

0. Introduction

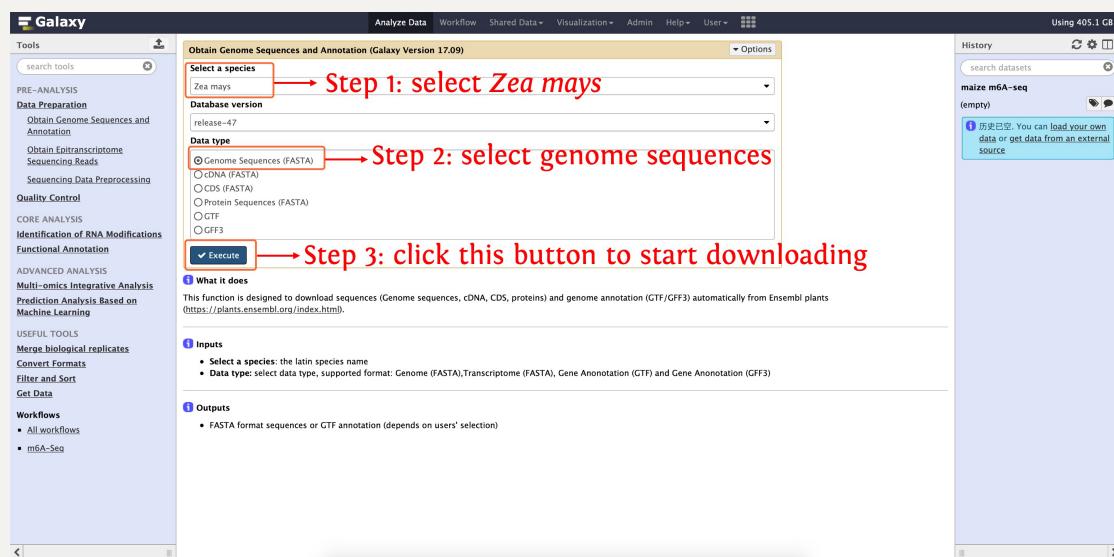
In this tutorial, we will show how to use deepEA to perform comprehensive m⁶A sequencing data analysis. The m⁶A -immunoprecipitated (IP) and input (non-IP) samples of *Zea mays* with two biological replicates were listed in the following table. More information regarding these maize m⁶A sequencing datasets is available in the reference ([Luo et al., 2020](#)).

SAMPLES	EXPERIMENTS	REPLICATES
SRR8383013	IP	Replicate 1
SRR8383014	IP	Replicate 2
SRR8383017	input	Replicate 1
SRR8383018	input	Replicate 2

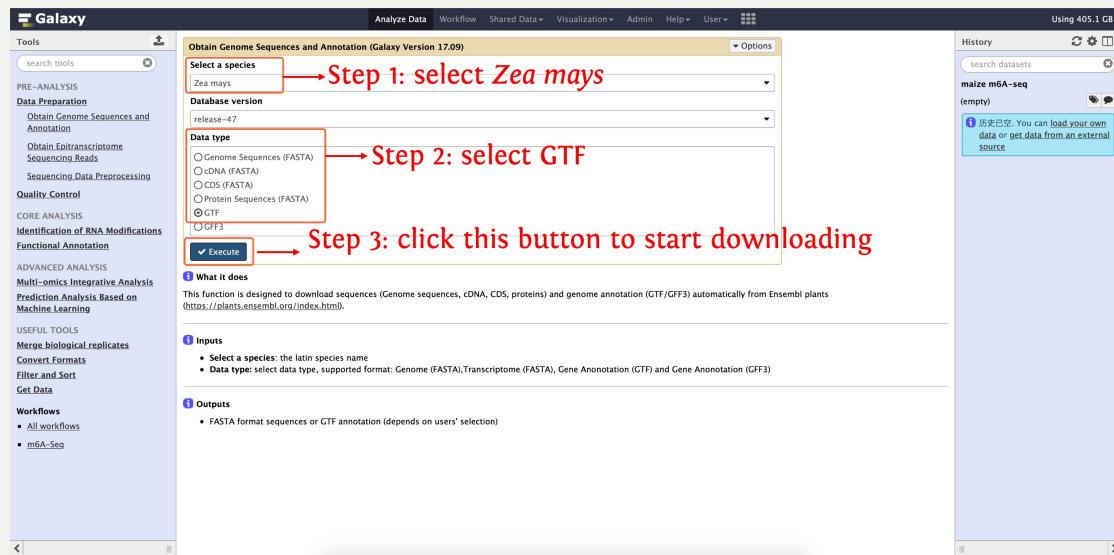
1. Download *Zea mays* reference genome sequences and annotation

Before analyzing m6A sequencing data, we firstly use the function **Obtain Genome Sequences and Annotation** in **Data Preparation** module to download *Zea mays* B73 reference genome sequences and GTF annotation, the following two screenshots shows details about how to execute this step:

Step 1: download *Zea mays* reference genome sequences in FASTA format



Step 2: download *Zea mays* genome annotation file in GTF format

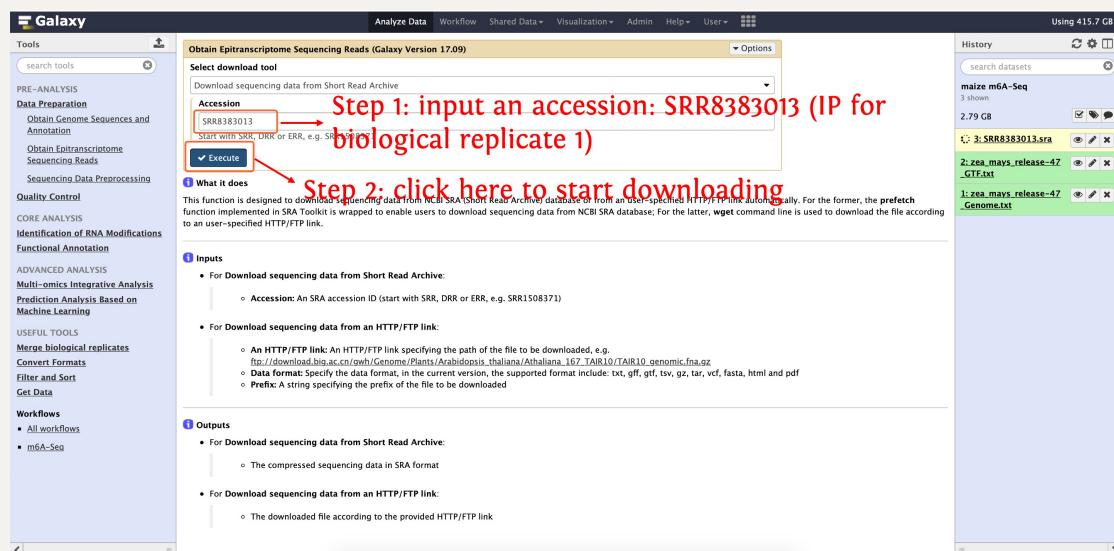


After that, reference genome sequences (named as **zea_mays_release-47_Genome.txt**) and annotation (named as **zea_mays_release-47_GTF.txt**) will be listed in your History Panel.

2. Download raw m⁶A sequencing reads

After finishing downloading *Zea mays* B73 reference genome sequences and annotation, we start to download raw m⁶A sequencing reads from NCBI SRA's database, this process can be finished by the function **Obtain Epitranscriptome Sequencing Reads** in Data Preparation module. Please see the following screenshots for details:

Step 1: download IP sample for biological replicate 1 (SRR8383013)



Step 2: download IP sample for biological replicate 2 (SRR8383014)

Step 1: input an accession: SRR8383014 (IP for biological replicate 2)

Step 2: click here to start downloading

Step 3: download input sample for biological replicate 1 (SRR8383017)

Step 1: input an accession: SRR8383017 (input for biological replicate 1)

Step 2: click here to start downloading

Step 4: download input sample for biological replicate 2 (SRR8383018)

