Chapter 13

Plant In Situ Hi-C Experimental Protocol and Bioinformatic Analysis

Francisco J. Pérez-de los Santos, Jesús Emiliano Sotelo-Fonseca, América Ramírez-Colmenero, Hans-Wilhelm Nützmann, Selene L. Fernandez-Valverde, and Katarzyna Oktaba

Abstract

Hi-C enables the characterization of the 0conformation of the genome in the three-dimensional nuclear space. This technique has revolutionized our ability to detect interactions between linearly distant genomic sites on a genome-wide scale. Here, we detail a protocol to carry out in situ Hi-C in plants and describe a straightforward bioinformatics pipeline for the analysis of such data, in particular for comparing samples from different organs or conditions.

Key words Hi-C, Chromosome conformation, Arabidopsis, Sequencing, Bioinformatics, Differential interactions

1 Introduction

Eukaryotic chromosomes repeatedly fold in the three-dimensional space of the cell nucleus. This organization is known as genome or chromosome topology and has several hierarchical levels [1]. At the most basic level, DNA is wrapped around histone proteins and arranged into loosely or tightly packed nucleosomal arrays (reviewed in [2]). Above the nucleosomal level, chromatin folds into loops that bring linearly distant regions of the genome into close spatial proximity, creating a multitude of functional interactions, for example enabling distal enhancers to interact with target promoters [3, 4]. Topologically associating domains (TADs), characterized by an increased interaction count between loci located in the same domain, and less frequent interactions with neighboring loci, comprise the next organization level [5, 6]. TADs themselves are organized into active (A) and inactive (B) compartments that share similar gene expression and epigenetic profiles [7]. Finally,

Alejandro Pereira-Santana, Samuel David Gamboa-Tuz and Luis Carlos Rodríguez-Zapata (eds.), *Plant Comparative Genomics*, Methods in Molecular Biology, vol. 2512, https://doi.org/10.1007/978-1-0716-2429-6_13,

individual chromosomes occupy distinct regions in the nucleus, known as chromosome territories [8]. This genomic architecture exhibits conserved folding patterns across species and cell types in both animals and plants [9].

The analysis of genome topology by chromosome conformation capture methods has drastically improved our ability to identify interactions between DNA domains that are seemingly distant in a linear representation of the genome [10]. The recent widespread use of chromatin conformation capture (3C) followed by unbiased high-throughput sequencing techniques (Hi-C) has provided comprehensive knowledge of chromatin interactions at a genome-wide level in a variety of organisms [11].

Here, we present a detailed in situ Hi-C protocol for plants along with a complete bioinformatic analysis workflow (Fig. 1), based on previously described wet-lab and bioinformatic methods [12–20]. This protocol was implemented in cotyledons and roots of *Arabidopsis thaliana*. The experimental protocol and the bioinformatics pipeline can be adapted for analysis in other organs and plant species.

Briefly, the experimental protocol involves formaldehyde fixation of cells, isolation of nuclei and digestion of chromatin with a restriction enzyme. The overhangs left by the restriction enzyme are then filled in with biotin-conjugated nucleotides, the fragments are religated, size-selected, and purified using streptavidin-coated magnetic beads. Sequencing libraries are amplified, undergo quality control and the reads are paired-end sequenced. The bioinformatics workflow details the steps required for filtering and aligning reads, generating contact matrices, and annotating compartments, TADs, interaction peaks, and differential interactions using publicly available bioinformatics packages.

2 Materials

2.1 Reagents

1. 2 M glycine.
2. DNase/RNase-free distilled water (Invitrogen).
3. Liquid nitrogen.
4. Vectashield Antifade Mounting Medium (Vector).
5. 4′,6-diamidino-2-phenylindole (DAPI).
6. DpnII restriction enzyme and DpnII buffer (NEB).
7. 0.4 mM biotin-14-dATP (Invitrogen).
8. DNA Polymerase I, Large (Klenow) Fragment (NEB).
9. T4 DNA Ligase and 10x T4 DNA Ligase buffer (Thermo Scientific).
10. 10 mg/ml bovine serum albumin (BSA; NEB).

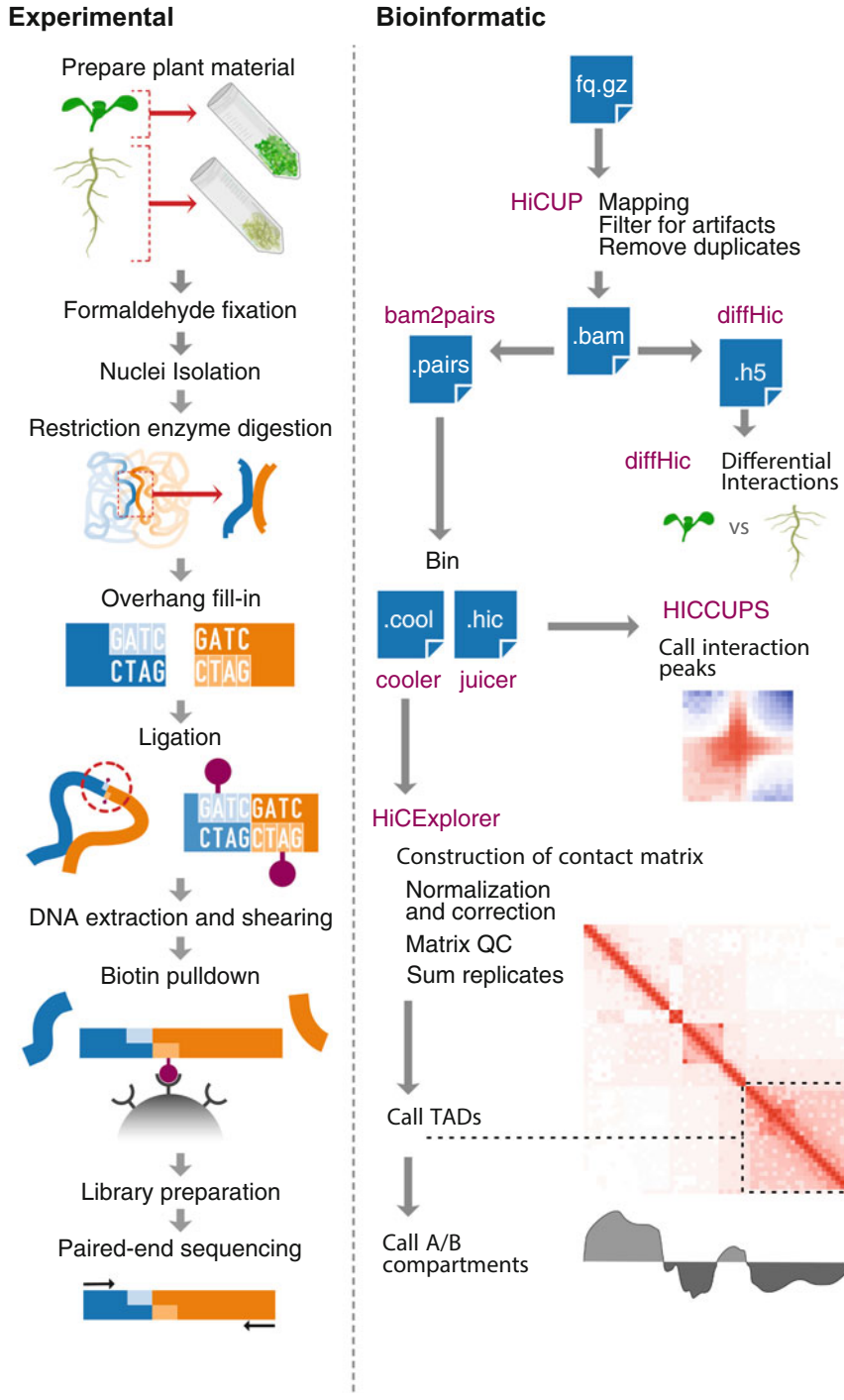

Fig. 1 Overview of experimental in situ Hi-C protocol and bioinformatic analysis of sequencing data. Left panel, simplified schematic representation of the Hi-C protocol. After obtaining plant material, nucleic acid–protein interactions are preserved by cross-linking with a formaldehyde solution, followed by nuclei extraction. DNA is subsequently digested with a restriction enzyme (DpnII), generating 5′ overhangs, which are filled with regular dNTPs and biotin-14-dATP, followed by blunt-end ligation. Next, DNA is purified and sonicated, producing small-sized DNA fragments that can be captured using streptavidin-coated beads. These fragments are used

11. 20 mg/ml Proteinase K (Thermo Scientific).
12. 5 M sodium chloride (NaCl).
13. 25:24:1 (v/v/v) phenol–chloroform–isoamyl alcohol.
14. Chloroform.
15. 20 mg/ml glycogen (Roche).
16. Ethanol.
17. 20 mg/ml RNase A (Invitrogen).
18. 1 Kb Plus DNA Ladder (Invitrogen).
19. Agarose.
20. T4 DNA polymerase and 10× NEB 2.1 buffer (NEB).
21. 0.5 M ethylenediaminetetraacetic acid (EDTA; Invitrogen).
22. Dynabeads MyOne Streptavidin C1 (Invitrogen).
23. NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB).
24. NEBNext Multiplex Oligos for Illumina (NEB).
25. SPRIselect beads (Beckman Coulter).
26. 2× NEBNext Ultra II Q5 Master Mix (NEB).
27. 100× SYBR Green I Nucleic Acid Gel Stain (Invitrogen).
28. Qubit dsDNA HS Assay Kit (Invitrogen).
29. Bioanalyzer High Sensitivity DNA Kit (Agilent).

