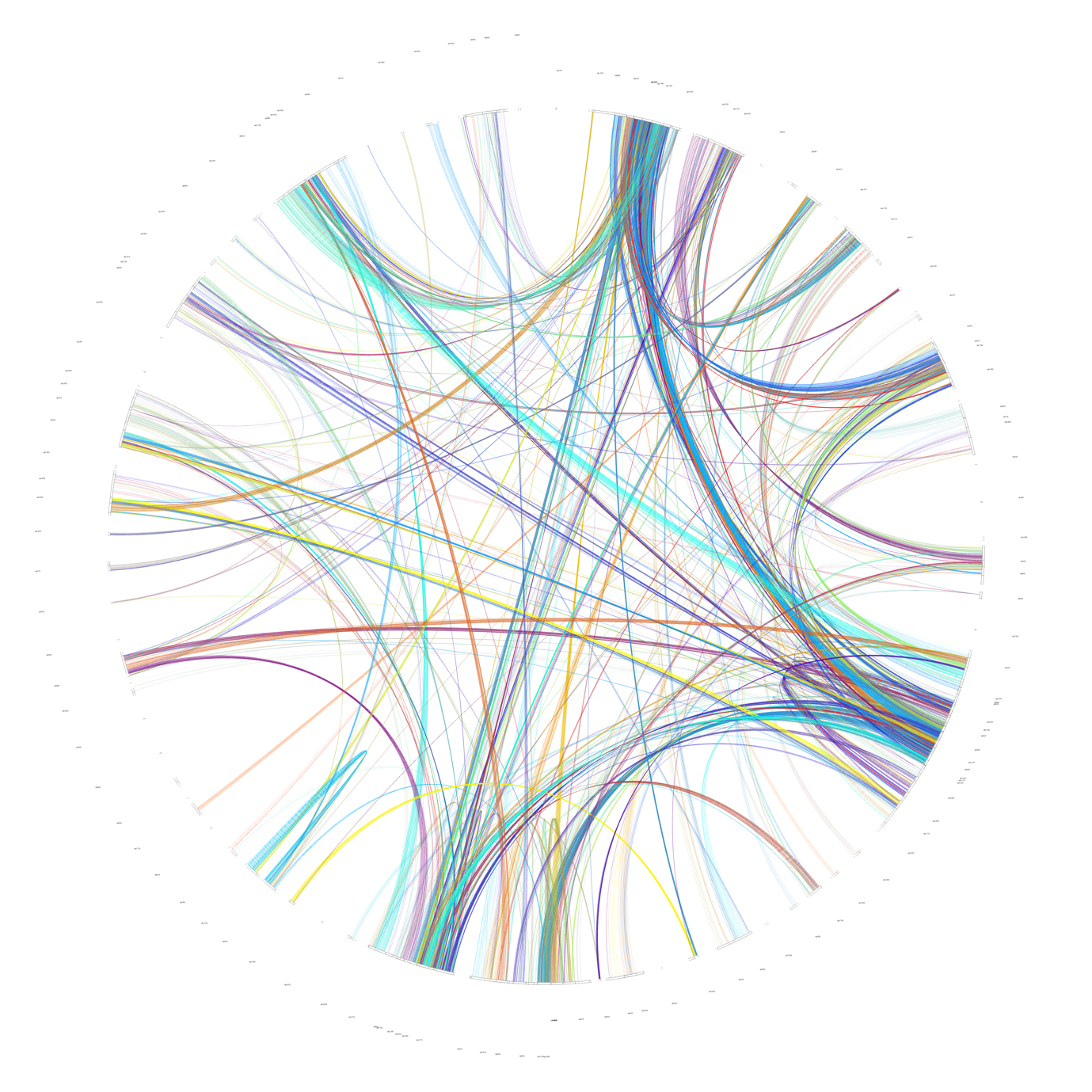
현유봉 교수님께.

안녕하세요 교수님. 인턴 임현수 입니다.

