Ver.2.0
2019. July

Prepared by:
Iwate Biotechnology Research Center
Kitakami, Iwate
Japan

&

Ishikawa Prefectural University
Nonoichi, Ishikawa
Japan

Copyright (C) <2019>

Ishikwa Prefectural University
Iwate Biotechnology Research Center

This program is free software:
you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

See the GNU General Public License for more details.
You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

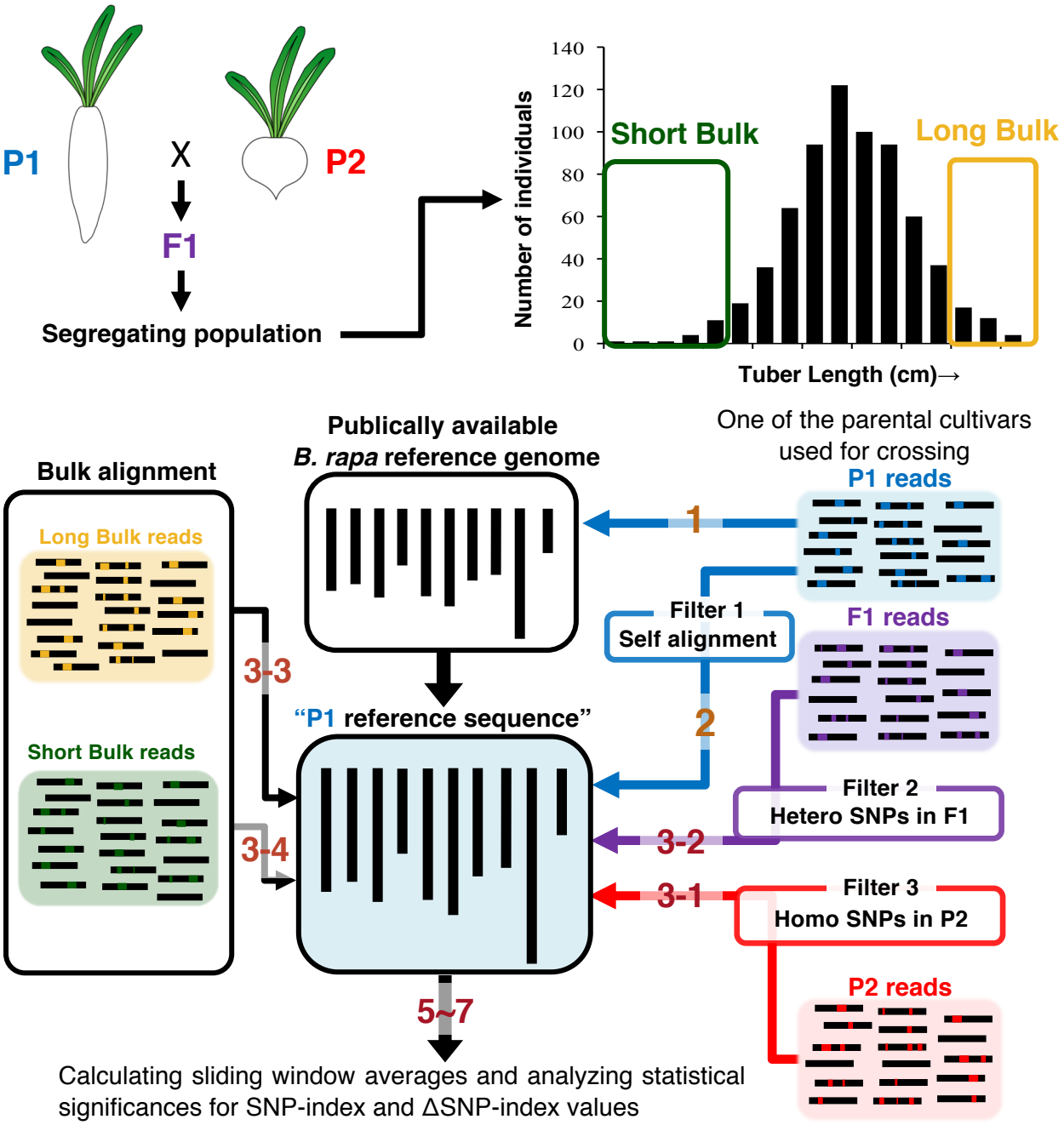
Preparation

- Install the following programs:
    1. Perl (v5.24.2)
    2. R (version 2.14.1)
    3. BWA (version 0.7.15-r1140)
    4. SAMtools (1.4.1 or after)

QTL-seq Pipeline

The identification of candidate QTL by the QTL-seq method involves following steps:

0. **Setup**
1. **Developing a "P1 reference sequence" for one of the parental cultivar/line (P1) used for crossing**
2. **Calculating SNP-index by aligning P1 sequence reads to "P1 reference sequence"**
3. **Calculating SNP-index by aligning bulk samples, F1 and P2 sequence reads to "P1 reference sequence"**
4. **Simulation test**
5. **Filtering by self alignment and sliding window analysis.**
6. **Filtering by F1 alignment and sliding window analysis.**
7. **Filtering by P2 alignment and sliding window analysis.**



1. 1_Q_Bat_Run_Parent_A_align_topublic_fasta.sh
2. 2_Q_Bat_Run_Parent_A_align_to_Parent_A_ref.sh
3-1. 3_Q_Bat_Run_Parent_B_align_to_Parent_A.sh
3-2. 3_Q_Bat_Run_Short_align_to_Parent_A.sh
3-3. 3_Q_Bat_Run_Long_align_to_Parent_A.sh
3-4. 3_Q_Bat_Run_F1_align_to_Parent_A.sh
5. 5_Q_Bat_Run_compare_Short_Long.sh
6. 6_Q_Bat_compare_Short_Long_select_F1_hetero_position.sh
7. 7_Q_Bat_compare_Short_Long_filtered_B_Parent_homo_SNP.sh

0. Setup

This step involves in setting reads and parameter for analysis.

Working place

myhome

(#) replace ”myhome” with the proper pathname according to your system environment

Works 1/4

- Install the pipeline framework to your system

ex)

```
$    cd myhome
$    unzip QTL-seq_v2.zip
$    mv QTL-seq_V2 QTL-seq_v2_test
```

0. Setup

Working place

myhome/QTL_seq_v2_test/read_ PATH.txt

Works 2/4

Input absolute path of your FASTQ file compressed in gzip file to proper position in “read_ PATH.txt” with tab separation. The multiple FASTQ file is allowed for the analysis.

E.x. <read_PATH.txt >

Input P1

xxx/xxx/aaa_R1.gz	xxx/xxx/aaa_R2.gz	The sequence read for P1
xxx/xxx/bbb_R1.gz	xxx/xxx/bbb_R2.gz	

Input P2

xxx/xxx/ccc_R1.gz	xxx/xxx/ccc_R2.gz	The sequence read for P2 If P2 sequence reads is not available, input dummy data.
xxx/xxx/ddd_R1.gz	xxx/xxx/ddd_R2.gz	

Input A_bulk

xxx/xxx/eee_R1.gz	xxx/xxx/eee_R2.gz	The sequence read for A bulk
xxx/xxx/fff_R1.gz	xxx/xxx/fff_R2.gz	

Input B_bulk

xxx/xxx/ggg_R1.gz	xxx/xxx/ggg_R2.gz	The sequence read for B bulk
xxx/xxx/hhh_R1.gz	xxx/xxx/hhh_R2.gz	

Input F1

xxx/xxx/iii_R1.gz	xxx/xxx/iii_R2.gz	The sequence read for F1
xxx/xxx/jjj_R1.gz	xxx/xxx/jjj_R2.gz	

First reads | Second reads

Input paired end data with tab separation

0. Setup

Working place

myhome/QTL-seq_v2_test/0_devloping_all_program_files.sh

Works 3/4

Input the parameter used for analysis in “0_devloping_all_program_files.sh”

<0_devloping_all_program_files.sh> の例

```
#! /bin/sh
#$ -S /bin/sh
#$ -cwd

used_cpu= 4

samtools_PATH="/usr/local/bin/samtools"
bwa_PATH=“/usr/local/bin/bwa"

P1_name=“Parent_A”
P2_name=“Parent_B”
A_bulk_name= “Short”
B_bulk_name= “Tall”

public_ref=“myhome/public.fasta"

filtered_depth= 10
```

- used_cpu= 4 → The number of the used core
- samtools_PATH="/usr/local/bin/samtools" / bwa_PATH=“/usr/local/bin/bwa" → Absolute path of samtools and bwa software
- P1_name=“Parent_A” / P2_name=“Parent_B” / A_bulk_name= “Short” / B_bulk_name= “Tall” → The used name for P1, P2, A_bulk and B_bulk
- public_ref=“myhome/public.fasta" → Absolute path of publically available reference genome (FASTA file)
- filtered_depth= 10 → The minimum depth for analyzed SNP positions (>8 recommended)

…continued

0. Setup

Working place

myhome/QTL_seq_v2_test/ 0_devloping_all_program_files.sh

Works 3/4

Input the parameter used for analysis in “0_devloping_all_program_files.sh”

<0_devloping_all_program_files.sh> の例

individual_number= 20	The number of individuals used for bulked samples
reprication= 10000	The number of times for simulation
population_structure= F2 #F2 or RIL	The used segregating population
filtered_mapping_score_in_bam=10	The minimum Mapping score (calculated by BWA) used for analysis
filtered_insert_size_in_bam= 10000	The minimum distance of the mapped paired end read used for analysis
window_size_Mb= 2	The window size (Mb) used for sliding window analysis
step_size_kb= 50	The step size (Mb) used for sliding window analysis
howmany_snp_number= 10 #----------	The minimum number of SNPs within the analyzed window for sliding window analysis.

0. Setup

Working place

myhome/QTL_seq_v2_test

Works 4/4

Run “0_devloping_all_program_files.sh”

```
$     cd myhome/QTL-seq_v2_test
$     ./0_devloping_all_program_files.sh
```

0. Setup

Result

Confirm the files described in red color. Each files shoud have the name defined in 0_devloping_all_program_files.sh.

<QTL-seq_v2_test>

```
├── 0_devloping_all_program_files.sh
├── 1_Q_Bat_Run_Parent_A_align_topublic_fasta.sh
├── 2_Q_Bat_Run_Parent_A_align_to_Parent_A_ref.sh
├── 3_Q_Bat_Run_Parent_B_align_to_Parent_A.sh
├── 3_Q_Bat_Run_Short_align_to_Parent_A.sh
├── 3_Q_Bat_Run_Long_align_to_Parent_A.sh
├── 3_Q_Bat_Run_F1_align_to_Parent_A.sh
├── 4_Q_Bat_Run_simulation.sh
├── 4_Q_Bat_Run_compare_depth_with_Parent_A.sh
├── 5_Q_Bat_Run_compare_Short_Long.sh
├── 6_Q_Bat_compare_Short_Long_select_F1_hetero_position.sh
├── 7_Q_Bat_compare_Short_Long_filtered_B_Parent_homo_SNP.sh
├── public.fasta_length.txt
├── original_scripts
│   └── Bat_file
├── read_information
│   ├── Parent_A_reads.txt
│   ├── Parent_B_reads.txt
│   ├── Short_reads.txt
│   ├── Long_reads.txt
│   └── F1_reads.txt
├── read_PATH.txt
├── script
└── simulation_result
```

1. Developing a “P1 reference sequence” for one of the parental cultivar/line (P1) used for crossing

In this step,
1: The SNPs are detected by alignment the P1 sequence reads to the publically available reference genome by BWA.

2: “P1 Reference sequence” is developed by replacing the detected SNPs with those from the publically available reference genome.

Working place

myhome/QTL-seq_v2_test

Works 1/1

- Run “1_Q_Bat_Run_Parent_A_align_topublic_fasta.sh”

```
$     cd myhome/QTL-seq_v2_test
$     ./1_Q_Bat_Run_Parent_A_align_topublic_fasta.sh
```

1. Developing a “P1 reference sequence” for one of the parental cultivar/line (P1) used for crossing

Result 1/2

Confirm the files described in red color.

<QTL-seq_v2_test>

```
⋮
├── script
├── simulation_result
├── Parent_A_ref_seq_development
│       ├── Parent_A_ref.fa.fai
│       ├── Parent_A_ref.fa.sa
│       ├── Parent_A_ref.fa.pac
│       ├── Parent_A_ref.fa.ann
│       ├── Parent_A_ref.fa.amb
│       ├── Parent_A_ref.fa.bwt
│       ├── Parent_A_ref.fa *1
│       ├── Parent_A_all_mapped_sort.pileup
│       ├── m_Parent_A_all_mapped_sort.pileup *2
│       ├── Parent_A_all_mapped_sort_indel.pileup
│       ├── Parent_A_all_mapped_sort.bam.bai
│       ├── Parent_A_all_mapped_sort.bam *3
│       ├── Parent_A_all_mapped.bam
│       ├── Parent_A_all_unmapped.bam
│       ├── Parent_A_all_merge_rmdup.bam
│       ├── Parent_A_all_merge.bam
│       ├── f_Parent_A_all_temp.bam
│       ├── ex_for_sam_development_Parent_A_reads.txt
│       ├── ex_for_bam_development_Parent_A_reads.txt
│       ├── ex_for_alignment_Parent_A_reads.txt *4
```

*1 Parent A (P1) reference sequence developed in this step

*2 The data for SNPs replaced with the publicly available genome for developing Parent A (P1) ref erence (Pileup Format type 1, p28).

*3 The alignment data of the sequence reads of Parent A to the publicly available genome

*4 The information of the sequence reads of Parent A used for alignment to the publicly available genome.

2. Calculating SNP-index by aligning P1 sequence reads to "P1 reference sequence"

In this step,
The sequence reads used for developoing "P1 reference sequence" are re-aligned to the developed "P1 reference sequence" to detect additional SNPs. Such additional SNPs should be either arise from misalignment or are due to heterozygous genomic regions in P1. It is recommended that these SNPs are excluded from QTL-seq analysis.

Working place

myhome/QTL-seq_v2_test

Works 2/2

Run "2_Q_Bat_Run_Parent_A_align_to_Parent_A_ref.sh"

```
$     cd myhome/QTL-seq_v2_test
$     ./2_Q_Bat_Run_Parent_A_align_to_Parent_A_ref.sh
```

2. Calculating SNP-index by aligning P1 sequence reads to "P1 reference sequence"

Result 2/2

Confirm the files described in red color.

<QTL-seq_v2_test>

```
:
:
:
:
:
:
├── simulation_result
├── Parent_A_ref_seq_development
├── Parent_A_self_alignment
│       ├── Parent_A_all_mapped_sort_cover_ratio.txt *1
│       ├── Parent_A_all_mapped_sort_depth.txt *2
│       ├── Parent_A_all_mapped_sort.bam.bai *3
│       ├── Parent_A_all_mapped_sort.bam *3
│       ├── Parent_A_all_unmapped.bam
│       ├── Parent_A_all_merge_rmdup.bam
│       ├── Parent_A_all_merge.bam
│       ├── f_Parent_A_all_temp.bam
│       ├── ex_for_sam_development_Parent_A_reads.txt
│       ├── ex_for_bam_development_Parent_A_reads.txt
│       ├── ex_for_alignment_Parent_A_reads.txt *4
```

*1 The cover ratio of aligned sequence reads in Parent A (P1) reference sequence
*2 The average depth of the aligned region in Parent A (P1) reference sequence
*3 The alignment data of the sequence reads of Parent A to Parent A (P1) reference sequence
*4 The information of the sequence reads of Parent A used for alignment to Parent A (P1) reference sequence

3. Calculating SNP-index by aligning bulk samples, F1 and P2 sequence reads to “P1 reference sequence”

In this step,
1: The SNPs are detected by alignment the bulk samples, F1 and P2 sequence reads to the publically available reference genome by BWA.

2: Calculating SNP-index value in each alignment data.

Working place

myhome/QTL-seq_v2_test

Works

Run “3_Q_Bat_Run_Short_align_to_Parent_A.sh”
Run “3_Q_Bat_Run_Long_align_to_Parent_A.sh”
Run “3_Q_Bat_Run_F1_align_to_Parent_A.sh”
Run “3_Q_Bat_Run_Parent_B_align_to_Parent_A.sh”

```
$    cd myhome/QTL-seq_v2_test
$    ./3_Q_Bat_Run_Parent_B_align_to_Parent_A.sh
$    ./3_Q_Bat_Run_Short_align_to_Parent_A.sh
$    ./3_Q_Bat_Run_Long_align_to_Parent_A.sh
$    ./3_Q_Bat_Run_F1_align_to_Parent_A.sh
```

3. Calculating SNP-index by aligning bulk samples, F1 and P2 sequence reads to “P1 reference sequence”

Result

Confirm the developed new directory (folder) described in red color.

<QTL-seq_v2_test>

├── Parent_A_ref_seq_development
├── Parent_A_self_alignment
├── Parent_B_to_Parent_A_alignment
├── Short_to_Parent_A_alignment
├── Long_to_Parent_A_alignment
├── F1_to_Parent_A_alignment

3. Calculating SNP-index by aligning bulk samples, F1 and P2 sequence reads to “P1 reference sequence”

Result 1/3

Confirm the files described in red color.

<QTL-seq_v2_test>

```
⋮
├── F1_to_Parent_A_alignment
├── Short_to_Parent_A_alignment
│       ├── Short_all_mapped_sort_cover_ratio.txt *1
│       ├── Short_all_mapped_sort_depth.txt*2
│       ├── Short_all_mapped_sort.bam *3
│       ├── Short_all_mapped_sort.bam.bai *3
│       ├── Short_all_merge_rmdup.bam
│       ├── Short_all_merge.bam
│       ├── Short_all_unmapped.bam
│       ├── f_Short_all_temp.bam
│       ├── ex_for_sam_development_Short_reads.txt
│       ├── ex_for_bam_development_Short_reads.txt
│       ├── ex_for_alignment_Short_reads.txt *4
```

*1 The cover ratio of aligned sequence reads in Parent A (P1) reference sequence
*2 The average depth of the aligned region in Parent A (P1) reference sequence
*3 The alignment data of the sequence reads of short bulk to Parent A (P1) reference sequence
*4 The information of the sequence reads of short bulk used for alignment to Parent A (P1) reference sequence

※ “Long_to_Parent_A_alignment” is similar format.

3. Calculating SNP-index by aligning bulk samples, F1 and P2 sequence reads to "P1 reference sequence"

Result 2/3

Confirm the files described in red color.

<QTL-seq_v2_test>

```
⋮
├── Parent_B_to_Parent_A_alignment
├── F1_to_Parent_A_alignment
|       ├── Parent_A_vs_F1.pileup *1
|       ├── m_Parent_A_vs_F1.pileup *2
|       ├── Parent_A_vs_F1_indel.pileup *3
|       ├── F1_all_mapped_sort_cover_ratio.txt *4
|       ├── F1_all_mapped_sort_depth.txt *5
|       ├── F1_all_mapped_sort.bam.bai
|       ├── F1_all_mapped_sort.bam *6
|       ├── F1_all_unmapped.bam
|       ├── F1_all_merge_rmdup.bam
|       ├── F1_all_merge.bam
|       ├── f_F1_all_temp.bam
|       ├── ex_for_sam_development_F1_reads.txt
|       ├── ex_for_bam_development_F1_reads.txt
|       ├── ex_for_alignment_F1_reads.txt *7
|       ├── Fishire_test_m_Parent_A_vs_F1.pileup *8
|       ├── sim_Fishire_test_m_Parent_A_vs_F1.pileup *9
```

*1 All SNP positions detected by aligning F1 sequence reads to Parent A (P1) reference sequence.

*2 The SNP positions showing SNP-index = 1 among the SNPs of *1.

*3 The short insertion and deletion detected by aligning F1 sequence reads to Parent A (P1) reference sequence.

*4 The cover ratio of aligned sequence reads in Parent A (P1) reference sequence

*5 The average depth of the aligned region in Parent A (P1) reference sequence

*6 The alignment data of the sequence reads of F1 to Parent A (P1) reference sequence

*7 The information of the sequence reads of F1 used for alignment to Parent A (P1)

*8 The SNP positions selected from *1 by *P* value <0.01 of Fishers exact test.

*9 Among of *8 SNP positions, the SNP positions selected by 95% confidence interval of F1 simulation test.

3. Calculating SNP-index by aligning bulk samples, F1 and P2 sequence reads to "P1 reference sequence"

Result 3/3

Confirm the files described in red color.

<QTL-seq_v2_test>

```
⋮
├── Parent_A_self_alignment
├── Parent_B_to_Parent_A_alignment
│       ├── Parent_A_vs_Parent_B.pileup *1
│       ├── m_Parent_A_vs_Parenr_B.pileup *2
│       ├── Parent_A_vs_Parent_B_indel.pileup *3
│       ├── Parent_B_all_mapped_sort_cover_ratio.txt *4
│       ├── Parent_B_all_mapped_sort_depth.txt *5
│       ├── Parent_B_all_mapped_sort.bam.bai
│       ├── Parent_B_all_mapped_sort.bam *6
│       ├── Parent_B_all_mapped.bam
│       ├── Parent_B_all_unmapped.bam
│       ├── Parent_B_all_merge_rmdup.bam
│       ├── Parent_B_all_merge.bam
│       ├── f_Parent_B_all_temp.bam
│       ├── ex_for_sam_development_Parent_B_reads.txt
│       ├── ex_for_bam_development_Parent_B_reads.txt
│       ├── ex_for_alignment_Parent_B_reads.txt *7
```

*1 All SNP positions detected by aligning Parent B (P2) sequence reads to Parent A (P1) reference sequence.
*2 The SNP positions showing SNP-index = 1 among the SNPs of *1.
*3 The short insertion and deletion detected by aligning Parent B (P2) sequence reads to Parent A (P1) reference sequence.
*4 The cover ratio of aligned sequence reads in Parent A (P1) reference sequence
*5 The average depth of the aligned region in Parent A (P1) reference sequence
*6 The alignment data of the sequence reads of Parent B (P2) to Parent A (P1) reference sequence
*7 The information of the sequence reads of Parent B (P2) used for alignment to Parent A (P1)

4. Simulation test

In this step,
Calculating 95 and 99% confidence interval of SNP-index value by simulation test under the null hypothesis of randomly bulked DNA.

Working place

myhome/QTL-seq_v2_test

Works

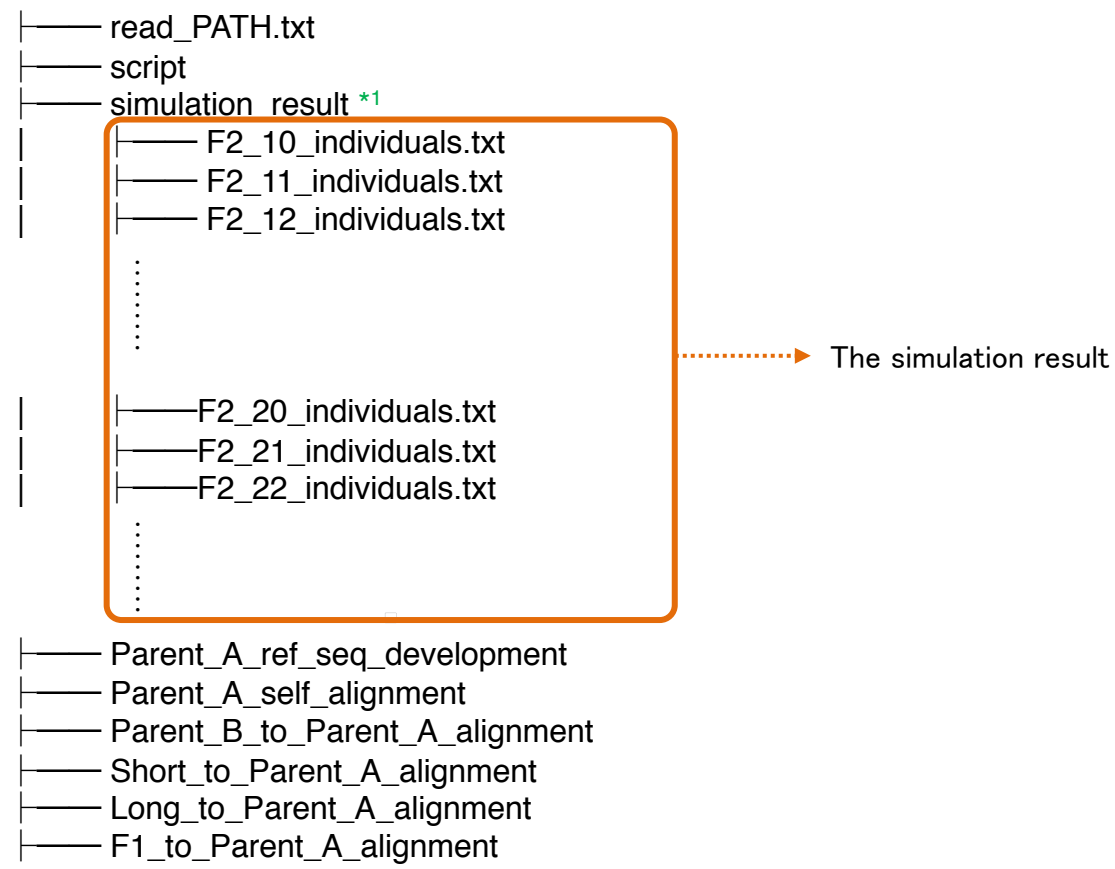
Run “4_Q_Bat_Run_simulation.sh”

```
$     cd myhome/QTL-seq_v2_test
$     ./4_Q_Bat_Run_simulation.sh
```

4. Simulation test

Result

Confirm the existence of the simulation result for the used bulk samples.



*1 If you input 20 individuals and F2 progeny type in the parameter at step 0, the file named “F2_20_individual.txt” should be existed.

5. Filtering by self alignment and sliding window analysis.

In this step,
1: Selecting SNPs by self alignment data

2: Applying sliding window analysis

3: Drawing graphs for SNP-index plot

Working place

myhome/QTL-seq_v2_test

Works

Run “5_Q_Bat_Run_compare_Short_Long.sh”

```
$     cd myhome/QTL-seq_v2_test
$     ./5_Q_Bat_Run_compare_Short_Long.sh
```

5. Filtering by self alignment and sliding window analysis.

Result

Confirm the files described in red color.

```
<QTL-seq_v2>
 ⋮
├── Long_to_Parent_A_alignment
├── F1_to_Parent_A_alignment
├── compare_Short_Long
│   ├── Parent_A_Short_Long_indel.pileup *1
│   ├── Parent_A_Short_Long.pileup *2
│   ├── m_Parent_A_Short_Long.pileup *3
│   ├── s_m_Parent_A_Short_Long.pileup *4
│   ├── sliding_window_2_Mb_s_m_Parent_A_Short_Long.pileup *5
│   ├── graph_Short_sliding_window_2_Mb_s_m_Parent_A_Short_Long.pileup.png *6
│   ├── graph_Long_sliding_window_2_Mb_s_m_Parent_A_Short_Long.pileup.png *7
│   ├── graph_Delta_Short_Long_sliding_window_2_Mb_s_m_Parent_A_Short_Long.
        pileup.png *8
```

*1 The short insertion and deletion detected by aligning sequence reads of bulk samples to Parent A (P1) reference sequence. (Pileup Format type 3 selected SNP and indel information, p32).

*2 The SNP positions eliminated the spurious SNPs by self-alignment based filter. (Pileup Format type 3, p31).

*3 Among of *2 SNP positions, the SNP positions exceeding depth decided at step 0 (Pileup Format type 4, p33).

*4 The SNP positions of *3 added confidence interval of SNP-index value by simulation test under the null hypothesis of randomly bulked DNA (Pileup Format type 5, p34).

*5 The data from sliding window analysis with given window size and step size (Sliding window average file, p35).

*6 The of SNP-index value of short bulk sample.

*7 The of SNP-index value of long bulk sample.

*8 The plot ΔSNP-index by subtracting the SNP-index values of the Short bulk from the Long bulk) and subtraction of SNP-index bulk sample.

* The details explanation of plot is described in pXXX~XXX.

6. Filtering by F1 alignment and sliding window analysis.

In this step,
1: Selecting SNPs by F1 alignment data

2: Applying sliding window analysis

3: Drawing graphs for SNP-index plot

Working place

myhome/QTL-seq_v2_test

Works

Run “6_Q_Bat_compare_Short_Long_select_F1_hetero_position.sh”.

```
$     cd myhome/QTL-seq_v2_test
$     ./6_Q_Bat_compare_Short_Long_select_F1_hetero_position.sh
```

6. Filtering by F1 alignment and sliding window analysis.

Result

Confirm the files described in red color.

<QTL-seq_v2_test>

```
:
:
:
:
├── F1_to_Parent_A_alignment
├── compare_Short_Long
├── compare_Short_Long_select_F1_hetero_SNP
|       ├── common_F1_hetero_snp_Parent_A_Short_Long.pileup *1
|       ├── m_common_F1_hetero_snp_Parent_A_Short_Long.pileup *2
|       ├── rm_F1_hetero_snp_Parent_A_Short_Long.pileup
|       ├── s_m_common_F1_hetero_snp_Parent_A_Short_Long.pileup *3
|       ├── sliding_window_2_Mb_s_m_common_F1_hetero_snp_Parent_A_Short_Lon
|           g.pileup *4
|       ├── graph_Short_sliding_window_2_Mb_s_m_common_F1_hetero_snp_Parent_
|           A_Short_Long.pileup.png *5
|       ├── graph_Long_sliding_window_2_Mb_s_m_common_F1_hetero_snp_Parent_
|           A_Short_Long.pileup.png *6
|       ├── graph_Delta_Short_Long_sliding_window_2_Mb_s_m_common_F1_hetero_
|           snp_Parent_A_Short_Long.pileup.png *7
```

*1 The SNP data eliminated the spurious SNPs from "compare_Short_Long/Parent_A_Short_Long.pileup" by F1 alignment based filter. (Pileup Format type 3, p31).

*2 Among of *1 SNP positions, the SNP positions exceeding depth decided at step 0 (Pileup Format type 4, p33).

*3 The SNP positions of *2 added confidence interval of SNP-index value by simulation test under the null hypothesis of randomly bulked DNA (Pileup Format type 5, p34).

*4 The data from sliding window analysis with given window size and step size (Sliding window average file, p35).

*5 The of SNP-index value of short bulk sample.

*6 The of SNP-index value of long bulk sample.

*7 The plot ΔSNP-index by subtracting the SNP-index values of the Short bulk from the Long bulk) and subtraction of SNP-index bulk sample.

* The details explanation of plot is described in in p36-38.

7. Filtering by P2 alignment and sliding window analysis.

In this step,
1: Selecting SNPs by P2 alignment data

2: Applying sliding window analysis

3: Drawing graphs for SNP-index plot

Working place

myhome/QTL-seq_v2

Works

Run “7_Q_Bat_compare_Short_Long_filtered_Parent_B_homo_SNP.sh”

```
$     cd myhome/QTL-seq_v2_test
$     ./7_Q_Bat_compare_Short_Long_filtered_Parent_B_homo_SNP.sh
```

7. Filtering by P2 alignment and sliding window analysis.

Result

Confirm the files described in red color.

```
<QTL-seq_v2_test>
 ⋮
 ├── compare_Short_Long
 ├── compare_Short_Long_select_F1_hetero_SNP
 ├── compare_Short_Long_filtered_Parent_B_homo_F1_hetero_SNP
 │   ├── common_Parent_B_homo_snp_common_F1_hetero_snp_Parent_A_Short_
         Long.pileup *1
 │   ├── m_common_Parent_B_homo_snp_common_F1_hetero_snp_Parent_A_Short
         _Long.pileup*2
 │   ├── rm_Parent_B_homo_snp_common_F1_hetero_snp_Parent_A_Short_Long.pile
         up
 │   ├── s_m_common_Parent_B_homo_snp_common_F1_hetero_snp_Parent_A_Sh
         ort_Long.pileup *3
 │   ├── sliding_window_2_Mb_s_m_common_Parent_B_homo_snp_common_F1_he
         tero_snp_Parent_A_Short_Long.pileup *4
 │   ├── graph_Short_sliding_window_2_Mb_s_m_common_Parent_B_homo_snp_co
         mmon_F1_hetero_snp_Parent_A_Short_Long.pileup.png*5
 │   ├── graph_Long_sliding_window_2_Mb_s_m_common_Parent_B_homo_snp_com
         mon_F1_hetero_snp_Parent_A_Short_Long.pileup.png*6
 │   ├── graph_Delta_Short_Long_sliding_window_2_Mb_s_m_common_Parent_B_ho
         mo_snp_common_F1_hetero_snp_Parent_A_Short_Long.pileup *7
```

*1 The SNP data eliminated the spurious SNPs from "compare_Short_Long/Parent_A_Short_Long.pileup" by P2 alignment based filter. (Pileup Format type 3, p31).

*2 Among of *1 SNP positions, the SNP positions exceeding depth decided at step 0 (Pileup Format type 4, p33).

*3 The SNP positions of *2 added confidence interval of SNP-index value by simulation test under the null hypothesis of randomly bulked DNA (Pileup Format type 5, p34).

*4 The data from sliding window analysis with given window size and step size (Sliding window average file, p35).

*5 The of SNP-index value of short bulk sample.

*6 The of SNP-index value of long bulk sample.

*7 The plot ΔSNP-index by subtracting the SNP-index values of the Short bulk from the Long bulk) and subtraction of SNP-index bulk sample.

* The details explanation of plot is described in p36-38.

File format

Pileup Format type 1

Ex) <Parent_A_ref_seq_development>

├—— **Parent_A_all_mapped_sort.pileup**

├—— **m_Parent_A_all_mapped_sort.pileup**

These file format is a TAB-delimited format with each data line consists of the following fields:

field#	Description
1	chromosome
2	coordinate
3	reference base
4	consensus base
5	consensus Depth
6	read bases
7	base qualities
8	SNP-index

filed #1 – 8 : based samtools pileup format (see the samtools manual)

Pileup Format type 1 selected SNP and indel information

Ex) < Parent_A_ref_seq_development >

├—— **Parent_A_all_mapped_sort_indel.pileup**

These file format is a TAB-delimited format with each data line consists of the following fields:

field#	Description
1	chromosome
2	coordinate
3	reference base
4	consensus Depth
5	read bases
6	base qualities
7	indel index

filed #1 – 7 : based samtools pileup format (see the samtools manual)

File format

Pileup Format type 2

Ex) < Parent_B_to_Parent_A_alignment >

├─── **Parent_A_vs_Parent_B.pileup**

├─── **m_Parent_A_vs_Parent_B.pileup**

< F1_to_Parent_A_alignment >

├─── **Parent_A_vs_F1.pileup**

├─── **m_Parent_A_vs_F1.pileup**

These file format is a TAB-delimited format with each data line consists of the following fields:

field#	Description
1	chromosome
2	coordinate
3	reference base
4	consensus base
5	reference Depth
6	reference read bases
7	reference base qualities
8	consensus Depth
9	consensus read bases
10	consensus base qualities
11	SNP-Index
12	The number of reads having SNP
13	The number of reads having same nucleotide with reference sequence

filed #1 – 13 : based samtools pileup format (see the samtools manual)

File format

Pileup Format type 2 selected SNP and indel information

Ex) < Parent_B_to_Parent_A_alignment >
├—— **Parent_A_vs_Parent_B_indel.pileup**
< F1_to_Parent_A_alignment >
├—— **Parent_A_vs_F1_indel.pileup**

These file format is a TAB-delimited format with each data line consists of the following fields:

field#	Description
1	chromosome
2	coordinate
3	reference base
4	reference Depth
5	reference read bases
6	base qualities
7	consensus Depth
8	consensus read bases
9	consensus base qualities
10	indel index
11	The number of reads having shot insertion and deletion
12	The number of reads having same nucleotide with reference sequnce

filed #1 – 12 : based samtools pileup format (see the samtools manual)

File format

Pileup Format type 3

Ex) < compare_Short_Long >
├── **Parent_A_Short_Long.pileup**

< compare_Short_Long_select_F1_hetero_SNP >
├── **common_Parent_B_homo_snp_Parent_A_Short_Long.pileup**
├── **rm_common_Parent_B_homo_snp_Parent_A_Short_Long.pileup**

< compare_Short_Long_filtered_with_F1_hetero_Parent_B_homo_SNP >
├── **common_Parent_B_homo_snp_common_F1_hetero_snp_Parent_A_Short_Long.pileup**
├── **rm_Parent_B_homo_snp_common_F1_hetero_snp_Parent_A_Short_Long.pileup**

These file format is a TAB-delimited format with each data line consists of the following fields:

field#	Description
1	chromosome
2	coordinate
3	reference base
4	reference Depth
5	Short bulk base
6	Short bulk Depth
7	Short bulk SNP-index
8	Long bulk base
9	Long bulk Depth
10	Long bulk SNP-index

filed #1 – 10: based samtools pileup format (see the samtools manual)

File format

Pileup Format type 3 selected SNP and indel information

Ex) < compare_Short_Long >

├—— **Parent_A_Short_Long_indel.pileup**

These file format is a TAB-delimited format with each data line consists of the following fields:

field#	Description
1	chromosome
2	coordinate
3	reference base
4	reference Depth
5	Short bulk base
6	Short bulk Depth
7	Short bulk indel-index
8	Long bulk base
9	Long bulk Depth
10	Long bulk indel-index
11	consensus base ?
12	indel-index ?

filed #1 – 10: based samtools pileup format (see the samtools manual)

File format

Pileup Format type 4

Ex) < compare_Short_Long >
├—— **m_Parent_A_Short_Long.pileup**

< compare_ Short_Long_select_F1_hetero_SNP >
├—— **m_common_F1_hetero_snp_Parent_A_Short_Long.pileup**

< compare_Short_Long_filtered_with_F1_hetero_Parent_B_homo_SNP >
├—— **m_common_Parent_B_homo_F1_hetero_snp_Parent_A_Short_Long.pileup**

These file format is a TAB-delimited format with each data line consists of the following fields:

field#	Description
1	chromosome
2	coordinate
3	reference Depth
4	Short bulk Depth
5	Short bulk SNP-index
6	Long bulk Depth
7	Long bulk SNP-index

filed #1 - 7: based samtools pileup format (see the samtools manual)

File format

Pileup Format type 5

Ex) <compare_Short_Long>
├── **s_m_Parent_A_Short_Long.pileup**

<compare_ Short_Long_select_F1_hetero_SNP>
├── **s_m_common_F1_hetero_snp_Parent_A_Short_Long.pileup**

<compare_Short_Long_filtered_with_F1_hetero_Parent_B_homo_SNP>
├── **s_m_common_Parent_B_homo_F1_hetero_snp_Parent_A_Short_Long.pileup**

