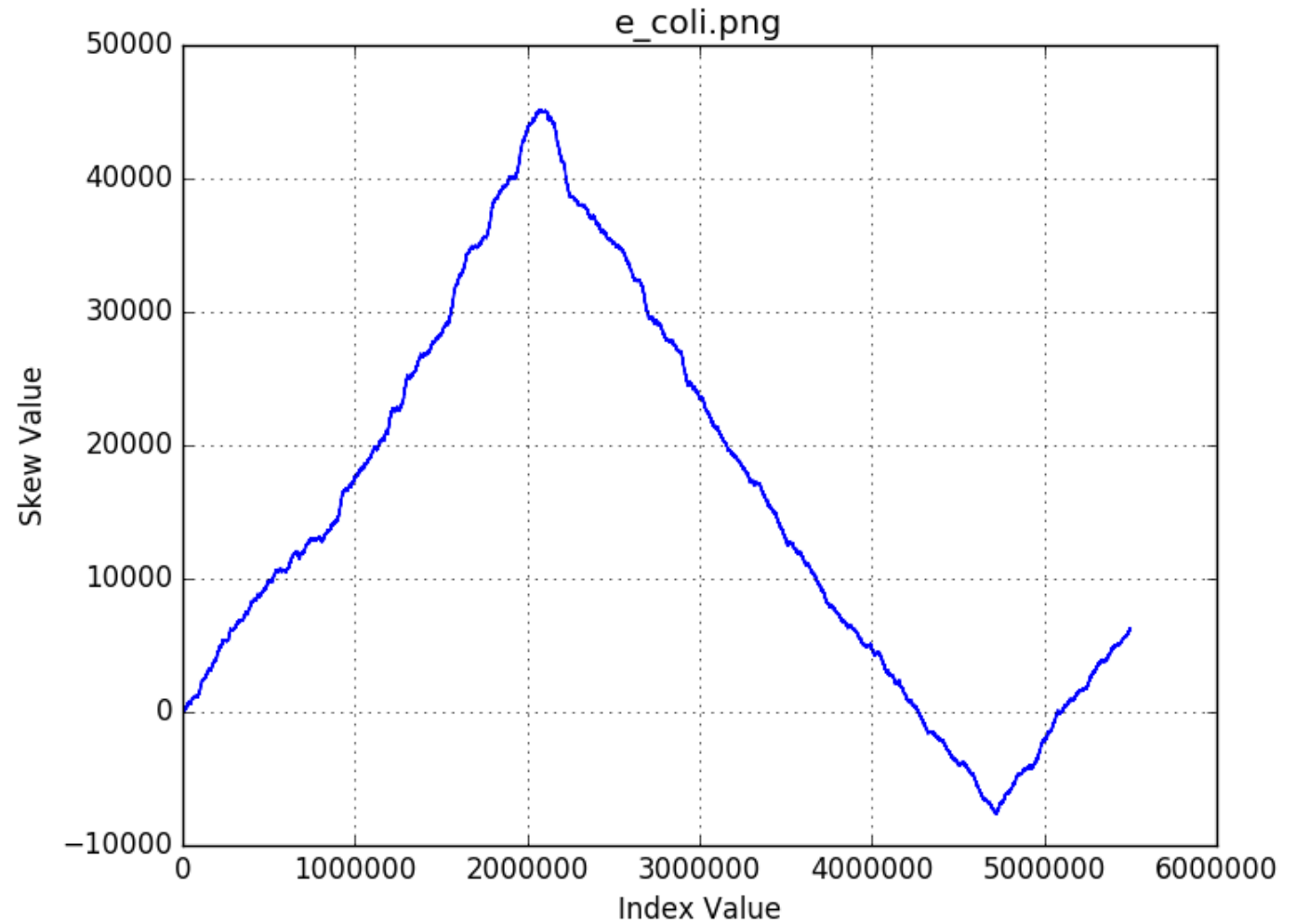


Identifying Origins of Replication Sites in Circular Genomes using a Scoring Method

