ABS ACTA BIOLOGICA SIBIRICA
A journal on the biodiversity of Siberia and the adjacent lands

Acta Biologica Sibirica 12: 1623–1632 (2026)
doi: 10.5281/zenodo.24286445
http://journal.asu.ru

RESEARCH ARTICLE

Genome Altai, Plants: the first database on plant genome sizes of Siberian flora

Mikhail V. Skaptsov[1], Alexei V. Vaganov[1], Sergey V. Smirnov[1], Sergey Y. Pankratov[1], Vladimir F. Zaykov[1], Tatiana A. Sinitsyna[1], Daria S. Demidova[1], Iliya A. Shestakov[1], Alexei A. Kechaykin[1], Anastasia M. Koltunova[1], Andrey V. Anisimov[1], Vitor O. Tolmachev[1], Alexander A. Batkin[1], Denis A. Dmitriev[1], Alexander I. Shmakov[1]

1 *Altai State University, 61 Lenin Ave., Barnaul, 656049, Russian Federation*

Corresponding author: Mikhail V. Skaptsov (mr.skaptsov@mail.ru)

Academic editor: R. Yakovlev | Received 14 October 2026 | Accepted 11 November 2026 | Published 18 November 2026

http://zoobank.org/6DE2F222-A715-4D4F-B0E6-C894A767019A

Citation: Skaptsov MV, Vaganov AV, Smirnov SV, Pankratov SY, Zaykov VF, Sinitsyna TA, Demidova DS, Shestakov IA, Kechaykin AA, Koltunova AM, Anisimov AV, Tolmachev VO, Batkin AA, Dmitriev DA, Shmakov AI (2026) Genome Altai, Plants: the first database on plant genome sizes of Siberian flora. Acta Biologica Sibirica 12: 1623–1632. https://doi.org/10.5281/zenodo.24286445

Abstract

Despite the global accumulation of genome size data, Siberia remains one of the most underrepresented regions in plant flow cytometry databases, creating a substantial geographic bias in comparative genomic research. To address this gap, we developed "Genome Altai, Plants" (https://genome.asu.ru/), the first public database of plant genome sizes in Russia, focused on the flora of Southern Siberia and adjacent territories. Genome size was estimated by flow cytometry with propidium iodide staining using internal standardization; herbarium specimens, silica gel-dried tissue, and fresh material from the living collection of the South-Siberian Botanical Garden were analysed. Each record is linked to a digitized voucher specimen in the ALTB Digital Herbarium, and for every measurement the internal standard and its assigned 2C value are specified, allowing recalculation between the classical Doležel and rice-based calibration systems. The database currently contains over 3,100 records covering 106 genera from 41 families. In this study, genome sizes are reported for 133 species from herbarium material and 136 species from the living collection, with 2C values ranging from 0.195 pg (*Selaginella shakotanensis*) to 156.48 pg (*Galanthus lagodechianus*). Because desiccation may distort DNA estimates by 15–20%, values obtained from dried material are flagged as provisional. The database supports ploidy interpretation, cytotype discovery, and biogeographic research, and will be regularly updated with new records and expanded taxonomic coverage.

Copyright Mikhail V. Skaptsov et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Keywords
Altai, C-value, cytotype, flow cytometry, genome size database, herbarium, Southern Siberia

Introduction

The study of nuclear DNA content (genome size, C-value) has become an integral part of modern plant sciences, providing fundamental insights into evolutionary biology, systematics, ecology, and practical applications such as the selection of species for genome sequencing projects, biological resource banks, genetics, and breeding (Loureiro et al. 2007; Siljak-Yakovlev et al. 2010). Several global databases have been established, including the Plant DNA C-values Database at the Royal Botanic Gardens, Kew - the most comprehensive compilation to date, which in its latest release (7.1, April 2019) contains data for 12,273 species. This database originated from a series of compilations of nuclear DNA amounts in angiosperms published by Bennett and colleagues since 1976 (Bennett and Smith 1976; Bennett and Leitch 1995, 2005, 2011). More specialised resources also exist, such as the GSAD database focused on the Asteraceae, which in its third release (2018) included 1,496 species from 4,350 records (Garnatje et al. 2011), and the Balkan flora database, which covers 821 taxa (Siljak-Yakovlev et al. 2010).

