first target site, and the site-specific reverse primer (Target-R) is designed 100~200 bp downstream of the last target site. The amplified fragments may be 1~10 kb (or better 2~6 kb). The site-specific pairing parts of the primers (18~20 nt, T_m = 60℃~63℃) need to be checked for specificity using the "Primer-BLAST" tool of NCBI, and then add the bridge sequence required for the second-round of PCR to the 5' end of the primer. The primer sequences are as following:

Target-F: 5'-atcgcctggctccacgctccgagttNNNNNNNNNNNNNNNNNNN-3' (NNN... is the site-specific binding sequences, with a length of 18~20 nt and Tm = 58℃~62℃; lowercase letters are the bridge sequences; the underlined sequence can form a stem-loop structure to inhibit the amplification of primer-dimers and shorter non-specific products, as referring to the STI PCR of Zhao et al., 2022, Molecular Plant).

Target-R: 5'-tacgcctggctccacgctcctagaNNNNNNNNNNNNNNNNNNN-3' (NNN⋯ is the site-specific binding sequence,s with a length of 18~20 nt and a Tm = 58℃~62℃; lowercase letters are the bridge sequences).

ii) Universal primers pos^N-F/col^N-F and plate^N-R/row^N-R

For 96 × n barcoding, pos^N-F and plate^N-R is used for library construction. pos^N-F contains position barcode (24 nt) and the binding sequence to the 5′ sequence of the first round of products. plate^N-R contains plate barcode (12 nt) and the binding sequence to the 3′ sequence of the first round of products. A total of 96 position barcodes (24 nt) and 6 plate barcodes (12 nt) for labeling the positions (A01 to H12) of samples in 96-well plates and the corresponding plate number, respectively, thus creating uniquely barcoded amplicons for up to 576 samples (96 × 6). Users can synthesize a certain number of plate-barcoded primers based on the number of samples.

pos^N-F: 5'-cttgNNNNNNNNNNNNNNNNNNNNNNatcgcctggctccacgctccgagtt -3' (NNN…is a 24-nt position barcode sequence, indicating the position of each sample on the 96-well PCR plate)

plate^N-R: 5'- cgacNNNNNNNNNNNtacgcctggctccacgctccctaga -3' (NNN… is a 12-nt plate barcode sequence, indicating different 96-well PCR plates, one 96-well PCR plate corresponds to one plate^N-R primer)

For convenience, the 96 position-barcoded primers can be prepared in a 96-well PCR plate and stored at -20℃ for routine use.

For 8×n barcoding, col^N-F and row^N-R is used for library construction. col^N-F contains column barcode (24 nt) and the binding sequence to the 5′ sequence of the first round of products. row^N-R contains row barcode (24 nt) and the binding sequence to the 3′ sequence of the first round of products. A total of 12 column barcodes (24 nt) and 8 row barcodes (24 nt) for

labeling the positions (A01 to H12) of samples in a 96-well plates. Users can synthesize a certain number of column-barcode primers and row-barcode primers based on the number of samples.

col^{N}-F: 5'-cttgNNNNNNNNNNNNNNNNNNNNNNNatcgcctggctccacgctccgagtt -3' (NNN…is a 24-nt column barcode sequence, indicating the column position of each sample on the 96- well PCR plate, one 96-well PCR plate corresponds to 12 col^{N}-F primer)

row^{N}-R: 5'- cgacNNNNNNNNNNNNNNNNNNNNNNNtacgcctggctccacgctccctaga -3' (NNN… is a 24-nt row barcode sequence, indicating the row position of each sample on the 96-well PCR plate, one 96-well PCR plate corresponds to 8 row^{N}-R primer)

For convenience, the 12 column-barcoded primers and 8 row-barcoded primers can be prepared in a 96-well PCR plate and stored at -20℃ for routine use.

For detailed sequences, see Appendix 2-LaDecode all primer sequences.xlsx

(2) Steps of amplification

i) Preparation of genomic DNA and various primers

All primers and genomic DNA working solutions are prepared with 0.3× TE (do not use 1× TE, to aviod too much EDTA that inhibits enzyme activity).

Site-specific primers Target-F, Target-R, and plate^{N}-R: diluting to 10 μM for use.

Prepare the working solution of pos^{N}-F into the concentration of 3 μM, and then take 20 μL of each primer into positions (from A1 to H12) of a 96-well plate. Store the at -20℃ for future use.

For 12×8 barcoding, prepare the working solution of col^{N}-F/row^{N}-R into the concentration of 6 μM, and then take 20 μL of each primer into positions (from A1 to H12) of a 96-well plate. Store the at -20℃ for future use.

ii) First-round of site-specific amplification

Use site-specific primers Target-F and Target-R to amplify the target-sites contained fragment, and then prepare the PCR reaction mixture according to the number of samples. The mixture was dispensed into 96-well PCR plates. The PCR reaction mixture is shown in Table 3-1 (the number of samples is *n*).

Table 3-1 Total reaction mixture of first-round PCR for LaDecode (15 μL)

2× Phanta Max Buffer	7.5 μL × *n*
Target-F (10 μM)	0.3 μL × *n* (final concentration 0.2 μM)
Target-R (10 μM)	0.3 μL × *n* (final concentration 0.2 μM)
dNTPs Mix (10 mM)	0.3 μL × *n*
Phanta Max Super-Fidelity DNA Polymerase (1U/μL)	0.3 μL × *n*
ddH_2O	6.3 μL × *n*

(Note: It is recommended to use a high-fidelity/high performance polymerase to amplify

fragments larger than 5 kb)

Dispense the template DNA of each sample (20 ng/μL) into a 96-well PCR plate, and use a 96-pin replicator (or a multi-channel pipette) to add ~1 μL of template DNA (approximately 20 ng) into the PCR mixture to amplify the target amplicon of the sample. The PCR program of the PCR reaction is shown in Table 3-2.

