Software note.

1. pauvre is useful to check the sequencing stats (pauvre stats)
   1. https://github.com/conchoecia/pauvre
2. many useful tools (Chopper, Cramino, NanoPlot!)
   1. <https://github.com/wdecoster/nanopack>
3. Long-read correction:
   1. <https://github.com/HaploKit/vechat>
   2. https://github.com/morispi/CONSENT
4. New assembler, no publication currently
   1. <https://github.com/Nextomics/NextDenovo>
5. NextPolish: a fast and efficient genome polishing tool for long-read assembly (>280 citations)
   1. <https://github.com/Nextomics/NextPolish>
6. LongStitch: A genome assembly correction and scaffolding pipeline using long reads
   1. <https://github.com/bcgsc/longstitch>
7. RagTag is a collection of software tools for scaffolding and improving modern genome assemblies.
   1. <https://github.com/malonge/RagTag>
8. QUAST stands for QUality ASsessment Tool. Genome assembly evaluation tool
   1. <https://github.com/ablab/quast>
9. Inspector: A reference-free assembly evaluator.
   1. <https://github.com/Maggi-Chen/Inspector>
10. Long-read mapping to repetitive reference sequences using Winnowmap2
    1. Slightly better than minimap2 generally. Better at repetitive regions, such centromeres.
    2. <https://github.com/marbl/Winnowmap>
11. Fast minimap2
    1. <https://github.com/bwa-mem2/mm2-fast>
12. Fast BWA-MEM2
    1. <https://github.com/bwa-mem2/bwa-mem2>
13. Good SNP caller!!
    1. <https://luntergroup.github.io/octopus/>
14. Indel caller. IMSindel: An accurate intermediate-size indel detection tool incorporating de novo assembly (18 citations)
    1. <https://github.com/NCGG-MGC/IMSindel>
15. SNP+ Indel caller: NanoCaller
    1. <https://github.com/WGLab/NanoCaller>
16. clair3 is efficient in calling SNPs from Nanopore data
17. NanoSNP in inferior to clair3 but superior to other SNP callers for long reads when the coverage is low. NanoCaller is much inferior to clair3 when the coverage is low.
18. SequelTools is a program that provides a collection of tools for working with multiple SMRTcells of BAM format PacBio Sequel raw sequece data
19. SMRT Tools is PacBio’s official tools working with their reads

* Structural variant calling: the long and the short of it
* Comprehensive evaluation and characterization of short read general-purpose structural variant calling software
* NextSV3: automated structrual variation detection from long-read sequencing using state-of-the-art tools.
* Dysgu: efficient structural variant calling using short or long reads
* Jasmine and Iris: population-scale structural variant comparison and analysis

1. Population genetics
   1. genodive version 3.0
   2. ANGSD 0.94 stable
2. Demography, Site frequency spectrum
   1. <https://dadi.readthedocs.io/en/latest/>
   2. <http://cmpg.unibe.ch/software/fastsimcoal26/>
   3. https://speciationgenomics.github.io/easysfs/
3. Long read assemble.
   1. BlasR [76], MUMmmer [77], or Minimap2
   2. specialized methods to align long reads such as BLASR [76], Minimap2 [35], and NGMLR
4. cuteSV, LRcaller, Sniffles, SVJedi, and VaPoR
5. SV caller or workflow, Short-read alignment approach
   1. Manta (v1.1.0), DELLY (v0.7.7), LUMPY (v0.2.13) and GRIDSS (v1.3.4)
   2. DELLY [41]: the larger events remain hard
   3. TARDIS
   4. sv-callers: a highly portable parallel workflow for structural variant detection in whole-genome sequence data
6. SV caller or workflow, LOOOOOOONG-read mapping-based approach
   1. SVIM: structural variant identification using mapped long reads (newer)
   2. Sniffles operates on a per read base, also capable of reporting very low-frequency SVs in the sample.
   3. PBHoney: For PacBio, relying on BLASR alignments
   4. PacBio structural variant calling and analysis tools (PBSV) : For PacBio, SVs within the range of 20+ bp
   5. SMRT-SV: For PacBio, includes de novo assembly and a specialized genotyping module
   6. NanoSV: For Oxford Nanopore, preferentially uses as input an alignment from LAST. NanoSV reports only breakpoints (BND) which again makes the interpretation of the SVs type difficult.
   7. Overall, long-read mapping-based methods for SV calling often show a better performance than short-read ones
7. SV caller or workflow, De novo assembly-based approach: heterozygous SVs are often missed
   1. SGVar [32] is a more recent string graph-based (see Table 2 for definition) de novo assembly pipeline based on the SGA assembler [75] that also uses short-read sequencing data
   2. Assemblytics [34] is a web application that relies on MUMmer and identifies insertions and deletions
   3. paftools.js [35] uses Minimap2 alignments
   4. SMARTie-SV was recently introduced to detect insertions, deletions, and inversions, using BlasR
8. SV caller or workflow, Other approach
   1. Hic\_breakfinder (1+ Mbp), can potentially identify all types of SVs
   2. Strand-Seq, Strandseq-InvertR [68] (min ~ 1 kbp)
9. misc
   1. BEDOPS v2.4.41: [https://bedops.readthedocs.io/en/latest/index.html#](https://bedops.readthedocs.io/en/latest/index.html)
   2. GenomeScope: Estimate genome heterozygosity, repeat content, and size from sequencing reads using a kmer-based statistical approach.
   3. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes