Additional workflow steps in grey are methods, that would have been nessecary for other tools (read trimming for Flye assembler), that could have optimized our results (repeat masking) or further analyses (variant calling and read mapping). Unfortunately, we did not have time to perform them. Alternative tools for certain workflow steps are also indicated in grey.

Workflow Steps	Tools	Input	Output	Commands for Tools
DNA-Extraction	CTAB-based DNA extraction	Plant material	High molecular DNA	-
Library Preparation & Sequencing	Oxford Nanopore Technologies (SQK-LSK109, MinION)	High molecular DNA	Raw sequencing data in FAST5 format	-
Basecalling	Guppy, Bonito	Raw sequencing data in FAST5 format	Long reads in FASTQ format	guppy_basecaller -i /path_to_fast5_directory -s /path_to_ouput_directory --compress_fastq -x 'cuda:0' --flowcell flow_cell_type --kit sequencing_kit_type
Quality Control	Customized python script FASTQ_stats3_graphical_v0.2.py by Samuel Nestor Meckoni, https://codeberg.org/snmeckoni/scripts	Long reads in FASTQ format	Histogram of read length distribution	
Read Trimming	(read trimming was needed for Flye, but not for Shasta as read trimming is integrated)	Long reads in FASTQ format	Long reads in FASTQ format, filtered for >10 kb	
Assembly	Shasta, Canu, Flye	Long reads in FASTQ format	Contiqs in FASTA format	shasta-Linux-0.11.1 --input /fastq_file --assemblyDirectory /output_directory --config /config_file

Quality Control	BUSCO, stats, QUAST, NLR	Contiqs in FASTA format	Percentage of BUSCOs found	busco --cpu number of cpu --download_path /path_to_busco_database -l /database_for_busco_search -i /path_to_assembly_fasta -o /output_directory -m mode/sequence_type -f (overwriting existing files) >> /path_to_log_file 2>&1 & (errors into log file)
Repeat Masking	RepeatMasker			RepeatMasker -e hmmer (engine) -pa 5 (no. of processors) -qq (rush job, less sensitive) -species arabidopsis (nearest model sp.) -no_is (no check for bacterial insertions) -dir /output_directory -small (type of masking) /path_to_assembly_fasta > /path_to_log_file &
Struct. Annotation	Augustus, GeMoMa, BRAKER3	Contiqs in FASTA format	mRNA, CDS, PEP sequences in FASTA and annotation GFF3 format	augustus --gff3=on --UTR=on --uniqueGeneId=true --protein=on --cds=on --species=arabidopsis /path_to_assembly_fasta > /path_to_output_gff3_file 2> /path_to_log_file &

Funct. Annotation	MYB annotator, KIPES, BLAST, InterProScan5, KEGG	Peptide sequences from structural annotation	Candidate sequences	python3 KIPEs3.py --baits /path_to_baits_directory --positions /path_to_residuals_directory --out /output_directory --subject /fasta_from_structural_annotation --seqtype type_of_sequence python3 MYB_annotator.py --baits /path_to_baits_fasta --info /path_to_baits_txt --out /output_directory --subject /fasta_from_structural_annotation --fasttree /path_to_FastTree
Variant Calling				
Read mapping				