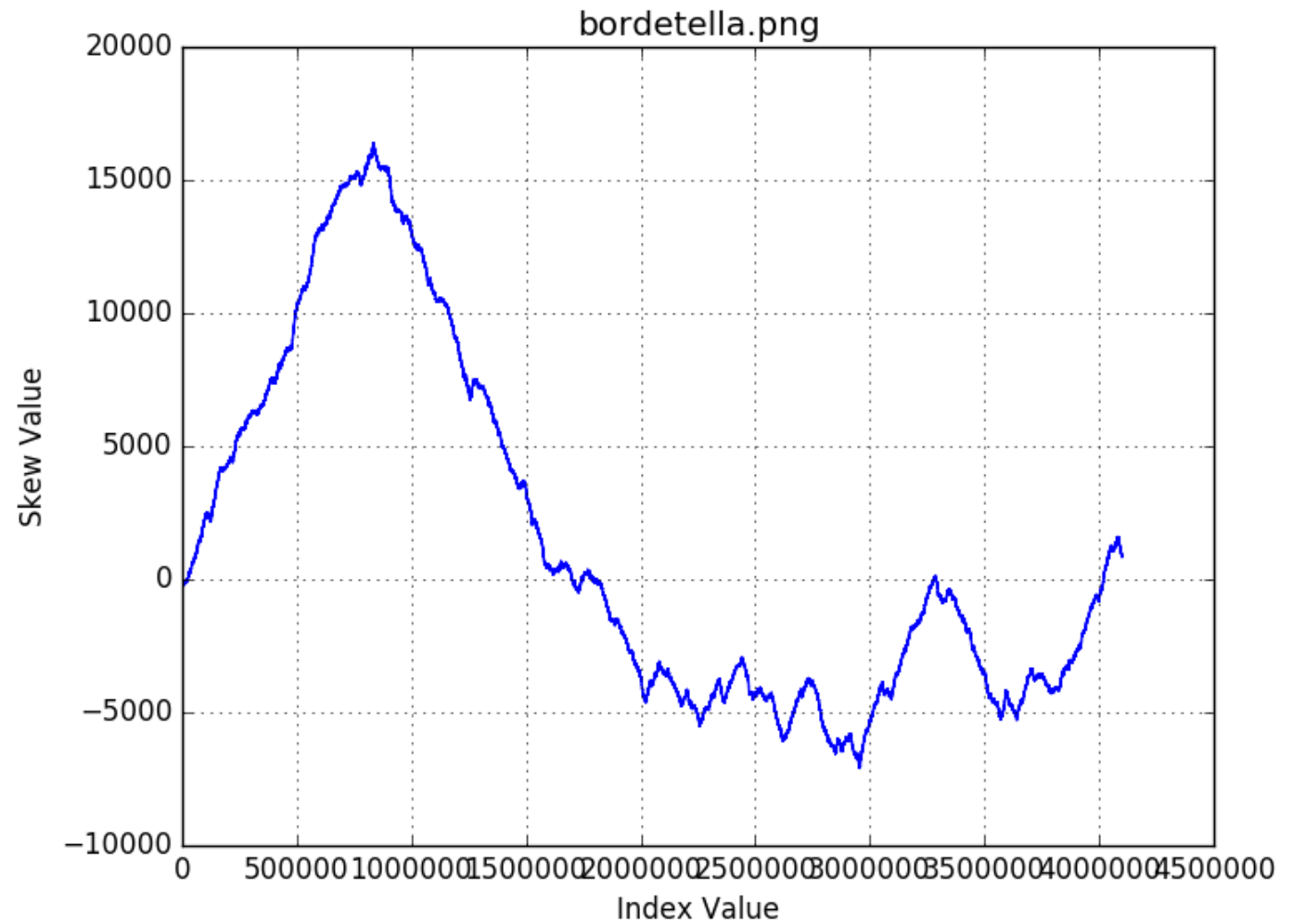
Or, alternatively: Cool Skew Graphs!

