Check CRISPR editing events in transgenic wheat with Next-Generation Sequencing

Junli Zhang, psjzhang@ucdavis.edu

Department of Plant Sciences, University of California, Davis, CA 95616, USA

Recently, more and more sequencing labs and companies provide CRISPR sequencing (or amplicon sequencing) services (*see* **Note 1**). PCR amplicons can be submitted and they will add barcodes and sequencing adapters to your samples before sequencing. Here we will describe how to use NGS to check the CRISPR editing efficiency in transgenic plants.

Materials

1.1 Used by all procedure

1. Pipettes and pipette tips (10 μL, 20 μL, 200 μL, 1 mL)

2. 1.5 mL Eppendorf tubes

3. Sterile deionized water

1.2 PCRs

1. PCR machine

2. Thin-walled PCR tubes

3. Taq polymerase buffer, 10×

4. Taq polymerase (commercial or homemade)

5. 50 mM $MgCl_2$

6. 10 μM dNTPs

7. 10 μM Forward primer with the left adapter

8. 10 μM Reverse primer with the right adapter

9. 10 μM Left barcodes

10. 10 μM Right barcodes

11. Isolated genomic DNA

1.3 PCR product recovery

1. Homemade beads (protocol can be found here: *https://ethanomics.files.wordpress.com/2012/08/serapure_v2-2.pdf* , you can also use the commercial Agencourt AMPure XP Beads (Beckman, A63881))

2. Freshly-prepared 70% ethanol: 1.0 ml per DNA sample

3. Elution buffer (10 mM Tris-HCl, pH 8.0): 30 µl per DNA pool

4. Magnet for 1.5 mL tubes

1.4 DNA concentration measurement

1. Qubit Fluorometer (version 2 or higher, ThermoFisher Scientific)

2. Qubit dsDNA HS Assay Kit (ThermoFisher Scientific, Q32851)

2 Methods

The CRISPR sequencing normally yield 50,000 to 100,000 reads per sample, our lab usually pool PCR amplicons of up to 300 sample. In order to pool multiple samples, barcodes need to be added to each sample via two rounds of PCRs (**Fig. 1**): the first around adds adapters to your PCR target (for priming of the 2nd round primers), and the 2nd round PCR adds barcodes to the ends of PCR amplicons. Most sequencing providers use Illumina PE150 (paired end 150 bp) for sequencing.

Fig. 1 Illustration of two rounds of PCR to add barcodes to samples before pooling

2.1 Design primers

1. PCR amplicon should be about 150 to 400 bp, ideally <300 bp to get overlap between the paired reads if using PE150, but at least the CRISPR editing site should be located within the first 100 base pairs (from either the 5'-end or the 3'-end) of the amplicon.

2. Primers can be designed to amplify homeologous genes in all sub-genomes.

3. Add adapters to the 5' end of the primers before ordering. (although you can design your own random adapters here, DO NOT use the Illumina NGS adapters which will be added later by the sequencing service provider).

```
Forward (Read1): 5’-TCCTCTGTCACGGAAGCG-Your forward primer -3’
Reverse (Read2): 5’-TTTAGCCTCCCCACCGAC-Your reverse primer -3’
```

4. Barcode primers (2nd round PCR): the total number of samples = [number of left barcode] x [number of right barcode]. For example, you can order 48 left barcode primers and 8 right barcode primers to be able to pool 48 x 8 = 384 samples. The 2nd round PCR primers only need to be ordered once. Usually, you just need to order more right barcode primers. See **File S1** for example barcodes.

```
Index 1 (XXXXXXXX: left barcoce):  5'-XXXXXXXX-TCCTCTGTCACGGAAGCG-3'

Index 2 (YYYYYYYY: right barcode): 5'-YYYYYYYY-TTTAGCCTCCCCACCGAC-3'
```

2.2 First-round PCR

Use the primers with adapters to do the first round PCR. Our lab has been using homemade Taq with the following PCR setup (**Table 1**).

