Variation in plastid genomes and transcriptomes in the gynodioecious species Silene vulgaris

CURRENT STATUS: ACCEPTED

Manuela Krüger
Ustav Experimentalni Botaniky Akademie ved Ceske republiky

Oushadee A.J. Abeyawardana
Ustav Experimentalni Botaniky Akademie ved Ceske republiky

Miloslav Juricek
Ustav Experimentalni Botaniky Akademie ved Ceske republiky

Claudia Krüger
Ustav Experimentalni Botaniky Akademie ved Ceske republiky

Helena Storchova  storchova@ueb.cas.cz
Ustav Experimentalni Botaniky Akademie ved Ceske republiky

Corresponding Author
ORCID: 0000-0001-6701-179X

DOI:
10.21203/rs.2.11276/v1

SUBJECT AREAS
Plant Molecular Biology and Genetics
Abstract

Background

Gynodioecious species exist in two genders – male-sterile females and hermaphrodites. Male sterility in higher plants often results from mitonuclear interaction between the CMS (cytoplasmic male sterility) gene(s) encoded by mitochondrial genome and by nuclear-encoded restorer genes. Mitochondrial and nuclear-encoded transcriptomes in females and hermaphrodites are intensively studied, but little is known about gender-specific gene expression in plastids. We have compared plastid transcriptomes between females and hermaphrodites in two haplotypes of a gynodioecious species Silene vulgaris with known CMS candidate genes.

Results

We generated complete plastid genome sequences from five haplotypes S. vulgaris including the haplotypes KRA and KOV, for which complete mitochondrial genome sequences were already published. We constructed a phylogenetic tree based on plastid sequences of S. vulgaris. Whereas lowland S. vulgaris haplotypes including KRA and KOV clustered together, the accessions from high European mountains diverged early in the phylogram. S. vulgaris belongs among Silene species with slowly evolving plastid genomes, but we still detected 212 substitutions and 112 indels between two accessions of this species. We estimated elevated Ka/Ks in the ndhF gene, which may reflect the adaptation of S. vulgaris to high altitudes, or relaxed selection. We compared depth of coverage and editing rates between female and hermaphrodite plastid transcriptomes and found no significant differences between the two genders. We identified 51 unique C to U editing sites in the plastid genomes of S. vulgaris, 38 of them in protein coding regions, 2 in introns, and 11 in intergenic regions. The editing site in the psbZ gene was edited only in one of two plastid genomes under study.
Conclusions

We revealed no significant differences between the genders in plastid transcriptomes of two haplotypes of S. vulgaris. It suggests that gene expression of plastid genes is not affected by CMS in flower buds of S. vulgaris, although both genders may still differ in plastid gene expression in specific tissues. We revealed the difference between the plastid transcriptomes of two S. vulgaris haplotypes in editing rate and in the coverage of several antisense transcripts. Our results document the variation in plastid genomes and transcriptomes in S. vulgaris.

Background

Gynodioecy is a plant breeding system, in which hermaphrodite (H) and female (F) individuals co-occur in the same population. It occurs in about 2% of all angiosperm genera [1]. Male sterility in most (but not in all) gynodioecious plant species is encoded by the interaction of mitochondrial-encoded cytoplasmic male sterility (CMS) genes and by nuclear restorer of fertility (Rf) genes [2]. CMS is used in agriculture to produce hybrid seed with high yield and it is therefore extensively studied in crops, e.g., rice [3, 4], sunflower [5], maize [6], or sugar beet [7]. Despite a widespread occurrence of gynodioecy among angiosperms, studies of CMS in wild species remain scarce. *Silene vulgaris* (bladder campion) emerged as a model for the study of gynodioecy and CMS in natural populations more than two decades ago [8]. The genus *Silene* is rich in species with diverse mating systems – hermaphroditism, gynodioecy and dioecy [9]. The mating system affects DNA sequence variation in organellar loci, which is often higher in gynodioecious *Silene* species than in their closely related dioecious or hermaphroditic congeners most likely owing to balancing selection acting on CMS loci and the whole organellar genome [10, 11].

Besides the mode of selection, substitution rate is another essential factor which
influences DNA sequence variation in plant organelles. Whereas the substitution rate is generally low in plant organellar genomes, it is highly elevated in some phylogenetic lineages, including the genus Silene [12]. Particularly two species - Silene noctiflora and Silene conica – achieved extreme rates of sequence and structural evolution in mitochondrial and plastid genomes [13, 14].

Silene vulgaris has an organellar genome substitution rate above the angiosperm average but is relatively slowly evolving compared to other Silene species [15, 16]. However, intraspecific structural rearrangements in the mitochondrial genome of S. vulgaris are extreme, involving not only frequent losses and gains of intergenic DNA, but also changes in coding sequences [15, 17].

The accessibility of completely sequenced mitochondrial genomes of S. vulgaris and progress in RNA-seq methods enable the construction of comprehensive mitochondrial transcriptomes in this species. Such comparisons of transcriptomes between F and H plants in this species have revealed candidate CMS genes in their mitochondrial genomes and found differences in RNA editing rates between the different haplotypes which increased intraspecific protein variation [17, 18].

The plastid genome of S. vulgaris has been studied to a lesser extent than the mitochondrial genome of this species. The complete sequence of the plastid genome of only a single haplotype has been published [14] and no comprehensive plastid transcriptome analysis of this species is available.

To gain more detailed insight into plastid genome evolution in S. vulgaris, we assembled complete plastid sequences from five haplotypes of this species, including two accessions from high mountains adapted to high altitudes and genetically distant from the remaining three haplotypes originating from lowland populations [19]. CMS is a complex phenotype, which is associated with profound changes in the expression of some mitochondrial and
many nuclear genes [20 - 22]. However, plastid transcriptomes are rarely studied in connection with CMS and it is not known, whether the transcription of plastid genes is affected by the male-sterile phenotype in gynodioecious plants. We therefore compared plastid transcriptomes of F and H plants in two of the haplotypes of *S. vulgaris*, for which mitochondrial transcriptomes are also available [17, 18]. We did not detect any significant distinctions between the genders but found interesting differences between transcriptomes of the two *S. vulgaris* haplotypes.

**Methods**

*Plant material*

We collected seeds of *S. vulgaris* from two populations occurring above a timber line in European mountains (Dachstein and Vřesová studánka, the haplotypes D11 and VS1, respectively), and from one population growing in sand dunes just above sea level in Netherlands (Zandvoort, ZE2) (**Table 1**). The mountain populations *S. vulgaris* exhibited morphological traits (floral color, leaf shape, procumbent growth) distinct from lowland plants. These populations were sometimes treated as the separate subspecies *S. vulgaris* subsp. *prostrata* or *S. vulgaris* subsp. *glareosa* [49], but unrestricted gene flow and clinal variation in floral color and flavonoid production along altitudinal gradient were documented [19, 32]. We therefore refer to the mountain populations simply as *S. vulgaris* in our study.

Seeds were germinated and cultivated in the greenhouse at the Institute of Experimental Botany (IEB) in Prague, as described previously [18]. *S. vulgaris* collected in Dachstein is procumbent with light violet flowers. It was sometimes classified as *S. vulgaris* subsp. *prostrata* [49]. However, individuals with intermediate phenotypes between higher altitude and lower altitude plants were reported [19], which suggests unlimited gene flow among the populations. The Dachstein plants are very sensitive to moisture, and they grew poorly
in the greenhouse. Thus they were cultivated under controlled conditions in IEB cultivation rooms at 21º C, 16/8 h light/ dark, in pots filled with perlite, vermiculite, and coconut coir (1:1:1), fertilized (Kristalon-start and Kristalon-fruit and flower, 1 : 1) once per week or every second week. The plant material was determined by Helena Štorchová. Seed of each S. vulgaris haplotype are deposited at IEB in Prague and are available upon request. All samples collected and used in this study did not require any special permission. Plant materials used in the current research complied with government regulations.

