Koninklijke
Nederlandse
Vereniging voor
Microbiologie

Scientific Spring Meeting
KNVM & NVMM 2022

NVMM

Bert Hubert
bert@hubertnet.nl / @bert_hu_bert
https://berthub.eu/ / https://skewdb.org

Disclosure of speaker's interests

No (potential) conflict of interests	Nothing!
Relations that could be relevant for the meeting	**Company names**
None!	None!

Who I am:

- Not some kind of weirdo!
  - Well…
- Been studying DNA since 2001
- 2011-2013 guest researcher over at Bertus Beaumont's lab, Bionanoscience TU Delft
- Co-author with Cees Dekker, Felix Hol and Juan Keymer of paper on antibiotic resistance (ISMEJ)
- Author of paper in Nature Scientific Data on the present research

Chargaff's **first** parity rule: %A = %T, %C = %G for all double stranded DNA

ACTCCCGGACTTGGA

TGAGGGCCTGAACCT

A: 6

C: 9

G: 9

T: 6

Chargaff's **second** parity rule (1968): %A ≅ %T, %C ≅ %G for **EACH** strand of DNA

ACTCCCGGACTTGGA

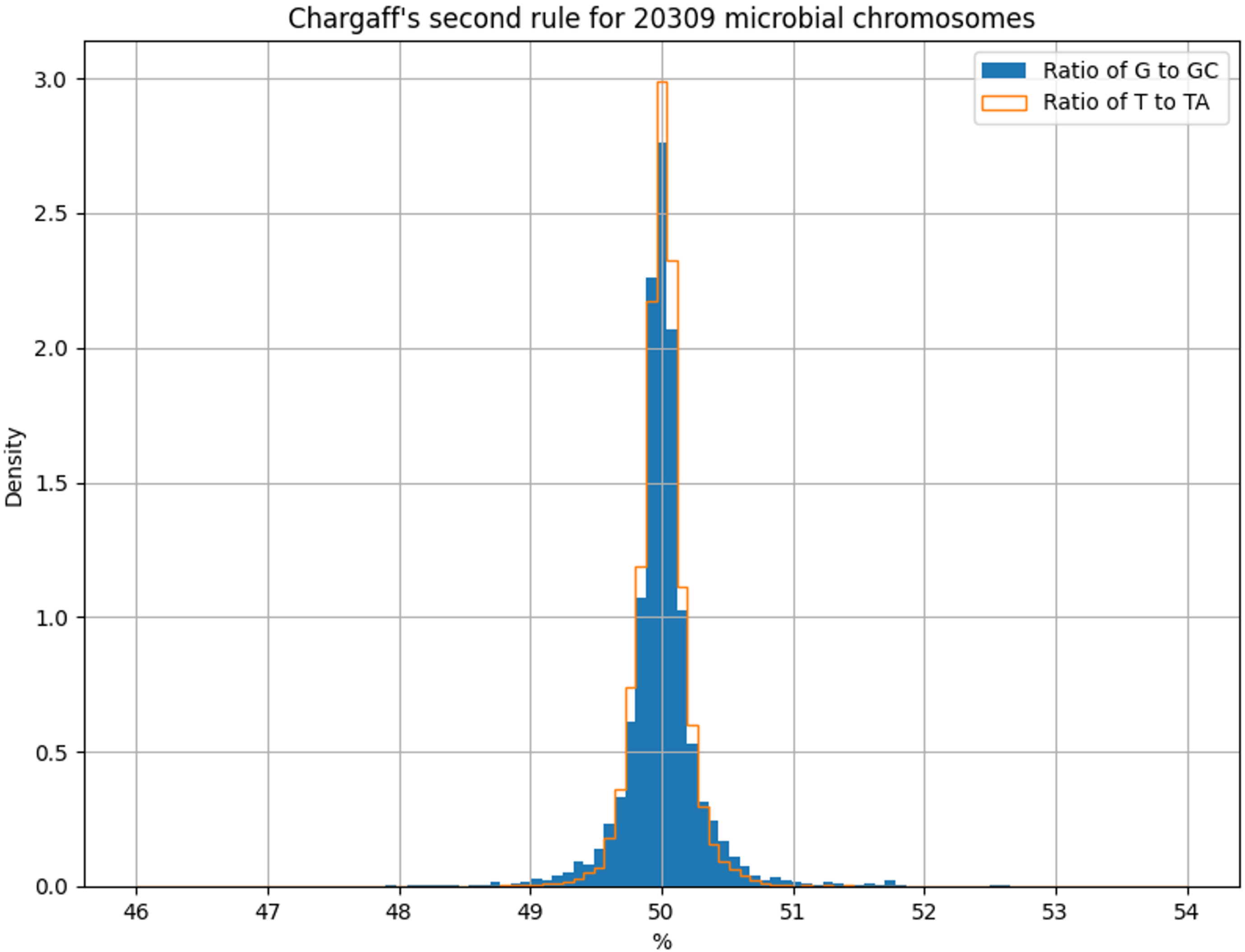
A: 3 T: 3

C: 5 G: 4

TGAGGGCCTGAACCT

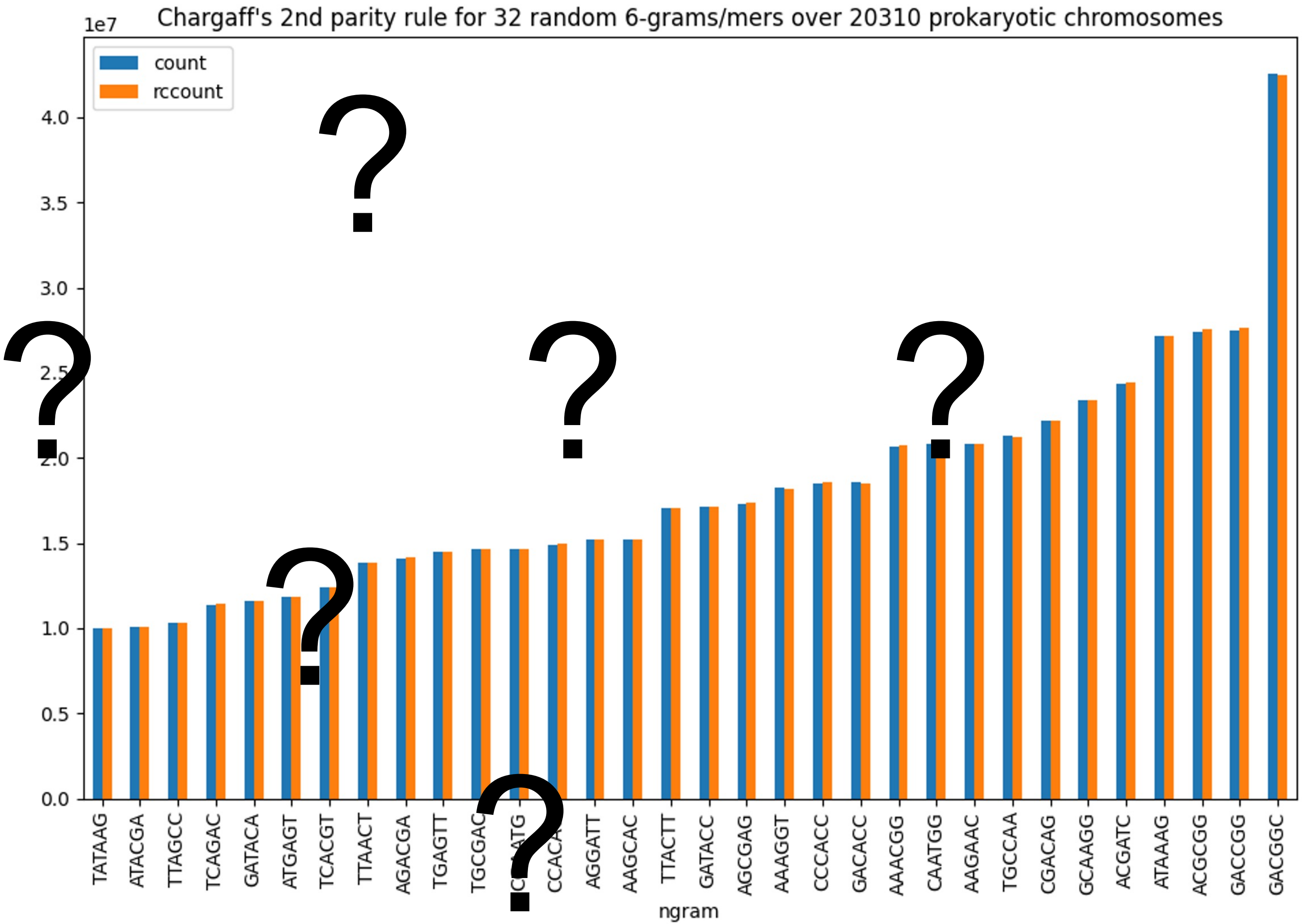
T: 3 A: 3

C: 4 G: 5


Chargaff's second rule for 20309 microbial chromosomes
Ratio of G to GC
Ratio of T to TA
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
46
47
48
49
50
51
52
53
54
%