Step 1: input an accession: SRR8383018 (input for biological replicate 2)

Step 2: click here to start downloading

The History panel on the right shows the download progress and results:

- 1: SRR8383017.sra (downloaded)
- 2: SRR8383014.sra (downloaded)
- 3: SRR8383013.sra (downloaded)
- 2: zea_mays_release-47_GTF.txt (downloaded)
- 1: zea_mays_release-47_Genome.txt (downloaded)

After finishing the above four steps, four raw m⁶A sequencing reads named as **SRR8383013.sra, SRR8383013.sra, SRR8383013.sra, SRR8383013.sra**, respectively will be listed in you History Panel

3. Convert raw m⁶A sequencing reads in SRA to FASTQ format

Here, let's convert raw m⁶A sequencing reads from SRA to FASTQ format by the function **Sequencing Data Preprocessing** in **Data Preparation** module, please see the following four screenshots for details:

Step 1: process SRR8383013.sra

Step 1: select SRR8383013.sra from History Panel

Step 2: select paired-end

Step 3: click here to run this function

The History panel on the right shows the conversion results:

- 1: SRR8383018.sra (downloaded)
- 2: SRR8383017.sra (downloaded)
- 4: SRR8383014.sra (downloaded)
- 3: SRR8383013.sra (downloaded)
- 2: zea_mays_release-47_GTF.txt (downloaded)
- 1: zea_mays_release-47_Genome.txt (downloaded)

Step 2: process SRR8383014.sra

The screenshot shows the Galaxy web interface. On the left, a sidebar lists various analysis categories like PRE-ANALYSIS, CORE ANALYSIS, and ADVANCED ANALYSIS. The main panel displays the 'Sequencing Data Preprocessing' tool. In the 'Input sra file' field, '4: SRR8383014.sra' is selected. Below it, 'Minimum reads length' is set to 20. Under 'Single-end or paired-end reads?', 'paired-end' is selected. A red box highlights the 'Execute' button. To the right is a 'History' panel showing a list of datasets: 4: SRR8383013.sra, 7: SRR8383013.sra.lfq, 6: SRR8383018.sra, 5: SRR8383017.sra, 4: SRR8383014.sra, 3: SRR8383013.sra, 2: zea_mays_release-47.GTF.txt, and 1: zea_mays_release-47.Genome.txt. A red arrow points from the 'History' panel to the 'Input sra file' field.

Step 1: select SRR8383014.sra from History Panel

Step 2: select paired-end

Step 3: click here to run this function

Step 3: process SRR8383017.sra

This screenshot is identical to the one above, showing the 'Sequencing Data Preprocessing' tool for SRR8383017.sra. The 'Input sra file' field contains '5: SRR8383017.sra'. The 'Execute' button is highlighted with a red box. The 'History' panel on the right shows the same list of datasets as the previous step.

Step 1: select SRR8383017.sra from History Panel

Step 2: select paired-end

Step 3: click here to run this function

Step 4: process SRR8383018.sra

Step 1: select SRR8383018.sra from History Panel

Step 2: select paired-end

Step 3: click here to run this function

For each SRA accession, two FASTQ format files (forward reads and reverse reads) will be generated as this experiment is paired-end sequencing.

4. Trim raw m⁶A sequencing reads

Before align m⁶A sequencing reads to genome, trim low-quality reads is necessary in all NGS (Next Generation Sequencing) analyses as which may cause incorrect mapping of reads to a reference genome, and even result in incorrect identification of RNA modifications. deepEA provided the function **Assess Reads Quality** in **Quality Control** module to filter raw reads to clean reads, please see the following screenshots for details:

Step 1: trim SRR8383013

Step 1: select paired-end

Step 2: select forward read for SRR8383013

Step 3: select reverse read for SRR8383013

Step 4: set the minimum read length as 30

Step 5: click this button to run this function

Step 2: trim SRR8383014

Step 1: select paired-end
Step 2: select forward read for SRR8383014
Step 3: select reverse read for SRR8383014
Step 4: set the minimum read length as 15
Step 5: click this button to run this function

