Application of next generation sequencing technologies in gene discovery in *Lolium perenne*

Mallikarjuna Rao Kovi

Norwegian University of Life Sciences, Ås, Norway

«EditGrass4Food» annual meeting

TalTech, Tallinn, 26th October, 2022

“EditGrass4Food”, ID No EEA-RESEARCH-64, Contract No EEZ/BPP/VIAA/2021/4 is financially supported by European Economic Area (EEA) grants

Outline

I. *Lolium perenne* NGS technologies for gene discovery

- *Lolium perenne* importance
- Evolution of sequencing technologies
- Third generation sequencing technologies, ONT and PacBIO
- Applications of these technologies: Genome sequence; SNP marker development (GBS), Transcriptomics
- Genomic resources in *Lolium perenne*

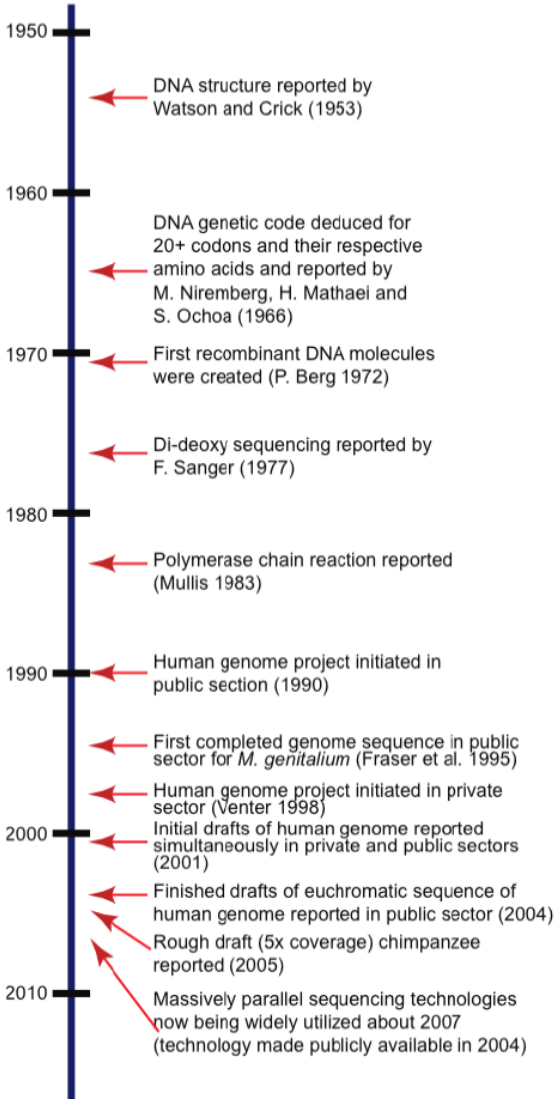
II. How we are implementing NGS in '*Editgrass4food*' project

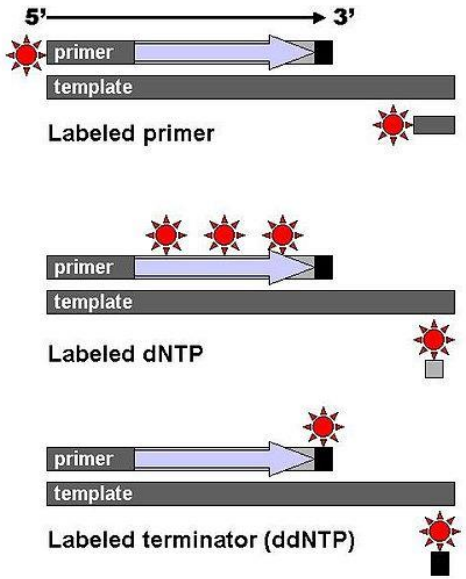
- Target sequencing for gene characterization
- Identification of genes and their expression for frost and drought by RNAseq.

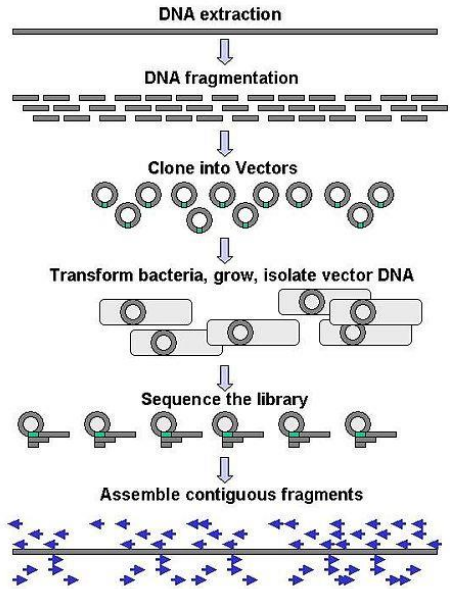
Lolium perenne

- Major forage crop for pasture and silage.
- High nutrition.
- Widely cultivated in temperate regions.
- Economic value

History of DNA sequencing



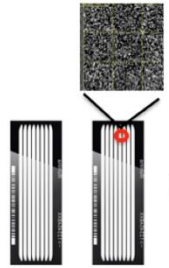




Illumina HiSeq2000

- 2 flow cells (can be run independently)
- Up to 320Gb mapped sequence per FC
- 64Gb sequence per day (2 flow cells)

c-Bot
(automated cluster generation)



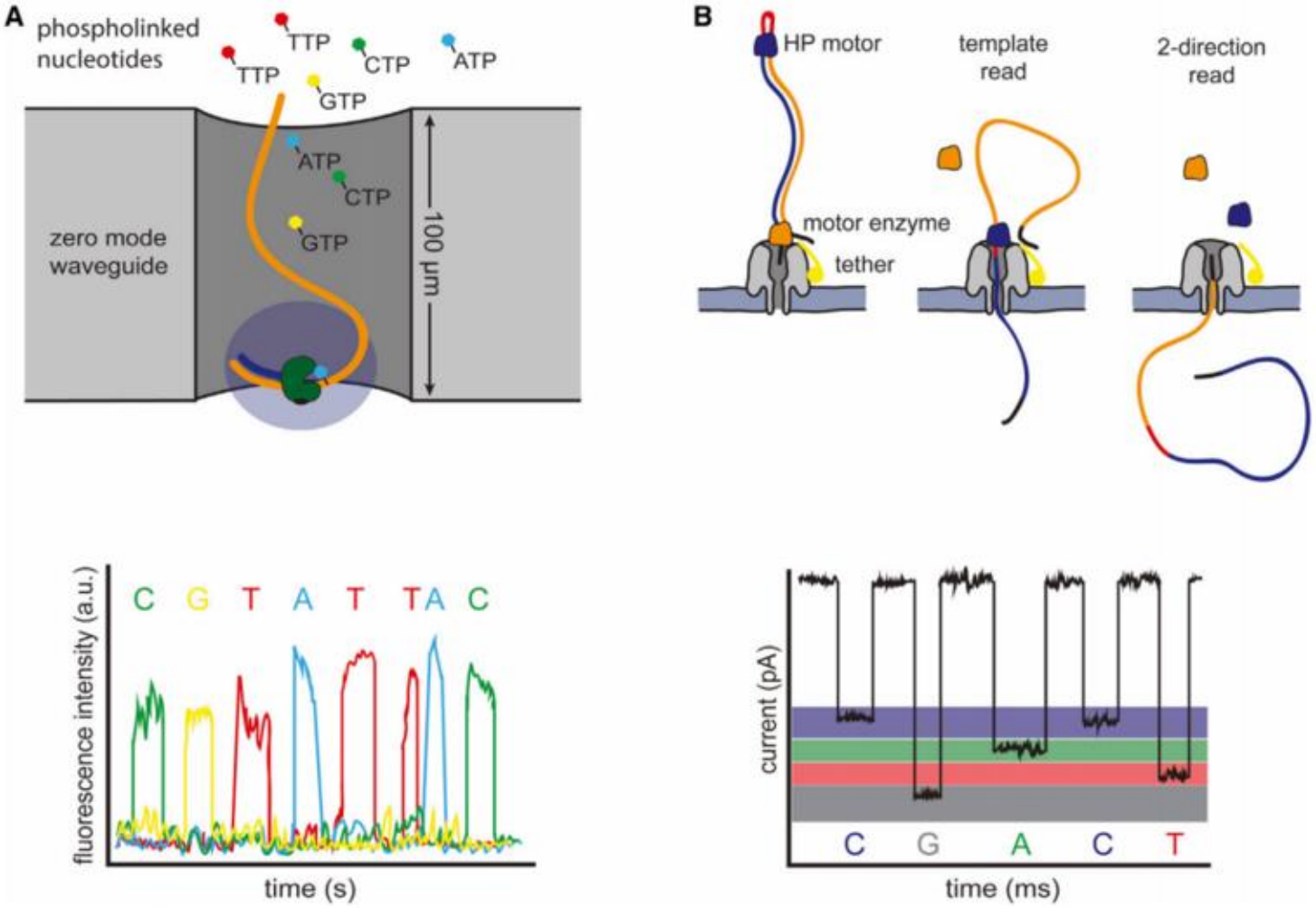
Cluster density
750-850/mm²

HiSeq Flow Cells

Evolution of sequencing technologies

- First generation sequencing
  - ➢Capillary sequencing technology : ABI Prism 3700
- Second generation sequencing
  - ➢Bridge PCR and emulsion PCR technology: Illumina and Ion Torrent
- Third generation sequencing
  - ➢Single molecule real time sequencing technology: PacBio and Oxford nanopore technology (MinION and promethION)