2.2 Labware and Equipment

1. 1.5 ml microcentrifuge tubes.
2. Greiner dishes (Sigma-Aldrich).
3. 50 ml centrifuge tubes.
4. Scalpel.
5. Nylon filters.
6. 10 ml serological pipettes.
7. Miracloth (Millipore).
8. Paper towels.
9. Mortar and pestle.

Fig. 1 (continued) to generate libraries for high-throughput paired-end sequencing. Right panel, the bioinformatics analysis begins with mapping sequencing read pairs to the reference genome and filtering out uninformative read pairs. The resulting bam file is converted to file formats compatible with HiC analysis tools. This information is used to generate a count matrix, which is then processed using HiCExplorer tools, to account for visibility biases. If the analysis aims to compare several matrices, a normalization step is necessary to account for differences in sequencing depth between samples. Once the matrices are corrected, topologically associating domains (TADs), compartments and interaction peaks can be determined. diffHiC can also be used to account for biases between libraries and to detect regions that interact differently in each experimental condition

10. Cell counting chamber.
11. Maxymum Recovery pipette filter tips (Axygen).
12. Phase Lock Gel Heavy tubes (Quantabio).
13. 1.5 ml LoBind microcentrifuge tubes (Eppendorf).
14. MicroTube AFA Fiber Pre-Slit Snap-Cap (Covaris).
15. 0.2 ml PCR tubes.
16. Plant growth incubator.
17. Desiccator connected to a vacuum pump with manometer.
18. Refrigerated centrifuge for 50 ml conical tubes.
19. Refrigerated centrifuge for 1.5 ml microcentrifuge tubes.
20. Epifluorescent microscope.
21. Thermomixer (Eppendorf).
22. NanoDrop spectrophotometer (Thermo Scientific).
23. Agarose gel casting tray and combs.
24. DNA electrophoresis chamber and power supply.
25. S2 Focused-Ultrasonicator (Covaris).
26. Magnetic rack for 1.5 ml microcentrifuge tubes.
27. PCR thermocycler.
28. Real-time PCR instrument.
29. Qubit Fluorometer (Thermo Scientific).
30. Bioanalyzer Instrument (Agilent).

2.3 Reagent Setup

1. MS medium: 4.3 g/l Murashige and Skoog basal salt mixture (MS; Sigma-Aldrich), 10 g/l sucrose, 6 g/l Phytagel (Sigma-Aldrich). Adjust pH to 5.8 with 1 M NaOH before adding Phytagel, store at 4 °C.
2. Nuclei isolation buffer (NIB): 20 mM HEPES pH 8.0, 250 mM sucrose, 1 mM $MgCl_2$, 5 mM KCl, 40% (v/v) glycerol, 0.25% (v/v) Triton X-100, 0.1 mM PMSF, and 0.1% (v/v) 2-mercaptoethanol (*see* **Note 1**) [12].
3. Nuclei isolation buffer with formaldehyde (NIB-FA): 20 mM HEPES pH 8.0, 250 mM Sucrose, 1 mM $MgCl_2$, 5 mM KCl, 40% (v/v) glycerol, 0.25% (v/v) Triton X-100, 4% (v/v) formaldehyde (ultrapure, methanol free; Polysciences), 0.1 mM PMSF, and 0.1% (v/v) 2-mercaptoethanol [12].
4. Nuclease isolation buffer with protease inhibitor cocktail (NIB-P): 20 mM HEPES pH 8.0, 250 mM Sucrose, 1 mM $MgCl_2$, 5 mM KCl, 40% (v/v) glycerol, 0.25% (v/v) Triton X-100, 1× cOmplete ULTRA protease inhibitor cocktail (Roche), 0.1 mM PMSF, and 0.1% (v/v) 2-mercaptoethanol [12].

5. 1× PBS: Diluted from 10× phosphate buffered saline (Gibco).
6. 0.5% (v/v) SDS: Prepare from 10% (v/v) ultrapure sodium dodecyl sulfate solution (Invitrogen).
7. 10% (v/v) Triton X-100: Prepare from Triton X-100 solution (Sigma-Aldrich).
8. 3.3 mM dCTP–dGTP–dTTP mix: Prepare from 100 mM dNTPs set (Invitrogen).
9. Extraction buffer: 50 mM Tris–HCl, pH 8.0, 10 mM EDTA, and 1% (v/v) SDS.
10. 3 M sodium acetate pH 5.2 ($C_2H_3NaO_2$).
11. 70% (v/v) ethanol.
12. EB buffer: 10 mM Tris–HCl pH 8.0.
13. 1 mM dATP: Prepare from 100 mM dNTPs set (Invitrogen).
14. 1 mM dGTP: Prepare from 100 mM dNTPs set (Invitrogen).
15. 2× B&W buffer: 10 mM Tris–HCl pH 7.5, 1 mM EDTA, 2 M NaCl, prepare fresh.
16. 1× B&W buffer +0.1% (v/v) Triton X-100, prepare fresh.
17. 80% (v/v) ethanol, prepare fresh.

3 Methods

The protocol below describes the preparation of Hi-C sequencing libraries from two plant organs: cotyledons and roots of *Arabidopsis thaliana* 8-day-old seedlings. All steps are described for processing a single sample (either from cotyledons or roots). If using other plant material or species, growth conditions, harvest, cross-linking, and nuclei isolation steps may have to be adapted and optimized.

3.1 Preparation of Plant Material

1. Using a 1.5 ml microcentrifuge tube, measure a volume of 300 μl of *Arabidopsis thaliana* seeds (approximately 6900 seeds). This amount of seeds is sufficient to collect 2–3 g of each organ, cotyledon or root, from 8-day-old seedlings.
2. Grow seedlings vertically on MS medium in Greiner dishes for 8 days in a plant growth incubator under long-day conditions (16 h light at 22 °C and 8 h darkness at 16 °C).

3.2 Formaldehyde Cross-Linking of Plant Material

1. Separate cotyledons from roots with a scalpel and transfer 2–3 g of each type of organ immediately into a 50 ml centrifuge tube containing 15 ml NIB. Store on ice (*see* **Note 2**).
2. To cross-link, add 15 ml NIB-FA (2% formaldehyde final concentration). Mix by swirling with a 1 ml pipette tip. Completely immerse the plant material in the solution using nylon filters. Vacuum-infiltrate in a desiccator for 1 h at room temperature

(RT). Release vacuum every 15 min to enhance the penetration of fixative (*see* **Note 3**).

3. To stop cross-linking reaction, release vacuum, remove the nylon filter and add 2 ml 2 M glycine. Mix by pipetting up and down with a 10 ml serological pipette. Reimmerse the sample and vacuum-infiltrate for 5 min.
4. Decant liquid and wash the sample three times with ice-cold 1× PBS or ultrapure water.
5. Carefully wrap the plant material in Miracloth and dry by pressing with paper towels.
6. Transfer the dried sample to a 50 ml centrifuge tube and flash-freeze in liquid nitrogen. Proceed to nuclei isolation.

3.3 Nuclei Isolation

1. Use mortar and pestle (precooled with liquid nitrogen) to grind the cross-linked plant material to a fine powder in liquid nitrogen. Transfer powder to a liquid nitrogen-cooled 50 ml centrifuge tube.
2. Resuspend the sample in 10 ml NIB-P. Mix by gentle agitation with a 1 ml pipette tip until the solution becomes homogeneous. Place the tube on ice (*see* **Note 4**).
3. Separate nuclei by filtering samples twice through a double layer of Miracloth. Collect the filtrate in a 50 ml centrifuge tube and wash Miracloth with an extra 5 ml of NIB-P to collect the remaining material. To avoid contamination with cell debris, do not squeeze the Miracloth.
4. Centrifuge nuclei suspension at 3000 × *g* for 15 min at 4 °C. Carefully remove and discard the supernatant using a 10 ml serological pipette.
5. To resuspend the nuclei pellet, add 1 ml NIB-P and mix by gently swirling with a pipette tip. To avoid mechanical damage of nuclei, continue to resuspend by pipetting gently using a cut-off pipette tip. Do not vortex.
6. Transfer sample to a 1.5 ml microcentrifuge tube using a cut-off pipette tip. Centrifuge at 1900 × *g* for 5 min at 4 °C. Discard supernatant.
7. Wash nuclei pellet twice with 1 ml NIB-P. Resuspend nuclei pellet by pipetting gently using a cut-off pipette tip. Centrifuge at 1900 × *g* for 5 min at 4 °C. Discard the supernatant.
8. Resuspend the nuclei in 100 µl NIB-P by pipetting gently using a cut-off pipette tip.
9. Assess nuclei quality by staining 1 µl nuclei suspension with Vectashield mounting medium with DAPI. Analyze the nuclei using epifluorescence microscopy [14]. Intact nuclei show sharp contours as described by [12]. Estimate nuclei quantity

by staining 3 μl nuclei suspension with DAPI and pipette sample onto a counting chamber. Count the individual nuclei using epifluorescence microscopy. Determine the concentration of nuclei and proceed with $>10^7$ nuclei [14].

