BMC Plant Biology

CrossMark

Comparative analysis of inverted repeats of polypod fern (Polypodiales) plastomes reveals two hypervariable regions

Maria D. Logacheva[1], Anastasiya A. Krinitsina[1], Maxim S. Belenikin[1,2], Kamil Khafizov[2,3], Evgenii A. Konorov[1,4], Sergey V. Kuptsov[1] and Anna S. Speranskaya[1,3*]

Abstract

Background: Ferns are large and underexplored group of vascular plants (~ 11 thousands species). The genomic data available by now include low coverage nuclear genomes sequences and partial sequences of mitochondrial genomes for six species and several plastid genomes.

Results: We characterized plastid genomes of three species of *Dryopteris*, which is one of the largest fern genera, using sequencing of chloroplast DNA enriched samples and performed comparative analysis with available plastomes of Polypodiales, the most species-rich group of ferns. We also sequenced the plastome of *Adianthum hispidulum* (Pteridaceae). Unexpectedly, we found high variability in the IR region, including duplication of *rrn16* in *D. blanfordii*, complete loss of *trnI*-GAU in *D. filix-mas*, its pseudogenization due to the loss of an exon in *D. blanfordii*. Analysis of previously reported plastomes of Polypodiales demonstrated that *Woodwardia unigemmata* and *Lepisorus clathratus* have unusual insertions in the IR region. The sequence of these inserted regions has high similarity to several LSC fragments of ferns outside of Polypodiales and to spacer between *tRNA-CGA* and *tRNA-TTT* genes of mitochondrial genome of *Asplenium nidus*. We suggest that this reflects the ancient DNA transfer from mitochondrial to plastid genome occurred in a common ancestor of ferns. We determined the marked conservation of gene content and relative evolution rate of genes and intergenic spacers in the IRs of Polypodiales. Faster evolution of the four intergenic regions had been demonstrated (*trnA- or*f42, *rrn16-rps12*, *rps7-psbA* and *ycf2-trnN*).

Conclusions: IRs of Polypodiales plastomes are dynamic, driven by such events as gene loss, duplication and putative lateral transfer from mitochondria.

Background

Chloroplast genomes (plastomes) of land plants are generally conserved in their size, gene content and order. The evolutionary origin of chloroplasts, as well as mitochondria, traces back to ancient endosymbiotic bacteria that consequently greatly reduced the genome, and contain only a small proportion of the ancestor's genes [1, 2].

Plant plastomes possess a quadripartite structure composed of large single-copy (LSC) and small single-copy (SSC) regions divided by two parts of inverted repeat (IR) [2, 3]. Plastome size of higher plants is usually around 150,000 bp in length and comprise approximately 120–130 genes, among which about 75 genes encode proteins of photosystem I and II, as well as for other proteins, involved in photosynthesis [see, for example, [4]], while other genes encode ribosomal RNA and proteins and transfer RNA. The highest deviations of the gene order and content in plastomes of land plants have been reported in non-photosynthetic species, in which extremely reduced plastomes were found - up to 11 Kbp [5]. However, in

* Correspondence: hanna.s.939@gmail.com
[1]M.V. Lomonosov Moscow State University, 119991 Moscow, Russia
[3]Federal Budget Institution of Science Central Research Institute of Epidemiology of The Federal Service on Customers, 111123 Moscow, Russia
Full list of author information is available at the end of the article

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 62 of 103

photosynthetic plants lineage-specific gene losses, as well as translocations that change the gene order were also observed [6–8].

IRs are usually regarded as the most stable part of the plastome. Indeed, the substitution rate of the IR sequences is the lowest compared to single copy regions [3]; the plastomes where IRs are absent (e.g. IRLC clade of Fabaceae) exhibit elevated substitution rates [9] and vice versa, genes that were translocated from SC to IR slow down their substitution rate [10]. IRs typically range in size from 15 to 30 kbp and contain a core set of genes consisting of four rRNA genes (*4.5S*, *5S*, *16S* and *23S rRNA*), tRNA genes (*trnAUG*, *trnI-GAU*, *trnN-GUU*, *trnR-ACG* and *trnV-GAC*). The IRs of many land plants also contain a number of other genes as a result of lineage-specific expansions and contractions [3]. Few evolutionary lineages demonstrate large-scale expansions (exceeding several kbp and containing numerous genes) of the IR [3, 11]. In particular, several overlapping inversions affect size, gene content and order of the IR in leptosporangiate ferns, the clade that includes most fern species [8, 10].

Another evolutionary event that can affect plastome structure is horizontal gene transfer (HGT). HGT between nucleus, mitochondria and plastids has been shown to occur with a high rate and contributed significantly to the plant genome evolution by relocating and refashioning of the genes and consequently contributing to genetic diversity. Transfers of DNA fragments from the mitochondria or plastids to the nucleus are the most common reported ones [12–16]. The mitochondrial genomes are also often invaded by plastome-derived sequences; the presence of DNA from nuclear genomes also has been shown in a number of lineages of flowering plants [17–24] and ferns [25]. Translocations of mtDNA fragments to plastid genome are much rarer. Currently, only two cases, all from flowering plants, are known. In *Daucus carota* (order *Apiales*, Apiaceae) plastome a ~1.5 Kb region with high similarity to *Vitis vinifera* (order *Vitales*) mitochondrial sequence was found [26]. This region did not contain any typical plastid genes. Characterization of *Daucus* mitogenome and screening of plastomes of other *Apiaceae* suggest that it was inserted from the mitochondrial genome to the plastome in a common ancestor of the genus *Daucus* [27]. Another example is the horizontal gene transfer is a 2.4-kb segment of mitochondrial DNA into the *rps2–rpoC2* intergenic spacer of the plastome of *Asclepias syriaca* (Apocynaceae) [28]. Thus, unlike the mitochondrial genomes, which are affected by insertions of plastid and nuclear sequences, the plastomes of flowering plants are infrequently profited by DNA transfer from the other cell compartments [22, 29, 30].

Extant ferns are non-seed vascular plants for which 45 families are currently known (with approx. 280 genera), of which more than half belong to the order Polypodiales,

in line with the classification of [31]. The majority of Polypodiales species fall into two large sister clades - Eupolypods I and Eupolypods II, and the remaining to families *Pteridaceae*, *Dennstaedtiaceae*, *Saccolomataceae*, *Lindsaeaceae* (basal clade) [32]. Wolf et al. [8] for the first time examined structure of the plastome across few widely ranged representative fern taxa but no one has represented the Polypodiales. Zhu et al. [3] analyzed the evolutionary rate and shifts in IR boundaries of land plants, including seven ferns but also no Polypodiales were considered. Raman et al. [33] characterized *Cyrtomium falcatum* plastid genome and found some differences with congeneric species *C. devexiscapulae* in tRNA gene content and start codons. For 24 fern samples, five of which were Polypodiales species (other represented ten extant orders), the part of LSC region (*rpoB-psbZ*) was analyzed and considerable genomic changes for distant species (belonging to different orders) were found [34]. A comparison of a few taxon-wide fern plastomes (of Lycopodiophyta, Psilotopsida, Equisetopsida, Marattiopsida and Polypodiopsida) showed that some lineages have experienced multiple IR changes including expansions and inversions while others demonstrated the stasis [35]. Recently many new sequences of fern plastomes (including Polypodiales species) were released. However, the corresponding study reports only the results of phylogenetic analysis of these sequences, without detailed analysis of their gene content and structure [36].

This clearly indicates that the diversity of plastome structures in ferns is insufficiently explored. With this premise, we characterized four additional plastome sequences from Polypodiales, three from *Dryopteris* and one from *Adianthum*, and performed comparative analysis of all available fern plastomes.

