

tutorial

October 17, 2024

1 Hicberg tutorial

This notebook aims to provide a simple tutorial on how to use the Hicberg package to analyze Hi-C data.

The package is designed to be user-friendly and easy to use, and it provides a wide range of tools for the analysis of Hi-C data.

The tutorial will cover all steps of the analysis, from data loading to visualization of results.

1.1 Importing libraries

```
[50]: # Import from standard libraries
import time
import glob, sys
from glob import glob
from shutil import which, rmtree
from os.path import join
from pathlib import Path
import subprocess as sp

## Multitprocessing puroposes
import multiprocessing
from multiprocessing import Process
from functools import partial

import numpy as np
import cooler
import matplotlib.pyplot as plt
from mpl_toolkits.axes_grid1 import make_axes_locatable

## Import from hicberg
import hicberg.align as hal
import hicberg.io as hio
import hicberg.utils as hut
import hicberg.plot as hpl
import hicberg.statistics as hst
```

1.2 Paths and parameters setting

Parameters are set to perform standard analysis of Hi-C data.

No blacklisted regions are used in this tutorial.

```
[2]: UNRESCUED_MATRIX = "unrescued_map.cool"
    RESTRICTION_MAP = "restriction_map.npy"

    # Directories
    output_directory = "./tmp"
    analysis_name = "my_library"

    # Data
    # genome = "data/V_cholerae_01.fa"
    # forward_reads = "data/SRR10394900_CC340_V.
    #                 ↪ cholera_HiC_using_HpaII_-_biological_replicate_1_1.fastq.gz"
    # reverse_reads = "data/SRR10394900_CC340_V.
    #                 ↪ cholera_HiC_using_HpaII_-_biological_replicate_1_1.fastq.gz"

    genome = "data/SC288_with_micron.fa"
    forward_reads = "/home/sardine/Bureau/tutorial/data/AC1.end1.fastq.gz"
    reverse_reads = "/home/sardine/Bureau/tutorial/data/AC1.end2.fastq.gz"

    # Experiment dependant parameters
    restriction_enzyme = ["DpnII", "HinfI"]
    circular = [""]
    rate = 1.0
    resolution = 2000
    mode = "standard"
    blacklist = None

    cpus = 16
```

1.3 Step 0: Creating analyses directories

```
[3]: output_folder = hio.create_folder(sample_name = analysis_name, output_dir =
    ↪ output_directory)
```

```
2024-10-17 -- 08:45:39 :: INFO :: Creating folder my_library in ./tmp
2024-10-17 -- 08:45:39 :: INFO :: Folder my_library in tmp/my_library created.
```

1.4 Step 1: Get chromosomes size and align reads

This step aims to define the size of the chromosomes (and relative binning), and align the reads to the reference genome.

Reads are going to be aligned in “very sensitive” mode for a better accuracy.

Mapping quality is set to 30 (default) and all reads are going to be kept (max_alignment = -1).

```
[4]: hut.get_chromosomes_sizes(genome = genome, output_dir = output_folder)
hut.get_bin_table(bins = resolution, output_dir = output_folder)
```

```
2024-10-17 -- 08:45:39 :: INFO :: Start getting chromosome sizes
2024-10-17 -- 08:45:39 :: INFO :: Chromosome sizes have been saved in
tmp/my_library/chromosome_sizes.npy
2024-10-17 -- 08:45:39 :: INFO :: Start getting bin table
```

```
[5]: # Index the genome
index = "SC288_with_micron"

hal.hic_build_index(genome = genome, output_dir = output_folder, cpus = cpus)

# Align reads
hal.hic_align(index = index, fq_for = forward_reads, fq_rev = reverse_reads,
↳sensitivity = "very-sensitive", max_alignment = -1,
        output_dir = output_folder, cpus = cpus)

# Convert to .bam
hal.hic_view(cpus = cpus, output_dir = output_folder)

# Sort reads by name
hal.hic_sort(cpus = cpus, output_dir = output_folder)
```

```
2024-10-17 -- 08:45:39 :: INFO :: Start building index for alignment
2024-10-17 -- 08:45:45 :: INFO :: Index built at
tmp/my_library/SC288_with_micron
2024-10-17 -- 08:45:45 :: INFO :: Start aligning reads
2024-10-17 -- 08:52:58 :: INFO :: 47310032 reads; of these:
    47310032 (100.00%) were unpaired; of these:
        2845860 (6.02%) aligned 0 times
        39601500 (83.71%) aligned exactly 1 time
        4862672 (10.28%) aligned >1 times
93.98% overall alignment rate
```

```
2024-10-17 -- 08:52:58 :: INFO :: 47310032 reads; of these:
    47310032 (100.00%) were unpaired; of these:
        2293189 (4.85%) aligned 0 times
        40099611 (84.76%) aligned exactly 1 time
        4917232 (10.39%) aligned >1 times
95.15% overall alignment rate
```

```
2024-10-17 -- 08:52:58 :: INFO :: Alignment saved at tmp/my_library
2024-10-17 -- 08:52:58 :: INFO :: Start converting .sam to .bam
2024-10-17 -- 08:53:45 :: INFO :: Compressed alignment done at tmp/my_library
2024-10-17 -- 08:53:45 :: INFO :: Start sorting .bam alignment files
[bam_sort_core] merging from 1 files and 16 in-memory blocks...
[bam_sort_core] merging from 1 files and 16 in-memory blocks...
```

```
2024-10-17 -- 08:56:17 :: INFO :: Sorted alignment done at tmp/my_library
```

1.5 Step 3: Classification of read pairs

All reads read pairs previously produced are going to be classified in the following categories:

- Unmapped pairs : read pairs that are not aligned to the reference genome (at least one read is unmapped)
- Unambiguous pairs : read pairs that are aligned to the reference genome and are uniquely mapped (both reads are uniquely mapped)
- Ambiguous pairs : read pairs that are aligned to the reference genome and are not uniquely mapped (at least one read is not uniquely mapped) or too poorly mapped (mapping quality < 35)

```
[6]: hut.classify_reads( output_dir = output_folder)
```

```
2024-10-17 -- 09:14:57 :: INFO :: Files for the different groups have been saved
in tmp/my_library
2024-10-17 -- 09:14:57 :: INFO :: Number of unmapped reads in forward file :
7580990
2024-10-17 -- 09:14:57 :: INFO :: Number of unmapped reads in reverse file :
8224108
2024-10-17 -- 09:14:57 :: INFO :: Number of uniquely mapped reads in forward
file : 35472762
2024-10-17 -- 09:14:57 :: INFO :: Number of uniquely mapped reads in reverse
file : 35472762
2024-10-17 -- 09:14:57 :: INFO :: Number of multi mapped reads in forward file :
59443286
2024-10-17 -- 09:14:57 :: INFO :: Number of multi mapped reads in reverse file :
56970596
```

1.6 Step 4: Build contact matrix

Hi-C contact matrix is going to be built using the read pairs classified in the previous step.