Single molecule real time sequencing (SMRT)



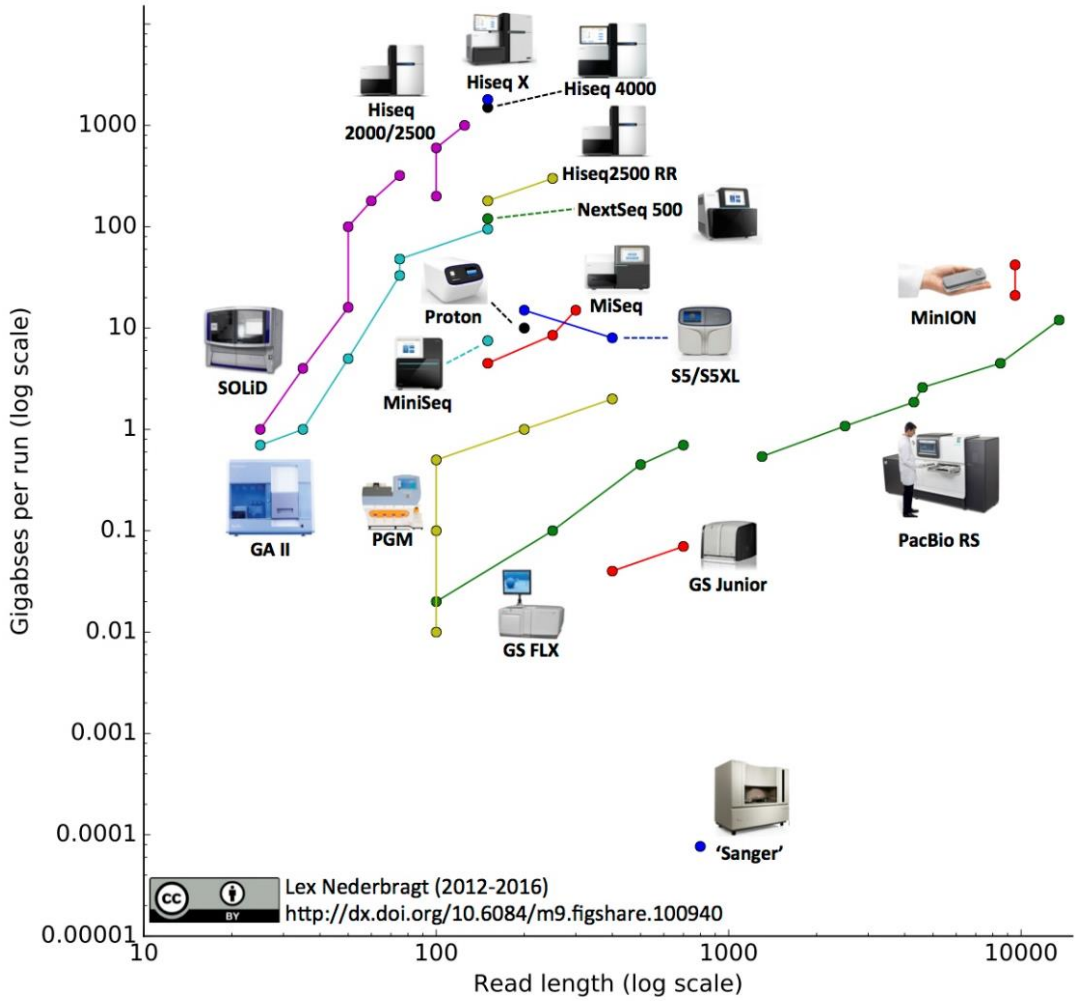
(A) Pacific Bioscience's SMRT sequencing. A single polymerase is positioned at the bottom of a ZMW. Phosphate-labeled versions of all four nucleotides are present, allowing continuous polymerization of a DNA template. Base incorporation increases the residence time of the nucleotide in the ZMW, resulting in a detectable fluorescent signal that is captured in a video.
(B) Oxford Nanopore's sequencing strategy. DNA templates are ligated with two adapters. The first adaptor is bound with a motor enzyme as well as a tether, whereas the second adaptor is a hairpin oligo that is bound by the HP motor protein. Changes in current that are induced as the nucleotides pass through the pore are used to discriminate bases. The library design allows sequencing of both strands of DNA from a single molecule (two-direction reads).

PacBIO

MinION

Iceland Liechtenstein Norway grants

EDIT GRASS 4FOOD



PacBIO Revio

15× increase in HiFi read throughput

With a high-density, 25 million ZMW SMRT Cell, up to 4 SMRT Cells in parallel, and 24-hour run times, the Revio system delivers 360 Gb of HiFi reads per day, equivalent to 1,300 human whole genomes per year.

The $1,000 complete, phased genome

HiFi sequencing provides small variants, structural variants, repeat expansions, methylation, and haplotype phasing from a single library and sequencing run.

50% fewer consumables and a vastly simplified experience

Operating the Revio system is simple, thanks to a combined reagent and sample plate, rapid run setup, flexibility to queue runs while sequencing is in progress, and no external nitrogen supply.

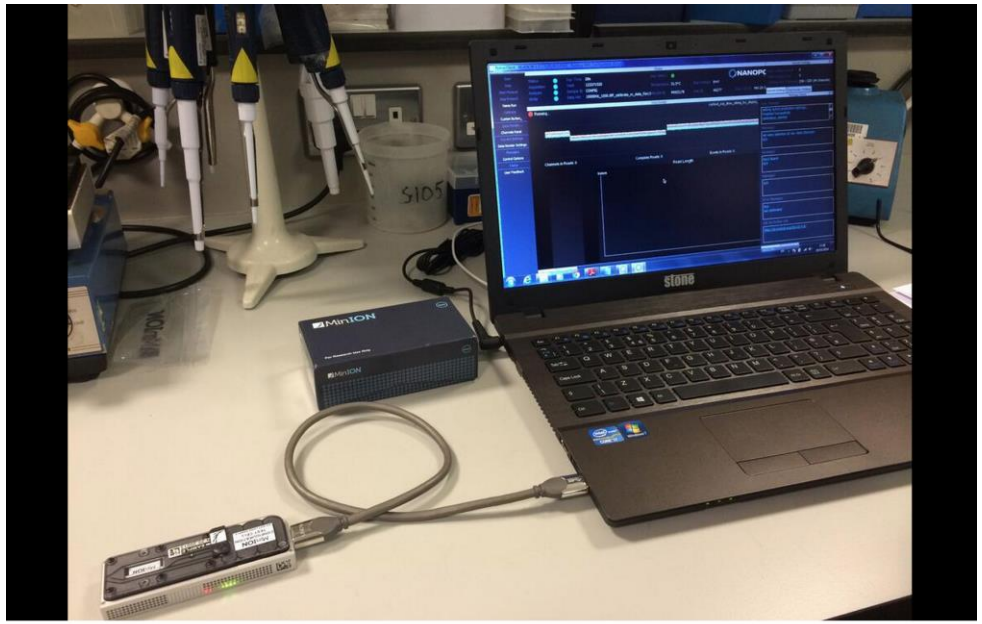
Iceland Liechtenstein Norway grants

EDIT GRASS 4FOOD

Oxford Nanopore MinION

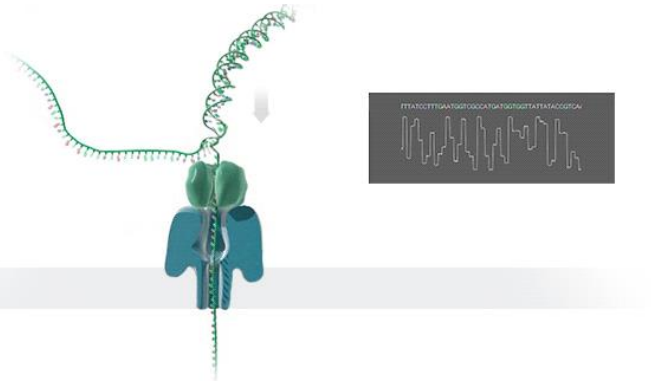
Unique properties:

- Real-time
- Portable
- Long reads



Iceland Liechtenstein Norway grants

EDIT GRASS 4FOOD



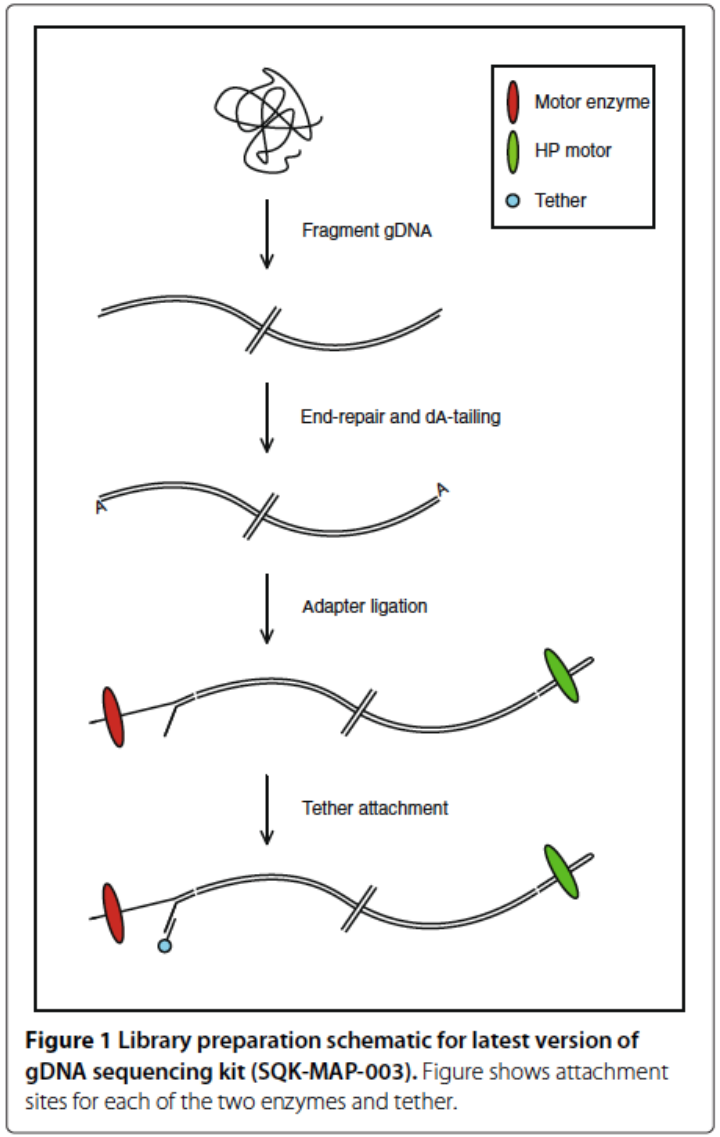
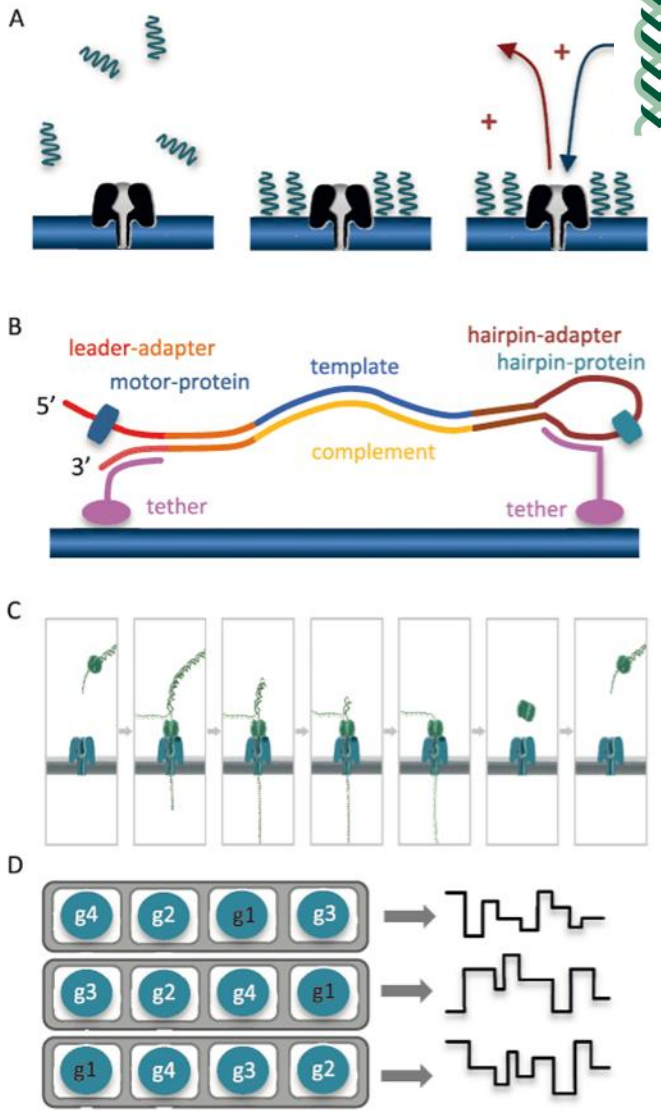


Figure 1 Library preparation schematic for latest version of gDNA sequencing kit (SQK-MAP-003). Figure shows attachment sites for each of the two enzymes and tether.



Library prep time:
92 minutes

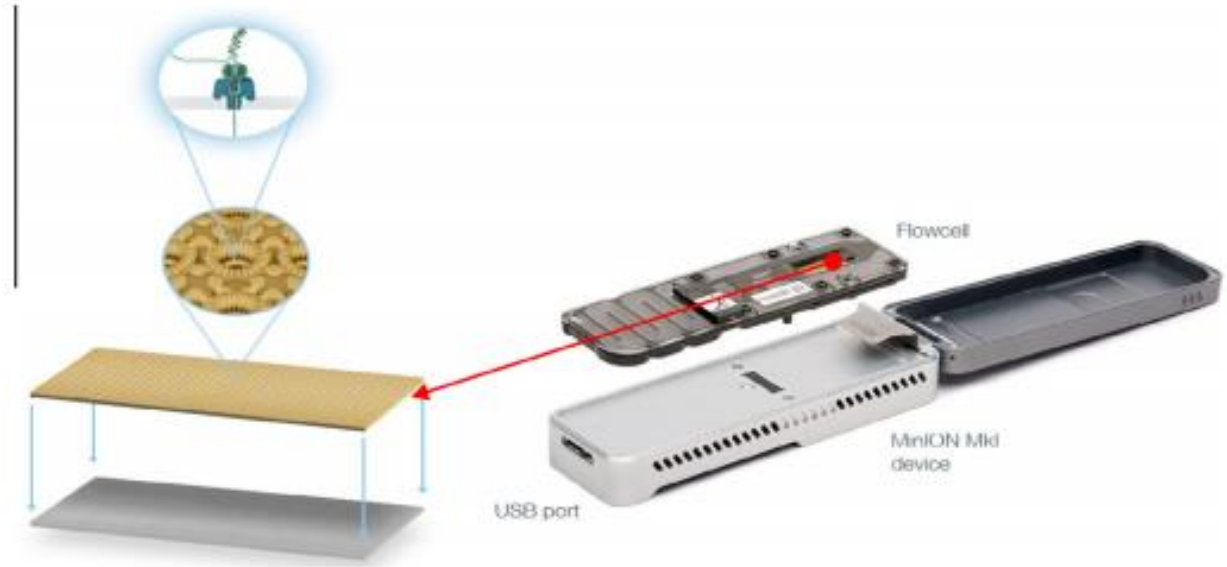


Figure 1 The MinION sequencing device

DNA sequencing is performed by adding the sample to the flowcell. When DNA molecules pass through or near the nanopore, there will be a change in the magnitude of the current in the nanopore, which is measured by a sensor. The data streams are passed to the ASIC and MinKNOW, the software that generates the signal-level data. ASIC, application-specific integrated circuit.

Future ☺

Applications

- Genome sequence
- Transcriptome sequence
- Genotyping by sequencing

Iceland Liechtenstein Norway grants

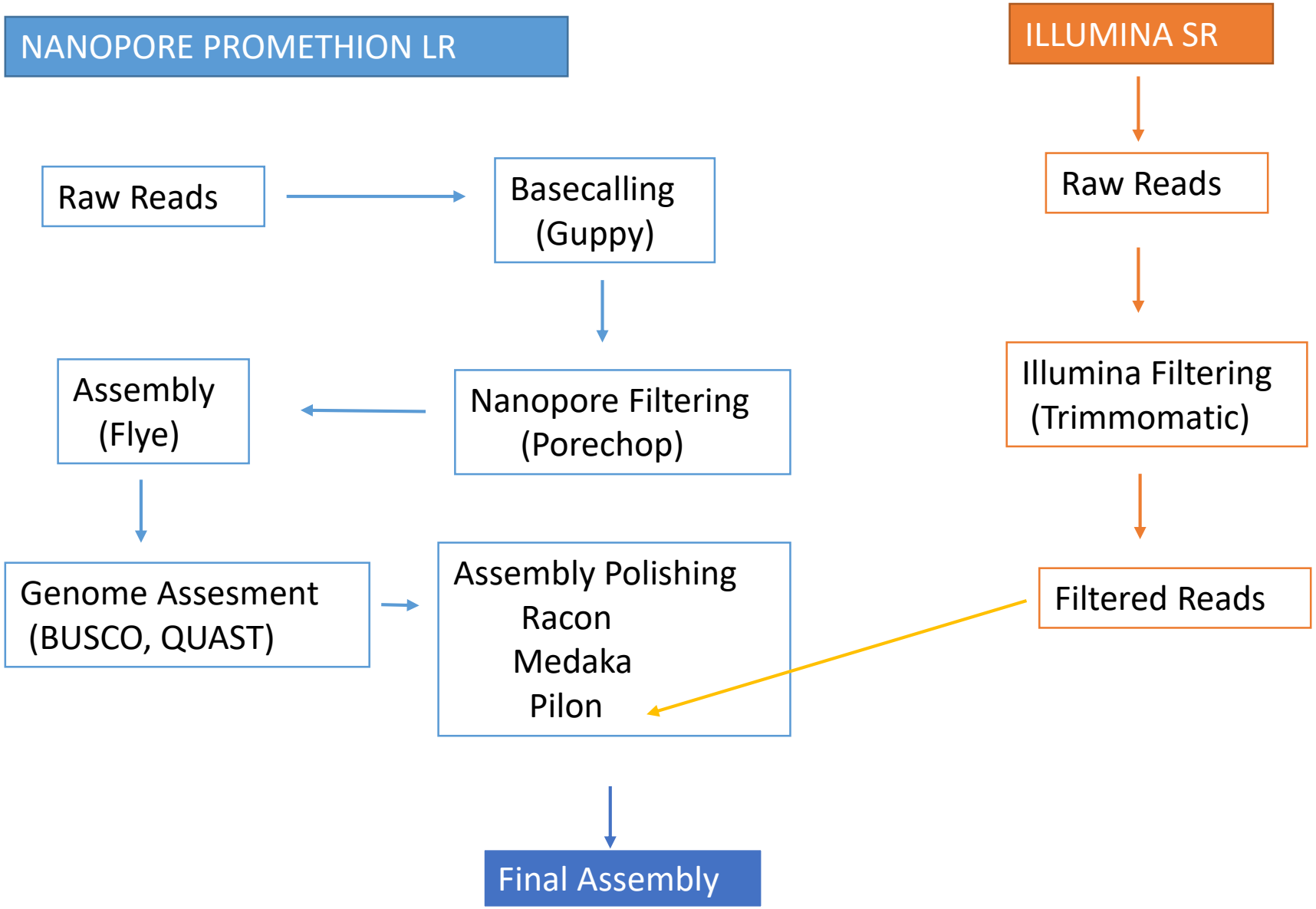
Lolium perenne Genome sequence

EDIT GRASS 4FOOD

Statistics of the *Lolium perenne* Kyuss Genome Assembly and Comparison with Other Public Ryegrass Assemblies

	Kyuss[a] *Lolium perenne*	P226/135/16 *Lolium perenne*	Rabiosa[a] *Lolium multiflorum*	M2289 *Lolium multiflorum*
Reference	This study	Byrne et al. (2015)	Copetti et al. (2021)	Knorst et al. (2019)
Est. genome size (Gb)	2.720	2.068	2.464	2.500
Assembly size (Gb)	2.281	1.128	4.531	0.585
% of genome assembled	83.9	54.6	183.9	23.4
# of sequences	1,935	48,415	226,949	129,579
N50 (kb)	11,276	70	2,941	5
N90 (kb)	3,320	14	283	2
L50 (#)	65	4,908	443	37,162
L90 (#)	209	16,951	1,984	103,446

Lolium perenne Genome sequence



Lolium perenne Genome sequence

Nagy *et al. BMC Genomics* (2022) 23:505
https://doi.org/10.1186/s12864-022-08697-0

BMC Genomics

RESEARCH Open Access

Chromosome-scale assembly and annotation of the perennial ryegrass genome

Istvan Nagy[1*], Elisabeth Veeckman[2,3,4], Chang Liu[5,6], Michiel Van Bel[3,7,8], Klaas Vandepoele[3,7,8], Christian Sig Jensen[9], Tom Ruttink[2] and Torben Asp[1]

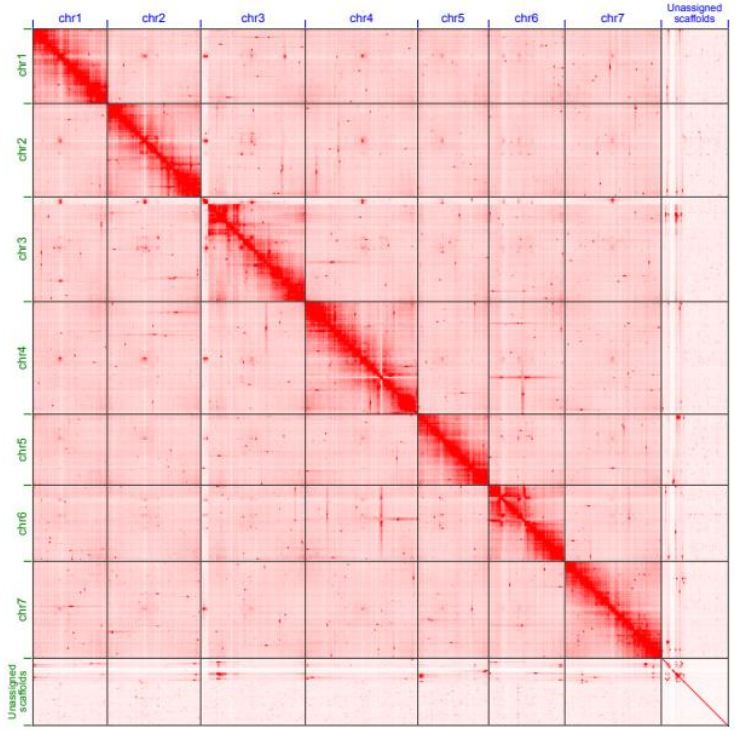


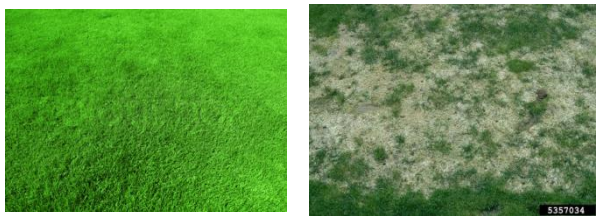
Figure S1 Hi-C contact map with Lolium_2.6.1 reference sequences

https://ryegrassgenome.ghpc.au.dk/

https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v5_monocots/

Transcriptome sequence

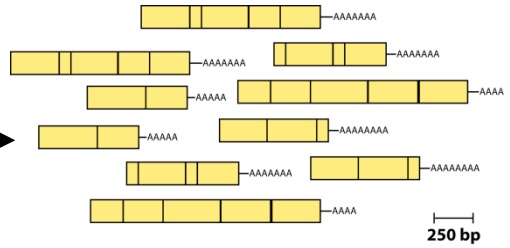
Samples of interest



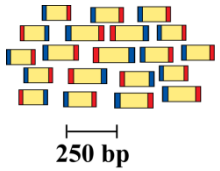
Condition 1 (healthy grass)

Condition 2 (Infected grass)

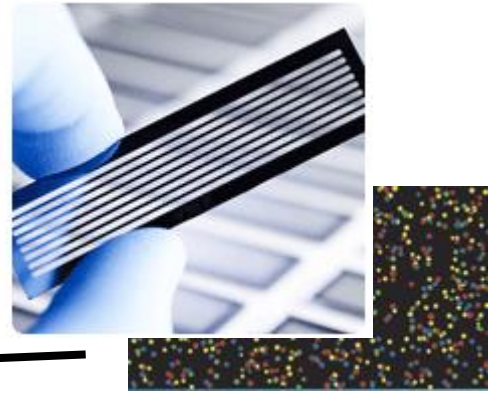
Isolate RNAs



Generate cDNA, fragment, size select, add linkers

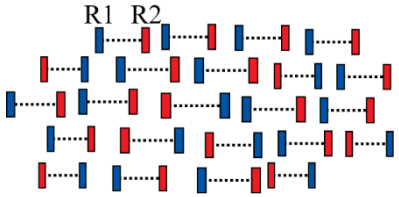


Sequence ends



100s of millions of paired reads
10s of billions bases of sequence

Paired-end reads (e.g. 2 x 100 bp)



