Sequencing *Linum usitatissimum* L. genomic DNA extracted from nuclei

Dvorianinova E.M.[1, 2*], Bolsheva N.L.[1], Rozhmina T.A.[1, 3], Sigova E.A.[1, 2], Pushkova E.N.[1], Melnikova N.V.[1], Dmitriev A.A.[1]

[1] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, Moscow, Russia*

[3] *Federal Research Center for Bast Fiber Crops, Torzhok, Russia*

* *dvorianinova.em@phystech.edu*

The work was financially supported by the Ministry of Science and Higher Education of the Russian Federation, grant number 075-15-2021-1064

Research objective

- **Contiguous plant genomes** are indispensable for various genetic applications (e.g. gene annotation, complex region resolution).
- The **Oxford Nanopore reads** can significantly improve assembly contiguity.
- The length of the sequenced ONT reads hinges on the employed **DNA isolation method**.
- Using **conventional methods** for the *Linum usitatissimum* L. DNA leads to inadequate DNA quality.
- The **nuclear membrane** preserves high-molecular-weight DNA, which is crucial for successful Nanopore sequencing.

To obtain pure long flax DNA, we **aimed** to create a protocol based on the nuclei isolation method

Materials and methods

- kept young shoots of the *L. usitatissimum* line #3896 in the dark for 1 week
- collected leaves
- isolated flax nuclei using a density-gradient method *
- extracted the nuclear DNA

DNA extraction

- prepared libraries (SQK-LSK-109)
- sequenced the DNA on a MinION instrument (R9.4.1)

sequencing

- basecalled fast5-files with Guppy 5.0.11 (dna_r9.4.1_450bps_sup.cfg)
- removed adapters with Porechop 0.2.4
- assembled raw genomes with Canu 2.2, Flye 2.9, NextDenovo 2.5.0, Raven 1.8.1, Shasta 0.10.0, wtdbg2 2.5.
- polished the raw assemblies with Racon 1.4.10, Medaka1.0.1/1.5.0, nextpolish 0.13.2, pepper 0.1.1
- calculated methylation frequencies with Megalodon 2.3.5 and megalodon.sh from METEORE

data processing

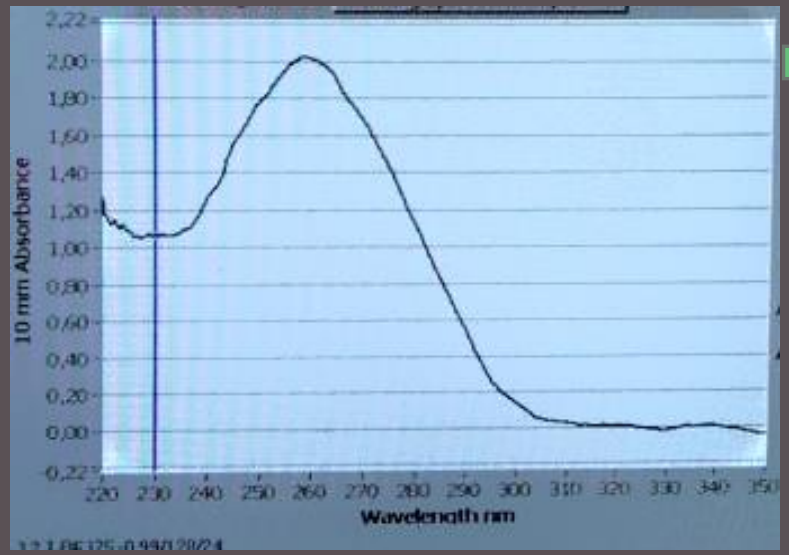
*- nuclei isolation:

Results

1.

$C_{Nanodrop}$=100 ng/µl

C_{Qubit}=85 ng/µl



A260/280= 1.8 A260/230= 1.9

2.

Raw data, 2 MinION runs

3.

Raw assemblies

Assembler	QUAST				BUSCO	
	length, Mb	contigs	N50, Mb	L50	compl., %	dupl., %
Canu	447.4	1728	6.2	26	93.3	59.8
Flye	345.1	7720	0.3	179	91.8	48.0
NextDenovo	289.5	248	3.1	26	91.1	44.5
Raven	269.3	1722	0.2	328	89.7	29.8
Shasta	372.5	6952	1.5	67	93.2	57.9
wtdbg2	243.7	3678	0.2	229	74.5	6.9

4.

