Phylogeographical patterns based on trnH-psbA plastid DNA shed light on evolution within Waldsteinia (Rosaceae)

M V Protopopova1, V V Pavlichenko1 and N V Stepanov2

1 Siberian Institute of Plant Physiology and Biochemistry SB RAS, Irkutsk, 664033 Russia
2 Siberian Federal University, Krasnoyarsk, 660041 Russia

E-mail: marina.v.protopopova@gmail.com

Abstract. Waldsteinia Willd. is a small herbaceous genus presumably of Neogene age and formerly wide-distributed around the Northern Hemisphere and now presents the remnants of the tertiary flora. According to the latest taxonomic revision, Waldsteinia is considered a group nested in Geum L. A comparatively low level of morphological divergence together with fuzzy ploidy patterns within Waldsteinia could not be reliable evidence to establish the former distributions and migration pathways of species. In the present study, using plastid DNA (trnH-psbA) data we tried to throw light on Waldsteinia history. Based on our data we believe that the taxonomic decision of nesting Waldsteinia in Geum does not reflect the complex structure of the obtained clade. The phylogenetic analysis showed that the objectivity of previous division Waldsteinia on two subgenera, Waldsteinia and Comaropsis (Rich. ex Nestl.) Teppner, which was based on morphological differences, is becoming controversial. We also suggest an East Asian origin of Waldsteinia and subsequent speciation and taxa distribution in the direction of Europe and North America. W. ternate s.l. (traditionally including W. maximowicziana, W. ternata, and W. trifolia) is appeared to be a polyphyletic group and at least W. maximowicziana should be considered a distinct species.

1. Introduction
Waldsteinia Willd. is a small herbaceous genus, which belongs to the family Rosaceae and consists of only a few species. The present geographical pattern of Waldsteinia exhibits a wide but highly disjunctive distribution and includes the following clearly detached fragments: Central and Southeast Europe (W. geoides Willd. and W. trifolia Rochel ex W.D.J.Koch), South Siberia (W. tanzybeica Stepanov and W. ternata (Stephan) Fritsch), Northeast Asia (W. maximowicziana (Teppner) Prob.), and eastern (W. doniana Tratt., W. fragarioides (Michx.) Tratt., and W. lobata (Baldwin) Torr. & A. Gray) and western (W. idahoensis Piper) parts of North America. Waldsteinia is presumably of Neogene age [1, 2] and formerly wide-distributed around the Northern Hemisphere and now presents the remnants of the tertiary flora [3-5], and at least a few species are considered as ‘true’ nemoral relics with narrow and fragmented ranges [1, 4-6]. This pattern of Waldsteinia disjunction correlates well with the long-known eastern North American – Eastern Asian floristic relationship involving migration and interchange of species of Asia and America across the region of the Bering Strait followed by disruption of the continuous ranges by the Pleistocene glaciations [4]. However, a comparatively low level of morphological divergence together with fuzzy ploidy patterns within Waldsteinia could not be reliable evidence to establish the former distributions and species migration pathways during the late Cainozoic. In the present study, we made an attempt to throw light on Waldsteinia history using DNA data.
According to the latest taxonomic revisions [7, 8], *Waldsteinia* together with the closely related genera *Coluria* R.Br. and *Taihangia* T.T.Yu & C.L.Li were nested in *Geum* L. Moreover, according to the new conception, *W. trifolia*, *W. ternata*, and *W. maximowicziana* are combined by the name *Geum ternatum* (Stephan) Smedmark (table 1). Fitting to the classical view, those species were considered as independent species or as intraspecific taxa of *W. ternata* [5, 9]. Although we do not entirely agree with the opinion on the inclusion of *Waldsteinia* in *Geum*, our study is not aimed at reviving *Waldsteinia* as a genus. Nevertheless, we will use the former names of *Waldsteinia* species in the text, so that we can distinguish the aforementioned intrageneric taxa which looks impossible following a new conception.

