



22BIO201: Intelligence of Biological Systems - 1

Complications in *oriC* Predictions

Dr. Manjusha Nair M
Amrita School of Computing, Amritapuri

Email : manjushanair@am.amrita.edu
Contact No: 9447745519

Contents

- Complications in *oriC* Predictions
- The skew diagram of *Thermotoga petrophila*
- Moral of the Chapter
- Open Problems

Complications in *oriC* Predictions

We have considered three genomes and found three different hypothesized 9-mers encoding DnaA boxes

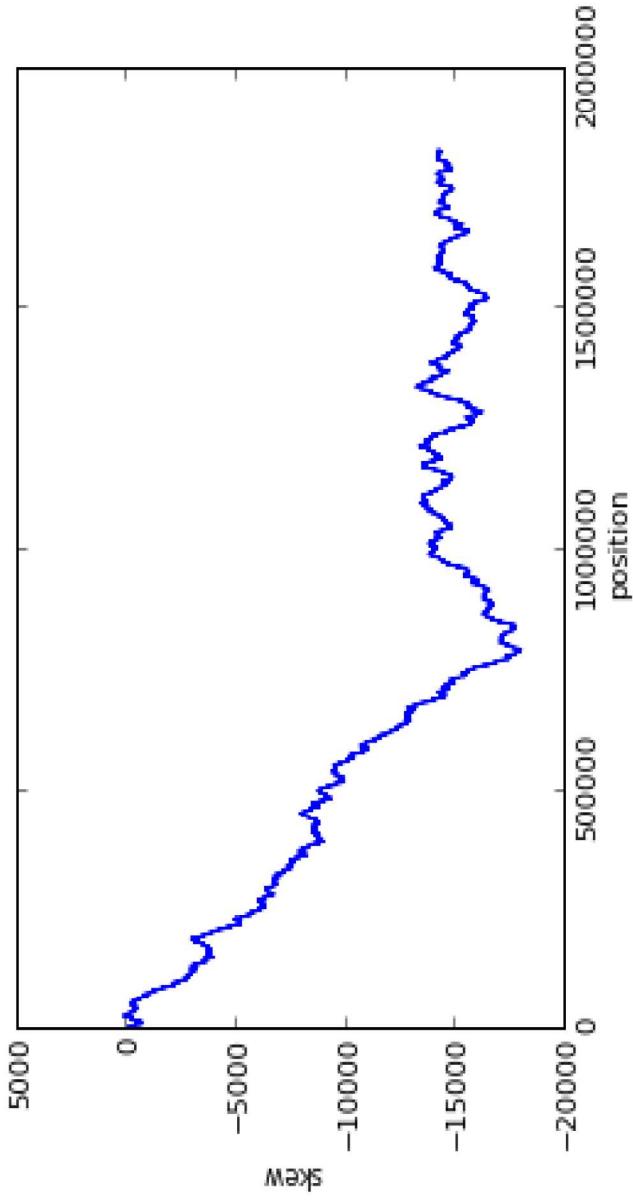
- Vibrio cholera : ATGATCAAG
- Thermotoga petrophila : CCTACCACC
- E. coli:TTATCCACA

Finding *oriC* is often more complex than in the three examples we considered

Complications in *oriC* Predictions

- Some bacteria have fewer *DnaA* boxes
 - making it difficult to identify them.
- *terC* is often not located directly opposite to *oriC*
 - resulting in reverse and forward half-strands having substantially different lengths.
- The position of the skew minimum is often only a rough indicator of *oriC* position
 - which forces researchers to expand their windows when searching for *DnaA* boxes, bringing in extraneous repeated substrings
- The skew diagram is often more complex than in the case of *E. coli*.
 - for example, the skew diagram for *Thermotoga petrophila* is more complex

The skew diagram of *Thermotoga petrophila*



- Skew diagram for *Thermotoga petrophila* is complex
 - oriC for this genome has not even been experimentally verified.

There is a chance that the region predicted as the oriC region for *Thermotoga petrophila* (or even for *Vibrio cholerae*) is actually incorrect!

