Benchmarking Oxford Nanopore Read Alignment-Based Structural Variant Detection Tools in Crop Plant Genomes

Gözde Yildiz, Silvia F. Zanini, Nazanin P Afsharyan, Christian Obermeier, Rod J Snowdon, and Agnieszka A. Golicz*

Department of Plant Breeding, Justus Liebig University Giessen, Giessen, Germany

*Correspondence: Agnieszka.Golicz@agrar.uni-giessen.de

ABSTRACT

Structural variations (SVs) are larger polymorphisms (>50 bp in length), which consist of insertions, deletions, inversions, duplications, and translocations. They can have a strong impact on agronomical traits and play an important role in environmental adaptation. The development of long-read sequencing technologies, including Oxford Nanopore, allows for comprehensive SV discovery and characterization even in complex polyploid crop genomes. However, many of the SV discovery pipeline benchmarks do not include complex plant genome datasets. In this study, we benchmarked popular long-read alignment-based SV detection tools for crop plant genomes. We used real and simulated Oxford Nanopore reads for two crops, allotetraploid *Brassica napus* (oilseed rape) and diploid *Solanum lycopersicum* (tomato), and evaluated several read aligners and SV callers across 5×, 10×, and 20× coverages typically used in re-sequencing studies. Our benchmarks provide a useful guide for designing Oxford Nanopore re-sequencing projects and SV discovery pipelines for crop plants.

Key words: Structural variants; benchmarking; long reads; Oxford Nanopore Technologies; crops; plants

1. INTRODUCTION

Structural variations (SVs) are a major type of polymorphisms, which consist of insertions, deletions, inversions, duplications, and translocations. SVs are larger polymorphisms (>50 bp) compared with single nucleotide polymorphisms (SNPs) and small indels (insertions and deletions). Copy number variations (CNVs) and presence/absence variations (PAVs) occur due to these genomic polymorphisms (Alkan et al., 2011; Sedlazeck et al., 2018a). Insertions and deletions can have a strong effect on crop traits and have been shown to play a role in domestication and environmental adaptation (Gill et al., 2021; Tao et al., 2019; Yildiz et al., 2022; Zanini et al., 2022; Żmieńko et al., 2014). Until recently, the lack of high-quality reference assemblies and the complex nature of often large, polyploid genomes made comprehensive SVs exploration challenging in crop genomic research (Meyers and Levin, 2006; Yuan et al., 2021).

Development of long-read sequencing technologies such as Oxford Nanopore Technologies (ONT) (Jain et al., 2016) and Pacific Bioscience (PacBio) (Roberts et al., 2013) provided new opportunities for comprehensive SV discovery in crop plants. The sequencing accuracy of these technologies is continuously improving. Currently, PacBio HiFi consensus reads exceed 99% accuracy (Wenger et al., 2019) while ONT R10.3 raw reads accuracy exceeds 95% (Delahaye and Nicolas, 2021). The reduction in error rates facilitates downstream applications, including the production of high-quality genome assemblies, and SV detection. ONT sequencing in particular is being adopted in crop plant research for large scale re-sequencing projects of tens to hundreds of individuals (Alonge et al., 2020; Chawla et al., 2021; Lemay et al., 2022; Vollrath et al., 2021; Zhang et al., 2022). Despite the constant decrease in sequencing error rate, long-read technologies require specialized computational approaches to take advantage of them efficiently.

The two main approaches for SV discovery are *de novo* assembly-based and read alignment-based. *De novo* assembly-based approaches assemble reads into longer contigs and identify SVs by aligning assemblies (Wenger et al., 2019). Read alignment-based approaches directly align reads to reference genomes to discover SVs. *De novo* assembly-based methods perform better at finding larger variants (tens to hundreds of kbp long; exceeding the length of individual reads) but require sufficient amount of data to produce high-quality assemblies, which leads to substantial increase in cost of the experiments for larger crop genomes. However, read alignment-based approaches can perform well even at modest sequencing depths of 5× to 10× and use less computational resources, but the discovered SVs are limited to differences with the reference genome which makes this approach more suitable for larger re-sequencing projects (Coster et al., 2021). Several algorithms were developed for SV discovery from long-reads including Sniffles (Sedlazeck et al., 2018b), NanoVar (Tham et al., 2019), SVIM (Heller and Vingron, 2019), cuteSV (Jiang et al., 2020), and dysgu (Cleal and Baird, 2022), which have been comprehensively reviewed recently (Mahmoud et al., 2019; Yuan et al., 2021). Additionally, several long-read aligners are available such as minimap2 (Li, 2018), NGMLR (Sedlazeck et al., 2018a), Vulcan (Fu et al., 2021), and lra (Ren and Chaisson, 2021). Considering the continued development and improvement in read-alignment and SV detection algorithms and multitude of their possible combinations, their combined performances in SV detection demand realistic and up-to-date benchmarks to guide the selection of SV discovery tools.

In this study, we hypothesized that certain combination(s) of read aligners and SV discovery software will have superior performance in datasets representing complex crop genomes. We used real and simulated ONT reads for two crop plant genomes and evaluated several mappers and SV callers across coverages including 5×, 10×, and 20× typically utilized in re-sequencing studies. We chose to perform benchmarking on allotetraploid *Brassica napus* (oilseed rape) and diploid *Solanum lycopersicum* (tomato) as these two species represent different ploidy, have different SV profiles, and were already studied using Oxford Nanopore Technology. Our benchmarks provide a useful guide for researchers designing Oxford Nanopore re-sequencing projects and those designing SV discovery pipelines.

2. MATERIALS AND METHODS

2.1 Read Aligners, SV Callers, and Benchmarking Datasets

The SV callers included in the study were selected using several criteria: (1) citation count (used as a proxy for popularity in the research community), (2) publication date and maintenance status (excluding older tools that were no longer maintained), (3) ability to detect both insertion and deletion SVs from ONT data. The benchmarking approach involved four long-read aligners, including minimap2 (Li, 2018), NGMLR (Sedlazeck et al., 2018a), lra (Ren and Chaisson, 2021), and Vulcan (Fu et al., 2021) as well as five SV calling software namely Sniffles (v2) (Sedlazeck et al., 2018b), NanoVar (Tham et al., 2019), SVIM (Heller and Vingron, 2019), cuteSV (Jiang et al., 2020), and dysgu (Cleal and Baird, 2022). All aligners and SV caller versions are provided in detail in (**Table S1**). Three simulated datasets (Sim_ONT_Bn1, Sim_ONT_Bn2, and Sim_ONT_Sl) and publicly available data, for *B. napus* and *S. lycopersicum* genomes, were used. The real-world datasets for whole genome Nanopore sequencing of *B. napus* cv. King 10 (accession number: SRR15731030) (Vollrath et al., 2021) and *S. lycopersicum* cv. M82 (accession number: SRR16966224) (Alonge et al., 2021) were downloaded from NCBI Sequencing Read Archive. The ONT reads were randomly subsampled to 5×, 10×, and 20× coverages using Rasusa (Hall, 2022) to test the effect of sequencing depth on SV discovery.

2.2 Simulated Dataset Generation

For three simulated datasets (workflow for all simulations is presented in (**Figure S1**), new haplotypes including SVs were generated, and synthetic ONT reads were simulated using VISOR v1.1 (Bolognini et al., 2020). For simulation one (Sim_ONT_Bn1) 20,000 genomic intervals (mean: 750 bp, SD: 500 bp) were randomly drawn from the *B. napus* genome (Express 617 v1). A subset of 10,000 were denoted as deletions. For the remaining 10,000, denoted as insertions, the genomic start coordinate was retained, while the sequences corresponding to the genomic intervals were extracted, randomly re-assigned to the coordinates, and served as insertion sequences at those coordinates (**Figure S1**).

Simulations two and three, denoted Sim_ONT_Bn2 and Sim_ONT_Sl, were designed to reflect SVs found in real-world datasets. For Sim_ONT_Bn2 the assembled *B. napus* genomes Express 617 v1 (Lee et al., 2020) and Westar (Song et al., 2020) were aligned using minimap2 v2.24. SVs were detected using SVIM-asm v1.0.2 (Heller and Vingron, 2020). To reduce the effect of using minimap2 for benchmarking dataset generation, the SV locations were shifted by a randomly selected number in the (-5000, 5000) interval. This changed the exact SV site while maintaining the realistic distribution of SV sizes and locations along the genome. A random subset of 10,000 insertions and 10,000 deletions was drawn from all SVs to create the benchmarking dataset. SNPs discovered from short reads using bcftools v1.15.1 were also included. The SVs and SNPs were provided to VISOR to generate new haplotypes, which in turn were used for Oxford Nanopore read simulation. Sim_ONT_Sl was generated using the same strategy as for Sim_ONT_Bn2 but

designed to reflect SVs of the *S. lycopersicum* genome. Heinz 1706 (Slycopersicum_691_SL4.0) and M82 (Alonge et al., 2021) assemblies were used for whole genome alignments. Due to smaller number of SVs, a random subset of 2,500 insertions and 2,500 deletions were drawn from all SVs.