Despite the substantial growth of these resources, they remain geographically biased: the vast majority of available genome size estimates come from European and Mediterranean regions of Eurasia, while vast territories such as Siberia remain severely underrepresented. This uneven coverage creates significant knowledge gaps and limits our understanding of genome size variation across the full geographic and ecological range of plant species (Garcia et al. 2014; Siljak-Yakovlev et al. 2010).

The "Genome Altai, Plants" database was developed to fill this gap and is the first public database in Russia dedicated to plant genome sizes. Genome size estimations are performed at the South-Siberian Botanical Garden of Altai State University using flow cytometry (Genome Altai, Plants 2025). Uniquely focused on the flora of Southern Siberia and adjacent regions, the database provides data for a territory that has been largely overlooked in previous genome size compilations.

Materials and methods

Database structure and Darwin Core compliance

The "Genome Altai, Plants" database (https://genome.asu.ru/) is built on a relational structure comprising 28 fields (terms), of which 19 correspond to the Darwin Core specification (https://dwc.tdwg.org/list/). The Darwin Core-compliant fields include, among others, occurrenceID (barcode/inventory number of the ALTB Herbarium voucher), references (link to the digitised herbarium sheet or to a field

observation with photo documentation), basisOfRecord (PreservedSpecimen or LivingSpecimen), scientificName, family, genus, specificEpithet, country, countryCode, decimalLatitude, decimalLongitude, minimumElevationInMeters, eventDate, recordedBy, identifiedBy, verbatimLocality, locationRemarks, catalogNumber, and typeStatus (Fig. 1).

Genome-size-specific fields include DNAamount2C (2C DNA content, pg), genomesize1C (1C genome size, Mbp), standartOrganism (internal standard used, with reference organism indicated), standartLinkReferences (link to the publication or resource describing the standard and/or methodology), associatedReferences and associatedLinkReferences (bibliographic citation and link to the source of the genome size value), as well as chromosomenumber2N, chromosomenumber2NReferences, and chromosomenumber2NLink (chromosome number and its source).