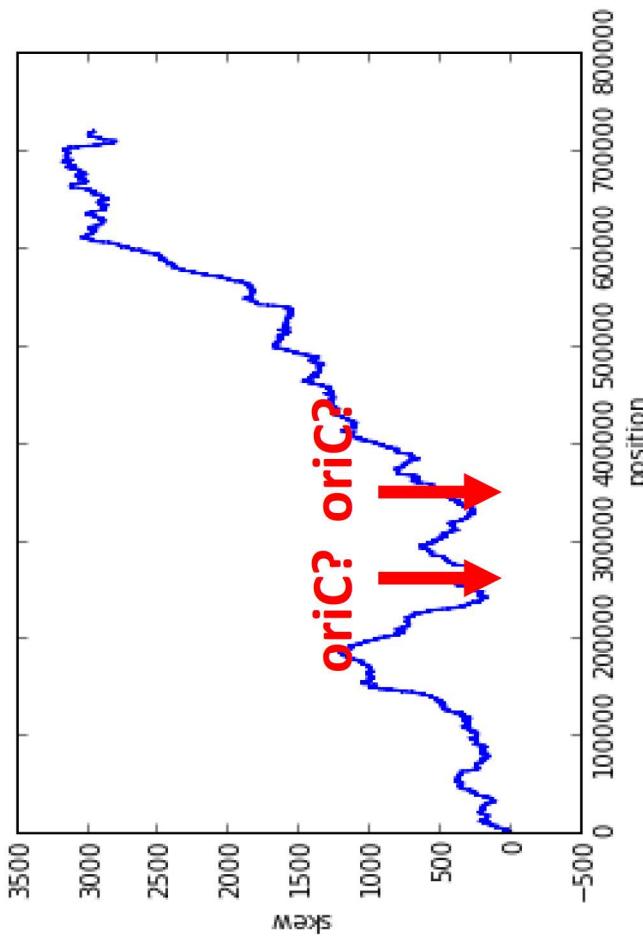
Moral of the Chapter

- There are many different types of hidden messages in genomes
 - and most of them have nothing to do with replication
 - One example is the regulatory DNA motifs responsible for gene expression
 - The important lesson is that existing approaches to *oriC* prediction remain imperfect and sometimes inconclusive.
 - However, even providing biologists with a small collection of 9-mers as candidate DnaA boxes is a great aid as long as one of these 9-mers is correct.
- Even though computational predictions can be powerful, bioinformaticians should collaborate with biologists to verify their computational predictions.

Open Problems

- Finding Multiple Origins of Replication in a Bacterial Genome.
- Finding *oriC* in Archaea
- Finding *oriC* in Yeast

Finding Multiple Origins of Replication in a Bacterial Genome



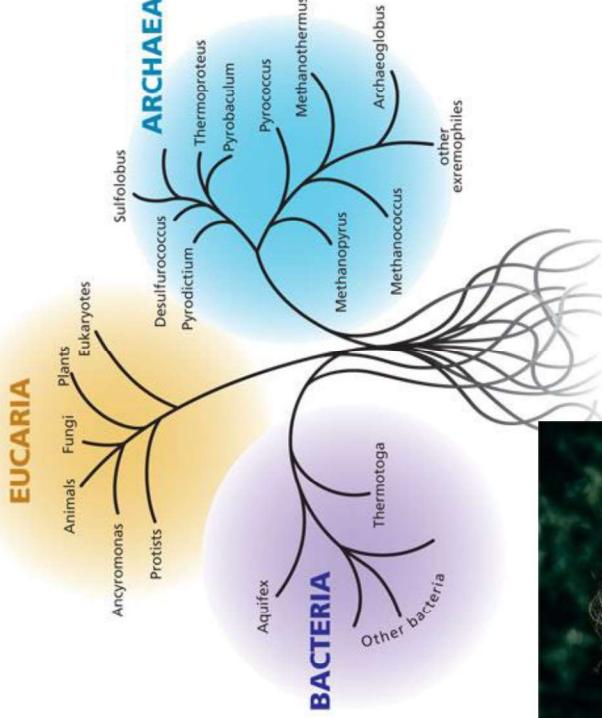
- Biologists long believed that each bacterial chromosome has a single replication origin.
- Xia (2012) argued that some bacteria may have multiple replication origins.