The produced contact matrix (in .cool) is going to be normalized using the ICE method.

```
[7]: hio.build_pairs(output_dir = output_folder)
hio.build_matrix(cpus = cpus, balance = True, output_dir = output_folder)
```

```
2024-10-17 -- 09:14:57 :: INFO :: Start building pairs file for unambiguously
aligned reads
2024-10-17 -- 09:16:27 :: INFO :: Pairs file successfully created in
tmp/my_library
INFO:cooler.create:Writing chunk 0:
/home/sardine/Bureau/tutorial/tmp/my_library/tmpt4gyxh59.multi.cool::0
INFO:cooler.create:Creating cooler at
"/home/sardine/Bureau/tutorial/tmp/my_library/tmpt4gyxh59.multi.cool::/0"
INFO:cooler.create:Writing chroms
INFO:cooler.create:Writing bins
```

```

INFO:cooler.create:Writing pixels
INFO:cooler.create:Writing indexes
INFO:cooler.create:Writing info
INFO:cooler.create:Writing chunk 1:
/home/sardine/Bureau/tutorial/tmp/my_library/tmpt4gyxh59.multi.cool::1
INFO:cooler.create:Creating cooler at
"/home/sardine/Bureau/tutorial/tmp/my_library/tmpt4gyxh59.multi.cool::/1"
INFO:cooler.create:Writing chroms
INFO:cooler.create:Writing bins
INFO:cooler.create:Writing pixels
INFO:cooler.create:Writing indexes
INFO:cooler.create:Writing info
INFO:cooler.create:Writing chunk 2:
/home/sardine/Bureau/tutorial/tmp/my_library/tmpt4gyxh59.multi.cool::2
INFO:cooler.create:Creating cooler at
"/home/sardine/Bureau/tutorial/tmp/my_library/tmpt4gyxh59.multi.cool::/2"
INFO:cooler.create:Writing chroms
INFO:cooler.create:Writing bins
INFO:cooler.create:Writing pixels
INFO:cooler.create:Writing indexes
INFO:cooler.create:Writing info
INFO:cooler.create:Merging into tmp/my_library/unrescued_map.cool
INFO:cooler.create:Creating cooler at "tmp/my_library/unrescued_map.cool::/"
INFO:cooler.create:Writing chroms
INFO:cooler.create:Writing bins
INFO:cooler.create:Writing pixels
INFO:cooler.reduce:nns: [3574146, 3570234, 1704044]
INFO:cooler.reduce:current: [3574146, 3570234, 1704044]
INFO:cooler.create:Writing indexes
INFO:cooler.create:Writing info
INFO:cooler.cli.balance:Balancing "tmp/my_library/unrescued_map.cool"
INFO:cooler.balance:variance is 122743344.90695125
INFO:cooler.balance:variance is 211147.15648302896
INFO:cooler.balance:variance is 68594.15576544323
INFO:cooler.balance:variance is 92976.99436259829
INFO:cooler.balance:variance is 41348.44191362295
INFO:cooler.balance:variance is 66348.47134112733
INFO:cooler.balance:variance is 32021.30267203885
INFO:cooler.balance:variance is 49995.11223916566
INFO:cooler.balance:variance is 25787.54204781249
INFO:cooler.balance:variance is 38216.275169273365
INFO:cooler.balance:variance is 20922.845733074057
INFO:cooler.balance:variance is 29408.517360463928
INFO:cooler.balance:variance is 16982.96122915793
INFO:cooler.balance:variance is 22727.146052963693
INFO:cooler.balance:variance is 13763.797008202846
INFO:cooler.balance:variance is 17619.696866678343
INFO:cooler.balance:variance is 11132.097258583519

```

INFO:cooler.balance:variance is 13695.548269823908
INFO:cooler.balance:variance is 8985.018719289526
INFO:cooler.balance:variance is 10668.970371191004
INFO:cooler.balance:variance is 7238.056394751527
INFO:cooler.balance:variance is 8327.312946987715
INFO:cooler.balance:variance is 5820.530165331506
INFO:cooler.balance:variance is 6510.692428708155
INFO:cooler.balance:variance is 4673.244711801411
INFO:cooler.balance:variance is 5098.060995896909
INFO:cooler.balance:variance is 3746.8197235468933
INFO:cooler.balance:variance is 3997.2843878046915
INFO:cooler.balance:variance is 3000.2790519712476
INFO:cooler.balance:variance is 3137.925648675021
INFO:cooler.balance:variance is 2399.7975669478014
INFO:cooler.balance:variance is 2465.9295239550943
INFO:cooler.balance:variance is 1917.5860657903845
INFO:cooler.balance:variance is 1939.672195019119
INFO:cooler.balance:variance is 1530.9106486638314
INFO:cooler.balance:variance is 1527.0054840908
INFO:cooler.balance:variance is 1221.242165208706
INFO:cooler.balance:variance is 1203.0324213274723
INFO:cooler.balance:variance is 973.527724955726
INFO:cooler.balance:variance is 948.4247409245171
INFO:cooler.balance:variance is 775.5734609617684
INFO:cooler.balance:variance is 748.144501437939
INFO:cooler.balance:variance is 617.5262994694666
INFO:cooler.balance:variance is 590.468767654965
INFO:cooler.balance:variance is 491.44218177019644
INFO:cooler.balance:variance is 466.24272491485
INFO:cooler.balance:variance is 390.9286455193502
INFO:cooler.balance:variance is 368.3057772204631
INFO:cooler.balance:variance is 310.85059121514956
INFO:cooler.balance:variance is 291.0491999895363
INFO:cooler.balance:variance is 247.08920806618917
INFO:cooler.balance:variance is 230.0742329027448
INFO:cooler.balance:variance is 196.34525826071837
INFO:cooler.balance:variance is 181.9271329755843
INFO:cooler.balance:variance is 155.97912219019793
INFO:cooler.balance:variance is 143.89339078176508
INFO:cooler.balance:variance is 123.88113218122673
INFO:cooler.balance:variance is 113.83756490030291
INFO:cooler.balance:variance is 98.36673901527269
INFO:cooler.balance:variance is 90.0783869143378
INFO:cooler.balance:variance is 78.09195237362141
INFO:cooler.balance:variance is 71.29120435071235
INFO:cooler.balance:variance is 61.9852731714646
INFO:cooler.balance:variance is 56.43166093988196
INFO:cooler.balance:variance is 49.19299913879837