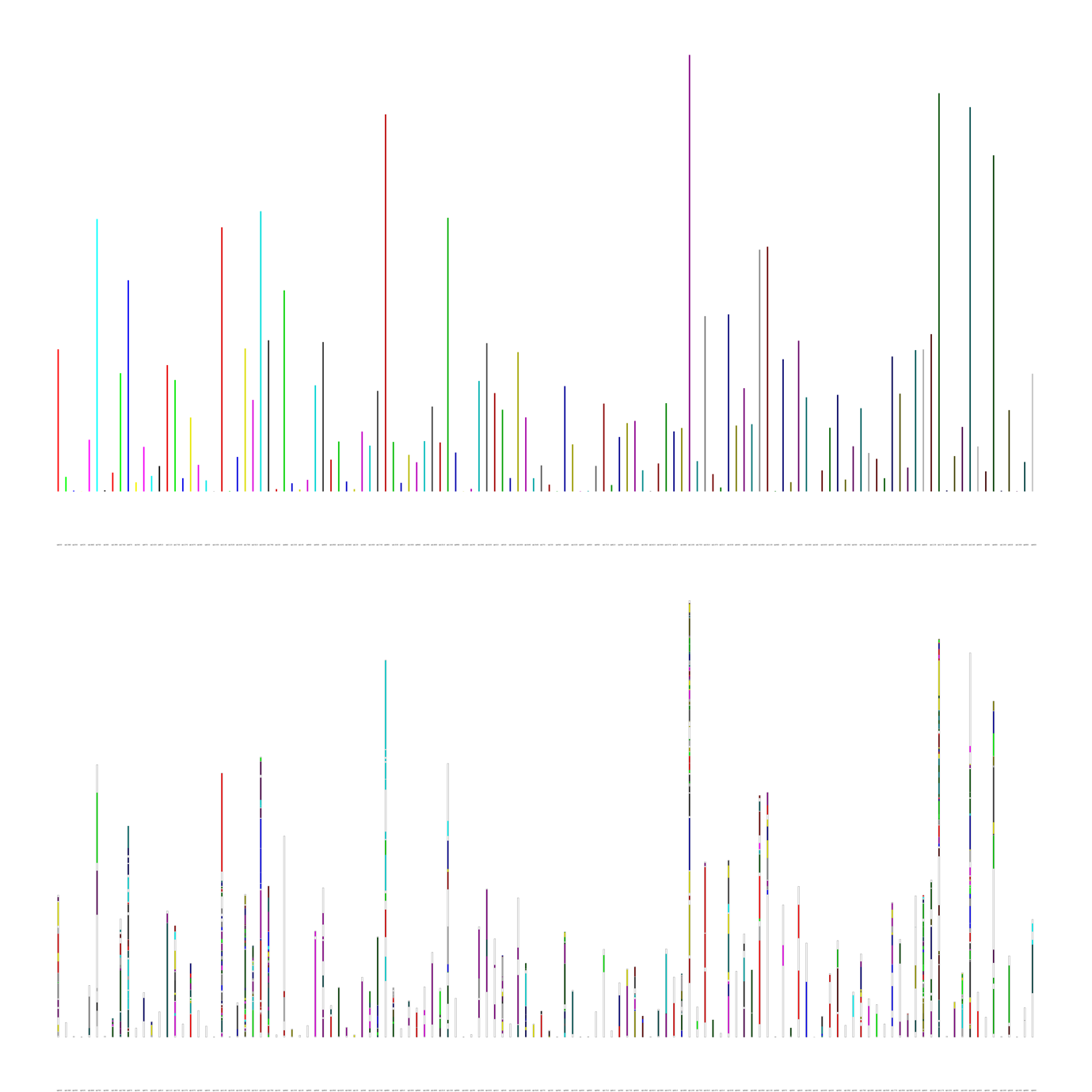
*Viola* 유전체를 공부하다가, 이번 데이터미팅 이전에 교수님께 조언을 구하기 위해 연락드리게 되었습니다.