Map to genome, transcriptome, and predicted exon junctions

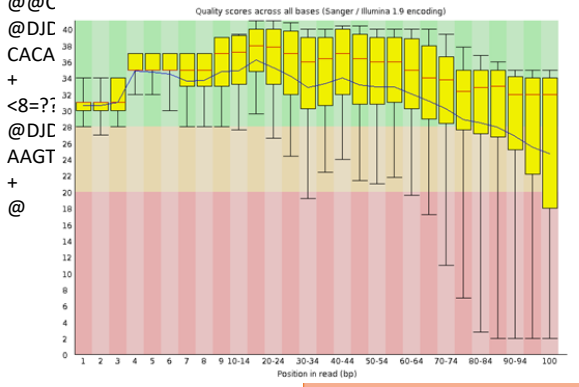
Downstream analysis

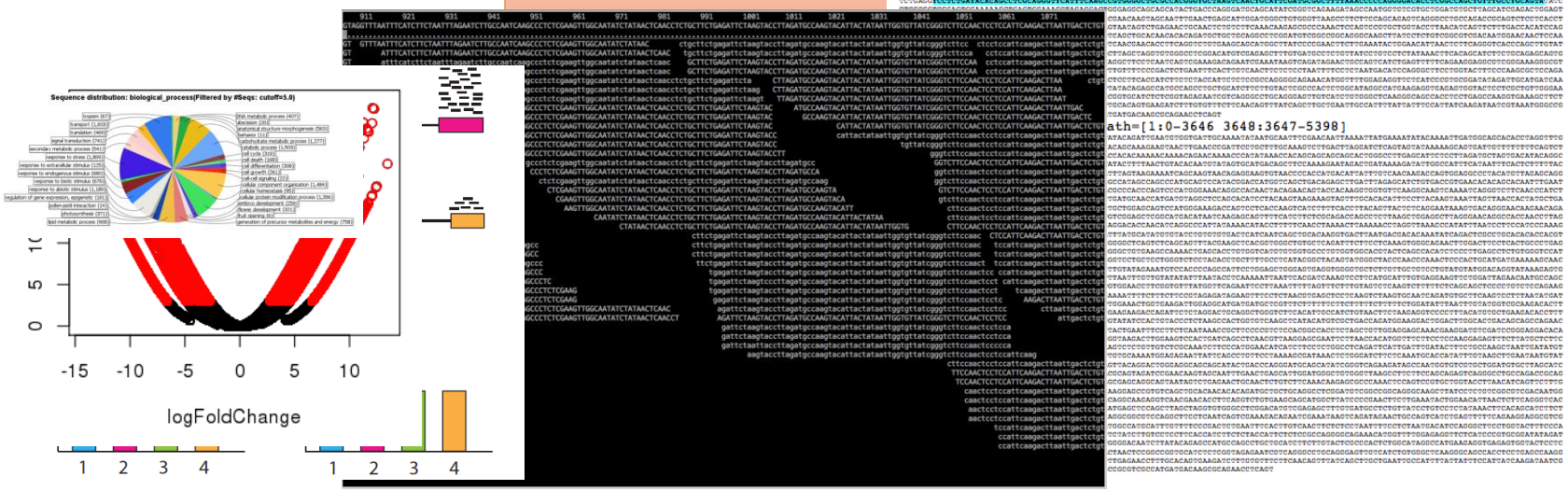
RNA seq. analysis pipeline

Millions of reads

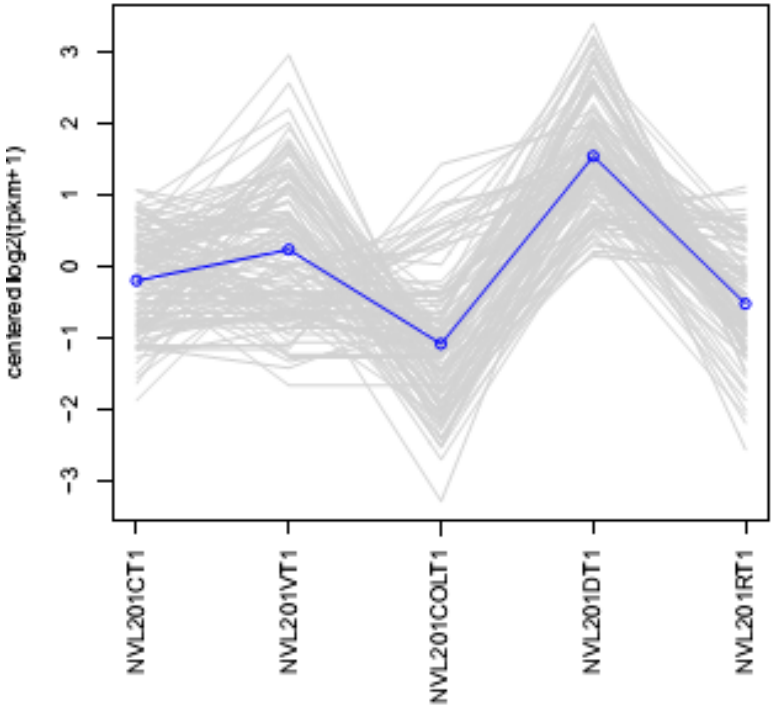
Quality control
FastQC

Trimmed reads

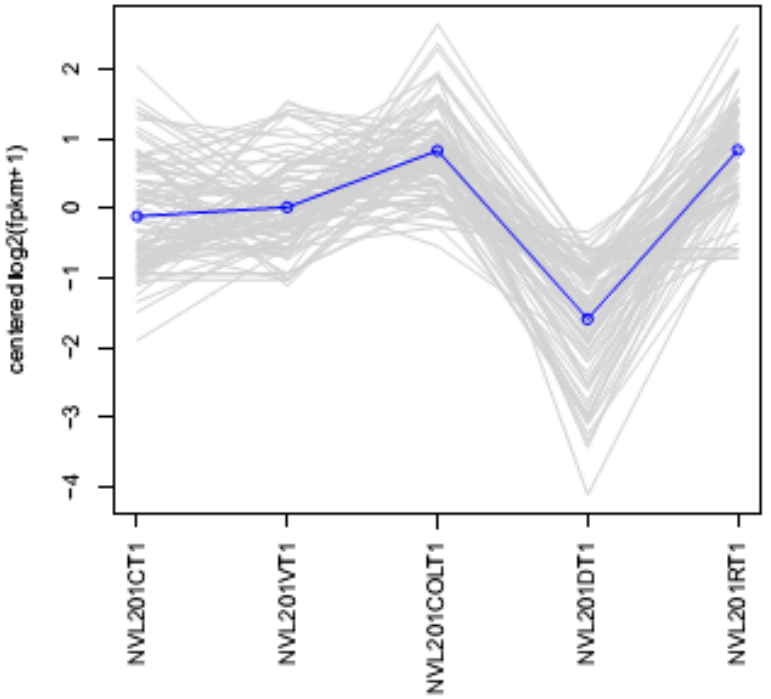




Subclusters for Lolium 201(V-) in vernalization treatment

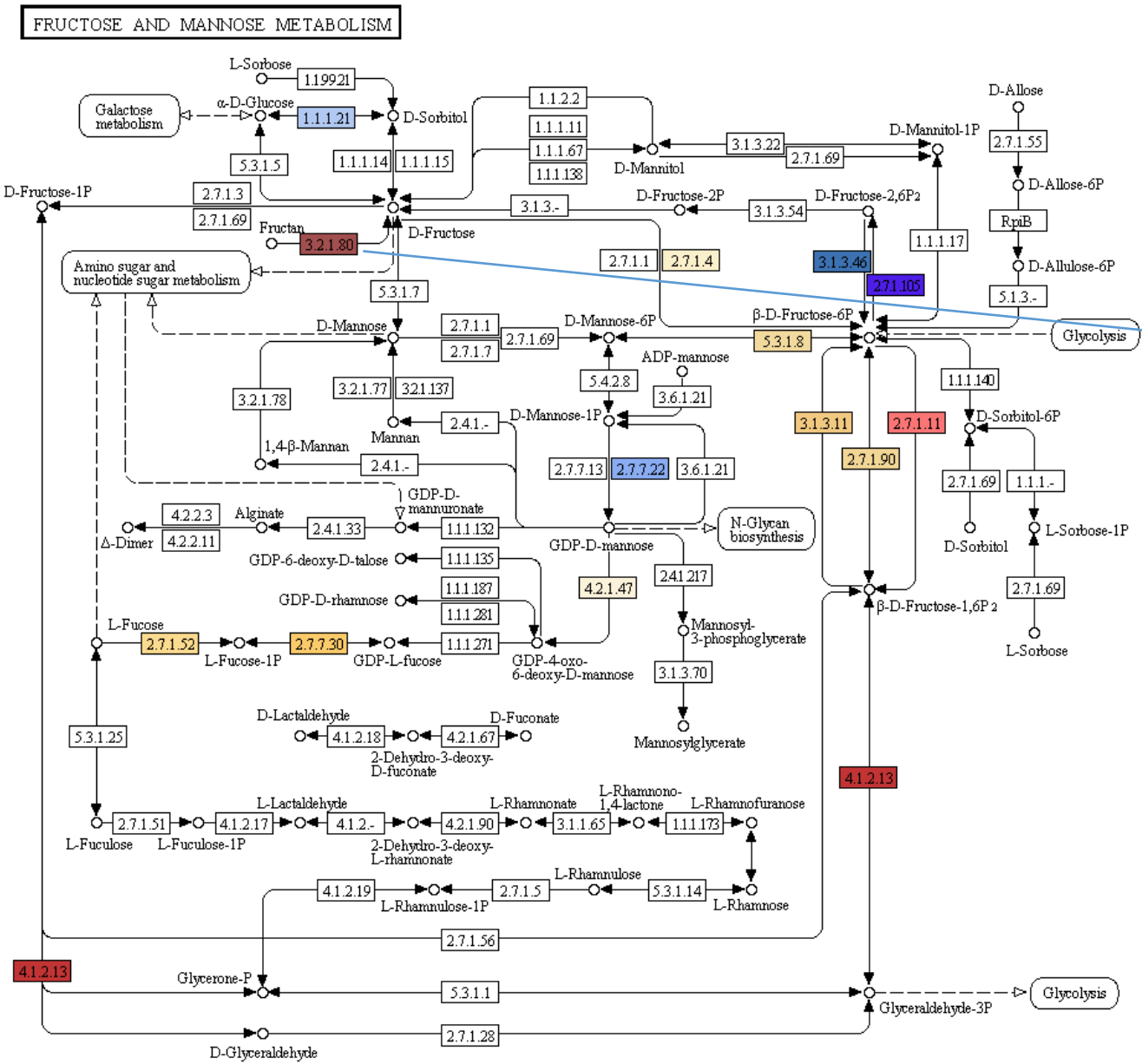


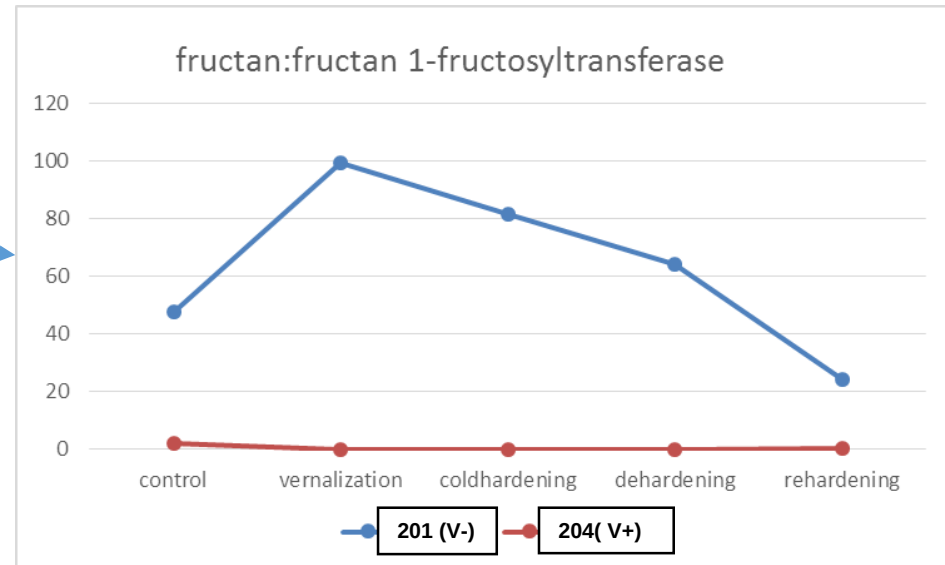
Calmodulin binding protein