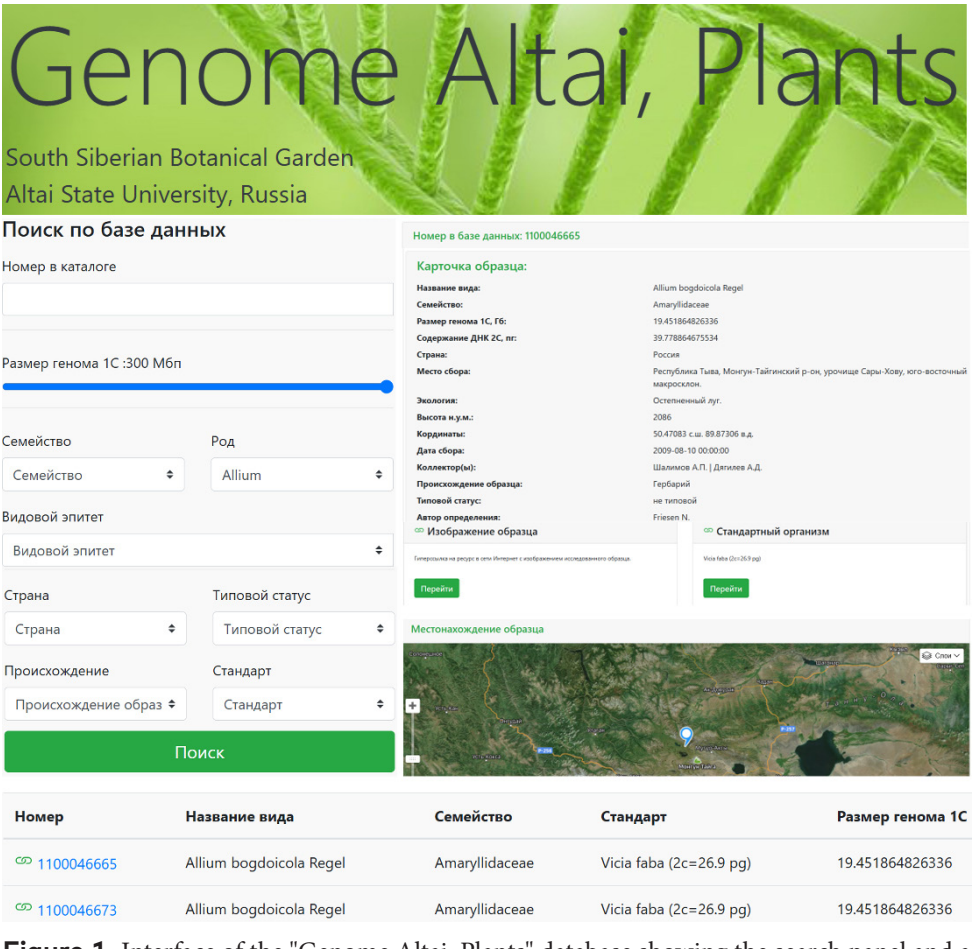


Figure 1. Interface of the "Genome Altai, Plants" database showing the search panel and a specimen record that illustrates the linkage of each measurement to a herbarium voucher and the specification of the internal standard used.

The use of Darwin Core terms ensures integration of the database with the Global Biodiversity Information Facility (GBIF, http://www.gbif.org). Altai State University is a data publisher in GBIF (IPT, http://altb.asu.ru/ipt/). The database is hosted on the national server of Altai State University and is accessible at genome.asu.ru.

Technical implementation

The database was developed using Apache web server, MySQL, PHP, Smarty 3.1, JavaScript, HTML, and CSS. Access to server files is organised via SFTP. Query formation and table storage are managed through MySQL (localhost) and PHPMyAdmin (https://hosting1.asu.ru/phpmyadmin). Geospatial data and cartographic visualisation are implemented through integration with the Yandex Maps module (https://yandex.ru/dev/maps/).

Digitisation and taxonomic verification

Herbarium specimens were digitised using an A3 scanner "Microtek Objectscan 1600". Taxonomic identification and cabinet work with herbarium and expedition material were carried out using stereoscopic microscopes "OLYMPUS" SZ61 and AO LOMO "MSP2v.2sd". Nomenclature and authorship of botanical names were verified against the International Plant Name Index (IPNI).

Plant material

For the present study, genome size measurements were performed on two categories of plant material:

1. Herbarium material (Suppl. material 1: Table S1). Dried leaf tissue from voucher specimens stored in the ALTB Herbarium was used. This approach allows retrospective analysis of archived collections and provides genome size data for species that are difficult to collect fresh. However, as demonstrated previously, dried material can yield overestimated (up to 15–20%) or, after long-term storage, underestimated values compared to fresh tissue (Bainard et al. 2011; Skaptsov et al. 2026). Therefore, values obtained from herbarium material are regarded as provisional genome size estimates and are flagged accordingly in the database.
2. Living collection of the South-Siberian Botanical Garden (Suppl. material 2: Table S2). Fresh leaf tissue from plants cultivated in the introduction nursery was used. This material provides more accurate absolute genome size values and is planned for inclusion in the database as it becomes available.

Flow cytometry

DNA content was determined by flow cytometry with propidium iodide (PI) staining. In brief, fresh or dried leaf tissue was co-chopped with an appropriate internal standard in LB01 (Doležel et al. 1992) supplemented with PI (50 µg/ml), RNase (10 µg/ml) and 12 mM sodium thiosulfate (Skaptsov et al. 2024). The following internal standards were used: *Vicia faba* 'Inovec' (V.f., 2C = 26.9 pg), *Allium fistulosum* (A.f., 2C = 25.0 pg), *Chlorophytum comosum* (C.c., 2C = 24.14 pg), *Secale cereale* 'Tatyana' (S.c., 2C = 16.19 pg), *Hordeum vulgare* 'Ditta' (H.v., 2C = 10.43 pg), *Pisum sativum* 'Citrad' (P.s., 2C = 9.09 pg), *Petroselinum crispum* (P.c., 2C = 4.5 pg), *Solanum pseudocapsicum* (S.p., 2C = 2.835 pg), *Glycine max* 'Polanka' (G.m., 2C = 2.5 pg) and *Ficus benjamina* (F.b., 2C = 1.07 pg) (Doležel et al. 1992; Doležel et al. 1994; Doležel et al. 1998; Obermayer et al. 2002; Hornych et al. 2019; Skaptsov et al. 2024). Analyses were performed on Cytoflex cytometer (Beckman Coulter, USA). Peaks with at least 1,000 nuclei and a coefficient of variation (CV) below 3% (for fresh material) or below 5% (for dried material) were used for analysis. For complex samples with high background noise, histograms were additionally analysed using GSAnalyser (Skaptsov 2026).