The optimal polishing scheme was Racon (2 iterations) + Medaka

Final assembly

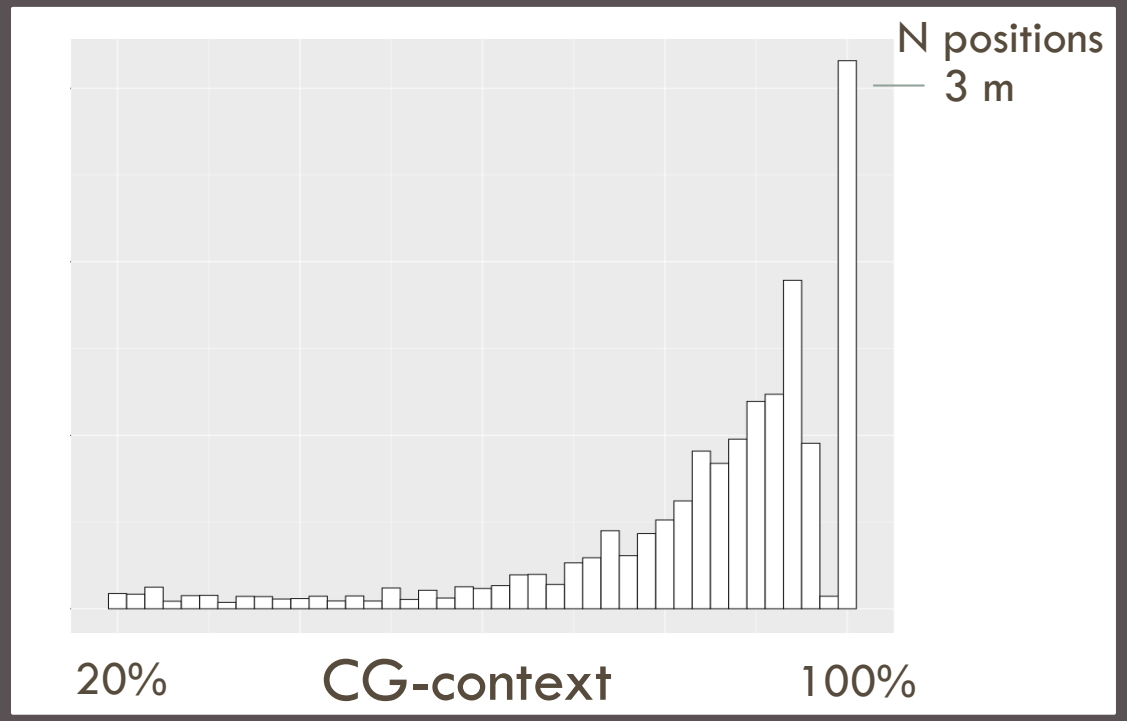
~447 Mb

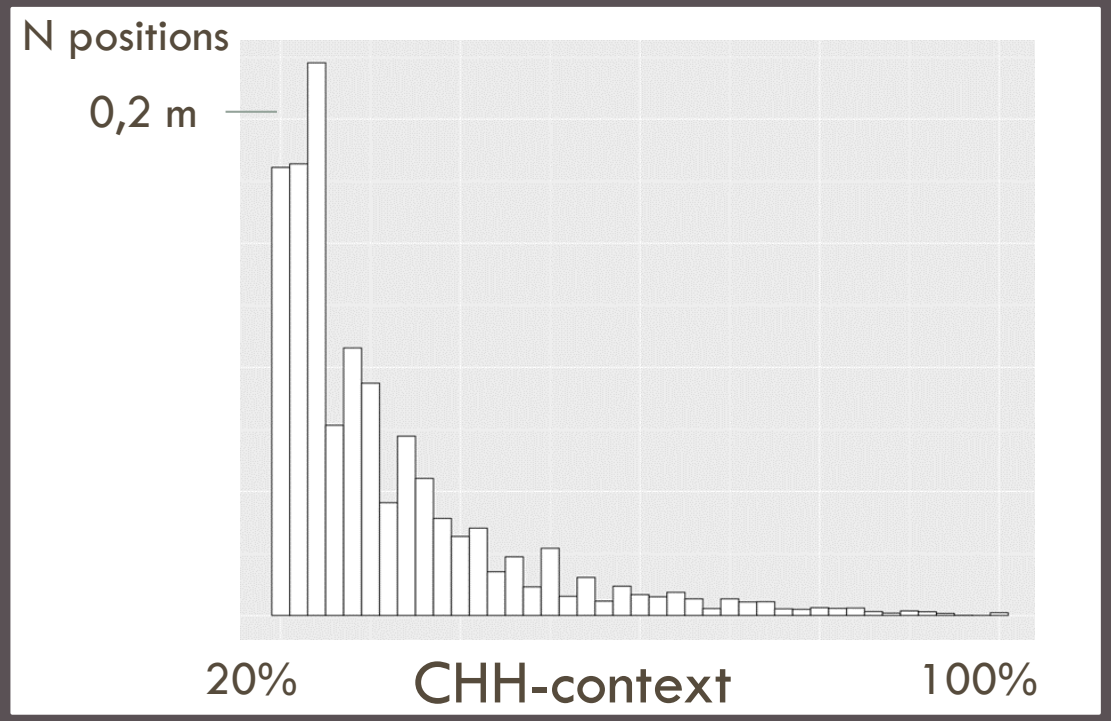
93.8% completeness (eudictos_odb10)

N50=6.2 Mb

Results

Methylation frequency distribution for cytosines with ≥ 10x coverage:





Methylation context	CG	CHG	CHH
Context abundance, % of the called CN sites	17.1	11.9	71.1
Percentage of sites with high methylation levels (≥ 50%)	53.8	2.8	0.02

Conclusions

- The developed protocol of extraction from **nuclei** yields **high-molecular-weight** DNA.
- We assembled a **contiguous flax genome** sequence (N50 = **6.21 Mb**) with 93.8% **completeness**.
- The received fast5-files can **already** be used in molecular genetic research on flax, such as DNA **methylation** studies.