Complete plastid genomic sequences from S. vulgaris D11, VS1, and ZE2

We performed de novo assembly of three plastid haplotypes of S. vulgaris. About 100 mg of young flower buds from a single individual S. vulgaris D11, VS1 or ZE2 (Table 2) were flash frozen in liquid nitrogen and ground with a china pestle and mortar in Lysis buffer (Qiagen Genomic DNA Buffer Set). The protocol for the preparation of high-molecular genomic DNA using Qiagen Genomic Tip (20G) was followed according to the manufacturer’s instructions. About 8 mg of DNA dissolved in 100 ml of 10 mM Tris-HCl buffer (pH = 8.3) was sent to GATC Biotech (Konstanz, Germany) for SMRT sequencing on the Pacific Bioscience RSII P. A similar aliquot of genomic DNA from each of the three accessions of S. vulgaris was sent to the Centre of Plant Structural and Functional Genomics IEB in Olomouc for Illumina MiSeq sequencing (2 x 300 cycles, fragment size about 1000 nt) using Nextera chemistry for DNA library preparation.

The SMRT sequencing generated around 39,000 reads with N50 Read length = 16,500 nt for S. vulgaris D11; 98,000 reads with N50 Read length = 16,800 nt for S. vulgaris VS1; 94,000 reads with N50 Read length = 18,100 nt for S. vulgaris ZE2. A hybrid correction pipeline proovread [50] was adopted to correct long but error-prone SMRT reads with short but accurate MiSeq reads. The proovread output ‘trimmed’ consisting of error corrected
reads was used as a local blast database and the reads homologous to the *S. vulgaris* plastid genome (JF715057) were identified by blastn search with a cutoff e -20. Canu v 1.3 [51] was applied for the assembly of proovread corrected reads. The resulting two contigs corresponded to a long single copy region (LSC) plus inverted repeat (IR), and a short single copy region (SSC). The complete plastid genomic sequences were deposited under the Genbank accession numbers MK473866-8 (ZE2, VS1,D11).

The complete plastid genomic sequence JF715057 [14] derived from *S. vulgaris* carrying the mitochondrial haplotype SD2 [15] served as a reference for the annotation of the newly assembled plastid genomes of *S. vulgaris*.

*Complete plastid genomic sequences from S. vulgaris KOV and KRA*

The data sets obtained from the Roche 454 GS-FLX platform with Titanium reagents (from constructed 3 kb paired-end libraries) previously used to assemble mitochondrial genomes of the *S. vulgaris* haplotypes KOV [15] and KRA [17] were utilized for the generation of plastid sequences for both haplotypes. Although the DNA specimens were enriched for mitochondrial DNA, they contained plenty of plastid reads, which provided 10 - 20 × coverage of the plastid genome. Roche’s GS de novo Assembler v.2.6 (‘Newbler’) was used for initial assembly. The resulting contigs were mapped against the available chloroplast genome of *S. vulgaris* (JF715057) and gaps were filled by individual trimmed 454 reads mapping against the same reference. The KOV and KRA plastid genomic sequences were confirmed by re-mapping of the 454 reads against them. Alignments were manually checked for potential SNPs, indels and insertions and edited respectively. Within-individual variation in A or T homopolymers > 5 was often observed, which might reflect possible heteroplasmy, or the co-existence of two or more variant sequences in the same individual. The resulting KOV and KRA complete plastid sequences (GenBank accession
numbers MH890612 and MH890613) were used as the reference genomes for the following transcriptomic analyses.

Chloroplast genome features

The distance matrix with the pairwise comparison of single nucleotide polymorphisms across all different plastid genomes was calculated using snp-dists (v. 0.6, github.com/tseemann/snp-dists), while for the pairwise comparison of indels the dist.dna function of the ape package within R was employed with the “indel” and “indelblock” model. The positions of simple sequence repeats (SSR) were estimated using the microsatellite identification software tool MISA-web [52] with thresholds of the repeat sequence length longer than five nucleotides for mononucleotides, four repeat units for dimer and trinucleotide SSRs, and three repeat units for tetra-, penta- and hexanucleotide SSRs.

Phylogenetic analyses

First, the six plastid genomes of *S. vulgaris* together with *Silene latifolia* as outgroup (JF715055) were aligned with MAFFT v.7.388 [53] using the L-INS-i mode with misaligned sites manually edited. The inverted repeat region A (IRA) region was cut from the resulting nucleotide alignment. For the calculation of phylogenetic trees two different alignments were compared; first with non-informative sites such as homopolymer regions masked when longer than five nucleotides in non-coding sequence (CDS); second with CDS only (without tRNAs and rRNAs). Additionally, indel characters were coded using the “simple indel coding” algorithm as described in [54] for both alignments with 2matrix [55]. The phylogenetic trees were calculated using RAxML [56] and MrBayes [57] at the CIPRES portal [57].
The maximum likelihood (ML) method was applied using the CIPRES webportal with RAxML v. 8.2.10 with 1000 bootstraps and the GTRGAMMA model for both bootstrapping and tree inference. Indels were given in a partition file as binary characters describing indel size and distribution throughout the respective sequence alignment. Alternatively, the Bayesian approach for phylogenetic tree construction was employed through MrBayes v. 3.2.6 [58] using the Markov chain Monte Carlo algorithm and the default model 4by4 for 5000 generations in two runs with trees sampled every 1000 generations. The different partitioned nexus file consisted of the sequence alignment and the indel coding, each. Stationary character frequency was fixed for the indel data set and dirichlet (1.0, 1.0, 1.0, 1.0) and the first 25 % of topologies were discarded (burnt in). The analysis was stopped when the standard deviation of split frequencies between the runs was lower than 0.01.

The numbers of synonymous and nonsynonymous substitutions within *S. vulgaris* plastid protein coding genes were determined with DnaSP v5 [59]. The segregating site was identified, if an alternative nucleotide was found in the respective position in at least one of six aligned plastid sequences of *S. vulgaris*.

**Illumina read mapping**

We used the reads stored under the Short Read Archive accession number PRJNA321915. They were obtained by Illumina sequencing of cDNA derived from total RNA extracted from flower buds of three F and three H individuals of *S. vulgaris* KRA [17] (GenBank accession numbers SRX3102769 – SRX3102774) and from flower buds of three F and three H individuals of *S. vulgaris* KOV [18] (GenBank accession numbers SRX272140 – SRX272145).

The initial alignment was performed with the assembler GSNAP v. 2017-05-03 [60] in paired-end mode with known splice sites [14]. The plastid genome sequences of the
haplotypes KOV and KRA of *S. vulgaris* were used as the references. IRA within the reference plastid genome was cut away to ensure proper read mapping with GSNAP. The resulting alignments were separated by strand using the view function in SAMtools v. 1.9 [61] by filtering according to read-pair orientation utilizing SAM flags as described in [18]. These alignments were filtered for potential mitochondrial reads derived from plastid DNA inserted to the mitochondrial genome of the corresponding haplotype (GenBank accession numbers JQ771300 and MH455602) (Table 4). At these known regions all reads not matching to the reference sequence and not presenting potential RNA editing were filtered deploying SAMtools view function, a custom AWK script and seqtk v. 1.2 (https://github.com/lh3/seqtk) for subsequent analyses. The mapped reads were visualized by means of the Integrative Genomic Viewer (IGV) [62].

For comparison we also used the ChloroSeq pipeline [24] to analyze the plastid transcriptomes of *S. vulgaris* KOV and KRA. This pipeline relies on several different, open-source bioinformatic programs, such as SAMtools and BEDtools v. 2.25.0 to run properly. The same reference, known splice sites and filtered reads as for the final GSNAP alignments were used with bowtie v. 2.2.6 [63] and tophat v. 2.1.1 [64] for read mapping as described in [24].

**RNA editing rates**

Initial variant discovery was performed with HaplotypeCaller in GenomeAnalysisTK v. 3.7 [65] on the minus- and plus-stranded alignments of *S. vulgaris* KOV and KRA with the minimum call and emit threshold (stand_call_conf) set to 20. All variant sites with C-to-T for the plus- and G-to-A alteration in the minus-strand were manually checked and verified for subsequent final calling of variants with SAMtools v. 1.2 mpileup using the DPR output tag (discontinued since v. 1.3) for the number of high-quality bases per observed allele. In
cases of low read mapping, but definite RNA editing the DP4 values were used to evaluate editing rates across all individuals. RNA editing rates were calculated based on these values as counts of Ts divided by the sum of Cs and Ts at the specific editing sites. The final editing sites were used as list for calculation of editing rates with ChloroSeq. Editing rates below a threshold of five percent (or 10 edited nucleotides, whatever is smaller), or low coverage (less than 200 reads mapped) are not shown in the final results.

Transcript abundance estimation

The estimation of transcript abundance was done as described in [18] using the coverage function within bedtools and a custom AWK script. The coverage was calculated per-base, averaged over the length of the respective feature of interest, first regardless of strand followed by strand-specific calculations for each sample and normalized as TPM as described in [66]. The average and standard deviation of TPM were calculated for both haplotypes and sexes. Antisense transcripts were recognized if their depth of coverage exceeded 300 – 500, which corresponded to the TPM values of the least expressed protein coding genes.

Results

Complete plastid genomes of S. vulgaris

We assembled five complete plastid genomes of S. vulgaris from Eurasia – the haplotypes D11, VS1, ZE2, KRA and KOV (Table 1). They ranged in length from 151,463 bp to 151,572 bp and contained a long single copy region (LSC), a short single copy region (SSC), and two inverted repeats (IRs) (Table 2). The boundaries between repeat and single copy regions and gene content were identical to the previously published plastid genome of S. vulgaris SD2 [14].

We identified the positions of simple sequence repeats (SSR), short arrays of tandem
repeat units, in the six plastid genomes of *S. vulgaris* under study. We found 871 mononucleotide repeats longer than five nucleotides, but only 46 of them were longer than nine nucleotides (Additional file 1: Data Set 1). In addition, there were 62 dinucleotides, eight trinucleotides, eight tetranucleotides, and only one pentanucleotide. SSRs represent useful markers in population genetic studies owing to their variability. Within-individual variations in the number of mononucleotide units, arisen due to heteroplasmy, was observed in most mononucleotide regions. SSRs with a repeat unit higher than two nucleotides, not affected by heteroplasmy, can therefore be recommended for plastid genotyping in *S. vulgaris*. We found 19 positions of di-, tri-, or tetranucleotides, which varied among the analyzed plastid genomes of *S. vulgaris*.

*Phylogenetic relationships and sequence polymorphism*

Plastid haplotypes VS1 and D11 of the high mountain *S. vulgaris* populations, occurring at the altitudes above 1200 m a. s. l, diverged first on the phylogenetic tree constructed on the basis of concatenated protein-coding sequences and with *S. latifolia* as an outgroup (Fig. 1b). The same topology was confirmed when entire plastid sequences except for homopolymers larger than five were used (Fig. 1a). In the latter phylogenetic tree, the position of the KRA haplotype from Siberia and D11 haplotype from the Alps was not resolved. The results obtained by maximum-likelihood (ML) method were consistent with the outputs generated by MrBayes, except that the position of KRA and D11 was always resolved (Additional file 2: Fig. 1, Additional file 3: Fig. 2).

The number of indels in pairwise alignments of individual plastid genomes of *S. vulgaris* varied from 33 to 112, and nucleotide substitutions ranged from 35 to 212 (Table 2). The highest number of the polymorphisms was found in the alignments with the haplotype VS1 from the Jeseníky Mts, consistent with its divergence’s basal position within phylogenetic
tree of *S. vulgaris* haplotypes. The indels occurred primarily in intergenic regions, but the deletion of nine bp in length, which shortened the rpl20 protein by the last three amino acids was identified in the haplotypes SD2 and ZE2. All other haplotypes (KOV, KRA, D11, VS1) harbored the longer version of the rpl20 protein. *S. vulgaris* plastid genes varied in their degree of polymorphism. Thirty of 77 unique protein coding genes were identical, and additional 25 genes carried only synonymous segregating sites and therefore encoded proteins identical among the six plastid haplotypes. Only 22 genes, including accD, matK, rpoB or ycf2, carried at least one non-synonymous segregating site (Fig 2). Two highly polymorphic genes – ycf1 and ndhF – contained the vast majority of polymorphisms: 48 of the total 69 non-synonymous segregating sites. (Additional file 1. Data Set 2). A detailed inspection of Ka/Ks for these two genes revealed values > 1.0, which indicate relaxed or positive selection. This was the case in all the pairwise comparisons among the *S. vulgaris* haplotypes for ycf1. However, only the ndhF alignments which comprised the mountain haplotypes VS1 and D11 exhibited high Ka/Ks (Additional file 1: Data Set 3).

**Plastid transcriptomes of *S. vulgaris* KRA and KOV**

We generated plastid transcriptomes of two haplotypes of *S. vulgaris* KRA [17, 18] using the data sets previously employed to construct the mitochondrial transcriptomes. We compared gene coverages and RNA editing rates in flower buds between F and H individuals and between the two haplotypes. We compared depth of coverage of protein coding genes, because rRNA was removed before cDNA library preparation and small RNAs (< 100 nt) including tRNAs were lost in the course of RNA extraction. Depth of coverage was similar in F and H plants in both haplotype KOV and KRA, no gene was differentially expressed between the genders. The
depth of coverage could not be directly compared between the KOV and KRA plastid genomes, because Illumina sequencing was performed on different platforms and produced reads of different lengths for each plastid transcriptome. We therefore compared the sets of highly and lowly covered genes between the two haplotypes. The genes psbA, rbcL, psbE, rps14 and rps16 were among the most highly expressed, whereas the ndhF and psbN were the least expressed genes both in KRA and KOV plants (Additional file 1: Data Set 4.), which indicates general similarity between the two plastid transcriptomes.

Introns were covered to a lower extent than adjacent exons in most plastid genes, but their coverage reached levels comparable to exons in some genes, for example in ndhB (Fig. 3a).

*Antisense non-coding transcripts in plastid transcriptomes of S. vulgaris KRA and KOV*

We identified five antisense non-coding transcripts > 100 nt with transcript abundance comparable to protein coding genes, overlapping trn genes (Additional file 1: Data Set 5.), three of them were found in both KRA and KOV transcriptomes, two were revealed in KOV only. The trnS-GGA and trnW-CCA - trnP-UGG antisense transcripts corresponded to the 3’UTRs of the rps4 gene and petG genes, respectively. The trnS-UGA antisense transcript colocalized with the 5’UTR of the psbZ gene (Fig. 3b, c, d).

The antisense transcript spanning the 5’UTR and the start of the rps14 gene (Fig. 4a), was the most abundant antisense transcript derived from protein coding genes in KOV, but it was absent in the KRA transcriptome. Another antisense transcript derived from the rps19 and rpl2 genes was revealed in both S. vulgaris transcriptomes under study (Fig. 4b). The psbN gene coding for a small transmembrane protein necessary for the assembly of photosystem II [23] was transcribed from minus DNA strand in sense orientation and from the opposite strand in antisense orientation as a part of the longer psbT-psbH
transcript. The antisense psbN transcript exhibited much higher depth of coverage than the sense psbN transcript coding for the psbN protein (Fig. 4c).

Similarly with protein-coding genes, no statistically significant differences in antisense transcript levels were found between F and H plants. In contrast, the abundance of antisense transcripts differed between the KOV and KRA transcriptomes of S. vulgaris more than the transcript levels of protein coding genes. The most remarkable distinction was found in the antisense rps14 transcript (Fig. 4a), which was highly covered in KOV, but completely missing in KRA.