Data presentation

Genome sizes in the database are presented on the basis of the classical Doležel values (Doležel et al. 1992, 1998) or new standards calculated on its basis. The Doležel set was calibrated against the human genome (2C = 7.0 pg), which is now known to be an overestimate (modern data: 6.15–6.41 pg). An alternative calibration system based on the fully sequenced rice genome (*Oryza sativa* ssp. *japonica* 'Nipponbare', 2C = 0.795 pg) is considered more accurate (Veselý et al. 2012). Both systems of standards are acceptable for use in the database, and users may choose the one that best suits their comparative purposes (Temsch et al. 2022). We retain the classical Doležel values for compatibility with the vast majority of published data, while the recording of the standard used and its assigned value allows all presented data to be easily recalculated to modern standards if needed.

Results and discussion

Taxonomic and geographic coverage

The database currently covering 106 genera from 41 families, primarily collected from the Altai region and adjacent areas of Southern Siberia. This represents a substantial contribution to the knowledge of plant genome sizes in a territory that has been largely overlooked in previous global compilations. The geographic coverage

extends from the Altai Mountains and the Salair Ridge to the Kuznetsk Alatau, the Tuva Republic, and adjacent regions of Kazakhstan and Mongolia.

Herbarium material

Supplementary material 1 presents estimated DNA content data for 133 species obtained from herbarium material. These species belong to 41 families and 106 genera, including economically valuable, rare, and endemic plants of the Siberian flora. Examples include *Achillea asiatica* (2C = 5.31 pg), *Allium amphibolum* (2C = 33.28 pg), *Anemone altaica* (2C = 39.83 pg), *Erythronium sibiricum* (2C = 51.10 pg), *Lilium pilosiusculum* (2C = 82.62 pg), and *Paeonia anomala* (2C = 27.08 pg). The range of 2C values among herbariumderived records spans from 0.473 pg (*Turritis glabra*) to 82.62 pg (*Lilium pilosiusculum*). As noted above, these values should be regarded as provisional estimates due to the potential effects of desiccation on DNA staining.

Living collection

Supplementary material 2 presents absolute DNA content data for 136 species from the living collection of the South-Siberian Botanical Garden that are planned for inclusion in the database. These species represent a wide taxonomic and geographic range, including plants from the Altai Republic, Primorsky Krai, the Jewish Autonomous Oblast, Kazakhstan, Kyrgyzstan, the Caucasus, and cultivated varieties. The 2C values in this dataset range from 0.195 pg (*Selaginella shakotanensis*) to 156.48 pg (*Galanthus lagodechianus*). This dataset is particularly valuable because it provides genome size data for species that are difficult to collect in the wild, including rare and endangered plants, and because it allows direct comparison between wild and cultivated accessions.

Importance of the database

It is important to emphasise that genome size data are essential not only for comparative genomics but also for the correct interpretation of ploidy levels. When combined with chromosome number databases – such as the general Index to Plant Chromosome Numbers or the familyspecific Index to Chromosome Numbers in Asteraceae (Watanabe 2004) – genome size estimates allow researchers to infer ploidy even in the absence of direct chromosome counts, and to distinguish between true polyploidy and genome size variation at the same ploidy level. By comparing our regional data with those from global resources, it becomes possible to identify previously unknown cytotypes, track their geographic distribution, and conduct biogeographic and phylogeographic studies that link genome size variation to environmental gradients, glacial refugia, and historical dispersal routes. The "Genome Altai, Plants" database thus serves as a crucial reference for uncovering cytotype diversity in a previously unstudied part of Eurasia.