먼저 전체 contig들의 synteny를 분석하여 circle plot으로 나타낸 그림은 **그림 1**, 그리고 **그림 2**와 같습니다. 분석 과정에서 blast를 선행하였고, e-value를 1e-80으로 잡아 거의 같은 서열을 가진 paralog만 검출되도록 하였습니다.

직관적으로 본 결과 분석한 *Viola*의 유전체 내부에서 유전자 대다수가 2 이상의 bridge가 아닌 1 bridge(한 쌍)로만 연결되어 두 근연종으로부터 유전체를 받아 형성된 alloploid나 autoploid라는 것을 관찰해 볼 수 있습니다. 그러나 busco를 비롯한 다른 결과로부터 alloploid라는 점을 이미 알고 있기 때문에 기존의 관찰 결과와 일치하는 결론이라는 것을 확인할 수 있습니다.

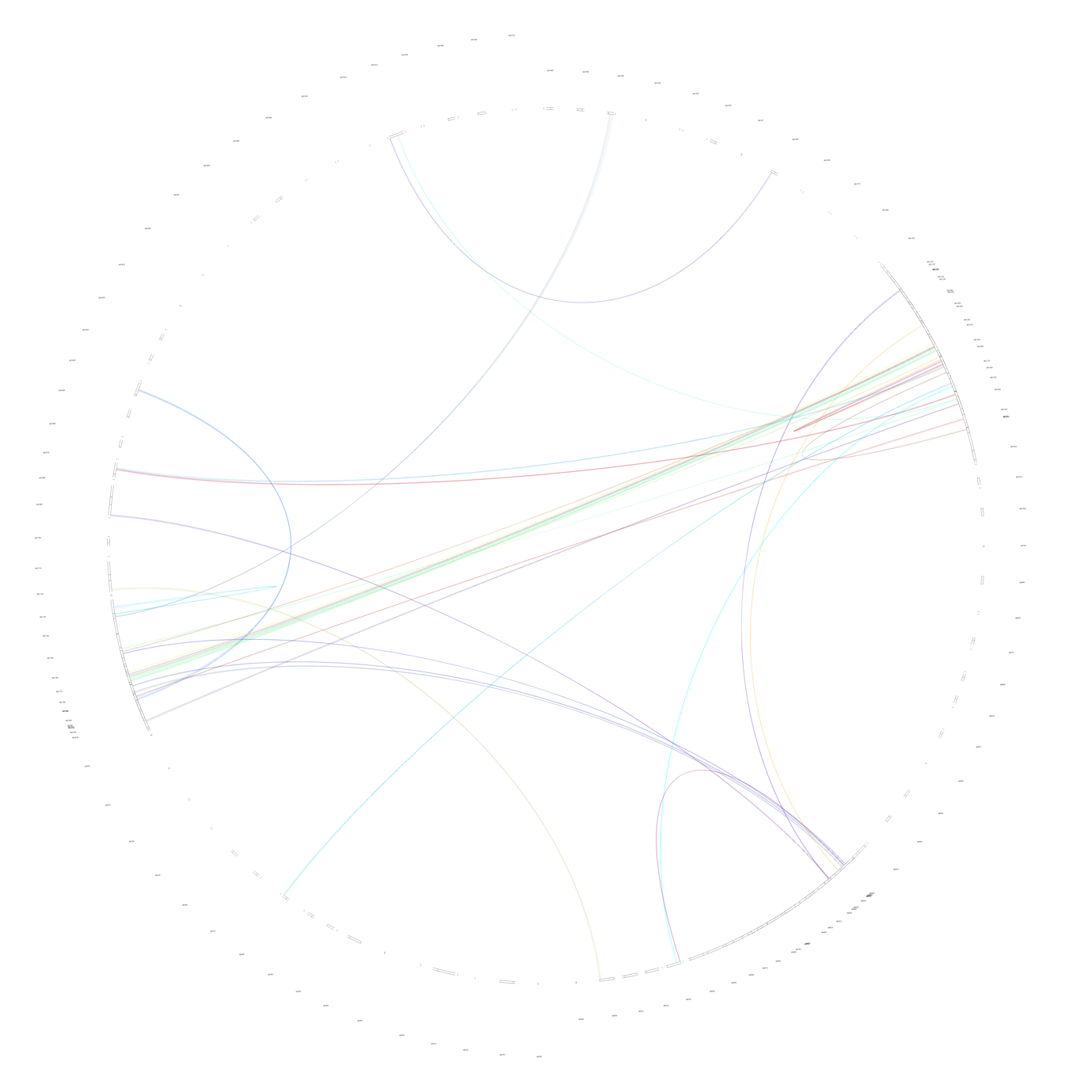


**그림 1.** 186개 contig들의 synteny scan 결과(circle plot)

또, **그림 2**를 통해 각 contig들의 어떤 부분이 다른 contig들에서 발견되는지 확인하여 볼 수 있습니다. 