These file format is a TAB-delimited format with each data line consists of the following fields:

field#	Description
1	chromosome
2	coordinate
3	reference Depth
4	Short bulk Depth
5	Short bulk SNP-index
6	Long bulk Depth
7	Long bulk SNP-index
8	95% confidence interval lower side of ΔSNP-index
9	95% confidence interval upper side of ΔSNP-index
10	99% confidence interval lower side of ΔSNP-index
11	99% confidence interval upper side of ΔSNP-index
12	95% confidence interval lower side of SNP_index for Short bulk
13	95% confidence interval upper side of SNP_index for Short bulk
14	99% confidence interval lower side of SNP_index for Short bulk
15	99% confidence interval upper side of SNP_index for Short bulk
16	95% confidence interval lower side of SNP_index for Long bulk
17	95% confidence interval upper side of SNP_index for Long bulk
18	99% confidence interval lower side of SNP_index for Long bulk
19	99% confidence interval upper side of SNP_index for Long bulk
20	95% confidence interval lower side of ΔSNP-index
21	95% confidence interval upper side of ΔSNP-index
22	99% confidence interval lower side of ΔSNP-index
23	99% confidence interval upper side of ΔSNP-index

filed #1 - 12: based samtools pileup format (see the samtools manual)
filed #8 - 11: 2つのバルクのうち、Depth が小さい方のシミュレーション結果
filed #20 - 23: 2つのバルクのうち、Depth が大きい方のシミュレーション結果

File format

Sliding window average File

Ex) <compare_Short_Long>
├── **sliding_window_2_Mb_s_m_Parent_A_Short_Long.pileup**

<compare_ Short_Long_select_F1_hetero_SNP>
├── **sliding_window_2_Mb_s_m_common_F1_hetero_snp_Parent_A_Short_Long.pileup**

<compare_Short_Long_filtered_with_F1_hetero_Parent_B_homo_SNP>
├── **sliding_window_2_Mb_s_m_common_Parent_B_homo_F1_hetero_snp_Parent_A_Short_Long.pileup**

These file format is a TAB-delimited format with each data line consists of the following fields:

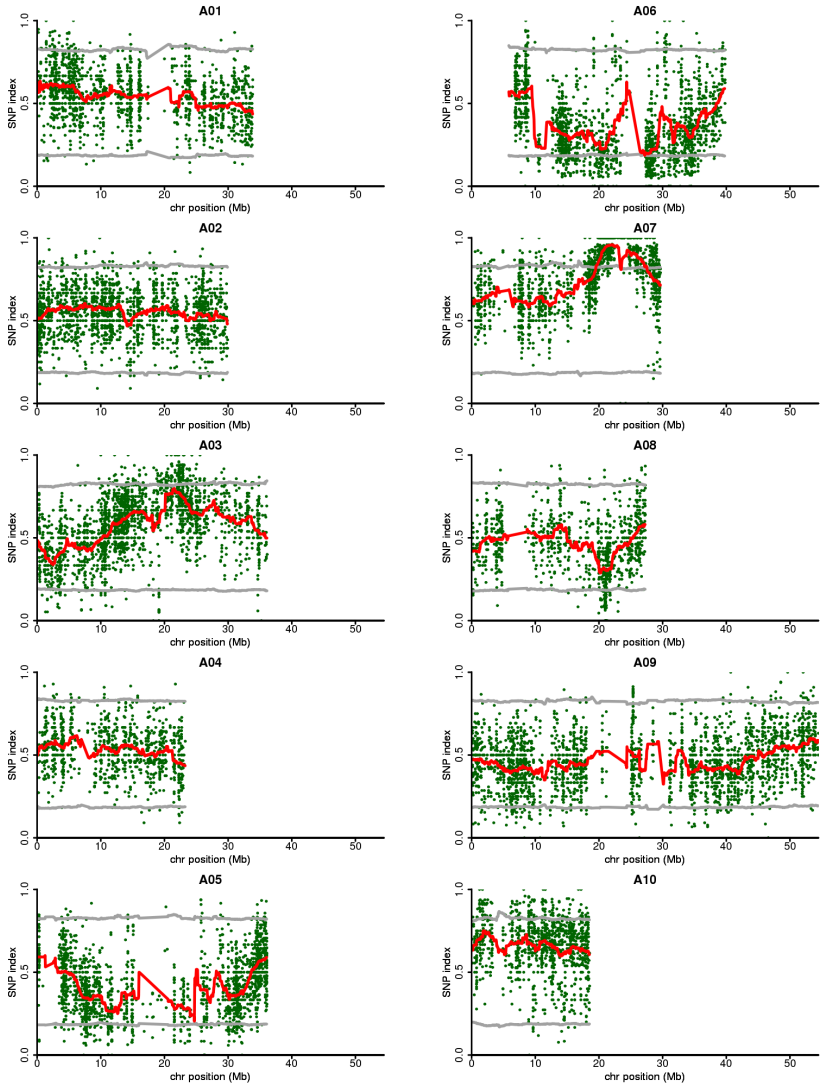
field#	Description
1	chromosome
2	coordinate
3	The average of SNP-index for Short bulk
4	The average of SNP-index for Short bulk
5	The average of SNP-index for Short ΔSNP-index
6	The average of 95% confidence interval lower side of ΔSNP-index
7	The average of 95% confidence interval upper side of ΔSNP-index
8	The average of 99% confidence interval lower side of ΔSNP-index
9	The average of 99% confidence interval upper side of ΔSNP-index
10	The average of 95% confidence interval lower side of SNP_index for Short bulk
11	The average of 95% confidence interval upper side of SNP_index for Short bulk
12	The average of 99% confidence interval lower side of SNP_index for Short bulk
13	The average of 99% confidence interval upper side of SNP_index for Short bulk
14	The average of 95% confidence interval lower side of SNP_index for Long bulk
15	The average of 95% confidence interval upper side of SNP_index for Long bulk
16	The average of 99% confidence interval lower side of SNP_index for Long bulk
17	The average of 99% confidence interval upper side of SNP_index for Long bulk

File format

Graph 2 : SNP-index distribution for bulk A sample.

Ex) ├ QTL-seq_v2_test
│ ├ compare_Short_Long_filtered_with_F1_hetero_Parent_B_homo_SNP
│ ├ **graph_Short_sliding_window_2_Mb_s_m_common_Parent_B_homo_snp_common_F1_hetero_snp_Parent_A_Short_Long.pileup.png**

This graph is drawn with both pileup format type 5 (p34) and sliding window average file (p35) .