Skew diagram of *Wigglesworthia glossinidia*

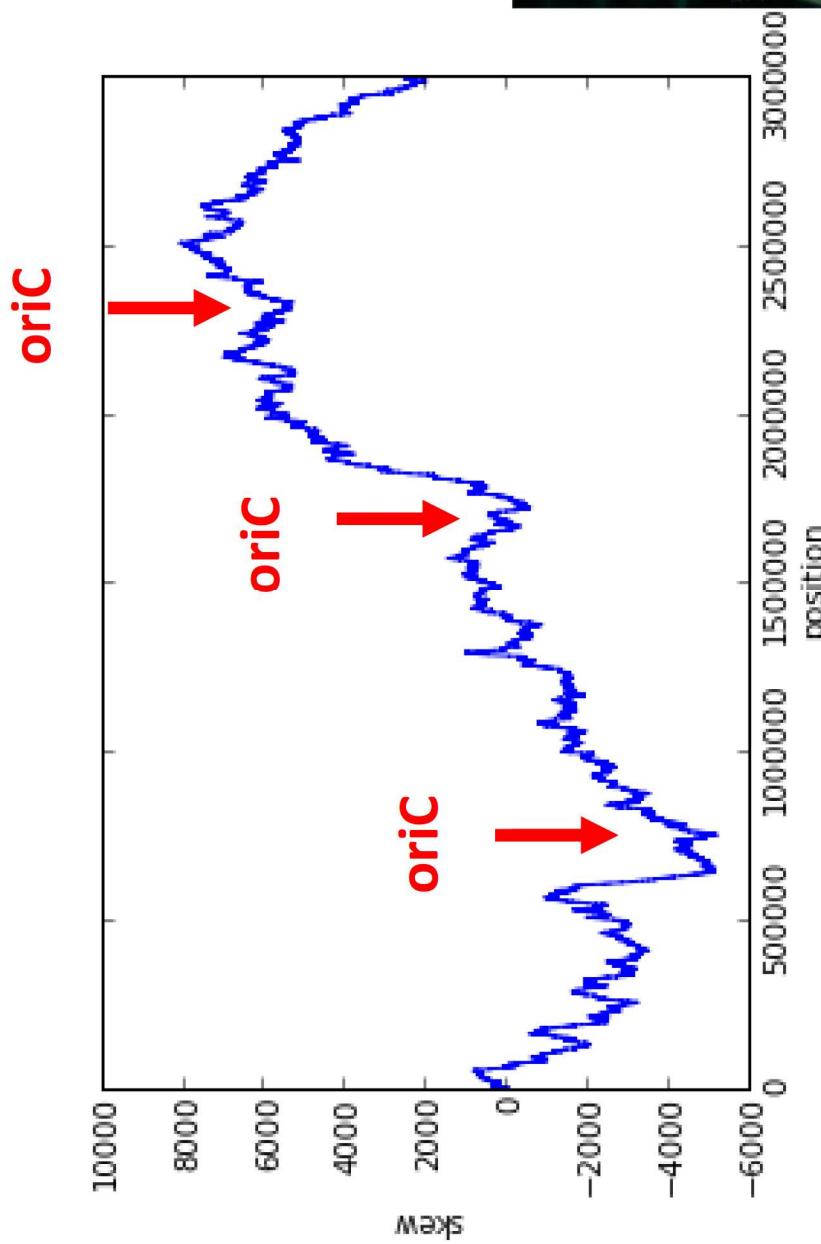


Open Problem: Can you confirm or refute the Xia conjecture that this bacterial genome indeed has multiple replication origins?

Finding oriC in Archaea



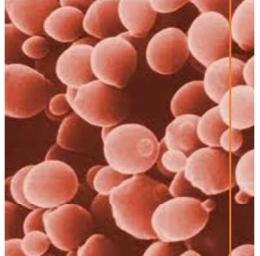
Project Director
Mikhail Gelfand



The skew diagram for *Sulfolobus salfataricus*

Open Problem: Archaea do have multiple origins of replication (3 in *Sulfolobus salfataricus*) but there is no algorithm and software tool yet to predict them reliably – can you develop it?

Finding *oriC* in Yeast



If you feel that finding bacterial replication origins is difficult, wait until you analyze replication origins in yeast or humans.



Open Problem: Yeast genomes have hundreds of origins of replication, but there is no software tool to predict them reliably – can you develop such a tool?

Project Director
Uri Keich

Extra Reading

- Recent development of Ori-Finder system and DoriC database for microbial replication origins :
<https://academic.oup.com/bib/article/20/4/1114/4796864>
- Ori-Finder 2022: A Comprehensive Web Server for Prediction and Analysis of Bacterial Replication Origins:
<https://www.sciencedirect.com/science/article/pii/S1672022922001309>

22BIO201: Intelligence of Biological Systems - 1

