ISSN 0096-3925, Moscow University Biological Sciences Bulletin, 2013, Vol. 68, No. 2, pp. 58–62. © Allerton Press, Inc., 2013.
Original Russian Text © O.A. Zverkov, L.Y. Rusin, A.V. Seliverstov, V.A. Lyubetsky, 2013, published in Vestnik Moskovskogo Universiteta. Biologiya, 2013, No. 1, pp. 8–13.

GENETICS

Study of Direct Repeats in Micro Evolution of Plant Mitochondria and Plastids Based on Protein Clustering

O. A. Zverkov[a], L. Y. Rusin[b], A. V. Seliverstov[a], and V. A. Lyubetsky[a]
[a] *Institute for Information Transmission Problems (Kharkevich Institute), Russian Academy of Sciences, Moscow, Russia*
[b] *Department of Biology, Moscow State University, Moscow, Russia*
e-mail: zverkov@iitp.ru, slvstv@iit.ru, rusin@iitp.ru, lyubetsk@iitp.ru
Received August 25, 2011

Abstract—The study focuses on insertions of perfect direct repeats of words of arbitrary length in plastomes and mitochondriomes. The approach is exemplified using seed plants. Plastomes of close species were analyzed to further develop and refine published evidence for the evolution of non-coding DNA. The results suggest that perfect repeats are common elementary events resulting from replication errors—duplication of DNA. The role of such duplications in the evolution of the plastome is discussed.

Keywords: inserts of perfect direct repeats, plastids, mitochondria, seed plants, microevolution of non-coding regions of DNA, clustering of plastid proteins.

DOI: 10.3103/S0096392513020119

INTRODUCTION

The problem of finding imperfect repeats of DNA ("words") is considered widely. The search is often performed with the REPuter program [1]. In [2], it was used for plastome of *Phoenix dactylifera* and 11 direct and inverted imperfect repeats with repeated words of 30 bp size was found. In [3], direct imperfect repeats with repeat words of 10 to 100-bases in plastomes of five species of *Oenothera* were detected. In [2], imperfect repeats in two species of *Silene* were discussed. During sequencing of a new plastome, imperfect repeats are usually analyzed [4–6]. There are many examples of this type of result that address individual genomes, in particular plastome, and not short-repeated words (otherwise there is exceeding expectance) and repeats are usually imperfect (accuracy is approximately 90%).

In [7], the evolution of large inverted repeats based on a large number of newly sequenced plastomes was reviewed. This is not related to the study of direct perfect repeats which, as our results demonstrated, have a relatively short length. Study [8] is devoted to polymorphisms in the almost identical mitochondrial genomes of related species *Oryza* and *Brassica*; previously, *Beta vulgaris* was also investigated. This study compares mitochondrial types of large taxonomic groups that are more distant from each other.

Investigation of words of *arbitrary lengths of perfect repeats* is needed, since the model of substitutions and insertions of independent single nucleotide discussed, for example in [9], cannot be applied for words with lengths of four or more bp in non-coding regions of the genome, while these inserts are common, for example in chloroplasts.

Studies [9, 10] describe major evolutionary events in non-coding chloroplast DNA and show a high frequency of perfect direct repeat insertions (PDRIs) and, especially, repeated single-base insertions. However, these analyses [9, 10] used only very peculiar short genome regions, e.g., the first group of gene *trnL* introns in asterids. These inserts were used for classification of tree species [11].