10. Centrifuge the nuclei suspension at 1900 × *g* for 5 min at 4 °C. Discard the supernatant.
11. Remove NIB-P traces by washing the nuclei pellet twice with 300 μl 1× DpnII buffer. Resuspend nuclei pellet by pipetting gently using a cut-off pipette tip. Centrifuge at 1900 × *g* for 5 min at 4 °C. Discard the supernatant. For all subsequent steps use pipette filter tips. In order to minimize sample retention and enhance yield, use Maxymum Recovery pipette filter tips or equivalent.
12. Permeabilize nuclei by gently resuspending the pellet in 100 μl 0.5% SDS using a cut-off pipette tip, be careful to avoid froth or bubbles. Incubate for 10 min at 65 °C (*see* **Note 5**).
13. Quench SDS by adding 70 μl ultrapure water and 50 μl 10% Triton X-100. Mix by pipetting up and down, be careful to avoid froth or bubbles. Incubate for 15 min at 37 °C while shaking at 450 rpm (*see* **Note 6**).

3.4 Restriction Enzyme Digestion

1. Add 25 μl 10× DpnII buffer and mix by pipetting. Collect 10 μl as undigested chromatin control, store at −20 °C until further processing. Digest chromatin by adding 100 U of DpnII enzyme. Mix by pipetting up and down. Incubate the digestion reaction for 3 h at 37 °C while shaking at 450 rpm. This digestion step may be done overnight (*see* **Note 7**).
2. Inactivate DpnII enzyme by incubating for 20 min at 62 °C. Collect 10 μl as digested chromatin control, store at −20 °C until further processing. Transfer the sample to ice (*see* **Note 8**).
3. Assess the quantity and quality of undigested and digested chromatin control samples by performing a cross-linking reversal, followed by DNA extraction, spectrophotometric quantification, and gel electrophoresis. Alternatively, perform rapid reversal of chromatin cross-linking (*see* **Note 9**).

3.5 Overhang Fill-in with a Biotinylated Nucleotide

1. The overhangs left by DpnII are filled-in by adding 19 μl 0.4 mM biotin-14-dATP (0.03 mM final concentration), 2.3 μl 3.3 mM dCTP/dGTP/dTTP (0.03 mM final concentration) and 50 U DNA Polymerase I Large (Klenow) Fragment. Mix by pipetting up and down. Incubate for 90 min to 2 h at 37 °C while shaking at 450 rpm.

3.6 In-Situ Ligation of Proximal Ends

1. Filled-in DNA fragments are ligated by adding 719 μl ultrapure water, 120 μl 10× T4 DNA ligase buffer, 100 μl 10% Triton

X-100, 50 Weiss U T4 DNA ligase and 12 μl 10 mg/ml BSA. Mix by inverting the tube 5 times. Incubate overnight at 16 °C with gentle rotation (*see* **Note 10**).

3.7 Cross-Linking Reversal

1. Centrifuge the nuclei suspension at 2500 × *g* for 10 min at 4 °C. Carefully remove and discard the supernatant.
2. Resuspend the nuclei in 380 μl extraction buffer.
3. To digest proteins, add 20 μl 20 mg/ml proteinase K and mix by pipetting. Incubate for 30 min at 55 °C while shaking at 1000 rpm.
4. Add 100 μl 5 M NaCl and mix by pipetting. Incubate at least 8 h or overnight at 68 °C.

3.8 DNA Extraction

1. Before use, spin Phase Lock Gel (Quantabio) tube at 12,000 × *g* for 30 s at RT. Transfer decross-linked sample to prespun Phase Lock Gel tube. Add 500 μl phenol–chloroform–isoamyl alcohol (25:24:1). Mix thoroughly by vigorous shaking for 2 min to form a transiently homogenous suspension, do not vortex. Centrifuge at 12,000 × *g* for 5 min at RT.
2. Perform a second extraction, by adding 500 μl chloroform to the same Phase Lock Gel tube. Mix thoroughly by vigorous shaking for 2 min, do not vortex. Centrifuge at 12,000 × *g* for 5 min at RT.
3. Transfer DNA-containing aqueous upper phase to a 1.5 ml LoBind microcentrifuge tube.
4. Precipitate DNA by adding 1/10 volume 3 M sodium acetate pH 5.2, 1 μl glycogen and 1200 μl ice-cold 100% ethanol.
5. Mix by inverting 5 times and incubate at −80 °C for at least 1 h. Centrifuge at 20,000 × *g* for 1 h at 4 °C. Remove the supernatant.
6. Wash the pellet twice with 1 ml 70% ethanol. Centrifuge at 20,000 × *g* for 5 min at RT. Remove the supernatant and air-dry pellet.
7. Dissolve the ligated Hi-C sample in 50 μl EB buffer.
8. Digest RNA by adding 1 μl 20 mg/ml RNase A. Mix by pipetting. Incubate at 37 °C for 30 min.
9. To assess ligation efficiency, measure DNA concentration using a NanoDrop spectrophotometer, and load 500 ng of sample and a 1 Kb Plus DNA Ladder on a 1.8% agarose gel for electrophoresis. Run the gel for 40 min at 90 V. Ligated Hi-C sample will appear as a smear of intermediate sizes between undigested and digested chromatin controls (assessed in Subheading 3.4).

3.9 Biotin Removal from Unligated DNA Ends

1. Remove biotin from unligated DNA ends, by adding to the 50 μl ligated Hi-C sample 12 μl 10× NEB 2.1 buffer, 3 μl 1 mM dATP, 3 μl 1 mM dGTP, 1.2 μl 10 mg/ml BSA,15 U T4 DNA polymerase, and ultrapure water up to 120 μl. Incubate for 30 min at 20 °C (*see* **Note 11**).
2. To stop the reaction, add 3 μl 0.5 M EDTA. Mix by pipetting up and down.

3.10 DNA Shearing

1. Transfer sample to a Covaris AFA fiber microtube; be careful to avoid bubbles.
2. Shear the DNA to ~200–400 bp (depends on sequencing read length). For Covaris S2 Focused-Ultrasonicator use the following program: 2 cycles of 50 s, 10% duty, intensity 5, 200 cycles per burst (*see* **Note 12**).
3. To assess the fragment size distribution of the sheared sample, load 300 ng and a 100 bp DNA Ladder on a 1.8% agarose gel for electrophoresis. Run the gel for 40 min at 90 V. The sheared sample will appear as a smear from 200 to 400 bp approximately.

3.11 Biotin Pulldown

1. To prepare Dynabeads MyOne Streptavidin C1 for biotin pulldown, vortex 10 mg/ml beads suspension for >30 s and transfer 60 μl to a 1.5 ml LoBind microcentrifuge tube. Wash twice by adding 1 ml 2× B&W buffer and resuspending by pipetting. Place the tube on a magnetic rack for 2 min. Carefully remove and discard the supernatant. Resuspend the beads in 120 μl 2× B&W.
2. To capture biotinylated DNA fragments, adjust the volume of the sheared Hi-C sample to 120 μl with ultrapure water, and add to the 120 μl 2× B&W bead suspension. Incubate the tube using gentle rotation for 20 min at RT.
3. Place the tube on a magnetic rack for 2 min. Carefully remove and discard the supernatant.
4. Wash beads twice with 600 μl 1× B&W buffer +0.1% Triton X-100. Mix by pipetting and incubate for 2 min at 55 °C while shaking at 1000 rpm. Place the tube on a magnetic rack for 2 min. Carefully remove and discard the supernatant.
5. Wash beads with 600 μl EB buffer. Mix by pipetting. Place the tube on a magnetic rack for 2 min. Carefully remove and discard the supernatant.
6. Resuspend bead-Hi-C sample in 50 μl EB buffer (*see* **Note 13**).

3.12 Sequencing Library Preparation

The following steps describe the Hi-C sequencing library preparation using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) and NEBNext Multiplex Oligos for Illumina (NEB).

3.13 End Repair and dA-Tailing

1. Transfer the 50 μl bead-Hi-C sample to a 0.2 ml PCR tube.
2. Add 7 μl NEBNext Ultra II End Prep Reaction Buffer and 3 μl NEBNext Ultra II End Prep Enzyme Mix. Pipette up and down at least 10 times to mix thoroughly, be careful to avoid bubbles. Perform a quick spin to collect all liquid from the sides of the tube.
3. Place in a PCR thermocycler with the heated lid set to ≥75 °C and run the following program: 30 min 20 °C, 30 min 65 °C, hold at 10 °C.