Table 3-2 First-round PCR program of LaDecode

Initial denaturing	95℃ 2 min	
Denaturing	96~98℃ 30 s	30~32 super-cycles
Annealing	60℃ 20 s	
Extension	(65℃ 10 s, 68℃ 10 s, 72℃ 15 s) × *n* (Set the nested cycle number *n* according to the fragment length, ca. 1 kb/*n*)	
Final extension	70 ℃, 2 min	

(Note: If the target sequences are long (>5 kb) with uneven GC content distributions, STI PCR programs (Zhao et al., 2022 Mol Plant) can be used to improve the amplification; If the sequence GC content is higher, use 97~98℃ for denature)

After the first-round of PCR, the amplified products of several samples can be randomly selected from each plate for agarose gel (0.8%) electrophoresis to check the specificity and concentration.

iii) Second-round PCR for barcoding

The purpose of the second-round PCR reaction is to add the barcode sequence. Dilute the first round of PCR amplification products 3 times by adding 30 μL ddH_2O to each first-round product. Prepare the second-round PCR total mixture according to the number of samples (Table 3-3), and dispense 15 μL of the total PCR reaction mixture into a 96-well PCR plate.

Table 3-3 Second-round PCR mixture (15 μL per reaction)

2× Phanta Max Buffer	7.5 μL × *n*
plate^{N}-R (10 μM) (Note: plate^{N}-R does not need to added in mixture for library constructed with 12×8 barcoding strategy)	0.3 μL × *n* (final concentration 0.2 μM)
dNTPs Mix (10 mM)	0.3 μL × *n*

Phanta Max Super-Fidelity DNA Polymerase (1U/μL)	0.3 μL × *n*
ddH_2O	6.8 μL × *n*

For 96×n barcoding, use a 96-pin replicator (or multichannel pipette) to take ~1 μL of pos^N-F (3 μM) and add it to the reaction mixture in each PCR well (final concentration of each pos^N-F is about 0.2 μM).

For 12×8 barcoding, use a 96-pin replicator (or multichannel pipette) to take ~1 μL of col^N-F/row^N-R (3 μM) and add it to the reaction mixture in each PCR well (final concentration of each col^N-F and row^N-R is about 0.2 μM).

Add ~1 μL of the first-round PCR product to the PCR reaction mixture (final concentration is about 45×). The PCR reaction program is shown in Table 3-4.

Table 3-4 Second-round PCR program of LaDecode

Denaturing	96~98℃ 30 s	10~12 cycles
Annealing	62℃ 20 s	
Extension	(65℃ 10 s, 68℃ 10 s, 72℃ 15 s) × *n* (Set the nested cycle number *n* according to the fragment length, 1 kb/*n*)	
Final extension	70 ℃ 3 min	

After the second-round of PCR reaction, several products can be randomly selected from each plate and run on 0.8% agarose gel to examine their quality. If the concentration of products is too low, 2~3 super-cycles can be increased to increase the product concentration.

Pool 3 μL of each sample from the same plate to generate a group. The PCR products are purified using 0.8% agarose gel electrophoresis and a DNA purification kit. Various types of mutation may be generated by multiplex editing, including large fragment deletions, inversions, and even duplications among target sites. The purpose of electrophoresis is mainly to remove primers; in order to ensure that all types of mutation fragments are collected, the time of electrophoresis should not be too long. Pool the products of each group in equal amounts and subject to TGS. The final concentration of the total mixed products is >20 ng/μL, and the product amount is > 2 μg.

2. TGS sequencing

The purified products are sent to a company for TGS library construction (including adding sequencing adapters and purification) and sequencing. It is recommended to use Pacbio HiFi

sequencing platform for the sequencing. The sequencing company will provide clean TGS read file(s) in Fasta, Fastq, or bam format, in which the sequencing adapters are removed. The sequencing sizes for different lengths of amplicons ranging from approximately 1–4 kb, 4–7 kb, and 7–10 kb can be ~1 Gb, 2 Gb, and 3 Gb of clean data per 96 samples, respectively, which will give sequencing depths of >3,000×.

3. Mutation analysis using LaDecode

(1) The interface of LaDecode

The main interface of LaDecode includes the menu, toolbar, file input panel, plate canvas, and decoding results panel (Figure 3-2). The menu bar provides commonly used setting functions and plugin. The toolbar provides some setting options such as setting mutation threshold and CPU thread number. The file input panel is used to input wild-type reference sequence file, target sequence file, and TGS sequencing files. In the barcode settings, user can select the used position- and plate-barcodes. The plate canvas is used to interactively display the analysis results of each sample.