Table 1: PCR setup for preparing NGS samples

Component	**1x (µL)**
H_2O	11.5
10x Buffer	2
10x dNTP	2
10x $MgCl_2$	2
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
Taq	0.5
DNA	1
Total	20

The PCR program is:

95 C 3 min
35 cycles of
    95 C 20 s
    55 C 20 s (depends on the Tm of your primers)
    72 C 30 s
72 C 5 min
12 C forever

2.3 Second-round PCR

First, do a quick PCR quality test. Check 5 µL of the 1st round PCR product on a 1.5% agarose gel to make sure you get the expected PCR product. Next, make a 1:10 dilution of the 1st round PCR and use 1 µL as the template of the 2nd round PCR. The setup is the same as the 1st round PCR above, except that you just add reverse barcode primers in the master mix and add the left barcode primers later, well by well. Use the PCR program:

95 C 3 min
15 cycles of
    95 C 20 s
    55 C 20 s
    72 C 30 s

72 C 5 min

12 C forever

After the 2nd round PCR, the PCR amplicons will look like this (52 bp longer than your PCR target):

```
5' - XXXXXXXX-TCCTCTGTCACGGAAGCG
- PCR target -
GTCGGTGGGGAGGCTAAA-YYYYYYYY - 3'
```

Do another quick PCR quality test before proceeding to the PCR clean-up step. Use 5 μL to check on a 1.5% agarose gel to verify that the PCR amplifications are in good quality.

2.4 PCR cleanup with homemade beads

1. After the 2nd round PCR, pool 2 μL of PCR product from each sample in a 1.5 mL tube (Note: may need to add more for low-efficiency primers based on the gel check). Make one pool for each plate. A whole plate will have 2 x 96 = 192 μL PCR product.
2. Add 0.9x volume of beads. Vortex well before using; a whole plate will need 192 x 0.9 = 172 μL beads. Mix thoroughly by pipetting up and down multiple times. Incubate on the bench for 5 min at room temperature.
3. Place the tube to the magnetic stand and let sit for 3 min until the solution is clear;
4. Remove all the supernatant by pipetting;
5. Wash with 500 μL of 70% ethanol and let sit for 30 sec;
6. Remove all the ethanol and the ethanol wash step (Step 5);
7. Remove all the ethanol and keep the tube on magnetic stand for about 5 min to dry all the leftover ethanol. Caution: Do not over dry the beads.
8. Remove the tube from the magnetic stand and add the elution buffer (30 μL of 10 mM Tris pH 8.0).
9. Mix well by pipetting a few times and let sit for 20 min;
10. Transfer the tubes to the magnetic stand for 2 min;
11. Collect the eluted DNA.
12. Load 2 μL on 1.5% agarose gel to check the cleaning results and DNA concentration [See **Note 2**].
13. For best results, measure the DNA concentration with Qubit instead of the Nanodrop which gives less accurate DNA concentrations. Mix equal amount of DNA from each tube for sequencing.

2.5 Data analysis

After the samples are sequenced, you will receive fastq files with all the reads. To get the reads for each sample in the pool, first demultiplex the fastq files and then check for editing events. All the analysis can be done in a web browser with web tools in the website *https://junli.netlify.app/apps/*.

2.5.1 Filter the raw output

1. Obtain the sequencing data from you service provider and get the fastq files ready (file names with extension .fastq or .fq or .fastq.gz or .fq.gz). You will get either an interleaved fastq (read 1 and read 2 are in the same file) or two separate files (read 1 and read 2 are separated, file names are like xxx_R1_001.fastq.gz and xxx_R2_001.fastq.gz).

2. Go to *https://junli.netlify.app/apps/filter-fastq-files-with-fastp/* (**Fig. 2**). This program uses software *fastp* **[3]** in the background.

3. Upload the raw fastq file(s). Check "**Interleaved PE?**" if you only receive one fastq file. Other options can remain default.

4. Next, go to **Step 2** and Click "**Start Filtering**" to start the program.

5. Once the filtering is completed, click the download button to retrieve the filtered fastq files and a summary html file. A summary about the filtering (e.g. read number before and after filtering) will also show up on the webpage.