To test the effect of sequencing depth on SV discovery, the datasets were simulated at 5×, 10×, and 20× coverage. The simulations provided the objective truth sets, which could be used to calculate SV precision, recall, and combined F1-scores. Precision describes the proportion of correct positive predictions among all positive predictions. It is calculated by dividing the true positives by overall positives. Recall describes the proportion of positive predictions made out of all positive elements in the dataset. It is calculated by dividing true positives by total number of relevant elements. F1-score combines precision and recall by taking their harmonic mean. Its value ranges from 0 to 1. F1-score close to 1 indicates high precision and recall. Using two different strategies for generating simulated datasets will make it possible to minimize analytical bias. If the same combination of tools performed best on all simulated datasets, this will likely reflect true superior performance.

2.3 Comparative Analyses

Express 617 v1 for the *B. napus* (Lee et al., 2020) and Slycopersicum_691_SL4.0 for the *S. lycopersicum* (Hosmani et al., 2019) were used as reference sequences. Simulated datasets and real subsampled reads at each coverage depth were aligned to respective reference genomes. The SV call sets were filtered using the following criteria: (1) number of minimum supporting reads: 5×: 3, 10×: 5, and 20×: 8, (2) SV type: INS or DEL (the most abundant SVs supported by all the benchmarked tools), (3) minimum SV length: 50 bp, (4) SV quality: SVs flagged as "PASS" (5) genotype: homozygous genotype for alternative allele ('1/1'). For simulated data, precision, recall, and F1-scores of the SVs were computed for each combination of coverage depth, read aligner, and SV caller using Truvari v3.0.0 (English et al., 2022). Comparisons between results from the same tool combination across different coverages and different tool combinations across the same coverages were performed using surpyvor v0.8.1 (Jeffares et al., 2017). For real datasets, where no truth sets were available, we focused on within-dataset comparisons and how those compared to the results from simulated data. For each coverage, 20 different read aligner/SV caller combinations were used which led to a total of 60 different combinations for three coverages. All the relevant commands for simulated data generation and SV discovery are available in the **Supplementary Note.**

3. RESULTS

3.1 Selecting the Benchmarking Datasets

We chose to focus on two crop plant species *B. napus* (oilseed rape; genome size ~1.1 Gbp) and *S. lycopersicum* (tomato; genome size ~900 Mbp) because they are both important crops and their structural variation was previously studied using Oxford Nanopore Technologies (Alonge et al., 2020; Chawla et al., 2021). Whole Genome Alignment (WGA)-based SV discovery also suggested that they have quite different SV profiles with 38,666 SVs (Real_WGA_Bn, mean size: 2,068 bp,

median size: 593 bp, 19,450 insertions and 19,216 deletions) discovered for *B. napus* and 7,108 SVs (Real_WGA_Sl, mean size: 3,029 bp, median size: 178 bp, 4,159 insertions and 2,949 deletions) discovered for *S. lycopersicum*.

Two simulated *B. napus* haplotypes (Sim_ONT_Bn1 and Sim_ONT_Bn2) and one simulated *S. lycopersicum* haplotype (Sim_ONT_Sl) were used to generate Oxford Nanopore reads at 5×, 10×, and 20× to test the effect of sequencing depth on SV discovery. The two publicly available real-world datasets, from *B. napus* (38×) and *S. lycopersicum* (68×), were subsampled with the same logic (Real_ONT_Bn, Real_ONT_Sl). The available graphical representation of a workflow for simulation and real data is shown in **Figure 1**.

3.2 Characteristics of Structural Variant Truth Sets

The SVs supplied to VISOR to generate Sim_ONT_Bn1, Sim_ONT_Bn2, and Sim_ONT_Sl haplotypes served as three truth sets for our comparisons. The truth sets included deletions and insertions. The length distribution of truth set SVs is presented in **Figure 2**. Sim_ONT_Bn1 is unbiased in terms of the bioinformatics tools used, as the regions representing SVs were entirely randomly drawn from the *B. napus* genome. For any simulated dataset to reflect realistic SV distribution, SVs have to be discovered first and provided to the simulation software. Any relationship between tools used for SV identification for long-read dataset simulation and tools used for SV detection from these simulated reads (for example use of similar/same mapping algorithm) can result in inflated performance and biased results. However, Sim_ONT_Bn1 does not reflect realistic SV length and genomic distribution. To mitigate that Sim_ONT_Bn2 and Sim_ONT_Sl were created using SVs derived from real-world datasets. The two simulation strategies are complementary and should allow both unbiased and realistic assessment of SV calls. The median (mean) sizes (bp) for insertions and deletions were 800 (834) and 795 (825) for Sim_ONT_Bn1, 629 (1,959) and 594 (1,904) for Sim_ONT_Bn2 and 162 (3,178) and 165 (2,477) for Sim_ONT_Sl. Overall, the Sim_ONT_Bn2 and Sim_ONT_Sl truth sets had a wider range of insertion and deletion sizes. They were more reflective of true biological variation, making them more realistic than the Sim_ONT_Bn1 truth set.

3.3 Performance of Long Read Aligners

Subsampled *S. lycopersicum, B. napus,* and simulated reads were aligned using lra, minimap2, Vulcan, and NGMLR to the Slycopersicum_691_SL4.0, and Express 617 v1 reference genomes. Mapping statistics and run times of alignment against relevant reference genomes with different coverages of Sim_ONT_Bn1, Sim_ONT_Bn2, Sim_ONT_Sl, *B. napus* (Real_ONT_Bn), and *S. lycopersicum* (Real_ONT_Sl) real-world datasets are given in **Table S2**. Minimap2 had the shortest run time across all coverages. Conversely, NGMLR had the longest run time and also the lowest mapping rate. **Figure 3** shows mapping runtime (h:mm:ss or m:ss) for both simulation and real-world datasets with eight CPUs. Real_ONT_Bn dataset with 20× coverage was aligned ~220 hours by NGMLR and ~119 hours by Vulcan, compared to ~4 hours by minimap2 and ~5 hours by lra. Therefore, minimap2 and lra provided a greater speed advantage than NGMLR and Vulcan.

The run times increased with the higher coverages (**Figure 3**). Processing of real data took substantially longer than processing of simulated data. Moreover, Vulcan and minimap2 produced the highest proportion of mapped reads in Real_ONT_Bn (>96%), Real_ONT_Sl (96%-98%), and all simulated data (>98%) (**Table S2**). NGMLR reported the lowest proportion of mapped reads for Real_ONT_Bn (~81%) and Real_ONT_Sl (~76%), while lra and NGMLR resulted in similar statistics (96%-97%) for Sim_ONT_Bn1, Sim_ONT_Bn2, and Sim_ONT_Sl at each coverage. The combination of fast run time, good mapping rate, and the SV calling results presented below suggest that minimap2 is the top-performing aligner for simulated and real reads.

3.4 Performance of SV Callers on Simulated Data

3.4.1 Performance using Sim_ONT_Bn1 as benchmark

We calculated the precision, recall, and F1-score of the SVs generated using different mapper and SV caller combinations using the Sim_ONT_Bn1 truth set. **Table S3** shows comparison of the precision, recall, and F1-scores for all mapper/SV caller combinations at the 5×, 10×, and 20× coverages. Each aligner/SV caller combination was evaluated with respect to total SVs, deletions, and insertions. **Figure 4** presents the corresponding F1-scores at 5× to 20× coverages. CuteSV after minimap2 alignment reached the highest F1-scores 5×:~0.90, 10×:~0.97, and 20×:~0.99 for total SVs, 5×:~0.91, 10×:~0.97, and 20×:~0.99 for deletions, and 5×:~0.89, 10×:~0.96, and 20×:~0.99 for insertions. At the lower end of coverage (5×), combination of minimap2/cuteSV provided a better advantage when compared to other mapper/SV caller combinations, especially in capturing insertions. Minimap2/Sniffles2 had second-best F1-scores (**Figure 4**). SVs detection by NanoVar was obtained directly from reads as NanoVar has its own internal mapping algorithm therefore the precision, recall, and F1-scores for different aligners are not included.

We also compared the total number of SVs, insertions, and deletions for all tested aligner/SV caller combinations. **Table S4** summarizes the number of SVs found at 5×, 10×, and 20× coverages. There were more discovered deletions than insertions regardless of coverage. The combinations of minimap2/cuteSV and minimap2/Sniffles2 detected the highest number of SVs at each coverage. We also analyzed how many of the SVs overlapped across different coverages while using the same tool combination and how many of the SVs overlapped across different tool combinations within the same coverage. **Data S1** shows the number of overlapping and unique SVs across coverages. Minimap2/cuteSV combination had the highest number of overlapping SVs. It also resulted in the highest proportion of overlapping SVs; 76.99% for all SVs, 79.19% for deletions, and 74.79% for insertions, while the minimap2/Sniffles2 combination (second best according to F1-scores) had the second highest percentage overlap; 75.35% for all SVs, 78.35% for deletions, and 72.33% for insertions **(Table S11** and **Figure 7)**. In addition, we performed comparisons across different tool combinations within the same coverage. **Data S2** displays the overlap, including the intersection sizes between SV calls and the Sim_ONT_Bn1 truth set. The highest number of overlapping SVs was found at 20x coverage, following minimap2 aligner. Our Sim_ONT_Bn1 results suggest that the combination of cuteSV and Sniffles2 with minimap2 alignment gave the

best results achieving high F1-scores and capturing the highest number of overlapping SVs across coverages.