Yurel Watson | Bio 131 | Anna Ritz | 5-3-2016

1. What is the problem?



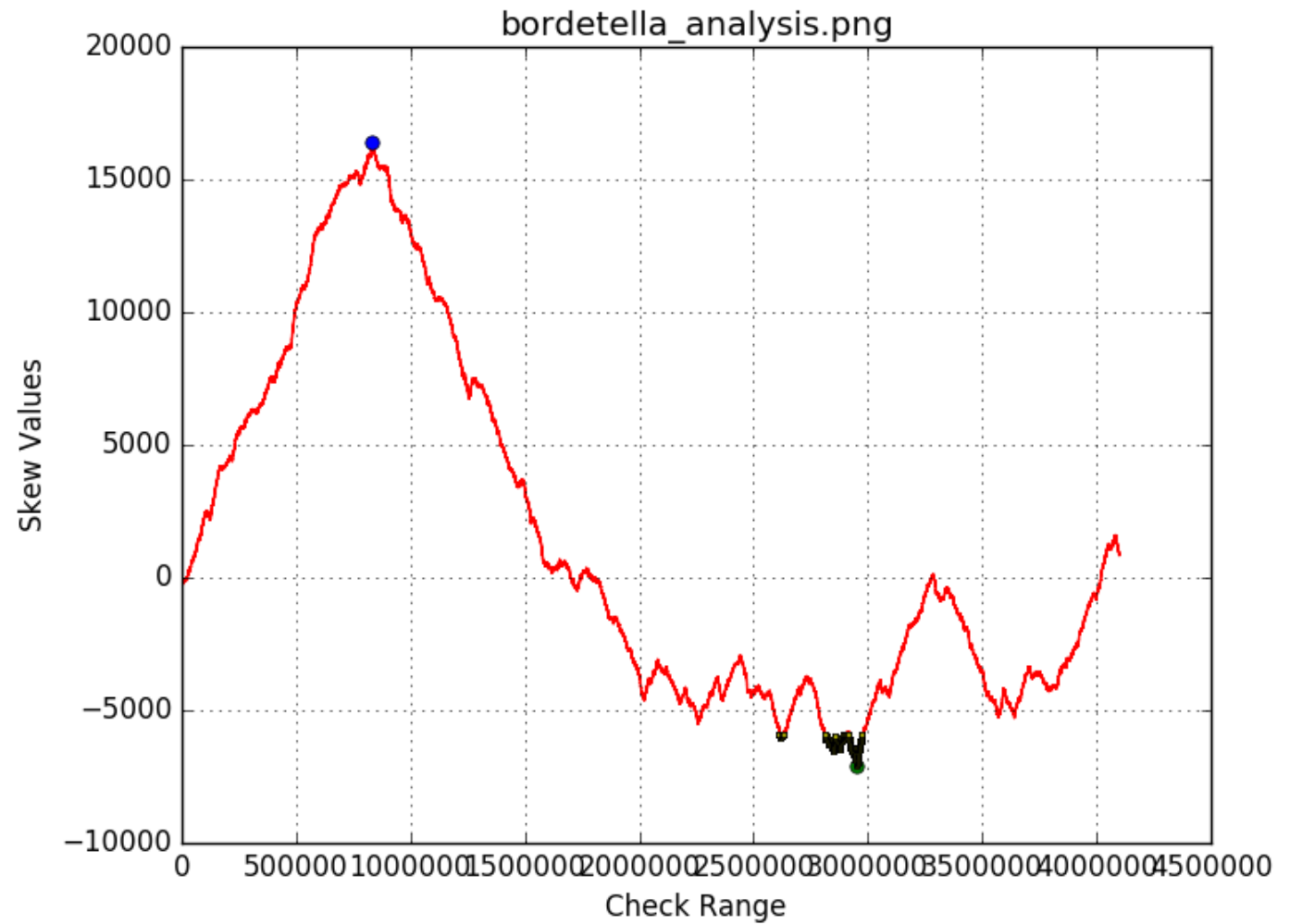
1. What is the problem?



2. How did you try to solve it?

- My program identifies candidate origin of replication sites by scoring the difference between the highest and lowest skew points, and comparing these against a cluster of low-skew data points.