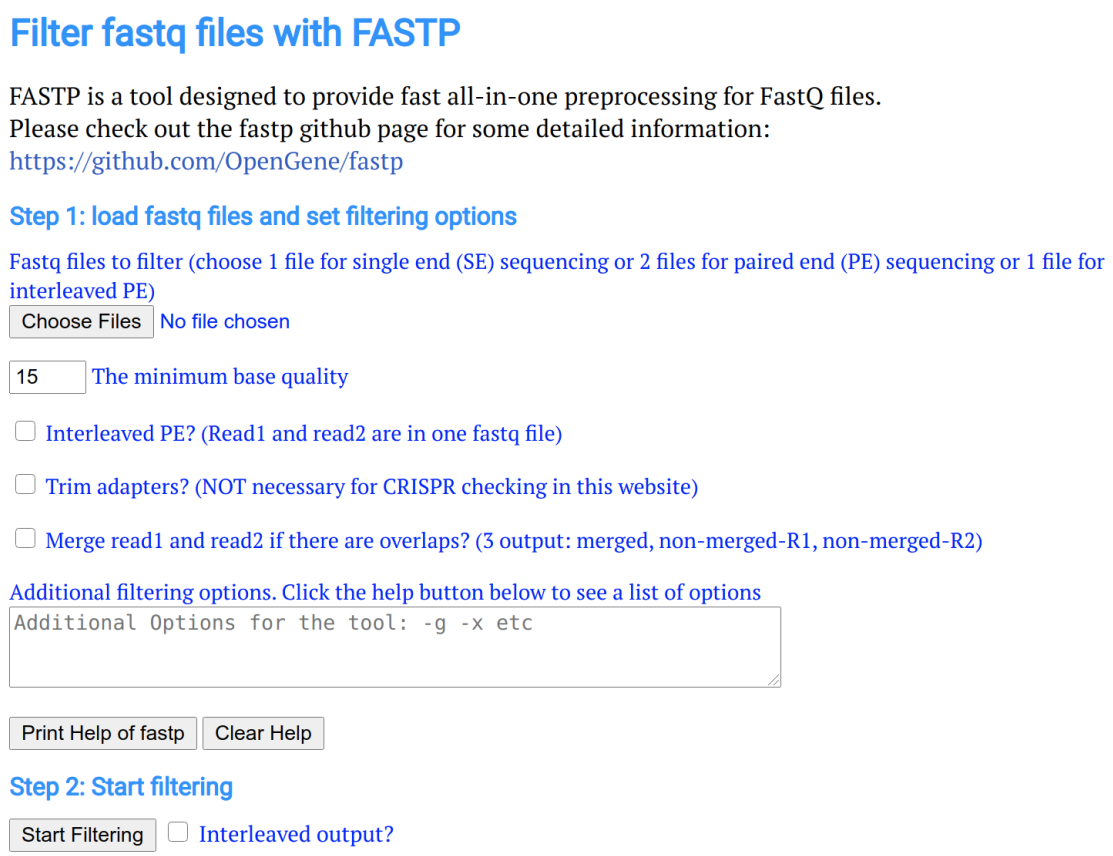

Fig. 2 The web app to filter fastq files (*https://junli.netlify.app/apps/filter-fastq-files-with-fastp/*)

2.5.2. Demultiplex the fastq file

1. Once you filtered the fastq files (I. Filter Raw Output), extract the downloaded zip file and prepare the filtered fastq file for demultiplexing.
2. Prepare the barcode file: a tab-delimited text file with 3 columns: sample ID, left barcode sequence, and right barcode sequence. The data can be prepared in an Excel spreadsheet.
3. Go to *https://junli.netlify.app/apps/demultiplex-a-fastq-file/* (**Fig. 3**)
4. Load your fastq files and barcode file. If the same barcode adapters listed above were used, just click "**Example Adapters**" to fill the adapter sequences.
5. Then, click "**Start Demultiplex**" to demultiplex your fastq file.
6. Download the demultiplexed fastq files when it is finished.

