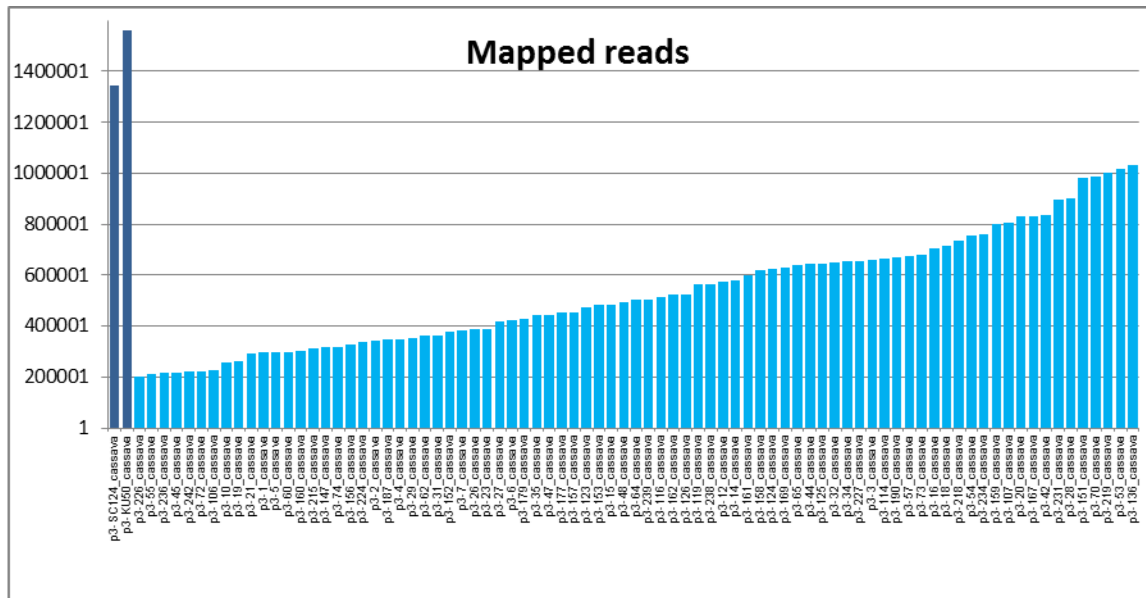
AFSM sequencing approach: a simple and rapid method for genome-wide SNP and methylation site discovery and genetic mapping

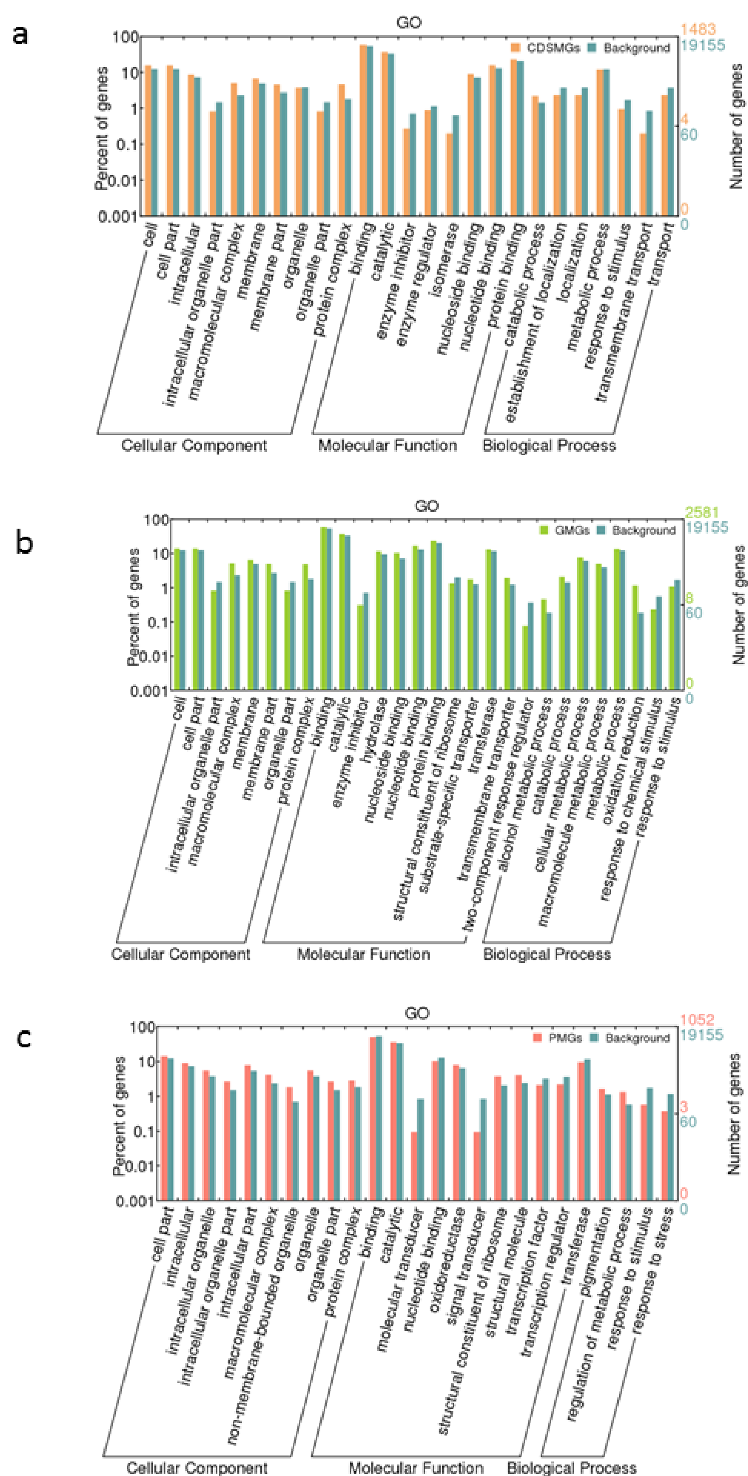
Zhiqiang Xia, Meiling Zou, Shengkui Zhang, Binxiao Feng & Wenquan Wang