3.14 Adaptor Ligation

1. To ligate the adaptor, add in the following order: 2.5 μl 15 μM NEBNext Adaptor for Illumina, 1 μl NEBNext Ligation Enhancer, and 30 μl NEBNext Ultra II Ligation Master Mix. Since the NEBNext Ultra II Ligation Master Mix is very viscous, ensure adequate mixing by pipetting up and down at least 10 times. Perform a quick spin to collect all liquid from the sides of the tube. Incubate for 15 min at 20 °C in a PCR thermocycler with the heated lid off (*see* **Note 14**).
2. Add 3 μl USER enzyme to the ligation mixture and mix by pipetting. Incubate 15 min at 37 °C in a PCR thermocycler with the heated lid set to ≥47 °C.
3. Transfer sample to 1.5 ml LoBind microcentrifuge tube. Place the tube on a magnetic rack for 2 min. Carefully remove and discard the supernatant.
4. Wash beads twice with 600 μl 1× B&W + 0.1% Triton X-100. Mix by pipetting. Place the tube on a magnetic rack for 2 min. Carefully remove and discard the supernatant.
5. Wash beads with 600 μl EB buffer. Mix by pipetting. Place the tube on a magnetic rack for 2 min. Carefully remove and discard the supernatant.
6. Resuspend beads in 40 μl EB buffer.

3.15 PCR Amplification (PCR 1)

1. Amplify the Hi-C library by setting up a first PCR reaction (PCR 1). Transfer 20 μl bead-Hi-C library suspension to a 0.2 ml PCR tube, add 25 μl 2× NEBNext Ultra II Q5 Master Mix, 2.5 μl 10 μM NEBNext Universal PCR Primer, and 2.5 μl 10 μM NEBNExt Index Primer (sample-specific). Mix by pipetting up and down at least 10 times. Perform a quick spin to collect all liquid from the sides of the tube. Store the remaining 20 μl Hi-C library bead suspension at −20 °C, for potential troubleshooting or later amplification.
2. Place in a PCR thermocycler and run the following program: 1 cycle: 3 min 98 °C (initial denaturation); 3–5 cycles: 30 s 98 °C (denaturation), 30 s 63 °C (annealing/extension), 40 s 72 °C (final extension); hold at 10 °C. The number of PCR

cycles should be chosen based on input amount and thus may need to be optimized.

3. Transfer the sample to a 1.5 ml LoBind microcentrifuge tube. Place the tube on a magnetic rack for 5 min. Transfer the supernatant containing the amplified Hi-C library to a new 1.5 ml Lo-Bind microcentrifuge tube (*see* **Note 15**).
4. Adjust the volume of the amplified Hi-C library to 50 μl with EB.

3.16 Removal of Adapter Dimers

1. To remove adapter dimers, add 45 μl well resuspended SPRI-select beads to 50 μl of amplified Hi-C library (0.9× ratio bead/sample). Pipette up and down at least 10 times to mix thoroughly. Incubate for 5 min at RT. Place the tube on a magnetic rack for 2 min. Carefully remove and discard the supernatant (*see* **Note 16**).
2. Wash the beads twice with 200 μl freshly prepared 80% ethanol while on the magnetic rack, do not resuspend. Incubate for 2 min at RT. Carefully remove and discard the supernatant.
3. After removal of all traces of ethanol with a 10 μl pipette, air-dry beads for 3–4 min while on the magnetic rack. Do not over-dry the beads, as this may result in a lower Hi-C library recovery.
4. Elute the amplified Hi-C library from the beads by adding 20 μl EB buffer. Mix thoroughly by vortex. Incubate for 5 min at 37 °C. Place the tube on a magnetic rack for 2 min. Transfer the supernatant to a 1.5 ml LoBind microcentrifuge tube.

3.17 Side qPCR

1. To reduce PCR-related artifacts during further amplification of the Hi-C library, the appropriate number of cycles for PCR 2 is determined by qPCR. Set up a qPCR reaction in a well of a qPCR plate by adding 5 μl 2× NEBNext Ultra II Q 5 Master Mix, 0.5 μl 10 μM NEBNext Universal Primer, 0.5 μl 10 μM NEBNext Index Primer (sample-specific, same as for PCR 1), 0.1 μl 100× SYBR Green I, 1.5 μl amplified (PCR 1) Hi-C library, and 2.4 μl ultrapure water.
2. Place in a Real-time PCR instrument and run using the following program: 1 cycle: 3 min 98 °C; 30 cycles: 30 s 98 °C, 30 s 63 °C, 40 s 72 °C.
3. Calculate the additional number of cycles for PCR 2 by plotting the linear Rn versus cycle number. The cycle number that corresponds to one-third of the maximum fluorescent intensity is the desired number of cycles for PCR 2. The number of cycles may vary between samples and some may not need additional amplification (*see* **Note 17**).

3.18 PCR Amplification (PCR 2)

1. If further amplification of the Hi-C library is needed, set up a second PCR reaction (PCR 2) in a 0.2 ml PCR tube by adding 25 µl 2× NEBNext Ultra II Q5 Master Mix, 2.5 µl 10 µM NEBNext Universal PCR Primer, 2.5 µl 10 µM NEBNExt Index Primer (sample-specific, same as for PCR 1), 18.5 µl amplified (PCR 1) Hi-C library and 1.5 µl ultrapure water. Mix by pipetting.
2. Place in a PCR thermocycler and run the following program: 1 cycle: 3 min 98; N cycles: 30 s 98 °C, 30 s 63 °C, 40 s 72 °C; hold at 10 °C, where N is the number of cycles calculated from the side qPCR (*see* **Note 18**).
3. Adjust the volume of the amplified Hi-C library to 50 µl with EB.

3.19 Size Selection of Amplified Hi-C Library

1. To remove adapter dimers and narrow the amplified Hi-C library size range (depends on sequencing read length), add 35 µl well resuspended SPRIselect beads to 50 µl amplified Hi-C library (0.7× ratio bead/sample, to remove <250 bp. Pipette up and down at least 10 times to mix thoroughly. Incubate for 5 min at RT. Place the tube on a magnetic rack for 2 min. Carefully remove and discard the supernatant (*see* **Note 19**).
2. Wash the beads twice with 200 µl freshly prepared 80% ethanol while on the magnetic rack, do not resuspend. Incubate for 2 min at RT. Carefully remove and discard the supernatant.
3. After removal of all traces of ethanol with a 10 µl pipette, air-dry the beads for 3–4 min while on the magnetic rack. Do not over-dry the beads, as this may result in a lower Hi-C library recovery.
4. Elute the final Hi-C library from the beads by adding 20 µl EB buffer. Mix thoroughly by vortex. Incubate for 5 min at 37 °C. Place the tube on a magnetic rack for 2 min. Transfer the supernatant to a 1.5 ml LoBind microcentrifuge tube.
5. Store the final Hi-C library at −20 °C.

3.20 Library Quality Assessment

1. Quantify the concentration of the final Hi-C library by Qubit or qPCR.
2. Analyze fragment size and molarity by Bioanalyzer using a High Sensitivity DNA Kit.

3.21 Sequencing

The final Hi-C library is sequenced using sequencing by synthesis on a standard Illumina platform. Paired-end sequencing enables both ends of the DNA fragment to be sequenced and reads of at least 100 bp are recommended. For *A. thaliana*, a minimum of 200 million paired-end sequenced reads is needed to obtain sufficient data for subsequent bioinformatic analysis steps. Higher

sequencing depth will generally result in higher resolution Hi-C interaction maps (*see* **Note 20**).

3.22 Bioinformatic Analysis

The following steps are described as if processing only one sequencing file but should be applied to each Hi-C sequenced library. For steps such as matrix quality control and correction, or differential interactions identification, sequencing of two biological replicates (independent plant material collections and Hi-C library preparations) for each sample (either from cotyledons or roots) are required.

3.22.1 Software Requirements for Data Analysis

Given the large size of Hi-C sequenced data, these analyses should be carried out in a high-performance computer suitable for genomic analysis. The first step is to install the following software on this computer. If the computer is centrally managed, there may be a protocol in place to request software installs. The person carrying out the analysis should have basic knowledge of the Unix shell Bash and the programming language R.

1. Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml#obtaining-bowtie-2).
2. HiCUP (https://www.bioinformatics.babraham.ac.uk/projects/hicup/read_the_docs/html/index.html#installation).
3. HiCExplorer (https://hicexplorer.readthedocs.io/en/latest/content/installation.html).
4. Juicer_tools (https://github.com/aidenlab/juicer/wiki/Download).
5. Cooler (https://cooler.readthedocs.io/en/latest/quickstart.html#installation).
6. Bam2pairs (https://github.com/4dn-dcic/pairix/tree/master/util/bam2pairs).
7. diffHic (https://www.bioconductor.org/packages/release/bioc/html/diffHic.html).
8. edgeR (http://bioconductor.org/packages/release/bioc/html/edgeR.html).
9. csaw (https://bioconductor.org/packages/release/bioc/html/csaw.html).
10. GenomicRanges (https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html).
11. samtools (https://github.com/samtools/samtools).
12. statmod (https://cran.r-project.org/web/packages/statmod/index.html).

