

Effective identification and annotation of fungal genomes

Jian Liu, Jialiang Sun, Yongzhuang Liu

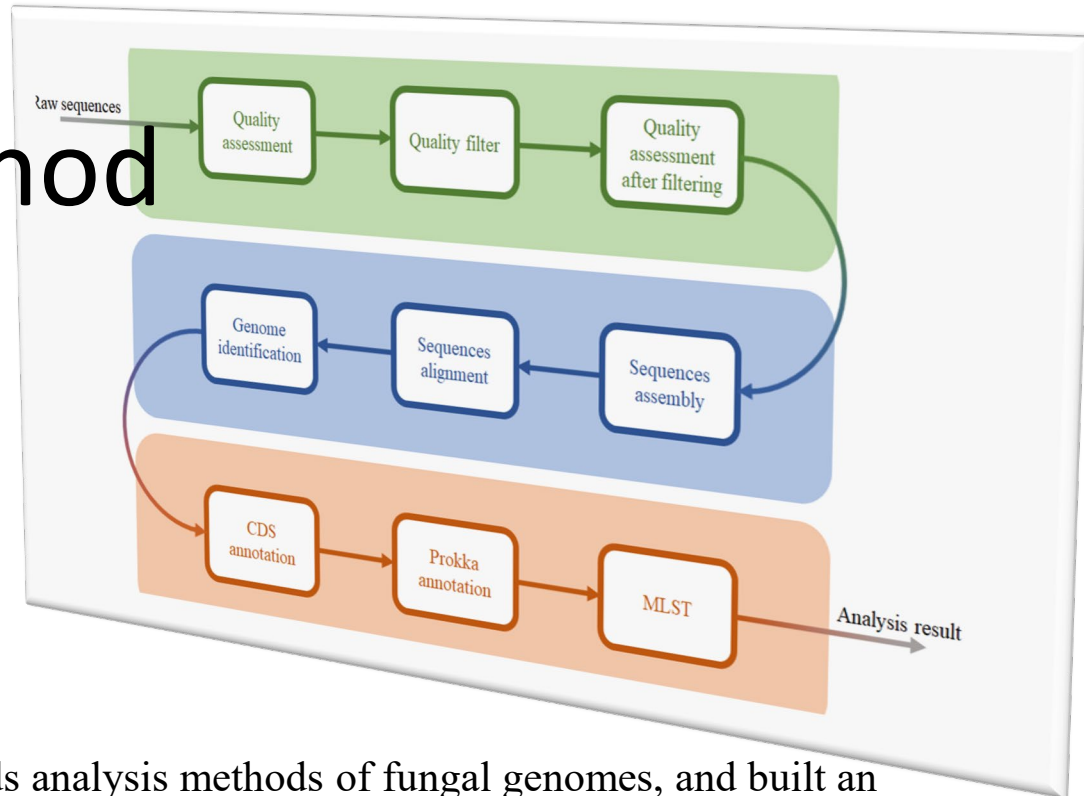
Liu J, Sun JL, Liu YZ. Effective identification and annotation of fungal genomes. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 36(2): 248{260 Mar. 2021. DOI 10.1007/s11390-021-0856-4



Research Objectives

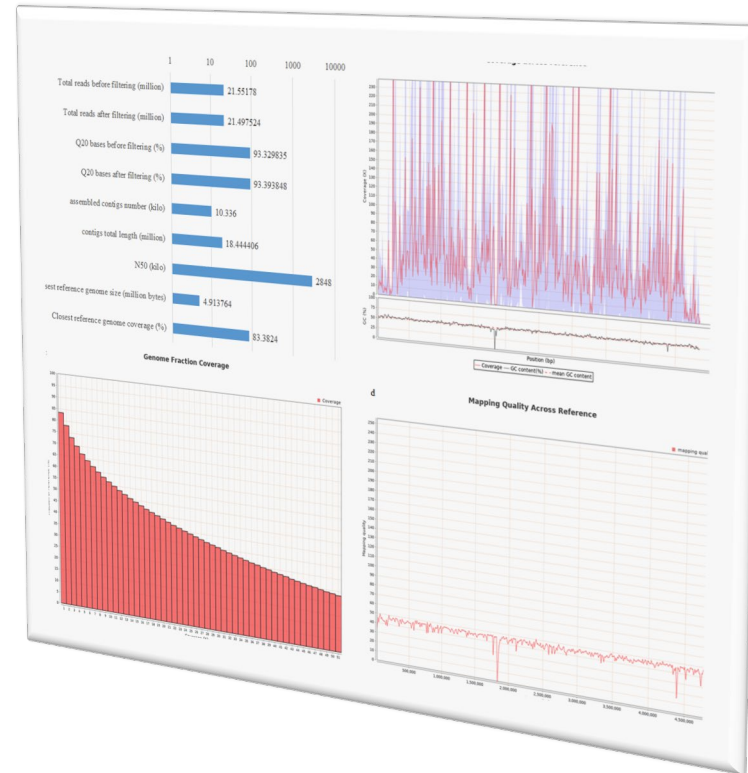
- With the decreasing price of sequencing, a large amount of genomic sequencing data has emerged. Faced with the increasing amount of genomic data, there is still a lack of easy-to-use, efficient and accurate pipelines for sequencing data analysis, especially for efficient identification and annotation analysis of large-scale fungal genomic sequencing data.
- In the study of new strain identification and annotation, the next-generation and the third-generation sequencing data are often taken into consideration for comprehensive analysis. Therefore, an effective analysis pipeline supporting both short reads and long reads analysis plays an important role in improving the quality and efficiency of microbial sequencing data analysis.

Research Method



- This work studies the short and long reads analysis methods of fungal genomes, and built an automated bioinformatics pipeline called PFGI to support the rapid identification and annotation of fungal genome sequencing data. PFGI firstly select short reads or long reads analysis mode, then it perform the pre-processing including the quality control, etc. On this basis, the fungal genome sequencing data identification is completed based on the sequence assembly, sequence alignment and similar reference genome identification. Besides, PFGI provides CDS annotations and also supports the prokka annotations and MLST annotations.

Research Results



- In order to verify the analytical performance of the PFGI, a series of experimental comparisons running on short and long genome sequencing data (e.g., as *aspergillus fumigatus*, *candida albicans*, *saccharomyces cerevisiae* and *verticillium dahlia* in EMBL Nucleotide Sequence Database) are provided Through experimental evaluation, it can see that PFGI has good analysis efficiency and high accuracy, and it can quickly and effectively complete the identification and annotations of short and long fungal genome sequencing data.



Research Conclusions

- In this paper, an effective pipeline called PFGI for efficient identification and annotation analysis of large-scale fungal genome data is provided. PFGI has the ability of handling sequencing data generated from currently popular sequencing platforms such as Illumina, PacBio, and Oxford Nanopore. PFGI also provides annotation functions such as CDS annotations, MLST annotations, etc.
- PFGI could provide easy-to-use, rapid and accurate bioinformatics analysis for biologists, clinicians and other researchers. It can be widely used in industrial microbial strain identification and transformation, clinical diagnosis and treatment and other application services.