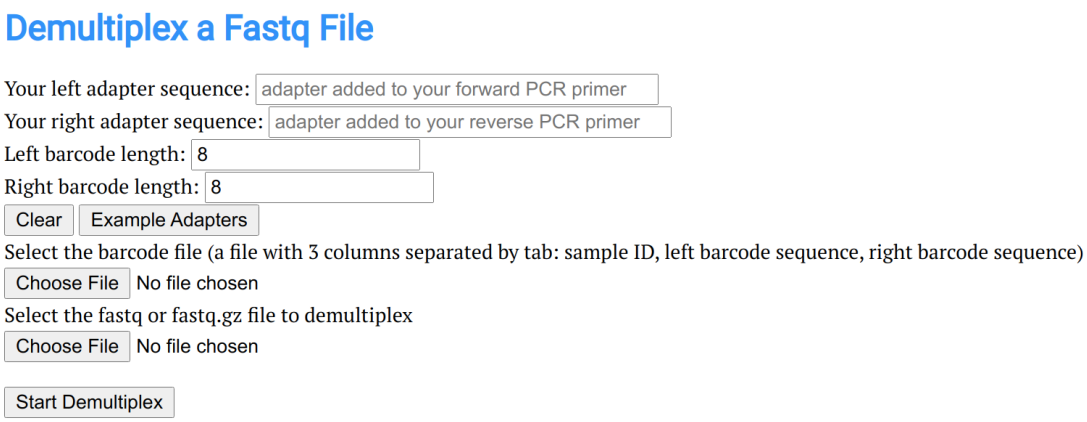

Fig. 3 The web app to demultiplex an interleaved fastq file (*https://junli.netlify.app/apps/demultiplex-a-fastq-file/*)

2.5.3. Make bam files for viewing the details in IGV

After getting the demultiplexed fastq.gz files, we can map all the reads to the PCR templates to review the real alignments of the reads (a bam file). The web app below uses BWA **[5]** and SAMTOOLS **[6]** to make indexed bam files and make variant calls from the bam files.

1. Prepare a fasta file including all your PCR templates for mapping reads.
2. Get ready of your demultiplexed fastq(.gz) files.
3. Go to *https://junli.netlify.app/apps/make-bam-files-with-bwa-and-samtools/* (**Fig. 4**)
4. At present, the compatible browsers for this tool are Firefox, Google Chrome, and the new version of Edge. Enable SIMD support based on the **Help** in the bottom of the web page.
5. This tool needs 3 inputs: a) the suffix of your R1 fastq files; most of the time it is "*_R1_001.fastq.gz*" (the default value); b) the fasta reference file in step 1, and c) all the fastq(.gz) files.

6. Click "**Map reads and make bam files**" to start. When completed, you will be prompted to download the zip file containing the bam files and indices. Depending on your data size, it may take some time. However, most of the CRISPR amplicon sequencing data only takes one or a few minutes.

7. You can view the bam files in *IGV* (**Fig. 5**), which can be downloaded from *https://software.broadinstitute.org/software/igv/download*. To load your reference: go to menu **Genomes** => Load Genome from File => select your fasta file used above; to load the bam files: go to menu **File** => Load from File => select one or more .bam files. Visit *IGV* website for more information.

Make bam files with BWA and SAMTOOLS

Before start: Please enable SIMD in your web brower [see Help below].

This tool is for **paired end** fastq files (for example, xxx_R1_001.fastq.gz and xxx_R2_001.fastq.gz). Please run the 3 steps below to get the indexed bams from a list of fastq files.

Provide the suffix of your read 1 (R1) fastq files:

_R1_001.fastq.gz

I. Choose reference file (a fasta file)

Choose File No file chosen

II. Choose demultiplexed fastq(.gz) files

Choose Files No file chosen

III. Map reads and create bam files

After loading the template fasta file and all the fastq files, now we will use `bwa` and `samtools` to create indexed bam files for viewing in the software IGV.

Map reads and Make bam files

Fig. 4 The web app to make indexed bam files (*https://junli.netlify.app/apps/make-bam-files-with-bwa-and-samtools/*)