2. How did you try to solve it?



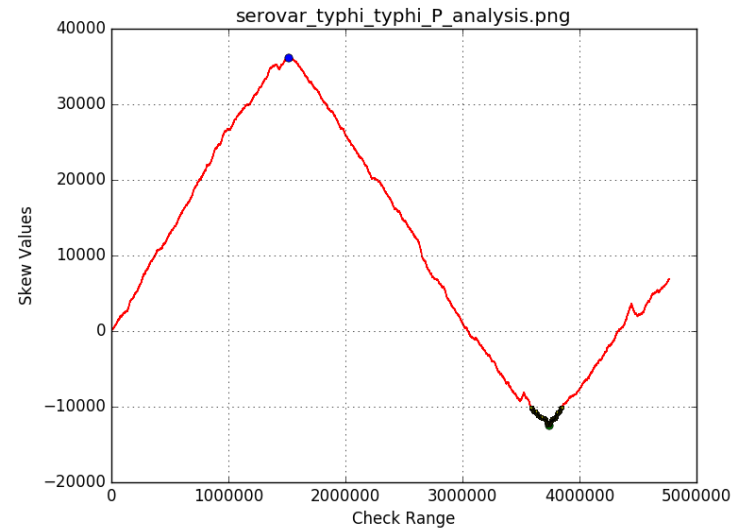
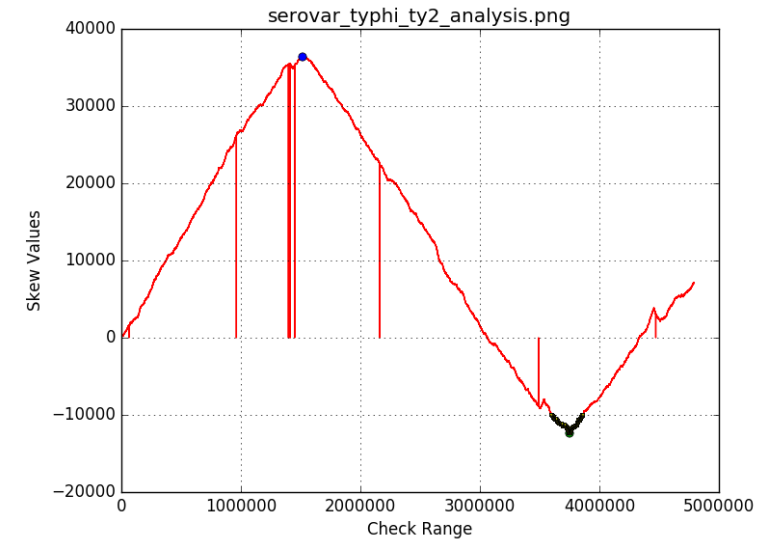
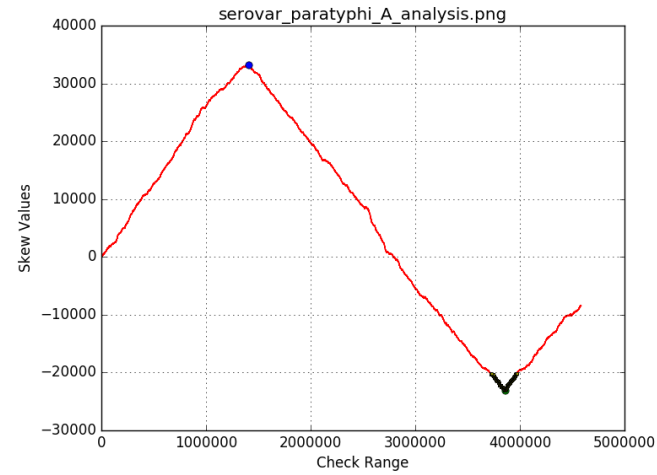
3. What data did I use?

- *achro.txt* - *Achromobacter xylosoxidans* strain FDAARGOS_150 (Accession: CP014028.1 GI: 991940861)
- *bordetella.txt* - *Bordetella pertussis* strain ATCC:BAA-1335D-5 (Accession: CP014153.1 GI: 991852837)
- *e_coli.txt* - *Escherichia coli* O157:H7 str. Sakai DNA (Accession: BA000007.2 GI: 47118301)
- *geo.txt* - *Geobacillus* sp. C56-T3 (Accession: NC_014206.1 GI: 297528376)
- *salmonella.txt* - *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. LT2 (Accession: AE006468.2 GI: 973795115)
- Sample and research datasets sources from GenBank

4. What are the high-level steps of the program?

- Input
- Input Processing
- Printing and Analysis
 - Input and Reading
 - Data Processing
`ore_sorted_tup = [x for x in cand_sorted_tup if score(x, highest_idx) >= comp_score - comp_score*degree_of_dif]`
 - Data Printing
- Exiting

5. What were the results?



6. What do the results mean?

- It is possible to use this program to represent genomic skew data
- The origin of replication appears to occur in similar locations in related species
- With enough tests, it theoretically should be possible to tackle larger statistical endeavors about the origin of replication in a species (specifically, where does the origin of replication typically end up being?)