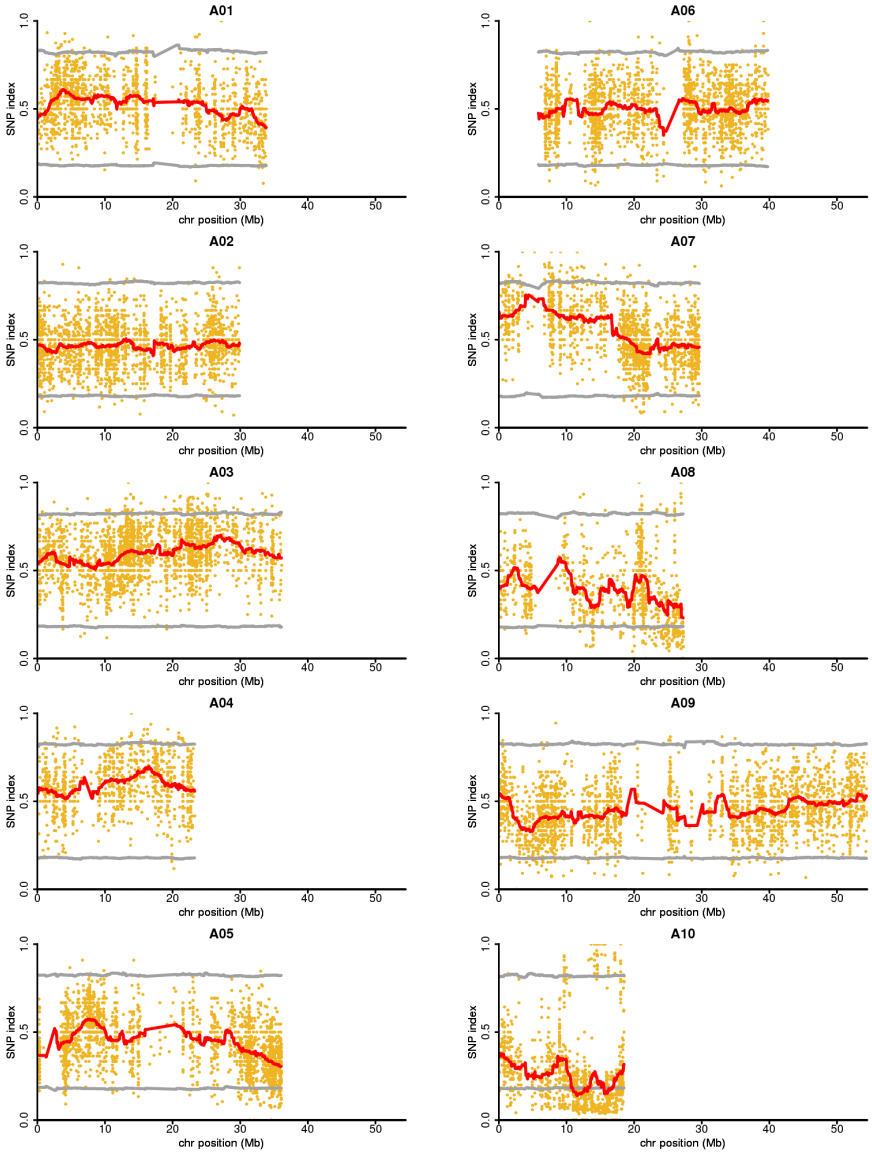
Green dot	SNP-index for Short bulk (The 4th row in pileup format type 5, p34)
Red line	Sliding window average of SNP-index for Short bulk (The 3th row in sliding window average file, p35)
Grey line	Sliding window average of 95%-confidence interval upper/lower side (The 9th and 10th row in sliding window average file, p35)

File format

Graph 3 : SNP-index distribution for bulk B sample.

Ex) ├ QTL-seq_v2_test
| ├ compare_Short_Long_filtered_with_F1_hetero_Parent_B_homo_SNP
| ├ **graph_Long_sliding_window_2_Mb_s_m_common_Parent_B_homo_snp_common_F1_hetero_snp_Parent_A_Short_Long.pileup.png**

This graph is drawn with both pileup format type 5 (p34) and sliding window average file (p35) .



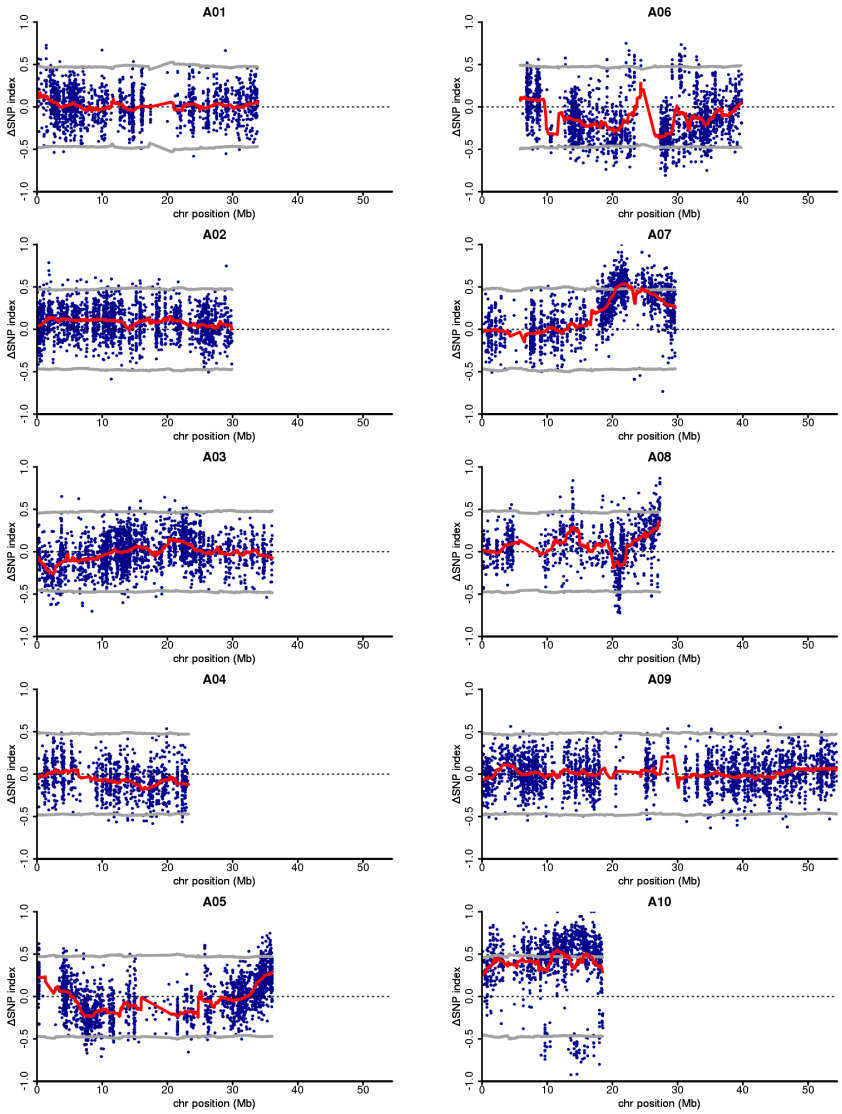
Orange dot	SNP-index for Long bulk (The 5th row in pileup format type 5, p34)
Red line	Sliding window average of SNP-index for Long bulk (The 4th row in sliding window average file, p35)
Grey line	Sliding window average of 95%-confidence interval upper/lower side (The 9th and 10th row in sliding window average file, p35)

