



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Chromatin structure profile data from DNS-seq: Differential nuclease sensitivity mapping of four reference tissues of B73 maize (*Zea mays* L)



Zachary M. Turpin<sup>a</sup>, Daniel L. Vera<sup>b</sup>, Savannah D. Savadel<sup>a</sup>,  
 Pei-Yau Lung<sup>c</sup>, Emily E. Wear<sup>d</sup>, Leigh Mickelson-Young<sup>d</sup>,  
 William F. Thompson<sup>d</sup>, Linda Hanley-Bowdoin<sup>d</sup>,  
 Jonathan H. Dennis<sup>a</sup>, Jinfeng Zhang<sup>c</sup>, Hank W. Bass<sup>a,\*</sup>

<sup>a</sup> Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, USA

<sup>b</sup> Center for Genomics and Personalized Medicine, Florida State University, Tallahassee, FL, USA

<sup>c</sup> Department of Statistics, Florida State University, Tallahassee, FL, 32306, USA

<sup>d</sup> Department of Plant and Microbial Biology, North Carolina State University, Raleigh 27606, USA

## ARTICLE INFO

## Article history:

Received 5 June 2018

Received in revised form

12 July 2018

Accepted 6 August 2018

Available online 10 August 2018

## ABSTRACT

Presented here are data from Next-Generation Sequencing of differential micrococcal nuclease digestions of formaldehyde-crosslinked chromatin in selected tissues of maize (*Zea mays*) inbred line B73. Supplemental materials include a wet-bench protocol for making DNS-seq libraries, the DNS-seq data processing pipeline for producing genome browser tracks. This report also includes the peak-calling pipeline using the iSeg algorithm to segment positive and negative peaks from the DNS-seq difference profiles. The data repository for the sequence data is the NCBI SRA, BioProject Accession *PRJNA445708*.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

\* Corresponding author.

E-mail addresses: [zmt11@my.fsu.edu](mailto:zmt11@my.fsu.edu) (Z.M. Turpin), [daniel@danielvera.org](mailto:daniel@danielvera.org) (D.L. Vera), [sds14d@my.fsu.edu](mailto:sds14d@my.fsu.edu) (S.D. Savadel), [pei-yau.lung@stat.fsu.edu](mailto:pei-yau.lung@stat.fsu.edu) (P.-Y. Lung), [emily\\_wear@ncsu.edu](mailto:emily_wear@ncsu.edu) (E.E. Wear), [lamicke@ncsu.edu](mailto:lamicke@ncsu.edu) (L. Mickelson-Young), [wftb@ncsu.edu](mailto:wftb@ncsu.edu) (W.F. Thompson), [lkhanley@ncsu.edu](mailto:lkhanley@ncsu.edu) (L. Hanley-Bowdoin), [dennis@bio.fsu.edu](mailto:dennis@bio.fsu.edu) (J.H. Dennis), [jinfeng@stat.fsu.edu](mailto:jinfeng@stat.fsu.edu) (J. Zhang), [bass@bio.fsu.edu](mailto:bass@bio.fsu.edu) (H.W. Bass).

<https://doi.org/10.1016/j.dib.2018.08.015>

2352-3409/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications Table

Subject area	Biology
More specific subject area	Plant Chromatin Structure Profiling using Nuclease Sensitivity Assay
Type of data	Next-Generation Sequencing Data
How data was acquired	NEBNext Ultra II DNA Library Prep Illumina HiSeq. 2500 (Paired-End 50 base reads)
Data format	Raw sequence non-interleaved.fastq files
Experimental factors	None, tissues were from normal (wild-type) plants
Experimental features	Formaldehyde-fixed nuclei were purified from maize tissues and subjected to light or heavy micrococcal nuclease digestion, followed by decrosslinking and purification of genomic DNA for NGS library preparation. Size-selected libraries were sequenced on the Illumina HiSeq. 2500 platform.
Data source location	Samples were from Tallahassee, FL, FSU research field (earshoot, endosperm), greenhouse (coleoptilar node), or from Raleigh NC, NCSU lab-grown (root-tip).
Data accessibility	Data is with this article and available at NCBI SRA repository under the Project Accession PRJNA445708 "Differential MNase Profiling of Select Zea Mays Tissues" <a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJNA445708">https://www.ncbi.nlm.nih.gov/bioproject/PRJNA445708</a>
Related research article	none