Step 3: trim SRR8383017

Step 1: select paired-end
Step 2: select forward read for SRR8383014
Step 3: select reverse read for SRR8383014
Step 4: set the minimum read length as 30
Step 5: click this button to run this function

Step 4: trim SRR8383018

For each SRA accession, three files will be output, please see the following screenshot for detail:

5. Align clean m⁶A sequencing reads to *Zea mays* B73 reference genome with HISAT2

Here, we start to align clean reads to reference genome with HISAT2 provided in module **Identification of RNA Modifications**, see the following screenshots for details:

Step 1: align SRR8383013

Step 1: select *Zea mays* reference genome

Step 2: select paired-end

Step 3: select clean forward read for SRR8383013

Step 4: select clean reverse read for SRR8383013

Step 5: select not to extract uniquely mapped reads

Step 6: click this button to run this function

Step 2: align SRR8383014

Step 1: select *Zea mays* reference genome

Step 2: select paired-end

Step 3: select clean forward read for SRR8383014

Step 4: select clean reverse read for SRR8383014

Step 5: select not to extract uniquely mapped reads

Step 6: click this button to run this function

Step 3: align SRR8383017

Step 1: select *Zea mays* reference genome

Step 2: select paired-end

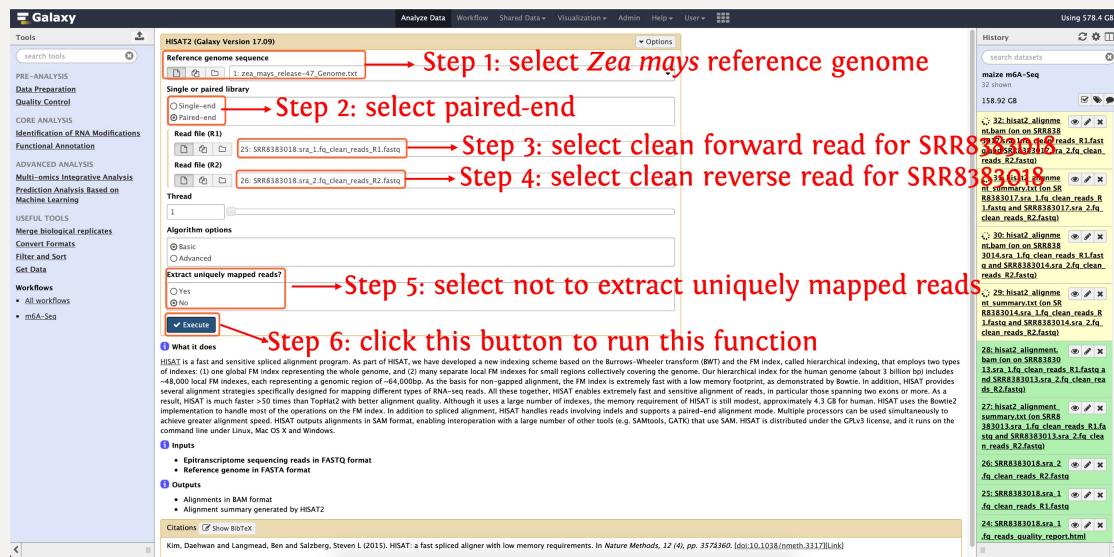
Step 3: select clean forward read for SRR8383017