3.22.2 Read Alignment and Filtering

The first step in Hi-C data analysis is to align the sequencing reads to the reference genome. HiCUP is an automatic pipeline that maps and filters paired-end reads derived from Hi-C ligation products [15]. Two files must be generated to map and filter Hi-C reads with HiCUP: (1) a mapping index and (2) a digested genome file. HiCUP can use either bowtie or bowtie2 to map reads. Here, we align the reads using bowtie2 [21].

In this protocol, we use generic file names for the commands which should be substituted with the names of actual files. For example, we use *your_genome.fa*, which should be substituted by the name of the file containing the genome of interest. These generic file names are in italics in the commands below.

1. Generate a bowtie2 mapping index for the genome.

```
$ bowtie2-build --threads 1 your_genome.fa your_genome_bt2idx
```

   Output files: Six bowtie2 index files whose names start with the prefix your_genome_bt2idx and end with the suffix bt2.

2. Generate an in silico digested genome file. Indicate with the character "^" the restriction enzyme recognition site and the restriction enzyme name separated by a comma.

```
$ hicup_digester --re1 ^GATC,DpnII --genome your_genome your_-
genome.fa
```

   Output file: A text file containing the digested genome, *Digest_your_genome_DpnII.txt*.

3. Run the complete pipeline with the `hicup` command to align and filter read pairs. The index files, the digested genome and the samples files should be in the working directory, otherwise indicate the path on the command line.

```
$ hicup --bowtie2 /your/path/to/bowtie2 --index your_gen-
ome_bt2idx --digest Digest_your_genome_DpnII.txt --longest
800 --shortest 100 --threads 1 --zip cotyledon_rep1_1.fq.gz
cotyledon_rep1_2.fq.gz
```

   Output files: A bam file ready for downstream analysis, *cotyledon_rep1_1_2.hicup.bam* and an html file summarizing the pipeline results, *cotyledon_rep1_1_2.HiCUP_summary_report.html*.

   Repeat this step for every library. Note that each library is made up of two fastq files (one per read pair).

3.22.3 Obtain Pairs File from the Bam Generated with HiCUP

Convert the bam containing filtered Hi-C pairs to the pairs format used by downstream Hi-C analysis tools. The pairs file is a standard format proposed by the 4DNucleome consortium [22]. We will use the bam2pairs command to obtain a .pairs file.

1. Generate a chromosome file containing the chromosome names and lengths, separated by a tab. An example file:

```
chr1    1000000
chr2    2000000
```

   This information is generally included in the header of the alignment bam files in the @SQ fields and can be visualized using the following command.

```
$ samtools view -H cotyledon_rep1_1_2.hicup.bam | grep @SQ
```

   Output file: *chr_file.txt*

2. Convert the bam file to a .pairs file.

```
$ bam2pairs -l -c chr_file.txt cotyledon_rep1_1_2.hicup.bam
cotyledon_rep1_1_2.hicup
```

   Output file: cotyledon_rep1_1_2.hicup.bsorted.pairs.

3.22.4 Bin Read Pairs to Obtain a Contact Matrix

The next step in the workflow is to aggregate the read level pairs into bins. Once the data is binned, other formats are used to store the matrix data. We will bin and store matrices using two common tools: juicer_tools, which stores the resulting matrix in the .hic format, and cooler, which uses the .cool format. Both formats are binary containers for Hi-C data [16, 18]. When binning the matrix, a high-resolution bin size (1–10 kb) is recommended because a lower resolution (>10 kb) can be easily obtained by summing adjacent bins. The .cool files store a single matrix at a particular resolution.

1. Convert the .pairs file to a .cool file binned at 10 kb resolution.

```
$ cooler cload pairs -c1 2 -p1 3 -c2 4 -p2 5 chr_file.txt:10000
cotyledon_rep1_1_2.hicup.bsorted.pairs  cotyledon_rep1_10k.
cool
```

   Output file: *cotyledon_rep1_10k.cool*.

3.22.5 Normalize Matrices to Account for Differences in Sequencing Depth Between Samples

When it is of interest to compare multiple matrices, differences in sequencing depth between experiments must be considered. To do this, we will use the hicNormalize function from HiCExplorer to adjust the matrices, so the total sum is equal to the matrix with lower sequencing depth [17].

In the following commands, we use 10 kb binned cool files for all the samples. In this case we have cotyledon_rep1, cotyledon_rep2, root_rep1, and root_rep2.

1. Normalize matrices to make them comparable in terms of sequencing depth.

```
$ hicNormalize --matrices cotyledon_rep1_10k.cool cotyledon_rep2_10k.cool root_rep1_10k.cool root_rep2_10k.cool --normalize smallest -o cotyledon_rep1_10k_norm.cool cotyledon_rep2_10k_norm.cool root_rep1_10k_norm.cool root_rep2_10k_norm.cool
```

Output files: One cool file with the _norm.cool suffix for each sample.

3.22.6 Correct Matrices to Account for Underlying Biases

Matrix correction is necessary to account for biases such as GC content or mappability. A typical matrix transformation is the iterative correction strategy [23]. It works under the assumption that, if an experiment was unbiased, all bins should have equal visibility of contacts. Iterative correction results in a matrix where the sum of each column and each row is equal.

Bin level filtering is necessary to remove low count bins before correction. To decide filtering values for this filter, we can run a diagnostic plot of a histogram of counts per bin. It is important to remove bins with a low number of contacts.

1. Generate a histogram of counts per bin.

```
$ hicCorrectMatrix diagnostic_plot --matrix cotyledon_rep1_10k_norm.cool -o cotyledon_rep1_10k_diagnostic.png
```

Output file: a histogram of counts per bin, *cotyledon_rep1_10k_diagnostic.png*.

Inspect the diagnostic plot and choose the cutoff values (*see* **Note 21**).

2. After deciding on minimum and maximum values, proceed to correct the matrix.

```
$ hicCorrectMatrix correct --matrix cotyledon_rep1_10k_norm.cool --correctionMethod ICE --outFileName cotyledon_rep1_10k_corrected.cool --filterThreshold -2.5 5
```

Output file: *cotyledon_rep1_10k_corrected.cool*.

Repeat **steps 1** and **2** for the rest of the samples.

3.22.7 Matrix QC

When working with an experimental design with multiple conditions and replicates, it is useful to assess how similar the replicates are and how different the conditions are (*see* **Note 22**).

1. Calculate the correlation of counts between replicates and conditions.

```
$ hicCorrelate --log1p --matrices cotyledon_rep1_10k.cool cotyledon_rep2_10k.cool root_rep1_10k.cool root_rep2_10k.cool --range 20000:500000 -oh between_matrix_cor_h.png -os between_matrix_cor_s.png
```

Output files: correlation heatmap, *between_matrix_cor_h.png* and correlation scatter plot *between_matrix_cor_s.png*.

3.22.8 Sum Replicate Matrices to Increase the Resolution

A common practice in Hi-C data analysis is to sum biological replicate matrices in order to increase sequencing depth and thus matrix resolution.

1. Sum matrices.

```
$ hicSumMatrices -m cotyledon_rep1_10k.cool cotyledon_rep2_10k.cool -o cotyledon_merge_10k.cool
$ hicSumMatrices -m root_rep1_10k.cool root_rep2_10k.cool -o root_merge_10k.cool
```

Output files: *cotyledon_merge_10k.cool* and *root_merge_10k.cool* contain the merged replicates for each sample.

For comparative analyses, the normalization and correction steps from Subheadings 3.22.5 and 3.22.6 should be applied to the merged matrices. Up to this point, we obtained matrices merged, binned and normalized at 10 kb. Repeat Subheadings 3.22.4 through 3.22.7, changing the resolution value to obtain matrices in other resolutions. In the following sections, we will be working with matrices with 50 kb and 500 kb resolution.

3.22.9 Build a .hic Matrix

Another Hi-C storage format is the .hic format, used by juicer, juicer_tools and juicebox [18]. It is a binary format that stores a Hi-C matrix with multiple bin sizes and corrections in a single file. The `juicer_tools pre` command bins and corrects the Hi-C matrix at several resolutions.

1. Generate the .hic matrix.

```
$ java -Xmx1G -jar /path/to/your/juicer_tools_1.13.02.jar pre cotyledon_rep1_1_2.hicup.bsorted.pairs cotyledon_rep1.hic chr_file.txt
```

Output file: The *cotyledon_rep1.hic* file stores the matrix at various resolutions, as well as different corrections. This file can be directly uploaded to juicebox for visualization and is ready to use with juicer_tools, which we will do in Subheading 3.22.13.