Results

We sequenced and assembled de novo new plastome sequences for three Eupolypods I species: *Dryopteris filix-mas*, *Dryopteris blanfordii* and *Dryopteris villarii*. The plastomes have typical quadripartite structure, are similar in their size (148,568, 152,945 and 148,727 bp, respectively). A total of 130 genes were annotated by DOGMA for *D. filix-mas*, including 91 protein-coding genes (5 of them are duplicated in IRs), 25 tRNA genes (5 of them are duplicated in IRs) and 4–5 rRNA genes (all of them are duplicated in IRs). Plastomes of *Dryopteris* species, including *D. blanfordii*, *D. villarii*, *D. filix-mas* and previously reported *D. decipiens* were identical to each other in gene content of LSC and SSC, but differed in IRs (Fig. 1).

We also sequenced one the plastome of *Adianthum hispidulum*, from Pteridaceae, the basal group relative to eupolypods. The complete plastome sequence of *A. hispidulum* was 151,327 bp in length, consisted of an LSC (83,188 bp), SSC (21,459 bp), and IRs (23,340 bp).
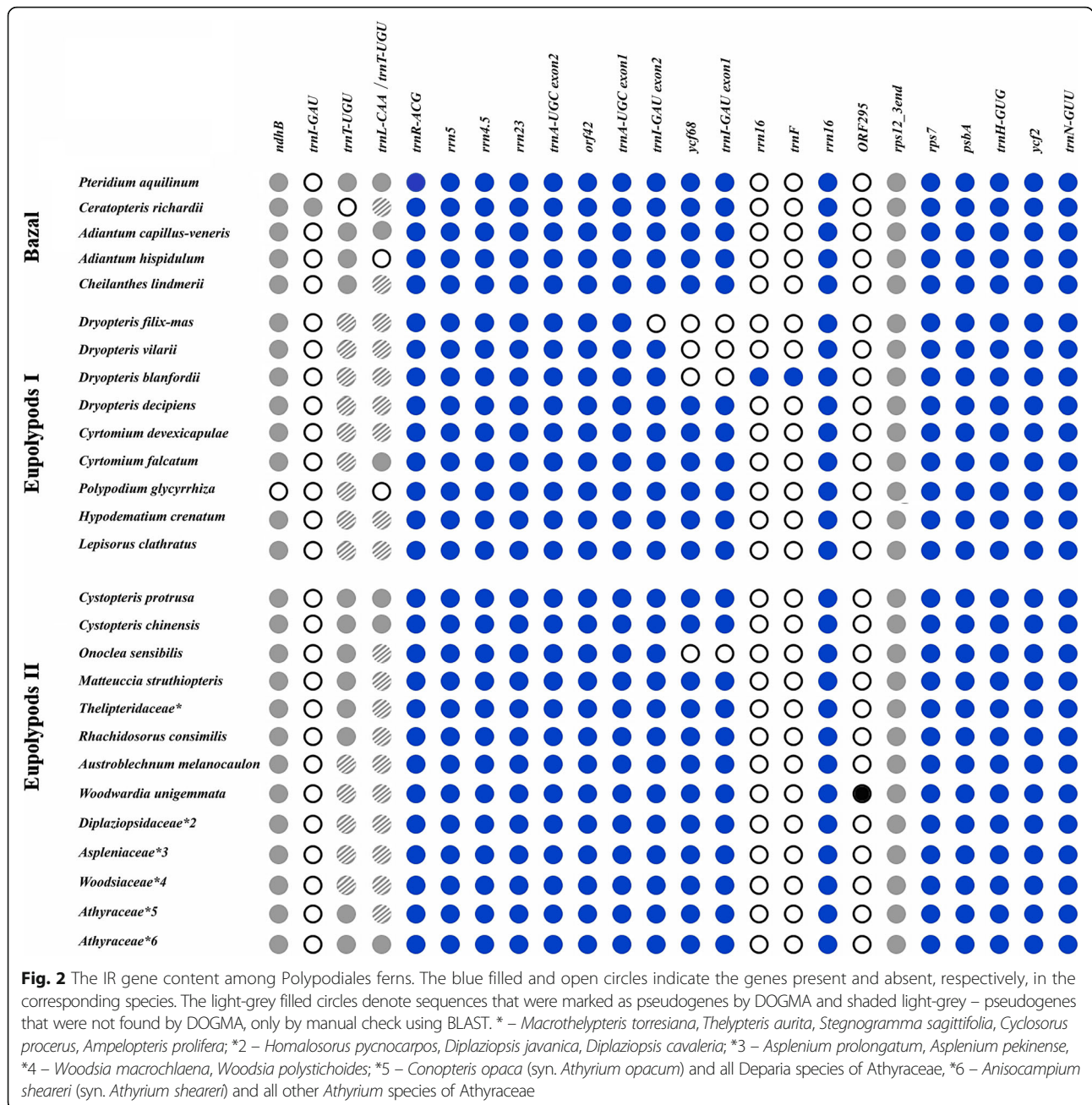
Logacheva et al. BMC Plant Biology 2017, 17(Suppl 2):255

Page 63 of 103
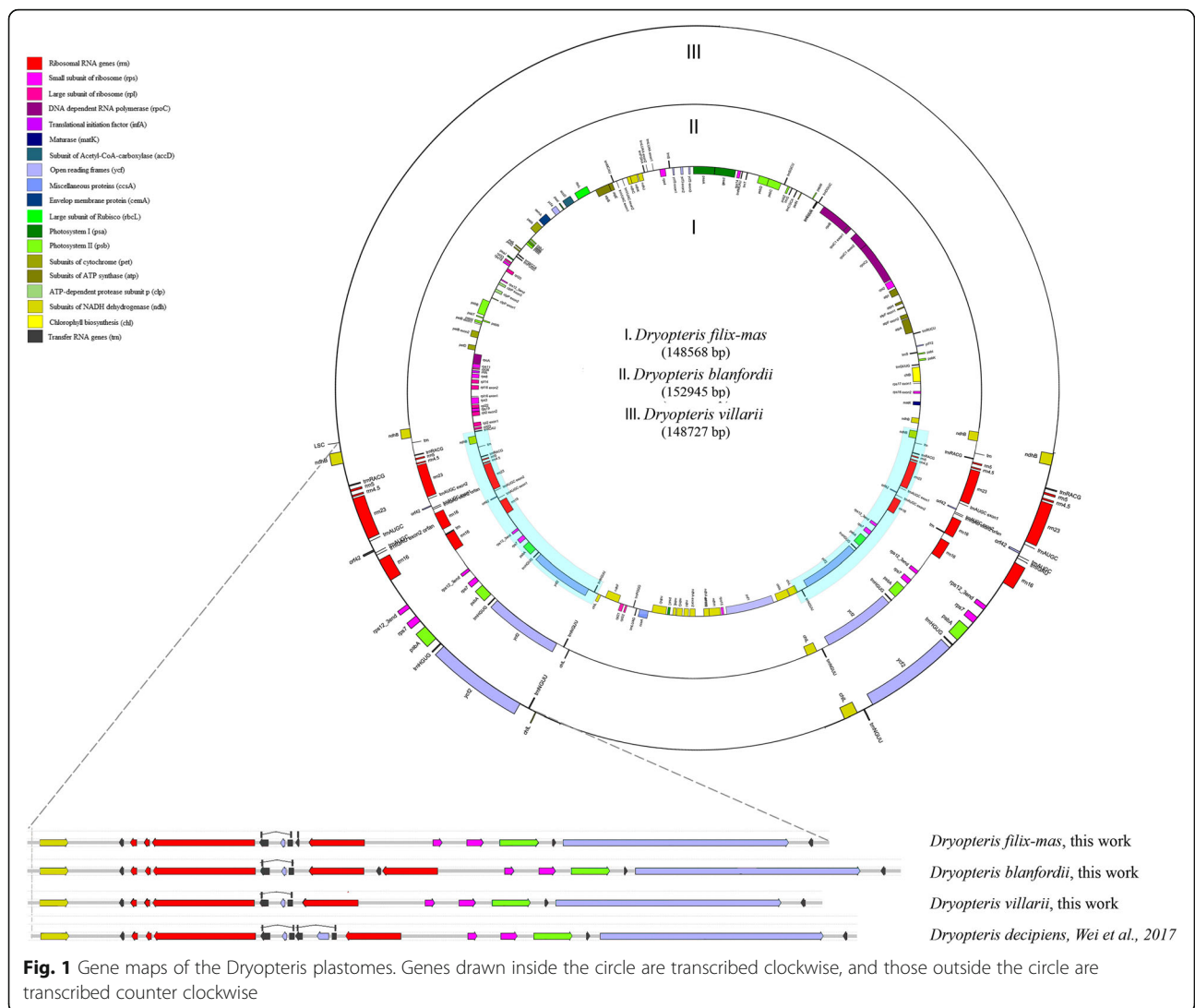
Fig. 1 Gene maps of the Dryopteris plastomes. Genes drawn inside the circle are transcribed clockwise, and those outside the circle are transcribed counter clockwise

The gene content in the plastome of *A. hispidulum* slightly differs from that of congeneric species *A. capillus-veneris* (published in [37]) – in *A. hispidulum trnT-UGU* gene (located in IR) is completely absent while in *A. capillus-veneris* it is represented by a pseudogene (Fig. 2).

Comparative dataset of IR sequences which includes all plastomes for Polypodiales available in the public sequences databases and our new data comprises 45 sequences: 31 belong to Eupolypods I, 9 – to Eupolypods II and the remaining 5 - to Pteridaceae or Dennstaedtiaceae. We have made re-annotation of published Polypodiales plastomes. The IR/SSC border of all species lies within the *ndhF* gene and IR/LSC - within *ndhB* gene. The similar IR borders were previously defined for many groups of monilophytes (Psilotales, Ophioglossales, Equisetales, Marattiales and Polypodiopsida) [35]. Commonly, chloroplast genomes of Polypodiales carry sequences (~ 246 bp) with high similarity to *ycf68* ORF in the IR regions, though

they were not previously annotated by the authors. The *ycf68* is a putatively functional gene located in the *trnI-GAU* intron, which is present in many land plant chloroplast genomes but often is not annotated as its function is still unknown [38, 39]. Another putative gene, *ORF42*, was annotated in the *trnI-GAU* intron for the all species included in the analysis. *ORF42* was found previously in the intron region of *trnA-UGC* plastid gene of some flowering plant species, for example, *Veratrum patulum* O. Loes. (Melanthiaceae) [40] and *Pelargonium* × *hortorum* L. H. Bailey (Geraniaceae) [11]. The sequence with high similarity to *ORF42* was found in the mitochondrial genome of *Phaseolus*; presumably as a result of plastid-to-mitochondrion lateral gene transfer [19].

The IR structures of Polypodiales are shown in Fig. 2 and Additional file 1. The number of genes normally varies from 14 to 16. The plastomes of Polypodiales mainly

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 64 of 103

Fig. 2 The IR gene content among Polypodiales ferns. The blue filled and open circles indicate the genes present and absent, respectively, in the corresponding species. The light-grey filled circles denote sequences that were marked as pseudogenes by DOGMA and shaded light-grey – pseudogenes that were not found by DOGMA, only by manual check using BLAST. * – *Macrothelypteris torresiana, Thelypteris aurita, Stegnogramma sagittifolia, Cyclosorus procerus, Ampelopteris prolifera*; *2 – *Homalosorus pycnocarpos, Diplaziopsis javanica, Diplaziopsis cavaleria*; *3 – *Asplenium prolongatum, Asplenium pekinense*, *4 – *Woodsia macrochlaena, Woodsia polystichoides*; *5 – *Conopteris opaca* (syn. *Athyrium opacum*) and all Deparia species of Athyraceae, *6 – *Anisocampium sheareri* (syn. *Athyrium sheareri*) and all other *Athyrium* species of Athyraceae

accumulated the gene number variability within IR in two areas: from 0 to ~3 Kbp and from ~7 to ~11,5 Kbp regions of IR. These regions also demonstrate lower sequence similarity. The other two regions - from ~3 to ~7 Kbp and from ~11,5 Kbp till the end of IR - were largely conservative in gene number and sequence.

The 0 to ~3 Kbp region contains tRNA (pseudo)gene (Fig. 2). Though it is annotated as functional two-exon *trnT-UGU* gene in many species (e.g. 33, 37), the tRNA structure prediction with tRNAscan-SE does not support its functionality. In other species it is annotated as pseudogene (*trnT-UGU* or *trnL-CAA*) or completely missing.

Manual check however shows that pseudogene is present in all species analyzed (Fig. 2) except for *Adianthum hispidulum* where it was completely lost.

In the ~7 to ~11,5 Kbp variable region three genes show partial or full deletions or duplications (*trnI-GAU*, *ycf68*, *rrn16*) (Fig. 2) (Additional file 1). Some species of Polypodiales have partially or completely lost *trnI-GAU* and *ycf68* (located in intron of *trnI-GAU*) namely: *O. sensibilis* (Onocleaceae) and three of four *Dryopteris* species (*D. blanfordii, D. villarii, D. filix-mas,* Dryopteridaceae). In *D. filix-mas trnI-GAU* and *ycf68* are completely lost. *D. blanfordii* has partially lost the *trnI-GAU* gene (the intron

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255
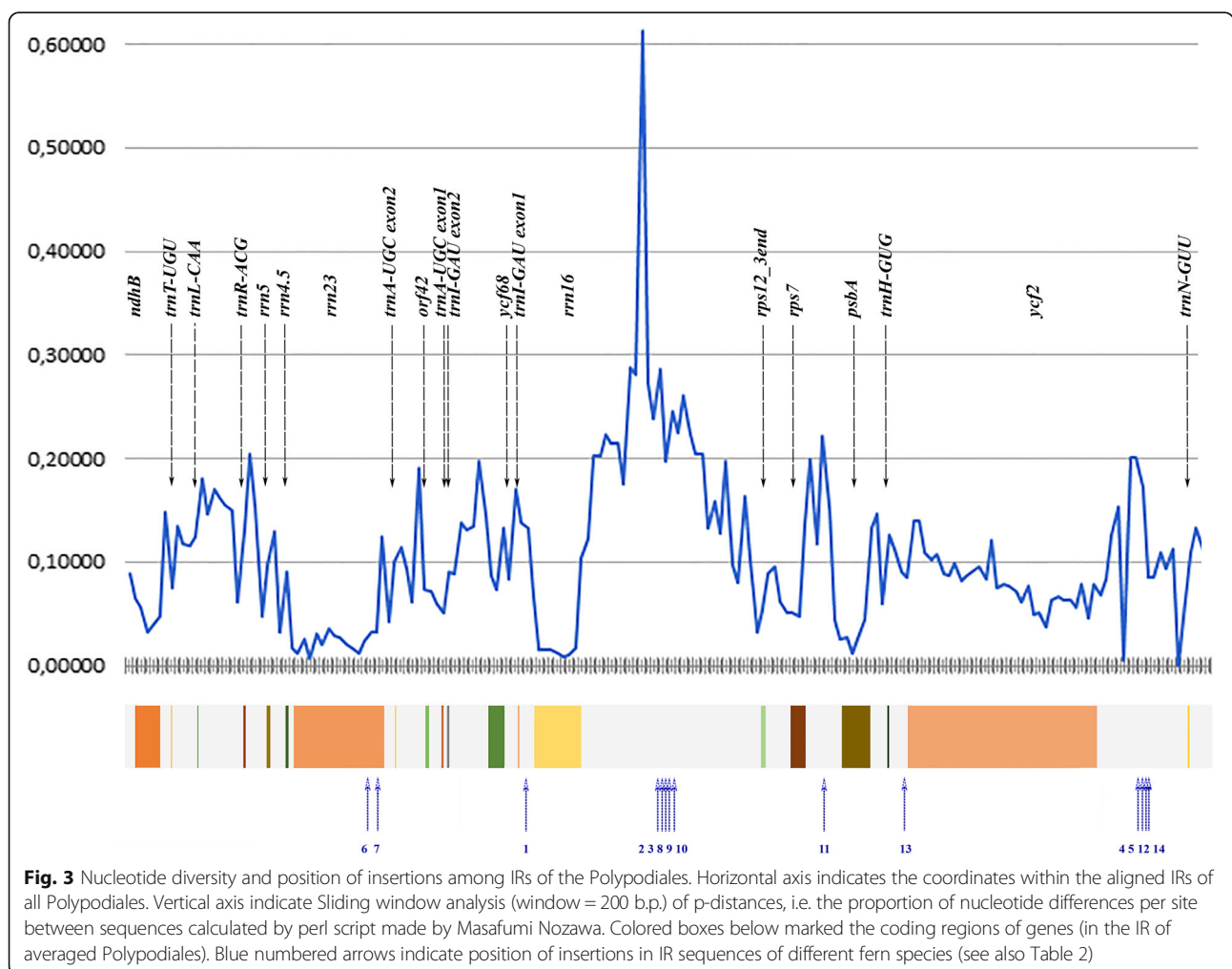
Page 65 of 103