INFO:cooler.balance:variance is 44.67590875835076
INFO:cooler.balance:variance is 39.03534517457282
INFO:cooler.balance:variance is 35.37371322711671
INFO:cooler.balance:variance is 30.971288695041842
INFO:cooler.balance:variance is 28.011629807508005
INFO:cooler.balance:variance is 24.570439407976828
INFO:cooler.balance:variance is 22.18406043802973
INFO:cooler.balance:variance is 19.490554123977855
INFO:cooler.balance:variance is 17.570483196971054
INFO:cooler.balance:variance is 15.45958085596085
INFO:cooler.balance:variance is 13.917524182407142
INFO:cooler.balance:variance is 12.261331895623426
INFO:cooler.balance:variance is 11.02483176472613
INFO:cooler.balance:variance is 9.724060929286216
INFO:cooler.balance:variance is 8.733939623889302
INFO:cooler.balance:variance is 7.711361545090666
INFO:cooler.balance:variance is 6.9194811521646855
INFO:cooler.balance:variance is 6.114919601850595
INFO:cooler.balance:variance is 5.48225521886585
INFO:cooler.balance:variance is 4.848744835685294
INFO:cooler.balance:variance is 4.343750656405688
INFO:cooler.balance:variance is 3.844581898966718
INFO:cooler.balance:variance is 3.441820833329779
INFO:cooler.balance:variance is 3.0482611656297123
INFO:cooler.balance:variance is 2.7272654988842846
INFO:cooler.balance:variance is 2.416797898268935
INFO:cooler.balance:variance is 2.1611287147615665
INFO:cooler.balance:variance is 1.9160870461341364
INFO:cooler.balance:variance is 1.7125622352847905
INFO:cooler.balance:variance is 1.5190718947685304
INFO:cooler.balance:variance is 1.3571355685882727
INFO:cooler.balance:variance is 1.2042895320315141
INFO:cooler.balance:variance is 1.0754990261572093
INFO:cooler.balance:variance is 0.9547158572819051
INFO:cooler.balance:variance is 0.8523258112705829
INFO:cooler.balance:variance is 0.7568486300503077
INFO:cooler.balance:variance is 0.6754747531167891
INFO:cooler.balance:variance is 0.5999796286025592
INFO:cooler.balance:variance is 0.535327558318567
INFO:cooler.balance:variance is 0.47561700820004166
INFO:cooler.balance:variance is 0.42426412715909645
INFO:cooler.balance:variance is 0.3770269278484935
INFO:cooler.balance:variance is 0.33624711849279565
INFO:cooler.balance:variance is 0.29886986509809377
INFO:cooler.balance:variance is 0.26649297755036794
INFO:cooler.balance:variance is 0.23691209090665216
INFO:cooler.balance:variance is 0.21121140447244133
INFO:cooler.balance:variance is 0.18779679446657518

INFO:cooler.balance:variance is 0.16739900584194456
INFO:cooler.balance:variance is 0.14886254189368928
INFO:cooler.balance:variance is 0.13267584728659915
INFO:cooler.balance:variance is 0.11799929090955748
INFO:cooler.balance:variance is 0.1051559782852972
INFO:cooler.balance:variance is 0.09353420015854391
INFO:cooler.balance:variance is 0.08334486143483974
INFO:cooler.balance:variance is 0.07414107342681185
INFO:cooler.balance:variance is 0.06605811081122955
INFO:cooler.balance:variance is 0.05876855083533943
INFO:cooler.balance:variance is 0.05235710912400359
INFO:cooler.balance:variance is 0.046583168782848516
INFO:cooler.balance:variance is 0.04149799543202271
INFO:cooler.balance:variance is 0.03692421188611946
INFO:cooler.balance:variance is 0.03289124506038918
INFO:cooler.balance:variance is 0.029267916247719773
INFO:cooler.balance:variance is 0.026069641181500087
INFO:cooler.balance:variance is 0.023199088134207797
INFO:cooler.balance:variance is 0.02066289503006984
INFO:cooler.balance:variance is 0.01838860263152753
INFO:cooler.balance:variance is 0.01637753400943464
INFO:cooler.balance:variance is 0.014575564567125322
INFO:cooler.balance:variance is 0.012980963861271249
INFO:cooler.balance:variance is 0.011553166025305244
INFO:cooler.balance:variance is 0.01028883831485582
INFO:cooler.balance:variance is 0.009157474695838795
INFO:cooler.balance:variance is 0.008155049629996927
INFO:cooler.balance:variance is 0.007258545853146243
INFO:cooler.balance:variance is 0.006463796034078234
INFO:cooler.balance:variance is 0.005753376512272479
INFO:cooler.balance:variance is 0.0051232950096208405
INFO:cooler.balance:variance is 0.00456032002787273
INFO:cooler.balance:variance is 0.004060800878834406
INFO:cooler.balance:variance is 0.0036146584803428768
INFO:cooler.balance:variance is 0.0032186560334891675
INFO:cooler.balance:variance is 0.0028650928962186796
INFO:cooler.balance:variance is 0.0025511613354756266
INFO:cooler.balance:variance is 0.002270961067938841
INFO:cooler.balance:variance is 0.0020220957261470876
INFO:cooler.balance:variance is 0.0018000321583674164
INFO:cooler.balance:variance is 0.001602750353836533
INFO:cooler.balance:variance is 0.001426758529531618
INFO:cooler.balance:variance is 0.0012703704781203639
INFO:cooler.balance:variance is 0.0011308900140818072
INFO:cooler.balance:variance is 0.00100692055775263
INFO:cooler.balance:variance is 0.0008963754928577477
INFO:cooler.balance:variance is 0.0007981054805005255
INFO:cooler.balance:variance is 0.0007104922160384205


```

INFO:cooler.balance:variance is 0.0006325948022053995
INFO:cooler.balance:variance is 0.0005631556519223578
INFO:cooler.balance:variance is 0.0005014078853351009
INFO:cooler.balance:variance is 0.00044637243514331656
INFO:cooler.balance:variance is 0.0003974265628632592
INFO:cooler.balance:variance is 0.00035380674522099735
INFO:cooler.balance:variance is 0.00031500887654480004
INFO:cooler.balance:variance is 0.00028043659713539946
INFO:cooler.balance:variance is 0.00024968292542445645
INFO:cooler.balance:variance is 0.00022228140145718852
INFO:cooler.balance:variance is 0.0001979042091125484
INFO:cooler.balance:variance is 0.00017618601667862186
INFO:cooler.balance:variance is 0.00015686329662998035
INFO:cooler.balance:variance is 0.00013964958014973922
INFO:cooler.balance:variance is 0.00012433338323829978
INFO:cooler.balance:variance is 0.00011068983272324492
INFO:cooler.balance:variance is 9.85494625051873e-05
INFO:cooler.balance:variance is 8.773557733556546e-05
INFO:cooler.balance:variance is 7.811255672953136e-05
INFO:cooler.balance:variance is 6.954143778559177e-05
INFO:cooler.balance:variance is 6.191380878026993e-05
INFO:cooler.balance:variance is 5.512029355564903e-05
INFO:cooler.balance:variance is 4.907431621669184e-05
INFO:cooler.balance:variance is 4.368972537519327e-05
INFO:cooler.balance:variance is 3.8897442922560096e-05
INFO:cooler.balance:variance is 3.462956622141431e-05
INFO:cooler.balance:variance is 3.0831020535413244e-05
INFO:cooler.balance:variance is 2.7448254891032806e-05
INFO:cooler.balance:variance is 2.443738866308234e-05
INFO:cooler.balance:variance is 2.1756166806235192e-05
INFO:cooler.balance:variance is 1.936964824392943e-05
INFO:cooler.balance:variance is 1.7244475200948022e-05
INFO:cooler.balance:variance is 1.5352839479180623e-05
INFO:cooler.balance:variance is 1.3668395879774661e-05
INFO:cooler.balance:variance is 1.2169023165610765e-05
INFO:cooler.balance:variance is 1.0833906523001572e-05
INFO:cooler.balance:variance is 9.64545582198916e-06
2024-10-17 -- 09:18:46 :: INFO :: Cooler matrix successfully created in
tmp/my_library

```

1.6.1 Visualizing the contact matrix

```

[43]: matrix = cooler.Cooler("./tmp/my_library/unrescued_map.cool")
      chromosome_I = matrix.matrix(balance = True).fetch("chr1")

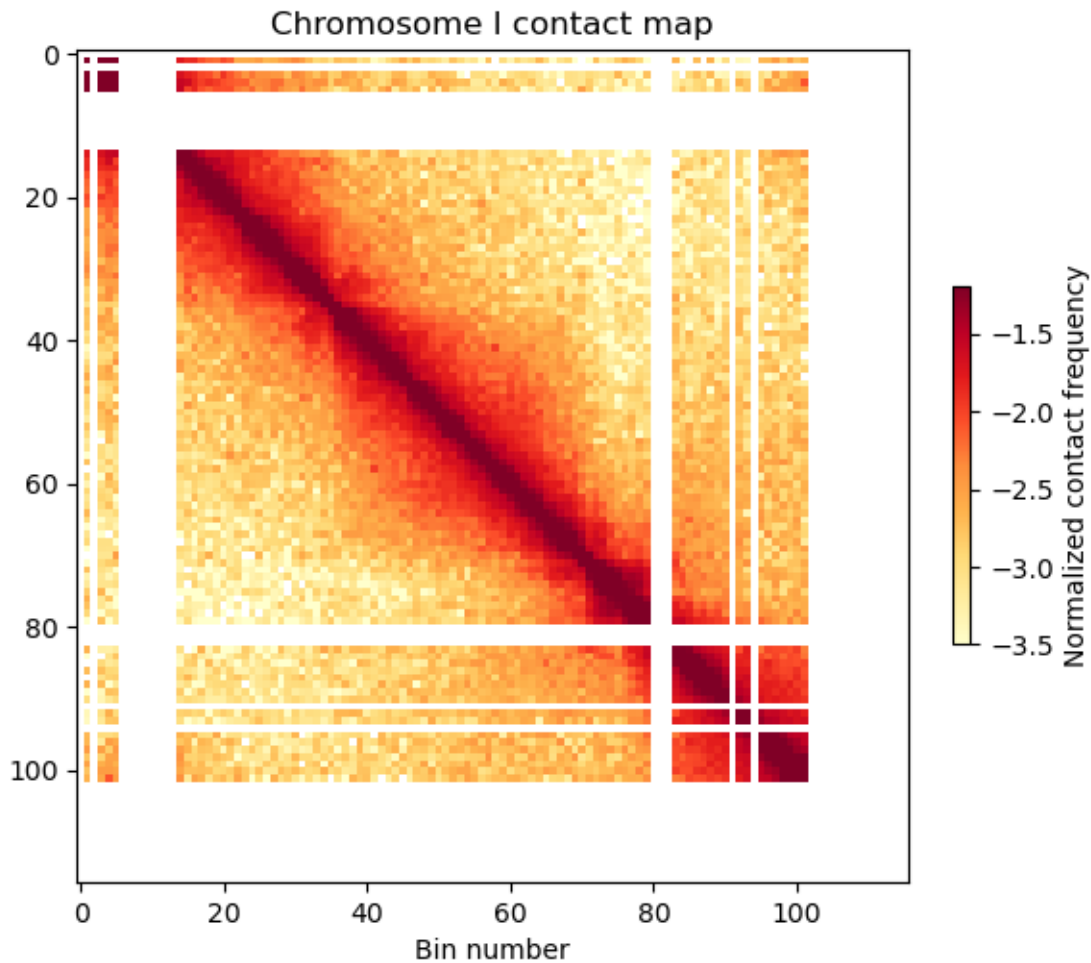
      plt.figure(figsize=(6,6))
      plt.imshow(np.log10(chromosome_I), cmap = "YlOrRd", vmin = -3.5, vmax = -1.2)

```

```
plt.colorbar(fraction = 0.02, label = "Normalized contact frequency")
plt.title("Chromosome I contact map")
plt.xlabel("Bin number")
plt.show()
```

/tmp/ipykernel_3133923/3058769468.py:5: RuntimeWarning: divide by zero encountered in log10

```
plt.imshow(np.log10(chromosome_I), cmap = "YlOrRd", vmin = -3.5, vmax = -1.2)
```



1.7 Step 5: Profiling Hi-C data

Statistical profiling of the Hi-C data is going to be performed.

The following metrics are going to be calculated (standard mode): - $P(s)$: the probability of contact between two loci separated by a genomic distance s - Coverage : the read coverage of the Hi-C data per bin

```
[9]: # Get the restriction map of the genome
restriction_map = hst.get_restriction_map(genome = genome, enzyme =
    ↪restriction_enzyme, output_dir = output_folder)

# Compute the distance between fragments and bin the genome
hst.get_dist_frags(genome = genome, restriction_map = restriction_map, circular
    ↪= circular, rate = rate, output_dir = output_folder)
hst.log_bin_genome(genome = genome, output_dir = output_folder)

# Define the profiles to be established
p1 = Process(target = hst.get_patterns, kwargs = dict(circular = circular,
    ↪output_dir = output_folder))
p2 = Process(target = hst.generate_trans_ps, kwargs = dict(output_dir =
    ↪output_folder))
p3 = Process(target = hst.generate_coverages, kwargs = dict(genome = genome,
    ↪bins = resolution, output_dir = output_folder))

# Start the processes
for process in [p1, p2, p3]:
    process.start()

for process in [p1, p2, p3]:
    process.join()
```

```
2024-10-17 -- 09:18:46 :: INFO :: Generating restriction map...
2024-10-17 -- 09:18:47 :: INFO :: Saved restriction map at : tmp/my_library
2024-10-17 -- 09:18:47 :: INFO :: Start generating distribution of fragments'
distance...
2024-10-17 -- 09:27:13 :: INFO :: Saved restriction map at :
tmp/my_library/dist.frag.npy
2024-10-17 -- 09:27:13 :: INFO :: Start log binning of genome...
2024-10-17 -- 09:27:13 :: INFO :: Log binning of genome
data/SC288_with_micron.fa saved in tmp/my_library/xs.npy.
2024-10-17 -- 09:27:13 :: INFO :: Start generating patterns distribution...
2024-10-17 -- 09:27:13 :: INFO :: Start getting trans-P(s)
2024-10-17 -- 09:27:13 :: INFO :: Start generating coverages...
2024-10-17 -- 09:27:22 :: INFO :: Trans P(s) saved in tmp/my_library
2024-10-17 -- 09:30:50 :: INFO :: Coverage dictionary saved in tmp/my_library
2024-10-17 -- 09:33:10 :: INFO :: Saved weirds.npy, uncuts.npy and loops.npy in
tmp/my_library
```

1.8 Step 6: Infer ambiguous read pairs from statistical profile

Previously profiled Hi-C data is going to be used to infer ambiguous read pairs.

```
[10]: # Reload restriction map
restriction_map = hio.load_dictionary(Path(output_folder) / RESTRICTION_MAP)
```

```

# Split the alignment files
hut.chunk_bam(nb_chunks = 32, output_dir = output_folder)

# Get chunks as lists
forward_chunks, reverse_chunks = hut.get_chunks(output_dir = output_folder)

# Reattribute reads
with multiprocessing.Pool(processes = cpus) as pool: # cpus

    results = pool.map(partial(hst.reattribute_reads, mode = mode,
↪restriction_map = restriction_map, output_dir = output_folder),
    zip(forward_chunks, reverse_chunks))
    pool.close()
    pool.join()

hio.merge_predictions(output_dir = output_folder, clean = True, cpus = cpus)

# Temporary files cleaning
folder_to_delete = Path(output_folder) / 'chunks'
rmtree(folder_to_delete)

```

```

2024-10-17 -- 09:33:10 :: INFO :: Start chunking BAM files
2024-10-17 -- 09:41:16 :: INFO :: Chunks saved in tmp/my_library/chunks
2024-10-17 -- 09:44:14 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:23 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:32 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:36 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:37 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:37 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:38 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:38 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:39 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:39 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:39 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:41 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:42 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:48 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:50 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:44:53 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:45:00 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:30 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:36 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:36 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:41 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:41 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:43 :: INFO :: Predictions written in tmp/my_library

```

```

2024-10-17 -- 09:47:43 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:44 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:45 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:47 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:47 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:49 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:52 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:56 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:59 :: INFO :: Predictions written in tmp/my_library
2024-10-17 -- 09:47:59 :: INFO :: Start merging predictions

```

1.9 Step 7: Build contact matrix with inferred read pairs

```

[11]: # Build pairs with both ambiguous and inferred reads
hio.build_pairs(mode = True, output_dir = output_folder)
# Build reconstructed matrix
hio.build_matrix(cpus = cpus, balance = True, mode = True, output_dir =
↳ output_folder)

```

```

2024-10-17 -- 09:48:06 :: INFO :: Start building pairs file for ambiguously
aligned reads
2024-10-17 -- 09:49:55 :: INFO :: Pairs file successfully created in
tmp/my_library
INFO:cooler.create:Writing chunk 0:
/home/sardine/Bureau/tutorial/tmp/my_library/tmp0fen71zc.multi.cool::0
INFO:cooler.create:Creating cooler at
"/home/sardine/Bureau/tutorial/tmp/my_library/tmp0fen71zc.multi.cool::/0"
INFO:cooler.create:Writing chroms
INFO:cooler.create:Writing bins
INFO:cooler.create:Writing pixels
INFO:cooler.create:Writing indexes
INFO:cooler.create:Writing info
INFO:cooler.create:Writing chunk 1:
/home/sardine/Bureau/tutorial/tmp/my_library/tmp0fen71zc.multi.cool::1
INFO:cooler.create:Creating cooler at
"/home/sardine/Bureau/tutorial/tmp/my_library/tmp0fen71zc.multi.cool::/1"
INFO:cooler.create:Writing chroms
INFO:cooler.create:Writing bins
INFO:cooler.create:Writing pixels
INFO:cooler.create:Writing indexes
INFO:cooler.create:Writing info
INFO:cooler.create:Writing chunk 2:
/home/sardine/Bureau/tutorial/tmp/my_library/tmp0fen71zc.multi.cool::2
INFO:cooler.create:Creating cooler at
"/home/sardine/Bureau/tutorial/tmp/my_library/tmp0fen71zc.multi.cool::/2"
INFO:cooler.create:Writing chroms
INFO:cooler.create:Writing bins
INFO:cooler.create:Writing pixels