We consider the problem of finding and counting the number of perfect direct repeat insertions (PDRIs) of arbitrary length in the non-coding regions, including introns, in plastids and mitochondria. The approach is exemplified with seed plants. We emphasize that this problem is not limited to finding repeats in individual sequences and requires *multiple sequence alignment* to be discerned, for example, an indigenous repeat from an acquired insertion in a sequence. To clarify the difference between repeats and acquired insertion, we give the following definition. Insertion of the direct repeat occurs as multiple alignments of the nucleotide sequences with two lines: one ("pattern") contains perfect direct repeat φ of word φ, in another line φ aligned to φ– or –φ, where '–' corresponds to complete absence of all letters in the word φ. For the pattern and each sequence the following condition met: φ aligned by one of the four types of sequence: φ, φ, –φ or ––. Figure 1 exemplifies alignment of 4-bp PDRI; it is part of the alignment in Fig. 2. Figure 2 shows an example of imperfect repeat in *Keteleeria davidiana* that probably resulted from the first inser-

```
TGACTCTTTCAAG----ATTCATTCATCTCT
TGAATCTTGAAAG--------ATTCATCTCT
TGGATCTTGAAAG--------ATTCATCTCT
TGGATCTTGAAAG----ATTCATTCATCTCT
TGGATCTTGAAAGAAAGATTCATTCATCTCT
```

Fig. 1. Multiple alignment of *psbM* 5'-leader regions. Repeated word φ is in bold.

tion of perfect repeat of words of length 11 bp, followed by a substitution in one of the positions.

MATERIALS AND METHODS

A large-scale search for PDRIs in plastids was conducted across families of seed plants with at least two species with a completely sequenced plastome. In brackets are plastome accession numbers in GenBank. **Acoraceae:** *Acorus calamus* (NCJ307407), *Acorus americanus* (NC_010093). Asteraceae: *Lactuca sativa* (NC_007578), *Helianthus annuus* (NC_007977), *Guizotia abyssinica* (NC_010601), *Parthenium argentatum* (NC_013553). **Brassicaccae:** *Arabidopsis thaliana* (NC_000932), *Aethionema cordifolium* (NC_009265), *Aethionema grandiflorum* (NC_009266), *Olimarabidopsis pumila* (NC_009267), *Arabis hirsute* (NC_009268), *Barbarea verna* (NC_009269), *Capsella bursa-pastoris* (NC_009270), *Crucihimala wallichii* (NC_009271), *Draba nemorosa* (NC_009272), *Lepidium virginicum* (NC_009273), *Lobularia