3.22.10 Visualize the Hi-C Matrices

Use the HiCExplorer command-line tools to generate visualizations of Hi-C matrices [17]. In this section, we are using matrices merged, binned, and corrected at 50 and 500 kb resolution.

1. Plot a large region.

```
$ hicPlotMatrix --perChromosome --log1p --matrix cotyledon_-
merge_500kb_corrected.cool --outFileName cotyledon_mer-
ge_500kb_corrected.png
```

Output file: *cotyledon_merge_500kb_corrected.png* is a plot of the corrected matrix per chromosome.

2. Plot a small region.

```
$ hicPlotMatrix --log1p --region 2:125000000-130000000 --ma-
trix cotyledon_merge_50kb_corrected.cool --outFileName coty-
ledon_merge_50kb_corrected.png
```

Output file: *cotyledon_merge_50kb_corrected.png* is a plot showing the region 2:125000000-130000000.

3. Obtain and plot a matrix containing the differences between conditions.

```
$ hicCompareMatrices --operation log2ratio --matrices coty-
ledon_merge_500kb_corrected.cool root_merge_500kb_corrected.
cool --outFileName root_cotyledon_500kb_log2.cool
```

Output file: *root_cotyledon_500kb_log2.cool* is a cool file containing the difference between root and cotyledon.

```
$ hicPlotMatrix --perChromosome --matrix root_cotyledon_500k-
b_log2.cool --outFileName root_cotyledon_500kb_log2.png
```

Output file: *root_cotyledon_500kb_log2.png* is a plot per chromosome of the difference matrix.

3.22.11 Identify A/B Compartments

Intrachromosomal contacts are segregated into transcriptionally active (A compartment) and inactive (B compartment) regions [7]. This is generally done using principal component analysis (PCA), a method that reduces the global interaction patterns to a single vector, or principal component, that captures most of the

variability between compartments. Each bin is labeled as either A or B compartment based on the sign of the principal component value for that bin (*see* **Note 23**).

1. Obtain the principal component to define compartments.

```
$ hicPCA -noe 1 --matrix cotyledon_merge_500kb_corrected.cool
--format bigwig -o cotyledon_merge_500kb_pca1.bw
```

Output file: *cotyledon_merge_500kb_pca1.bw* is a bigWig file containing the PC1 of the matrix.

2. Plot the first principal component along with the matrix.

```
$ hicPlotMatrix -m cotyledon_merge_500kb_corrected.cool -o
cotyledon_merge_compartments.png --log1p --bigwig cotyledon_-
merge_500kb_pca1.bw --perChromosome
```

Output file: *cotyledon_merge_compartments.png* is a plot of the whole genome matrix per chromosome along with the PC1.

3. If histone modification information is available, for example, H3K4me3, which is generally associated with active transcription (A compartment), this information can be displayed alongside the compartments.

```
$ hicPlotMatrix --log1p -m cotyledon_merge_500kb_corrected.
cool -o cotyledon_merge_500kb_histonemod.png --perChromosome
--bigwig cotyledon_H3K4me3.bw
```

Output file: *cotyledon_merge_500kb_histonemod.png* is a plot showing the whole genome matrix per chromosome together with the H3K4me3 signal.

3.22.12 Identify TADs

Topologically Associating Domains are defined as regions of increased self-interaction in Hi-C maps [6]. Several computational approaches have been developed to identify them. Here, we use hicFindTADs from HiCExplorer to identify TADs [17].

1. Identify TADs with hicFindTADs.

```
$ hicFindTADs -m cotyledon_merge_50kb_corrected.cool --out-
Prefix cotyledon_merge_50kb_tads --correctForMultipleTesting
fdr
```

Output files: *cotyledon_merge_50kb_tads* is a directory containing a list of domain boundaries in bed and gff formats, a list of domains in bed format, the TAD separation score in

bedgraph and a z score matrix calculated during the TAD calling procedure.

2. The hicPlotTADs function requires a file with the track information. The extension of this track configuration file is ".ini". Generate a .ini configuration file for TAD visualization. Copy the information below, between the hashes (#) to a text file called hic_tads.ini.

```
#######
[x-axis]
where=top
[hic matrix]
file = cotyledon_merge_50kb_tads
title = Hi-C data
depth = 1000000
transform = log1p
file_type = hic_matrix
[tads]
file = cotyledon_merge_50kb_tads_domains.bed
file_type = domains
border color = black
overlay previous = share-y
[spacer]
[tad score]
file = cotyledon_merge_50kb_tads_tad_score.bedgraph
title = "TAD separation score"
file_type = bedgraph
#######
```

3. Visualize TADs alongside the matrix.

```
$ hicPlotTADs --tracks hic_tads.ini -o cotyledon_50k_tads.png
--region 2:122000000-126000000
```

Output file: *cotyledon_50k_tads.png* is a plot showing the TADs and TAD separation score along the matrix focused in the region from position 122,000,000 to 126,000,000 of chromosome 2.

3.22.13 Identify Interaction Peaks

Interaction peaks are regions of high interaction frequency between two distant genomic regions. A standard tool for peak calling is HiCCUPs, available from the juicer_tools toolkit [18]. To use this tool, we need the .hic matrix generated previously in Subheading 3.22.9.

1. Identify peaks using HiCCUP.

```
$ java -jar /usr/local/src/juicer/juicer_tools_1.13.02.jar
hiccups --cpu --threads 2 -r 10000 cotyledon_rep1.hic -k KR
cotyledon_hiccups_loops
```

Output files: *cotyledon_hiccups_loops* is a directory containing the merged_loops file containing the final list of identified loops. Intermediate processing files will also be saved in this output directory.

2. Generate an aggregated peak plot. This plot is useful to get an overview of all the peaks at once.

```
$ java -jar /usr/local/src/juicer/juicer_tools_1.13.02.jar apa
-r 10000 cotyledon_rep1.hic cotyledon_hiccups_loops cotyle-
don_hiccups_apa
```

Output files: the *cotyledon_hiccups_apa* directory will contain an APA.png file with the aggregate signal across all loops.

3.22.14 Identifying Statistically Significant Differential Interactions

An additional strategy when analyzing chromatin conformation data is to identify changes in interaction intensity that are statistically significant between two or more biological conditions. Various publicly available tools identify these Differential Interactions (Dis) from Hi-C data, such as FIND [24], HOMER [25], and HiBrowse [26]. We will identify differential interactions using the diffHic package [19].

1. Sort bam files by read name. Do this for each bam file before going to the next step.

```
$ samtools sort -n my_bam_file.bam > my_bam_file.hicup.sorted.
bam ; done
```

In the following steps, use the R console.

2. Load required libraries.

```
> Packages <- c("diffHic", "GenomicRanges", "edgeR", "csaw")

> lapply(Packages, library, character.only=T)
```

3. Import HiCUP file with digested genome into R (generated in Subheading 3.22.2).

```
> digest <- read.csv("Digest_your_genome_DpnII.txt.", head-
er=T, sep="\t", skip=1)
```

4. Generate the object hic_experiment.frag with the digested genome in the format required by diffHic.

```
> hic_experiment.frag <- with(digest,GRanges(Chromosome, IR-
anges(Fragment_Start_Position,Fragment_End_Position)))
```

5. Generate a pairParam object to store the fragments and other parameters.

```
> hic_experiment.param <- pairParam(hic_experiment.frag)
```

6. Create h5 files to count Hi-C reads into bins. This process matches the mapping location of each read to a restriction fragment in the reference genome.

```
# Cotyledon samples.

> preparePairs("cotyledon_rep1_1_2.hicup.sorted.bam", hic_experiment.param, file="cotyledon1.h5")
> preparePairs("cotyledon_rep2_1_2.hicup.sorted.bam", hic_experiment.param, file="cotyledon2.h5")

# Root samples.

> preparePairs("root_rep1_1_2.hicup.sorted.bam", hic_experiment.param, file="root1.h5")
> preparePairs("root_rep2_1_2.hicup.sorted.bam", hic_experiment.param, file="root2.h5")

# Generate input object.

> input <- c("cotyledon1.h5","cotyledon2.h5","root1.h5","root2.h5")
```

7. Count reads that fall within each genomic bin. Choose a bin size and count read pairs between paired bins for the four libraries with squareCounts using input, contained in the hic_experiment_data object.

```
> bin.size <- 50000
> hic_experiment_data <- squareCounts(input, hic_experiment.param , width=bin.size, filter=1)
```

The following steps are necessary to filter noninformative bin pairs.

8. Plot the log-NBmean-per-million (average abundance).

```
> ave.ab <- aveLogCPM(asDGEList(hic_experiment_data))
> hist(ave.ab, xlab="Average abundance", col="powderblue")
```

9. One filtering strategy is to keep only bin pairs with abundances x-times higher (3 in this example) than the median abundance across interchromosomal bin pairs, as the majority of these represent false interactions.

```
> direct <- filterDirect(hic_experiment_data)
> direct.keep <- direct$abundances > log2(3) + direct$thresh-
old
> summary(direct.keep)
> log2(3) + direct$threshold
```

10. Apply filter to data object. This will eliminate all rows that are not named in the object `direct.keep`.

```
> hic_experiment_data <- hic_experiment_data[direct.keep, ]
```

11. Visualize filtered data.

```
> ave.ab <- aveLogCPM(asDGEList(hic_experiment_data))
> hist(ave.ab, xlab="Average abundance", col="blue")
```

diffHic allows for various library normalization strategies, the simplest being library size normalization. In the next steps, we will use nonlinear normalization (LOESS) to account for trended biases between libraries.