The depth of coverage in the KRA and KOV plastid transcriptome estimated by the ChloroSeq pipeline [24] was in agreement with the results stated above, which were obtained using GSNAP according to [18] (Additional file 1: Data Set 4.).

RNA editing positions in plastid genomes of S. vulgaris KRA and KOV

We identified 51 unique C to U editing sites in the plastid genomes of S. vulgaris KRA and KOV, 38 of them located in protein coding regions, 2 of them in introns, and 11 in intergenic regions. Editing sites in rRNAs and tRNAs were not evaluated due to their biased coverage caused by the sample preparation methods. Most edits (95%) in coding sequences were non-synonymous, changing the amino acid composition; only two editing sites were silent. The most frequently edited genes were ndhB (9 edits), ndhD (4 sites), and ndhA (4 sites).

We compared S. vulgaris editing sites with eight angiosperm species, for which plastid editome had comprehensively been studied: Amborella trichopoda, Cucumis sativus [25], Spirodea polyrhiza [26], Aegilops tauschii [27], Arabidopsis thaliana [28], Hevea brasiliensis [29], Nicotiana tabacum [30], Vigna radiata [31] (Table 3). The majority of the 38 edits in protein coding regions identified in S. vulgaris were either edited, or C was
replaced with T at the DNA level in most angiosperms under comparison. The two silent
edits were not conserved across angiosperms. A highly edited position in the rps16 intron
was also edited in *A. tauschii* and replaced with T in DNA of *A. trichopoda* and *A. thaliana*,
which may indicate its functional importance. The intergenic regions could not have been
reliably aligned across the angiosperm species under comparison.

The rate of RNA editing in plastid transcriptomes of *S. vulgaris* KRA and KOV

The editing rate higher than 80% in at least one of the two *S. vulgaris* transcriptomes was
determined in 26 of 38 edits in protein coding genes, all of them were non-synonymous
*(Additional file 1: Data Set 6)*. Both silent sites were edited only about 50% or less. An
editing event introduced a premature stop codon in about 10% of the *ndhJ* transcripts, but
this position was not edited in other angiosperms under comparison. Editing is necessary
to create a start codon in the *ndhD* gene in all angiosperms, where C is present in the
second position of the coding region. However, all KRA and KOV plants were edited < 15%
in this position, which means that only a small portion of the *ndhD* transcripts encoded a
functional protein.

In contrast with protein coding genes, only two of 11 editing positions in intergenic
regions were edited more than 80%. One of them was located in 3′UTR of the *atpH* gene,
the second one in the position 64,933 of the KRA plastid genome in the *trnW-CCA - trnP-
UGG* antisense transcript.

Editing rates in the KRA and KOV plastid transcriptomes were mutually congruent,
exhibiting moderate differences in the positions with intermediate rates 40 – 70% *(Fig. 5)*.

The most remarkable difference was found in the position 50 of the *psbZ* coding region,
which changed leucine for serine. No editing was observed in this position in the KRA
haplotype, whereas approximately 6% of *psbZ* transcripts, which represented about four
hundred reads, were edited in each of six KOV plants (Table 3). Editing of this position varied across angiosperms. The same position was edited in A. thaliana and S. polymorpha, while no editing was reported in A. tauschii, H. brasiliensis, or C. sativus. T replaced C in this position in the plastid genomes of N. tabacum and V. radiata. To verify editing of this position in Caryophyllales, we downloaded the transcriptomic data of four Silene species, Agrostemma githago and Spinacia oleracea from the SRA archive and mapped them against the psbZ sequence. We found high editing of the position 50 of psbZ in S. conica and no editing in spinach, S. noctiflora and Silene paradoxa. This position was edited to a lower extent in Silene latifolia and A. githago (Additional file 1: Data Set 7). Although the coverage of psbZ was low in most data sets, it showed variable editing across close relatives of S. vulgaris, the pattern similar to scattered editing at high taxonomic level.

No statistically significant differences of editing rates between F and H individuals were observed between the KOV or KRA haplotypes. The estimates of editing rates provided by the GSNAP [18] and the ChloroSeq pipeline [24] were consistent (Additional file 4: Figure S3).

Discussion

Variation in six plastid genomes of S. vulgaris

This comparison of six completely sequenced plastid genomes of S. vulgaris revealed identical structures and moderate sequence differences, which is in line with the previously reported slow plastid evolutionary rate of this Silene species, unlike fast evolving S. noctiflora and S. conica [14]. Our study involved two accessions from mountain populations and one accession collected in sandy beach close to the sea. The mountain haplotypes diverged first in the phylogenetic tree, confirming their genetic distance from lowland populations. In contrast, the beach accession ZE2 clustered together with the remaining populations, suggesting it recently diverged likely becoming isolated in the
process of adaptation to the ecologically divergent seashore habitat with an increased salinity.

High variation in plastid evolutionary rates exist not only among the *Silene* species, but also among individual plastid genes [16]. We found prominent distinctions in substitution rate among the plastid genes also at the within-species level in *S. vulgaris*. Mainly two genes, *ycf1* and *ndhF*, accumulated a high number of non-synonymous substitutions. Whereas the *ycf1* gene exhibited high acceleration in substitution rate across all *Silene* species and also other angiosperms, the *ndhF* gene belonged among the more slowly evolving genes [16]. Our study found elevated Ka/Ks in the *ndhF* pair-wise alignments which included at least one mountain haplotype of *S. vulgaris*. The Ka/Ks values close to 1 indicate relaxed selection, the values >1 may imply positive selection. Our results suggest distinct selection acting on the *ndhF* gene in mountain populations compared to lowland plants of *S. vulgaris*. The habitats above timberline are exposed to intense light including UV. The increased content of flavonoids in the mountain populations of *S. vulgaris* has been interpreted as a defense against excessive light [32]. It is possible that the non-synonymous substitutions in the *ndhF* may be also related to adaptation to a higher light intensity in high altitudes.

*RNA editing in plastid transcriptomes of S. vulgaris*

We found no significant difference in the plastid transcriptomes of flower buds in terms of coverage and RNA editing rate between F and H plants neither in the KRA nor KOV haplotype of *S. vulgaris*. This indicates that plastid transcriptomes were not affected by the processes underlying pollen abortion in two haplotypes with distinct CMS types [17, 18]. However, as we analyzed total RNA from entire flower buds, it cannot be excluded that the transcription in plastids of some specific tissues, e.g., tapetum is still influenced
Thirty eight RNA editing sites were revealed in the protein coding genes in the *S. vulgaris* transcriptomes. This value is comparable to those reported in other eudicots: 34 edits *A. thaliana* [28], 40 edits in *V. radiata* [31], or 51 edits in *C. sativus* [25]. This number is much lower than 138 editing sites discovered in the plastid protein coding genes of the basal angiosperm *A. trichopoda*, which is in agreement with a general trend of gradual loss of plastid edits in the course of the evolution of flowering plants [25]. The loss of editing by the replacement of C for T in genomic DNA occurred in parallel to a similar extent, but it affected distinct sites in various lineages (*Table 3*).

The vast majority of editing sites in protein coding genes (36) were non-synonymous, changing the encoded amino acid. With a single exception (low-rate edit introducing a premature stop codon in *ndhJ*), all non-synonymous sites were conserved - they were either edited or C was replaced with T in at least one species under comparison. Most, but not all non-synonymous sites were edited more than 80%. An interesting example of a largely unedited essential position was observed in the *ndhD* gene, where editing established the start codon in only about 15% of transcripts, which strongly reduced the abundance of correct mRNA and might have decreased the production of functional NdhD protein. The same position was low-edited in non-photosynthetic organs (roots, etiolated seedlings) of *A. thaliana* [33]. As our transcriptomes were derived from young flower buds, which contained both photosynthetic and non-photosynthetic tissues, low editing of the *ndhD* start codon might have reflected a lack of editing in some floral tissues. Owing to its possible strong impact on the NdhD protein abundance and the function of NDH complex, editing of the start codon of might have been employed to regulate the expression of the *ndhD* gene. Although the primary function of organellar RNA editing in plants resides in the restoration of conserved amino acids [34], its role in
posttranscriptional gene expression control shall be considered[35]. Additional examples of developmental stage- and tissue-specific RNA editing were previously described in plastids of tomato [36] or potato [37].

