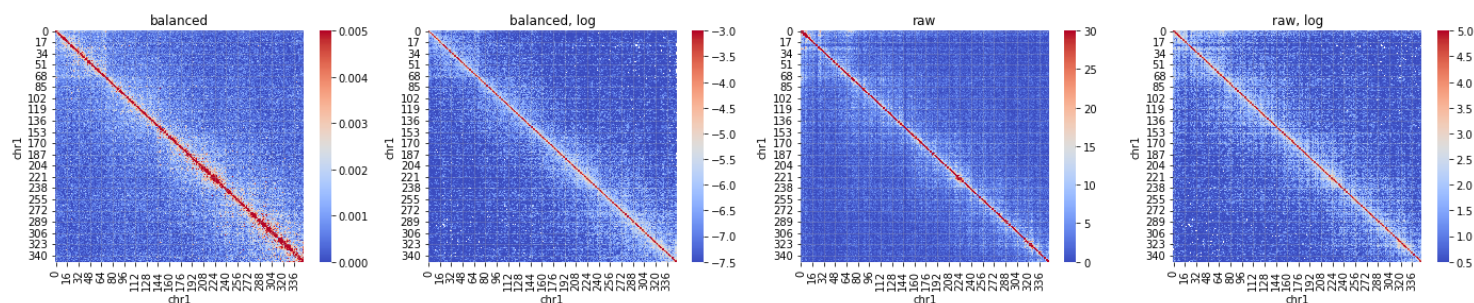
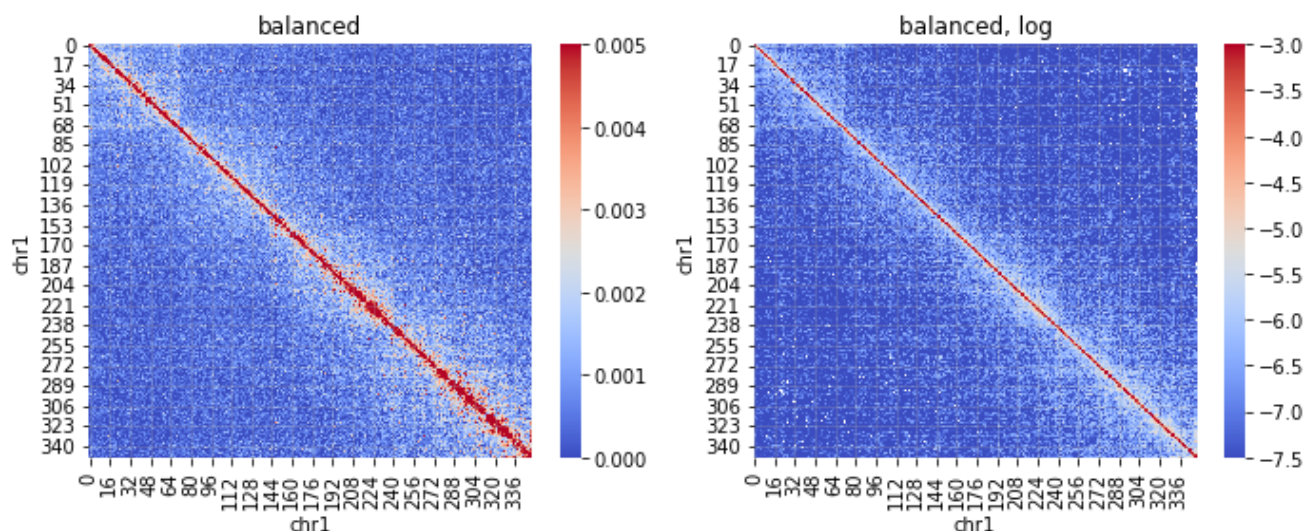


1) Цель задания - получить Hi-C карту для PAO1. Включите в отчет картинку получившейся карты (отрисовка полученной матрицы любым удобным способом в питоне) для 2х вариантов: до итеративной коррекции и после. Сравните их. Опишите итеративно-корректированную карту Hi-C, какие характерные для бактерий особенности можно наблюдать на полученной карте?