3.4.2 Performance using Sim_ONT_Bn2 as benchmark

While Sim_ONT_Bn1 represents relatively short SVs randomly distributed along the genome, Sim_ONT_Bn2 reflects true biological variation in *B. napus*. **Table S5** presents comparison of the precision, recall, and F1-scores for all mapper/SV caller combinations at the 5×, 10×, and 20× coverages. **Figure 5** presents the F1-scores of SVs (total, insertions, and deletions) obtained using different combinations of aligners and variant callers across coverages. CuteSV following minimap2 alignment again was the top performing combination with the highest overall F1-score values 5×:~0.87, 10×:~0.93, and 20×:~0.96 for total SVs, 5×:~0.90, 10×:~0.96, and 20×:~0.98 for deletions, and 5×:~0.83, 10×:~0.89, and 20×:~0.94 for insertions. Especially, at low 5× coverage, this combination performed better than others. Minimap2/Sniffles2 had the second highest F1-scores at 20× coverage as in Sim_ONT_Bn1. However, minimap2/dysgu F1-score for insertions at 5× and 10× was higher than Sniffles2 after the minimap2 alignment.

In addition, the total number of SVs, the total number of insertions, and deletions for all combinations of tested aligners and SV callers were compared. **Table S6** summarizes the total number of SVs detected at 5×, 10×, and 20× coverages. Minimap2/cuteSV found the highest number of SVs at each coverage like in Sim_ONT_Bn1. Again, more deletions than insertions were found for all aligner and SV caller combinations across different coverages. We also analyzed how many of the SVs overlapped across different coverages while using the same tool combination and how many of the SVs overlapped across different tool combinations within the same coverage. **Data S3** lists the number of overlapping SVs across different coverages using the same tool combination. Minimap2/cuteSV combination had the highest number of overlapping SVs. It also had the highest proportion of overlapping SVs; 73.95% for all SVs, 80.05% for deletions, and 67.44% for insertions. The minimap2/dysgu combination was second best detecting 73.23% for all SVs, and 67.28% for insertions. Minimap2/Sniffles2 combination was the second best for deletions with 79.14% overlap **(Table S11** and **Figure 7)**. **Data S4** displays overlap between results from different SV callers within the same coverage after each aligner, including the intersection with the Sim_ONT_Bn2 truth set. The highest number of overlapping SVs was found at 20x coverage, following minimap2 aligner. Overall, in Sim_ONT_Bn2, the combination of cuteSV after minimap2 alignment gave the best results both in terms of F1-Scores and concordance across coverages.

3.4.3 Performance using Sim_ONT_Sl as benchmark

Sim_ONT_Sl represents the true biological variation of *S. lycopersicum*. **Table S7** presents comparison of the precision, recall, and F1-scores for all mapper/SV caller combinations at the 5×, 10×, and 20× coverages. **Figure 6** shows the F1-score of SVs (total, insertions, and deletions) identified using combinations of the different aligners and variant callers. CuteSV and Sniffles2 with minimap2 alignment were top performers with the highest F1-score values (5×:~0.85,

10×:~0.92, and 20×:~0.94) for total SVs, (5×:~0.88, 10×:~0.95, and 20×:~0.97) for deletions, and (5×:~0.81, 10×:~0.88, and 20×:~0.91) for insertions. Lra/Sniffles2 combination had the best F1-score for insertions for each coverage.

In addition, the total number of SVs, the total number of insertions, and deletions for all tested aligner/SV caller combinations were compared. **Table S8** summarizes the total number of SVs at 5×, 10×, and 20× coverages. Again, more deletions than insertions were found for all aligner and SV caller combinations across coverages like in the previous simulated datasets. The number of SVs overlapping across coverages while using the same tool combination and the number of SVs overlapping across different tool combinations but within the same coverage were also calculated. **Data S5** shows the number of overlapping SVs across different coverages using the same tool combination. Minimap2/dysgu combination had the highest number of overlapping SVs. However, minimap2/cuteSV combination found the highest proportion of overlap; 73.49% for all SVs, 77.52% for deletions, and 68.98% for insertions, while the minimap2/Sniffles2 combination was second best detecting 72.73% for all SVs, 76.32% for deletions, and 68.72% for insertions **(Table S11 and Figure 7)**. Although minimap2/dysgu found the highest number of SVs at each coverage in Sim_ONT_Sl, the proportion of overlapped SVs was reported as 68.82%. **Data S6** displays overlap between results from different SV callers within the same coverage after each aligner, including the intersection with Sim_ONT_Sl truth set. The highest number of overlapping SVs was found at 20x coverage, following minimap2 aligner. Overall, in Sim_ONT_Sl, the combination of cuteSV and Sniffles2 after minimap2 alignment gave the best results both in terms of F1-Scores and concordance across coverages.

3.5 Performance of SV Callers on Real-World Data

While tool performance on simulated data provides a useful guide, real-world datasets usually provide additional unaccounted-for complexity and challenges. After finding the best combinations in simulated data, we investigated whether the pattern would be similar in real-world datasets. Since for the real-world data we do not have an objective truth set, they were only evaluated from two perspectives which are the congruence of results when using the same tool combination across different coverages and when using different tool combinations within the same coverage.

3.5.1 Performance on *B. napus* Real-World ONT Data

B. napus ONT real dataset (Real_ONT_Bn) was evaluated using the above-described strategy. **Table S9** shows the number of SVs from all tested combinations at different coverages in *B. napus*. The minimap2/cuteSV and minimap2/dysgu combinations within all coverages captured the highest number of total SVs, deletions, and insertions. Overall, a higher number of deletions than insertions was detected for all aligner and SV caller combinations at different coverages. The number of overlapped SVs across coverages for the same SVs caller/aligner combinations was calculated (**Data S7**). Minimap2/cuteSV combination found the highest proportion of overlapping SVs discovered at different coverages using the same combination of tools (51.53% of total SVs, 54.52% of deletions, and 47.91% of insertions), while the minimap2/sniffles2 combination was

second best, detecting overlap of 50.1% for all SVs, 54.56% for deletions, and 44.92% for insertions across coverages **(Table S12 and Figure 7)**. Although the minimap2/dysgu combination found more SVs, the percentage of intersecting SV was low. NanoVar detected the lowest proportion of overlapping SVs across coverages (19.04% of total SVs, 25.07% of deletions, and 10.21% of insertions) and discovered more unique SVs. Surprisingly we noticed a high proportion of heterozygous genotypes (0/1) in SV calling results for Real_ONT_Bn, considering that the data represented a highly inbred elite line (Vollrath et al., 2021). **Table S14** shows the number of SVs genotyped as homozygous and heterozygous in real-world data. As our SV filtering required the genotypes to be homozygous for the alternative allele (1/1) these heterozygous calls were removed prior to analysis. We also investigated the overlap in SV calls across different tool combinations within the same coverage (**Data S8**). We observed that a substantial proportion of deletions and insertions were shared by most SV callers, with the largest number of overlapping SVs at 20×, following minimap2 alignment.

3.5.2 Performance on *S. lycopersicum* Real-World ONT Data

We performed a similar evaluation for the real-world dataset of *Solanum lycopersicum* (Real_ONT_Sl). **Table S10** shows the number of SVs found from all tested combinations at different coverages. The minimap2/dysgu combinations at 5×, 10×, and 20× captured the most SVs. Additionally, for *S. lycopersicum* all tool combinations with the exception of NanoVar found more insertions than deletions at each coverage. We also calculated the number of overlapping SVs while using the same tool combination across different coverages (**Data S9**). Minimap2/cuteSV combination found the highest proportion of overlapping SVs; 49.34% for all SVs, 49.63% for deletions, and 49.16% for insertions, while the minimap2/sniffles2 combination detected 47.80% for all SVs, 49.41% for deletions, and 46.61% for insertions. Even though the minimap2/dysgu combination found more SVs, the percentage of common SVs (40.82%) was low like Real_ONT_Bn data. NanoVar again detected the lowest proportion of overlapping SVs (21.57% for all SVs, 31.20% for deletions, and 12.16% for insertions), and it discovered more unique SVs like for the Real_ONT_Bn dataset **(Table S12 and Figure 7)**. Again, we also tested overlaps between SV calls within the same coverage, but across different tool combinations (**Data S10**). The largest number of overlapping SVs was found at 20×, following minimap2 alignment.