File format

Graph 1 : QTL-seq v2.0 filter 3 result

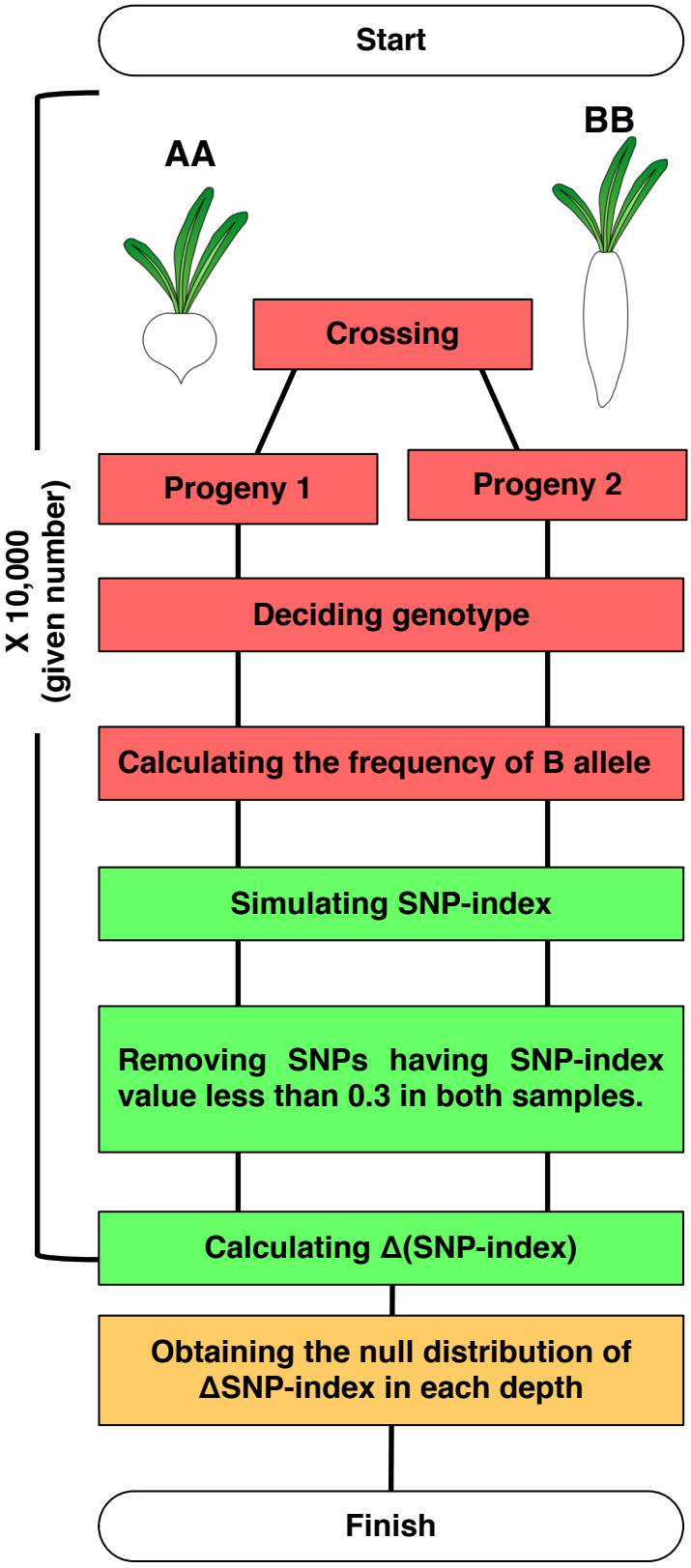
Ex)　├ QTL-seq_v2_test
　　│　├ compare_Short_Long_filtered_with_F1_hetero_Parent_B_homo_SNP
　　│　　├ **graph_Delta_Short_Long_sliding_window_2_Mb_s_m_common_Parent_B_homo_snp_common_F1_hetero_snp_Parent_A_Short_Long.pileup.png**

This graph is drawn with both pileup format type 5 (p34) and sliding window average file (p35) .



Blue dot	Δ(SNP-index) (The 6th row in pileup format type 5, p34)
Red line	Sliding window average for Δ(SNP-index) (The 5th row in sliding window average file, p35)
Grey line	Sliding window average of 95%-confidence interval upper/lower side (The 9th and 10th row in sliding window average file, p35)

Simulation test



Flow chart of simulation test for null distribution.

Simulation for random sampling

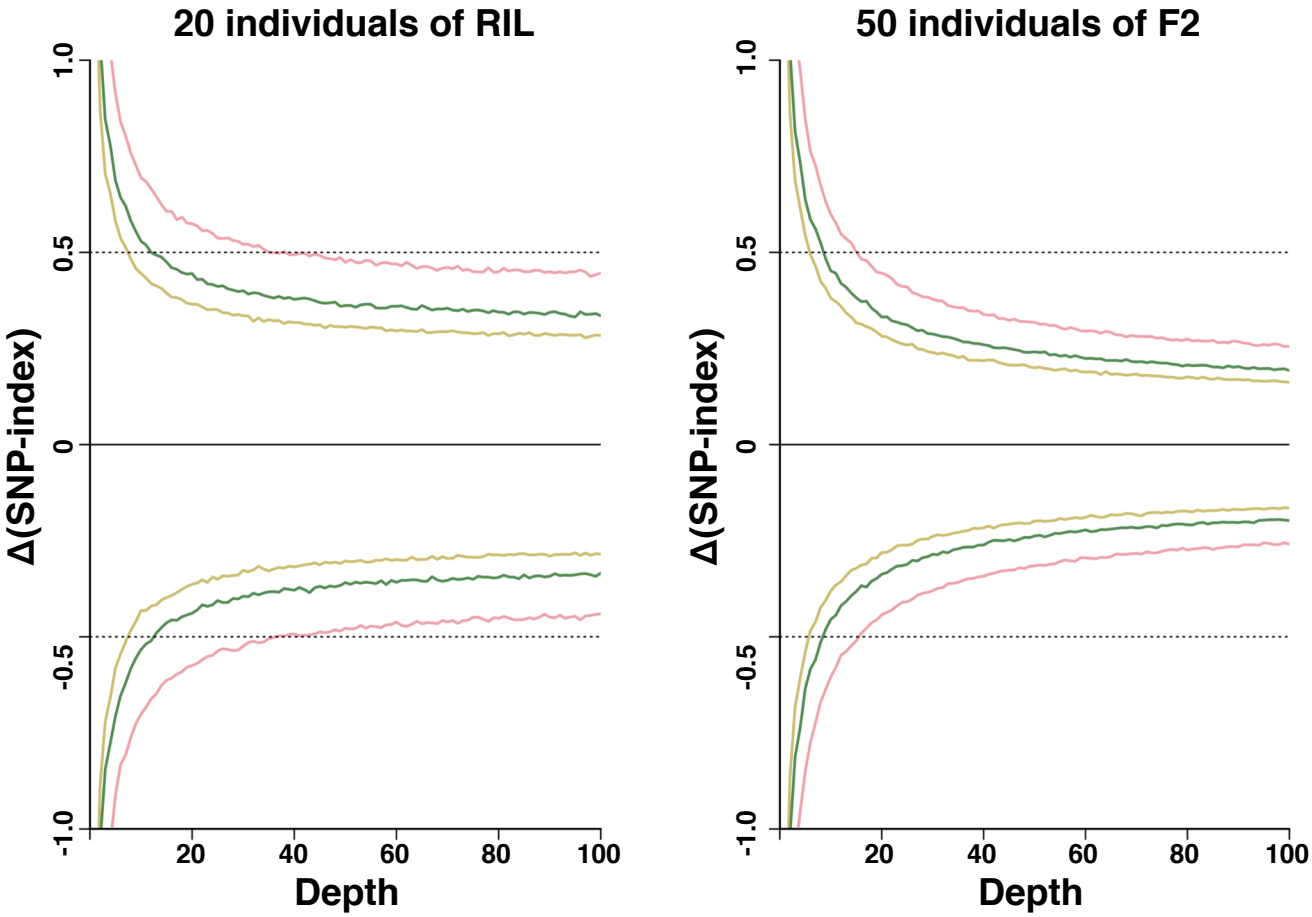
We assume the progenies were obtained by crossing two cultivars whose genotypes were represented by AA and BB, respectively. The number of individuals in both progeny is equal and corresponded to the QTL-seq samples presented in the main text. The genotype of the selected individuals is decided at a random position in the genome. The type of progeny is decided as either RILs or F2 as per the examples presented in the main text..

Simulation for calculating Δ(SNP-index)

SNP-index is simulated by the probability distribution of SNP-index, which corresponded to the binomial distribution of B (given read depth, B allele frequency). Δ(SNP-index) is calculated by subtraction of Progeny 2 SNP-index from Progeny 1 SNP-index.

Calculation of null distribution

The 90, 95 and 99% confidence intervals of Δ(SNP-index) are calculated based on the result obtained from 10,000 times replication. We use this confidential interval as null distribution of the non-selected bulked DNA.

Simulation test

(**left**) and (**right**) The result of simulation test of 10,000 replications at each read depth. The x- and y-axis represent the read depth and Δ(SNP-index), respectively. Light yellow, light green and light red colors indicated 90, 95 and 99% confidential interval for null distribution, respectively. (**left**) the result for the condition simulating RIL type progeny and 20 individuals, and represents the progeny derived from a cross between Nortai and Hitomebore in the main text. (**right**)) the result for the condition simulating F2 type progeny and 50 individuals, and represents the progeny derived from a crosses between Hitomebore and Dunghan shali in the main text.