```

INFO:cooler.create:Writing indexes
INFO:cooler.create:Writing info
INFO:cooler.create:Merging into tmp/my_library/rescued_map.cool
INFO:cooler.create:Creating cooler at "tmp/my_library/rescued_map.cool:/"
INFO:cooler.create:Writing chroms
INFO:cooler.create:Writing bins
INFO:cooler.create:Writing pixels
INFO:cooler.reduce:nnzs: [3574146, 3570234, 2935339]
INFO:cooler.reduce:current: [3574146, 3570234, 2935339]
INFO:cooler.create:Writing indexes
INFO:cooler.create:Writing info
INFO:cooler.cli.balance:Balancing "tmp/my_library/rescued_map.cool"
INFO:cooler.balance:variance is 222461452.62033328
INFO:cooler.balance:variance is 265627.0744999617
INFO:cooler.balance:variance is 203037.49208577862
INFO:cooler.balance:variance is 68317.08886766927
INFO:cooler.balance:variance is 34611.153812396326
INFO:cooler.balance:variance is 36346.949268297096
INFO:cooler.balance:variance is 20394.799365397907
INFO:cooler.balance:variance is 25664.679316998878
INFO:cooler.balance:variance is 15513.78533982416
INFO:cooler.balance:variance is 19231.42455331384
INFO:cooler.balance:variance is 12207.20489575824
INFO:cooler.balance:variance is 14590.926346117578
INFO:cooler.balance:variance is 9638.025839863622
INFO:cooler.balance:variance is 11118.084032059145
INFO:cooler.balance:variance is 7601.712161122348
INFO:cooler.balance:variance is 8493.722588675077
INFO:cooler.balance:variance is 5985.6618435290675
INFO:cooler.balance:variance is 6501.814767760592
INFO:cooler.balance:variance is 4705.443155774952
INFO:cooler.balance:variance is 4985.443123997591
INFO:cooler.balance:variance is 3693.518781678391
INFO:cooler.balance:variance is 3828.318788600394
INFO:cooler.balance:variance is 2895.356701564661
INFO:cooler.balance:variance is 2943.528515069523
INFO:cooler.balance:variance is 2267.0038448588325
INFO:cooler.balance:variance is 2265.7736544576296
INFO:cooler.balance:variance is 1773.1716866482163
INFO:cooler.balance:variance is 1745.7983638718365
INFO:cooler.balance:variance is 1385.641781918313
INFO:cooler.balance:variance is 1346.3234742539407
INFO:cooler.balance:variance is 1081.931979701887
INFO:cooler.balance:variance is 1039.0520548750083
INFO:cooler.balance:variance is 844.188401592781
INFO:cooler.balance:variance is 802.4500038364816
INFO:cooler.balance:variance is 658.272879559998
INFO:cooler.balance:variance is 620.0924919205228

INFO:cooler.balance:variance is 513.0173874049375
INFO:cooler.balance:variance is 479.4264515620404
INFO:cooler.balance:variance is 399.6191527606135
INFO:cooler.balance:variance is 370.84074565410003
INFO:cooler.balance:variance is 311.15283239990066
INFO:cooler.balance:variance is 286.9650340838589
INFO:cooler.balance:variance is 242.17909891048436
INFO:cooler.balance:variance is 222.13940387824437
INFO:cooler.balance:variance is 188.43195200424995
INFO:cooler.balance:variance is 172.01202733839793
INFO:cooler.balance:variance is 146.56985390457223
INFO:cooler.balance:variance is 133.23316277863609
INFO:cooler.balance:variance is 113.97829847097003
INFO:cooler.balance:variance is 103.2218962700649
INFO:cooler.balance:variance is 88.61361905986729
INFO:cooler.balance:variance is 79.98796979555587
INFO:cooler.balance:variance is 68.87971843947356
INFO:cooler.balance:variance is 61.99543891213742
INFO:cooler.balance:variance is 53.53098450005511
INFO:cooler.balance:variance is 48.05817049069423
INFO:cooler.balance:variance is 41.59596796475303
INFO:cooler.balance:variance is 37.259629427814474
INFO:cooler.balance:variance is 32.317476879878576
INFO:cooler.balance:variance is 28.89123007413104
INFO:cooler.balance:variance is 25.105621505863
INFO:cooler.balance:variance is 22.40490183159967
INFO:cooler.balance:variance is 19.501054315609725
INFO:cooler.balance:variance is 17.376554765074047
INFO:cooler.balance:variance is 15.146221802469277
INFO:cooler.balance:variance is 13.477915332515327
INFO:cooler.balance:variance is 11.762902656456946
INFO:cooler.balance:variance is 10.454797418448969
INFO:cooler.balance:variance is 9.134671840197043
INFO:cooler.balance:variance is 8.11032650024919
INFO:cooler.balance:variance is 7.093220000835592
INFO:cooler.balance:variance is 6.291979840848916
INFO:cooler.balance:variance is 5.50768723515839
INFO:cooler.balance:variance is 4.881569176500869
INFO:cooler.balance:variance is 4.276351562903511
INFO:cooler.balance:variance is 3.7874940571915348
INFO:cooler.balance:variance is 3.3201553749541253
INFO:cooler.balance:variance is 2.9387485769733237
INFO:cooler.balance:variance is 2.577665533618847
INFO:cooler.balance:variance is 2.2802828741808137
INFO:cooler.balance:variance is 2.0011510731414295
INFO:cooler.balance:variance is 1.7694119577286622
INFO:cooler.balance:variance is 1.553531662903707
INFO:cooler.balance:variance is 1.3730345634422854