3.5.3 The Unique Features of Real-World Datasets

We found a surprisingly high proportion of heterozygous calls in the real-world datasets given the highly inbred nature of the material used for sequencing. A high proportion of those is therefore likely SV discovery/genotyping errors. More heterozygous calls were found in the *B. napus* than the *S. lycopersicum* dataset. *B. napus* is an allotetraploid species, which undergoes reciprocal and non-reciprocal homeologous exchanges (HEs; exchanges of large corresponding chromosome segments between subgenomes). Non-reciprocal HEs could potentially cause erroneous SV calls if there are HE present in the reference, but absent in the sample. As a result, reads will have no corresponding mapping location and may be mis-mapped. To test such a scenario, we used the Sim_ONT_Bn2 dataset (20×, minimap2 for mapping, and cuteSV for SV detection) and two

versions of the modified Express 617 reference. In the first version, we replaced chromosome A01 by C01 (two C01 chromosomes and no A01). In the second version, we replaced chromosome C01 by A01 (two A01 chromosomes and no C01). In both cases, the use of the modified reference resulted in an increased number of heterozygous (162.3% for reference with A01 missing, and 237.1% for reference with C01 missing), but not homozygous calls across all chromosomes (**Figure 8**), suggesting the non-reciprocal HEs can contribute to produce erroneous heterozygous calls.

4. DISCUSSION

Many of the SV detection tools are benchmarked primarily on human/animal datasets, (Bolognini and Magi, 2021; Coster et al., 2019; Dierckxsens et al., 2021; Jiang et al., 2020; Jiang et al., 2021; Zhou et al., 2019), however the complexity and different SV profiles of crop plant genomes might bring unique challenges. Therefore, to guide the design of large-scale long-read re-sequencing studies, this study performed comprehensive benchmarking of popular SV calling tools with a focus on tool performance at lower sequencing coverage. For this purpose, we designed two data simulation strategies representing both unbiased and realistic benchmarking datasets reflecting structural variation for two major crops oilseed rape (*B. napus*) and tomato (*S. lycopersicum*).

Four long-read aligners (minimap2, NGMLR, lra, and Vulcan) and five SV callers (Sniffles2, SVIM, cuteSV, dysgu, and NanoVar) were tested to detect SVs, particularly deletions and insertions. Alignment time varied widely between the four aligners, while differences in the proportion of mapped reads were moderate. As expected, higher sequencing coverage and reference genome size length increased the run time of the mapping algorithms. The real-world datasets required more time at the same coverage and reference genome size, which most likely reflected additional complexity not captured in simulations. Overall, the results found minimap2 to be the best performing aligner for SV calling applications, which also had the fastest run time and the most mapped bases. Recent benchmarking studies on human data also recommended minimap2 among tested aligners such as GraphMap, LAST, and NGMLR (Bolognini and Magi, 2021; Coster et al., 2019; Zhou et al., 2019).

We found that similar tool combinations (especially cuteSV, followed closely by Sniffles2 and dysgu after minimap2 alignment) had superior performance across all the simulated datasets. The findings are in line with a recent study reporting that cuteSV performed better than other tested SV tools such as Sniffles1, SVIM, and pbsv for precision and recall at both SV calling and genotyping in human datasets (Bolognini and Magi, 2021). Increasing coverage improved recall and F1-scores for all tested SVs calling combinations, confirming that the probability of detecting quality SVs increases with more sequencing coverage (Jiang et al., 2021). However, even at low coverages (5×) using cuteSV, Sniffles2, and dysgu for SV detection from reads aligned by minimap2 achieved >0.8 F1-scores on simulated datasets, suggesting that Oxford Nanopore technology might be suitable for large-scale low coverage re-sequencing projects. While the lack of objective truth sets

for real-world datasets precludes similar comparisons, the results revealed that tool combinations with best performance for simulated datasets also had the most consistent outcome across the range of coverages.

The criteria for filtering SV in this study were quite stringent, including retaining only SV genotyped as homozygous for alternative allele (1/1). While in simulated datasets the number of SV genotyped as heterozygous was relatively low, the proportion was much higher for real-world datasets, especially in *B. napus*. We found that in *B. napus*, the presence of homeologous exchanges will likely contribute to the erroneous discovery of heterozygous SV. *B. napus* is well known to harbour wide-spread non-reciprocal homeologous chromosomal exchanges even extending to whole chromosomes, e.g. for chromosomes A01 and C01 as simulated here (Udall et al., 2005). The finding underlies the importance of species-specific consideration when interpreting SV discovery results. The presence of HEs likely explains only a proportion of the observed heterozygous calls and other factors need to be considered as well, including other sources of mis-mappings, genotyping errors, and residual heterozygosity in samples.

In conclusion, we found that for homozygous/inbred genotypes often used in crop studies a substantial proportion of SVs can be discovered/genotyped at coverages as low as 5×, making Oxford Nanopore technology a suitable option for larger-scale re-sequencing studies. At this time, following our benchmarks we recommend using the minimap2/cuteSV combination as it achieves good precision and recall at SV calling and found the highest overlap between SVs across coverages. The performance of minimap2/cuteSV was followed closely by minimap2/Sniffles2 for both simulated and real datasets.

FIGURES

Figure 1: Graphical overview of the benchmarking workflow.
Figure 2: Size distribution of the real-world SV and SV from three benchmarking datasets.
Figure 3: Read aligner run time (h:mm: ss or m: ss) for both simulation and real-world datasets with 5×, 10×, and 20× coverages (8 CPU). The reads were simulated with a mean length of 15,000 bp. The real-world datasets had a mean read length of 12,553 bp for *B. napus* and 22,339 bp for *S. lycopersicum*.
Figure 4: F1-scores of Sim_ONT_Bn1 including total SVs, deletions, and insertions at 5×, 10×, and 20× coverages for different combinations of read aligners and SV callers.
Figure 5: F1-scores of Sim_ONT_Bn2 including total SVs, deletions, and insertions at 5×, 10×, and 20× coverages for different combinations of read aligners and SV callers.
Figure 6: F1-scores of Sim_ONT_Sl including total SVs, deletions, and insertions at 5×, 10×, and 20× coverages for different combinations of read aligners and SV callers.
Figure 7: Proportion of overlapped SVs (%), across 5×, 10×, and 20× coverages for simulated and real-world datasets.

Figure 8: The effect of non-reciprocal homeologous exchanges on SV discovery. Non-reciprocal homeologous exchanges were simulated by replacing chromosome A01 by C01 and C01 by A01.

DATA ACCESS

Supplementary Data and Files can be accessed under the following link: https://osf.io/9c5hz/

ACKNOWLEDGMENTS

This work was supported by the Alexander von Humboldt Foundation in the framework of Sofja Kovalevskaja Award to AAG and the German Research Foundation (DFG) project number 458716530 to RJS.

AUTHOR CONTRIBUTIONS

GY performed research, wrote the manuscript
SFZ assisted in the analysis, wrote the manuscript
NPA provided critical comments, wrote the manuscript
CO provided critical comments, wrote the manuscript
RJS edited the manuscript
AAG conceived research, supervised research, wrote the manuscript, acquired funding

REFERENCES

Alkan, C., Coe, B. P. and Eichler, E. E. (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* **12**, 363–376.

Alonge, M., Lebeigle, L., Kirsche, M., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C. and Soyk, S. (2021) Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. *bioRxiv*, 2021.11.18.469135.

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., Levy, Y., Harel, T. H., Shalev-Schlosser, G., Amsellem, Z., Razifard, H., Caicedo, A. L., Tieman, D. M., Klee, H., Kirsche, M., Aganezov, S., Ranallo-Benavidez, T. R., Lemmon, Z. H., Kim, J., Robitaille, G., Kramer, M., Goodwin, S., McCombie, W. R., Hutton, S., van Eck, J., Gillis, J., Eshed, Y., Sedlazeck, F. J., van der Knaap, E., Schatz, M. C. and Lippman, Z. B. (2020) Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell* **182**, 145-161.e23.

Bolognini, D. and Magi, A. (2021) Evaluation of Germline Structural Variant Calling Methods for Nanopore Sequencing Data. *Frontiers in Genetics* **12**.

Bolognini, D., Sanders, A., Korbel, J. O., Magi, A., Benes, V. and Rausch, T. (2020) VISOR: a versatile haplotype-aware structural variant simulator for short- and long-read sequencing. *Bioinformatics* **36**, 1267–1269.

Chawla, H. S., Lee, H., Gabur, I., Vollrath, P., Tamilselvan-Nattar-Amutha, S., Obermeier, C., Schiessl, S. V., Song, J.-M., Liu, K., Guo, L., Parkin, I. A. P. and Snowdon, R. J. (2021) Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. *Plant Biotechnol J* **19**, 240–250.

Cleal, K. and Baird, D. M. (2022) Dysgu: efficient structural variant calling using short or long reads. *Nucleic acids research*.

493 Coster, W. de, Rijk, P. de, Roeck, A. de, Pooter, T. de, D'Hert, S., Strazisar, M., Sleegers, K. and van
494 Broeckhoven, C. (2019) Structural variants identified by Oxford Nanopore PromethION sequencing of
495 the human genome. *Genome Res* **29**, 1178–1187.
496 Coster, W. de, Weissensteiner, M. H. and Sedlazeck, F. J. (2021) Towards population-scale long-read
497 sequencing. *Nature Reviews Genetics* **22**, 572–587.
498 Delahaye, C. and Nicolas, J. (2021) Sequencing DNA with nanopores: Troubles and biases. *PLOS ONE* **16**,
499 e0257521.
500 Dierckxsens, N., Li, T., Vermeesch, J. R. and Xie, Z. (2021) A benchmark of structural variation detection
501 by long reads through a realistic simulated model. *Genome Biology* **22**, 342.
502 English, A. C., Menon, V. K., Gibbs, R., Metcalf, G. A. and Sedlazeck, F. J. (2022) Truvari: Refined Structural
503 Variant Comparison Preserves Allelic Diversity. *bioRxiv*, 2022.02.21.481353.
504 Fu, Y., Mahmoud, M., Muraliraman, V. V., Sedlazeck, F. J. and Treangen, T. J. (2021) Vulcan: Improved
505 long-read mapping and structural variant calling via dual-mode alignment. *Gigascience* **10**, giab063.
506 Gill, R. A., Scossa, F., King, G. J., Golicz, A. A., Tong, C., Snowdon, R. J., Fernie, A. R. and Liu, S. (2021) On
507 the Role of Transposable Elements in the Regulation of Gene Expression and Subgenomic
508 Interactions in Crop Genomes. *Critical Reviews in Plant Sciences* **40**, 157–189.
509 Hall, M. B. (2022) Rasusa: Randomly subsample sequencing reads to a specified coverage. *Journal of*
510 *Open Source Software* **7**, 3941.
511 Heller, D. and Vingron, M. (2019) SVIM: structural variant identification using mapped long reads.
512 *Bioinformatics* **35**, 2907–2915.
513 Heller, D. and Vingron, M. (2020) SVIM-asm: Structural variant detection from haploid and diploid
514 genome assemblies. *Bioinformatics* **36**, 5519–5521.
515 Hosmani, P. S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., van Haarst,
516 J., Cordewener, J., Sanchez-Perez, G., Peters, S., Fei, Z., Giovannoni, J. J., Mueller, L. A. and Saha, S.
517 (2019) An improved de novo assembly and annotation of the tomato reference genome using single-
518 molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv*, 767764.
519 Jain, M., Olsen, H. E., Paten, B. and Akeson, M. (2016) The Oxford Nanopore MinION: delivery of
520 nanopore sequencing to the genomics community. *Genome Biology* **17**, 239.
521 Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J. and
522 Sedlazeck, F. J. (2017) Transient structural variations have strong effects on quantitative traits and
523 reproductive isolation in fission yeast. *Nature Communications* **8**, 14061.
524 Jiang, T., Liu, S., Cao, S., Liu, Y., Cui, Z., Wang, Y. and Guo, H. (2021) Long-read sequencing settings for
525 efficient structural variation detection based on comprehensive evaluation. *BMC Bioinformatics* **22**,
526 552.
527 Jiang, T., Liu, Y., Jiang, Y., Li, J., Gao, Y., Cui, Z., Liu, Y., Liu, B. and Wang, Y. (2020) Long-read-based human
528 genomic structural variation detection with cuteSV. *Genome Biology* **21**, 189.
529 Lee, H., Chawla, H. S., Obermeier, C., Dreyer, F., Abbadi, A. and Snowdon, R. (2020) Chromosome-Scale
530 Assembly of Winter Oilseed Rape Brassica napus. *Frontiers in Plant Science* **11**.
531 Lemay, M.-A., Sibbesen, J. A., Torkamaneh, D., Hamel, J., Levesque, R. C. and Belzile, F. (2022) Combined
532 use of Oxford Nanopore and Illumina sequencing yields insights into soybean structural variation
533 biology. *BMC Biology* **20**, 53.
534 Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100.
535 Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C. and Sedlazeck, F. J. (2019)
536 Structural variant calling: the long and the short of it. *Genome Biology* **20**, 246.

Meyers, L. A. and Levin, D. A. (2006) On the abundance of polyploids in flowering plants. *Evolution* **60**, 1198–1206.

Ren, J. and Chaisson, M. J. P. (2021) lra: A long read aligner for sequences and contigs. *PLOS Computational Biology* **17**, e1009078.

Roberts, R. J., Carneiro, M. O. and Schatz, M. C. (2013) The advantages of SMRT sequencing. *Genome Biology* **14**, 405.

Sedlazeck, F. J., Lee, H., Darby, C. A. and Schatz, M. C. (2018a) Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics* **19**, 329–346.

Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Haeseler, A. von and Schatz, M. C. (2018b) Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* **15**, 461–468.

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., Xie, W.-Z., Cheng, Y., Zhang, Y., Liu, K., Yang, Q.-Y., Chen, L.-L. and Guo, L. (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. *Nature Plants* **6**, 34–45.

Tao, Y., Zhao, X., Mace, E., Henry, R. and Jordan, D. (2019) Exploring and Exploiting Pan-genomics for Crop Improvement. *Molecular Plant* **12**, 156–169.

Tham, C. Y., Tirado-Magallanes, R., Goh, Y., Fullwood, M. J., Koh, B. T., Wang, W., Ng, C. H., Chng, W. J., Thiery, A., Tenen, D. G. and Benoukraf, T. (2019) NanoVar: Accurate Characterization of Patients' Genomic Structural Variants Using Low-Depth Nanopore Sequencing. *bioRxiv*, 662940.

Udall, J. A., Quijada, P. A. and Osborn, T. C. (2005) Detection of Chromosomal Rearrangements Derived From Homeologous Recombination in Four Mapping Populations of Brassica napus L. *Genetics* **169**, 967–979.

Vollrath, P., Chawla, H. S., Schiessl, S. V., Gabur, I., Lee, H., Snowdon, R. J. and Obermeier, C. (2021) A novel deletion in FLOWERING LOCUS T modulates flowering time in winter oilseed rape. *Theoretical and Applied Genetics* **134**, 1217–1231.

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N. D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A. M., Schatz, M. C., Myers, G., DePristo, M. A., Ruan, J., Marschall, T., Sedlazeck, F. J., Zook, J. M., Li, H., Koren, S., Carroll, A., Rank, D. R. and Hunkapiller, M. W. (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* **37**, 1155–1162.

Yildiz, G., Zanini, S. F., Knight, P. and Golicz, A. A. (2022) Pangenomics in Agriculture. *CABI Biotechnology Series*.

Yuan, Y., Bayer, P. E., Batley, J. and Edwards, D. (2021) Current status of structural variation studies in plants. *Plant Biotechnol J* **19**, 2153–2163.

Zanini, S. F., Bayer, P. E., Wells, R., Snowdon, R. J., Batley, J., Varshney, R. K., Nguyen, H. T., Edwards, D. and Golicz, A. A. (2022) Pangenomics in crop improvement—from coding structural variations to finding regulatory variants with pangenome graphs. *Plant Genome* **15**, e20177.

Zhang, F., Xue, H., Dong, X., Li, M., Zheng, X., Li, Z., Xu, J., Wang, W. and Wei, C. (2022) Long-read sequencing of 111 rice genomes reveals significantly larger pan-genomes. *Genome Res* **32**, 853–863.

Zhou, A., Lin, T. and Xing, J. (2019) Evaluating nanopore sequencing data processing pipelines for structural variation identification. *Genome Biology* **20**, 237.

Żmieńko, A., Samelak, A., Kozłowski, P. and Figlerowicz, M. (2014) Copy number polymorphism in plant genomes. *Theoretical and Applied Genetics* **127**, 1–18.