The two haplotypes of \textit{S. vulgaris} differed in editing extent of the \textit{psbZ} gene. One haplotype was edited to a low extent, but consistently across all six individuals of both genders, the other one was not edited at all. Editing rate of the position 50 of \textit{psbZ} varied across angiosperms under comparison, where all the possibilities, namely editing, replacement of C for T, and zero editing, were observed (\textbf{Table 3}). No comprehensive plastid transcriptome of \textit{Silene} has been published, but mining available transcriptomic data from GenBank showed similar pattern of the variation in \textit{psbZ} editing rate across close relatives of \textit{S. vulgaris} as across angiosperms. The \textit{psbZ} protein is an important component the supramolecular architecture of photosystem II [38, 39], whose subunits belong among the least divergent genes in \textit{Silene} [16], most likely owing to the action of purifying selection. The variation in \textit{psbZ} RNA editing across angiosperms, Caryophyllaceae, and even within a single species \textit{S. vulgaris} is therefore surprising. The editing of the position 50 replaces S for L in the middle of the first transmembrane domain of \textit{psbZ} [39]. It is possible that this exchange does not have noticeable impact on the protein function. The variation in RNA editing among \textit{Arabidopsis} species, affecting functionally less important sites, was recently described [40], but the position 50 of \textit{psbZ} was uniformly edited in the \textit{Arabidopsis} species according to this study. As the position 50 of \textit{psbZ} is completely edited in the model plant \textit{A. thaliana}, the identification of nuclear factors responsible for the \textit{psbZ} editing will be possible, which may help to clarify the function of this editing event.

\textit{Antisense RNAs in plastid transcriptomes of \textit{S. vulgaris}}
We found the long antisense transcript of the *psbN* gene, which was more abundant than the sense transcript of this gene. The *psbN* gene is located on the strand complementary to the *psbT-psbH* intergenic region, which is a part of the conserved *psbB* operon. The transcription of the *psbN* gene in *A. thaliana* is controlled by a specific promoter recognized by the plastid-encoded RNA polymerase together with nucleus-encoded sigma factor SIG3 [41]. The antisense *psbN* transcript was found to affect the cleavage of the *psbT-psbH* intercistronic RNA [42] and to influence the translation of *psbT* mRNA in *A. thaliana* [43]. It is therefore possible that antisense *psbN* transcript has a similar regulatory function in *S. vulgaris*. On the contrary, we found only a very low or zero coverage of the strand complementary to the *ndhB* gene. The antisense *ndhB* transcript was observed in *A. thaliana*, tobacco and poplar and may play a role in mRNA stability control [44]. Its expression is influenced by temperature and developmental stage. It may not be expressed in floral buds, or in *S. vulgaris* at all.

We did not estimate the expression of small RNAs including tRNAs, owing to a size limitation, but we detected longer antisense RNAs transcribed from the strand complementary to the *trn* genes. The antisense *trnS-GGA* and antisense *trnW-CCA* are located in 3′UTR of *rps4* and *petG*, respectively. They form secondary structures, which may be recognized by RNA-binding proteins that regulate transcription of plastid mRNAs [45, 46]. Similarly, the antisense *trnS-UGA* may stabilize the 5′end of the *psbZ* transcript and influence its translation.

Numerous antisense RNAs were described in plastid transcriptomes, for example 107 putative antisense transcripts in *A. thaliana* [47], or 137 antisense candidates in *Salvia miltiorrhiza* [48]. We detected only eight putative long antisense RNAs in *S. vulgaris*, which might have been caused by two factors. First, we narrowed our search by raising the coverage threshold to the level of protein-coding genes. Second, we carefully...
eliminated reads derived from plastid inserts in the mitochondrial genome, which can be erroneously recognized as plastid-encoded transcripts.

The accumulation of antisense RNA can be influenced by the environment and developmental stage [43, 44], which may explain, why some antisense RNAs were expressed only in one haplotype of S. vulgaris. On the other hand, all the putative antisense RNAs recognized in the S. vulgaris transcriptomes were also found in A. thaliana [47], which suggests their evolutionary conservation and possible functional importance.

Conclusions

We found no significant differences between F and H individuals in the plastid transcriptomes prepared from flower buds (where the differences between both genders may be expected) of two haplotypes of gynodioecious plant S. vulgaris, which suggests that CMS was not associated with the changes in plastid gene expression in this species. However, we cannot exclude, that differences in plastid transcriptomes exist in specific tissues of floral buds. We observed differences between the two haplotypes of S. vulgaris in the rate of RNA editing of position 50 psbZ gene, which is edited in some angiosperms including A. thaliana, but not in the others. Differences in the levels of expression of antisense transcripts were also detected among haplotypes. Our results document the variation in plastid transcriptomes at the intraspecific level in S. vulgaris.

The plastid haplotypes KRA and KOV, from which the transcriptomes were constructed belong to a main cluster in the phylogenetic tree constructed of complete plastid genome sequences. S. vulgaris populations collected in the high European mountains, for which complete plastid genomes were sequenced, were added to phylogenetic analyses to increase intraspecific sampling. They occupy basal positions on the phylogram and may be closer to the ancestor of S. vulgaris.
Abbreviations

accD  acetyl-CoA carboxylase subunit D
bp  base pair
CMS  cytoplasmic male sterility
cDNA  complementary DNA
F  female
H  hermaphrodite
kb  kilobase
LSC  long single copy
matK  maturase K
ndhA  NADH dehydrogenase A
ndhB  NADH dehydrogenase B
ndhD  NADH dehydrogenase D
ndhF  NADH dehydrogenase F
rbcL  rubisco subunit L
rpoB  RNA polymerase B
psbA  photosystem II protein A
psbE  photosystem II protein E
psbH  photosystem II protein H
psbN  photosystem II protein N
psbT  photosystem II protein T
psbZ  photosystem II protein Z
rps14  ribosomal protein S14
rps16  ribosomal protein S16
rRNA  ribosomal RNA
tRNA  transfer RNA
SSC  short single copy
SSR  simple sequence repeat
UTR  untranslated region

Declarations

**Ethics approval and consent to participate**

Not applicable

**Consent to publish**

Not applicable

**Availability of data and materials**

The data of this study data have been deposited in the NCBI with BioProject accession number PRJNA321915. The RNA-seq reads from hermaphrodites of *S. vulgaris* KRA are stored under the number SRS2438489, the reads from females under the number SRS2438490. The RNA-seq reads from six individuals of *S. vulgaris* KRA are deposited under GenBank accession numbers SRX272140 - SRX272145. The complete plastid genomes can be found under GenBank accession numbers MH890612 and MH890613 (KOV and KRA), and MK473866-8 (ZE2, VS1,D11).

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This project was funded by the grant of the Grant Agency of the Czech Republic 16-09220S to HŠ. Additional support was provided by European Regional Development Fund-Project...
"Centre for Experimental Plant Biology" (No. CZ.02.1.01/0.0/0.0/16_019/0000738). Funders provided financial support only, they had no role in the design of the study, analysis and interpretation of data, decision to publish, or the preparation of the manuscript.

**Authors' contributions**

MK performed most data analyzes and interpretations, MJ participated in data analyzes, OAJA and CK performed wet lab work, HŠ conceived and designed the experiments and provided interpretation, HŠ and MK wrote the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

The authors thank Dan B. Sloan and James D. Stone for valuable comments and linguistic corrections. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005), is greatly appreciated.

**References**

1. Renner SS. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. Am J Bot. 2014;101:1588-96. doi:10.3732/ajb.1400196

2. Hanson MR, Bentolila S. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 2004;16:S154-S169. doi:10.1105/tpc.015966.