and one of exons were deleted), but has a duplication of the large part of *rrn16* gene. *D. villarii* has lost intron and one exon of *trnI-GAU* but no *rrn16* duplication. As result of deletions/duplications the IR size of *Dryopteris* species varied: IR of *D. filix-mas* and *D. villarii* were ~1450–1570 bp shorter but *D. blanfordii*, on the contrary, had IRs 647 bp longer (*Cyrtomium* species were used as reference). Surprisingly *D. decipiens,* reported in [36], has no deletions in this region. *O. sensibilis* (Onocleaceae) plastome has a deletion of *ycf68-trnI-GAU* region. *Onoclea* and *Dryopteris* species belong to different clades: Eupolypod II and Eupolypod I. Therefore, the losses of *ycf68-trnI-GAU* regions are independent events.

Due to the deletions and duplications IR size in Polypodiales varies from 22 Kbp (*Ceratopteris richardii* [KM052729]) to 26,9 Kbp (*Cystopteris protrusa,* [KP136830]. In particular, we found the most of the large insertions (370 bp and more) in the highly variable intergenic spacers mentioned above: intergene 14 (between *rrn16* and *rps12*), intergene 16 (between *rps7* and *psbA*) and intergene 19 (between *ycf2* and *trnN-GUU*), see Fig. 3. For the insertion's

length, localization and similarity to known high plant sequences see Table 1.

An unusual 1663 bp insertion was found *rrn16* and *rps12* genes of *Woodwardia unigemmata* (coordinates 93,404... 95067). *W. unigemmata* is a fern of family Blechnaceae (Eupolypods II), whose plastome was sequenced by Lu et al., 2015 [41], and no genes were annotated in this region previously. The large part of *W. unigemmata* insertion has high sequence similarity to the insertion of *Lepisorus clathratus* (Polypodiaceae, Eupolipods I) located in the same region of IR, between *rrn16* and *rps12* genes. About 1160 bp of *W. unigemmata* and *L. clathratus* insertions demonstrated 65–78% similarity to each other. No similar sequences were found in IR of other Polypodiales, except for small (about 150 bp) sequence in *Matteuccia struthiopteris* (Onocleaceae, Eupolypods II) which has 62–66% similarity to *L. clathratus* and *W. unigemmata* insertions (further called WL-sequences).

Surprisingly the WL-like sequences were found in LSC regions of species from distant taxonomic groups of ferns, outside Polypodiales (Table 2). Firstly, small part

Fig. 3 Nucleotide diversity and position of insertions among IRs of the Polypodiales. Horizontal axis indicates the coordinates within the aligned IRs of all Polypodiales. Vertical axis indicate Sliding window analysis (window = 200 b.p.) of p-distances, i.e. the proportion of nucleotide differences per site between sequences calculated by perl script made by Masafumi Nozawa. Colored boxes below marked the coding regions of genes (in the IR of averaged Polypodiales). Blue numbered arrows indicate position of insertions in IR sequences of different fern species (see also Table 2)

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 66 of 103

Table 1 Relative evolution rate *(ERaBLE output was normalized by rate of evolution in concatenated RNA- and protein-coding sequences)* in different regions of IR

Region	All Polypodiales	Eupolypods 1	Eupolypods 2	Dennstaedtiaceae + Pteridium
Intergene1	1,27	1,43	1,22	1,65
ndhB	2,46	3,50	2,78	2,83
Intergene2	11,52	12,55	10,86	14,59
trnT-UGU	16,59	26,18	3,11	31,63
Intergene3	2,68	3,32	2,20	3,45
trnL-CAA	2,20	3,24	2,44	2,53
Intergene3p	*5,43*	*8,53*	*6,47*	*5,13*
trnR-ACG	85,25	0,00	138,59	0,00
Intergene4	10,11	13,42	12,65	10,99
rrn5	0,15	0,22	0,27	0,11
Intergene5	7,25	0,46	8,02	7,12
Intergene6	2,35	2,49	3,46	0,93
rrn23	0,78	9,81	1,77	0,30
Intergene7	2,97	2,51	4,30	2,26
trnA-UGC	0,00	0,00	0,00	0,00
Intergene8	**22,93**	**10,48**	**59,98**	**11,34**
orf42	530,31	1221,25	3,38	3,77
Intergene9	0,74	1,05	0,86	0,73
trnA-UGC2	0,02	0,04	0,00	0,00
Intergene10	2,67	16,92	1,71	1,70
trnI-GAU	0,00	0,00	0,00	0,00
Intergene11	2,07	3,40	2,61	2,22
ycf68	2,87	2,55	8,39	1,07
Intergene12	68,77	41,48	245,28	15,63
trnI-GAU2	2,31	5,57	0,00	0,00
Intergene13	3,01	4,05	3,76	3,10
rrn16	0,42	0,18	0,97	0,20
Intergene14	**9,37**	**27,47**	**11,23**	**8,26**
rps12	1,88	2,13	2,77	1,44
Intergene15	1,43	1,48	1,24	1,59
rps7	2,20	2,32	2,65	1,83
Intergene16	**7,31**	**7,69**	**7,79**	**7,57**
psbA	1,42	1,89	1,78	1,30
Intergene17	2,57	4,21	2,63	2,91
trnH-GUG	0,35	0,86	0,00	0,01
Intergene18	3,14	1,92	3,36	4,20
Intergene19	**5,47**	**5,07**	**4,97**	**6,56**
trnN-GUU	0,00	0,00	0,00	0,00
Intergene20	*3,02*	*5,42*	*3,27*	*4,03*

The bolded rows marked intergenic spacers there LRT for the branch site model A was significantly better than the model B. (Model A was used for the ribosomal RNA and non-coding regions and supposes that substitutions in a DNA sequences of different tree branches are homogenous, model B – the same but supposes non-homogenous substitutions, see also materials and methods)