## Value of the data

- These data provide a public resource for integrative and comparative analysis of chromatin structure in maize, a major crop and model organism.
- These data provide basic information about chromatin structure that could accelerate crop improvement strategies to address current and future challenges in food and biofuel production.
- These data define genome-wide differences in chromatin structure in four diverse tissues of maize.
- These data allow for discovery of so-called open chromatin linked to gene regulation.
- These data can be used for comparative analysis of MNase-seq to other chromatin accessibility assays such as ATAC-seq, DNaseI-seq, and FAIRE-seq.

## 1. Data

The sequence data are 50 bp paired-end reads from Illumina Hi-Seq. 2500 and the files are named as follows. *FRT1Ha\_R1.fastq.gz* refers to FSU-grown 1 mm Root Tips, Biological Replicate 1, Heavy digest, technical replicate a, read 1. For each tissue, there are four sets of libraries, heavy and light digests, and their replicates. These digest pairs are later used to produce the "difference" files which capture the differential nuclease sensitivity, DNS. For comparison, the heavy digests alone are typical of conventional nucleosome occupancy mapping data, whereas the light digests are needed for the difference calculation.

In order to facilitate subsequent analysis, each data file (.fastq formatted) described in the NCBI SRA BioProject Accession *PRJNA445708* can be named using the above schema to capture uniquely identifying information about each sample in the file name.

## 2. Experimental design, materials, and methods

### 2.1. Development of DNS-seq to study maize chromatin structure

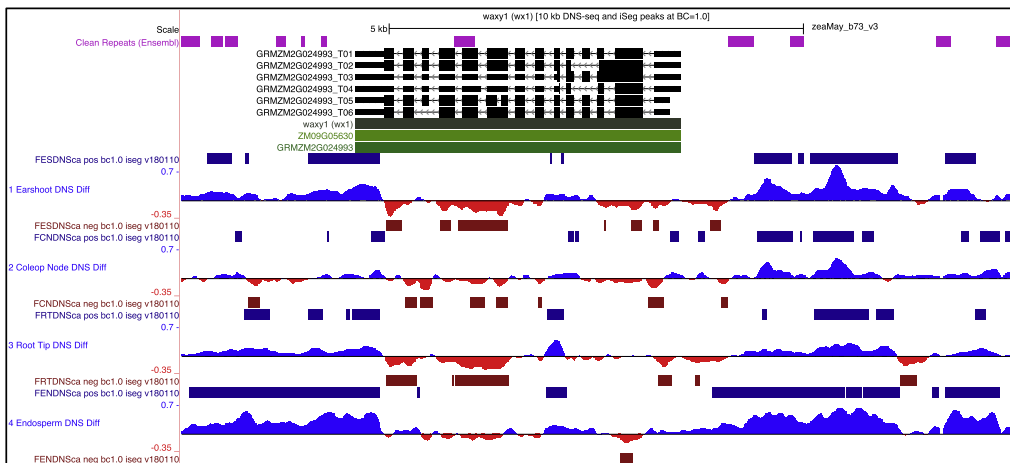
We have developed and refined Differential Nuclease Sensitivity (DNS) as a procedure for mapping nucleosome occupancy and open chromatin in maize [1–3]. Briefly, fixed nuclei are digested with a diffusible enzymatic probe, micrococcal nuclease (MNase, EC 3.1.31.1). Two digest conditions, “light” and “heavy” are employed and the resulting genomic DNA fragments are quantified by Next Generation Sequencing (NGS) to obtain the relative abundance of aligned fragments. The wet-bench protocol used to produce the DNS sequencing libraries is provided in Supplemental file 1. This protocol describes many tissue- and digestion condition- specific considerations for isolation of fixed nuclei, selection of appropriate MNase digestion conditions, and construction of Next-Generation Sequencing library pairs.

The Differential Nuclease Sensitivity (DNS) values are calculated as the difference (Light digest normalized reads “minus” Heavy digest normalized reads) between light and heavy relative read coverage. As shown in Fig. 1, plotting the DNS profiles along a genomic region reveals areas with both positive (MNase Sensitive - blue) or negative (MNase Resistant - red) values. DNS-seq profiles were calculated in this way for four reference tissues (15 DAP endosperm, 1 mm terminal root tips, 1–2 cm earshoots, and 4–5 day old seedling coleoptilar nodes). The DNS-seq data processing pipeline provided in Supplemental file 2 describes the sequential computational steps required to produce UCSC genome-browser-ready data tracks from NGS paired-end Illumina reads.

In order to delineate significant differences in global MNase-sensitivity, we segmented the profiles using the peak-calling program iSeg [4]. Descriptive statistics of these DNS-seq peak segments are shown in Fig. 2. The computational pipeline used to produce the iSeg peak calls is described in Supplemental file 3.

### 2.2. Plant materials

Maize (*Zea mays L*) B73 seeds were obtained from WF Thompson (NC State University, Raleigh, NC, USA, accession “27 FARMS”, bulked for NCSU, denoted B73<sup>NC27</sup>). Earshoot and endosperm tissues were collected from plants that were grown in the field at the Florida State University Mission Road



**Fig. 1.** Browser screenshot from genome maize.org, a UCSC browser for maize B73. A screenshot showing a 10 kilobase pair region around the *waxy1* (*wx1*) gene on chromosome 9. Repeat DNA (pink), gene models (black), gene identifiers (green) are shown above the DNS-seq profiles (blue/positive and red/negative difference value histograms) and their associated peak calls at a stringency of “BC=1.0” [4]. Track names for the DNS-seq profiles are shown at left and the data are scaled from -0.35 to 0.7 with smoothing = 6.

**A**

TOTAL BP in <i>i</i> SEG PEAKS			
TISSUE SAMPLE	Biological Cutoff		
	BC=1.0	BC=2.0	BC=3.0
	MSFs, MNase hyper-Sensitive Footprint peaks		
EARSHOOT	137,557,638	14,850,480	3,147,565
ENDOSPERM	119,354,275	16,277,529	2,378,279
COLEOPTILAR NODE	110,197,936	18,196,335	5,959,881
ROOT TIP	113,095,636	12,180,580	1,105,298
	MRFs, MNase Resistant Footprint peaks		
EARSHOOT	117,880,942	5,621,653	736,803
ENDOSPERM	102,249,575	8,498,076	1,041,557
COLEOPTILAR NODE	115,891,410	9,923,017	1,254,223
ROOT TIP	100,557,717	4,913,530	684,909

**B**

FRACTION (%) of mapped bp in <i>i</i> SEG PEAKS			
TISSUE SAMPLE	Biological Cutoff		
	BC=1.0	BC=2.0	BC=3.0
	MSFs, MNase hyper-Sensitive Footprint peaks		
EARSHOOT	7.5 %	0.81 %	0.17 %
ENDOSPERM	6.6 %	0.90 %	0.13 %
COLEOPTILAR NODE	6.0 %	1.0 %	0.33 %
ROOT TIP	6.1 %	0.66 %	0.06 %
	MRFs, MNase Resistant Footprint peaks		
EARSHOOT	6.4 %	0.31 %	0.04 %
ENDOSPERM	5.6 %	0.47 %	0.06 %
COLEOPTILAR NODE	6.3 %	0.54 %	0.07 %
ROOT TIP	5.5 %	0.27 %	0.04 %

**C**

TOTAL NUMBER <i>i</i> SEG PEAKS			
TISSUE SAMPLE	Biological Cutoff		
	BC=1.0	BC=2.0	BC=3.0
	MSFs, MNase hyper-Sensitive Footprint peaks		
EARSHOOT	450,718	87,444	29,030
ENDOSPERM	466,412	114,340	25,556
COLEOPTILAR NODE	728,915	128,937	53,494
ROOT TIP	426,160	82,074	12,266
	MRFs, MNase Resistant Footprint peaks		
EARSHOOT	804,159	61,671	6,633
ENDOSPERM	661,936	73,440	9,955
COLEOPTILAR NODE	1,045,286	117,746	13,998
ROOT TIP	620,006	48,616	5,817

**Fig. 2.** Summary statistics for *i*Seg peaks for the DNS-seq profiles of the four core reference tissues. Segments at three "biological cutoffs" (BC=1.0, 2.0, and 3.0) are tabulated separately for positive (MSF) and negative (MRF) peaks for each of the four tissues. Values summed over the whole genome are given for segment (A) total bp, (B) fraction of mapped bp, and (C) total number.

**Table 1**  
Sample information.

Harvest location	Tissue type	Biological replicate number	Digest level Heavy (H) or Light (L)	Technical replicate (a or b)	Read number (R1 or R2)
FSU (F)	coleoptilar node (CN)	1 or 2	L or H	N/A	R1 or R2
FSU (F)	endosperm (EN), 15 days after pollination	1 or 2	L or H	N/A	R1 or R2
FSU (F)	earshoot (ES), 1–2 cm	1 or 2	L or H	N/A	R1 or R2
FSU (F)	root tip (RT), 1 mm	1, 2, 3, or 4	L or H	a or b (for bioreps 1 and 2 only)	R1 or R2

In order to facilitate subsequent analysis, each data file (.fastq formatted) described in the NCBI SRA BioProject Accession HYPERLINK "<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA445708>" \h PRJNA445708 can be named using the above schema to capture uniquely identifying information about each sample in the file name.

Research Facility (Tallahassee, FL, USA). Tissue harvest for field-grown tissues was done at 9–11 A.M. by flash-freezing in liquid nitrogen. Coleoptilar nodes were collected from greenhouse-grown (King Greenhouse-room A, Biological Science Dept., FSU, Tallahassee, FL, USA) seedlings. Root tips were collected from lab-grown (WF Thompson Lab, NC State University, Raleigh, NC, USA) seedlings as described below.

### 2.3. Earshoot, 1–2 cm, Field-grown

Earshoots were harvested in the field at 9–11 A.M. and immediately frozen in liquid nitrogen. For each earshoot collected, stalks were cut at the base, the top or second from the top earshoot was removed, the husks were peeled off, the silks were gently but quickly rubbed off, the earshoot was cut at the base and measured. Earshoot samples were immediately frozen in liquid nitrogen, pooled by date of harvest, and transferred to  $-80^{\circ}\text{C}$  freezers for storage.

### 2.4. Endosperm, 15 days after pollination, field-grown

Endosperm from self-pollinated ears was harvested at 15 days after pollination by manual microdissection in the field at 9–11 A.M. and immediately frozen in liquid nitrogen. For each ear, husks were removed, kernels detached, and the endosperm was separated from the embryo, scutellum, and pericarp. For each ear, only 10–20 kernels were quickly removed in order to limit the time elapsed between ear removal and tissue freezing to less than 5 min. The process was repeated for multiple ears during the 9am–11am time window for each day of harvest. Endosperm samples were immediately frozen in liquid nitrogen, pooled by date of harvest, and transferred to  $-80^{\circ}\text{C}$  freezers for storage.

### 2.5. Coleoptilar Node, 4–5 days after planting, 5 mm section of greenhouse-grown, spear stage seedlings

Coleoptilar nodes were harvested from greenhouse-grown seedlings 4–5 days after planting. Seeds were planted 2/3 deep in 2 in. of soil ("Maize Soil Mix": per batch = 10 gallon Fafard #3 mix: 5 gal coarse sand, and 200 mL Osmocote fertilizer pellets (Scott's, 18.6.1)) spread into planting flats on greenhouse tables with natural plus supplemental lighting. The soil was well watered on day 1 and kept moist until harvest. Coleoptilar nodes were harvested between 10 AM and 12 PM on the day that "spears" showed above ground. Seedlings were quickly removed from the soil and first cut with a razor blade just above the seed. A 5 mm segment centered on the CN was then excised and immediately transferred to liquid nitrogen. Seedling removal and CN collection was done sequentially in order to limit the handling time to less than 1 min per plant. The CN samples were pooled by date of harvest, and transferred to  $-80^{\circ}\text{C}$  freezers for storage (Table 1).

## 2.6. Root Tip, 3.5 days after imbibition, terminal 1 mm tips of primary and seminal roots

Root tips were harvested, fixed, and frozen from laboratory-grown seedlings using a modified protocol adapted from a technique developed to investigate plant DNA replication [5]. Seeds were imbibed in running water overnight, and germinated in sterile Magenta boxes (Sigma-Aldrich) containing a damp paper towel, and seedlings were grown under constant light at 28 °C for 72 h. The root tips were collected by cutting off the terminal 1 mm of primary and seminal root tips, fixing in 1% formaldehyde for 20 min, rinsing 3X in Buffer A [2], and flash freezing in liquid nitrogen. This fixed and frozen tissue, unlike the other tissues, was broken open via gentle brief polytron (speed 2–3, 15 s) of root tips from ~ 160 seedlings directly in MNase digestion buffer [2] followed by microfiltration through 50 µm Partec filter.

## 3. Methods

A general methodology for the isolation of plant nuclei, MNase digestion, and subsequent NGS library preparation are described in detail in supplemental file 1, “DNS Bench Protocol”.

Methodology for the processing of raw NGS data to browser-ready DNS profile data tracks are described in detail in supplemental file 2, “DNS Pipeline”.

Methodology for the segmentation of the DNS profiles are described in detail in supplemental file 3, “iSeg Pipeline”.

## Acknowledgements

This work was supported by grants from the NSF Plant Genome Research Program (IOS Award 1444532 to H.W.B.; IOS Award 1025830 to L.H.-B. and W.F.T.) and from the NIH National Institute of General Medical Sciences (R01GM126558 to J.Z.).

## Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.08.015>.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.08.015>.

## References

- [1] J.A. Fincher, D.L. Vera, D.D. Hughes, K.M. McGinnis, J.H. Dennis, H.W. Bass, Genome-wide prediction of nucleosome occupancy in maize reveals plant chromatin structural features at genes and other elements at multiple scales, *Plant Physiol.* 162 (2013) 1127–1141.
- [2] D.L. Vera, T.F. Madzima, J.D. Labonne, M.P. Alam, G.G. Hoffman, S.B. Girimurugan, J. Zhang, K.M. McGinnis, J.H. Dennis, H.W. Bass, Differential nuclease sensitivity profiling of chromatin reveals biochemical footprints coupled to gene expression and functional DNA elements in maize, *Plant Cell.* 26 (2014) 3883–3893.
- [3] E. Rodgers-Melnick, D.L. Vera, H.W. Bass, E.S. Buckler, Open chromatin reveals the functional maize genome, *Proc. Natl. Acad. Sci. USA* 113 (2016) E3177–E3184.
- [4] S.B. Girimurugan, Y. Liu, P.-Y. Lung, D.L. Vera, J.H. Dennis, H.W. Bass, J. Zhang, iSeg: an efficient algorithm for segmentation of genomic and epigenomic data, *BMC Bioinforma.* 19 (2018) 131.
- [5] H.W. Bass, E.E. Wear, T.-J. Lee, G.G. Hoffman, H.K. Gumber, G.C. Allen, W.F. Thompson, L. Hanley-Bowdoin, A maize root tip system to study DNA replication programmes in somatic and endocycling nuclei during plant development, *J. Exp. Bot.* 65 (2014) 2747–2756.