Material type and data reliability

In addition to the type of plant material (fresh, silica gel-dried, herbarium, seeds), the database records the internal standard used and its exact 2C value for every measurement. This information is critical because, as demonstrated in numerous studies (Bainard et al. 2011; Wang and Yang 2016; Skaptsov et al. 2024), dried material often yields unpredictable genome size estimates compared to fresh samples – in most cases overestimating by up to 15–20%, while long-term herbarium storage may conversely lead to underestimation. Therefore, values obtained from non-fresh material should be regarded as provisional DNA content estimates or provisional genome size estimates. Users are encouraged to consult the original metadata to assess the reliability of each record for their specific comparative purposes.

Conclusion

The "Genome Altai, Plants" database represents a significant step towards a comprehensive understanding of plant genome size variation in Russia. By combining regional floristic coverage, voucher-linked records, detailed standardisation metadata, and clear indication of material type, it complements existing global databases and provides a reliable tool for comparative research, biodiversity conservation, and the development of genetic passports for plant species. Our future plans include regular updates with new data, expansion of taxonomic coverage, and targeted sampling of additional species across Southern Siberia and neighbouring regions, with the aim of further reducing the geographic bias in plant genome size knowledge and fostering collaborative studies on the evolution and ecology of plant genomes.

Acknowledgements

This work was supported by the State Assignment of Altai State University (project FZMW20260006).

References

Bainard JD, Husband BC, Baldwin SJ, Fazekas AJ, Gregory TR, Newmaster SG, Kron P (2011) The effects of rapid desiccation on estimates of plant genome size. Chromosome Research 19(6): 825–842. https://doi.org/10.1007/s10577-011-9232-5

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philosophical Transactions of the Royal Society of London B 274(933): 227–274. https://doi.org/10.1098/rstb.1976.0044

Bennett MD, Leitch IJ (1995) Nuclear DNA amounts in angiosperms. Annals of Botany 76(2): 113–176. https://doi.org/10.1006/anbo.1995.1085

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. Annals of Botany 95(1): 45–90. https://doi.org/10.1093/aob/mci003

Bennett MD, Leitch IJ (2011) Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. Annals of Botany 107(3): 467–590. https://doi.org/10.1093/aob/mcq258

Doležel J, Sgorbati S, Lucretti S (1992) Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. Physiologia Plantarum 85: 625–631. https://doi.org/10.1111/j.1399-3054.1992.tb04764.x

Doležel J, Doleželová M, Novák F (1994) Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). Biologia Plantarum 36: 351–357. https://doi.org/10.1007/BF02920930

Doležel J, Greilhuber J, Lucretti S, Meister A, Lysák M, Nardi L, Obermayer R (1998) Plant genome size estimation by flow cytometry: inter-laboratory comparison. Annals of Botany 82: 17–26. https://doi.org/10.1093/oxfordjournals.aob.a010312

Garcia S, Leitch IJ, Anadon-Rosell A, Canela MÁ, Gálvez F, Garnatje T, Gras A, Hidalgo O, Johnston E, Mas de Xaxars G, Pellicer J, Siljak-Yakovlev S, Vallès J, Vitales D, Bennett MD (2014) Recent updates and developments to plant genome size databases. Nucleic Acids Research 42(D1): D1159–D1166. https://doi.org/10.1093/nar/gkt1195

Garnatje T, Canela MÁ, Garcia S, Hidalgo O, Pellicer J, Sánchez-Jiménez I, Siljak-Yakovlev S, Vitales D, Vallès J (2011) GSAD: A Genome Size in the Asteraceae Database. Cytometry A 79(6): 401–404. https://doi.org/10.1002/cyto.a.21060

Genome Altai, Plants (2025) Available from: https://genome.asu.ru/

Hornych O, Ekrt L, Riedel F, Koutecký P, Košnar J (2019) Asymmetric hybridization in central European populations of the *Dryopteris carthusiana* group. American Journal of Botany 106: 1477–1486. https://doi.org/10.1002/ajb2.1369

