



Introduction to Bioinformatics

Instructor : Sakhaa Alsaedi

Sakhaa.Alsaedi@kaust.edu.sa

Day 4: De novo Genome Assembly
12th January 2024



Check Galaxy Slides 😊



<https://training.galaxyproject.org/training-material/topics/sequence-analysis/tutorials/quality-control/slides.html#1>



Hands-on and Practical Part



GitHub



جامعة الملك عبد الله
للعلوم والتقنية
King Abdullah University of
Science and Technology



Introduction to Bioinformatics

Application: De novo Genome Assembly

Part 2: De Bruijn graph



De Bruijn graph



“tomorrow and tomorrow and tomorrow”



De Bruijn graph



“tomorrow and tomorrow and tomorrow”

tomorrow

and



De Bruijn graph

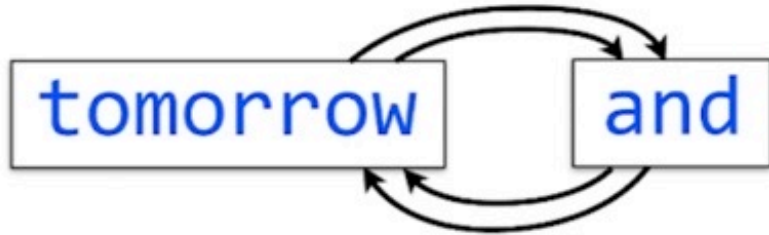
“tomorrow and tomorrow and tomorrow”



De Bruijn graph



“tomorrow and tomorrow and tomorrow”





De Bruijn graph

genome: AAABBBBA



De Bruijn graph



genome: AAABBBBA

3-mers: AAA, AAB, ABB, BBB, BBB, BBA



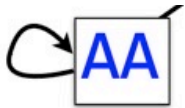


De Bruijn graph

genome: AAABBBBA

3-mers: AAA, AAB, ABB, BBB, BBB, BBA

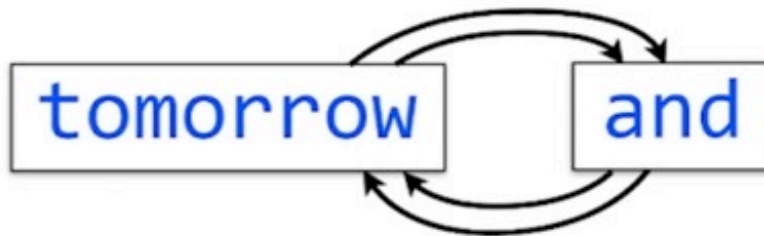
L/R 2-mers: AA, AA





De Bruijn graph

“tomorrow and tomorrow and tomorrow”





De Bruijn graph

genome: AAABBBBA

3-mers: AAA, AAB, ABB, BBB, BBB, BBA

L/R 2-mers: AA, AA

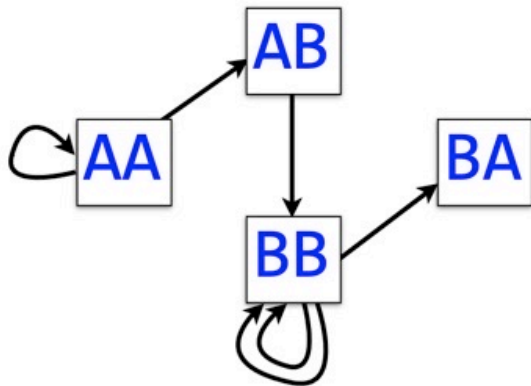
De Bruijn graph



genome: **AAABBBBA**

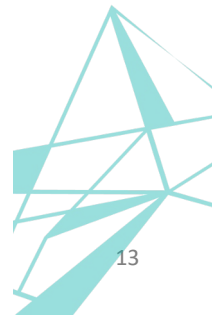
3-mers: **AAA, AAB, ABB, BBB, BBB, BBA**

L/R 2-mers: **AA, AA AA, AB AB, BB BB, BB BB, BB BB, BA**

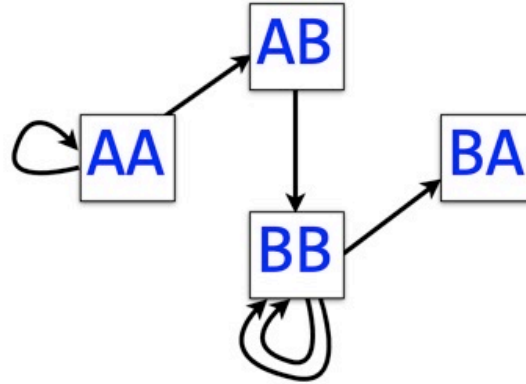


One edge per k -mer

One node per distinct $k-1$ -mer

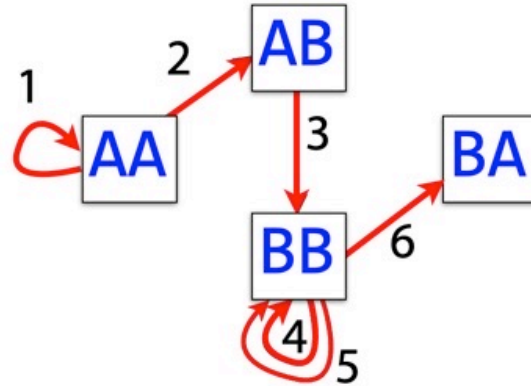


De Bruijn graph



Walk crossing each edge exactly once gives a reconstruction of the genome

De Bruijn graph



AAABBBBA

Walk crossing each edge exactly once gives a reconstruction of the genome. This is an *Eulerian walk*.

Hands-on and Practical Part



GitHub



جامعة الملك عبد الله
للعلوم والتقنية
King Abdullah University of
Science and Technology



Done with Day 4, Heyyyyyy!

Thank You !



جامعة الملك عبد الله
للعلوم والتقنية
King Abdullah University of
Science and Technology