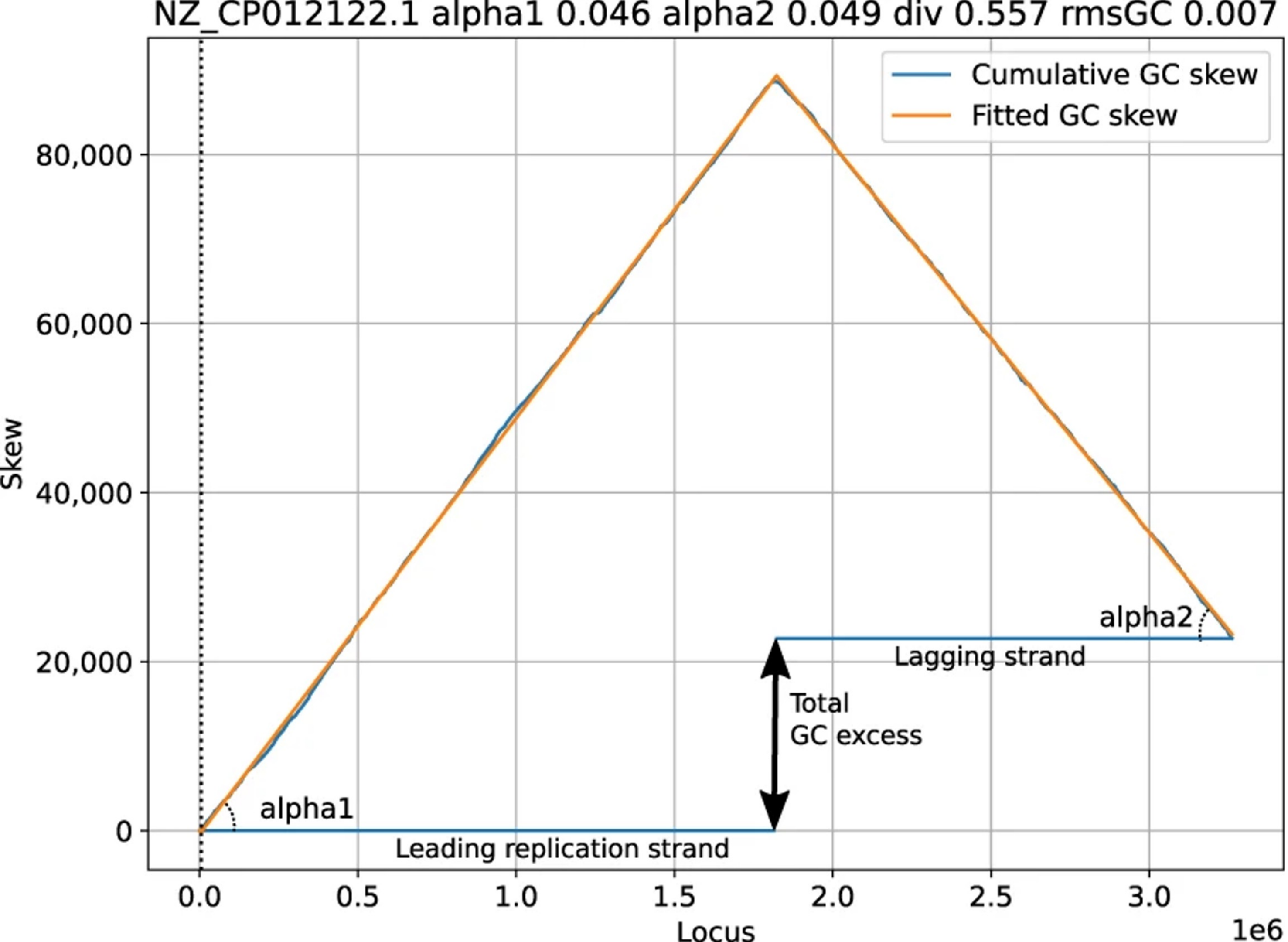
% ACCGTC ≅
% GACGGT

% of any k-mer = equal to % of its reverse complement, up to large k

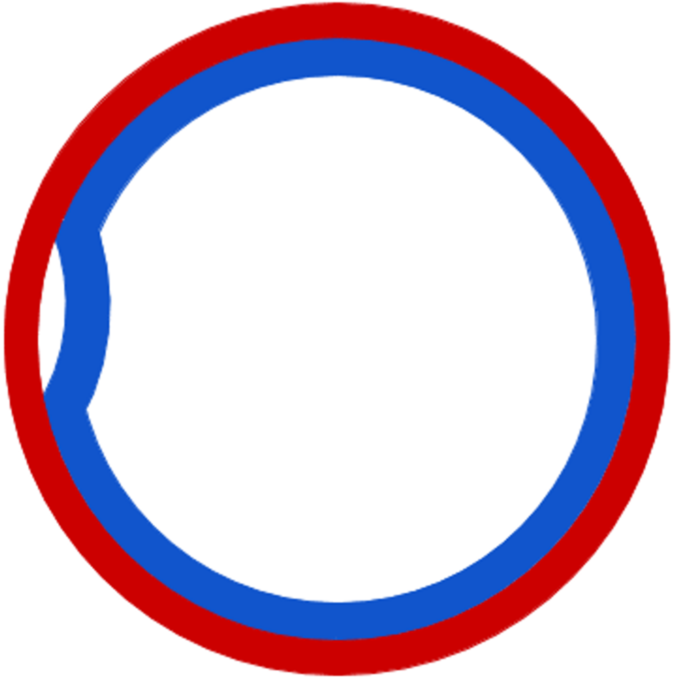


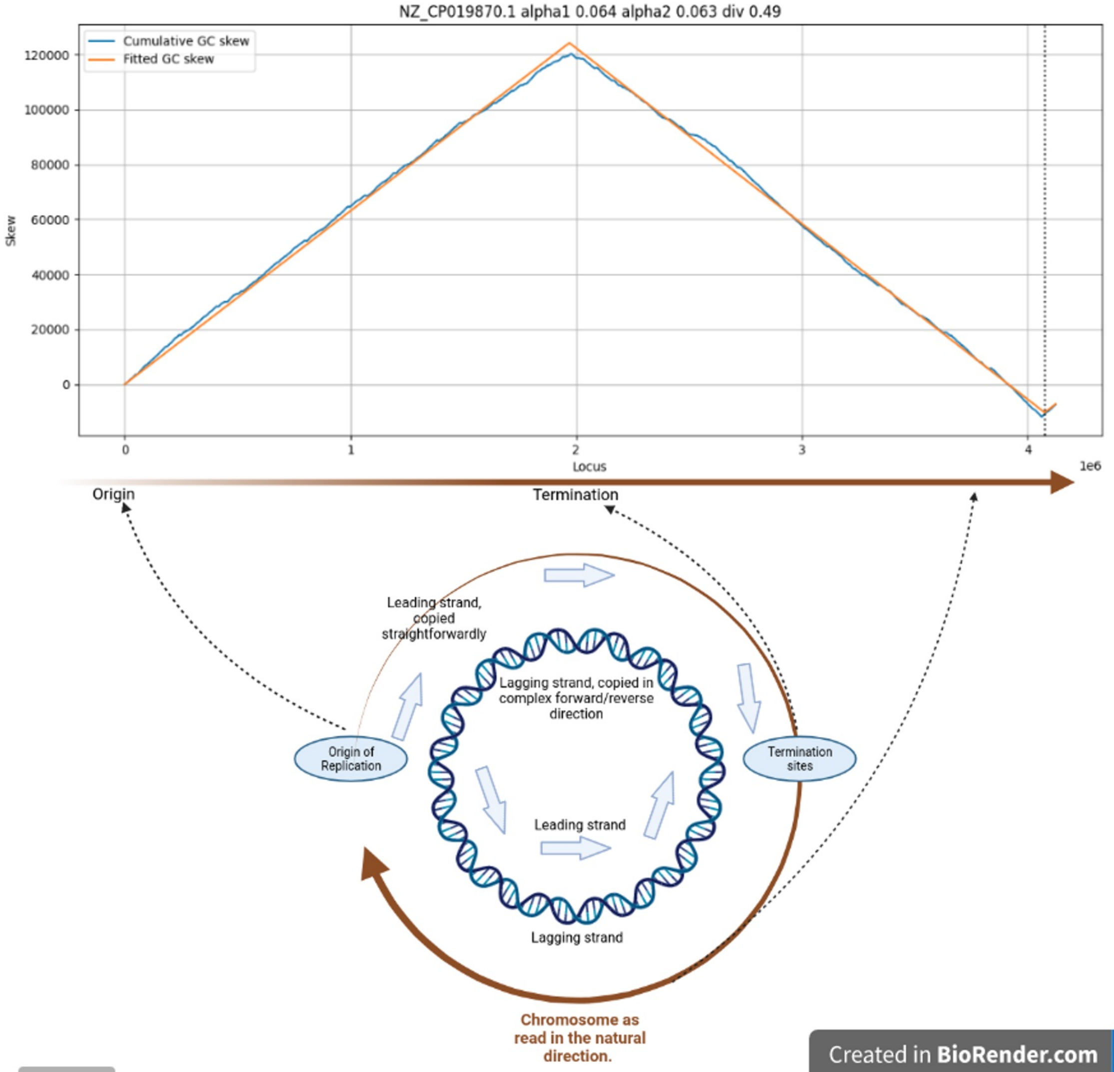
GC Skew

- Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. Molecular biology and evolution 13, 660-665 (1996)
- “One half of almost all circular chromosomes has an excess of G over C, compensated in the other half”
- 2.5% on average, robust, pervasive
  - Not in archaea!



Sample graph showing *Skew*DB data for *Lactiplantibacillus plantarum* strain LZ95 chromosome.