12. Compare one library of each sample group using an MA plot (in this case 1 and 4). The fitted line on the plot shows that there is an abundance-dependent trend.

```
> ab <- aveLogCPM(asDGEList(hic_experiment_data))
> o <- order(ab)
> adj.counts <- cpm(asDGEList(hic_experiment_data), log=TRUE)
> mval <- adj.counts[,1]-adj.counts[,4]
> smoothScatter(ab, mval, xlab="A", ylab="M", main="Cotyledon
vs Root")
> fit <- loessFit(x=ab, y=mval)
> lines(ab[o], fit$fitted[o], col="red")
```

13. Apply normalization.

```
> hic_experiment_data <- normOffsets(hic_experiment_data)
```

14. Store the matrix of offsets in a separate object.

```
> nb.off <- assay(hic_experiment_data, "offset")
```

15. Adjust the log counts with the offsets and generate another MA plot to evaluate the normalization. We should see that the trend is removed.

```
> ab <- aveLogCPM(asDGEList(hic_experiment_data))
> o <- order(ab)
> adj.counts <- log2(assay(data) + 0.5) - nb.off/log(2)
> mval <- adj.counts[,1]-adj.counts[,4]
> smoothScatter(ab, mval, xlab="A", ylab="M", main="Cotyledon
vs Root after NLN")
> fit <- loessFit(x=ab, y=mval)
> lines(ab[o], fit$fitted[o], col="red")
```

16. Create a design matrix that describes the experimental setup. In this case we have two conditions (cotyledon and root) with two replicates each.

```
> design <- model.matrix(~factor(c("cotyledon", "cotyledon",
"root", "root")))
> colnames(design) <- c("Intercept", "root")
```

17. Convert the `hic_experiment_data` object to a DGEList object to analyze it with edgeR.

```
> y <- asDGEList(hic_experiment_data)
```

18. The variability between replicates of the same condition is estimated using the dispersion parameter of the Negative Binomial (NB) distribution. Estimate the dispersion.

```
> y <- estimateDisp(y, design)
> plotBCV(y)
```

19. Estimate the Quasi-likelihood dispersion.

```
> fit <- glmQLFit(y, design, robust=TRUE)
> plotQLDisp(fit)
```

20. Identify differential interactions with the quasi-likelihood *F*-test. This test will evaluate the statistical significance of each differential interaction and provide a *p*-value and an adjusted *p*-value, or false discovery rate (FDR) for each of them.

```
> result <- glmQLFTest(fit, coef=2)
> topTags(result)
```

21. Save significance statistics in the variable rowData of the InteractionSet object.

```
> rowData(hic_experiment_data) <- cbind(rowData(hic_experi-
ment_data), result$table)
```

22. Plot the total of differential interactions in a smear MA plot.

```
> de <- decideTestsDGE(result, p.value=0.05, adjust.method="BH")
> debins <- rownames(result)[as.logical(de)]
> plotSmear(result, de.tags=debins)
```

23. We can cluster those bin pairs that are adjacent and significant to avoid redundancy so that each cluster contains only statistically significant bins that correspond to a differential interaction. Here, we are clustering bins that are right next to one another (tol = 1).

```
> clustered.sig <- diClusters(hic_experiment_data, result$table, target=0.05, cluster.args=list(tol=1))
> length(clustered.sig$interactions)
> head(clustered.sig$interactions)
> clustered.sig$FDR
```

24. Using the indices of the bin pairs we can use the combineTests function to calculate the combined p-value for the cluster.

```
> tabcomdata <- combineTests(clustered.sig$indices[[1]], result$table)
> head(tabcomdata)
```

25. Using the same indices, we can also use getBesetTest to identify the bin pair with the most significant p-value within a cluster.

```
> tabbestdata <- getBestTest(clustered.sig$indices[[1]], result$table)
> head(tabbestdata)
```

26. Save the coordinates and statistics for each differential interaction.

```
> tabstat <- data.frame(tabcomdata[,,], logFC=tabbestdata$logFC, FDR=clustered.sig$FDR)
> result.d <- as.data.frame(clustered.sig$interactions)[,c("seqnames1","start1","end1","seqnames2","start2","end2")]
> result.d <- cbind(result.d, tabstat)
> o.d <- order(result.d$PValue)
> write.table(result.d[o.d,], file="DI_ClustersData.tsv", sep="\t", quote=FALSE, row.names=FALSE)
```

4 Notes

1. When preparing NIB, NIB-FA and NIB-P, add all the components in the indicated order. PMSF, 2-mercaptoethanol, and formaldehyde should be added prior to use, under a fume hood. Prepare the protease inhibitor cocktail using cOmplete ULTRA Tablets (Roche) and add prior to use.
2. It is recommended to collect plant material quickly and cross-link with formaldehyde immediately. In this case, harvesting and separating organs of approximately 6900 seedlings is done in maximum 20 min, involving two persons.
3. Formaldehyde is oxidized to formic acid under normal atmospheric oxygen concentrations. Therefore, preferably use pure, methanol-free, ampule-sealed formaldehyde solution. Opened ampules should be resealed using Parafilm and stored at 4 °C for no longer than a week. Poor-quality formaldehyde will adversely affect the experiment.
4. To avoid air bubbles, add NIB-P slowly to the sample. To prevent chromatin degradation, precool 50 ml centrifuge tubes, 1.5 ml microcentrifuge tubes, and NIB-P buffer on ice. Always keep the samples on ice.
5. Incubation with SDS will increase chromatin accessibility for better restriction digestion and inactivation of endogenous nucleases. The duration of incubation may need to be optimized in a sample dependent manner. Shorter incubation time may result in inefficient or partial digestion due to the chromatin being inaccessible to the restriction enzyme. Longer incubation time may lead to excessive digestion, alteration of chromatin territories and may even reverse cross-links [14].
6. It is important to maintain a 6–10× ratio Triton X-100/SDS, as nonadequate SDS quenching can inhibit the enzymatic activities of the enzymes used in downstream steps.
7. Hi-C experiments can be done using 6-cutter (i.e., HindIII) or 4-cutter (i.e., DpnII) restriction enzymes, the latter generating genome-wide chromosomal contact maps with higher resolution [27]. The MboI 4-cutter enzyme, that recognizes the same sequence as DpnII, has been used in plants such as rice, foxtail millet, sorghum, tomato, and maize [28]. When using different enzymes, concentrations and incubation times may need to be optimized.
8. The method of enzyme inactivation is specific for DpnII. Use appropriate inactivation conditions if using other enzymes.
9. Thaw 10 μl control sample, add 82 μl EB buffer and 1 μl 20 mg/ml RNase A. Incubate for 30 min at 37 °C. Add 5 μl 10% SDS and 2 μl 20 mg/ml proteinase K. Incubate overnight

at 37 °C and 6 h at 65 °C. Before use, spin Phase Lock Gel (Quantabio) tube at 12,000 × *g* for 30 s at RT. Transfer sample to prespun Phase Lock Gel tube. Add 100 μl EB buffer and 200 μl phenol–chloroform–isoamyl alcohol (25:24:1). Mix thoroughly by vigorous shaking for 2 min to form a transiently homogenous suspension, do not vortex. Centrifuge at 12,000 × *g* for 5 min at RT. Perform a second extraction, by adding 200 μl chloroform to the same Phase Lock Gel tube. Mix thoroughly by vigorous shaking for 2 min, do not vortex. Centrifuge at 12,000 × *g* for 5 min at RT. Transfer DNA-containing aqueous upper phase to a 1.5 ml microcentrifuge tube. Precipitate the DNA by adding 1/10 volume 3 M sodium acetate pH 5.2, 1 μl glycogen, and 2.5 volumes ice-cold 100% ethanol. Mix by inverting 5 times and incubate at −80 °C for 1 h. Centrifuge at 20,000 × *g* for 20 min at 4 °C. Remove the supernatant. Wash the pellet twice with 1 ml 70% ethanol. Centrifuge at 20,000 × *g* for 5 min at RT. Remove the supernatant and air-dry pellet. Dissolve DNA pellet in 10 μl EB buffer. Measure DNA concentration using a NanoDrop spectrophotometer. Load 500 ng of sample and a 1 Kb Plus DNA Ladder on a 1.8% agarose gel for electrophoresis. Run the gel for 40 min at 90 V. Undigested (or intact) chromatin will be seen as a tight high molecular weight band (>10 kb), while digested chromatin will appear as a smear from 100 bp to 3 kb approximately. For rapid reversal of chromatin cross-linking, thaw 10 μl of control sample, add 83 μl EB buffer, 4 μl 5 M NaCl, 2 μl 20 mg/ml proteinase K, and 1 μl 20 mg/ml RNase A. Incubate for 1 h at 65 °C. Proceed as above with DNA extraction using Phase Lock Gel tubes.