INFO:cooler.balance:variance is 1.206004198853658
INFO:cooler.balance:variance is 1.0654787679861017
INFO:cooler.balance:variance is 0.936197289349774
INFO:cooler.balance:variance is 0.826832730971304
INFO:cooler.balance:variance is 0.7267365377723248
INFO:cooler.balance:variance is 0.6416510610481222
INFO:cooler.balance:variance is 0.5641293984671298
INFO:cooler.balance:variance is 0.49795210207674995
INFO:cooler.balance:variance is 0.4378985714261102
INFO:cooler.balance:variance is 0.38644059294050076
INFO:cooler.balance:variance is 0.3399086437137925
INFO:cooler.balance:variance is 0.29990496137602984
INFO:cooler.balance:variance is 0.26384295428023746
INFO:cooler.balance:variance is 0.2327499619525307
INFO:cooler.balance:variance is 0.20479721533873968
INFO:cooler.balance:variance is 0.18063422602779278
INFO:cooler.balance:variance is 0.15896386424211642
INFO:cooler.balance:variance is 0.14018914657064085
INFO:cooler.balance:variance is 0.12338691097800654
INFO:cooler.balance:variance is 0.10880082774295192
INFO:cooler.balance:variance is 0.09577155221062855
INFO:cooler.balance:variance is 0.08444093890449494
INFO:cooler.balance:variance is 0.07433632749583541
INFO:cooler.balance:variance is 0.06553549617329493
INFO:cooler.balance:variance is 0.057698318605541844
INFO:cooler.balance:variance is 0.050863061200990975
INFO:cooler.balance:variance is 0.04478401166359734
INFO:cooler.balance:variance is 0.03947575825734514
INFO:cooler.balance:variance is 0.03476008867871069
INFO:cooler.balance:variance is 0.03063799225916596
INFO:cooler.balance:variance is 0.026979694934820428
INFO:cooler.balance:variance is 0.02377889889414335
INFO:cooler.balance:variance is 0.020940723195859153
INFO:cooler.balance:variance is 0.018455448464227427
INFO:cooler.balance:variance is 0.01625342833163224
INFO:cooler.balance:variance is 0.014323815569577111
INFO:cooler.balance:variance is 0.0126152861866686
INFO:cooler.balance:variance is 0.011117162238936005
INFO:cooler.balance:variance is 0.009791476731893659
INFO:cooler.balance:variance is 0.008628397428428998
INFO:cooler.balance:variance is 0.007599733637628835
INFO:cooler.balance:variance is 0.006696797948948116
INFO:cooler.balance:variance is 0.0058985836189839585
INFO:cooler.balance:variance is 0.005197625770542358
INFO:cooler.balance:variance is 0.004578217280341328
INFO:cooler.balance:variance is 0.004034070504246006
INFO:cooler.balance:variance is 0.0035534027575559765
INFO:cooler.balance:variance is 0.0031309962543518517


```
INFO:cooler.balance:variance is 0.0027579851223802124
INFO:cooler.balance:variance is 0.0024300887161933784
INFO:cooler.balance:variance is 0.0021406167430841447
INFO:cooler.balance:variance is 0.0018860889253746512
INFO:cooler.balance:variance is 0.0016614431650792962
INFO:cooler.balance:variance is 0.0014638703093149638
INFO:cooler.balance:variance is 0.0012895306954268852
INFO:cooler.balance:variance is 0.0011361701973132596
INFO:cooler.balance:variance is 0.0010008697132425617
INFO:cooler.balance:variance is 0.0008818292403333155
INFO:cooler.balance:variance is 0.0007768248652724758
INFO:cooler.balance:variance is 0.0006844250114445625
INFO:cooler.balance:variance is 0.0006029321354685735
INFO:cooler.balance:variance is 0.0005312115266200924
INFO:cooler.balance:variance is 0.00046796515708374834
INFO:cooler.balance:variance is 0.0004122961897095803
INFO:cooler.balance:variance is 0.00036321050842921907
INFO:cooler.balance:variance is 0.0003200010032721617
INFO:cooler.balance:variance is 0.0002819052187005473
INFO:cooler.balance:variance is 0.0002483667897369543
INFO:cooler.balance:variance is 0.000218800178140602
INFO:cooler.balance:variance is 0.00019276840476488156
INFO:cooler.balance:variance is 0.00016982127224280855
INFO:cooler.balance:variance is 0.00014961609355915994
INFO:cooler.balance:variance is 0.00013180636669140404
INFO:cooler.balance:variance is 0.00011612370452332751
INFO:cooler.balance:variance is 0.00010230116538439575
INFO:cooler.balance:variance is 9.012879244733173e-05
INFO:cooler.balance:variance is 7.94007637882386e-05
INFO:cooler.balance:variance is 6.995299454005778e-05
INFO:cooler.balance:variance is 6.162666936868221e-05
INFO:cooler.balance:variance is 5.429366335540948e-05
INFO:cooler.balance:variance is 4.7831350255492554e-05
INFO:cooler.balance:variance is 4.213975916168431e-05
INFO:cooler.balance:variance is 3.7124150210507775e-05
INFO:cooler.balance:variance is 3.2706570852012235e-05
INFO:cooler.balance:variance is 2.8813787252046603e-05
INFO:cooler.balance:variance is 2.538505032624985e-05
INFO:cooler.balance:variance is 2.2363723783837456e-05
INFO:cooler.balance:variance is 1.9702488898759993e-05
INFO:cooler.balance:variance is 1.7357527033179515e-05
INFO:cooler.balance:variance is 1.529199668550602e-05
INFO:cooler.balance:variance is 1.3471983508729229e-05
INFO:cooler.balance:variance is 1.1868814304950752e-05
INFO:cooler.balance:variance is 1.0456231763266501e-05
INFO:cooler.balance:variance is 9.211927404493302e-06
2024-10-17 -- 09:52:28 :: INFO :: Cooler matrix successfully created in
tmp/my_library
```

1.10 Step 8: Plot the reconstructed matrix and profiles

```
[12]: p1 = Process(target = hpl.plot_laws, kwargs = dict(output_dir = output_folder))
      p2 = Process(target = hpl.plot_trans_ps, kwargs = dict(output_dir = output_folder))
      p3 = Process(target = hpl.plot_coverages, kwargs = dict(bins = resolution,
      output_dir = output_folder))

      for process in [p1, p2, p3]:

          process.start()
      # Launch processes
      for process in [p1, p2, p3]:

          process.join()
```

```
2024-10-17 -- 09:52:30 :: INFO :: Saved coverages at : tmp/my_library
2024-10-17 -- 09:52:35 :: INFO :: Saved plots of patterns at : tmp/my_library
2024-10-17 -- 09:52:36 :: INFO :: Saved pseudo P(s) of patterns at :
tmp/my_library
```

```
[40]: # Load reconstructed matrix
      matrix_reconstructed = cooler.Cooler("./tmp/my_library/rescued_map.cool")

      chromosome_I_reconstructed = matrix_reconstructed.matrix(balance = True).
      fetch("chr1")
```

1.10.1 Visualizing the contact matrix

Chromosome I exhibits a clear diagonal pattern, which is expected for Hi-C data.

