Complete sequencing of barley organellar genomes: new data for intraspecific differentiation

Yermakovich (Makarevich) Anna, Siniauskaya Maryna, Halayenka Innesa, Liaudanski Aleh, Davydenko Oleg
Laboratory of cytoplasmic inheritance, IGS NAS of Belarus

Organelle genomes are an important tool to investigate:

- domestication,
- distribution,
- microevolution of plant species.

However, they have found limited use in cereal **intraspecific studies** so far.

In the present study, organelle genomes of wild (***Hordeum vulgare subsp. spontaneum***) and cultivated (***H. vulgare subsp. vulgare***) barley samples were sequenced.
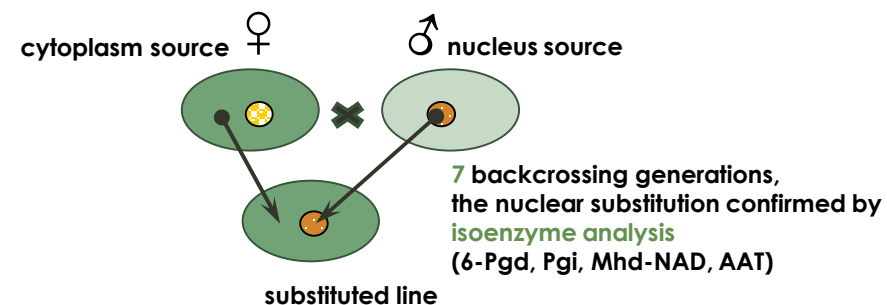
The study aimed:

1) To evaluate the diversity of organellar genomes within the ***Hordeum vulgare*** species;

2) To estimate the prospects of using chloroplast (**cp**) and mitochondrial (**mt**) genome sequences for intraspecific barley identification;

3) To re-evaluate the phylogenetic structure of *H. vulgare* species using complete **cp** and **mt** genome sequences of the presented set of wild and cultivar barley samples.

bio.makarevich@gmail.com
cytoplasmic@mail.ru

The study sample contained:

- **5** barley cultivar varieties (H. vulgare subsp. vulgare): Vezha, Roland, Vizit, Sobolyok, Maresi;

- **12** substituted lines whith **cp** and **mt** genomes from: H. vulgare subsp. vulgare (varieties Atlas and Himalaya) and H. vulgare subsp. spontaneum (wild samples W1, W3, W4, W8, W9)

The substituted barley lines creating processs

cytoplasm source ♀   ♂ nucleus source

7 backcrossing generations, the nuclear substitution confirmed by isoenzyme analysis (6-Pgd, Pgi, Mhd-NAD, AAT)

substituted line

The collection of **substituted barley lines** was created earlier to study nuclear-cytoplasmic interactions and includes 84 combinations of nuclear and cytoplasmic genomes. The presence of lines bearing the cytoplasmic genomes from a **wide range of wild and cultivated barley** makes the collection a suitable object to investigate the diversity of organelle genomes within *Hordeum vulgare* species.
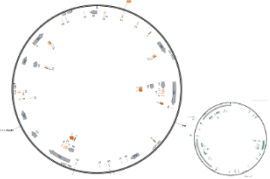
Methods

1) Isolation of the chloroplast fraction with the admixture of mitochondria
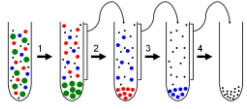
3-days barley seedlings, differential centrifugation method