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 67 of 103

Table 2 Coordinates of insertions with size 45 bp more inside Polypodiales IRs

№	Clades	Species	Insertion IR coordinates		Insertion size	Location	Genbank accession number	Description
			start	end				
1	1Eu*	*Dryopteris blanfordii*	7615	10,000	2385	Large insertion between *trnI-GAU* and *rrn16*	This work	Contains duplicated rrn16 and trnF-GAA genes
2	1Eu	*Polypodium glycyrrhiza*	9546	11,764	2218	intergene14 (*rrn16-rps12*)	KP136832	~ 700 bp part of insertion has homology to insertion of *L. clathratus*
3	1Eu	*Lepisorus clathratus*	10,863	14,073	3210	intergene14 (*rrn16-rps12*)	KY419704	Part if insertion has homology to part of WL and another to insertion of *P. glycyrrhiza*
4	1Eu	*Cyrtomium devexiscapulae*	22,596	22,968	372	intergene19 (*ycf2-trnN-GUU*)	KT599100	No significant homology to known sequences was found
5	1Eu	*Cyrtomium falcatum*	22,570	22,942	372	intergene19 (*ycf2-trnN-GUU*)	KP189363	No significant homology to known sequences was found
6	2Eu	*Cystopteris protrusa*	6100	6286	186	*rrn23*	KP136830	Duplication inside IR (homologous to 6403–6560)
7	2Eu	*Cystopteris protrusa*	6897	9498	2631	*rrn23*	KP136830	~ 950 bp fragment of insertion has 75–76% homology to mitochondrion of different mosses
8	2Eu	*Asplenium pekinense*	10,447	12,688	2316	intergene14 (*rrn16-rps12*)	KY427331	~ 200 bp of insertion has homology to *rrn16-rps12* spacer of *L. clathratus*
9	2Eu	*Rhachidosorus consimilis*	10,780	11,683	903	intergene14 (*rrn16-rps12*)	KY427356	No significant homology to known sequences was found
10	2Eu	*Woodwardia unigemmata*	11,015	12,678	1663	intergene14 (rrn16-rps12)	KT599101	WL sequence
11	2Eu	*Matteuccia struthiopteris*	13,885	14,255	370	intergene16 (*rps7-psbA*)	KY427353	Part if insertion has homology to part of WL sequence
12	2Eu	*Asplenium prolongatum*	22,083	23,414	1331	intergene19 (*ycf2-trnN-GUU*)	KY427332	The most of insertion has homology to part of *rrn16-rps12* spacer of *Polypodium glycyrrhiza*
13	Pt	*Ceratopteris richardii*	14,185	14,499	314	intergene18 (*trnH-ycf2*)	KM052729	No significant homology to known sequences was found
14	Pt	*Cheilanthes lindheimeri*	22,714	25,096	2382	intergene19 (*ycf2-trnN-GUU*)	HM778032	No significant homology to known sequences was found

1Eu Eupolypods I, *2Eu* Eupolypods II, *Pt* Pteridaceae

of WL sequence has similarity to the fragment of LSC plastome of *Plagiogyria* species (*Plagiogyria* is a single genus in monotypic family Plagiogyriaceae, Cyatheales, see for example «Flora of China» [42]). To be more precise, the 772 bp part of WL sequence has 67% identity to the region of the *P. glauca* plastome (coordinates 29,128...29895, KP136831) and to *P. japonica* plastome (partial sequence, coordinates 4503–5273, HQ658099). In both species sequences with homology to WL lie within *trnD-GUC-psbM* intergenic spacer. Other ferns that also contain WL-like sequences in LSC are *Ophioglossum californicum* (KC117178) [34] and *Mankyua chejuensis* (KP205433). Both these species belong to Ophioglossales (basal ferns); they are distant from Polypodiales (core leptosporangiate ferns) and from Cyatheales. The fragments with similarity to WL sequences in Ophioglossales ferns have about 255–275 bp length and are located in the different regions of LSC. In *O. californicum* (KC117178) it is found in intergenic spacer between the *trnT-GGU* and

trnfM-CAU genes, in *M. chejuensis* (KP205433, JF343520) it is also located in LSC region but in the other intergenic spacer (between *trnL-UAA* and *rps4*). It is annotated as *ORF295* (Fig. 4). Altogether, this suggests that there is a translocation of DNA fragments from LSC to IR (or vice versa) during evolution of fern plastomes.

Interestingly, the part of the WL sequence has high identity (67%) to the 203-bp fragment of about 5,9 Kb intergenic spacer between *tRNA-CGA* and *tRNA-TTT* genes of mitochondrial genome of *Asplenium nidus* (partial sequence, coordinates 3188...2986, AM600641) (Table 2). *Asplenium* is a genus that belongs to the same clade of Polypodiales - eupolypods II as *Woodwardia* but to the other family - Aspleniaceae. We performed similarity search of WL-sequence against fern mitochondrial contigs available in Utah State University Repository [45; http://digitalcommons.usu.edu/fern_genome/]. Two hits were found in *Plagiogyria formosana* - 244 bp (identity 72%, contig №439, coordinates 5336–5585) and 277 bp (identity

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 68 of 103

Fig. 4 Location of WL-homologous fragments in plastomes of Ophioglossales, Cyatheales and Polypodiales ferns

72%, contig №93, coordinates 1–277). No hits were detected in the contigs of another available fern species (*D. conjugata, Pteridium aquilinum, C. richardii, Polypodium glycyrrhiza, C. protrusa*).

We have estimated relative evolution rate of genes and intergenic spacers of Polypodiales IR. For protein-coding genes we analyzed dN/dS using branch-site model, for non-coding regions two models: homogenous substitution parameters and non-homogenous were used (see materials and methods for details). We found that for four non-coding (intergenic) regions the model, which has different substitution parameters for different branches, had significantly higher likelihood (LRT) compared to model based on assumption of identical substitution parameters in different lineages (see Table 3). All these regions are spacers: between *trnA-UGC* and *orf*42

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 69 of 103

Table 3 Coordinates, size and identity of WL-homologous regions in different species

Species	GenBank accession number	Region	Coordinates	Annotation	% Identity to Wl-homologous fragment of *W. unigemmata*	b.p.
Plagiogyria glauca plastid, partial	KP136831	LSC	29,128...29895	intergenic spacer	67%	768
Plagiogyria japonica plastid, partial	HQ658099	LSC	4503…5273	intergenic spacer	67%	770
Mankyua chejuensis plastid, complete genome	KP205433	LSC	62,408...62158	ORF295	66%	255
Ophioglossum californicum plastid, complete genome	KC117178	LSC	49,930...49660	intergenic spacer	65%	270
Asplenium nidus mitochondrion, partial	AM600641	Mitochondrion	3188…2986	intergenic spacer	67%	202

(intergene8), between *rrn16* and *rps12* (intergene14), between *rps7* and *psbA* (intergene16), and the last one between *ycf2* and *trnN-GUU* (intergene19).

Discussion

In contrast to previous observations on the stability of the IR region, we found high variability in IR sequence and gene content in Polypodiales ferns. There are two hypervariable regions – one located at the beginning of IR, 0–3 Kb, and the second is 7–11 Kb region. These regions are the subject to the similar evolutionary changes occurred independently in the different clades. The first region in most species contains tRNA pseudogene. The members of both Eupolypods I and Eupolypods II demonstrated independent deletion of *trnI-GAU-ycf68* region (i.e. *Dryopteris* and *Onoclea*). Polypodiales (together with Salviniales and Cyatheales) belong to a clade called core leptosporangiates [43]. Their plastomes acutely differ from those of eusporangiate ferns (Psilotales or whisk ferns, Ophioglossales, Marattiales, Equisetales) [8, 33]. It should be noted however that comparative analysis of fern plastomes is obfuscated by the uncertainty of the annotation of tRNA genes. This concerns, in particular, the intron-containing *trnT-UGU*, which was reported in IR (between *ndhB* and *trnR-ACG*) of several fern plastomes [37, 43, 44] and thought to be specific feature of core leptosporangiates. But intron-containing *trnT-UGU* was not found in other fern lineages [45, 46] or in any other plants outside ferns; only intronless *trnT-UGU* is present. This is unusual given that plastid tRNA genes, in contrast to protein-coding genes, have highly conserved exon-intron structure. Gao and coworkers supposed that tRNA genes may be lost repeatedly independently during evolution of ferns and probably the loss of *trnT-UGU* is the one of those events [45]. Our analysis which included manual reannotation and check using tRNA prediction program tRNAscan-SE however does not support the functionality of intron-containing IR-located *trnT-UGU* in any Polypodiales species where it was reported. In contrast, we found the *trnT-UGU* pseudogene in the IR of almost all Polypodiales. Most likely, this pseudogene is difficult to be recognized and therefore results of automatic annotation could be interpreted as gene loss. Notably, Gao and co-workers [45] compared the sequences of putative intron-containing and intronless *trnT-UGU* and it can be seen that the former are unusually divergent, much higher than expected for a functional tRNA gene. We conclude that intron-containing IR-located *trnT-UGU* is an artifact caused by the shortcomings of the automatic annotation. Moreover, two parts ("exons") of this pseudogenes can be recognized by automation annotation programs, such as DOGMA, as different tRNAs - *trnT-UGU* and *trnL-CAA*.

In the second hypervariable region, 7–11 Kb, we found an unusual insertion (the WL-sequence) in two unrelated Polypodiales – *Woodwartia unigemmata* and *Lepisorus clathrathus*. Smaller insertion with high similarity to the WL-sequence was found in the same region in *Mattheucia sthruttiopteris*. In addition, the insertions with high similarity to WL-sequence were found in plastomes of Ophioglossales (basal ferns, distant from Polypodiales) but in different position – in LSC region.

The WL-sequence has high similarity with the region of mitochondrial genome of *Asplenium nidus*. This has two possible explanations: that it is either the sequence of mitochondrial origin, which was integrated in the plastome, or the sequence of plastid origin, which was integrated into mitochondrial genome and lost from the plastid genomes of most ferns, with exception of *W. unigemmata* and *L. clathrathus*. By now we can't make a conclusion about which of these two hypotheses is true, due to the unavailability of fern complete mitochondrial genome sequences.

In any case, it is a result of the horizontal genome fragment transfer between mitochondria and plastids. Horizontal genome fragment transfer is a phenomenon, commonly observed in the pro- and eukaryotes. In plants, the presence of three genomes within a cell compartments (mitochondria, chloroplast and nucleus) leads to different possible types of intracellular genome fragments exchange: between organelles and nucleus and between mitochondria and chloroplasts, bidirectional [47]. The transfer of genetic material from organelles to the nucleus seems to be a continuing evolutionary process of the prokaryotic ancestors' genome reduction [48, 49]. Many reports asserted that

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 70 of 103

plant mitochondrial genomes are unusually prone to the introgression of alien sequences compared to chloroplast and nuclear genomes [47, 50]. There are only few data on mitochondria of ferns. No complete mitochondrial genome assemblies are available, only contigs. Multiple regions with strong sequence similarity to plastid DNA were detected by [51] but they didn't relate to the plastome sequences in the total genomic contigs of six ferns species *Dipteris conjugata* (Gleicheniales), *Plagiogyria formosana* (Cyatheales), *Pteridium aquilinum* (Dennstaedtiaceae), *Ceratopteris richardii* (Pteridaceae), *Polypodium glycyrrhiza* (eupolypods) and *Cystopteris protrusa* (eupolypods). Authors speculated that the plastome-like sequences reside within the nuclear or mitochondrial genomes [42]. Assuming this is the case, it implies that the horizontal transfer of organelle genome fragments are not rare events in the evolution of ferns.

Conclusions

In this study we investigated the structure and evolutionary stability of IRs of plastomes in Polypodiales ferns. The two regions of IRs were found to be highly variable: (i) the sequences between *ndhB* and *trnR-ACG* genes (~3 Kbp) and (ii) the fragment including the *rrn16* gene and flanking vicinity regions (~4,5 Kbp). This blinking of *trnI-CAU*, *trnT-UGU*, *ndhB* and *rps12, trnI-GAU, ycf68, rrn16* genes related to these regions was observed in different species. The plastomes of three *Dryopteris* species demonstrate dynamic process of trnI-GAU elimination/*rrn16* duplication.

Two Polypodiales species - *W. unigemmata* and *L. clathratus* - have an unusual sequence in the IR region. It demonstrates similarity to LSC spacers *trnL-rps4* of Ophioglossales and *pbsM-trnD* of Cyatheales and with the part of mitochondrial genome of *Asplenium* (Polypodiales). We suppose these features are a consequence of intraplastomic rearrangements as well as of the transfer between the chloroplast and mitochondrial genomes during the evolution of ferns.

Methods
Plant material

Mature fronds of both *Dryopteris filix-mas* (L.) Schott, *Dryopteris blanfordii* (C. Hope) C. Christensen and *Dryopteris villarii (Bellardi)* Woyn. ex Schinz & Thell were sampled from outdoor section of the Moscow State University Botanical Garden.

Dryopteris filix-mas (L.) Schott is a common fern species in the Russian forests, therefore the specimen's collection locality was stated only approximately as "in the vicinity of Moscow".

Dryopteris blanfordii (C. Hope) C. Christensen grows in *Picea* or *Abies* forests at 2900–3500 m AMSL in China (Gansu, Sichuan, Xizang, Yunnan), Afghanistan, India, Kashmir, Nepal, and Pakistan [52–54]. The parent

plant was collected in 2003 in India. Spores of the specimen were germinated under artificial conditions of the greenhouse of Botanical Garden of the Moscow State University. Developed sporophytes were then transplanted to the outdoor section of the Botanical Garden.

Dryopteris villarii (Bellardi) Woyn. ex Schinz & Thell. - subalpine species, grows on outcrops of hills, limestone cliffs, including high-mountain in Central and SouthEurope [55]. The spores, courtesy of Zürich Botanical Garden seed department (collected in natural habitat of Swiss Confederation), was germinated and specimen was germinated and grown in small greenhouse of Moscow State University Botanical Garden during 2013–2017.

Adiantum hispidulum Sw. pantropical, paleotropical species, it is distributed from eastern Africa through southern India, Thailand and the Ceylon to Pacific islands, Polynesia, New Zealand and Australasia [56–59]. The adult frond of *Adiantum hispidulum* was collected from greenhouse of Botanical Garden of Moscow State University. The voucher specimen was kept in Herbarium of Biology Department of Moscow State University.

Chloroplast genome sequencing, de novo assembly and annotation

The chloroplast DNA (cpDNA) were sequenced using the Illumina MiSeq high-throughput sequencing platform. For a sample preparation, the adult live plants were taken from the collection of the Moscow State University Botanical Garden. cpDNA was extracted from 2,6 g. (fresh weight) of fronds using the cpDNA extraction protocol [60, 61] with small modifications: after cleaning with a distilled water, the fronds were homogenized in 35 ml isolation buffer at +4 °C (Tris-HCl (pH 8,0) 50 mM, EDTA 7 mM, 1% PVP-40, NaCl 1,25 M, ascorbic acid 0,25 M, sodium metabisulfite 10 mM, Borax 0,0124 M) and the homogenate was filtered using soft wipes. The homogenate was then successively centrifuged at 200 g for 15 min at 4 °C (cell wall debris was discarded), at 1000 g for 20 min at +4 °C (the precipitate was discarded) and finally at 2000 g for 20 min at +4 °C. In the latter case, the precipitate was resuspended in 3 ml of wash buffer (Tris-HCl (pH 8,0) 50 mM, EDTA 25 mM) and carefully loaded into a 15 ml tube containing sucrose gradient consisting of 7 ml of 52% sucrose in wash buffer and overlaid 4 ml of 52% sucrose in wash buffer. The tube with the sample and sucrose gradient was centrifuged at 3500 g for 60 min at 4 °C. The interface between 52% and 30% sucrose (about 1 ml) was collected, centrifuged at 12,000 g. The pellet was resuspended in 900 μl of wash buffer and 100 μl of 10% CTAB was added for lysis (1 h, 55 °C). Then the DNA purification step was carried out using the protocol described in [62].