3. Kazama T, Nakamura, Watanabe, M. Sugita KT. Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. Plant J. 2008;55:619-28. doi: 10.1111/j.1365-313X.2008.03529.x

4. Wang K, Gao F, Ji Y, Liu Y, Dan Z, Yang P, et al. ORFH79 impairs mitochondrial
function via interaction with a subunit of electron transport chain complex III in Honglian cytoplasmic male sterile rice. New Phytol. 2013;198:408–18. doi: 10.1111/nph.12180

5. Sabar M, Gagliardi D, Balk J, Leaver C. ORFB is a subunit of F1FO-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. EMBO Rep. 2003;4:381–6. doi: 10.1038/sj.embor.embor800

6. Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Guan NL, et al. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics. 2007;177:1173–92. doi: 10.1534/genetics.107.073312

7. Darracq A, Varré JS, Maréchal-Drouard L, Courseaux A, Castric V, Saumitou-Laprade P, et al. Structural and content diversity of mitochondrial genome in beet: A comparative genomic analysis. Genome Biol Evol. 2011;3:723–36. doi: 10.1093/gbe/evr042

8. Charlesworth D, Laporte V. The male-sterility polymorphism of *Silene vulgaris*: Analysis of genetic data: From two populations and comparison with *Thymus vulgaris*. Genetics. 1998;150:1267–82.

9. Desfeux C, Maurice S, Henry JP, Lejeune B, Gouyon PH. Reproductive Systems in the Genus *Silene*. Evolution of reproductive systems in the genus *Silene*. Proc R Soc B Biol Sci. 1996;263:409–14.

10. Städler T, Delph LF. Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. Proc Natl Acad Sci U S A. 2002;99:11730–5. doi: 10.1073/pnas.182267799.

11. Touzet P, Delph LF. The effect of breeding system on polymorphism in mitochondrial genes of silene. Genetics. 2009;181:631–44. doi:10.1534/genetics.108.092411

12. Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. Extensive variation in
synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol. 2007;7:7. doi: 10.1186/1471-2148-7-135

13. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, et al. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 2012;10: e1001241. doi:0.1371/journal.pbio.1001241

14. Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. Genome Biol Evol. 2012;4:294-306. doi:10.1093/gbe/evs006

15. Sloan DB, Müller K, McCauley DE, Taylor DR, Storchová H. Intraspecific variation in mitochondrial genome sequence, structure, and gene content in *Silene vulgaris*, an angiosperm with pervasive cytoplasmic male sterility. New Phytol. 2012;196:1228-39. doi:10.1111/j.1469-8137.2012.04340.x.

16. Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Mol Phylogenet Evol. 2014;72:82-9. doi:10.1016/j.ympev.2013.12.004

17. Štorchová H, Stone JD, Sloan DB, Abeyawardana OAJ, Muller K, Walterová J, Pažoutová M. Homologous recombination changes the context of Cytochrome b transcription in the mitochondrial genome of *Silene vulgaris* KRA. BMC Genomics. 2018;19:874. doi: 10.1186/s12864-018-5254-0

18. Stone JD, Koloušková P, Sloan DB, Štorchová H. Non-coding RNA may be associated with cytoplasmic male sterility in *Silene vulgaris*. J Exp Bot. 2017;68:1599-612. doi: 10.1093/jxb/erx057
19. Abbate JL, Antonovics J. Elevational disease distribution in a natural plant-pathogen system: insights from changes across host populations and climate. Oikos. 2014;123:1126-36. doi: 10.1111/oik.01001

20. Li ZF, Zhang YC, Chen YQ. MiRNAs and IncRNAs in reproductive development. Plant Sci. 2015;238:46-52. doi:10.1016/j.plantsci.2015.05.017

21. Wu J, Zhang M, Zhang B, Zhang X, Guo L, Qi T, et al. Genome-wide comparative transcriptome analysis of CMS-D2 and its maintainer and restorer lines in upland cotton. BMC Genomics. 2017;18:454. doi: 10.1186/s12864-017-3841-0

22. Hamid R, Tomar RS, Marashi H, Malekzadeh S, Golakiya BA, Mohsenpour M. Transcriptome profiling and cataloging differential gene expression in floral buds of fertile and sterile lines of cotton (Gossypium hirsutum L.). Gene. 2018;660:80-91. doi: 10.1016/j.gene.2018.03.070

23. Plöchinger M, Schwenkert S, von Sydow L, Schroder WP, Meurer J. Functional update of the Auxiliary TerC and ALB3 in maintenance and assembly of PSII. Front Plant Sci. 2016;7:423. doi:10.3389/fpls.2016.00423

24. Castandet B, Hotto AM, Strickler SR, Stern DB. ChloroSeq, an optimized chloroplast RNA-Seq bioinformatic pipeline, reveals Remodeling of the organellar transcriptome under heat stress. G3-Genes Genomes Genet. 2016;6:2817–27. doi:10.1534/g3.116.030783.

25. Hein A, Polsakiewicz M, Knoop V. Frequent chloroplast RNA editing in early-branching flowering plants: Pilot studies on angiosperm-wide coexistence of editing sites and their nuclear specificity factors. BMC Evol Biol. 2016;16:23. doi:10.1186/s12862-016-0589-0.

26. Wang W, Zhang W, Wu Y, Maliga P, Messing J. RNA editing in chloroplasts of Spirodela polyrhiza, an aquatic monocotelydonous species. PLoS One. 2015;10:e0140285.
27. Wang M, Liu H, Ge L, Xing G, Wang M, Weining S, et al. Identification and Analysis of RNA editing sites in the chloroplast transcripts of Aegilops tauschii L. Genes. 2017;8:13. doi: 10.3390/genes8010013

28. Ruwe H, Castandet B, Schmitz-Linneweber C, Stern DB. Arabidopsis chloroplast quantitative editotype. FEBS Lett. 2013;587:1429–33. doi:10.1016/j.febslet.2013.03.022.

29. Tangphatsornruang S, Uthaipaisanwong P, Sangsrakru D, Chanprasert J, Yoocha T, Jomchai N, et al. Characterization of the complete chloroplast genome of Hevea brasiliensis reveals genome rearrangement, RNA editing sites and phylogenetic relationships. Gene. 2011;475:104–12. doi:10.1016/j.gene.2011.01.002.

30. Hirose T, Kusumegi T, Tsudzuki T, Sugiura M. RNA editing sites in tobacco chloroplast transcripts: editing as a possible regulator of chloroplast RNA polymerase activity. Mol Gen Genet. 1999;262:462–7.

31. Lin C, Ko C, Kuo C, Liu M, Schafleitner R. Transcriptional slippage and RNA Editing Increase the diversity of transcripts in chloroplasts: Insight from deep sequencing of Vigna radiata genome and transcriptome. PLoS One 2015;10: e0129396. doi:10.1371/journal.pone.0129396.

32. Berardi AE, Fields PD, Abbate JL, Taylor DR. Elevational divergence and clinal variation in floral color and leaf chemistry in Silene vulgaris. Am J Bot. 2016;103:1508–23. doi: 10.3732/ajb.1600106

33. Tseng CC, Lee CJ, Chung YT, Sung TY, Hsieh MH. Differential regulation of Arabidopsis plastid gene expression and RNA editing in non-photosynthetic tissues. Plant Mol Biol. 2013;82:375–92. doi: 10.1007/s11103-013-0069-5

34. Maier RM, Neckermann K, Igloi GL, Kossel H. Complete sequence of the maize
chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol. 1995;251:614–28.