maritime* (NC_009274), *Nasturtium officinale* (NC_009275). **Convolvulaceae:** *Cuscuta gronovii* (NC_009765), *Cuscuta reflexa* (NC_009766), *Cuscuta obtusiflora* (NC_009949), *Cuscuta exaltata* (NC_009963), *Ipomoea purpurea* (NC_009808). **Fabaceae:** *Lotus japonicas* (NC_002694), *Medicago truncatula* (NC_003119), *Glycine max* (NC_007942), *Phaseolus vulgaris* (NC_009259), *Cicer arietinum*

```
Ca  TTAGTTAGATTTTGTTATTCGAACCAATTTTATAAATT---------------------------ACATCTCTTTCA
Cd  TTAGTTAGATTCCGTTACTCGAACCGATTCTATTAATT----------------------CAATCACATCTATTTCA
Kd  TTAGTTATATTCCATTACTCGAACCGATTCTATTCATTTTCTATTCATTTTCTATT[redacted]ATTCAATCATATCTATTTCA
Ps  TTAGTTAGATTCCGTTACTCGAACCGATTCTATCAATT---------------------------ACATCTATTTCA
Pc  TTAGTTAGATTCCGTTACTCGAACCAATTCTATCAATG-----------------------ACATAGATCTATTTCA
Pg  TTAGTTAGATTCCGTTACTCGAACCGATTCCATAGATA-----------------------AAATACATCTATTTCA
Pkr TTAGTTAGATTCCGTTACTCGAACCGATTCCATAGATA-----------------------AAATACATCTATTTCA
Pk  TTAGTTAGATTCCGTTACTCGAACCGATTCCATAGATA-----------------------AAATACATCTATTTCA
Pt  TTAGTTAGATTCCGTTACTCGAACCAATTCTATCAATG-----------------------ACATACATCTATTTCA
Cj  TCAGATTGATCCTATTGATGGAATT-ACTCCATGGATT---------------------------CAATCCATTTTT

Ca  TGACTCTTTCAAG----ATTCATTCATCTCT----------------------ATGAGATCAAATTCGAGCTATTTT
Cd  TGAATCTTGAAAG--------ATTCATCTCT----------------------ATGAGATCAAATCTCGAGCTATTT
Kd  TGGATCTTGAAAG--------ATTCATCTCT----------------------ATGAGATAAAATCTCGAGCTATTT
Ps  TGGATCTTGAAAG----ATTCATTCATCTCT----------------------ATGAAATTAAATCTCGAGCTATTT
Pc  TGGATCTTGAAAGAAAGATTCATTCATCTCTATGAGATCAAAATGAGATCAAAATGAGATCAAATCTCGAGCTATTT
Pg  TGGATCTTGAAAG----ATTCATTCATCTCT----------------------ATGAGATAAAATCTCGAGCTATTT
Pkr TGGATCTTGAAAG----ATTCATTCATCTCT----------------------ATGAGATAAAATCTCGAGCTATTT
Pk  TGGATCTTGAAAG----ATTCATTCATCTCT----------------------ATGAGATAAAATCTCGAGCTATTT
Pt  TGGATCTTGAAAGAAAGATTCATTCATCTCTATGAGATCAAAATGAGATCAAAATGAGATCAAATCTCGAGCTATTT
Cj  TTTGAC-------AAAAAAATGAGATACTCT----------------------ATGAGATCAAATCTCGAGTTATTG

Ca  T-GAACGAAGTAAAAATAAGGAGATC
Cd  TAGAACGAAGTAAAAATAAGGAGATC
Kd  T-AAACGAAGTAAAAATCAGGGGATC
Ps  TTTAACGAAGTAAAGATCAGGAGATC
Pc  TTGAACAAAGTTAAAATAAGGAGATC
Pg  TGGAACAAAGTGAAAATCAGGAGATC
Pkr TGGAACAAAGTGAAAATCAGGAGATC
Pk  TGGAACAAAGTGAAAATCAGGAGATC
Pt  TTGAACAAAGTTAAAATAAGGAGATC
Cj  T----AAAACGAAGGGAAAATCAATC
```

Fig. 2. Multiple alignment of *psbM* 5'-leader regions with triple 11 bp PDRIs and double 4 bp PDRIs. PDRIs are underlined; for *K. davidiana*, the underlined is also continuation of the repeat, which is different from repeated word in one position (it is grayed). Repeated φ words are in bold. Species notations: Ca—*C. argyrophylla*; Cd—*C. deodara*; Kd—*K. davidiana*; Ps—*P. sitchensis*; Pc—*P. contorta*; Pg—*P. gerardiana*; Pkr—*P. krempfii*; Pk—*P. koraiensis*; Pt—*P. Thunbergii*; Cj—*C. japonica*.

(NC_011163), *Trifolium subterraneum* (NC_011828), *Vigna radiate* (NC_013843), *Pisum sativum* (NC_014057), *Lathyrus sativus* (NC_014063). **Geraniaceae:** *Pelargonium x hortorum* (NC_008454), *Erodium texanum* (NC_014569), *Geranium palmatum* (NC_014573), *Monsonia speciosa* (NC_014582). *Malvaceae*: *Gossypium hirsutum* (NC_007944), *Gossypium barbadense*(NC_008641). *Myrtaceae*: *Eucalyptus globules*(NC_008115), *Eucalyptus grandis* (NC_014570). **Nymphaeaceae**: *Nymphaea alba* (NC_006050), *Nuphar advena* (NC_008788). **Oleaceae:** *Jasminum nudiflorum* (NC_008407), *Olea europaea* (NC_013707). **Onagraceae:** *Oenothera elata* (NCJ)02693), *O. argillicola* (NC_010358), *O. glazioviana* (NC_010360), *O. biennis* (NC_010361), *O. parviflora* (NC_010362). **Pinaceae:** *Pinus thunbergii* (NC_001631), *Pinus koraiensis* (NCJ304677), *Picea sitchensis* (NC_011152), *Pinus contorta* (NC_011153), *Pinus gerardiana* (NC_011154), *Pinus krempfii* (NC_011155), *Keteleeria davidiana* (NC_011930), *Cedrus deodara* (NC_014575), *Cathaya argyrophylla* (NC_014589). *Poaceae* (BEP): *Oryza sativa* Japonica Group (NC_001320), *Triticum aestivum* (NC_002762), *O. nivara* (NC_005973), *O. sativa* Indica Group (NC_008155), *Hordeum vulgare* (NC_008590), *Agrostis stolonifera* (NC_008591), *Lolium perenne* (NC_009950), *Brachypodium distachyon* (NC_011032), *Festuca arundinacea* (NC_011713), *Bambusa oldhamii* (NC_012927), *Dendrocalamus latiflorus* (NC_013088). *Poaceae* (PACCAD): *Zea mays* (NC_001666), *Saccharum hybrid* SP80-3280 (NC_005878), *Saccharum officinarum* (NC_006084), *Sorghum bicolor* (NC_008602), *Coix lacryma-jobi* (NC_013273). **Ranunculaceae:** *Ranunculus macranthus* (NC_008796), *Megaleranthis saniculifolia* (NC_012615). **Saliceae:** *Populus alba* (NC_008235), *Populus trichocarpa* (NC_009143). **Solanaceae:** *Nicotiana tabacum* (NC_001879), *Atropa belladonna* (NC_004561), *Nicotiana sylvestris* (NC_007500), *Nicotiana tomentosiformis* (NC_007602), *Solanum lycopersicum* (NC_007898), *Solanum bulbocastanum* (NC_007943), *Solanum tuberosum* (NC_008096). The plastome of *Cryptomeria japonica* (NC_010548), family *Cupressaceae*, is more divergent and was chosen as the outgroup (Fig. 2).

Instead of the mitochondria of the family of *Brassicaceae*, the mitochondria of order *Brassicales* were analyzed: *Brassica napus* (NC_008285), *Arabidopsis thaliana* (NC_001284), *Carica papaya* (NC_012116); *Caryophyllales*: *Beta vulgaris subsp. vulgaris* (NC_002511), *Beta vulgaris subsp. maritime* (NC_015099). Mitochondria of family *Poaceae* were also analyzed: *Triticum aestivum* (NC_007579), Sorghum bicolor *(NC_008360)*, *Oryza sativa Indica Group* (NC_007886), *Oryza sativa Japonica Group* (NC_011033), *Oryza rufipogon* (NC_01 J816), *Zea mays subsp. mays* (NC_007982), *Zea perennis* (NC_008331), *Zea mays subsp. parviglumis* (NC_008332), *Zea luxurians* (NC_008333), *Tripsacum dactyloides* (NC_008362).

To solve this problem, we used a database. This database contains the results of clustering (subdivision into families based on related sequences) plastid proteins from three groups: rodophytes, chlorophytes, and monocots. A new original clustering algorithm was developed for production of the database.

Inserts of direct repeats were searched as follows: for each pair of species, pairs of positionally linked genes or exons were searched for closely related species, and very short nonconservative genes were ignored. After it aggregated, non-coding regions between these genes or exons of all the species of the same family were aligned and inserts of direct repeats were detected. Programs [12–14] were used.

RESULTS

A database reflecting the clustering of proteins encoded in the plastids of plant groups [15] was produced. A search of clusters by phylogenetic profile of protein is available at http://lab6.iitp.ru/ppc/. Description of the algorithm and manual is also available there. Main algorithm parameters were: Parameter H—upper threshold of similarity of proteins from different clusters (proteins with similarity above H were not separated). Parameter L—lower limit of similarity (proteins with similarity below L are considered as different). Parameter p—upper threshold of cluster size of N species (clusters with number of proteins greater than $p \times N$ were always divided if they have an edge with a weight lower than H). The results for monocotyledonous plants (best results were obtained with parameters: $p = 2$, $L = 0$, $H = 0.41$) are shown. Clusters produced by using these values were corrected manually: in cluster PetG, proteins YP_654227.1 from *Oryza sativa* Indica Group and YP_358627.1 from *Phalaenopsis aphrodite* were added; in cluster RpL23, proteins YP_874745.1 from *Agrostis stolonifera* and YP_899416.1 from *Sorghum bicolour* were added; in cluster RpL2, paralogs YP_654244.1 and YP_654261.1 from *Oryza sativa* Indica Group were added. By this method, 105 nonsingleton clusters and 20 singleton clusters were produced. Out of nonsingleton clusters, 71 contain no more than one protein from the same species, 30 contain pairs of proteins from the same species, two contain three proteins from the same species, and 2 contain 4 proteins from the same species. In 29 (nonsingleton) clusters, from 1 to 12 species were present; there are no clusters which contain 13 to 30 species, 31 to 36 species were present in 76 clusters (range limits all included). Maximum distribution: 30 clusters with 35 species.

The following dependencies between the word φ size and PDRI numbers were found (the notation is "word size: PDRIs number"). Family Acoraceae: 1: 9. Family Asteraceae: 1: 153; 2: 10; 3: 5; 4: 9; 5: 27; 6: 31; 7: 9; 8: 4; 9: 1; 10: 5; 11: 1; 12: 2; 13: 1; 17: 2; 18: 2; 21:

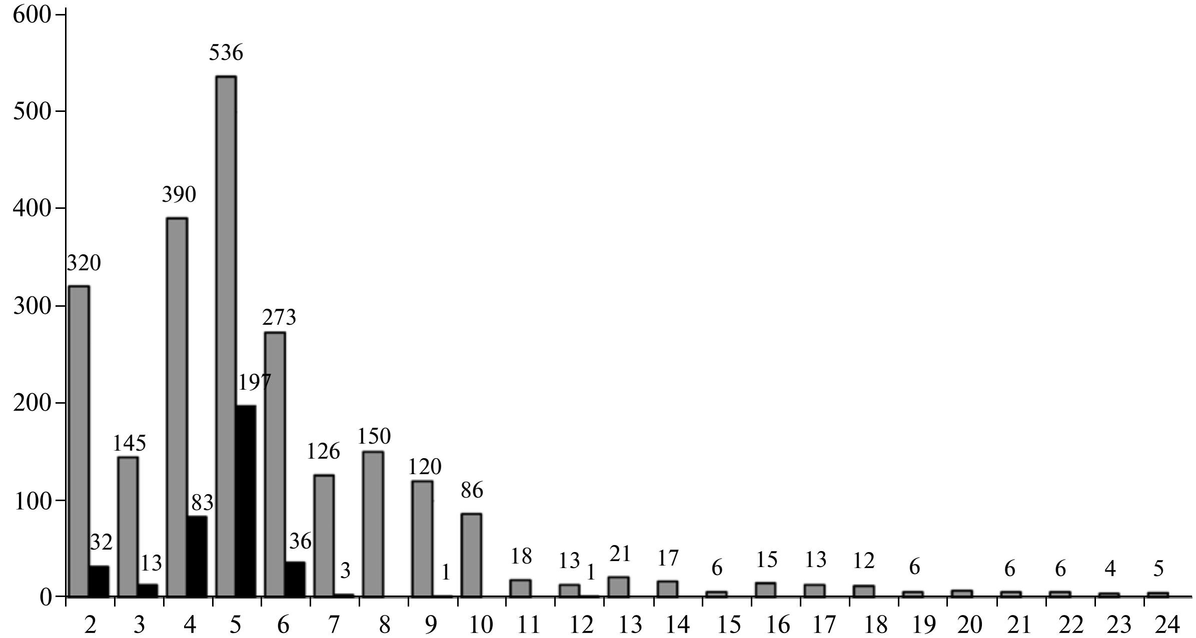

Fig. 3. Occurrences of PDRIs and repeated words (2–24 bp length). PDRIs in plastomes are in black; they are grey in mitochondria.