Figure 1

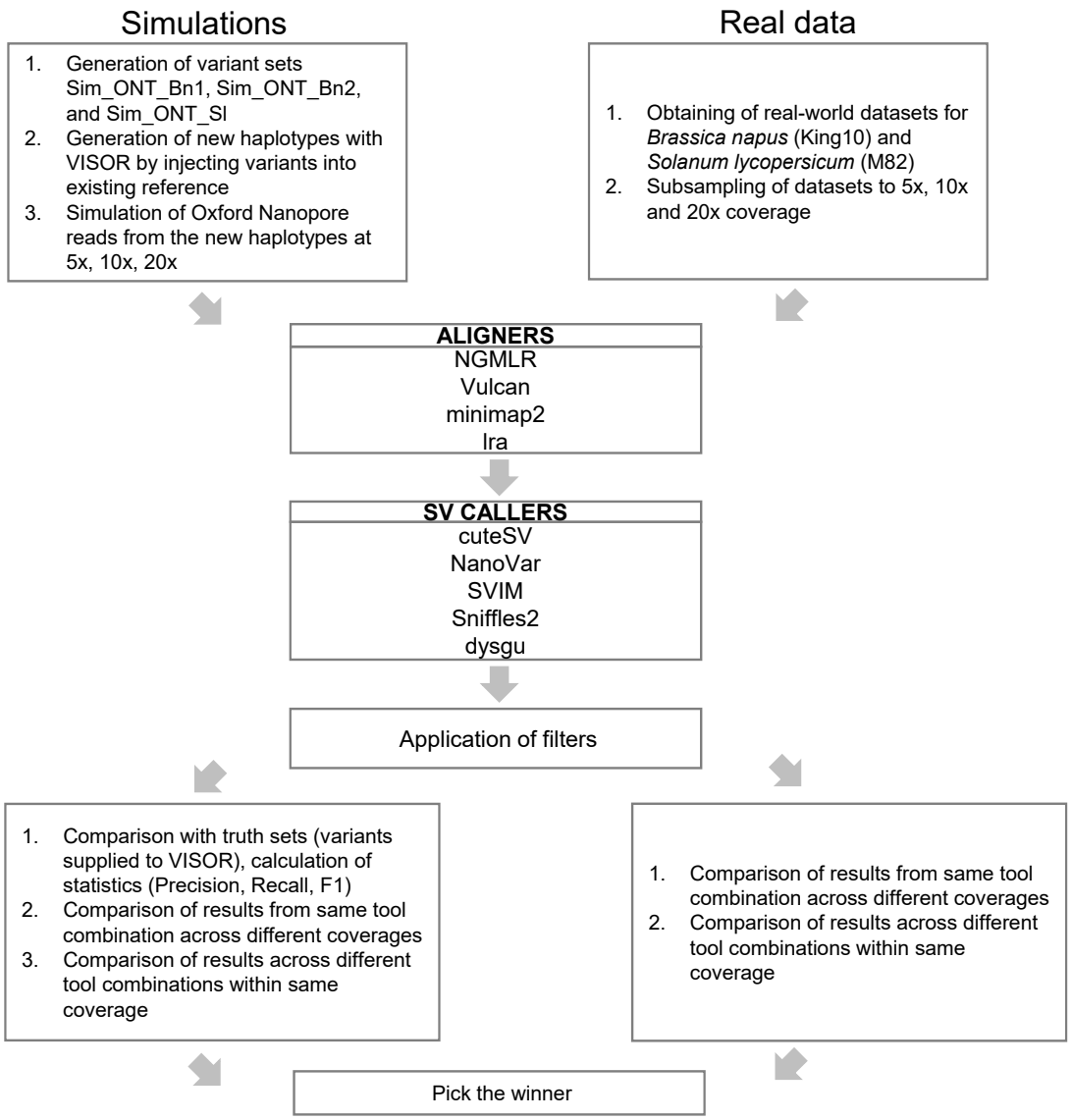


Figure 2

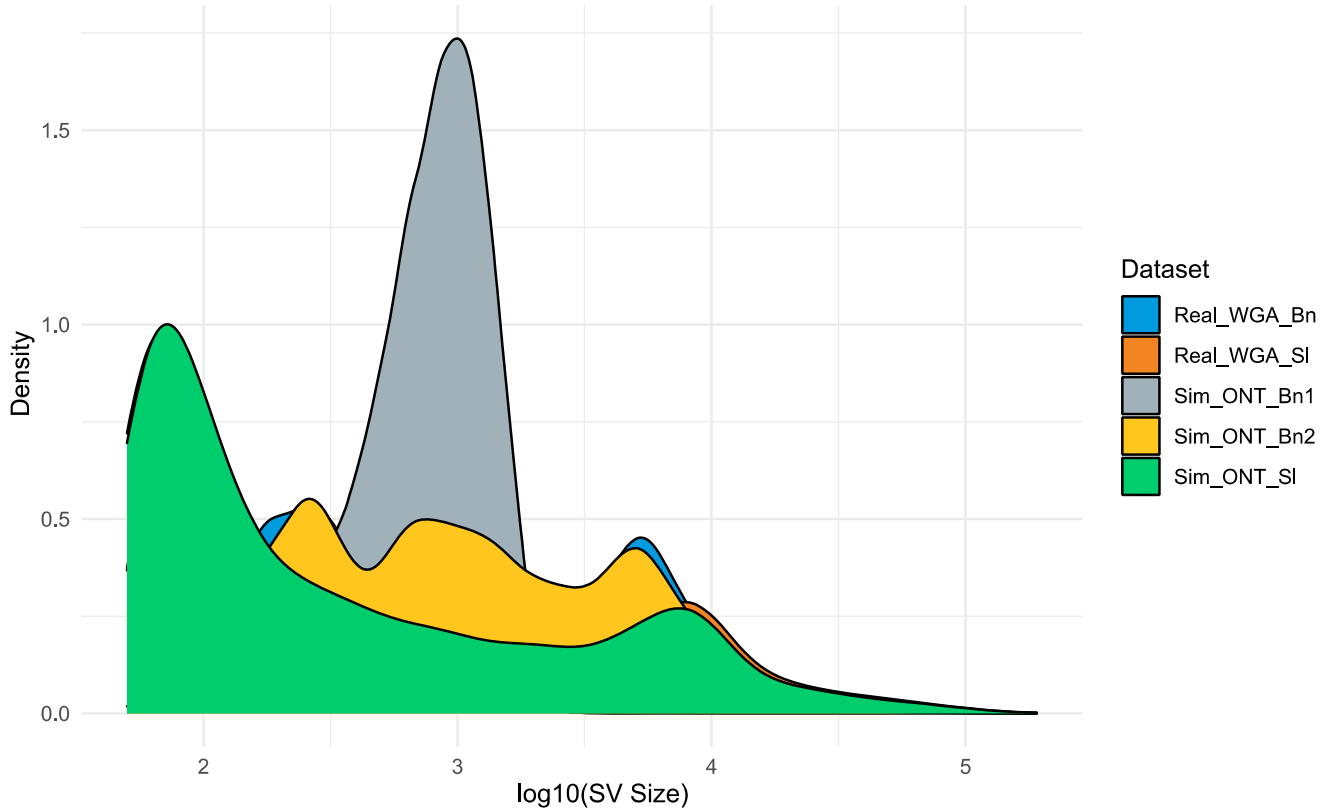
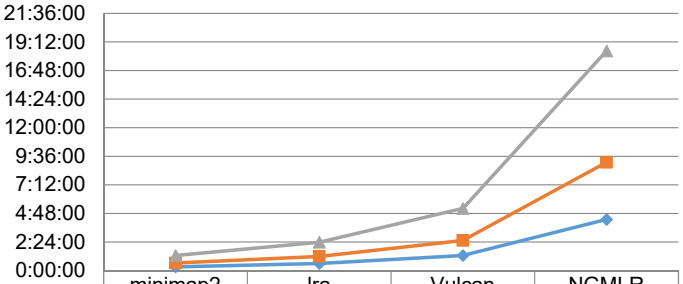


Figure 3

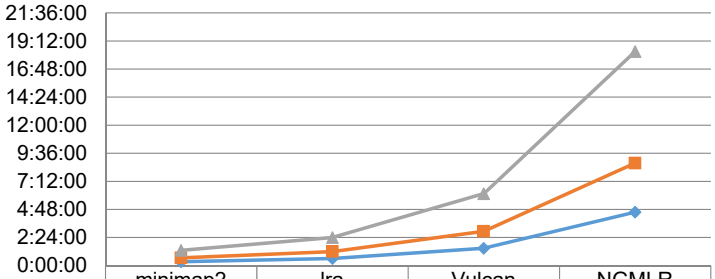
Simulated data

Runtimes (h:mm:ss or m:ss) of Sim_ONT_Bn1



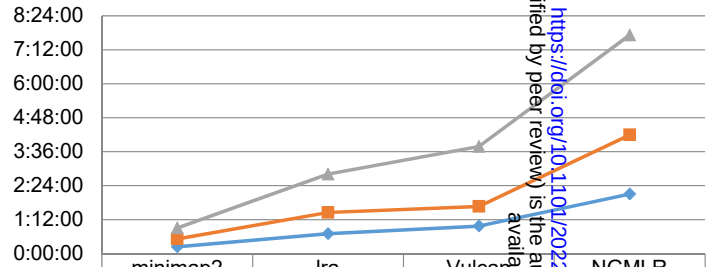
	minimap2	lra	Vulcan	NGMLR
5x	19:18.1	36:10.9	01:16:23	04:18:05
10x	38:07.1	01:11:49	02:32:03	09:05:54
20x	01:16:06	02:23:16	05:13:13	18:27:09

Runtimes (h:mm:ss or m:ss) of Sim_ONT_Bn2



	minimap2	lra	Vulcan	NGMLR
5x	19:57.9	36:36.8	01:29:29	04:34:22
10x	39:48.7	01:12:28	02:56:15	08:45:46
20x	01:18:45	02:24:05	06:08:58	18:18:12