Hicberg was able to reconstruct the contact matrix with the inferred read pairs, and allow better visualization of (sub)telomere contact behavior.

```
[75]: plt.figure(figsize=(12,6))

      plt.subplots_adjust(wspace=0.3)
      ax1 = plt.subplot(1, 2, 1)
      plt.imshow(np.log10(chromosome_I), cmap="YlOrRd", vmin=-3.5, vmax=-1.2)
      plt.title("Native contact map")
      plt.xlabel("Bin number", fontsize = 12)

      ax2 = plt.subplot(1, 2, 2)
      plt.imshow(np.log10(chromosome_I_reconstructed), cmap="YlOrRd", vmin=-3.5,
      vmax=-1.2)
      plt.title("Reconstructed contact map")
      plt.xlabel("Bin number", fontsize=12)
```

```

divider1 = make_axes_locatable(ax1)
cax1 = divider1.append_axes("right", size="10%", pad=0.1) # Adjust size and
↳pad as needed
plt.colorbar(cax=cax1, label="Normalized contact frequency")

divider2 = make_axes_locatable(ax2)
cax2 = divider2.append_axes("right", size="10%", pad=0.1)
plt.colorbar(cax=cax2, label="Normalized contact frequency")
plt.suptitle("Comparison between the original and the reconstructed contact
↳maps of chromosome I", position=(0.5, 0.9), fontsize = 14)

plt.show()

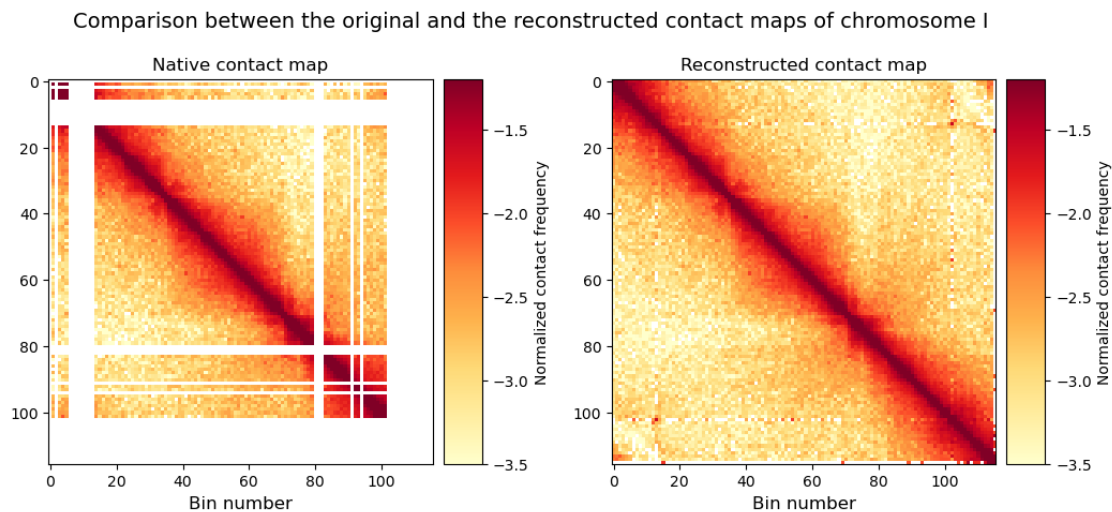
```

/tmp/ipykernel_3133923/382376023.py:5: RuntimeWarning: divide by zero encountered in log10

```
plt.imshow(np.log10(chromosome_I), cmap="YlOrRd", vmin=-3.5, vmax=-1.2)
```

/tmp/ipykernel_3133923/382376023.py:10: RuntimeWarning: divide by zero encountered in log10

```
plt.imshow(np.log10(chromosome_I_reconstructed), cmap="YlOrRd", vmin=-3.5,
vmax=-1.2)
```



1.10.2 Visualizing the profiling results

Here we can see the $P(s)$ and coverage profiles of the Hi-C data before and after the inference of ambiguous read pairs.

For tutorial purposes, only chromosome I will be visualized.

```

[98]: # reload the statistics
xs = hio.load_dictionary("./tmp/my_library/xs.npy")

```

```

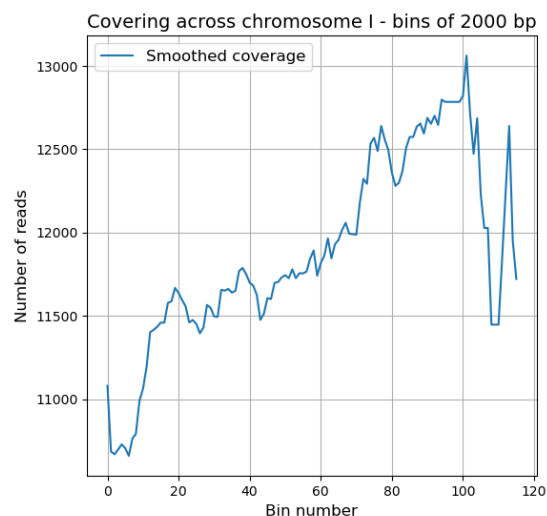
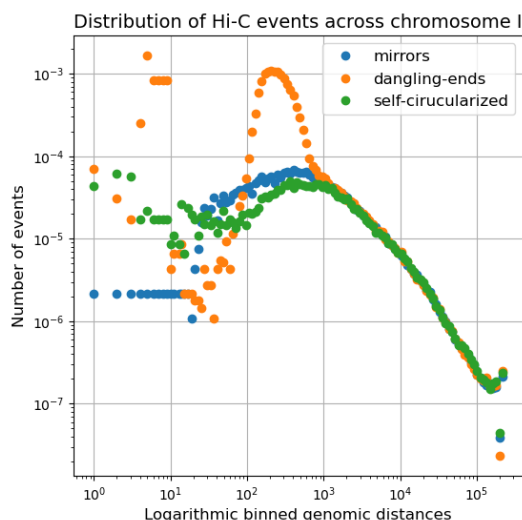
weirds = hio.load_dictionary("./tmp/my_library/weirds.npy")
uncuts = hio.load_dictionary("./tmp/my_library/uncuts.npy")
loops = hio.load_dictionary("./tmp/my_library/loops.npy")

chromosome = "chr1"
plt.figure(figsize=(14, 6))
plt.subplots_adjust(wspace=0.3)
ax1 = plt.subplot(121)

ax1.loglog(xs[chromosome], weirds[chromosome], "o", label="mirrors")
ax1.loglog(xs[chromosome], uncuts[chromosome], "o", label="dangling-ends")
ax1.loglog(xs[chromosome], loops[chromosome], "o", label="self-circularized")
ax1.set_title(f"Distribution of Hi-C events across chromosome I", fontsize = 14)
ax1.set_xlabel("Logarithmic binned genomic distances", fontsize = 12)
ax1.set_ylabel("Number of events", fontsize = 12)
ax1.grid()
ax1.legend(fontsize = 12)

ax2 = plt.subplot(122)
ax2.plot(coverage[chromosome], label="Smoothed coverage")
ax2.set_title(f"Covering across chromosome I - bins of {resolution} bp",
    ↪ fontsize = 14)
ax2.set_xlabel(f"Bin number", fontsize = 12)
ax2.set_ylabel("Number of reads", fontsize = 12)
ax2.legend(fontsize = 12)
ax2.grid()
plt.show()

```



1.11 Step 9: Clean temppoary files

```
[15]: hio.tidy_folder(output_dir = output_folder)
```