2; 22: 3; 23: 1. Family Brassicaceae: 1: 1373; 2: 82; 3: 26; 4: 31; 5: 66; 6: 57; 7: 32; 8: 16; 9: 5; 10: 3; 11: 3; 12: 1; 13: 6; 15: 1; 16: 1; 17: 2; 20: 3; 22: 1; 23: 1. Family Fabaceae: 1: 783; 2: 84; 3: 41; 4: 91; 5: 71; 6: 25; 7: 10; 8: 3; 10: 1; 12: 2; 13: 1; 14: 1; 16: 2. Family Geraniaceae: 1: 186; 2: 14; 3: 6; 4: 25; 5: 34; 6: 21; 7: 9; 8: 2; 9: 1. Family Convolvulaceae: 1: 168; 2: 8; 3: 4; 4: 6; 5: 4; 6: 7; 7: 3; 8: 2; 15: 1; 18: 1; 19: 1. Family Malvaceae: 1: 45; 2: 1; 4: 2; 5: 8; 6: 4; 7: 2; 8: 2; 51: 1. Family Myrtaceae: 1: 34; 6: 1; 18: 1; 19: 1. Family Nymphaeaceae: 1: 58; 2: 8; 3: 3; 4: 22; 5: 32; 6: 8; 7: 5; 9: 2; 11: 1; 14: 2; 24: 1. Family Oleaceae: 1: 46; 2: 4; 3: 3; 4: 5; 5: 6; 6: 3; 7: 1; 8: 2; 11: 1; 12: 1; 14: 1; 17: 1. Family Onagraceae: 1: 114; 2: 6; 3: 1; 4: 5; 5: 14; 6: 10; 7: 4; 9: 1; 11: 1; 12: 3; 13: 2; 15: 2; 16: 2; 17: 1; 19: 1; 20: 1; 21: 2; 22: 2; 23: 2; 24: 2; 26: 1; 29: 1; 35: 1; 50: 1; 78: 1. Family Pinaceae: 1: 378; 2: 41; 3: 38; 4: 104; 5: 99; 6: 23; 7: 6; 8: 4; 9: 4; 10: 2; 11: 3; 12: 3; 13: 2; 16: 1. Family Poaceae; the BEP clade: 1: 695; 2: 22; 3: 9; 4: 47; 5: 108; 6: 52; 7: 15; 8: 6; 9: 6; 10: 1; 11: 1; 13: 2; 14: 2; 15: 1; 16: 1; 18: 1; 19: 1; 21: 2; 24: 1; 27: 1. Family Poaceae; the ÐÀÑÑAD clade: 1: 202; 2: 7; 3: 4; 4: 11; 5: 30; 6: 7; 7: 5; 8: 2; 9: 2; 14: 3; 16: 3; 17: 3; 18: 3; 19: 1; 20: 1; 26: 1. Family Ranunculaceae: 1: 111; 2: 13; 3: 3; 4: 12; 5: 22; 6: 7; 7: 5. Family Saliceae: 1: 49; 2: 2; 4: 1; 6: 4; 7: 6; 8: 1; 9: 1; 10: 1; 11: 4; 13: 2; 14: 4; 16: 1; 17: 1; 18: 1; 20: 1. Family Solanaceae: 1: 238; 2: 18; 3: 2; 4: 19; 5: 15; 6: 13; 7: 14; 8: 10; 9: 7; 10: 3; 11: 3; 12: 1; 13: 5; 14: 4; 15: 1; 16: 4; 17: 3; 18: 3; 19: 1; 20: 1; 24: 1; 30: 1.