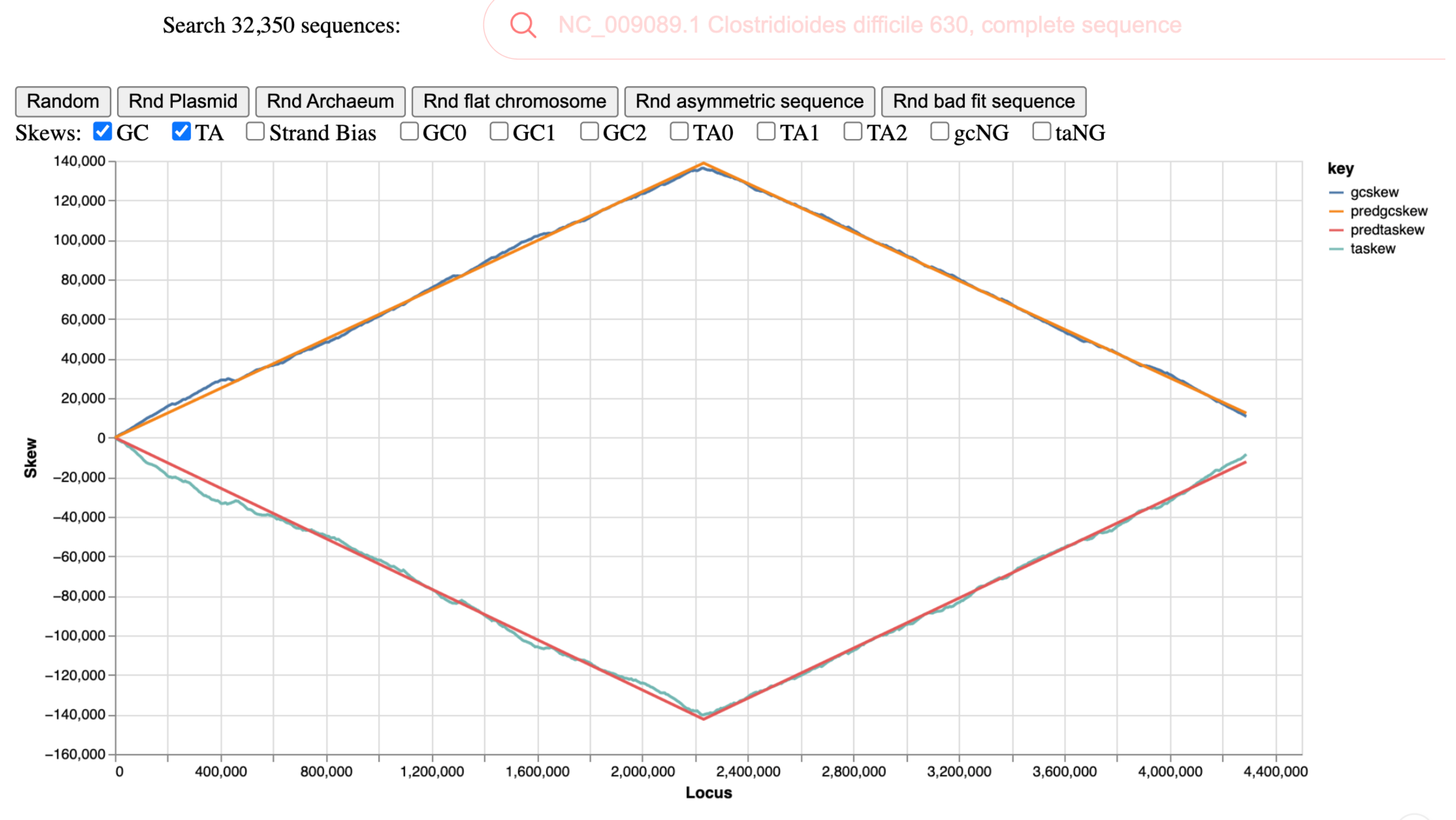
Created in BioRender.com

GC Skew

- De-amination
  - More on the lagging strand
  - DNA repair happens more frequently while transcribing genes, and this happens more on the leading strand (where genes like to huddle)
- “Codon bias * strand bias”
- Third codon position selective pressure
- *None of these explanations on their own is sufficient*

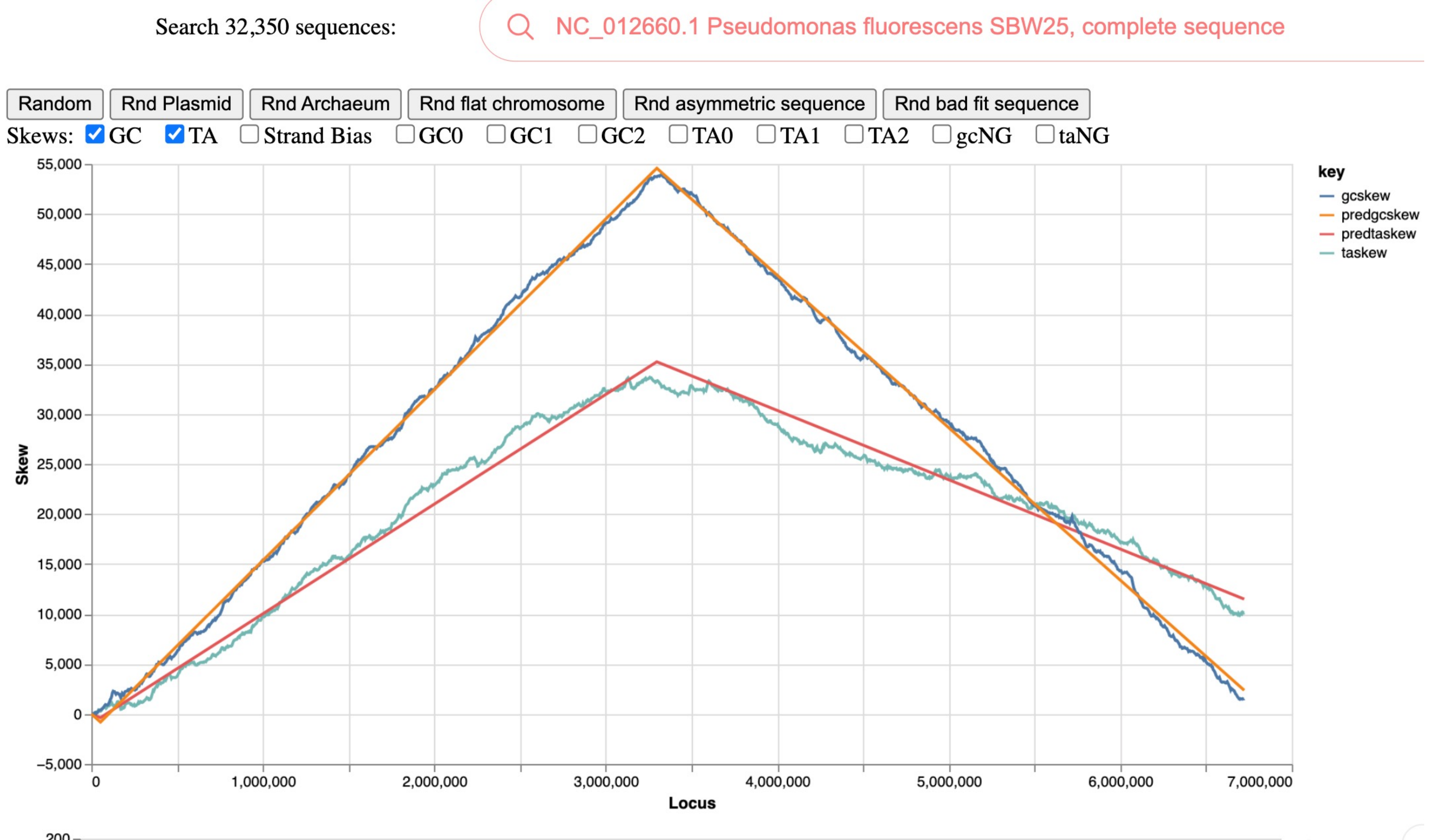
SkewDB: 32,000 chromosomes & plasmids

- Skew information on:
  - GC / TA, leading & lagging strand
  - Split out over:
    - Whole chromosome/plasmid
    - 1st, 2nd, 3rd codon position, non-coding regions
  - Gene strand bias
- Information on: codon biases
- Easy to use, Jupyter notebook available
- https://skewdb.org/ + https://skewdb.org/view