2) Phenol-chloroform DNA extraction

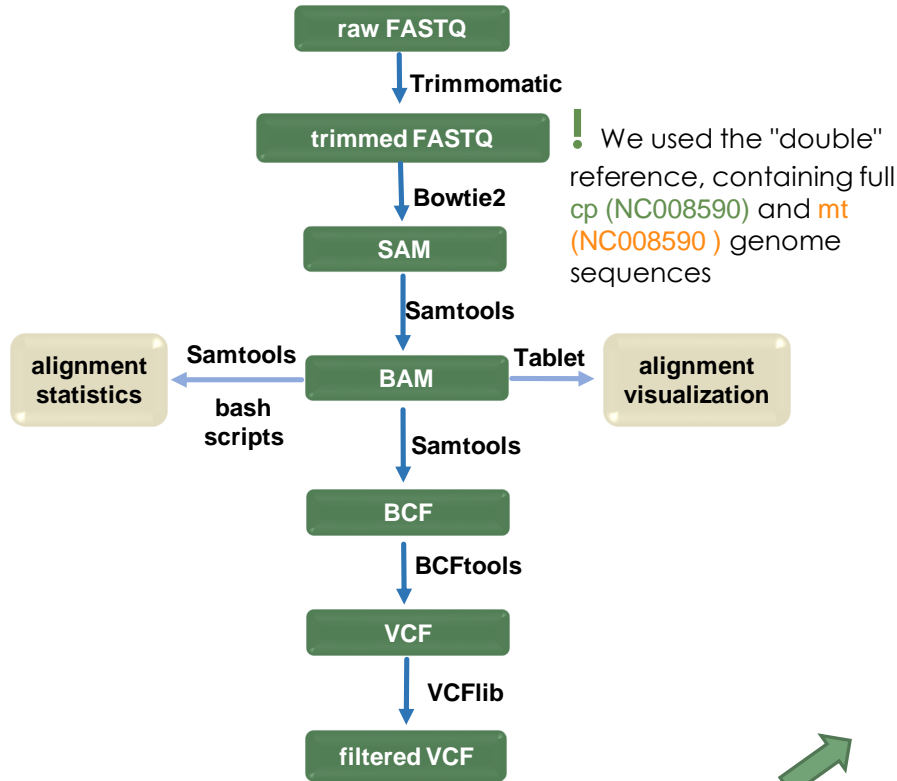The mixture of cp and mt DNA

3) NGS

MiSeq System

NexteraXT library preparation kit, MiSeq Reagent Kits v.3 and v.2.

4) The NGS data processing

The data processing algorithm was optimized for **cpDNA** and **mtDNA** mixtures **(taking into account the homology areas)** фтв tested with **ART** set of simulation tools.

raw FASTQ
→ Trimmomatic →
trimmed FASTQ

! We used the "double" reference, containing full cp (NC008590) and mt (NC008590) genome sequences

→ Bowtie2 →
SAM
→ Samtools →
BAM

Samtools / bash scripts → alignment statistics

Tablet → alignment visualization

→ Samtools →
BCF
→ BCFtools →
VCF
→ VCFlib →
filtered VCF

- Resulted **VCF files** contained all the polymorphic loci of the **cp** and **mt** genomes of each sample.
- Complete **FASTA sequences** of these genomes were generated on the base of these files.

SAM – Sequence Alignment Map
BAM – Binary Alignment Map
BCF – Binary compressed VCF
VCF – Variant Call Format

5) Obtained complete cp and mt genome sequences of study barley set

Sample type	Sample Name	NCBI GenBank ID		Country of origin	The source of cytoplasm
		Cp-sequence	Mt-sequence		
Substituted lines (BC⁷)*	Vizit (W3)	MN171383	MN127980	Israel	*Hordeum vulgare subsp. spontaneum*
	Vezha (W3)	MN171381	MN127977		
	Roland (W3)	MN171377	MN127971		
	Vizit (W4)	MN171384	MN127981	Israel	
	Roland (W4)	MN171378	MN127972		
	Vizit (W8)	MN171385	MN127982	Israel	
	Roland (W8)	MN171379	MN127973		
	Vezha (W8)	MN171382	MN127978		
	Roland (W1)	MN171376	MN127970	Israel	
	Roland (W9)	MN171380	MN127974	Israel	
	Roland (Atlas)	MN171387	MN127967	California	
	Roland (Himalaya)	MN171389	MN127968	Nepal	
Cultivars	Vezha	MN171391	MN127976	Belarus	*Hordeum vulgare subsp. vulgare*
	Roland	MN171388	MN127969	Sweden	
	Vizit	MN171392	MN127979	Belarus	
	Sobolyok	MN171390	MN127975	Russia	
	Maresi	MN171386	MN127966	German	

*All the substituted barley lines were obtained by seven backcrossings (BC⁷)

- Obtained **FASTA** as well as complete chloroplast and mitochondrial genome sequences of *H. vulgare subsp. vulgare* and *H. vulgare subsp. spontaneum* accessible in NCBI GenBank database were involved in the full-genome comparative analysis. **In total, 22 complete chloroplast genome sequences and 19 complete mitochondrial genome sequences of barley were analyzed.**

- The phylogenetic analyses of the **cp** and **mt** genomes was carried out with Bayesian statistics methods (*MrBayes*) using the GTR (General Time-Reversible) model.

Results

- Comparative analysis of 22 **cp** and 19 complete **mt** genome sequences of barley revealed **107** polymorphic sites in the chloroplast genome and **23** polymorphisms in the mitochondrial genome.

- The phylogenetic analysis of these genomes resulted into the the trees, which were consistent with each other. The trees for cp and mitochondrial genomes subdivided on **two big clades,** both clades contained **as wild** (*H. vulgare subsp. spontaneum*), **as cultivated** (*H. vulgare subsp. vulgare*) barley forms.

- These results indicated the presence of at least two centres of *H. vulgare subsp. vulgare* origin and conformed with the prevailing hypothesis of several domestication centres of barley. They also provided a direct evidence of a higher rate of nucleotide substitutions in the chloroplast genomes as compared to that of mitochondria on a microevolution scale. Revealed high level of variability of chloroplast genomes makes it possible to use them for intraspecific barley differentiation.

The chloroplast and mitochondrial genomes variability of *Hordeum vulgare*

| Variation type	Total variation number (including var. of protein coding sequence)	
	Cp genome (136 462 bp)	**Mt** genome (525 599 bp)
INDELs	9 (0)	1 (0)
SNPs	79 (20)	22 (4*)
SSRs	19 (0)	0

*In the **mt** genome there were detected:
2 amino acid substitution-causing mutations in ribosomal protein small subunit 4 gene;
2 SNPs in ribosomal protein S3A pseudogene.