The longest φ words were found in Onagraceae: 50 bp in between genes *petA* and *psbJ* and 78 bp in between *atpH* and *atpI*. In the Malvaceae, the spacer between *petN* and *psbM* contains a 51 bp φ. Single-base insertions are most common; 4642 were detected. Among insertion of direct repeats, 5 bp are more common (Fig. 3). Insertion of repeats longer than 24 bp are rare: two cases of 26 bp and single cases of 27, 29, 30, 35, 50, 51, and 78 bp. The search was not limited by the length of the repeat.

Consider an exemplary PDRI in the 5’-leader region of gene *psbM* in the Pinaceae (Fig. 2). Thrice-repeated 11-bp words in *K. davidiana* and *P. thunbergii* differ in nucleotide composition and position relative to the start codon *psbM*, which suggests their independent origin. *C. argyrophylla*, *C. deodara*, *P. sitchensis*, *P. gerardiana*, *P. krempfii*, and *P. koraiensis* lack this repeat, while *P. contorta* and *P. thunbergii* have the repeated ATGAGATCAAA motif absent in other species. In *K. davidiana*, the word TTCTATTCATT contains a substitution in its third instance; the repeat is absent from the two pine species and more distanced from the gene start. *P. contorta* and *P. thunbergii* have the AAAG repeat; *C. argyrophylla*, *P. sitchensis* and *Pinus spp* possess the ATTC repeat. The emergence of repeats within the *psbM* 5’-leader region is likely to be evolutionary favored, regardless of the nucleotide composition of the repeat. Thus, in *C. japonica*, a close relative of the Pinaceae, repeats are absent upstream of *psbM*, which suggests their emergence in the Pinaceae rather than a loss from elsewhere. The neighborhood of repeats is quite conserved.

In mitochondria, the following dependencies between the φ word size and PDRI numbers were found. The order Brassicales: 1: 159; 2: 20; 3: 7; 4: 12; 5: 16; 6: 5; family Poaceae: 1: 164; 2: 12; 3: 6; 4: 71; 5: 181; 6: 31; 7: 3; 9: 1; 12: 1. Similarly to plastids, most common are single-base insertions, followed in number by 5.

DISCUSSION

In this study, plastomes and mitochondria of close species were analyzed to further develop and refine inferences on the evolution of non-coding DNA regions. The results suggest that perfect direct repeat insertions are common elementary events in microevolution of short non-coding DNA regions of plastomes and mitochondria. The repeated word length is usually 5. The word length distribution is similar between plastomes and mitochondriomes (Fig. 3). The imperfect repeat in *Keteleeria davidiana* is likely to have resulted from a perfect 11-bases PDRI followed by a single mutation (Fig. 2). Instant emergence of direct repeat insertions is proposed to be a result of replication errors leading to duplications of non-coding DNA regions. The work was presented during conference [16] (by O. Zverkov). Part of the results was presented in [17]. The study was performed with partial support from the Ministry of Education and Science of the Russian Federation (14.740.11.0624, 14.740.11.1053, NK-421P, 14.740.12.0830).

