Proceedings of the LXII SIGA Annual Congress
Verona, Italy – 25/28 September, 2018
ISBN **978-88-904570-8-1**

Poster Communication Abstract – 6.13

IDENTIFICATION OF SUITABLE TOOLS TO IDENTIFY GENOMIC VARIANTS IN POLYPLOID SPECIES

FERRIZ A., AIESE CIGLIANO R., SANSEVERINO W.

Sequentia Biotech SL, Campus UAB, Edifici Eureka, Barcelona (Spain)

variant calling, polyploids, bioinformatics, benchmark

Variant discovery is one of the most important aspects to study differences among individuals, and can have great implications in evolutionary discoveries. Next-generation sequencing has made possible to perform variant calling analysis in a great scale and has propitiated the development of many bioinformatics tools along the years. This study is intended to be a benchmark of the current available open-source variant calling software, evaluating precision, recall and f-score of the detected variants and a revision on the ease of use and convenience of each of the programs to detect variants on plant polyploid genomes. The variant callers analyzed in this study include the recently released GATK4, Freebayes, Octopus and VarDict. Not only the output of the different programs varies greatly but also the necessary inputs and the simplicity of each of them. To conduct the study, a tetraploid version from the plant model organisms *Arabidopsis thaliana* was simulated placing randomly distributed variants along the genome consisting of SNP and *indels*. A file containing the golden variants produced by the simulator was then used to evaluate the different metrics of the variant callers. Our preliminary results show that regarding SNPs, Freebayes achieved the highest F-score (0.92) marked by an excellent precision of 0.999. Considering InDels, GATK4 obtained the highest F-score (0.71) with a balance between sensitivity and recall.