HiSpike method for high-throughput cost effective sequencing of the SARS-CoV-2 spike gene

Fass Ephraim[1], Zizelski Valenci Gal[1], Rubinstein Mor[1], Freidlin Paul J[1], Rosencwaig Shira[1], Kutikov Inna[1], Werner Robert[1], Ben-Tovim Nofar[1], Bucris Efrat[2], Erster Oran[2], Zuckerman Neta S[2], Mor Orna[2,3], Mendelson Ella[2,3], Dveyrin Zeev[1], Rorman Efrat[1], and Nissan Israel[1,*]

[1]National Public Health Laboratory, Public Health Services, Ministry of Health, Tel Aviv, Israel
[2]Central Virology Laboratory, Public Health Services, Ministry of Health, Chaim Sheba Medical Center, Tel-Hashomer, Ramat Gan, Israel
[3]School of Public Health, Sackler faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[*]Corresponding author: israel.nissan@moh.gov.il

ABSTRACT

Emerging of SARS-CoV-2 spike gene variants jeopardize global efforts to produce immunity and reduce morbidity and mortality. These challenges require effective real-time genomic surveillance solutions that the medical community can quickly adopt. The SARS-CoV-2 spike protein mediates host receptor recognition and entry into the cell and, therefore, it is most susceptible to generation of variants with increased transmissibility and pathogenicity. The spike protein is also the primary target of neutralizing antibodies in COVID-19 patients and the most common antigen for induction of effective vaccine immunity. Currently, SARS-CoV-2 sequencing methods are labor intensive and expensive. Consequently, most SARS-CoV-2 strains are sequenced in a few developed countries and rarely in developing regions. This poses the risk that dangerous variants will emerge undetected. In this work, we present HiSpike, a method for high-throughput cost-effective targeted next generation sequencing of the spike gene.

Figure 1

Schematic representation of the major steps in the gold standard ARTIC method (A) and in our new Hispike method (B). The cost per sample for library preparation using ARTIC is 54$ versus 1.7$ using HiSpike. Notably, the spike gene represents less than 13% of the viral genome, allowing running more samples on the same Illumina flow cell.

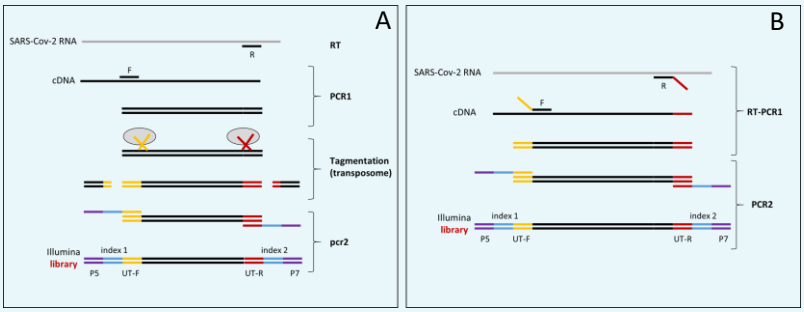


Figure 2

Comparison between ARTIC and HiSpike methods. Spike gene sequences of 70 samples (with >90% spike sequence coverage in both ARTIC and HiSpike methods) were compared using a heat-map. Grey shades between white and black indicate 0 to 10 SNPs respectively. HiSpike sequences were aligned based on an hierarchical clustering tree (shown to the right) vertically and the ARTIC sequences of the same samples were aligned horizontally. Pairs of the same sample are represented in the diagonal line (outlined in red) indicating perfect match between ARTIC to HiSpike results.

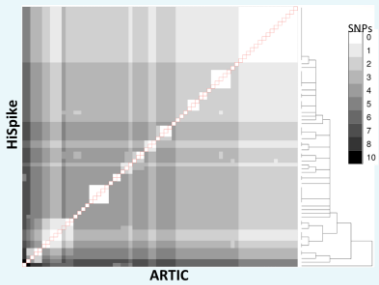
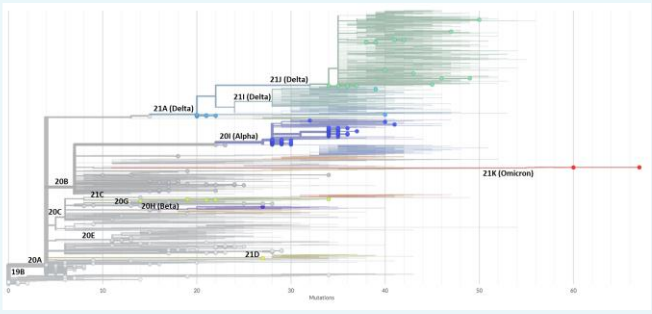


Figure 3

Phylogenetic tree based on HiSpike sequencing of 441 positive SARS-CoV-2 samples, including various Delta clades (21A, 21I, 21J) and the emerging Omicron variants (21K, Red). Sequences generated by HiSpike method were uploaded to Nextclade platform. SARS-CoV-2 clades are illustrated by different colors and locations on the rectangular tree. HiSpike sequences are represented by the full circles on a background of the Nextclade's global clade tree.



Conclusions

The HiSpike method is able to sequence all tested SARS-CoV-2 variants, including Delta and the emerging Omicron. It is based on simple three-step protocol, can be completed in less than 30 hours at a fraction of the cost relative to gold standard ARTIC and Illumina COVIDseq methods. HiSpike method was proven valid, and has identified multiple spike variants from real-time field samples. HiSpike provides affordable sequencing options to help laboratories conserve resources for widespread monitoring of emerging and circulating variants through spike gene sequencing.