Supplementary File	Title
Supplementary Figure 1	Distribution of reads across the 85 barcoded samples obtained from the cassava KS population that were sequenced in one lane.
Supplementary Figure 2	Annotation of methylated genes (CDSMGs, GMGs, and PMGs).
Supplementary Table 1	The full list of barcode adapters for EcoRI used in the AFSM approach.
Supplementary Table 2	Adapters and PCR primers used in the AFSM library preparation.
Supplementary Table 3	High-throughput sequencing of AFSM tags by Illumina Hiseq 2500.
Supplementary Table 4	Relative levels of cytosine methylation at the CCGG sites in cassava KS population.
Supplementary Table 5	AFSM markers (including SNPs, indels, and methylated sites) obtained in the cassava population.
Supplementary Table 6	Methylation patterns in the SC12 and, KU50 parents and their F1 progenies.
Supplementary Table 7	Comparison of six high-throughput polymorphism detection approaches.
Supplementary Protocol	Detailed protocol describing AFSM sample preparation.

Supplementary Figure 1 Distribution of reads across the 85 barcoded samples obtained from the cassava KS population that were sequenced in one lane.



Supplementary Figure 2 Annotation of methylated genes (CDSMGs, GMGs, and PMGs).



Annotation of methylated genes: CDSMGs (a), GMGs (b), PMGs (c) with BGI WEGO[31]. Of the 19,155 genes with GO annotations, 1483 CDSMGs, 2581 GMGs and 1052 PMGs showed significant enrichment differences ($P<0.05$, χ^2 test) compared with total analysed cassava genes. Annotations are grouped by cellular component, molecular function or biological process based on the JGI cassava gene GO annotation information. Gene numbers and percentages (on a log scale) are listed for each category.

Supplementary Table 1 The full list of 96 developed barcode adapters for EcoRI in AFSM approach.

No.	Barcode Adapter Sequence	No.	Barcode Adapter Sequence	No.	Barcode Adapter Sequence	No.	Barcode Adapter Sequence	No.	Barcode Adapter Sequence	No.	Barcode Adapter Sequence
1	TGGCT	17	CGCGT	33	GCGGA	49	TAACA	65	AAGCG	81	CTTAT
2	GAGAT	18	TCACA	34	CAATA	50	TCTCT	66	CGCCG	82	GTAGA
3	GCCTT	19	CCGCA	35	TCTGA	51	ACCTG	67	GGATA	83	CCTCG
4	CAGCT	20	CATGG	36	ACTAA	52	CGGTG	68	TGTAG	84	GAACT
5	TGCTT	21	TAGTT	37	GGCAT	53	GATTA	69	TCCAA	85	TTACT
6	ATAGG	22	CTCTA	38	CTACA	54	GCACT	70	TCATT	86	ATCCG
7	TGTTG	23	TCCGT	39	TACCT	55	ACTGG	71	TATCA	87	CGTCA
8	CTTAG	24	TATGT	40	TCGCG	56	ATTCA	72	GGAGG	88	CAGAA
9	GTCGT	25	TCGTA	41	CTCAT	57	TGAGA	73	CACCA	89	ACAGT
10	AATGA	26	GAATT	42	TTCCA	58	GGCCA	74	TGCGG	90	GGTTG
11	AACTT	27	ACCGA	43	ATATT	59	GTGTT	75	ATGCT	91	GCTAT
12	CTCGG	28	ACACG	44	GCGAG	60	CCATG	76	GTTCT	92	ATGTG
13	TGTGT	29	GCTCA	45	TTCTG	61	CCTTA	77	ATCAA	93	GTGCG
14	GTAAG	30	CTTGA	46	GAGTG	62	TTGAA	78	CCGAT	94	TGAAT
15	ACGTT	31	TAGAG	47	CATAT	63	TGCAA	79	AATTG	95	CAAGT
16	AATCT	32	ATTAT	48	GTTAA	64	TCCAG	80	AAGCA	96	GGTGT

A total of 85 barcode adapters (No. 1-85) were used in this study.

Supplementary Table 2 Adapters and PCR primers used in AFSM library preparation.

Name	Sequence (5'to3')
Adaptor Type	
“barcode” adapter	TAGCTCGTAGACACCGTCAG[barcode]G
	AATTC[barcode]CTGACGGTGTCTACGAGCTA
“HpaII-methylation” adapter	CGGTGAGATGAGGCATGAC
	GTCATGCCTCATCTCAC
“MspI-methylation” adapter	CGGACTAATGAGGCATGAC
	GTCATGCCTCATTAGTC
Primer Name	
AFSM primer EcoRI_all	TAGCTCGTAGACACCGTCAG
AFSM primer HpaII	GTCATGCCTCATCTCACCGG
AFSM primer MspI	GTCATGCCTCATTAGTCCGG

Supplementary Table 3 High-throughput sequencing of AFSM tags by Illumina Hiseq 2500.

	Number	Mapping efficiency (% HQ reads)
Raw reads	113,472,592	/
HQ reads	99,446,059	/
The highest HQ reads in one sample	3,164,835	/
The lowest HQ reads in one sample	428,131	/
The average HQ reads per sample	1,169,954	/
[a] CV in HQ reads	45.33%	/
HQ reads with “HpaII-methylation” adapter and HpaII cut-site	50,780,017	/
HQ reads with “MspI-methylation” adapter and MspI cut-site	48,457,678	/
Mapped reads	48,253,199	48.52%
The highest mapped reads in one sample	1,561,416	49.34%
The lowest mapped reads in one sample	201,558	47.08%
The average mapped reads per sample	567,685	48.52%
[a] CV in mapped reads	44.44%	/
Mapped reads with “HpaII-methylation” adapter and HpaII cut-site	25,055,929	49.34%
Mapped reads with “MspI-methylation” adapter and MspI cut-site	24,179,781	49.90%

[a] CV: The coefficient of variation

Supplementary Table 4 Relative levels of cytosine methylation at the CCGG sites in cassava KS population.

KS population	Methylated CCGG sites		
	Total	Hemi-methylated sites	Fully methylated sites
All	13598	9193（67.61%）	4405（32.39%）
Promoter	1354（9.96%）	658（7.16%）	696（15.80%）
Gene	2937（21.60%）	1092（11.88%）	1845（41.88%）
CDS	1661（12.22%）	561（6.10%）	1100（24.97%）

Supplementary Table 5 AFSM markers (including SNP, indel, and methylated sites) obtained in cassava population.

	Total	SNP	Indel	Methylated Site[a]	AFSM Marker		
					Gene	CDS	Promoter
All	316,276	278,737	23,941	13,598	113,907	49,831	48,058
Filtered	15,176	10,685	3,246	1,245	5,566	2,621	2,685

[a] Hemi-methylated or full methylated sites > 2/3.

Supplementary Table 6 Methylation patterns in the SC12 and, KU50 parents and their F1 progenies.

Pattern	SC124		KU50		F1		Number of sites
	Hemi-methylation	Full methylation	Hemi-methylation	Full methylation	Hemi-methylation	Full methylation	
A1	+	-	+	-	+	-	534
A2	-	+	-	+	-	+	182
B1	+	-	-	+	+	-	0
B2	+	-	-	+	-	+	0
B3	-	+	+	-	+	-	0
B4	-	+	+	-	-	+	0
C1	+	-	+	-	-	+	0
C2	-	+	-	+	+	-	0
D1	+	-	+	-	+	+	333
D2	-	+	-	+	+	+	197
D3	+	-	-	+	+	+	1148
D4	-	+	+	-	+	+	890

Supplementary Table 7 Comparison of six high-throughput polymorphism detection approaches.

	2b-RAD[22]	RAD[19]	GBS[20]	RSTA[32]	SFP[33]	AFSM
Methylation site detection	No	No	No	No	No	Yes
Fully methylated site detection	No	No	No	No	No	Yes
Hemi-methylated site detection	No	No	No	No	No	Yes
Number of cytosine methylated site surveyed	/	/	/	/	/	13,598
Number of cytosine fully methylated site surveyed	/	/	/	/	/	4,405
Number of cytosine hemi-methylated site surveyed	/	/	/	/	/	9,193
Marker type	33~36bp Sequence data: SNPs	Sequence data: SNPs next to restriction cut sites	Sequence data: SNPs and indels next to restriction cut sites	Restriction cut site polymorphisms: distinguishes SNPs and indels	SNP and indels	Sequence data: SNPs, indels and methylated sites
High quality reads	11,829,959	41,622	102,505,713	50,935	92,924	99,446,059
Read length (bp)	33~36 [b]	26 [b]	86[b]	50 [b]	25 [b]	2X132[a]
The coefficient of variation (CV)	Not mentioned	Not mentioned	43%	Not mentioned	Not mentioned	45.33%
Number of polymorphisms identified	5,051,475	Approximately 13,000	Not mentioned	12,431	3,806	48,253,199
Informative marker rate	42.7%	31%	Not mentioned	24%	4% at a 5% false discovery rate cut off	48.52%
Prior information required	Not necessary	EST or genome sequence to map short sequence reads	Not necessary	EST, 454 or genome sequence	EST, 454 or genome sequence	Not necessary

Polymorphism identification	Custom Perl scripts for sequence alignment	Custom Perl scripts for sequence alignment	Custom Perl scripts for sequence alignment	Genotype clusters across study individuals	Hybridization signal difference among study individuals	Custom Perl scripts for sequence alignment
Individual genotype data	Not mentioned	Not mentioned	Not mentioned	Yes	No	Yes
Organisms studied	*Arabidopsis*	*Neurospora*	Barley	Purple sea urchin	Yeast, *Arabidopsis*, *Anopheles*, several seed plants	Cassava
Sequencing platform	Illumina sequencing	Illumina sequencing	Illumina HiSeq2000	Custom high-density oligonucleotid e array (Agilent)	Custom high-density oligonucleotide array (Affymetrix)	Illumina HiSeq2500
Sonication	Not required	Yes	Not required	\	\	Not required
de novo analysis without sequenced genome still work	No	No	Yes	No	No	Yes

[a] Paired-end read sequencing

[b] Single-end read sequencing

Supplementary Protocol. Detailed protocol describing AFSM sample preparation.

The protocol can be adopted for detection of genome-wide SNPs, indels and cytosine methylations (including full methylations and hemi-methylations) in 85 samples in a simple and easy way.

The collected samples should be fresh and in the same organization part, and the growth conditions and the growth period should be consistent. Long-term preservation samples to liquid nitrogen or below -70℃ refrigerator.

High quality DNA is required for AFSM. DNA need be quantified using NANODROP 2000C spectrophotometer instrument and DNA concentration were normalized to 100ng/μL. Note-before processing a new set of samples, digested and non-digested total DNA need to run an agarose gel, to test whether DNA extraction are of sufficient quality for library construction. As control, non-digested sample should look identical to the original DNA after 8 hours at 37℃, while the digested samples should be clearly degraded.

Three types of adapters were designed: Barcodes Adapter, HpaII-Methylation Adapter and MspI-Methylation Adapter. Besides, two primer pairs were designed for PCR.

Combined two restriction enzyme pairs (EcoRI-MspI and EcoRI-HpaII) are used in this protocol.

Digest DNA

1、In two 85 well plates set up each of the 2 reactions in each well plate. A reaction mix is in a total volume of 20μL as follows. Tap to mix. Spin briefly in tabletop centrifuge. Incubate at 37 ℃ for 8 hours.