35. Sloan DB. Nuclear and mitochondrial RNA editing systems have opposite effects on protein diversity. Biol Lett. 2017;13:20170314. doi:10.1098/rsbl.2017.0314

36. Kahlau S, Bock R. Plastid transcriptomics and translatomics of tomato fruit development and chloroplast-to-chromoplast differentiation: chloroplast gene expression largely serves the production of a single protein. Plant Cell. 2008;20:856–74. doi:10.1105/tpc.107.055202

37. Valkov VT, Scotti N, Kahlau S, Maclean D, Grillo S, Gray JC, et al. Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: Transcriptional and posttranscriptional control. Plant Physiol. 2009;150:2030–44. doi:10.1104/pp.109.140483

38. Swiatek M, Kuras R, Sokolenko A, Higgs D, Olive J, Cinque G, et al. The chloroplast gene ycf9 encodes a photosystem II (PSII) core subunit, PsbZ, that participates in PSII supramolecular architecture. Plant Cell. 2001;13:1347–67.

39. Wei X, Su X, Cao P, Liu X, Chang W, Li M, et al. Structure of spinach photosystem II – LHCII supercomplex at 3.2 Å resolution. Nature. 2016;534:69–74. doi:10.1038/nature18020.

40. Kawabe A, Furihata HY, Tsujino Y, Kawanabe T, Fujii S, Yoshida T. Divergence of RNA editing among Arabidopsis species. Plant Sci. 2019;280:241–7. doi: 10.1016/j.plantsci.2018.12.009

41. Zghidi W, Merendino L, Cottet A, Mache R, Lerbs-Mache S. Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the psbN gene in plastids. Nucleic Acids Res. 2007;35:455-464. doi: 10.1093/nar/gkl1067

42. Chevalier F, Ghulam MM, Rondet D, Pfannschmidt T, Merendino L, Lerbs-Mache S.
Characterization of the psbH precursor RNAs reveals a precise endoribonuclease cleavage site in the psbT/psbH intergenic region that is dependent on psbN gene expression. Plant Mol Biol. 2015;88:357–67. doi: 10.1007/s11103-015-0325-y

43. Zghidi-Abouzid O, Merendino L, Buhr F, Ghulam MM, Lerbs-Mache S. Characterization of plastid psbT sense and antisense RNAs. Nucleic Acids Res. 2011;39:5379–87. doi: 10.1093/nar/gkr143

44. Georg J, Honsel A, Rennenberg H, Hess WR. Rapid report A long antisense RNA in plant chloroplasts. New Phytol 2010;4:615–22. 10.1111/j.1469-8137.2010.03203.x

45. Bollenbach TJ, Sharwood RE, Gutierrez R, Lerbs-Mache S, Stern DB. The RNA-binding proteins CSP41a and CSP41b may regulate transcription and translation of chloroplast-encoded RNAs in Arabidopsis. Plant Mol Biol. 2009;69:541–52. doi:10.1007/s11103-008-9436-z

46. Manavski N, Schmid LM, Meurer J. RNA-stabilization factors in chloroplasts of vascular plants. Essays Biochem. 2018;62:51–64. 10.1042/EBC20170061

47. Hotto AM, Schmitz RJ, Fei Z, Ecker JR, Stern DB. Unexpected diversity of chloroplast noncoding RNAs as revealed by deep sequencing of the Arabidopsis transcriptome. G3-Genes Genomes Genet. 2011;1:559–70. doi:10.1534/g3.111.000752.

48. Chen H, Zhang J, Yuan G, Liu C. Complex interplay among DNA modification, noncoding RNA expression and protein-coding RNA expression in Salvia miltiorrhiza chloroplast genome. PLoS One. 2014;9: e99314. doi:10.1371/journal.pone.0099314

49. Marsden-Jones EM, Turrill WB. The bladder campions. London: The Ray Society; 1957.

50. Hackl T, Hedrich R, Schultz J, Forster F. Sequence analysis proovread : large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics. 2014;30:3004–11. doi:10.1093/bioinformatics/btu392

51. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and
accurate long-read assembly via adaptive k-mer weighting and repeat separation. 
Genome Res. 2017;27:722-36. doi/10.1101/gr.215087.116

52. Beier S, Thiel T, Munich T, Scholz U, Mascher M. Sequence analysis MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583-5. doi: 0.1093/bioinformatics/btx198

53. Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability article fast track. Mol Biol Evol. 2013;30:772-80. doi:10.1093/molbev/mst010

54. Simmons MP, Ochoterena H. Society of Systematic Biologists Gaps as Characters in Sequence-Based Phylogenetic Analyses. Syst. Biol. 2000;49:369-81.

55. Salinas DR, Little DP. 2MATRIX: A utility for indel coding and phylogenetic matrix concatenation. Appl Plant Sci. 2014;2:1300083. doi:10.3732/apps.1300083

56. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312-3. 
 doi:10.1093/bioinformatics/btu033

57. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, et al. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539-42. doi:10.1093/sysbio/sys029

58. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA p. 1-8.

59. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451-2. 
 doi:10.1093/bioinformatics/btp187

60. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in
short reads. Bioinformatics. 2010;26:873–81. doi:10.1093/bioinformatics/btq057

61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment / map format and SAMtools. Bioinformatics. 2009;25:2078-9.
   doi:10.1093/bioinformatics/btp352

62. Thorvaldsdottir P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2012;14:178–92.
   doi:10.1093/bib/bbs017

63. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. doi:10.1038/nmeth.1923.

64. Trapnell C, Pachter L, Salzberg SL. TopHat : discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–11. doi:10.1093/bioinformatics/btp120

65. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis A, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297-303. doi:10.1101/gr.107524.110.

66. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131:281-5. doi: 10.1007/s12064-012-0162-3

Tables

Table 1. Collection sites of *Silene vulgaris* accessions analyzed in this study
| Haplotype | Region | Location                                      | Latitude | Longitude |
|-----------|--------|-----------------------------------------------|----------|-----------|
| D11       | Europe | Austria, the Alps Dachstein                   |          | 47.456802 |
| VS1       | Europe | Czech Republic Jeseníky Mts. Vřesová studánka (Heidebrünnel) | 50.145429 |           |
| ZE2       | Europe | Netherlands, Zandvoort                        | 52.374868 |           |
| KOV       | Europe | Czech Republic Kováry near Prague             | 50.185833 |           |
| KRA       | Asia   | Russia Krasnoyarsk, Siberia                   | Not known |           |

Table 1. Collection sites of *Silene vulgaris* accessions analyzed in this study
Table 3. The comparison of editing between the plastid genomes of *S. vulgaris* KRA and *S. vulgaris* KOV and eight angiosperm species. *Ambo.* – *Amborella trichopoda*, *Spiro.* – *Spirodela polyrhiza*, *Aegilops tauschii*, *Arab.* – *Arabidopsis thaliana*, *Hevea brasiliensis*, *Nicot.* – *Nicotiana tabacum*, *Vigna radiata*, *Cucumis sativus*.