Step 4: select clean reverse read for SRR8383017

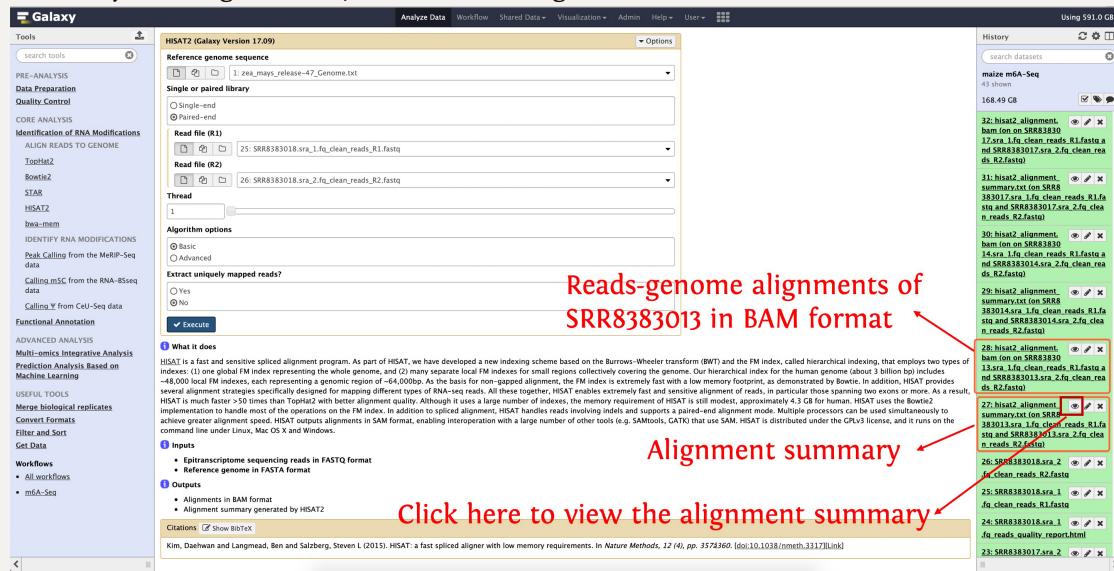
Step 5: select not to extract uniquely mapped reads

Step 6: click this button to run this function

Step 4: align SRR8383018



Then for each SRA accession, reads-genome alignments in BAM format and alignment summary will be generated, see the following screenshot:



6. Call m⁶A enriched peaks with macs2

After finishing aligning reads to genome, let's start to call m⁶A enriched peaks with macs2, the following screenshots show details about parameter settings:

Step 1: call m⁶A peaks for biological replicate 1

The screenshot shows the Galaxy interface with the 'Peak Calling from the MeRIP-Seq data (Galaxy Version 17.09)' tool selected. The workflow is as follows:

- Step 1: input IP and input alignment respectively for biological replicate 1
- Step 2: select macs2 to call peaks
- Step 3: set effective genome size
- Step 4: set significant threshold
- Step 5: click here to run this function

Annotations in red highlight the 'Input the BAM file in IP sample' dropdown, the 'Peak calling methods' section (with 'MACS2' selected), the 'Effective genome size' input field (set to 417272442), the 'Minimum FDR (q-value) cutoff for peak detection' input field (set to 0.01), and the 'Execute' button.

This step will generate ~16,400 peaks for biological replicate 1

Step 2: call m⁶A peaks for biological replicate 2

The screenshot shows the Galaxy interface with the 'Peak Calling from the MeRIP-Seq data (Galaxy Version 17.09)' tool selected. The workflow is as follows:

- Step 1: input IP and input alignment respectively for biological replicate 2
- Step 2: select macs2 to call peaks
- Step 3: set effective genome size
- Step 4: set significant threshold
- Step 5: click here to run this function

Annotations in red highlight the 'Input the BAM file in IP sample' dropdown, the 'Peak calling methods' section (with 'MACS2' selected), the 'Effective genome size' input field (set to 417272442), the 'Minimum FDR (q-value) cutoff for peak detection' input field (set to 0.01), and the 'Execute' button.

This step will generate ~16,100 peaks for biological replicate 2

Step 3: obtain consistent peaks between two biological replicates

After finishing peak calling for two replicates, deepEA also provided a function **Merge two biological replicates** to obtain consistent peaks between two biological replicates, see the following screenshots for details:

Then the consistent peaks named as **intersect.bed** (about ~14,000 consistent peaks) will be shown in you History Panel.