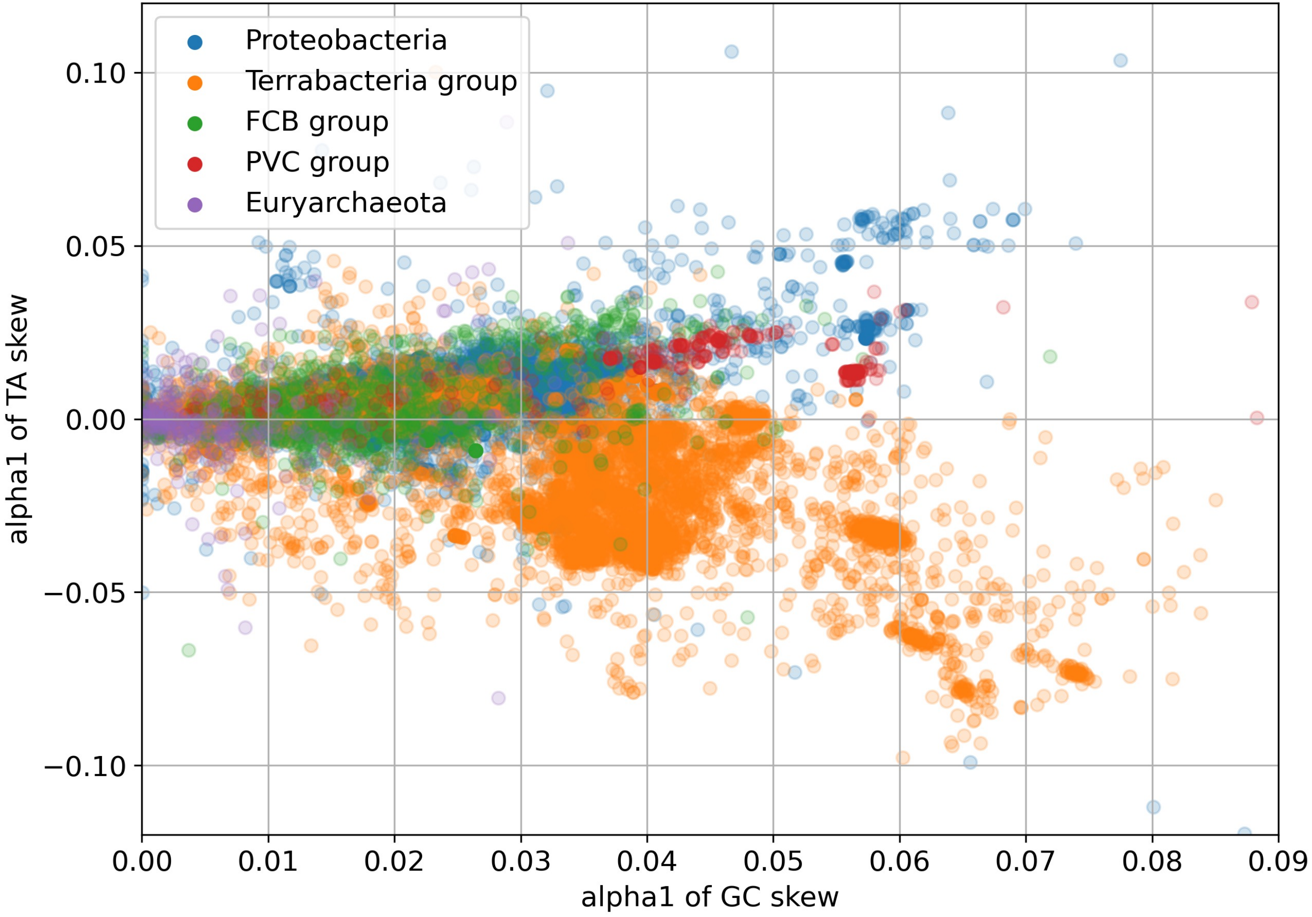
Search 32,350 sequences: NC_009089.1 Clostridioides difficile 630, complete sequence

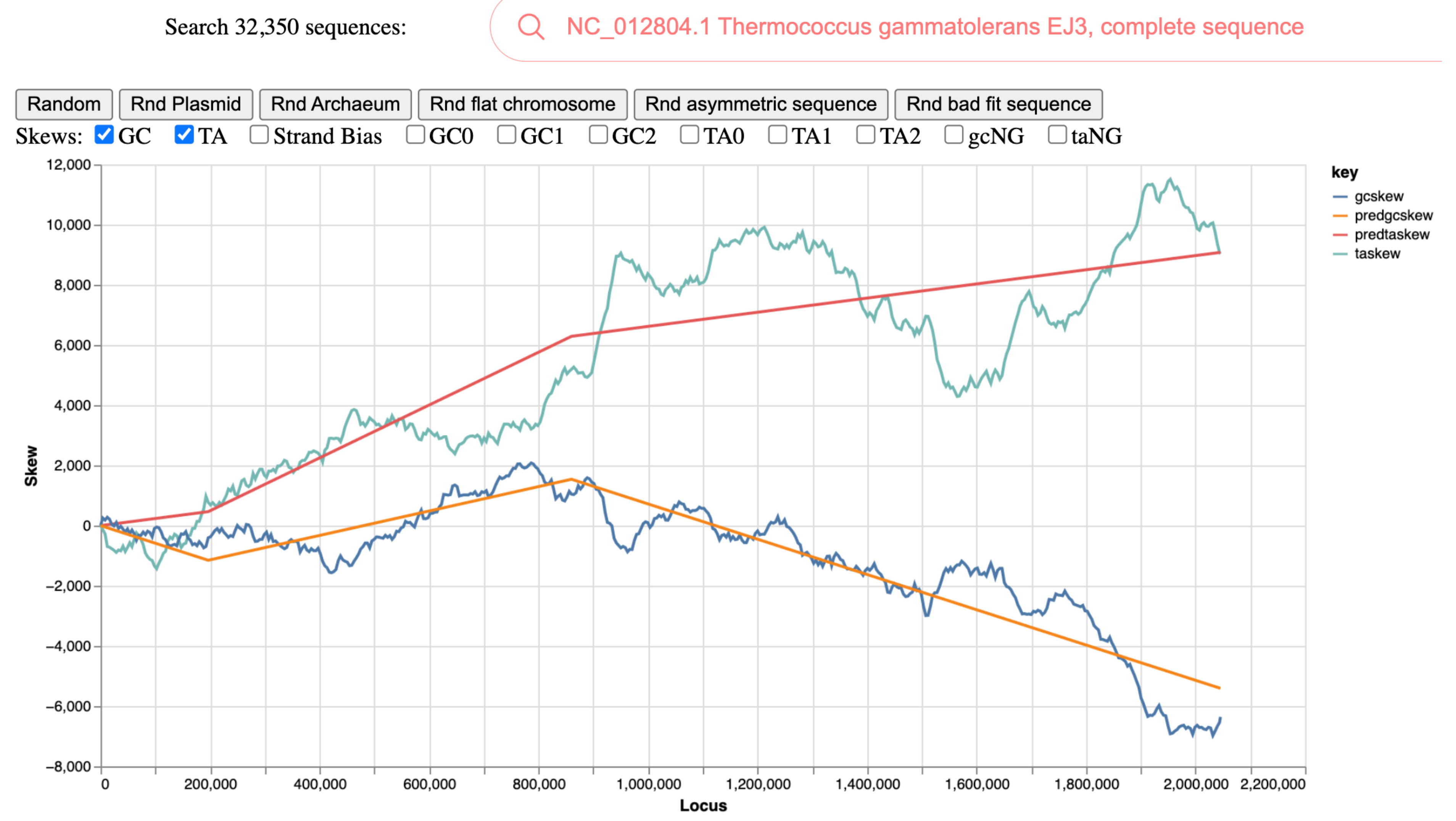
Skews: ☑ GC ☑ TA ☐ Strand Bias ☐ GC0 ☐ GC1 ☐ GC2 ☐ TA0 ☐ TA1 ☐ TA2 ☐ gcNG ☐ taNG



Search 32,350 sequences: NC_012660.1 Pseudomonas fluorescens SBW25, complete sequence

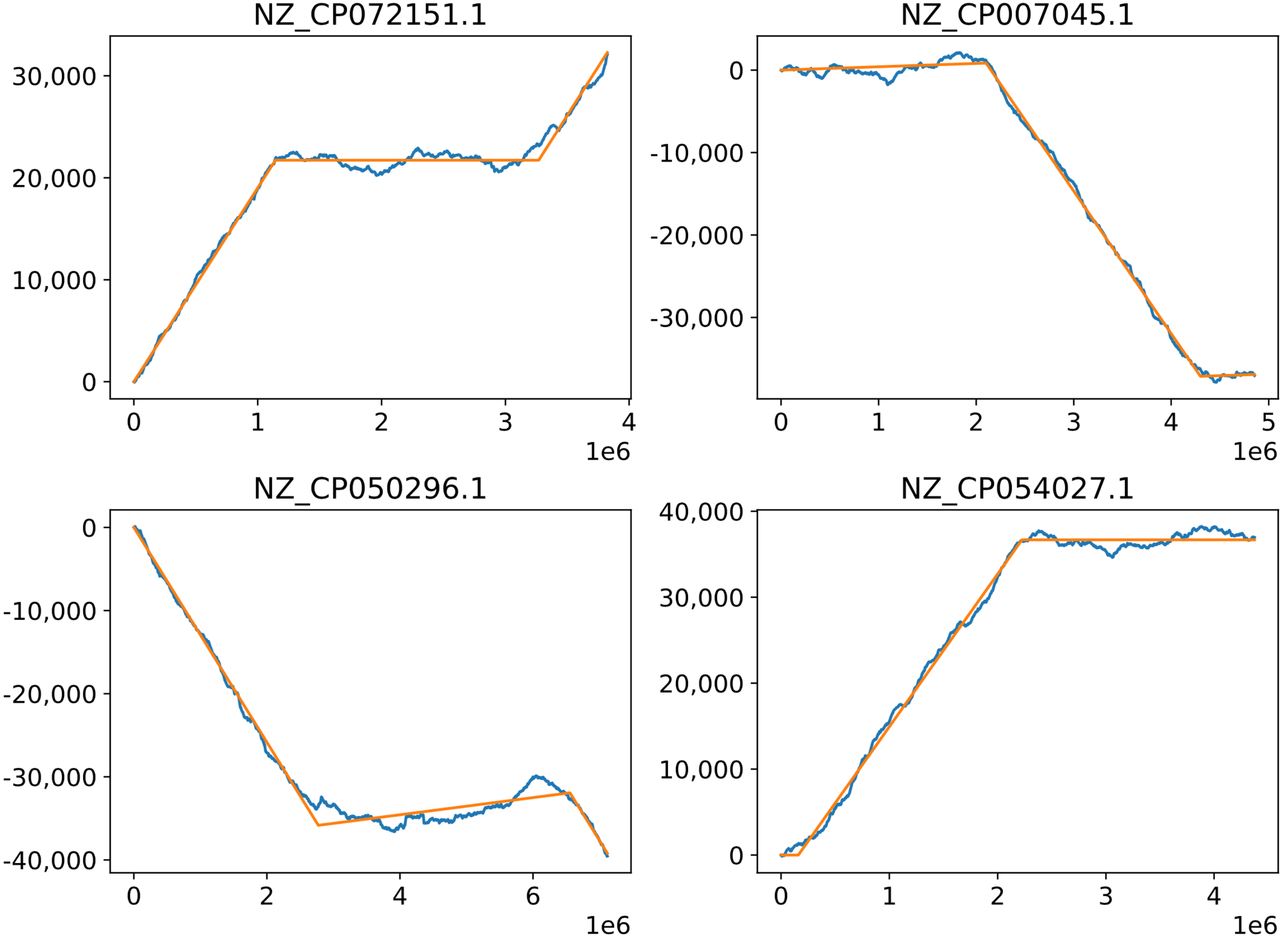
Skews: ☑ GC ☑ TA ☐ Strand Bias ☐ GC0 ☐ GC1 ☐ GC2 ☐ TA0 ☐ TA1 ☐ TA2 ☐ gcNG ☐ taNG