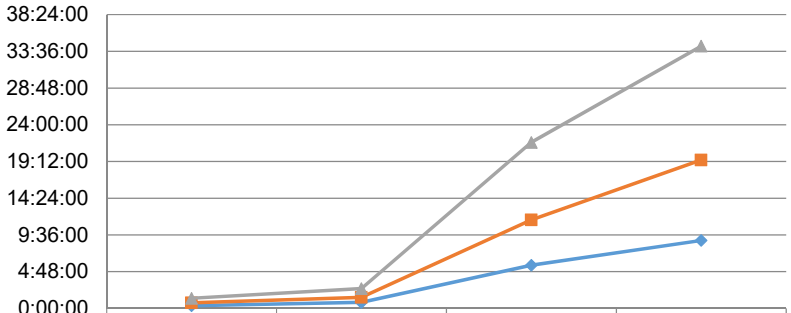
Runtimes (h:mm:ss or m:ss) of Sim_ONT_Sl



	minimap2	lra	Vulcan	NGMLR
5x	14:31.9	42:20.4	00:58:2[illegible]	02:06:16
10x	31:09.0	01:27:02	01:40:0[illegible]	04:11:44
20x	00:54:23	02:48:33	03:47:0[illegible]	07:43:21

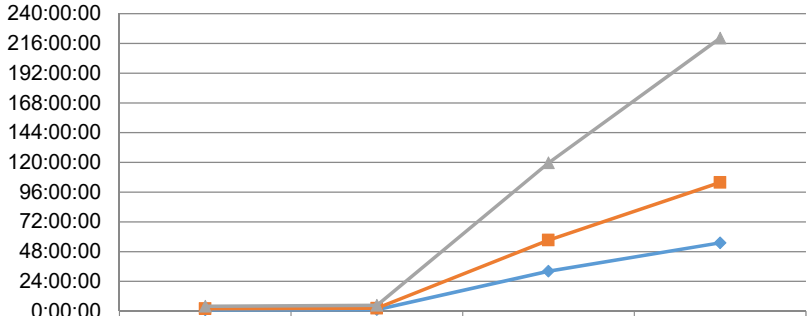
Real-world data

Runtimes (h:mm:ss or m:ss) of Real_ONT_Sl



	minimap2	lra	Vulcan	NGMLR
5x	21:17.4	46:34.3	05:38:56	08:52:39
10x	43:40.8	01:25:41	11:33:37	19:23:59
20x	01:19:18	02:35:05	21:42:18	34:17:16

Runtimes (h:mm:ss or m:ss) of Real_ONT_Bn



	minimap2	lra	Vulcan	NGMLR
5x	59:15.0	01:13:05	32:10:56	55:06:47
10x	01:55:55	02:18:07	57:15:57	103:46:26
20x	03:52:07	04:44:12	119:39:11	220:11:56

Figure 4

Total

	5x-F1	10x-F1	20x-F1
minimap2			
cuteSV	0.9003	0.9676	0.9955
Sniffles2	0.8928	0.9635	0.9948
SVIM	0.7825	0.9645	0.9869
dysgu	0.8618	0.9417	0.9776
Ira			
cuteSV	0.8665	0.9417	0.9829
Sniffles2	0.8578	0.9352	0.9801
SVIM	0.7291	0.9354	0.9793
dysgu	0.7593	0.8718	0.9148
Vulcan			
cuteSV	0.8495	0.9256	0.9751
Sniffles2	0.8000	0.8787	0.9463
SVIM	0.6864	0.9024	0.9345
dysgu	0.7441	0.8253	0.8639
NGMLR			
cuteSV	0.8001	0.8691	0.9220
Sniffles2	0.6524	0.7174	0.7689
SVIM	0.6295	0.8496	0.8980
dysgu	0.6275	0.7120	0.7415

	5x-F1	10x-F1	20x-F1
NanoVar	0.8950	0.9593	0.9848

Deletions

	5x-F1	10x-F1	20x-F1
minimap2			
cuteSV	0.9074	0.9733	0.9961
Sniffles2	0.9037	0.9724	0.9963
SVIM	0.7970	0.9715	0.9957
dysgu	0.9057	0.9721	0.9952
Ira			
cuteSV	0.8836	0.9562	0.9860
Sniffles2	0.8821	0.9557	0.9865
SVIM	0.7696	0.9563	0.9857
dysgu	0.8783	0.9552	0.9852
Vulcan			
cuteSV	0.8707	0.9441	0.9823
Sniffles2	0.8544	0.9323	0.9780
SVIM	0.7325	0.9389	0.9809
dysgu	0.8695	0.9484	0.9814
NGMLR			
cuteSV	0.8490	0.9152	0.9496
Sniffles2	0.8198	0.8924	0.9338
SVIM	0.7116	0.9110	0.9477
dysgu	0.8428	0.9260	0.9534

	5x-F1	10x-F1	20x-F1
NanoVar	0.9012	0.9676	0.9913

Insertions

	5x-F1	10x-F1	20x-F1
minimap2			
cuteSV	0.8931	0.9620	0.9948
Sniffles2	0.8818	0.9544	0.9933
SVIM	0.7676	0.9574	0.9778
dysgu	0.8140	0.9092	0.9593
Ira			
cuteSV	0.8488	0.9267	0.9798
Sniffles2	0.8324	0.9138	0.9736
SVIM	0.6857	0.9135	0.9728
dysgu	0.6112	0.7735	0.8336
Vulcan			
cuteSV	0.8275	0.9065	0.9678
Sniffles2	0.7401	0.8191	0.9124
SVIM	0.6367	0.8632	0.8834
dysgu	0.5866	0.6689	0.7150
NGMLR			
cuteSV	0.7465	0.8187	0.8927
Sniffles2	0.4282	0.4753	0.5424
SVIM	0.5358	0.7805	0.8431
dysgu	0.3116	0.3894	0.4193

	5x-F1	10x-F1	20x-F1
NanoVar	0.8886	0.9509	0.9784

Figure 5

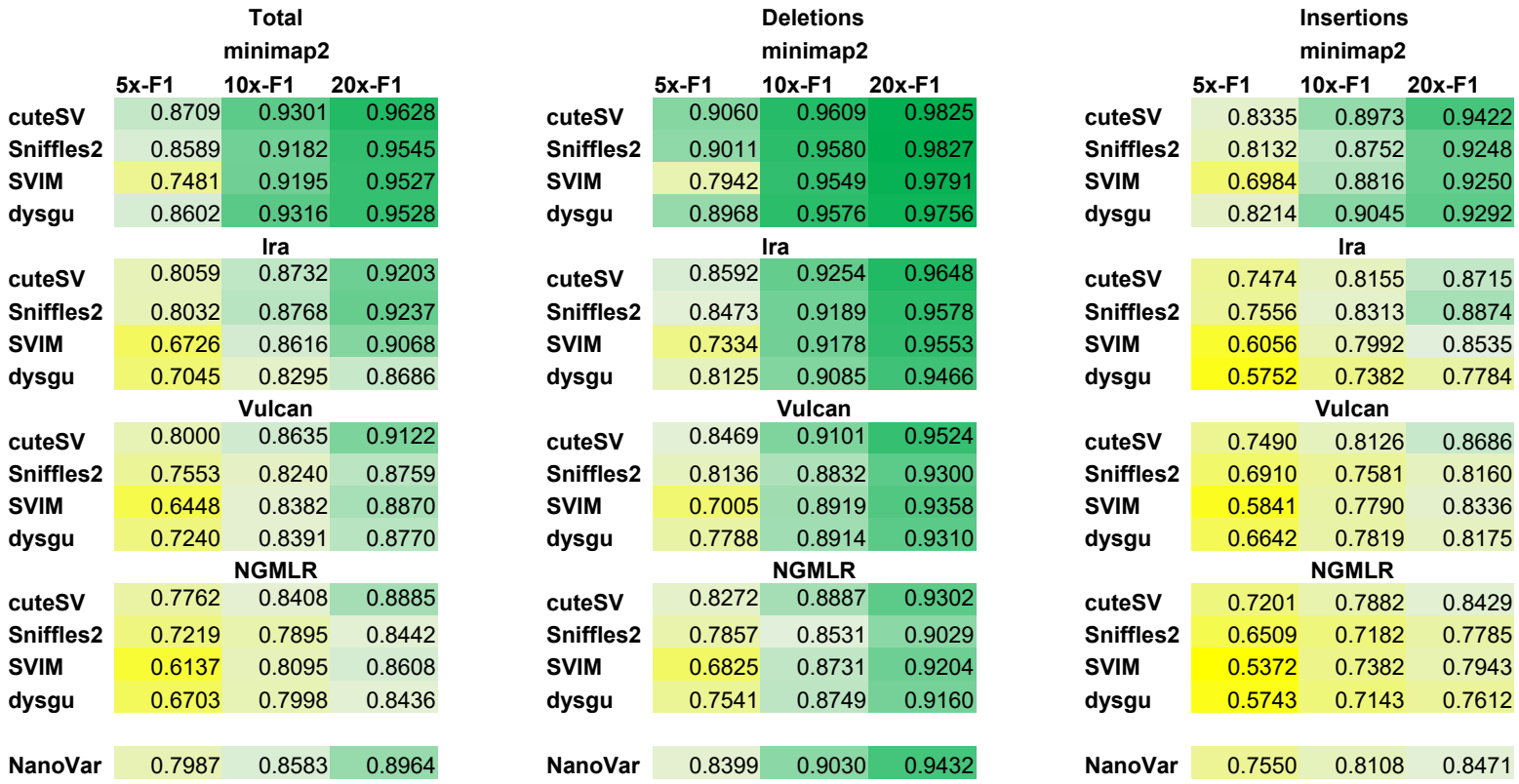


Figure 6

Total

minimap2

	5x-F1	10x-F1	20x-F1
cuteSV	0.8467	0.9167	0.9375
Sniffles2	0.8432	0.9174	0.9394
SVIM	0.7520	0.9184	0.9377
dysgu	0.8371	0.9008	0.9043
lra			
cuteSV	0.8158	0.8936	0.9278
Sniffles2	0.8334	0.9170	0.9515
SVIM	0.7158	0.8955	0.9315
dysgu	0.7682	0.8884	0.9149
Vulcan			
cuteSV	0.8128	0.8881	0.9161
Sniffles2	0.8012	0.8785	0.9120
SVIM	0.6994	0.8804	0.9140
dysgu	0.7852	0.8759	0.8992
NGMLR			
cuteSV	0.8002	0.8693	0.9001
Sniffles2	0.7889	0.8615	0.9015
SVIM	0.6832	0.8662	0.9050
dysgu	0.7636	0.8626	0.8881
NanoVar	0.7504	0.8098	0.8103