Leitch IJ, Johnston E, Pellicer J, Hidalgo O, Bennett MD (2019) Plant DNA C-values Database (release 7.1, April 2019). Available from: https://cvalues.science.kew.org/

Loureiro J, Suda J, Doležel J, Santos C (2007) FLOWER: A Plant DNA Flow Cytometry Database. In: Doležel J, Greilhuber J, Suda J (Eds) Flow Cytometry with Plant Cells. Wiley-VCH, Weinheim, 423–438. https://doi.org/10.1002/9783527610921.ch18

Siljak-Yakovlev S, Pustahija F, Šolić EM, Bogunić F, Muratović E, Bašić N, Catrice O, Brown SC (2010) Towards a genome size and chromosome number database of Balkan flora: C-values in 343 taxa with novel values for 242. Advanced Science Letters 3: 190–213. https://doi.org/10.1166/asl.2010.1115

Skaptsov M (2026) GSAnalyser: an open-source tool for objective DNA content estimation and quality control in plant flow cytometry (Version v1.1.0). Zenodo. https://doi.org/10.5281/zenodo.23163447

Skaptsov M, Degtyareva G, Smirnov S, Kutsev M, Shmakov A (2026) Plant flow cytometry: current problems, aspects of application and data analysis. Ecological Genetics 24(3). https://doi.org/10.17816/ecogen707965

Skaptsov MV, Kutsev MG, Smirnov SV, Vaganov AV, Uvarova OV, Shmakov AI (2024) Standards in plant flow cytometry: an overview, polymorphism and linearity issues. Turczaninowia 27(2): 86–104. https://doi.org/10.14258/turczaninowia.27.2.6

Temsch EM, Koutecký P, Urfus T, Šmarda P, Doležel J (2022) Reference standards for flow cytometric estimation of absolute nuclear DNA content in plants. Cytometry Part A 101(9): 710–724. https://doi.org/10.1002/cyto.a.24495

Veselý P, Bureš P, Šmarda P, Pavlíček T (2012) Genome size and DNA base composition of geophytes: the mirror of phenology and ecology? Annals of Botany 109(1): 65–75. https://doi.org/10.1093/aob/mcr267

Wang G, Yang Y (2016) The effects of fresh and rapid desiccated tissue on estimates of Ophiopogoneae genome size. Plant Diversity 38(4): 190–193. https://doi.org/10.1016/j.pld.2016.08.001

Watanabe K (2004) Index to chromosome numbers in the Asteraceae on the web. Compositae Newsletter 41: 64.

Supplementary material 1

Table S1. Genome size estimates for species of the Siberian flora obtained from herbarium material (ALTB Digital Herbarium)

Authors: Mikhail V. Skaptsov, Alexei V. Vaganov, Sergey V. Smirnov, Sergey Y. Pankratov, Vladimir F. Zaykov, Tatiana A. Sinitsyna, Daria S. Demidova, Iliya A. Shestakov, Alexei A. Kechaykin, Anastasia M. Koltunova, Andrey V. Anisimov, Vitor O. Tolmachev, Alexander A. Batkin, Denis A. Dmitriev, Alexander I. Shmakov

Data type: table

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://journal.asu.ru/biol/article/view/20673/17598

Supplementary material 2

Table S2. Genome size estimates for species from the living collection of the South-Siberian Botanical Garden (Altai State University), planned for inclusion in the Genome Altai, Plants database

Authors: Mikhail V. Skaptsov, Alexei V. Vaganov, Sergey V. Smirnov, Sergey Y. Pankratov, Vladimir F. Zaykov, Tatiana A. Sinitsyna, Daria S. Demidova, Iliya A. Shesta-

kov, Alexei A. Kechaykin, Anastasia M. Koltunova, Andrey V. Anisimov, Vitor O. Tolmachev, Alexander A. Batkin, Denis A. Dmitriev, Alexander I. Shmakov
Data type: table
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://journal.asu.ru/biol/article/view/20673/17599