После итеративной коррекции		До итеративной коррекции	
не логарифмированная	логарифмированная	не логарифмированная	логарифмированная



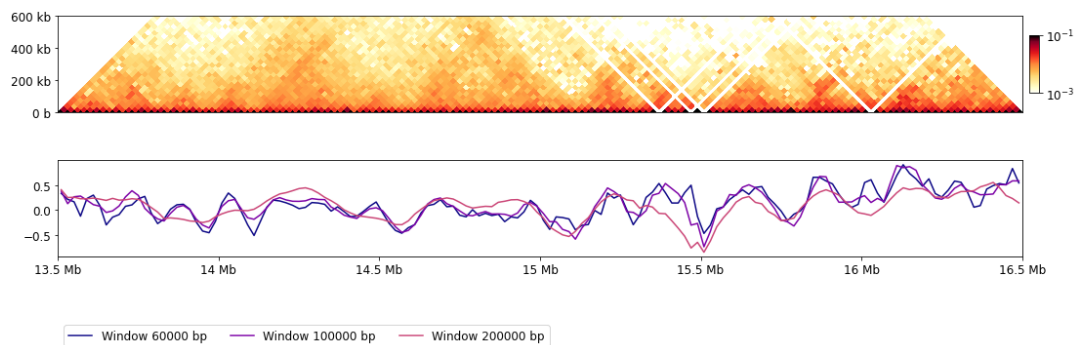
Видно, что после итеративной коррекции карта получается более “сглаженной”. Насколько я поняла, убираются артефакты, образующиеся при пробоподготовке в эксперименте. Часто на “сырых” Hi-C картах заметны места разрезания рестриктазой, места, на которые не картировались риды, а также есть регионы, на которые попало относительно больше ридов, чем на остальные регионы. Такие неровности можно принять за нативные структуры хроматина, однако это не так. На корректированной карте, неровности эксперимента сглаживаются. Рассмотрим ближе итеративно-корректированную карту Hi-C.



Можно видеть наличие небольшого количества хромосомных компартментов, однако четких TADов, какие можно замечать на Hi-C картах эукариот нет - нет инсуляторных белков CTCF, образующих границы TADов в эукариотических клетках. На логарифмированной карте можно видеть небольшие сгущения контактов в углах, противоположных углам главной диагонали. Это характерно для бактериальных кольцевых геномов.

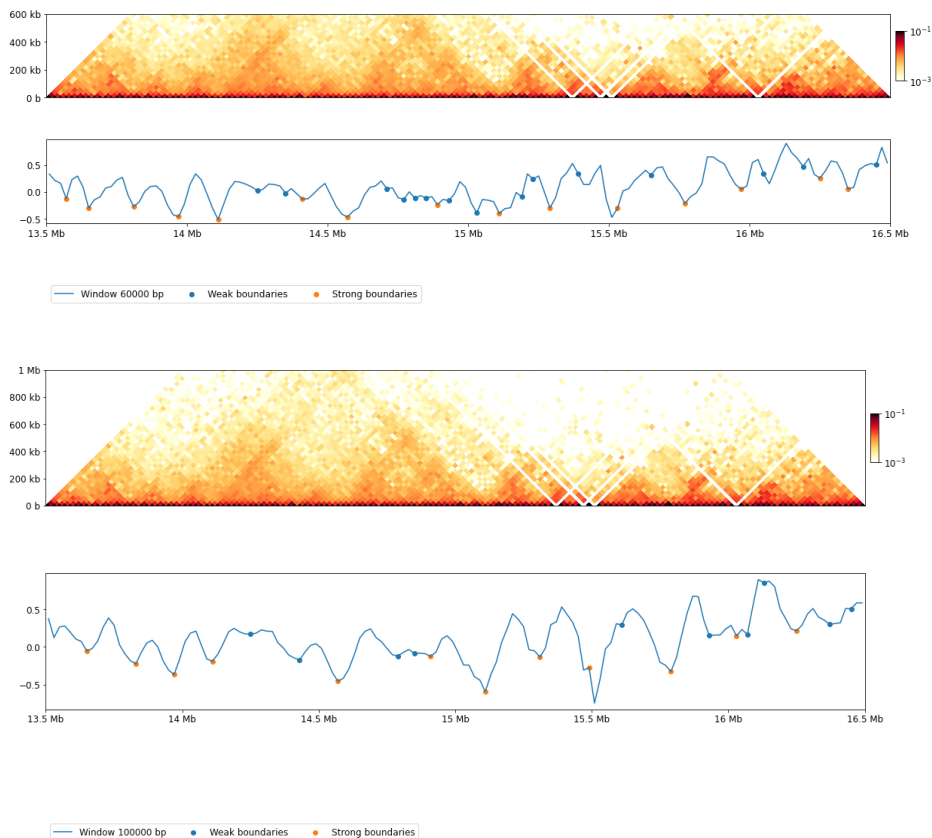
2) Детекция границ с помощью метода insulation из инструмента cooltools. Для готовой Hi-C карты HiC_hESC_merged.hg38.marq_30.10000.mcool (250kb) определите и визуализируйте границы. Получите таблицу с информацией о значении Insulation score и детектированными границами для трех вариантов окна (см ноутбук) и двух разрешений - 20000 и 40000. Визуализируйте границы вместе с картой предложенными в ноутбуке методами. Оцените разметку для перечисленных вариантов. Какую бы вы выбрали для дальнейшего анализа? Стоит ли, по вашему мнению, включать в финальную разметку слабые границы?