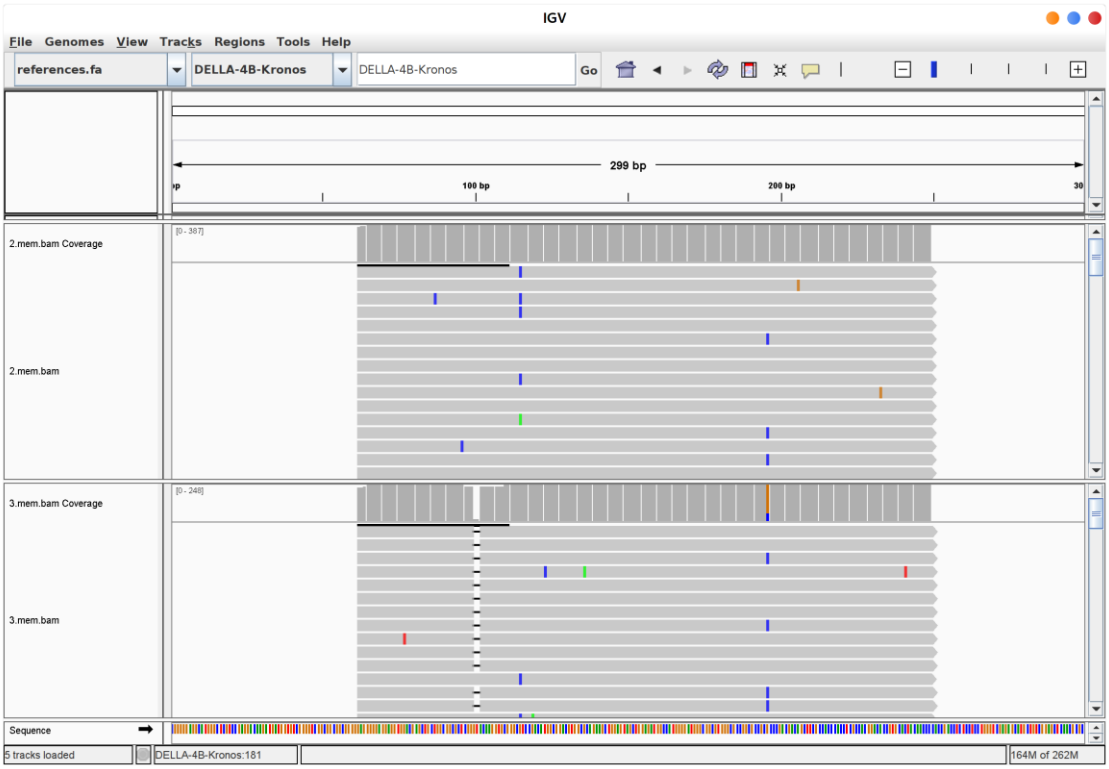

Fig. 5 Use IGV to check the alignments and editing details

3 Notes

1. We have been using the sequencing services in the U.S. provided by **MGH CCIB DNA Core** (*https://dnacore.mgh.harvard.edu/new-cgi-bin/site/pages/crispr_sequencing_main.jsp*) and **Genewiz** (*https://www.genewiz.com/Public/Services/Next-Generation-Sequencing/Amplicon-Sequencing-Services/Amplicon-EZ*).

2. If there are still dimers after the cleaning, you can clean up again with 0.8x of homemade beads. Adjust the final concentration and submit samples based on the instructions of your sequencing service provider. You can also use gel recovery to clean up your PCR product instead of beads.

3. All these analyses are done in your computer memory. It may freeze your browser if overloading, but there should be no problem for regular amplicon sequencing that normally yield 50,000 to 100,000 reads per sample. Refresh the page if you need to redo the analysis or there are problems.

4. There might be great variations on the number of reads for each individual in the pooled samples, which are caused by the DNA concentration variations among samples and primer efficiency differences. It might be not easy to normalize hundreds of DNA samples, but make sure to

collect a similar amount of tissue samples. If you need to pool PCR amplicons of different primers, you can pool the PCR product of each pair of primers first, measure the concentration of each pool after cleaning, and then pool an equal amount of DNA / primer. For example, you have two pairs of primers (P1 and P2). P1 amplified 40 samples and P2 amplified 56 samples. After cleaning, P1 pool has a concentration of 30 ng/ul and P2 pool has 40 ng/ul. You need to submit 25 ul DNA at 20 ng/ul (500 ng DNA totally). Then you will need 500 ng * 40 / (40 + 56) = 208 ng DNA from P1 and 500 - 208 = 292 ng from P2, that is 208 ng / 30 ng/ul = 6.9 ul of P1 DNAs and 292 ng / 40 ng/ul = 7.3 ul. Just add 25 - 6.9 - 7.3 = 10.8 ul of water.

5. Random mutations may be introduced during PCR especially using homemade Taq. That is okay for our indel check. If the same SNP shows up in almost all the reads, that is possibly a real variation between your material and the reference sequence.

6. Another useful tool the readers can try: *https://crispresso.pinellolab.partners.org/submission*.

References

1. Chen K, Wang Y, Zhang R, et al (2019) CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. Annual Review of Plant Biology 70:667–697.

2. Debernardi JM, Tricoli DM, Ercoli MF, et al (2020) A GRF–GIF chimeric protein improves the regeneration efficiency of transgenic plants. Nature Biotechnology 38:1274–1279.

3. Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890.

4. Connelly JP, Pruett-Miller SM (2019) CRIS.py: A Versatile and High-throughput Analysis Program for CRISPR-based Genome Editing. Scientific Reports 9:4194.

5. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

6. Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.