| Gene position | Effect/Notes | Genome position KRA | Edit rate | Genome position KOV | Edit rate | Ambič |
|---------------|--------------|---------------------|-----------|---------------------|-----------|-------|
| rps 16 intron | TCA (S) => TTA (L) | 5004 | 69% | 4996 | 30% | NA |
| rps 16 intron | TCA (S) => TTA (L) | 5325 | 93% | 5317 | 93% | T |
| atpAeU914SL | intergenic | 10003 | 96% | 10000 | 96% | T |
| rps2eU248SL | TCA (S) => TTA (L) | 14920 | 98% | 14920 | 97% | C |
| rpoC2eU2968HY | CAT (H) => TAT (Y) | 16608 | 70% | 16586 | 85% | C |
| rpoBeU566SL | TCA (S) => TTA (L) | 25178 | 52% | 25156 | 80% | C |
| rpoBeU551SL | TCA (S) => TTA (L) | 25193 | 49% | 25171 | 76% | Edit |
| rpoBeU473SL | TCA (S) => TTA (L) | 25271 | 78% | 25249 | 83% | Edit |
| psbZeU50SL | TCA (S) => TTA (L) | 34009 | 0 | 33958 | 6% | Edit |
| ndhJeU160QX | CAA(Q) => TAA(Stop) | 47529 | 13% | 47486 | 8% | C |
| ndhKeU65SL | TCA (S) => TTA (L) | 48408 | 65% | 48365 | 77% | Edit |
| accDeU1337PL | CCA (P) => CTA (L) | 57221 | 95% | 57185 | 96% | Edit |
| psaleU85HY | CAT (H) => TAT (Y) | 57642 | 71% | 57601 | 82% | Edit |
| psbLeU2TM | ACG(T) => ATG (Start) | 62667 | 96% | 62630 | 98% | Edit |
| petLeU5PL | CCT (P) => CTT (L) | 64268 | 86% | 64236 | 90% | Edit |
| clpPeU559HY | CAT (H) => TAT (Y) | 68351 | 97% | 68321 | 98% | Edit |
| petBeU12VV | Silent GTC (V) => GTT (V) | 73958 | 32% | 73935 | 56% | A |
| petBeU418RW | CGG (R) => TGG (W) | 74364 | 97% | 74341 | 93% | Edit |
| rpoAeU200SF | TCT (S) => TTT (F) | 77018 | 93% | 76983 | 89% | Edit |
| ndhBeU1481PL | CCA (P) => CTA (L) | 91996 | 89% | 91961 | 80% | Edit |
| ndhBeU836SL | TCA (S) => TTA (L) | 92641 | 90% | 92606 | 88% | T |
| Strain      | Site     | Other            | Ratio (%) | Consensus    | Direction  | Tc       | Consensus    | Direction  |
|-------------|----------|------------------|-----------|--------------|------------|---------|--------------|------------|
| TCA (S) => TTA (L) | 92647   | 88%              | 92612     | 84%          | Edit       | T        |              |            |
| CCA (P) => CTA (L)  | 93399   | 74%              | 93364     | 80%          | T          |         |              |            |
| CAT (H) => TAT (Y)  | 93408   | 87%              | 93373     | 81%          | Edit       |         |              |            |
| TCT (S) => TTT (F)  | 93559   | 60%              | 93524     | 56%          | Edit       |         |              |            |
| ACG (T) => ATG (M)  | 93603   | 30%              | 93568     | 45%          | Edit       |         |              |            |
| TCA (S) => TTA (L)  | 93678   | 73%              | 93643     | 79%          | Edit       |         |              |            |
| TCA (S) => TTA (L)  | 113839  | 96%              | 113810    | 96%          | T          |         |              |            |
| CCC (P) => CTC (L)  | 114250  | 84%              | 114221    | 87%          | T          |         |              |            |
| CAT (H) => TAT (Y)  | 114754  | 78%              | 114722    | 87%          | Edit       |         |              |            |
| ACG (T) => ATG (Start) | 115135  | 13%              | 115103    | 16%          | Edit       |         |              |            |
| ACA (T) => ATA (I)  | 116860  | 78%              | 116828    | 86%          | C          |         |              |            |
| TCC (S) => TTC (F)  | 117937  | 83%              | 117905    | 75%          | Edit       |         |              |            |
| CCT (P) => TCT (S)  | 118049  | 74%              | 118017    | 83%          | Edit       |         |              |            |
| TCA (S) => TTA (L)  | 118444  | 6%               | 118412    | 21%          | Edit       |         |              |            |
| TCA (S) => TTA (L)  | 119692  | 89%              | 119660    | 73%          | T          |         |              |            |
| Silent ATC (I) => ATT (I) | 120913  | 29%              | 120881    | 7%           | C          |         |              |            |

**Editing below threshold)**

| Strain      | Site     | Other            | Ratio (%) | Consensus    | Direction  | Tc       | Consensus    | Direction  |
|-------------|----------|------------------|-----------|--------------|------------|---------|--------------|------------|
| CTT (L) => TTT (F) | 57639   | 5%               | 57598     | 4%           | C          |         |              |            |
| intergenic  | 57712   | 9%               | 57671     | 15%          | NA         |         |              |            |
| intergenic  | 57717   | 7%               | 57676     | 12%          | NA         |         |              |            |
| intergenic  | 57740   | 4%               | 57699     | 9%           | NA         |         |              |            |
| intergenic  | 57745   | 5%               | 57704     | 9%           | NA         |         |              |            |
| intergenic  | 57892   | 4%               | 57854     | 8%           | NA         |         |              |            |
| intergenic  | 57897   | 2%               | 57859     | 13%          | NA         |         |              |            |

**C? information about editing not available**

*) Editing rate did not achieve the threshold in most plants (using GSNAP or ChloroSeq), but editing events observed across all 12 individuals of *S. vulgaris* KO and KRA

 **) GCA triplet in Arabidopsis encoding alanine

 ***) TCG triplet in Nicotiana, encoding serine

 ****) TCA triplet in Vigna, encoding serine

Table 4. Plastid DNA insertions in mitochondrial genomes of *S. vulgaris* KRA and *S. vulgaris*....
KOV with more than 95% of sequence similarity.

|                | S. vulgaris KRA | S. vulgaris KOV |
|----------------|-----------------|-----------------|
| Mitochondrial  | Plastid genome  | Mitochondrial   | Plastid genome  |
| KRA-1 coordinates | coordinates | KOV-1 coordinates | coordinates |
| Size (nt) | 100,495 - 100,872 | 35,346 - 35,723 | 196,739 - 203,697 |
|              | 295,098 - 295,716 | 40,335 - 40,982 | 189,999 - 195,890 |
|              | 368,306 - 371,081 | 50,900 - 53,664 | 156,787 - 169,357 |
|              | 135,024 - 135,444 | 73,826 - 74,243 | 93,157 - 105,727 |
|              | 318,767 - 321,559 | 75,549 - 78,345 | 2,796 |
|              | 78,752 - 79,770  | 81,520 - 82,538 | 1,018 |

**Figures**
Figure 1

Maximum likelihood phylogenetic trees for various plastid haplotypes of Silene vulgaris. a based on all sites of the plastid genome except for homopolymer sites larger than five nucleotides; b plastid coding regions only. Silene latifolia was used as outgroup. Long branches were shortened by 50%, marked by two diagonal slashes. The scale bar indicates the number of substitutions per site. Branches with bootstrap support below 60% were collapsed to polytomies.

Phylogenetic trees were computed through the CIPRES webportal with RAxML v. 8.2.10 using 1000 bootstraps and the GTRGAMMA model.
Figure 2

Variation in synonymous and nonsynonymous single nucleotide polymorphisms among S. vulgaris plastid protein genes. The number of segregating sites per synonymous or nonsynonymous site was calculated by DnaSp.
Figure 3

RNA-seq read coverages for genes and antisense transcripts in the plastid genome of S. vulgaris KRA. Mean DOC values across individuals are given by the plots’ lines, while standard deviations in DOC are given by the violet (plus strand) or green (minus strand) bands. a. The ndhB, rps7, and rps12 genes are transcribed from the minus strand. b. trnS-GGA is transcribed from the plus strand, but its coverage was not estimated owing to a small size. The antisense trnS-GGA transcript has DOC comparable to the adjacent rps4 gene. c. The antisense trnW-CCA and trnP-UGG co-transcript overlaps the petG 3’UTR. d. The antisense trnS-UGA transcript overlaps the psbZ 5’UTR.
RNA-seq read coverages for genes and antisense transcripts in the plastid genome of S. vulgaris. Mean DOC values across individuals are given by the plots’ lines, while standard deviations in DOC are given by the violet (plus strand) or green (minus strand) bands. a. The antisense RNA spans the 5’UTR and 5’ end of the rps14 gene, which is transcribed from the minus strand. b. The antisense transcripts of rps19 and rpl2 in S. vulgaris KRA. c. The psbN gene is transcribed in sense orientation less than in antisense orientation.
RNA editing extent in S. vulgaris plastids. The haplotypes KOV and KRA are compared. Mean values (±SD) calculated from six individuals are plotted, a 95% confidence band for the observed data is given in gray within plot.

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

Additional_file_1.xlsx
Figure_S3.pdf
Figure_S1.pdf
Figure_S2.pdf