The TruSeq protocol (NEBNext® DNA Library Prep Master Mix Set for Illumina, E6040, NEB reagents) was

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 71 of 103

used for preparing the genomic libraries. We made PE sequence (2 × 300 bp.) with a double number of each library reads about 1.2–1.97 M. After the quality trimming with Trimmomatic [63], sequencing reads were filtered using 13 complete and 5 partial fern chloroplast genome sequences from RefSeq database and Bowtie2 [64]. Then the two contig sets were produced for both filtered and unfiltered reads sets using the Velvet Assembler [65] and MIRA4 [66]. Assembled contigs and scaffolds were selected for the next assembly if they showed similarity to the published fern chloroplast genomes. The final de novo assembly was finished through a few iterative steps. The draft sequence was manually corrected by the PE reads mapping.

We have obtained the reads of complete circulated chloroplast genomes comprising the large single-copy region (LSC), small single-copy region (SSC) and the two inverted repeat (IR) regions. Finally, mapping of the initial reads was performed to the assembly in order to check for the potential assembly artefacts. Protein-coding gene annotation in the assembled chloroplast genome was annotated by DOGMA [67]. Bowtie2, VarScan (v.2.3.7) and SAMtools/BCFtools software packages were used for mapping of the reads and variant calling [64, 68, 69].

Chloroplast genomes analysis

Genbank or ENA accession numbers of sequences included in this study are listed in Table 2.

Analysis of the complete chloroplast genomes was carried out on species sequenced in this study together with previously reported species. Nine plastomes were downloaded from the GenBank. A complete list of the analyzed species can be found in the Table 1. Firstly, sequences of all the chloroplast sequences were pair-wise aligned against each other by Kalign (www.ebi.ac.uk/Tools/msa/kalign). Phylogenetic analysis was carried out by a maximum likelihood (ML) using Mega 6.0 [70]. Comparative analysis of chloroplast genome sequences was performed by the mVista web-tool (http://genome.lbl.gov/vista/mvista/submit.shtml).

For evolution rate analysis for the each region of IRs (genes and intergenic regions separately), as well as for the concatenate of all coding sequences, the alignments was built using MUSCLE [71]. An ML tree was constructed using concatenate alignment. Substitution model with lowest BIC score was chosen using modelTest function from phangorn package [72]. The tree topology was optimized using follow parameters: the nucleotide substitution matrix, gamma, the proportion of invariant sites and gamma distribution parameter. For non-protein coding regions the tree branch lengths were calculated by two models (homogenous substitution parameters - nhomo = 1, non-homogenous - nhomo = 4) using baseml [73], then models were compared by LRT. For protein-coding regions, dN and dS were estimated for each gene using codeml from PAML package [73], dN/dS ratio in each lineage was estimated by branch and M0 model. For both baseml and codeml analysis phagorn concatenate tree with nearest neighbour interchange was used. Distance matrices were calculated using baseml/codeml trees in the ape package [74]. Then relative evolution rate for each region (coding and non-coding) was calculated using ERaBLE [75]. Sliding window analysis (window = 200 b.p.) of p-distances, i.e. the proportion of nucleotide differences per site between sequences was calculated by perl script made by Masafumi Nozawa [76].

Additional file

Additional file 1: List of genes annotated for *D. filix-mas* and *D. blanfordii* chloroplast genome by DOGMA. (XLSX 46 kb)

Availability of data and materials
The sequences of full chloroplast genomes have been deposited in the European Nucleotide Archive (ENA). Other data used in the analysis are included within the article and the additional files.

About this supplement
This article has been published as part of *BMC Plant Biology* Volume 17 Supplement 2, 2017: Selected articles from Belyaev Conference 2017: plant biology. The full contents of the supplement are available online at https://bmcplantbiol.biomedcentral.com/articles/supplements/volume-17-supplement-2.

Authors' contributions
SK selected and provided the plant material for this study; AK isolated chloroplasts and extracted cpDNA, performed phylogenetic analysis; AS and ML prepared library and performed sequencing work; AS, ML and AK performed comparative analysis of genomic data; KK and MB performed the bioinformatics analysis, assembled chloroplast genome; MB and AS prepared genomes annotation; EK – conducted evolution rate analysis; AS and ML conceived the study and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 72 of 103

Author details
[1]M.V. Lomonosov Moscow State University, 119991 Moscow, Russia.
[2]Moscow Institute of Physics and Technology, Dolgoprudny 141700, Moscow Region, Russia. [3]Federal Budget Institution of Science Central Research Institute of Epidemiology of The Federal Service on Customers, 111123 Moscow, Russia. [4]Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Moscow, Russia.

