

# An exploration of assembly strategies and quality metrics on the accuracy of the *Knightia excelsa* (rewarewa) genome.

Ann McCartney<sup>1</sup>, Elena Hilario<sup>1</sup>, Seung-Sub Choi<sup>1</sup>, Joseph Guhlin<sup>1</sup>, Jessie Prebble<sup>1</sup>, Gary Houlston<sup>2</sup>, Thomas Buckley<sup>3</sup>, and David Chagné<sup>4</sup>

<sup>1</sup>Affiliation not available

<sup>2</sup>Landcare Research New Zealand

<sup>3</sup>Landcare Research

<sup>4</sup>New Zealand Institute for Plant and Food Research Ltd

January 12, 2021

## Abstract

We used long read sequencing data generated from *Knightia excelsa* R.Br, a nectar producing Proteaceae tree endemic to Aotearoa New Zealand, to explore how sequencing data type, volume and workflows can impact final assembly accuracy and chromosome construction. Establishing a high-quality genome for this species has specific cultural importance to Māori, the indigenous people, as well as commercial importance to honey producers in Aotearoa New Zealand. Assemblies were produced by five long read assemblers using data subsampled based on read lengths, two polishing strategies, and two Hi-C mapping methods. Our results from subsampling the data by read length showed that each assembler tested performed differently depending on the coverage and the read length of the data. Assemblies that used longer read lengths (>30 kb) and lower coverage were the most contiguous, kmer and gene complete. The final genome assembly was constructed into pseudo-chromosomes using all available data assembled with FLYE, polished using Racon/Medaka/Pilon combined, scaffolded using SALSA2 and AllHiC, curated using Juicebox, and validated by synteny with *Macadamia*. We highlighted the importance of developing assembly workflows based on the volume and type of sequencing data and establishing a set of robust quality metrics for generating high quality assemblies. Scaffolding analyses highlighted that problems found in the initial assemblies could not be resolved accurately by utilizing Hi-C data and that scaffolded assemblies were more accurate when the underlying contig assembly was of higher accuracy. These findings provide insight into what is required for future high-quality de-novo assemblies of non-model organisms.

## Hosted file

Rewarewa\_genome\_Final.pdf available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>

## Hosted file

Figure1.pdf available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>

## Hosted file

Figure2.pdf available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>

### Hosted file

Figure3.pdf available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>

### Hosted file

Figure4.pdf available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>

### Hosted file

Figure5.pdf available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>

### Hosted file

Figure6.pdf available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>

### Hosted file

Table1.xlsx available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>

### Hosted file

Table2.xlsx available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>

### Hosted file

Table3.xlsx available at <https://authorea.com/users/389088/articles/503727-an-exploration-of-assembly-strategies-and-quality-metrics-on-the-accuracy-of-the-knightia-excelsa-rewarewa-genome>