PB-40

Pplex: An Available Program to Demultiplex Genome Sequencing Data

Dong Su Yu[1]*, Manjung Kang[1]

[1]National Institute of Agricultural Sciences, RDA, Jeonju 54874, Republic of Korea

[Introduction]
When sequencing genomes of multiple samples with Next-Generation Sequencing such as Illumina shotgun, generally a barcoding technology called 'Multiplexing' is used and then separates multiplexed genome data to individual genomes as 'Demultiplexing'. When demultipliexing genome sequencing data, some bioinformatics' tools such as Ultraplex and fastq-multix are available but they don't results significant performances in sensitivity and accuracy on sequencing error including null bases (N) in barcode sequences in spite of fast running time.

[Program]
We developed a program, a fast Parallel demultiplexing program (Pplex), accurately separating barcoded sequences (reads) including null bases such as 'N' using hash algorithm and rule based modeling with parallel process.

[Results and Discussion]
When comparing performance of Pplex with other program, Ultraplex and fastq-multx, our program showed higher sensitivity (0.94) and accuracy (0.95) than Ultraplex (0.67 and 0.74) and fastq-multx (0.38 and 0.5) in read data set including 'N' bases such as NNATGCNN and NNTAGCNN although Pplex was moderately fast. From these results, we expect that Pplex algorithm will be useful as a bioinformatics' tool that can treat multiplexed sequencing data sets with the various unique barcodes.

[Acknowledgement]
This work was supported by a project grant from "DB development and services for enhancing the utilization of genetic resources", RDA Project No. PJ014226 Rural Development Administration, Republic of Korea.

*Corresponding author: E-mail. dsyu216@korea.kr Tel. +82-63-238-4910