2. Materials and methods

2.1. Sampling
The present study considered seven *Waldsteinia* taxa including all Eurasian species. *W. ternata* was collected from a South Siberian population on the Khamar-Daban Ridge (floodplains of the Bezymynnaya and Snezhnaya Rivers, the Lake Baikal basin) and Western Sayan Mountains (floodplain of the Kaldar River). *W. tanzybeica* was sampled from Western Sayan Mountains (the Bolshoy Kebezh River floodplain). *W. maximowicziana* was collected in the Russian Far East (from two populations of Nadezhdinsky and one of Partizansky areas, Primorsky Krai). Samples of European *W. geoides* and *W. trifolia* were collected from living collections in the Botanic garden of Irkutsk State University (Irkutsk, Russia) and the Central Botanical Garden of the National Academy of Sciences of Belarus (Minsk, Republic of Belarus), respectively.

The samples of North American *Waldsteinia* species were collected from the herbarium of the Komarov Botanical Institute RAS (LE; Saint-Petersburg, Russia), i.e. *W. fragarioides* (near Douro-Dummer Township, Peterborough Country, Ontario, Canada; the Green River, Greenfield, Massachusetts, USA; near Hamburg, New Jersey, USA; and Oatka Creek Park near Rochester, New York, USA) and *W. doniana* (the Sipsey River near Addison, Winston County, Alabama, USA). *Geum rivale* as an additional reference species was collected from the Khamar-Daban Ridge. At least six separate individuals were collected from each natural population and one specimen of each species was taken from the living collections and herbaria. Each sample was kept in an individual filter paper bag prior to DNA isolation. Fresh plant material was dried and stored in silica gel.

2.2. DNA isolation, PCR and sequencing
Total DNA was isolated from silica-dried leaf tissue following the cetyltrimethylammonium bromide (CTAB) method [10], with modifications [11] and RNase A treatment for RNA removal.

For our study the region containing trnH-psbA intergenic spacer and short fragments of flanking trnH and psbA genes of plastid DNA as a molecular marker was used. The DNA region was amplified using primers which successfully applied for *W. fragarioides* [12]. PCR was performed in a reaction mixture of 20 µL contained 1x Green GoTaq Flexi Buffer, 1 unit of GoTaq Flexi DNA polymerase (Promega) and final concentrations of 2.5 mM of MgCl₂, 250 µM of each dNTP, and 250 nM of each primer. PCR products were visualized in 1% agarose gel stained by ethidium bromide after electrophoresis, gel-purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) and directly sequenced by Sanger method using BigDye Terminator Cycle Sequencing kit v. 3.1 (Applied Biosystems) on 3500 Genetic Analyzer (Applied Biosystems).

2.3. Sequence alignment and phylogenetic analysis
The raw sequencing data were primary edited using SnapGene Viewer software v.2.6.2 (GSL Biotech). The primer regions were completely removed. GenBank accession numbers of new and reference sequences are presented in table 1. Several *Geum* species in the phylogenetic reconstruction were also included. As an out-group *Fallugia paradoxa* was chosen based on the tree of J E E Smedmark [7].
Table 1. The taxa and DNA sequences used for the phylogenetic reconstructions.

| Former species name                             | Accepted name [13] | GenBank accession number  |
|--------------------------------------------------|--------------------|--------------------------|
| *Fallugia paradoxa* (D.Don) Endl. ex Torr.       | *F. paradoxa* (D.Don) Endl. ex Torr. | KY419999.1               |
| *Geum aleppicum* Jacq.                          | *G. aleppicum* Jacq. | HQ596717.1               |
| *Geum canadense* Jacq.                          | *G. canadense* Jacq. | KP402565.1               |
| *Geum macrophyllum* Willd.                      | *Geum macrophyllum* Willd. | KY419936.1               |
| *Geum rivale* L.                                | *G. rivale* L.      | MZ346036                 |
| *Geum triflorum* Torr.                          | *Geum triflorum* Torr. | KY419977.1               |
| *Geum urbanum* L.                               | *G. urbanum* L.     | FJ395499.1               |
| *Taihangia rupestris* T.T.Yu & C.L.Li           | *G. rupestris* (T.T.Yu & C.L.Li) Smedmark | h1 KC573071.1           |
|                                                  |                    | h2 NC_037392.1           |
| *Waldsteinia doniana* Tratt.                    | *G. donianum* (Tratt.) Weakley & Gandhi | MZ346037                |
| *W. fragarioides* (Michx.) Tratt.               | *G. fragarioides* (Michx.) Smedmark | MZ346038                |
| *W. geoides* Willd.                             | *Geum waldsteiniae* Smedmark | MZ346039                |
| *W. maximovicziana* (Teppner) Prob.             | *G. ternatum* (Stephan) Smedmark | h1 MZ346040              |
|                                                  |                    | h2 MZ346041              |
| *W. tanzypeica* Stepanov                        | *G. tanzybeicum* (Stepanov) Smedmark | MZ346042                |
| *W. ternata* (Stephan) Fritsch                  | *G. ternatum* (Stephan) Smedmark | h1 MZ346043              |
|                                                  |                    | h2 MZ346044              |
| *W. trifolia* Rochel ex *W.D.J.Koch*             | *G. ternatum* (Stephan) Smedmark | MZ346045                |

