

n-gram analysis of biological sequences in R

Michał Burdukiewicz¹, Piotr Sobczyk², Małgorzata Kotulska³, Paweł Mackiewicz¹

¹University of Wrocław, Department of Genomics,

²Wrocław University of Science and Technology, Faculty of Pure and Applied Mathematics,

³Wrocław University of Science and Technology, Department of Biomedical Engineering

Intro

n-grams (k-tuples) are vectors of n characters derived from input sequence(s).