

A beginners guide to phylogenetic analysis next-generation-sequencing data

Thaabet Parker^{1,2}

¹Foundational Biodiversity Research, The Compton Herbarium, South African National Biodiversity Institute, Private Bag X7, Newlands, Cape Town 7700, South Africa

²Department of Biological Sciences and Bolus Herbarium, University of Cape Town, Private Bag X3, 7701, Rondebosch, South Africa

1 Getting started

1.1 Terminology

K-mers A k-mer refers to all possible substrings of length **k** that are contained in a string e.g. sequence **ATGCA** has 4 different sets of k-mers. They are useful for computation purposes, error-correction (rare k-mers are more likely to be caused by sequence errors and improvement of alignment)