REFERENCES

1. Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R., REPuter: The Manifold Applications of Repeat Analysis on a Genomic Scale, *Nucleic Acids Res.*, 2001, vol. 29, pp. 4633–4642.
2. Yang, M. Xiaowei, ZhangX., et al., The Complete Chloroplast Genome Sequence of Date Palm (*Phoenix dactylifera* L.), *PLoS ONE*, 2010, vol. 5, no. 9, p. E12762.
3. Greiner, S., Wang, X., Rauwolf, U., Silber, M.V., Mayer, K., Meurer, J., Haberer, G., and Herrmann, R.G., The Complete Nucleotide Sequences of the Five Genetically Distinct Plastid Genomes of *Oenothera*, Subsection *Oenothera*: I. Sequence Evaluation and Plastome Evolution, *Nucleic Acids Res.*, 2008, vol. 36, no. 7, pp. 2366–2378.
4. Ogihara, Y., Terachi, T., and Sasakuma, T., Intramolecular Recombination of Chloroplast Genome Mediated by Short Direct-Repeat Sequences in Wheat Species, *Proc. Natl. Acad. Sci. USA*, 1988, vol. 85, no. 22, pp. 8573–8577.
5. Cai, Z., Guisinger, M., Kim, H.G., Ruck, E., Blazier, J.C., McMurtry, V., Kuehl, J.V., Boore, J., and Jansen, R.K., Extensive Reorganization of the Plastid Genome of *Trifolium subterraneum* (Fabaceae) Is Associated with Numerous Repeated Sequences and Novel DNA Insertions, *J. Mol. Evol.*, 2008, vol. 67, no. 6, pp. 696–704.
6. Timme, R.E., Kuehl, J.V., Boore, J.L., and Jansen, R.K., A Comparative Analysis of the *Lactuca* and *Helianthus* (Asteraceae) Plastid Genomes: Identification of Divergent Regions and Categorization of Shared Repeats, *Am. J. Bot.,* 2007, vol. 94, no. 3, pp. 302–312.
7. Moore, M.J., Hassan, N., Gitzendanner, M.A., Bruenn, R.A., Croley, M., Vandeventer, A., Horn, J.W., Dhingra, A., Brockington, S.F., Latvis, M., Ramdial, J., Alexandre, R., Piedrahita, A., Xi, Z., Davis, C.C., Soltis, P.S., and Soltis, D.E., Phylogenetic Analysis of the Plastid Inverted Repeat for 244 Species: Insights into Deeper-Level Angiosperm Relationships from a Long, Slowly Evolving Sequence Region, *Int. J. Plant Sci.,* 2011, vol. 172, no. 4, pp. 541–558.
8. Honma, Y., Yoshida, Y., Terachi, T., Toriyama, K., Mikami, T., and Kubo, T., Polymorphic Minisatellites in the Mitochondrial DNAs of *Oryza* and *Brassica*, *Curr. Genet.*, 2011, vol. 57, no. 4, pp. 261–270.
9. Borsch, T. and Quandt, D., Mutational Dynamics and Phylogenetic Utility of Noncoding Chloroplast DNA, *Plant Syst. Evol.*, 2009, vol. 282, pp. 169–199.
10. Kelchner, S.A., The Evolution of Non-Coding Chloroplast DNA and Its Application in Plant Systematics, *Ann. Missouri Botanical Garden,* 2000, vol. 87, pp. 482–498.
11. Ingvarsson, P.K., Ribstein, S., and Taylor, D.R., Molecular Evolution of Insertions and Deletion in the Chloroplast Genome of Silene, *Mol. Biol. Evol.*, 2003, vol. 20, no. 11, pp. 1737–1740.
12. Official Website of the Laboratory of the Institute for Information Transmission Problems. http://lab6.iitp. ru/ru/repeats/ (25.08.2011).
13. Edgar, R.C., MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput, *Nucleic Acids Res.*, 2004, vol. 32, no. 5, pp. 1792–1797.
14. Official Website of the Laboratory of the Institute for Information Transmission Problems. http://lab6.iitp. ru/en/treeal/ (25.08.2011).
15. Zverkov, O.A., Seliverstov, A.V., and Lyubetskii, V.A., Plastid-Encoded Protein Families Specific for Narrow Taxonomic Groups of Algae and Protozoa, *Mol. Biol.*, 2012, vol. 46, no. 5, pp. 717–726.
16. Seliverstov, A.V. and Lyubetskii, V.A., Direct Repeats in Noncoding Regions of Chloroplasts in Seed Plants, in *Tr. 52-i Nauch. Konf. Moskovskogo Fiziko-Tekhnicheskogo Universiteta* (Proc. 52nd Sci. Conf. Moscow Physico-Technical University), 2009, vol. 1, no. 1, pp. 116–117.
17. Zverkov, O.A., Rusin, L.Yu., Seliverstov, A.V., and Lyubetskii, V.A., Insertions of Direct Repeats in the Microevolution of Plastids and Mitochondria of Seed Plants, *Inform. Prots.*, 2012, vol. 12, no. 3, pp. 191–197.