7. Perform functional annotation for m⁶A

m⁶A distribution

The distribution of m⁶A in the genome and transcriptome can be visualized by the function **RNA Modification Distribution** in **Functional Annotation** module, see the folloioing screenshot for detail:

Then an interactively HTML will be generated, please see xxxxxxxxxxxxxxxxxxxxxxx to preview this results

Link m⁶A modifications with gene

Step 1: consistent m⁶A peaks between two biological replicates

Step 2: input genome annotation

Step 3: click here to run this function

This output for this function are shown in the following the screenshot

The screenshot shows the Galaxy web interface with the following details:

- Top Bar:** Galaxy, Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, User.
- Left Sidebar:** Tools, search tools, PRE-ANALYSIS, Data Preparation, Quality Control, CORE ANALYSIS, Identification of RNA Modifications, Functional Annotation, RNA Modification Distribution, Motif Analysis, Link RNA Modifications to Genes, Functional Enrichment Analysis, ADVANCED ANALYSIS, Multi-miRNA Integrative Analysis, Prediction Analysis Based on Machine Learning, USEFUL TOOLS, Merge biological replicates, Convert Formats, Filter and Sort, Get Data, Workflows, All workflows, miRNA-Seq.
- Central Content:** A table titled "Analyze Data" showing RNA modification distribution. The columns are Chrom, Start, End, Name, Score, Strand, ThickStart, ThickEnd, ItemRcsB, BlockCount, BlockSizes, and BlockStarts. The rows show data for various genomic features across different chromosomes.
- Right Sidebar:** History, search datasets, miRNA-Seq, 43 samples, 16.49 Gb, 40 items. A red arrow points from the "miRNA-Seq" item to the "38. RNA modifications genes.txt" file in the history list.

De-novo motif discovery

This following screenshot shows how to use homer to perform *de-novo* motif discovery

Step 1: input peak regions with strand information

Step 2: input reference genome sequences

Step 3: select HOMER

Step 4: set the motif length

Step 5: click here to run this function

Then an HTML document will be generated, please click xxxxxxxxxxxxxxxxxx to preview this results

GO functional enrichment analysis

deepEA provided the function **Functional Enrichment Analysis** to perform GO enrichment analysis, see the following screenshot for detail:

Step 1: input latin species name

Step 2: select RNA modifications gene list (generated by function Link RNA Modifications to Genes)

Step 3: click here to run this function

7. Multi-omics integrative analysis

To run this module, you have to download test data provided by deepEA, and then upload the data in directory `test_data/Multi-omics Integrative Analysis/` to deepEA server. If you are not sure how to upload local data into deepEA server, please see [here](#) for details. Then you can run the function **Integrative Analysis of Three Omics Data Sets** in **Multi-omics Integrative Analysis** module as the following screenshot shows:

Step 1: input the quantification matrix

Step 2: input duplicated gene pairs

Step 3: click here to run this function

Then an interactively HTML document will be output, click here xxxxxxxx to preview this results.

8. Build an m⁶A predictor based on machine learning

Step 1: generate positive and negative samples for m⁶A predictor construction

Step 1: input peak regions in BED6 format

Step 2: input reference genome sequences

Step 3: input genome annotation

Step 4: click here to run this function

Step 2: encoding positive samples

Step 1: select positive samples from History Panel

Step 2: select Zea mays reference genome sequence

Step 3: select feature encoding methods

Step 4: input Zea mays genome annotation

Step 5: select nucleic acid composition related features

Step 6: select autocorrelation-based features (GAC is not included as it runs particularly slow)

Step 7: select all features here

Step 8: click here to run this function

Step 3: encoding negative samples

Step 1: select negative samples from History Panel

Step 2: select Zea mays reference genome sequence

Step 3: select feature encoding methods

Step 4: input Zea mays genome annotation

Step 5: select nucleic acid composition related features

Step 6: select autocorrelation-based features (GAC is not included as it runs particularly slow)

Step 7: select all features here

Step 8: click here to run this function

Step 4: m⁶A predictor construction and evaluation

Step 1: select feature matrix of positive samples

Step 2: select feature matrix of negative samples

Step 3: set the number of threads as 5

Step 4: click here to run this function