Calreticulin2
Zinc finger (C2H2) family proteins
Glutathione-s-transferase
homeobox.-leucine zipper protein
Pyroline 5 carboxylate synthase



Photosnthesis related genes encoding D2 subunits of PSII complex
AP2/EREB transcription factors
Arabinogalactan
Chlorophyll a-b binding protein

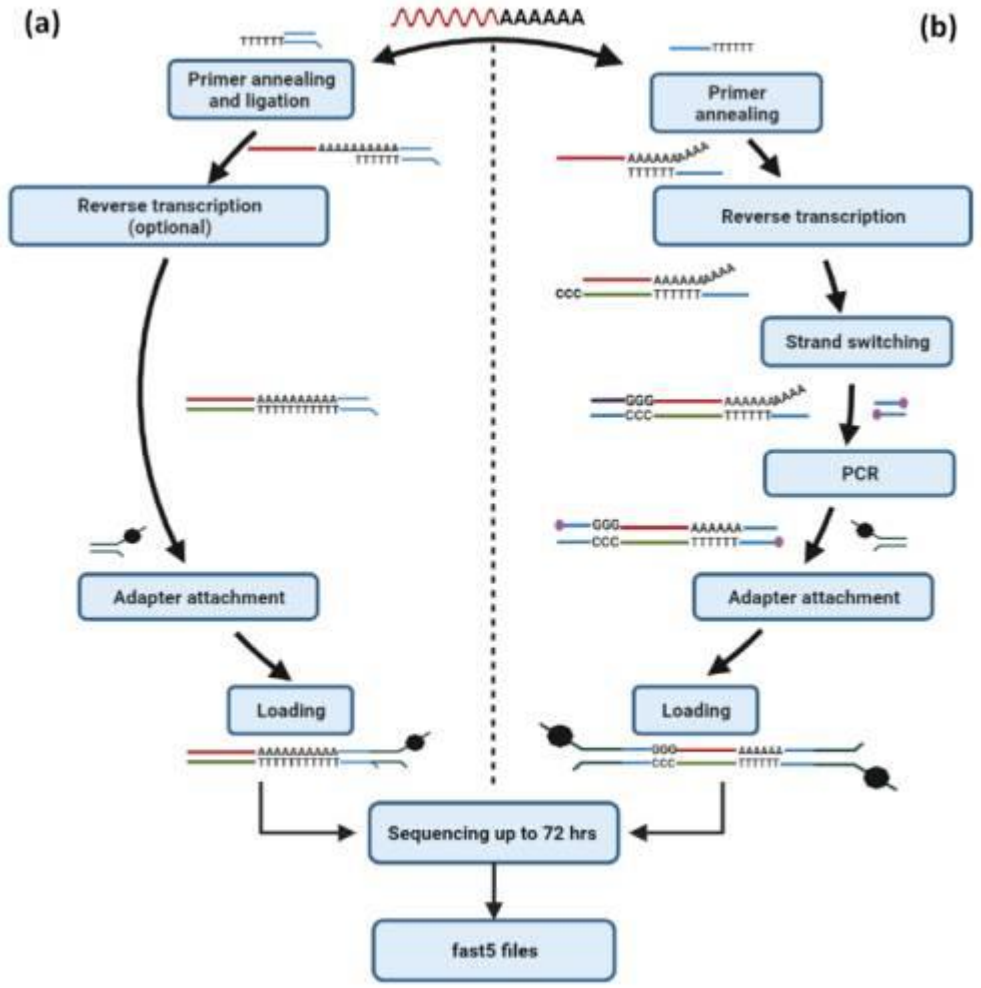
Genes involved in fructose metabolic pathway





Direct RNA sequencing

- Direct RNA sequencing — unbiased, fulllength transcript and base modification analysis