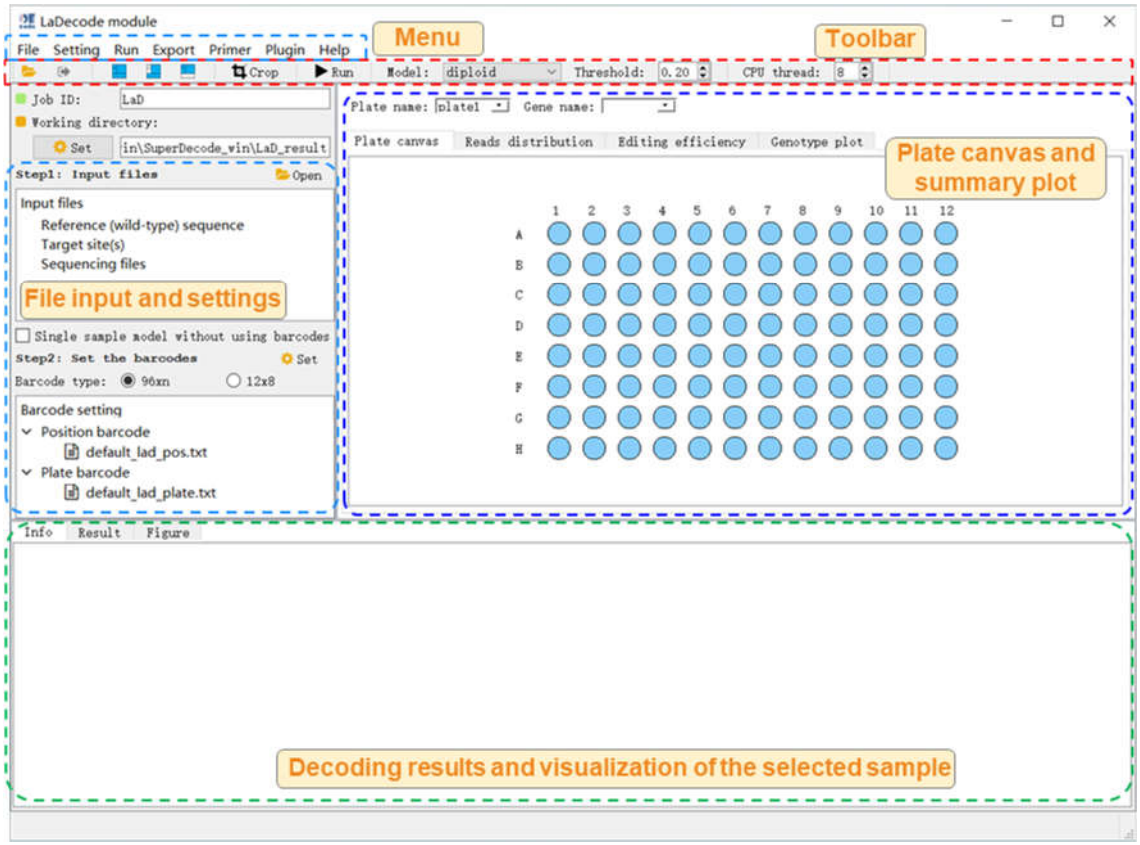

Figure 3-2 The interface of LaDecode

(2) Decoding operation of LaDecode

The use of LaDecode mainly includes four main steps: file input, barcode setting,

parameter setting and decoding. The detailed operations are as follows:

i) File iniput: Clicking the “File” button in the file input panel or on the toolbar. Next, input the files, including: the reference file in Fasta format (covering the wild-type genome reference sequence between Target-F and Target-R), optional target sequence file (with or without PAM sequence), and sequencing file (Fasta, Fastq, or bam format file) (Figure 3-3). The format of reference and target site files can be viewed by clicking the provided example files.

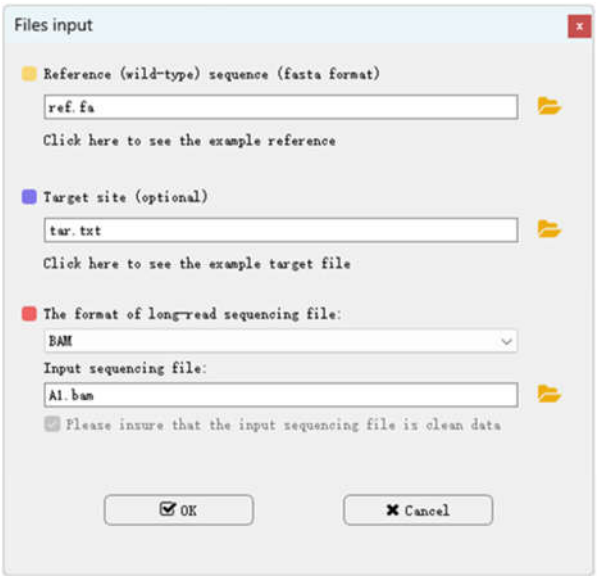

Figure 3-3 The interface of file input in LaDecode

ii) For 96 × n barcoding, click the 96 × n barcode type and double-click the “default_lad_pos.txt” and “default_lad_plate.txt” in the barcode setting panel to select the used position barcode and plate barcode. For 12×8 barcoding, click the 12×8 barcode type and double-click the “default_lad_col.txt” and “default_lad_row.txt” in the barcode setting panel to select the used column barcode and row barcode.

iii) Set the parameters for data analysis. LaDecode provides some built-in threshold values for different analysis models, including the diploid, polyploid, and low-frequency models, based on our laboratory tests. The low-frequency model is mainly used to identify mutations in cell lines, calli, or plant protoplasts; the output is all possible variations at the threshold. In addition to the preset threshold, users can adjust the threshold as needed. When the proportion of reads of a certain variant site in the sample is greater than the set threshold, the program considers the genotype as a realiable mutation and outputs the mutation site as well the frequency.

iv) Click the “Run” button on the toolbar or from “Run” menu to start the analysis task. It will take a few minutes to run the decoding.

(3) Analysis results

The analysis results of each well will be displayed in the decoding result panel. Yellowed position(s) indicate no result was produced for the sample(s) (Figure 3-4). When a sample is selected, the mutation result will be displayed in the result panel at the bottom, including the mutation type, total number of reads covered, sequence alignment, and supported read count of each allele. By draging the slider above, user can adjust the sequence range to be displayed.

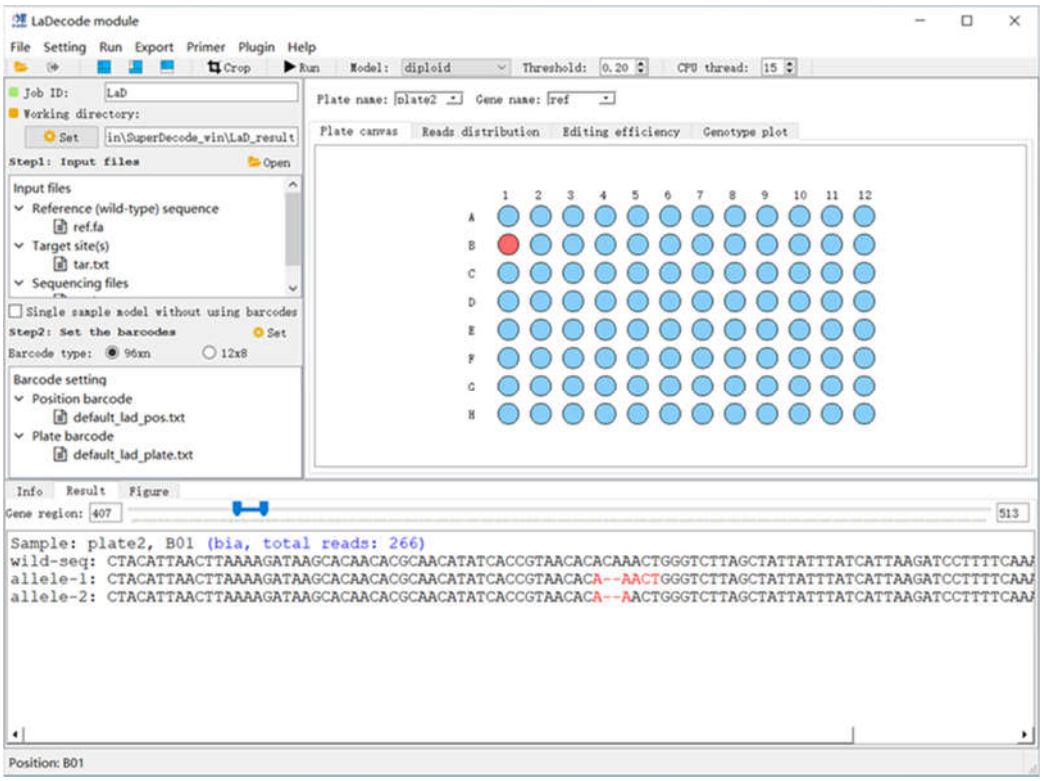

Figure 3-4 The result of LaDecode

LaDecode modules offer comprehensive analysis results that can be viewed in an interactive mode. After the analysis, plate canvas generates plots summarizing the distribution of covered reads, editing efficiency, and mutation genotype for the samples in the selected plate (Figure 3-5).

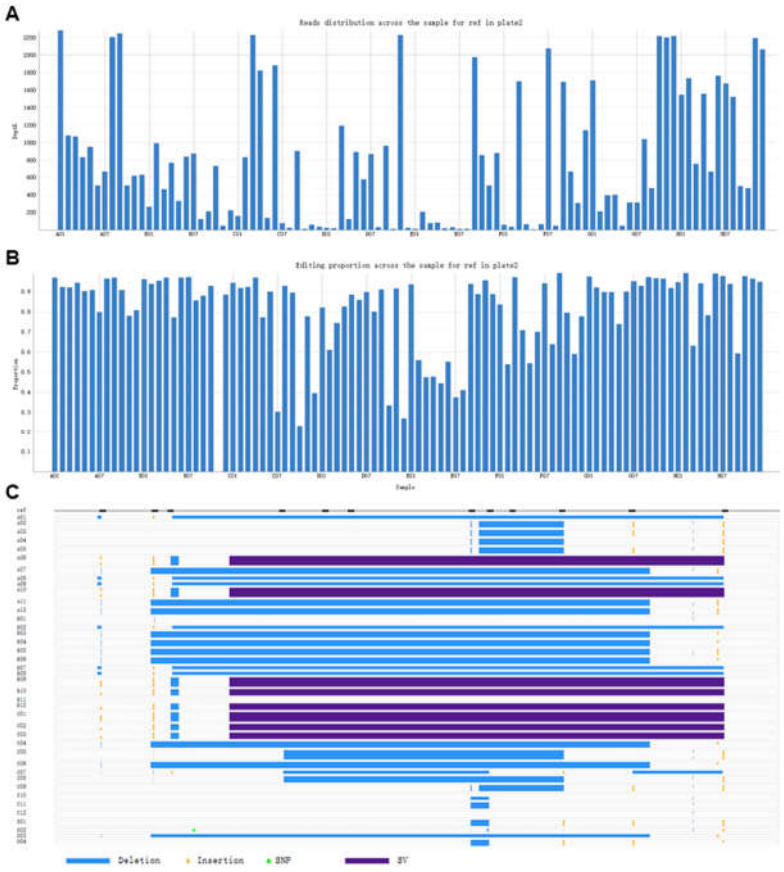

Figure 3-5 The reads distribution, editing efficiency and mutation genotype for the samples in the selected plate of LaDecode

Moreover, LaDecode modules generate figures depicting the mutation ratio, InDel size distribution, and mutation frequency at each position across the reference sequence for the selected sample (Figure 3-6).