Discovered nonsynonymous SNPs of the barley chloroplast genome

Location in the reference sequence (NC_008590)	Ancestral allele	Derivative allele	Amino acid substitution	Gene (NCBI ID)	Gene function	Samples carrying the derivative allele
2634	C	T	Val370Ile	matK (4525145)	trnK intron maturase	Roland(W9)
3242	A	G	Leu167Pro			H. spontaneum KC912688
25067	T	G	Asn571Lys	rpoC1 (452512)	RNA polymerase β′ subunit	Roland(W9)
49408	T	C	Lys233Asn	ndhK (4525141)	NADH dehydrogenase subunit K	Roland(W9)
76884	G	A	Ser87Leu	infA (4525109)	translation initiation factor 1	Vizit(W3), Vezha(W3), Roland(W3), Vizit(W4), Roland(W4), Roland(W1)
111390	C	T	Gly16Glu	ndhG (4525170)	NADH dehydrogenase subunit 6	Roland(W9)
115228	T	G	Arg136Arg	ndhH (4525173)	NADH dehydrogenase subunit 7	Vizit(W3), Vezha(W3), Roland(W3), Vizit(W4), Roland(W4), Roland(W1), Vizit(W8), Vetha(W8), Roland(W8), Vezha

The phylogenetic trees of the complete chloroplast (a) and mitochondrial (b) genome sequences of the studied barley samples

The phylogenetic analysis involved 21 complete cp genome sequences (8 *H. vulgare subsp. vulgare*, 11 *H. vulgare subsp. spontaneum* and 2 outgroups) and 19 complete mt genome sequences (8 *H. vulgare subsp. vulgare*, 11 *H. vulgare subsp. spontaneum*).
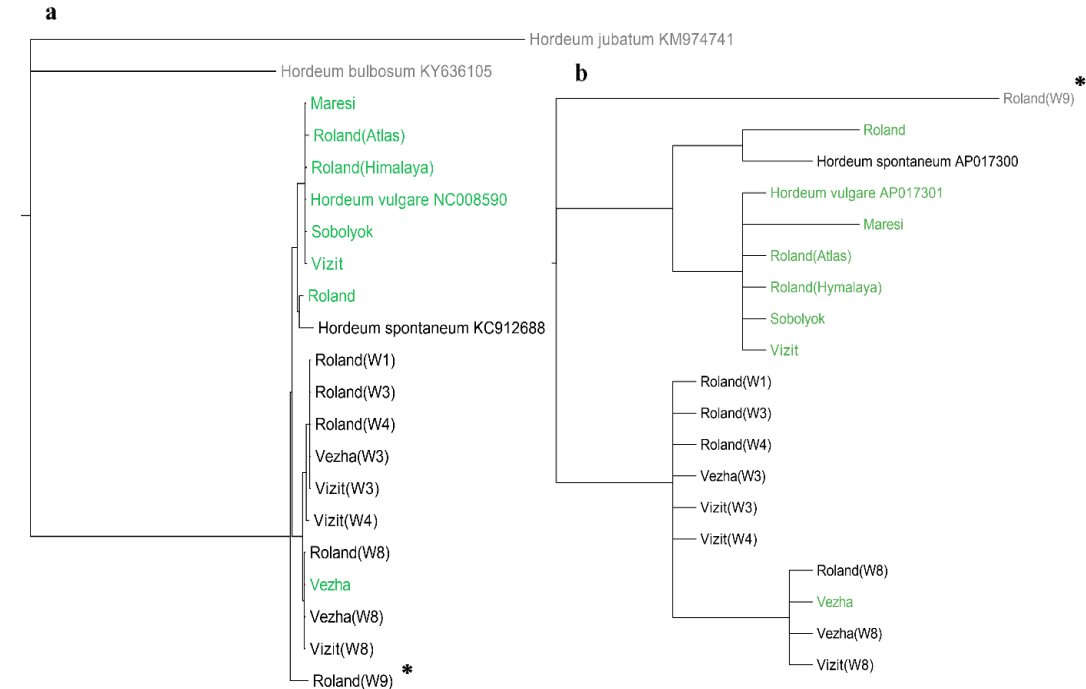
*Roland(W9) showed the maximum number of differences from the other samples in the comparative analysis: 20 unique polymorphisms of cp genome and 10 of mt genome. So, this sample formed a separate clade on the cp phylogenetic tree and also was used as an outgroup in the mt genomes phylogeny.

The samples carrying the wild barley organellar genomes **marked green**.

The results became a starting point for a broader study of the intraspecific diversity of chloroplast and mitochondrial genomes of H. vulgare. At the present moment, we are conducting NGS of other substituted lines from our collection and also for some cultivars and wild samples of barley from different countries.