10. The temperature and incubation time of the ligation reaction can be optimized.
11. Inefficient removal of biotin from unligated ends can lead to sequencing of reads from unwanted dangling-end products and not from real interactions [14].
12. DNA shearing target length depends on sequencing strategy and sequencing read length. The shearing settings depend on the equipment used and can be optimized.
13. Samples can be stored at −20 °C.
14. The appropriate adaptor concentration may need to be optimized depending on sample input amount.
15. The original Hi-C library is bound to the magnetic beads as a single biotinylated strand of the hybrid molecule. After PCR 1 the beads can be resuspended in 200 μl of EB buffer and stored at 4 °C for later troubleshooting. If needed, the Hi-C library can be amplified again setting up a new PCR reaction on these beads.

16. It is recommended to remove primers (<85 bp) and adaptor dimers (~127 bp) from the amplified Hi-C library. Due to their short size, the latter may be preferentially amplified during the following qPCR based library quantification, impairing the accurate determination of the library concentration.
17. This step helps minimize PCR-related artifacts, such as over-amplification and reduced library complexity, or GC and size bias, during further Hi-C library amplification [29].
18. The Hi-C library amplification should provide sufficient fragments for high-throughput sequencing, minimizing PCR-related artifacts.
19. The amplified Hi-C library size range depends on the sequencing strategy. For 2x150 bp (paired-end) Illumina sequencing the optimal library fragment size (insert +120 bp adaptors) is ~320–520 bp. A double size selection with SPRIselect beads may be performed if removing larger fragments (>600 bp) is needed for narrowing the library size. Importantly, it is recommended to remove primers (<85 bp) and adaptor dimers (~127 bp) from the library. Primers cannot cluster or be sequenced but can bind to the flow cell and reduce cluster density. On the other hand, adapter dimers will cluster and be sequenced if present in the library. The beads from all steps can be stored in 80% ethanol at 4 °C for later troubleshooting.
20. If other plant species are used, the sequencing depth has to be adjusted according to the genome size. It is worth noting that the resolution obtained in Hi-C does not vary linearly with genome size (our observations). Thus, it is recommended to use a higher sequencing depth than inferred by extrapolating the number of recommended reads for the *A. thaliana* genome.
21. The histogram of counts per bin should show two modes in the distribution, the first one around zero and the second one around the mean number of contacts per bin. To filter bins with low counts, the lower threshold value selected should be in the valley between the zero and the mean. The upper threshold should be selected based on the upper bound of the counts distribution. Note that the diagnostic plot will include suggested values for each threshold.
22. Replicates should have a higher correlation than conditions. It can be useful to make this analysis with different bin sizes to see if the correlation holds, as higher resolutions may be noisier. Correlation analysis of Hi-C counts is challenging because, as the distance between interacting bins increases, the average counts decrease and are more variable. For this reason, hicCorrelate has the --range option, to limit the distance range of the comparison. Other correlation strategies explicitly designed for

Hi-C data that take into account the distance effect are discussed in [30].

23. An alternative approach for compartment identification in *A. thaliana* is to analyze each chromosomal arm separately [31]. This is because performing principal component analysis in the whole chromosome generally identifies only three compartments, separating the euchromatic arms from the heterochromatin. Excluding the pericentromeric region improves the identification of informative subcompartments on the chromosome arms.

Acknowledgments

FJP-d, JES-F, and AR-C were funded by fellowships from the Consejo Nacional de Ciencia y Tecnología (CONACYT). N-HW, SF-V, and KO are funded by the Newton Advanced Fellowship (No. NAF\R1\180303) awarded to SF-V. KO is supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT, CB-2016-01/285847).

References

1. Gibcus JH, Dekker J (2013) The hierarchy of the 3D genome. Mol Cell 49:773–782. https://doi.org/10.1016/j.molcel.2013.02.011
2. Felsenfeld G, Groudine M (2003) Controlling the double helix. Nature 421:448–453. https://doi.org/10.1038/nature01411
3. Deng W, Lee J, Wang H et al (2012) Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell 149:1233–1244. https://doi.org/10.1016/j.cell.2012.03.051
4. Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76:8–32. https://doi.org/10.1086/426833
5. Dixon JR, Selvaraj S, Yue F et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376–380. https://doi.org/10.1038/nature11082
6. Nora EP, Lajoie BR, Schulz EG et al (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485:381–385. https://doi.org/10.1038/nature11049
7. Lieberman-Aiden E, van Berkum NL, Williams L et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293. https://doi.org/10.1126/science.1181369
8. Cremer T, Cremer C (2006) Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories. Eur J Histochem 50:161–176. https://www.ncbi.nlm.nih.gov/pubmed/16920639
9. Doğan ES, Liu C (2018) Three-dimensional chromatin packing and positioning of plant genomes. Nat Plants 4:521–529. https://doi.org/10.1038/s41477-018-0199-5
10. Dekker J, Marti-Renom MA, Mirny LA (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet 14:390–403. https://doi.org/10.1038/nrg3454
11. Eagen KP (2018) Principles of chromosome architecture revealed by Hi-C. Trends Biochem Sci 43:469–478. https://doi.org/10.1016/j.tibs.2018.03.006
12. Hövel I, Louwers M, Stam M (2012) 3C technologies in plants. Methods 58:204–211. https://doi.org/10.1016/j.ymeth.2012.06.010
13. Liu C (2017) In situ hi-C library preparation for plants to study their three-dimensional

chromatin interactions on a genome-wide scale. Methods Mol Biol 1629:155–166. https://doi.org/10.1007/978-1-4939-7125-1_11

14. Padmarasu S, Himmelbach A, Mascher M et al (2019) In situ hi-C for plants: an improved method to detect long-range chromatin interactions. Methods Mol Biol 1933:441–472. https://doi.org/10.1007/978-1-4939-9045-0_28
15. Wingett S, Ewels P, Furlan-Magaril M et al (2015) HiCUP: pipeline for mapping and processing Hi-C data. F1000Res 4:1310. https://doi.org/10.12688/f1000research.7334.1
16. Abdennur N, Mirny LA (2020) Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics 36: 311–316. https://doi.org/10.1093/bioinformatics/btz540
17. Ramírez F, Bhardwaj V, Arrigoni L et al (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun 9:189. https://doi.org/10.1038/s41467-017-02525-w
18. Durand NC, Shamim MS, Machol I et al (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst 3:95–98. https://doi.org/10.1016/j.cels.2016.07.002
19. Lun ATL, Smyth GK (2015) diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. BMC Bioinformatics 16:258. https://doi.org/10.1186/s12859-015-0683-0
20. Dong P, Zhong S (2020) Characterization of plant 3D chromatin architecture, in situ Hi-C library preparation, and data analysis. Methods Mol Biol 2093:147–167. https://doi.org/10.1007/978-1-0716-0179-2_11
21. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. Nat Methods 9: 357–359. https://doi.org/10.1038/nmeth.1923
22. Lee S, Bakker CR, Vitzthum C et al (2022) Pairs and Pairix: a file format and a tool for efficient storage and retrieval for Hi-C read pairs. Bioinformatics, 38:1729–1731
23. Imakaev M, Fudenberg G, McCord RP et al (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods 9:999–1003. https://doi.org/10.1038/nmeth.2148
24. Djekidel MN, Chen Y, Zhang MQ (2018) FIND: difFerential chromatin INteractions detection using a spatial Poisson process. Genome Res 28:412–422. https://doi.org/10.1101/gr.212241.116
25. Heinz S, Benner C, Spann N et al (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cel 38:576–589. https://doi.org/10.1016/j.molcel.2010.05.004
26. Paulsen J, Sandve GK, Gundersen S et al (2014) HiBrowse: multi-purpose statistical analysis of genome-wide chromatin 3D organization. Bioinformatics 30:1620–1622. https://doi.org/10.1093/bioinformatics/btu082
27. Wang C, Liu C, Roqueiro D et al (2015) Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. Genome Res 25: 246–256. https://doi.org/10.1101/gr.170332.113
28. Dong P, Tu X, Chu P-Y et al (2017) 3D chromatin architecture of large plant genomes determined by local a/B compartments. Mol Plant 10:1497–1509. https://doi.org/10.1016/j.molp.2017.11.005
29. Buenrostro JD, Wu B, Chang HY et al (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol 109:21.29.1–21.29.9. https://doi.org/10.1002/0471142727.mb2129s109
30. Yardımcı GG, Ozadam H, Sauria MEG et al (2019) Measuring the reproducibility and quality of Hi-C data. Genome Biol 20:57. https://doi.org/10.1186/s13059-019-1658-7
31. Grob S, Schmid MW, Grossniklaus U (2014) Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Mol Cell 55: 678–693. https://doi.org/10.1016/j.molcel.2014.07.009