

A beginners handbook to species delimitation using next-generation-sequencing data

Thaabiet Parker^{1,2}

¹Department of Biological Sciences and Bolus Herbarium, University of Cape Town, Private Bag X3, 7701, Rondebosch, South Africa

²Foundational Biodiversity Research, The Compton Herbarium, South African National Biodiversity Institute, Private Bag X7, Newlands, Cape Town 7700, South Africa

26 October 2023

1 Introduction

2 How you get your data

2.1 Illumina sequencing requirements

2.2 Illumina sequencing workflow

2.2.1 Library preparation

2.2.2 Sequencing

3 How you analyse your data

3.1 Data formats

3.2 Linux and high performance computing clusters

3.3 Bioinformatic workflow

3.3.1 Quality control reads

3.3.2 Align reads

3.3.3 Call variants

3.3.4 Filter data

3.3.5 Assess evolutionary questions

4 Detailed workflows

4.1 Complete workflows

dDocent

Stacks

FastGBS

4.2 Personal workflows

4.2.1 GBS with no reference genome

4.2.2 GBS with reference genome

5 Glossary