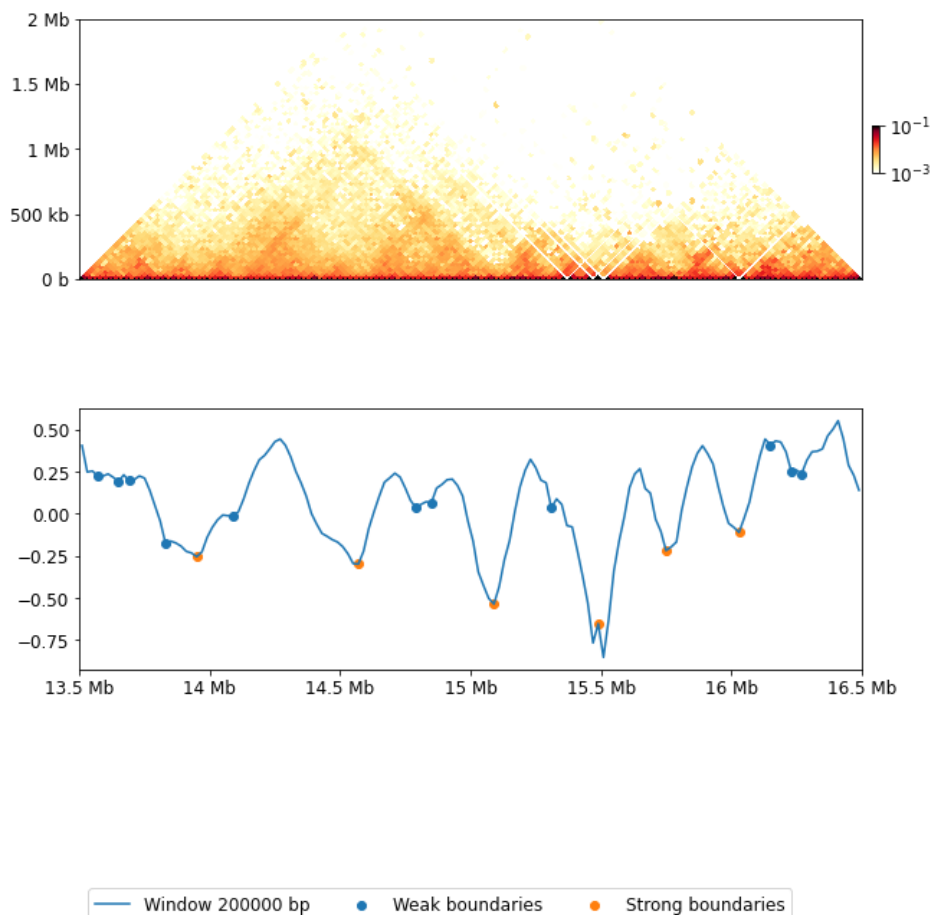
- Для разрешения 20000:



По моему мнению для дальнейшего анализа стоит брать разметку, посчитанную для среднего окна 100Kb: для окна 60Kb видно, что будет детектировано много мелких компартментов, которые, вероятно не так интересны для анализа - это просто соседние участки хромосомы. Для наибольшего окна график выглядит слишком сглаженным, можно что-то упустить при анализе.

