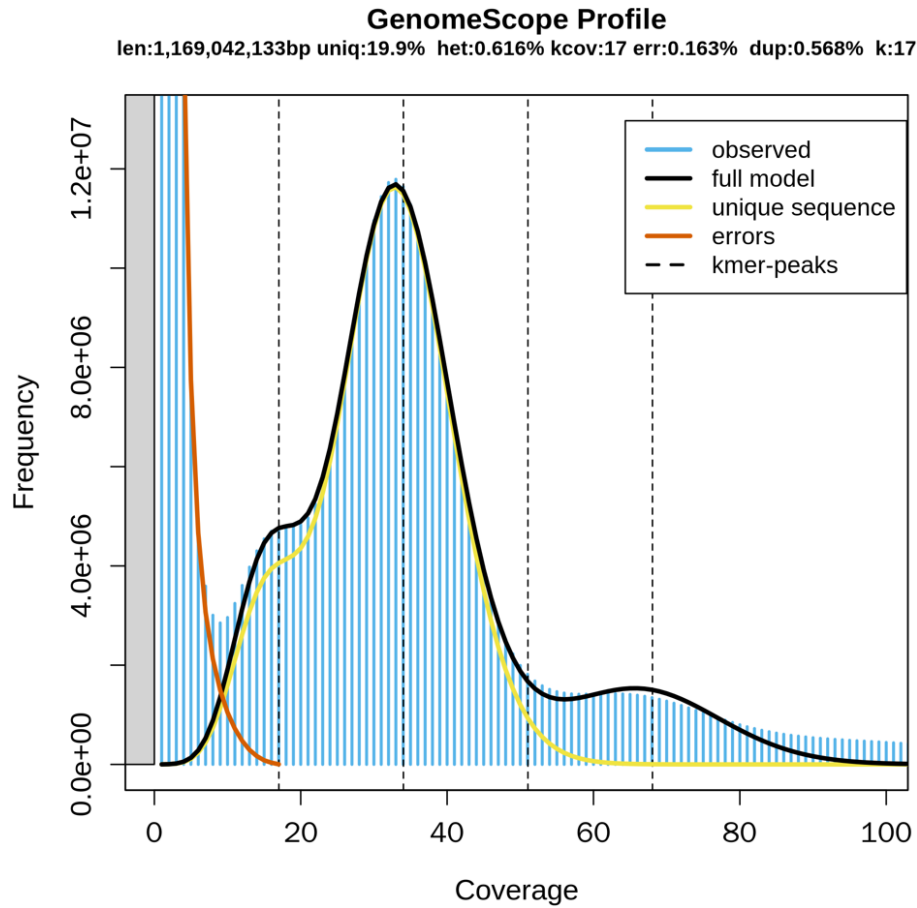




Supplementary Fig. 2 K-mer analysis of *C. monnieri* genome. The blue columns represent the observed k-mer frequency, while the black fitting line represents the remaining k-mers after removing the portion considered to be sequencing errors (the red fitting line). The yellow fitting lines represent k-mer frequency of non-repeating regions of the genome. The red fitting lines represent to low depth k-mers, which are mostly due to sequencing errors. The vertical dashed black line indicates the integral coverage of the predicted k-mer frequency peaks. The x-axis represents sequencing coverage. The y-axis represents the frequency of occurrence for 17 k-mers in the genome.