**그림 2.** 186개 contig들의 synteny scan 결과(bar plot)

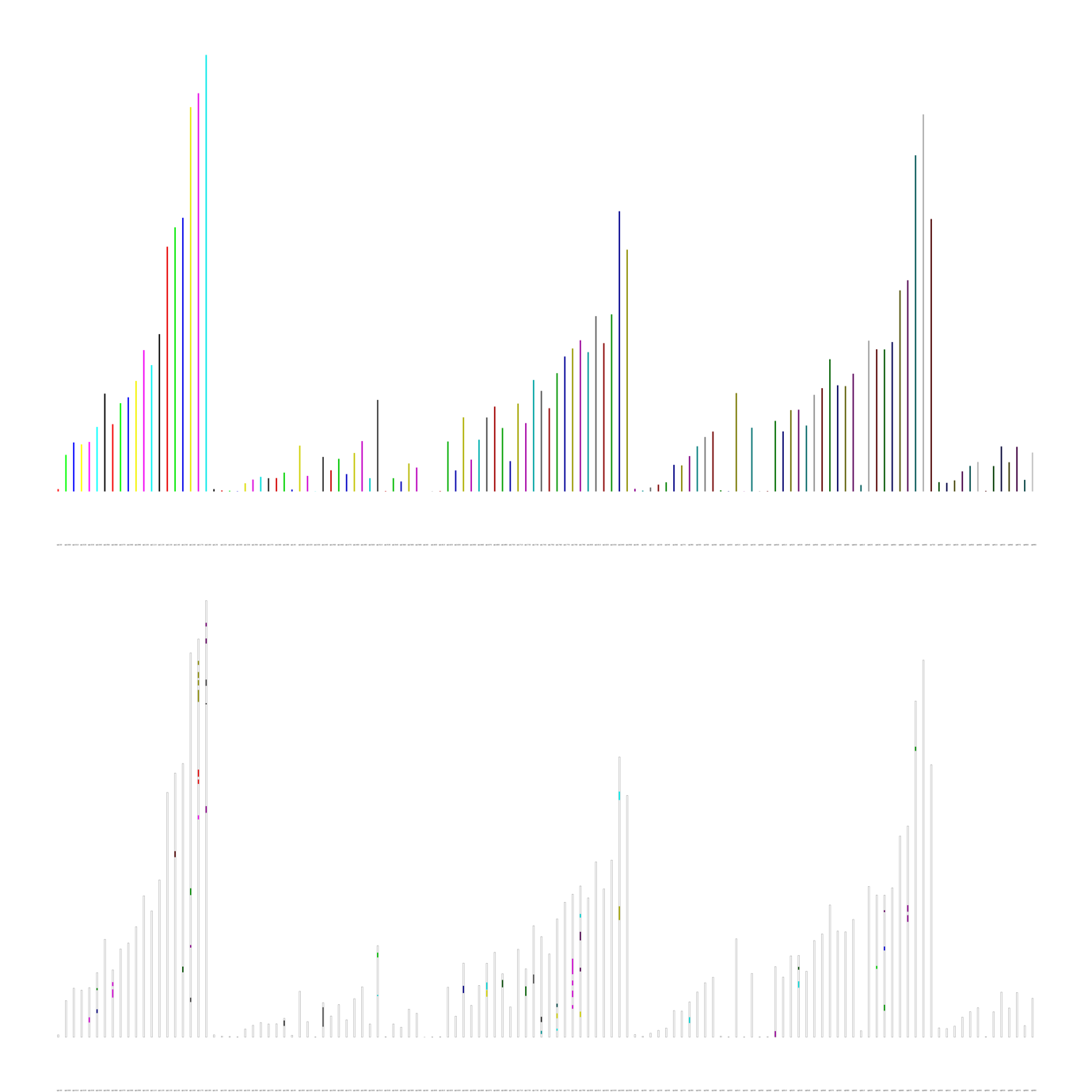
일차적으로 저는 위 **그림 1**과 **그림 2** 를 바탕으로 synteny를 통해 인위적으로 두 그룹으로 분류하고자 하였습니다. 왜냐하면 분석에 사용한 *Viola philippica*는 알고 있는 대로 두 “근연종” 이 alloploid가 된 것이기 때문에 두 근연종에서 발견되는 한 쌍의 synteny가 *V.philippica*에서 관찰될 것이기 때문입니다. 그러나 엮인 정도가 너무도 복잡하여 busco 유전자(busco eudicot-10 database 기반 hmm v3.0 scan)들 가운데 dual copy로 나타난 유전자들만 대상으로 synteny scan을 다시 시도해 보았습니다. 이에 대한 결과는 **그림 3**과 **그림 4**에 있습니다.