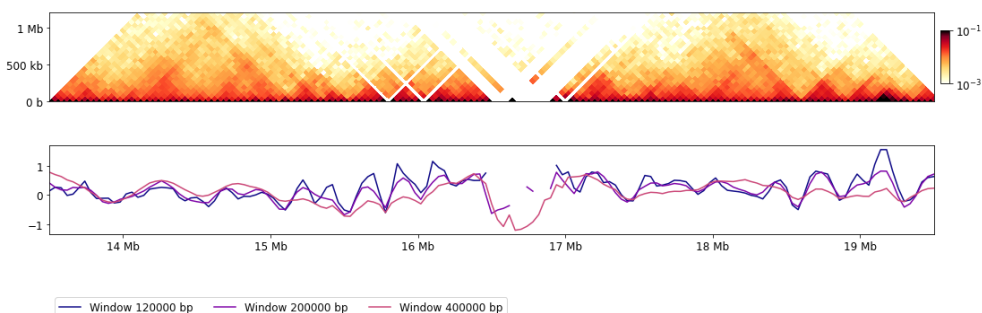
Оценка границ для трех вариантов окна (60Kb/100Kb/200Kb):



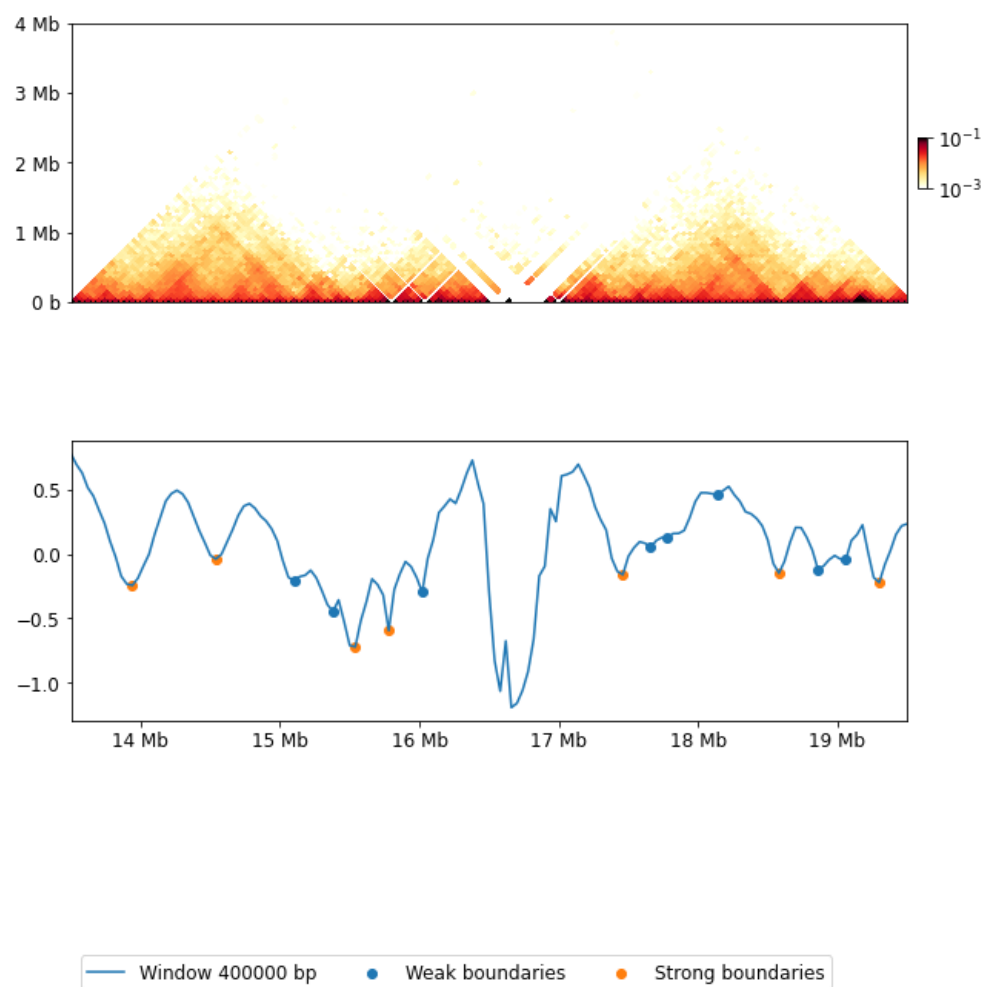
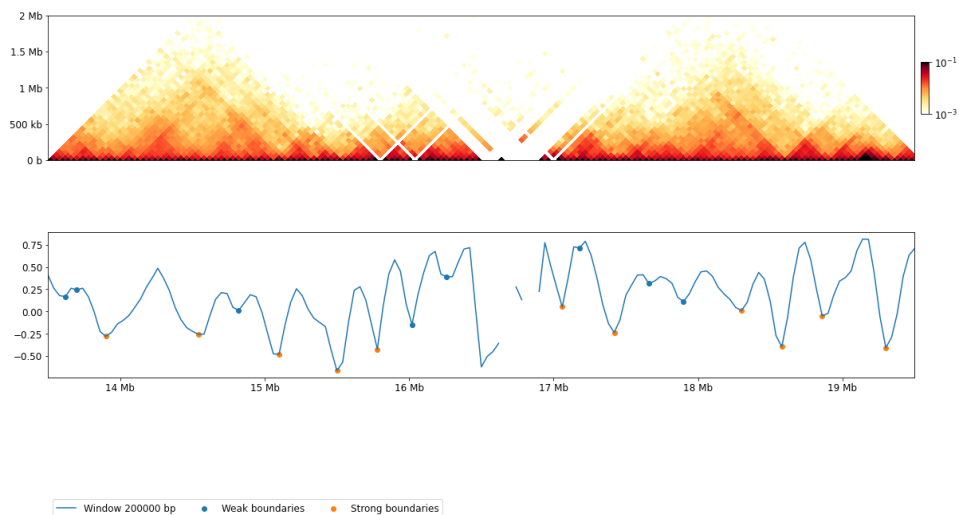
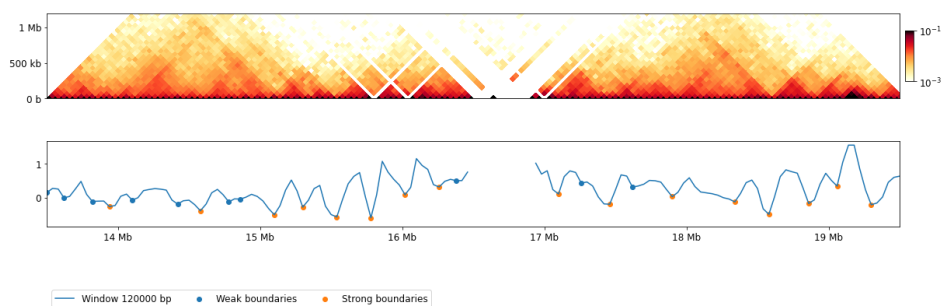