Reagent	**Volume (μL)**
Genomic DNA (100ng/μL)	2.0
NEBuffer 4 (10x)	2.0
EcoRI (20U/μL)	0.5
MspI (20U/μL)	0.5
dH_20	15.0
Total	**20.0**

Reagent	**Volume (μL)**
Genomic DNA (100ng/μL)	2.0
NEBuffer 1 (10x)	2.0
EcoRI (20U/μL)	0.5
HpaII (10U/μL)	1.0
dH_20	14.5
Total	**20.0**

2、Inactive the enzyme at 65℃ for 30 min. Hold samples to ice.

3、For each sample, load 8μL digested DNA on an agarose gel alongside a comparable amount of

intact DNA from the same sample as control.

Double stranded adaptors preparation

Combine the two oligonucelotides for each adaptor (Adaptor_1 and Adaptor_2). Standard, unmodified oligos were ordered in complimentary pairs and annealed in a high-salt solution to form the double-stranded adapter prior to use. The adapter were annealed by heating to 95℃ and then slowly cooling to 30℃ at a rate of -1℃/minute by programming a single step PCR cycle at 95℃ for 1 minute and then decreasing the temperature by 1℃ each cycle for 65 cycles. After ligation, the adapters were adjusted to a uniform concentration of 1.0μM each.

AFSM

Type	Adapter_1 (5'to3')	Adapter_2 (5'to3')
“Barcode” Adapter	TAGCTCGTAGACACC GTCAG[barcode]G	AATTC[barcode]CTGACG GTGTCTACGAGCTA
“MspI-Methylation” Adapter	CGGACTAATGAGGC ATGAC	GTCATGCCTCATTAGTC
“HpaII-Methylation” Adapter	CGGTGAGATGAGGC ATGAC	GTCATGCCTCATCTCAC

Ligation

1、In two 85 well plates set up each of the 2 reactions in each well plate. A reaction mix is in a total volume of 20μL as follows. Tap to mix. Spin briefly in tabletop centrifuge. Incubate at 16℃ over night.

Reagent	**Volume (μL)**
Digested DNA	10.0
T4 ligase buffer (10x)	2.0
1.0μM “Barcodes” Adapter	1.5
1.0μM “MspI-Methylation” Adapter	1.5
NEB T4 DNA ligase (200 U/μL)	0.5
dH_20	4.5
Total	**20.0**

Reagent	**Volume (μL)**
Digested DNA	10.0
T4 ligase buffer (10x)	2.0
1.0μM “Barcodes” Adapter	1.5
1.0μM “HpaII-Methylation” Adapter	1.5
NEB T4 DNA ligase (200 U/μL)	0.5
dH_20	4.5
Total	**20.0**

2、Inactive the enzyme at 65℃ for 20 min. Hold samples to ice.

Clean-up and Multiplexing

1、EcoRI-HpaII ligated samples were pooled and clean-up using E.Z.N.A. Cycle-pure Kit (Omega Bio-tek, Inc., US, D6492), then elute in 50μL elution buffer in one tube .

2、EcoRI-MspI ligated samples were pooled and clean-up, then elute in 50μL elution buffer in another tube.

PCR Amplification

1、Prepare the following master mix:

Reagent	**Volume (μL)**
DNA from previous step	10.0
NEB 2x Taq Master Mix	25.0
AFSM primer EcoRI_all (5μM)	4.0
AFSM primer HpaII (5μM)	4.0
dH_20	7.0
Total	**50.0**

Reagent	**Volume (μL)**
DNA from previous step	10.0
NEB 2x Taq Master Mix	25.0
AFSM primer EcoRI_all (5μM)	4.0
AFSM primer MspI (5μM)	4.0
dH_20	7.0
Total	**50.0**

2、Amplify using the following PCR cycling protocol:

a、30 seconds at 95℃

b、23 cycles of:

30 seconds at 95℃

30 seconds at 55℃

30 seconds at 68℃

c、5 minutes at 72℃

d、Hold at 16℃

3、Load these PCR products on a 2% agarose gel.

PCR product purification and size select

Purified the PCR product and receive the products with the size 250 to 500bp using the E.Z.N.A. Gel Extraction Kit (Omega Bio-tek, Inc., US, D2500-1).

After EcoRI-MspI 85-plex library and EcoRI-HpaII 85-plex library had been constructed, equal amount of the two libraries were combined into one 85-plex library.