Proceedings of the LXVI SIGA Annual Congress
Bari, 5/8 September, 2023
ISBN: **978-88-944843-4-2**

Poster Communication Abstract – 6.14

A COMPARISON STUDY BETWEEN TORRENT SUITE PIPELINE AND COMMON VARIANT CALLING ALGORITHMS

VETTORI C.***, GAROSI C.**, CONTI S.***, PAFFETTI D.**

*) National Research Council, Institute of Bioscience and Bioresources, Division of Florence, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy
**) University of Florence, Department of Agriculture, Food, Environment and Forestry, Piazzale delle Cascine, 18 – 50144 Florence, Italy
***) Life Technologies Italia part of ThermoFisher Scientific, Fil. Life Technologies Europe BV; c/o Segreen Business Park, Via S. Bovio 3, 20054 Segrate (MI)

Ion Torrent, Single Nucleotide Polymorphisms (SNPs), bioinformatics, plant genomics, Fagus sylvatica L.

Next-generation sequencing (NGS) has revolutionized plant research in many ways, including new high-throughput genotyping methods. Genotyping by sequencing (GBS) has been proven to be a robust and cost-effective method capable of producing thousands or millions of SNPs in a wide range of species. Undoubtedly, the greatest barrier to its wider use is the challenge of data analysis. With the advancement of technology in recent years, different sequencing technologies (Ion Torrent, Illumina, and PacBio) have emerged, resulting in different methods to obtain reads and thus different data types. Frequently applied pipelines often start with the pre-processing of sequence reads, followed by the alignment (mapping) of these reads to a reference sequence. Finally, the identification (calling) of sequence variants is performed based on alignments. Each of these three steps can be carried out by various alternative programs using different algorithms, which influence the accuracy and sensitivity of the resulting variant set. Genomic variants like single nucleotide variants (SNVs) or small insertions/deletions (InDels) can be inferred by variant callers based on sequence read alignment. Popular variant callers like SAMtools/BCFtools, CLC-GWB, FreeBayes, GATK, LoFreq, SNVer, VarDict, and VarScan use a variety of different approaches to call variants.

Consequently, resulting variant sets differ depending on the employed methods (Bayesian, likelihood, or machine learning), which come with strengths and weaknesses concerning the identification of specific variant types. As shown in recent comparative studies, despite the undoubted performance, there are no software that, with equal analysis efficiency, can be used to complete all the tasks to be done for genomic variant analysis. The purpose of this study is to analyse the performance of a direct pipeline, based on Ion Torrent technology, that allows obtaining genomic variant analysis (SNVs, MNVs, InDels) as a direct sequencing output, without the use of external software: Torrent suite software. In this study, we compared the performance of Torrent suite's direct variant calling pipeline with the most commonly used pipelines on target sequencing data from 18 genomic regions of *Fagus sylavtica* L. genome in 7 Italian populations. Our results demonstrated variant calling efficiency comparable to the various workflows that are commonly adopted by the scientific community.