Наверное, для разных применений подойдет анализ разных окон: для определения глобальных хромосомных компартментов использование большого окна, для локальных компартментов - меньшего окна. При этом, кажется, что для большого окна стоит использовать все детектированные границы, а для меньшего только сильные границы. Однако, это опять таки будет зависеть от цели и задач исследования. Также, вероятно, использование меньшего окна с малым разрешением приведет к ошибкам в детекции границ из-за малого покрытия ридами.

- Для разрешения 40000:



Большее разрешение более целесообразно использовать для определения глобальных хромосомных компартментов. На большем разрешении четче видны границы по графикам, мне кажется, что можно использовать не только сильные границы для дальнейшего анализа. При этом график отрисованный для границ при наибольшем окне 400000bp определяет не все границы. Вероятно, стоит брать окно поменьше



3) Детекция петель. Получите разметку петель 2 инструментами (cooltools и mustache) для хромосомы 1 и разрешения 20kb. Сравните кол-во петель, полученное двумя способами. Отрисуйте средние хитмэпы вокруг полученных петель и определите, какой инструмент дает более яркий паттерн для "средней петли". Какой инструмент вы бы предпочли для дальнейшей работы? Отрисованные средние хитмэпы и комментарии включите в отчет.

```
✓ [56] cooltools_dots = pd.read_csv("cooler_cooltools_dots.tsv", sep="\t")
    cooltools_dots.shape

(620, 22)
```