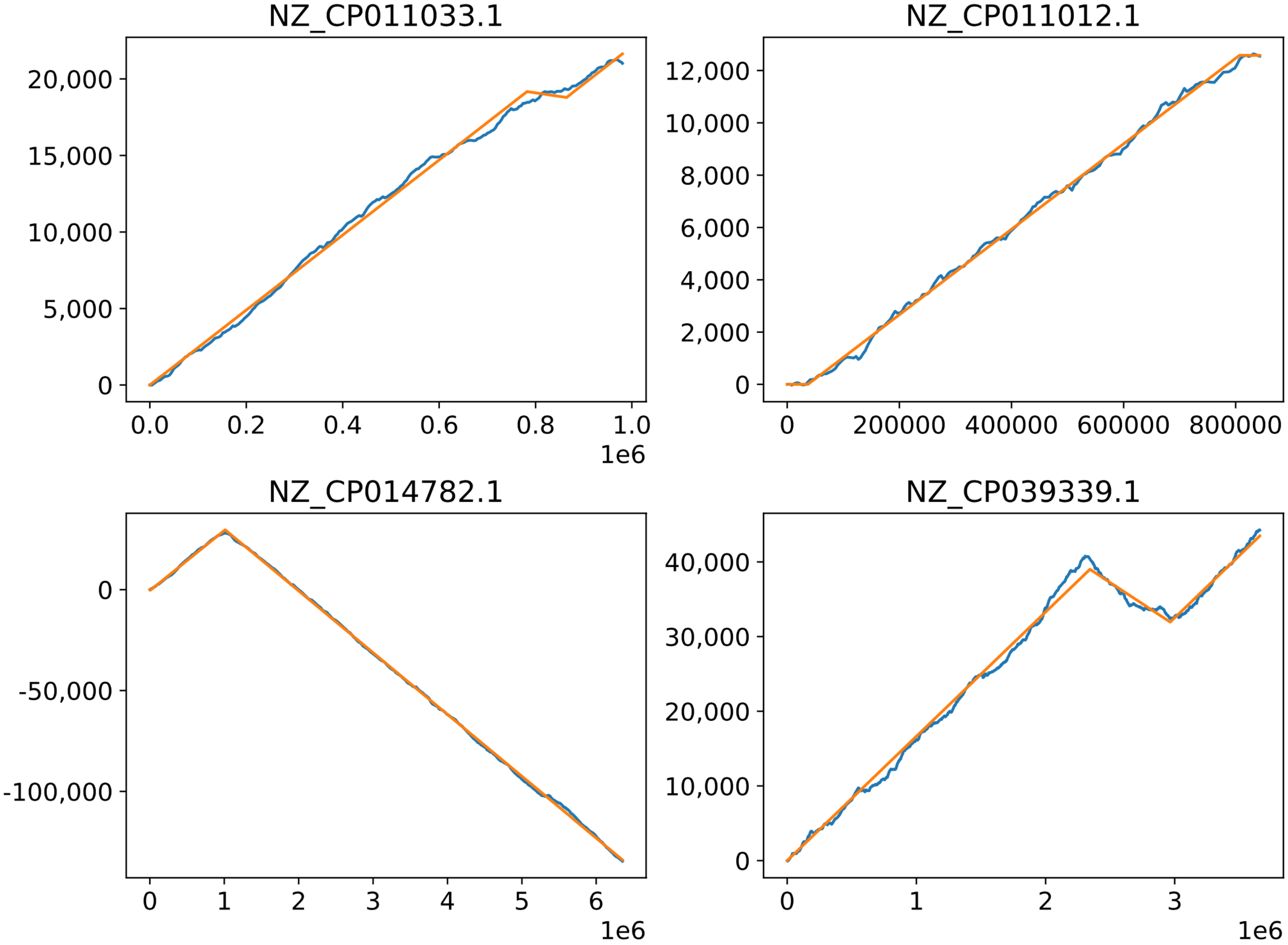

Proteobacteria
Terrabacteria group
FCB group
PVC group
Euryarchaeota
0.10
0.05
0.00
−0.05
−0.10
alpha1 of TA skew
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
alpha1 of GC skew

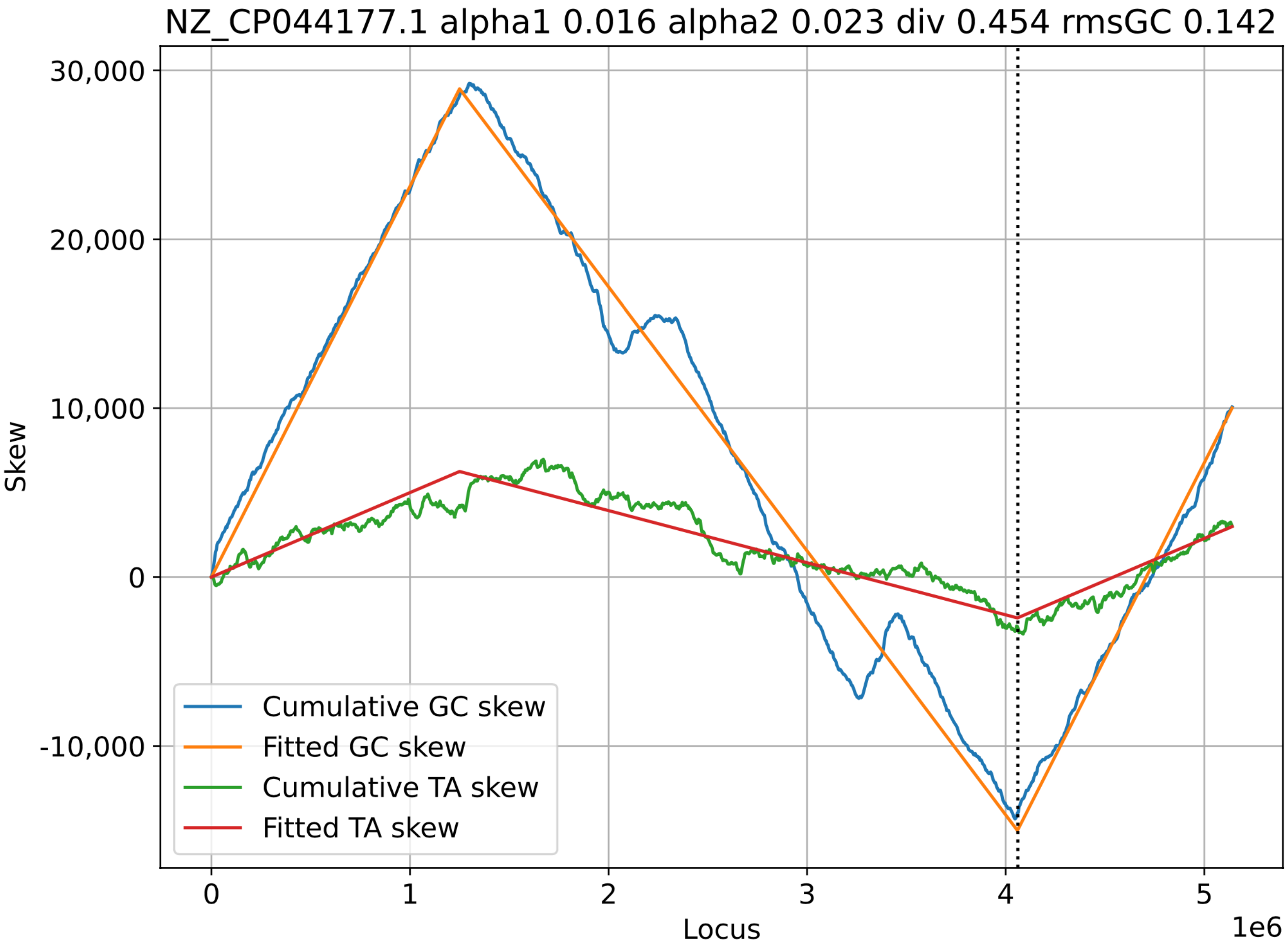


Search 32,350 sequences: NC_012804.1 Thermococcus gammatolerans EJ3, complete sequence