Deletions

minimap2

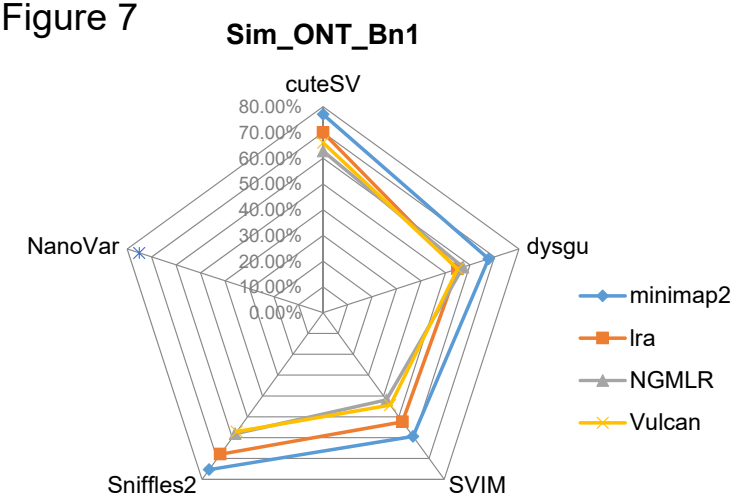
	5x-F1	10x-F1	20x-F1
cuteSV	0.8831	0.9477	0.9654
Sniffles2	0.8795	0.9492	0.9671
SVIM	0.7944	0.9488	0.9647
dysgu	0.8594	0.9194	0.9226
lra			
cuteSV	0.8547	0.9308	0.9628
Sniffles2	0.8519	0.9329	0.9652
SVIM	0.7591	0.9322	0.9639
dysgu	0.8040	0.9215	0.9462
Vulcan			
cuteSV	0.8490	0.9255	0.9525
Sniffles2	0.8324	0.9123	0.9448
SVIM	0.7396	0.9188	0.9476
dysgu	0.8005	0.9074	0.9241
NGMLR			
cuteSV	0.8370	0.9114	0.9380
Sniffles2	0.8178	0.8924	0.9257
SVIM	0.7279	0.9106	0.9412
dysgu	0.7908	0.9002	0.9194
NanoVar	0.8488	0.9093	0.9232

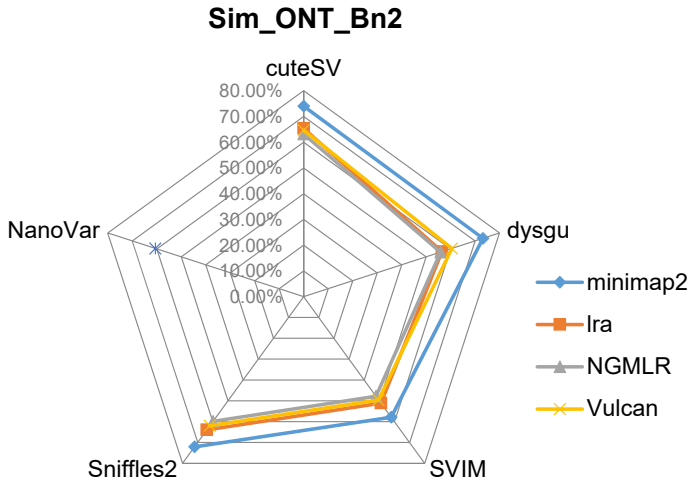
Insertions

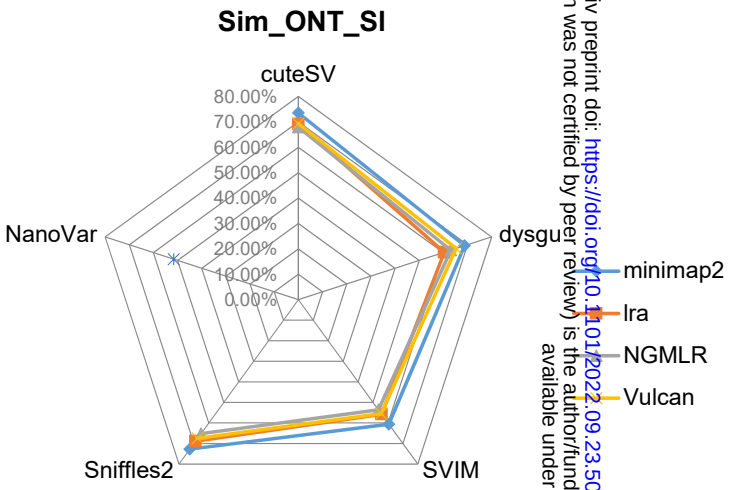
minimap2

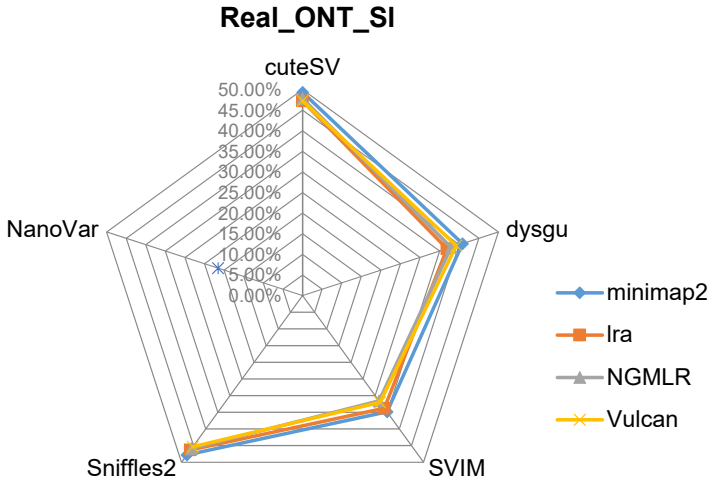
	5x-F1	10x-F1	20x-F1
cuteSV	0.8073	0.8833	0.9077
Sniffles2	0.8041	0.8835	0.9099
SVIM	0.7060	0.8858	0.9089
dysgu	0.8134	0.8814	0.8852
lra			
cuteSV	0.7736	0.8530	0.8897
Sniffles2	0.8141	0.9007	0.9376
SVIM	0.6688	0.8558	0.8966
dysgu	0.7295	0.8524	0.8808
Vulcan			
cuteSV	0.7736	0.8472	0.8763
Sniffles2	0.7678	0.8419	0.8768
SVIM	0.6562	0.8387	0.8780
dysgu	0.7694	0.8421	0.8730
NGMLR			
cuteSV	0.7601	0.8226	0.8581
Sniffles2	0.7580	0.8282	0.8758
SVIM	0.6347	0.8174	0.8658
dysgu	0.7348	0.8214	0.8541
NanoVar	0.6282	0.6841	0.6608

Figure 7









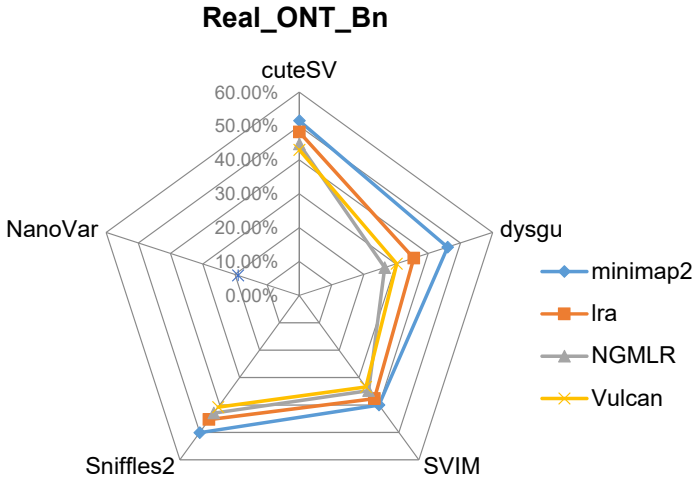


Figure 8