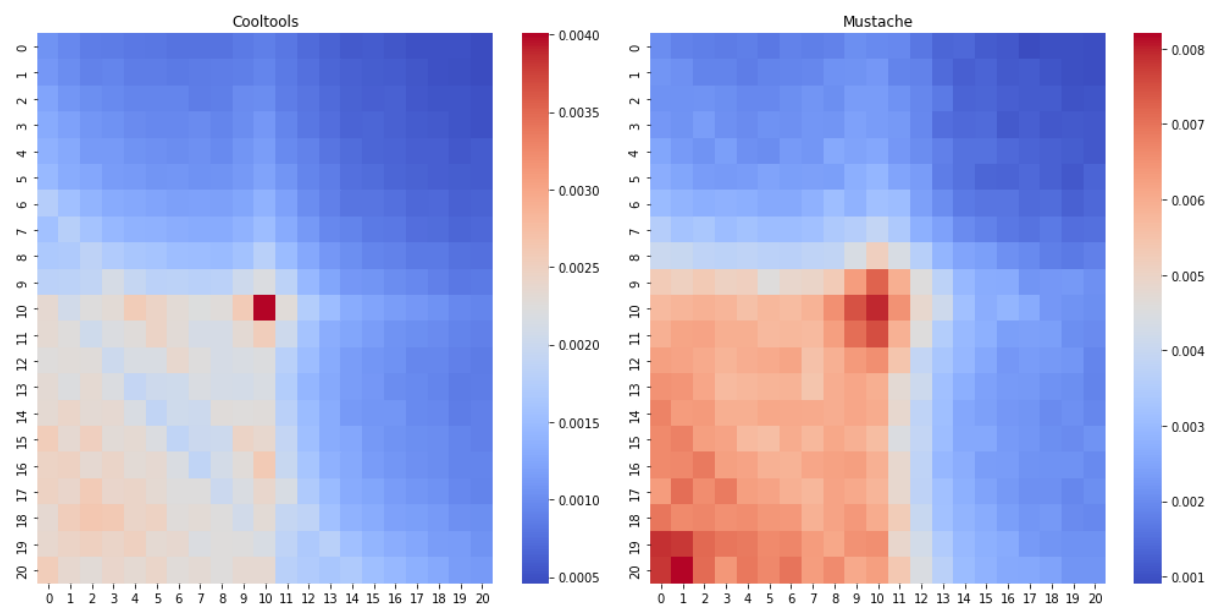
```
✓ [55] mustache_dots = pd.read_csv("cooler_mustache_dots.tsv", sep="\t")
    mustache_dots.shape

(77, 8)
```

Cooltools определяет на порядок больше петель (620), чем mustache (77).
Средние хитмэпы:

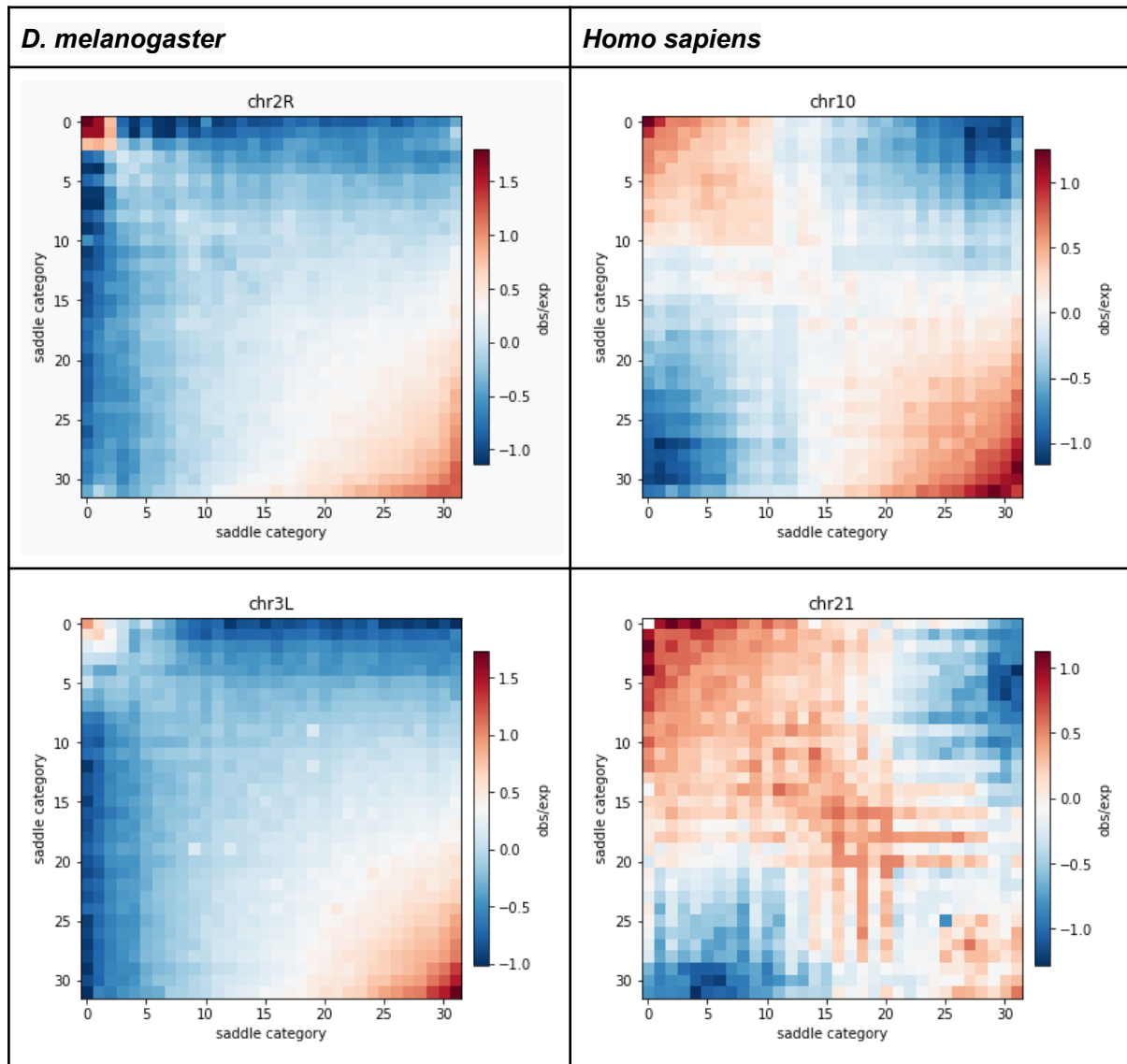
```
✓ [43] def count_matrix(region_matrix: np.ndarray, dots_df: pd.DataFrame, binsize: int, n: int = 21):
    n_s = n // 2
    n_e = (n // 2) + 1

    meanloop = np.zeros((dots_df.shape[0], n, n))
    coordinate = lambda start, end: int((start + end) / 2 / binsize)
    for dot in dots_df.itertuples():
        start1, start2 = coordinate(dot.start1, dot.end1), coordinate(dot.start2, dot.end2)
        meanloop[dot.Index] = region_matrix[start1-n_s:start1+n_e, start2-n_s:start2+n_e]
    return np.nanmean(meanloop, axis=0)
```



Несмотря на более яркую картинку для средней петли, полученную с помощью mustache, cooltools дает более качественную картинку с четко видной петлей. Хотя mustache и работает быстрее, его можно использовать, например для быстрого анализа данных. Однако для полного анализа, стоит использовать cooltools, потому что на одинаковых данных он способен детектировать больше петель и корректнее их определить, раз картинка для средней петли у него выглядит менее размыто.

4) Получение разметки компартментов и построение saddle plot. Для двух организмов (мушка и человек) выберите какую-нибудь хромосому, получите первый собственный вектор, характеризующий принадлежность к компартментам, и постройте saddle plot (просто отрисуйте полученную матрицу любым способом в питоне). Опишите, что отображено на saddle plot и какие различия в структуре компартментов наблюдаются у мушки и человека.



У человека наблюдается большее количество контактов, как между А-А компартментами, так и между В-В. У Дрозофилы же в принципе меньше контактов, особенно мало контактов неактивного хроматина, самого неактивного хроматина также меньше.

У меня есть 2 гипотезы:

1) вероятно, хроматин человека сильнее компактизирован, т.к. он на порядок длиннее, состоит из большего количества хромосом, при этом размер эукариотической клетки и ядра ограничен. У человека также больше замалчиваемого хроматина, составляющего большую часть генома (интроны, транспозоны, SINE, LINE)

2) на клетках человека доказана модель экструзии петли, позволяющая усиливать компактизацию хроматина и создавать большое количество компартментов. Не нашла информацию, как у Дрозофилы происходит компактизация хроматина и образование компартментов, однако по графикам видно, что слабее, чем у человека. Возможно, это объясняется отсутствием механизма экструзии петли.