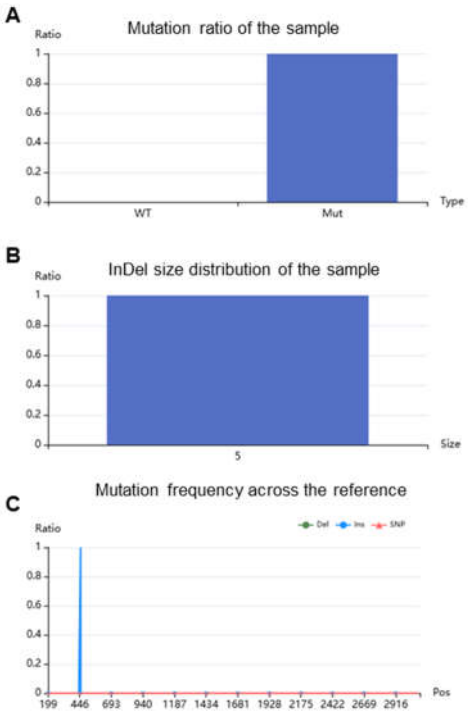

Figure 3-6 The mutation ratio, InDel size distribution, and mutation frequency at each position across the reference sequence for the selected sample of LaDecode

The results can be exported to an Excel file via the “Export” menu. Two types of result files can be exported. The function of “Export all results to excel file” is to export the detailed decoding results of all wells and plates (including the number of reads, mutant sequences, and the sequence alignments) (Figure 3-7), and “Fetch mutations to a single file” is to export simplified information, including mutant sequences, the count of supported reads, and mutant types.

plate2_B01		bia	266	(427 bp)GCACAACACGCAACATATCACCGTAACACACAAACTGGGTCTTAGCTATTATTTATCATTAAGATC(2840 bp)
	1	457:ACA>A;2938:AGA>A	138(0.51880)	(427 bp)GCACAACACGCAACATATCACCGTAACACA--AACTGGGTCTTAGCTATTATTTATCATTAAGATCTCTTTTCAAAAAACTGCTACT
	2	457:ACA>A	118(0.44361)	(427 bp)GCACAACACGCAACATATCACCGTAACACA--AACTGGGTCTTAGCTATTATTTATCATTAAGATC(2840 bp)
plate2_B02		hom	993	(164 bp)AATTTATTTTTGGCATTGCGATGGATTTTGATATCTCATTTACCGCAATAGCTTTGAGACATTTTGGCGGTTGGCAAGCTA(3089 bp)
	1	194:GATATCTCATTTACCGC	932(0.93857)	(164 bp)AATTTATTTTTGGCATTGCGATGGATTTTG----------GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTAT
plate2_B03		bia	470	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAATAGCTTTGAGACATTTTGGCGGTTGGCAAGCTATA(3179 bp)
	1	212:ATA>A;440:ACATATCAC	349(0.74255)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAA--GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTATATATATTG
	2	212:ATA>A;440:ACATATCAC	100(0.21277)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAA--GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTATATATATTG
plate2_B04		bia	765	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAATAGCTTTGAGACATTTTGGCGGTTGGCAAGCTATA(3179 bp)
	1	212:ATA>A;440:ACATATCAC	576(0.75294)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAA--GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTATATATATTG
	2	212:ATA>A;440:ACATATCAC	167(0.21830)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAA--GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTATATATATTG
plate2_B05		bia	333	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAATAGCTTTGAGACATTTTGGCGGTTGGCAAGCTATA(3179 bp)
	1	212:ATA>A;440:ACATATCAC	145(0.43544)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAA--GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTATATATATTG
	2	212:ATA>A;440:ACATATCAC	112(0.33634)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAA--GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTATATATATTG
plate2_B06		bia	836	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAATAGCTTTGAGACATTTTGGCGGTTGGCAAGCTATA(3179 bp)
	1	212:ATA>A;440:ACATATCAC	632(0.75598)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAA--GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTATATATATTG
	2	212:ATA>A;440:ACATATCAC	179(0.21411)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAA--GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTATATATATTG
plate2_B07		hom	875	(164 bp)AATTTATTTTTGGCATTGCGATGGATTTTGATATCTCATTTACCGCAATAGCTTTGAGACATTTTGGCGGTTGGCAAGCTA(3089 bp)
	1	194:GATATCTCATTTACCGC	850(0.97143)	(164 bp)AATTTATTTTTGGCATTGCGATGGATTTTG----------GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTAT
plate2_B08		hom	125	(164 bp)AATTTATTTTTGGCATTGCGATGGATTTTGATATCTCATTTACCGCAATAGCTTTGAGACATTTTGGCGGTTGGCAAGCTA(3089 bp)
	1	194:GATATCTCATTTACCGC	107(0.85600)	(164 bp)AATTTATTTTTGGCATTGCGATGGATTTTG----------GCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTAT
plate2_B09		chi	217	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAATAGCT----------
	1	453:A>AAA;533:AATTAAAGT	87(0.40092)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAATAGCT----------
	2	212:ATAGCT>AGCTATATAT	58(0.26728)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAAGCTATATATATATATATATATATATATATATAGAGAGAGAGAGAGAGAGATCT
	3	212:ATAGCT>AGCTATATAT	46(0.21198)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAAGCTATATATATATATATATATATATATATAGAGAGAGAGAGAGAGAGATCTGTC
plate2_B10		bia	730	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAATAGCT----------
	1	453:A>AAA;533:AATTAAAGT	477(0.65342)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAATAGCT----------
	2	212:ATAGCT>AGCTATATAT	201(0.27534)	(182 bp)CGATGGATTTTGATATCTCATTTACCGCAAGCTATATATATATATATATATATATATATATAGAGAGAGAGAGAGAGAGATCT
plate2_B11		wt	52	(179 bp)TTGCGATGGATTTTGATATCTCATTTACCGCAATAGCTTTGAGACATTTTGGCGGTTGGCAAGCTATATTTCTTTCAATCTATATAT
	1	wt	52(1.00000)	