References
1. Gray MW. Origin and evolution of organelle genomes. Curr Opin Genet Dev. 1993;3(6):884–90.
2. Olejniczak SA, Lojewska E, Kowalczyk T, Sakowicz T. Chloroplasts: state of research and practical applications of plastome sequencing. Planta. 2016; 244(3):517–27.
3. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016;209(4):1747–56.
4. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134.
5. Bellot S, Renner SS. The Plastomes of two species in the Endoparasite genus Pilostyles (Apodanthaceae) each retain just five or six possibly functional genes. Genome Biol Evol. 2015;8(1):189–201.
6. Haberle RC, Fourcade HM, Boore JL, Jansen RK. Extensive rearrangements in the chloroplast genome of Trachelium caeruleum are associated with repeats and tRNA genes. J Mol Evol. 2008;66(4):350–61.
7. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 2011;28(1):583–600.
8. Wolf PG, Roper JM, Duffy AM. The evolution of chloroplast genome structure in ferns. Genome. 2010;53(9):731–8.
9. Perry AS, Wolfe KH. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. J Mol Evol. 2002;55(5):501–8.
10. Li FW, Kuo LY, Pryer KM, Rothfels CJ. Genes Translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. Genome Biol Evol. 2016;8(8):2452–8.
11. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. The complete chloroplast genome sequence of pelargonium x hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 2006;23(11):2175–90.
12. Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J. 1991;10(10):3073–8.
13. Rice DW, Palmer JD. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. BMC Biol. 2006;4:31.
14. Bock R. The give-and-take of DNA: horizontal gene transfer in plants. Trends Plant Sci. 2010;15(1):11–22.
15. Smith DR. Extending the limited transfer window hypothesis to inter-organelle DNA migration. Genome Biol Evol. 2011;3:743–8.
16. Cusimano N, Wicke S. Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. New Phytol. 2016;210(2):680–93.
17. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 2003;424(6945): 197–201.
18. Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm Amborella. Proc Natl Acad Sci U S A. 2004;101(51): 17747–52.
19. Woloszynska M, Bocer T, Mackiewicz P, Janska H. A fragment of chloroplast DNA was transferred horizontally, probably from non-eudicots, to mitochondrial genome of Phaseolus. Plant Mol Biol. 2004;56(5):811–20.
20. Xi Z, Wang Y, Bradley RK, Sugumaran M, Marx CJ, Rest JS, Davis CC. Massive mitochondrial gene transfer in a parasitic flowering plant clade. PLoS Genet. 2013;9(2):e1003265.
21. Hepburn NJ, Schmidt DW, Mower JP. Loss of two introns from the Magnolia Tripetala mitochondrial cox2 gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. Mol Biol Evol. 2012; 29(10):3111–20.
22. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW, et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. Science. 2013;342(6165):1468–73.
23. Park S, Grewe F, Zhu A, Ruhlman TA, Sabir J, Mower JP, Jansen RK. Dynamic evolution of geranium mitochondrial genomes through multiple horizontal and intracellular gene transfers. New Phytol. 2015;208(2):570–83.
24. Wang D, Wu YW, Shih AC, Wu CS, Wang YN, Chaw SM. Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. Mol Biol Evol. 2007;24(9):2040–8.
25. Davis CC, Anderson WR, Wurdack KJ. Gene transfer from a parasitic flowering plant to a fern. Proc Biol Sci. 2005;272(1578):2237–42.
26. Goremykin VV, Salamini F, Velasco R, Viola R. Mitochondrial DNA of Vitis Vinifera and the issue of rampant horizontal gene transfer. Mol Biol Evol. 2009;26(1):99–110.
27. Iorizzo M, Grzebelus D, Senalik D, Szklarczyk M, Spooner D, Simon P. Against the traffic: the first evidence for mitochondrial DNA transfer into the plastid genome. Mobile Genet Elem. 2012;2(6):261–6.
28. Straub SC, Cronn RC, Edwards C, Fishbein M, Liston A. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae). Genome Biol Evol. 2013;5(10):1872–85.
29. Koulintchenko M, Konstantinov Y, Dietrich A. Plant mitochondria actively import DNA via the permeability transition pore complex. EMBO J. 2003;22(6):1245–54.
30. Knoop V. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet. 2004;46(3):123–39.
31. Christenhusz MJM, Zhang XC, Schneider H. A linear sequence of extant families and genera of lycophytes and ferns. Phytotaxa. 2011;19:7–54.
32. Rothfels CJ, Sundue MA, Kuo LY, Larsson A, Kato M, Schuettpelz E, Pryer KM. A revised family-level classification for eupolypod II ferns (Polypodiidae: Polypodiales). Taxon. 2012;61(3):515–33.
33. Raman G, Choi KS, Park S. Phylogenetic relationships of the Fern Cyrtomium Falcatum (Dryopteridaceae) from Dokdo Island based on chloroplast genome sequencing. Genes. 2016;7(12):115. doi:10.3390/genes7120115. http://www.mdpi.com/2073-4425/7/12/115.
34. Gao L, Zhou Y, Wang ZW, Su YJ, Wang T. Evolution of the rpoB-psbZ region in fern plastid genomes: notable structural rearrangements and highly variable intergenic spacers. BMC Plant Biol. 2011;11:64.
35. Grewe F, Guo W, Gubbels EA, Hansen AK, Mower JP. Complete plastid genomes from Ophioglossum Californicum, Psilotum Nudum, and Equisetum Hyemale reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. BMC Evol Biol. 2013;13:8.
36. Wei R, Yan Y-H, Harris AJ, Kang J-S, Shen H, Xiang Q-P. Plastid Phylogenomics resolve deep relationships among Eupolypod II ferns with rapid radiation and rate heterogeneity. Genome Biol Evol. 2017;9(6):1646–57.
37. Wolf PG, Rowe CA, Sinclair RB, Hasebe M. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. DNA Res. 2003;10:59–65.
38. Logacheva MD, Shipunov AB. Phylogenomic analysis of Picramnia, Alvaradoa, and Leitneria supports the independent Picramniales. J Syst Evol. 2017;55(3):171–6.
39. Roper JM, Hansen SK, Wolf PG, Karol KG, Mandoli DF, Everett KDE, Kuehl J, Boore JL. The complete plastid genome sequence of Angiopteris Evecta (G. Forst.) Hoffm. (Marattiaceae). Am Fern J. 2007;97(2):95–106.
40. Do HD, Kim JS, Kim JH. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of Veratrum Patulum O. Loes. (Melanthiaceae). Gene. 2013;530(2):229–35.
41. Lu JM, Zhang N, Du XY, Wen J, Li DZ. Chloroplast phylogenomics resolves key relationships in ferns. J Syst Evol. 2015;53(5):448–57.
42. Lycophytes & ferns. In: Zhengyi W, Raven PH, Deyuan H, eds. Flora of China vol. 2–3. Beijing and St. Louis: Science Press and Missouri Botanical Garden Press.
43. Wolf PG, Der JP, Duffy AM, Davidson JB, Grusz AL, Pryer KM. The evolution of chloroplast genes and genomes in ferns. Plant Mol Biol. 2011;76(3–5):251–61.
44. Gao L, Yi X, Yang YX, Su YJ, Wang T. Complete chloroplast genome sequence of a tree fern Alsophila Spinulosa: insights into evolutionary changes in fern chloroplast genomes. BMC Evol Biol. 2009;9:130.
45. Gao L, Wang B, Wang ZW, Zhou Y, Su YJ, Wang T. Plastome sequences of Lygodium Japonicum and Marsilea Crenata reveal the genome organization transformation from basal ferns to core leptosporangiates. Genome Biol Evol. 2013;5(7):1403–7.
46. Kim HT, Chung MG, Kim KJ. Chloroplast genome evolution in early diverged leptosporangiate ferns. Mol Cells. 2014;37(5):372–82.

Logacheva *et al. BMC Plant Biology* 2017, **17**(Suppl 2):255

Page 73 of 103

47. Gao C, Ren X, Mason AS, Liu H, Xiao M, Li J, Fu D. Horizontal gene transfer in plants. Funct Integr Genomics. 2014;14(1):23–9.

48. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 2004; 5(2):123–35.

49. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. Nat Rev Genet. 2015;16(8):472–82.

50. Archibald JM, Richards TA. Gene transfer: anything goes in plant mitochondria. BMC Biol. 2010;8:147.

51. Wolf PG, Sessa EB, Marchant DB, Li FW, Rothfels CJ, Sigel EM, Gitzendanner MA, Visger CJ, Banks JA, Soltis DE, Soltis PS, Pryer KM, Der JP. An exploration into Fern genome space. Genome Biol Evol. 2015;7(9):2533–44.

52. Mir SA, Mishra AK, Reshi ZA, Sharma MP. Four newly recorded species of Dryopteridaceae from Kashmir valley, India. Biodiversitas. 2014;15(1):6–11.

53. Mir SA, Mishra AK, Pala SA, Reshi ZA, Sharma MP. Ferns and fern allies of district Shopian, Kashmir Valley, India. Biodiversitas. 2015;16(1):27–43.

54. Fraser-Jenkins CR. A monograph of Dryopteris (Pteridophyta: Dryopteridaceae) in the Indian subcontinent, Bulletin of the British museum (natural history). London: British Museum (Natural History); 1989. p. 386–9.

55. Olsen S: Dryopteris villarii. In: Encyclopedia of Garden Ferns. China: Timber Press Inc.; 2007.

56. Parris BS. Adiantum Hispidulum Swartz and A. Pubescens Schkuhr (Adiantaceae: Filicales) in New Zealand. N Z J Bot. 1980;18:503–6.

57. Hemp A. Ecology of the pteridophytes on the southern slopes of Mt. Kilimanjaro. Plant Ecol. 2002;159:211–39.

58. Lu JM, Wen J, Lutz S, Wang YP, Li DZ. Phylogenetic relationships of Chinese Adiantum based on five plastid markers. J Plant Res. 2012;125(2):237–49.

59. Boonkerd T, Pollawatn R. Note on Adiantum Hispidulum (Pteridaceae), a new record species to Fern Flora of Thailand Songklanakarin. J Sci Technol. 2013;35(5):513–6.

60. Shi C, Hu N, Huang H, Gao J, Zhao YJ, Gao LZ. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. PLoS One. 2012;7(2):e31468.

61. Vieira LD, Faoro H, Fraga HPD, Rogalski M, de Souza EM, Pedrosa FD, Nodari RO, Guerra MP. An improved protocol for intact chloroplasts and cpDNA isolation in conifers. PLoS One. 2014;9(1):e84792.

62. Krinitsina AA, Sizova TV, Zaika MA, Speranskaya AS, Sukhorukov AP. A rapid and cost-effective method for DNA extraction from archival herbarium specimens. Biochemistry(Mosc). 2015;80(11):1478–84.

63. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

64. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–U354.

65. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18(5):821–9.

66. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WEG, Wetter T, Suhai S. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 2004;14(6):1147–59.

67. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004;20(17):3252–5.

68. Koboldt DC, Zhang QY, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568–76.

69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

70. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.

71. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

72. Schliep KP. Phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27(4):592–3.

73. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.

74. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20(2):289–90.

75. Binet M, Gascuel O, Scornavacca C, Douzery EJ, Pardi F. Fast and accurate branch lengths estimation for phylogenomic trees. BMC Bioinformatics. 2016;17:23.

76. Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in drosophila species. Genome Biol Evol. 2010;2:180–9.