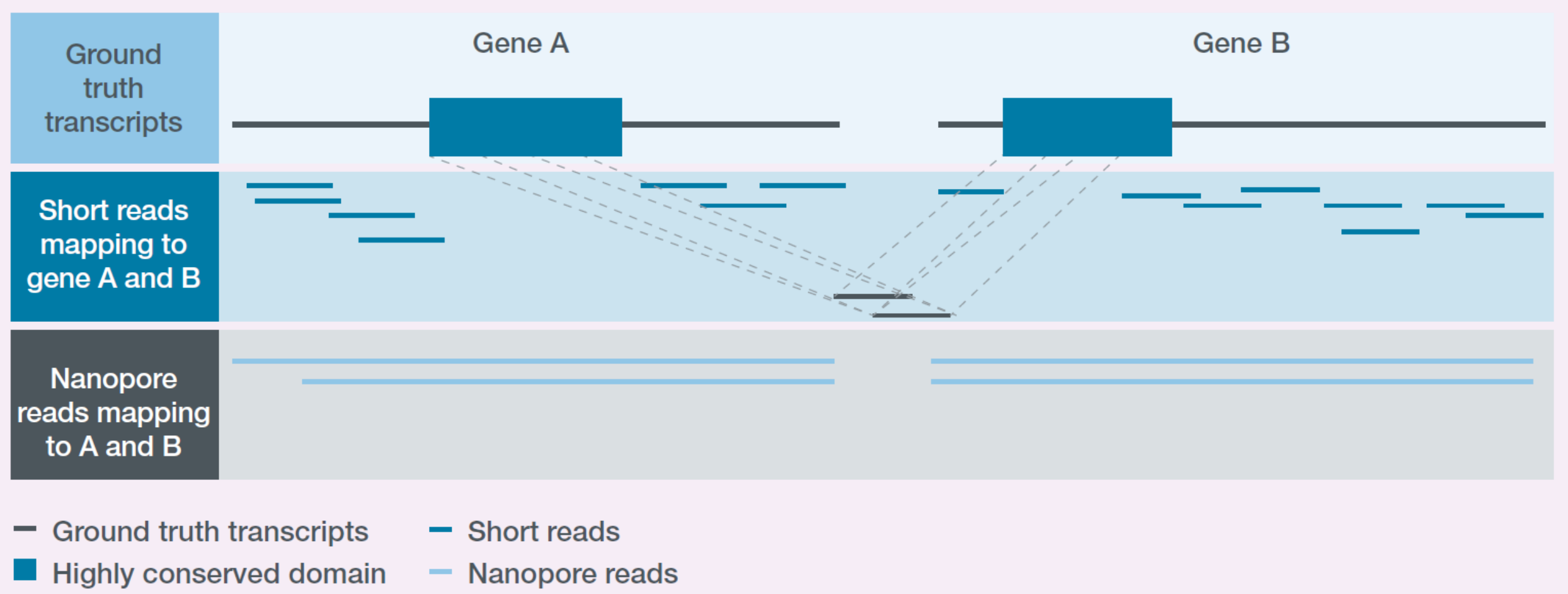


Nanopore website

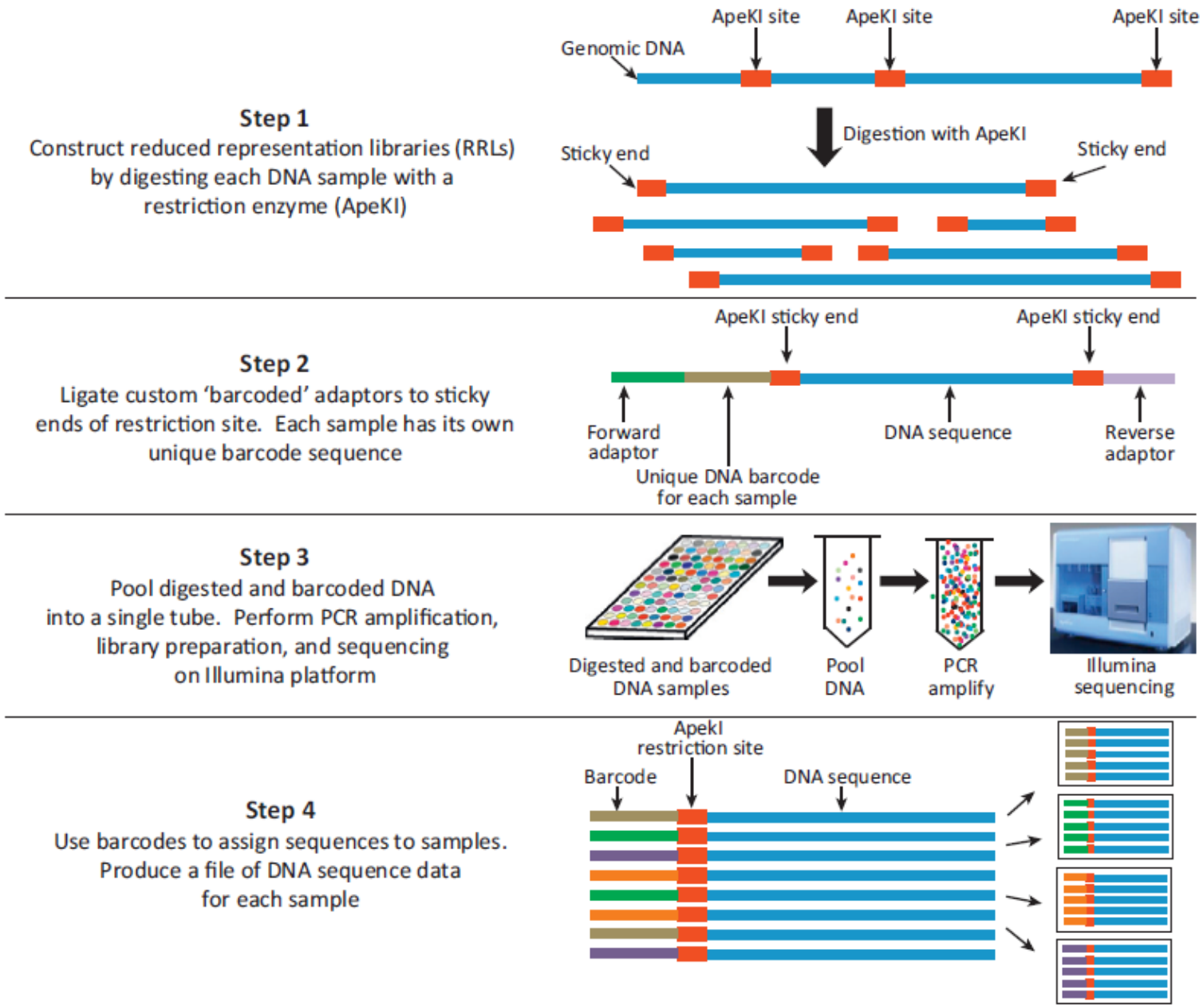
Iceland Liechtenstein Norway grants

EDIT GRASS 4FOOD

Assembly of full-length transcripts with short reads and long nanopore sequencing reads

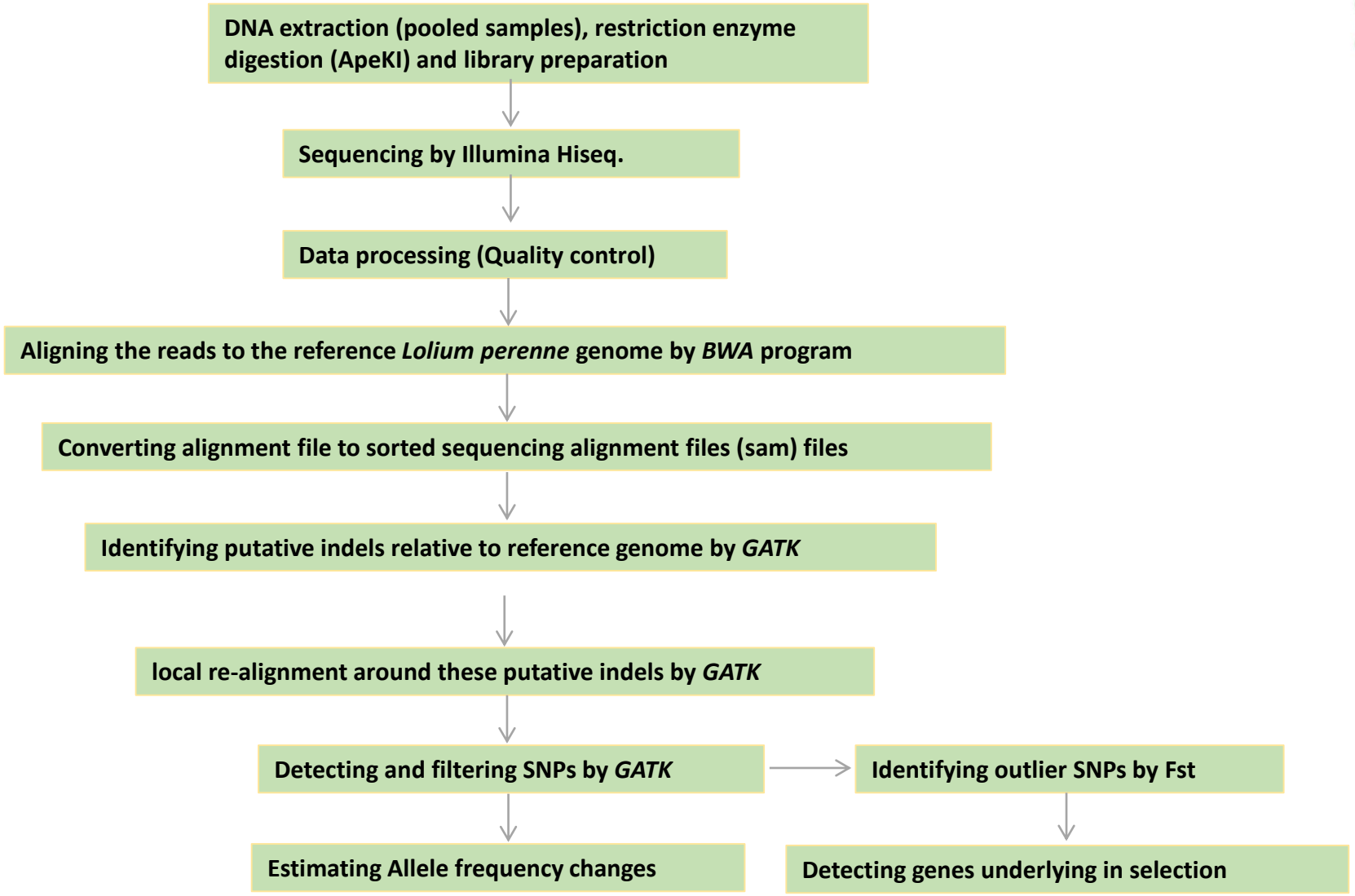


Genotyping by sequencing



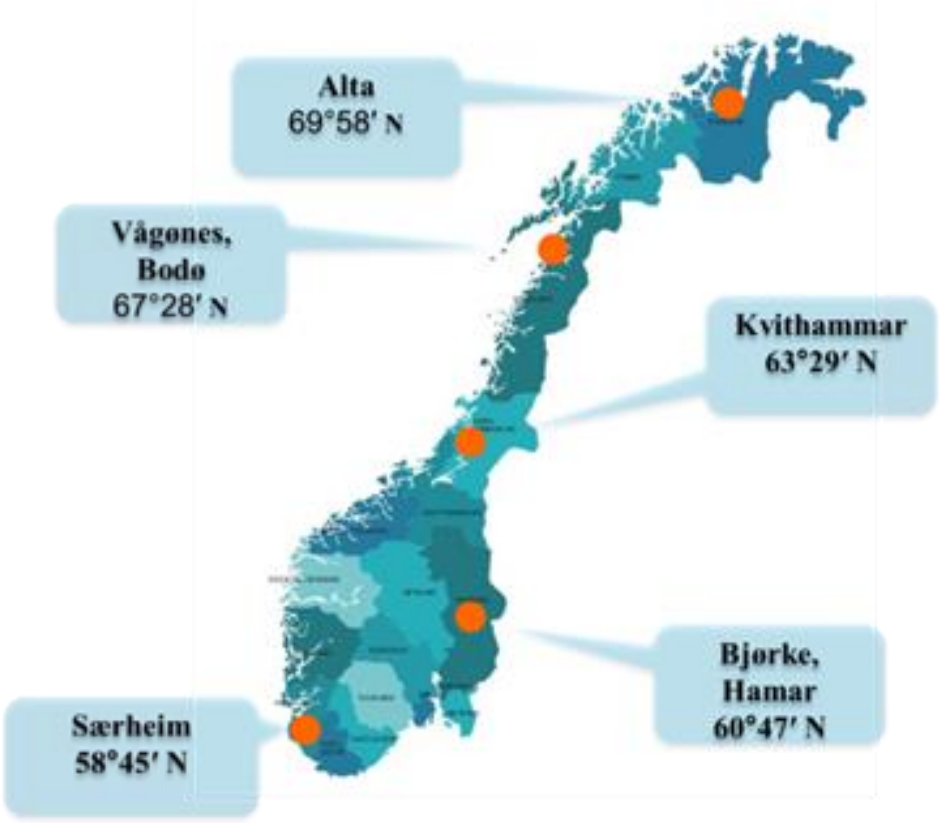
Elshire et al.