Figure 3-7 Detailed mutation result table of LaDecode

In addition, the “Export mutation genotypes of all samples” in “Export” can output the genotype plot of all samples, allowing users to intuitively view the mutation results of the samples (Figure 3- 8).

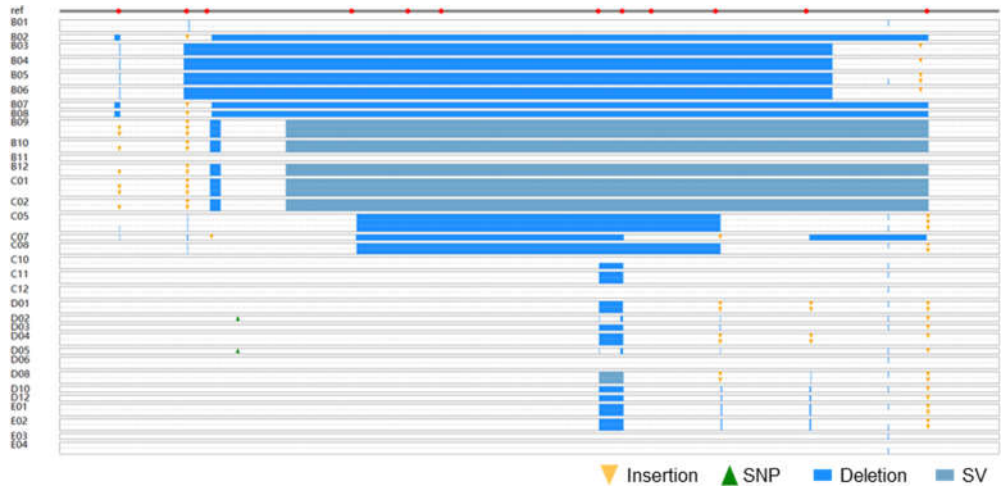

Figure 3-8 The mutation genotype plot of each sample in LaDecode

(4) Other handy functions

i) Mutation analysis from demultiplexed sequencing reads

Sometimes, sequencing companies can provide demultiplexed sequencing reads from pooled amplicon samples according to the user’s barcode sequence. To analyze the mutations

directly from the demultiplex files, LaDecode provides a plugin tool “Muataion analysis from demultiplexed fasta” from the “Plugin” menu. In the opened dialog, enter the wild-type reference sequence file, the optional target sequence file, and the directory path of the demultiplexed file (usually in Fasta format). Click the “OK” button for decoding analysis (Figure 3-9). The output result file is automatically saved to the same directory of the read files.

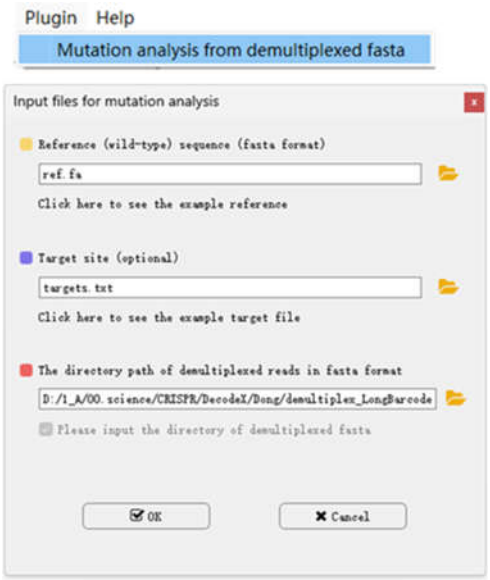

Figure 3-9 The interface of demultiplexed data analysis

ii) Generation of site-specific primers

LaDecode can automatically add the bridge sequences to the site-specific primers for first-round PCR. The program allows users to modify the bridge sequences by clicking the “edit” button (Figure 3-10). After entering the bridge sequence, the target name, and the specific sequence (5′-3′), click the “Generate primers” button to generate site-specific primers containing bridge sequences for the first-round PCR reaction. (Note: this function does not have the function to detect the specificity of the primer. It is recommended that use other tools such as the “Primer-BLAST” of NCBI or “primerDesign” of CRISPR-GE to detect the specificity of the primer.)

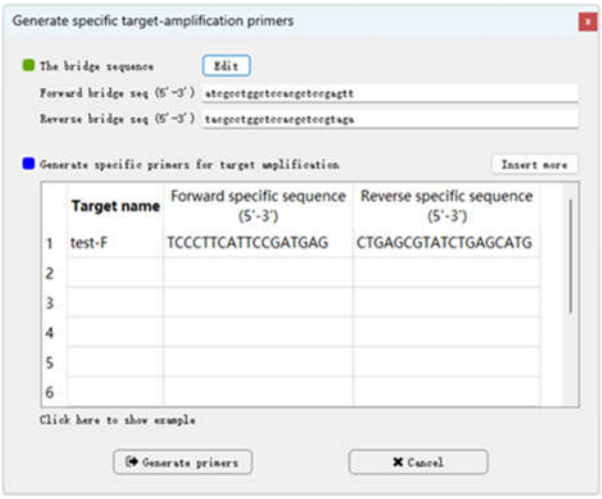

Figure 3-10 Primer generation dialog windows for LaDecode.

iii) Barcode setting

Insteading using the default postion and plate barcodes, users can define their own barcodes according to actual use during PCR.

For 96×n barcoding, open the barcode setting dialog by clicking the “Reset the default 96×n barcodes (define your own barcodes)” from the “Setting” menu to define new barcodes in batch. Alternatively, one by one modification is allowed via the barcode selection dock (Figure 3-11).

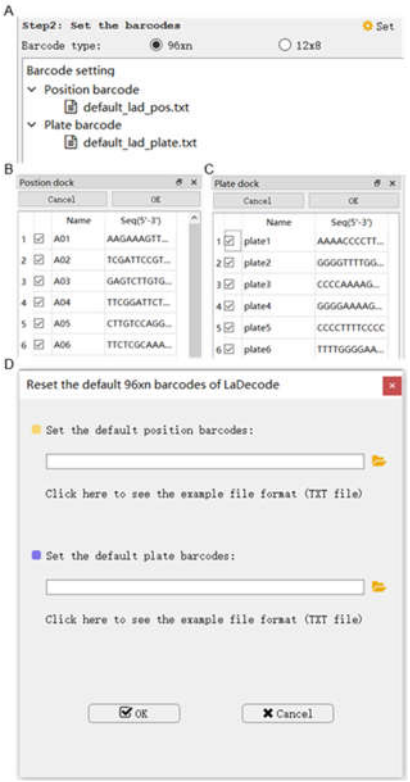

Figure 3-11 Barcode setting of Ladecode for 96×n barcoding

For 12×8 barcoding, open the barcode setting dialog by clicking the “Reset the default 12×8 barcodes (define your own barcodes)” from the “Setting” menu to define new barcodes in batch. Alternatively, one by one modification is allowed via the barcode selection dock (Figure 3-12).

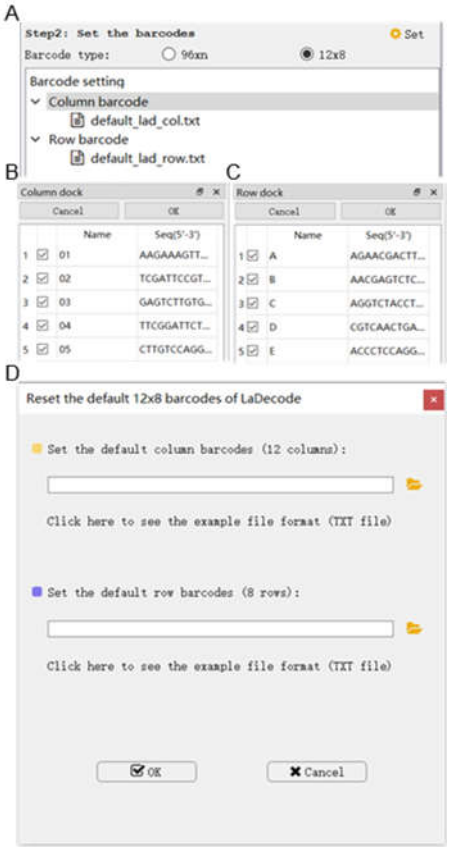

Figure 3-12 Barcode setting of Ladecode for 12×8 barcoding

iv) Mutation analysis from TGS data for a single sample

In addition to handling sequencing data from pooled amplicons, LaDecode is capable of analyzing TGS data from a single sample in which the target-site(s)-containing amplicons are directly subjected to sequencing without the barcoding step using the “Single sample model without using barcodes” function on the bottom of the file input and setting panel. In the dialog (Figure 3-13), input the reference sequence, target sequence (optional), and sequencing read files for starting the analysis.

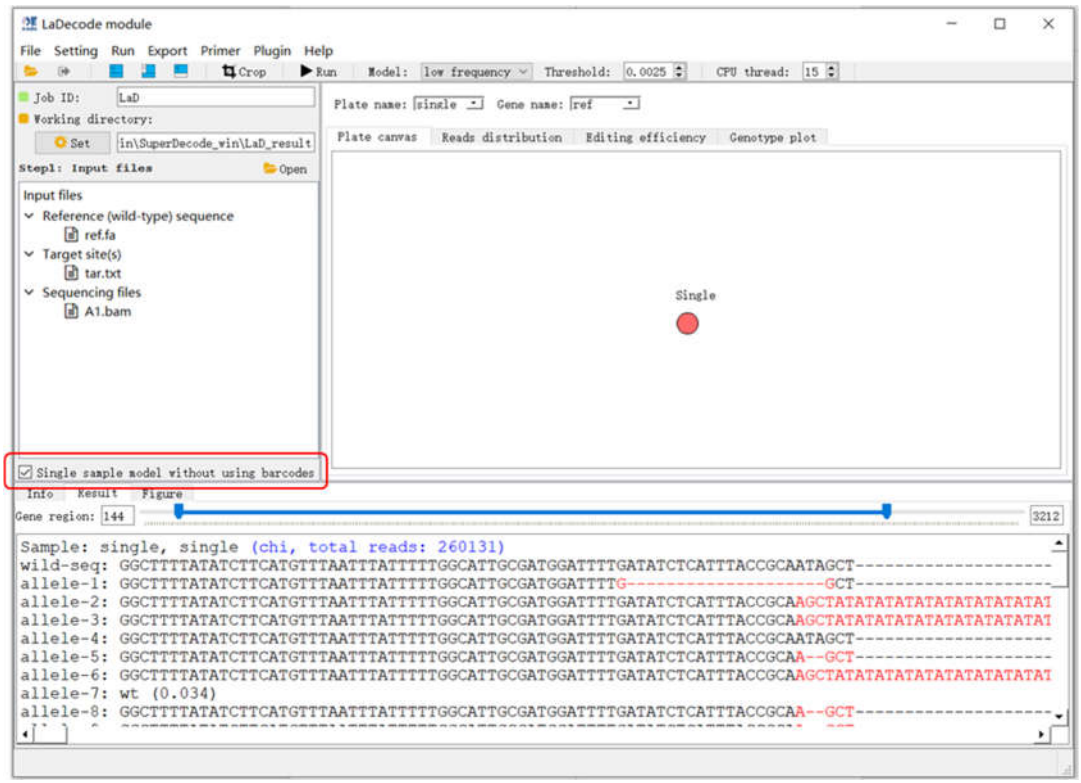

Figure 3-13 The mutation analysis from TGS data for a single sample

5. Mutation analysis using web-based software tool LaDecode

(1) The interface of web-based software tool LaDecode

The interface of web-based software tool LaDecode mainly includes a includes file input panel and barcode settings panel. (Figure 3-14).

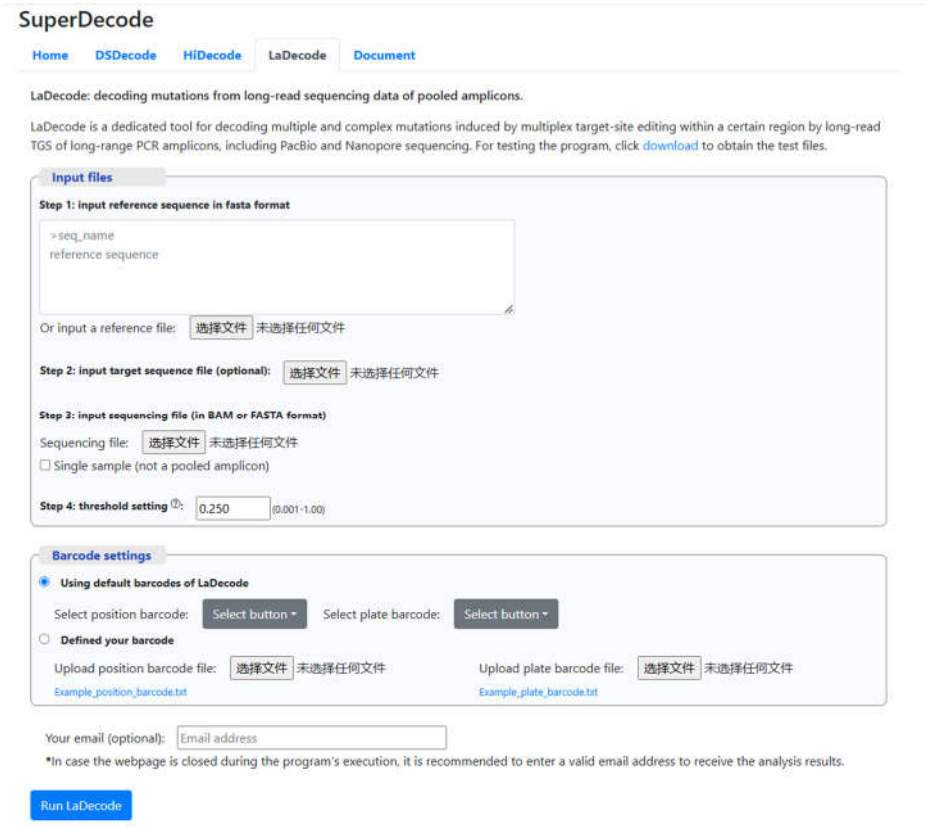

Figure 3-14 The interface of web-based software tool LaDecode

(2) Decoding operation of web-based software tool LaDecode

The use of web-based software tool LaDecode is mainly includes following four steps:

i) Input the files, including: the reference file in Fasta format (covering the wild-type genome sequence between Target-F and Target-R), optional target sequence file (with or without PAM sequence), and sequencing file (Fasta, Fastq, or bam format file).

ii) Setting the threshold and model of the task, if analysis the mutation from TGS data for a single sample, click the “Single sample (not a pooled amplicon)” button in the file input panel.

iii) Setting the barcode sequence in the barcode settings panel.

iv) Click the “Run LaDecode” button to start the analysis task. (Note: before starting the analysis task, it is recommendes to enter the email to receive the decoding results in the email input box.)

After decoding, the results will output in a new webpage (Figure 3-15), user can select the sample to display the decoding result, including the the sequence alignments, mutation efficiency, InDel size distribution, and mutation frequency at each position across the reference sequence for the selected sample. Users also can save the decoding results via click the download button.

Figure 3-15 The decoding result of web-based software tool LaDecode

- **Contact**

We highly appreciate your valuable suggestions for improving the experimental procedures and software packages. If you have any questions or suggestions, please contact Xianrong Xie (xiexianrong@scau.edu.cn)