Skews: ☑ GC ☑ TA ☐ Strand Bias ☐ GC0 ☐ GC1 ☐ GC2 ☐ TA0 ☐ TA1 ☐ TA2 ☐ gcNG ☐ taNG


NZ_CP072151.1
30,000
20,000
10,000
0
0
1
2
3
4
1e6
NZ_CP007045.1
0
-10,000
-20,000
-30,000
0
1
2
3
4
5
1e6
NZ_CP050296.1
0
-10,000
-20,000
-30,000
-40,000
0
2
4
6
1e6
NZ_CP054027.1
40,000
30,000
20,000
10,000
0
0
1
2
3
4
1e6

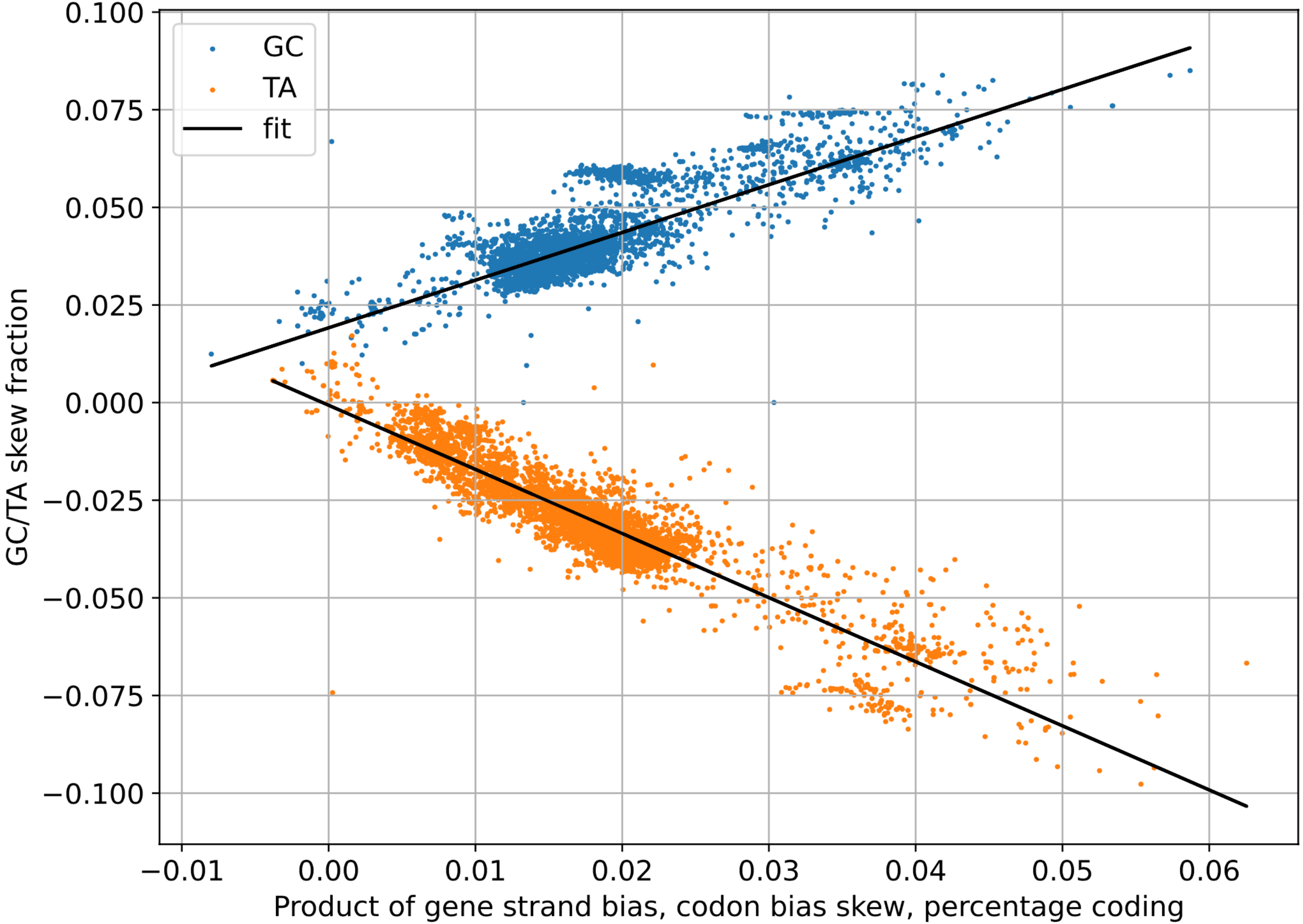

NZ_CP011033.1
20,000
15,000
10,000
5,000
0
0.0
0.2
0.4
0.6
0.8
1.0
1e6
NZ_CP011012.1
12,000
10,000
8,000
6,000
4,000
2,000
0
0
200000
400000
600000
800000
NZ_CP014782.1
0
-50,000
-100,000
0
1
2
3
4
5
6
1e6
NZ_CP039339.1
40,000
30,000
20,000
10,000
0
0
1
2
3
1e6



NZ_CP044177.1 alpha1 0.016 alpha2 0.023 div 0.454 rmsGC 0.142
Skew
30,000
20,000
10,000
0
-10,000
Locus
0
1
2
3
4
5
1e6
Cumulative GC skew
Fitted GC skew
Cumulative TA skew
Fitted TA skew

Possible uses

- https://skewdb.org/view is an easy way to check out your favorite organism
  - Autocomplete!
- Is my bacterium weird? Is the chromosome assembly plausible? Is my organism more archaeal or more eubacterial?
- Can I discover why GC skew happens like this?
  - And learn something about DNA replication!


GC
TA
fit
GC/TA skew fraction
0.100
0.075
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
−0.01
0.00
0.01
0.02
0.03
0.04
0.05
0.06
Product of gene strand bias, codon bias skew, percentage coding

Bert Hubert
bert@hubertnet.nl / @bert_hu_bert
https://berthub.eu/ / https://skewdb.org

Disclosure of speaker's interests

No (potential) conflict of interests	Nothing!
Relations that could be relevant for the meeting	**Company names**
None!	None!