Genotyping by Sequencing

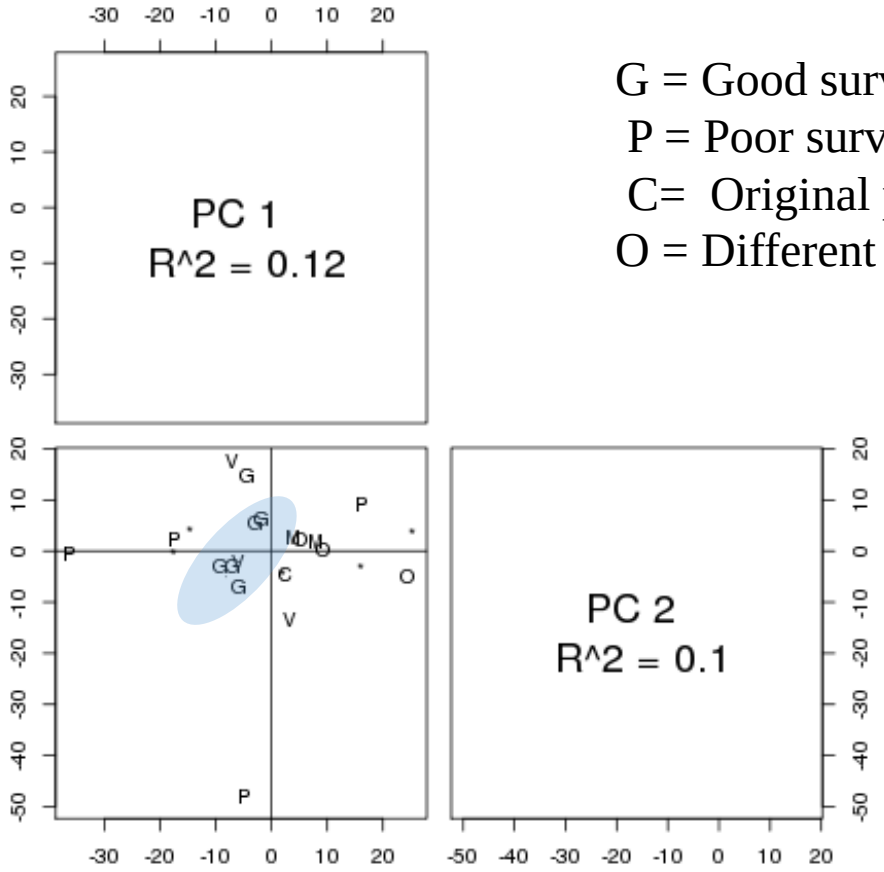


Selection studies

- Differential selection in the different climates at the 5 locations will generate changes in genetic composition
- Sampling for 3 years and phenotype
- Explore associations between allelic/haplotype shifts and location specific (climate) changes in phenotype for utilization in breeding in perennial ryegrass.
- Methods:
  - Genotyping by sequencing



Distinguishing survival ability groups based on GWAFFs



G = Good survival
P = Poor survival
C= Original population
O = Different combination mix of 5 varieties

Based on genome wide allele frequencies at 101375 SNP positions.

Iceland Liechtenstein Norway grants

EDIT GRASS 4FOOD

Implementing NGS technologies in EditGrass4Food project for gene discovery and characterization

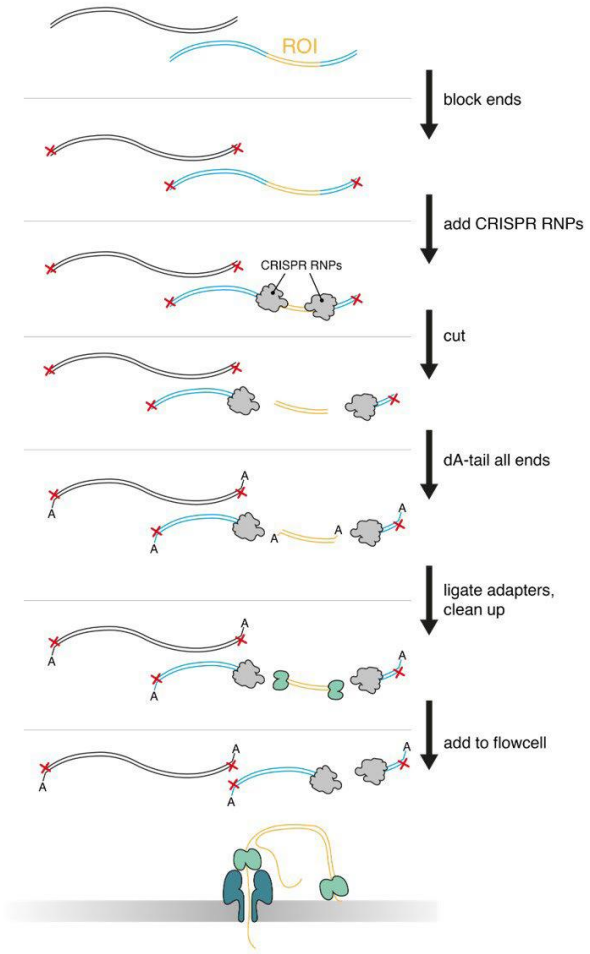
Target sequencing

- 10 genes (vernalization, freezing tolerance and fructan biosynthesis genes) involved in freezing tolerance will be selected based on previous transcriptome studies (VARCLIM project)

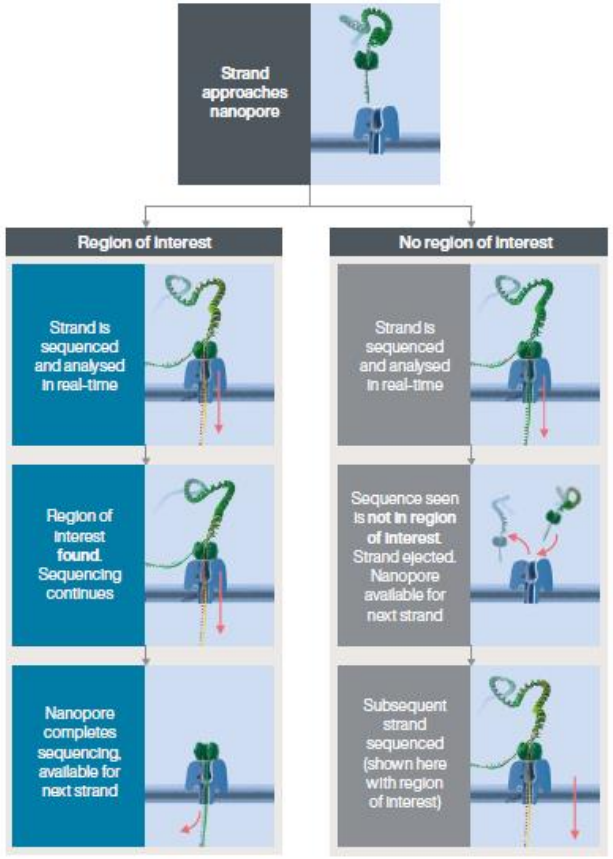
- 10 genes, responsible for leaf growth under water deficit conditions (GrowGene project)

Target sequence by oxford nanopore sequence technology

Cas9 targeted sequencing



Adaptive sampling sequencing



Transcriptome regulation of freezing and drought tolerance in perennial ryegrass.

- Two freezing tolerant and two susceptible genotypes will be selected and grown at short days (8 h) at 18°C to gain biomass. The plants will be pre-acclimated for two weeks at 4°C before subjected to freezing at -8 or -12°C.

- Freezing tests will be performed in growth chambers at LAMMC

- Samples of leaf tissue of each genotype will be taken for RNA extraction at the day before stress onset and 8 hours after onset of low temperatures.

Take home messages

- Sequencing technologies evolved so fast, that we need to update to tackle the enoromous data and make sense out of it.
- PacBio HiFi sequencing is recommended over Oxford nanopore for phased genomes in out crossing species
- Linux and R Programming can boost your bioinformatics skills
- Dont panic of data handling (bioinformatics). Its easy to learn ☺

Iceland Liechtenstein Norway grants

EDIT GRASS 4FOOD