**그림 3.** 186개 contig의 Busco 유전자들에 대한 synteny scan 결과(circle plot)

Busco 유전자들을 대상으로 하는 스캔에서는 28개의 쌍이 나왔으며, 그 결과는 **그림 3**과 같습니다. **그림 1**에 비해 많이 축약된 형태이며, dual copy busco 유전자들만 검색했기 때문에 정확하게 한 bridge(한 쌍) 만 나타납니다. **(**반면 **그림 1**에서는 대다수가 1 bridge이지만 몇 가지 유전자는 복수의 bridge로 나타납니다. 예컨대 11시 방향에서 출발하는 하늘색 synteny가 있습니다.)

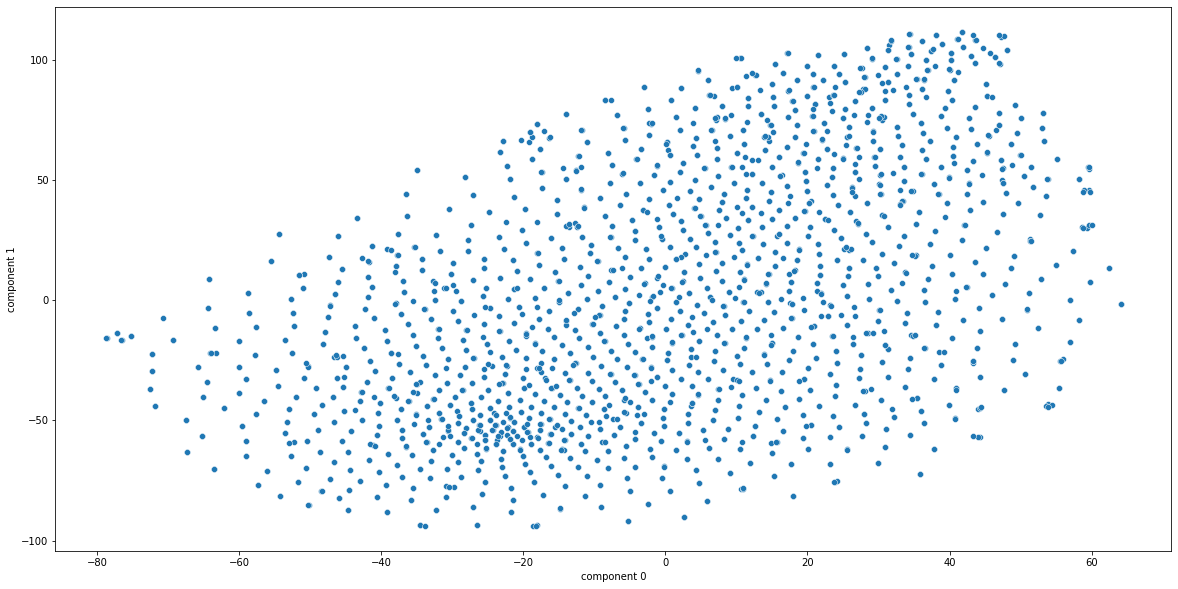
이 분석 결과를 바탕으로 최대 56개(겹치는 contig들이 있기 때문에 최대 56개입니다. 아직 정확히 몇 개의 contig들을 분류할 수 있는지는 세어보지 못 했습니다.)의 contig을 두 그룹으로 나눌 수 있으며, sequence identity를 느슨하게 하여 e-value를 높인다면, 조금 더 많은 homology를 찾아낼 수 있을 것으로 보입니다.



**그림 4.** 186개 contig의 Busco 유전자들에 대한 synteny scan 결과(bar plot)

더불어, **그림 4**에서 확인할 수 있는 각 contig들의 다른 contig 포함 정보를 바탕으로 **그림 3**을 보조하여 contig들의 그룹을 나누는 데 도움이 될 것이라고 생각합니다.

또한, 지난번 데이터미팅 때 보여드렸던 kmer 벡터화를 통한 t-SNE 분석에 대하여 busco 유전자들에 대한 같은 분석을 수행하여 보았습니다. 그 결과는 **그림 5**와 같고, 저는 이를 통해 별다른 메시지를 얻을 수 없었습니다. 가설로는 두 근연종에 대하여 genome kmer profile이 비슷할 것으로 예상하여 두 그룹으로 클러스터링 될 것을 기대했습니다만, 그러한 흔적을 fine-funing을 어느 정도 마친 이후에도 찾아볼 수 없어서 novel한 방법이었지만 별다를 수확은 없었던 것으로 결론 내리고 있습니다.



**그림 5.** *Viola philippica*의 Busco double copy유전자들에 대한 kmer tsne plot

여기서 제 질문은, 최대 56개의 contig들을 분류한 이후

1) 어떤 아이디어를 통해 contig들의 순서를 결정할 수 있는지(현재는 MUMmer이라는 소프트웨어를 사용해 contig들을 permutation 조합하여 모든 경우의 수에 대해 가장 타당하게 dot plot이 직선으로 나타나는 경우를 확인해 보고자 합니다.)

2) scaffolding할 때 gap의 길이는 어떤 아이디어로 파악할 수 있는지

3) **그림 5**에서 다른 메시지를 발견할 수 있는지

4) 현재까지 수행한 분석에 대하여 진화적으로 타당하지 않거나 생물학적으로 의미가 없는 것들이 있는지

궁금합니다.

감사합니다.

좋은 하루 되세요!!

임현수 올림.

---

길해찬 선배께

안녕하세요 선배님. 현유봉 교수님 연구실 인턴 임현수 입니다.

현재까지 수행한 연구는 교수님께 송부드린 내용과 같으며, 혹 이와 관련하여 커멘트 해주실 내용이 있으신지, 제가 놓치고 있는 부분이 있는지 의견 부탁드립니다.

감사합니다. 좋은 하루 되세요.

임현수 드림.