*a The GenBank accession numbers of the original sequences obtained in the present study are highlighted in bold typeface.*

The multiple alignment of nucleotide sequences by the MUSCLE application with gap opening penalty equalled 500 and extension penalty equalled 4.01 was conducted in MEGA software v. 7.0.16 [14] followed by manual editing. The first 15 positions of the alignment, corresponding to *trnH* gene, were completely removed from the analysis since some of the reference sequences had missing data. Phylogenetic reconstructions were performed by both the maximum likelihood method (ML) in MEGA and the Bayesian inference method implemented in MrBayes v. 3.2.5 [15]. The best-fit model of nucleotide substitutions by the lowest Bayesian Information Criterion (BIC) was selected using the ‘Find Best DNA Models’ tool in MEGA. For ML analysis the indels were completely excluded from the alignment, which final length made 171 positions. For ML-phylogeny the Tamura 3-parameter model [16] with no among-site rate variation was used. A bootstrap of 1,000 replicates was used as a test of the phylogeny. For Bayesian inference analysis insertion/deletion regions in alignments were considered as separated evolutionary events, most indels were coded as binary characters (presence ‘1’ or absence ‘0’ of the gap) and included as a separate binary data partition to the end of the matrix. In general, 21 indels from the total length of 548 alignment positions (28-42, 43-44, 45-52, 90-103, 114-167, 172-269, 270-279, 280-288, 289, 290-295, 299, 304-306, 307, 324-329, 330-335, 361-398, 371-380, 409-414, 420-430, 431-439, 467-489) were coded as binary data. Analysis for combining data for DNA (excluding gaps) and binary (gaps) sets using HKY [17] like model (*iset nst = 2*) with identical parameters of the model used in ML analysis and F81 like model [18] with the equal stationary state frequencies, respectively, was performed in MrBayes. For the HKY like model, the following parameters were used: the *nucleotide frequencies* (in order A, C, G, and T) were fixed at 0.33, 0.17, 0.17, and 0.33; the parameter setting the prior for the transition/transversion rate ratio (*prset tratiopr*) was fixed at 0.63. Markov chain Monte Carlo (MCMC) analyses were run for 1,000,000 generations, with four simultaneous chains with sampling every 100 generations and diagnostic calculation every 1,000 generations. Finally, a phylogram was constructed from the posterior distribution of trees and edited in FigTree v.1.4.3 [19].
The haplotype network was performed with Haplotype Viewer software [20]. The analysis was based on the alignment previously used for the phylogenetic Bayesian tree (see above). The alignment included the DNA sequences of all studied *Waldsteinia* species with *Fallugia paradoxa* as an outgroup. Sequences of other taxa used in the phylogenetic trees were excluded from the analysis of haplotypes. The tree was built using the Bayesian inference method with the HKY model for DNA sequence set and the F81 like models for the binary set in MrBayes with the adjustments described above.

3. Results and discussion

The phylogenetic analysis showed that all studied *Waldsteinia* species were combined in a well-supported clade (figure 1, node A). Most of the reference species formed the clade corresponding to *Geum* in sensu stricto sense (node B). Two *trnH-psbA* haplotypes of *G. rupestris* (h. 1 and h. 2) combined together and representing the formerly monotypic genus *Taihangia* (node C).

**Figure 1.** Phylogram based on Bayesian analysis of the *trnH-psbA* region in *Waldsteinia* and closely related species.

The posterior probabilities are marked below branches, bootstrap values of respective nodes obtained by ML analysis are above branches. The colours of the terminal branches conventionally mean the geographical origin of haplotypes (the designations are given directly in the figure: C – central; E – the East, eastern; N – the North, northern; S – the South, southern). The nodes discussed in the text are indicated by capital letters. The scale indicates the branch length equivalent to the distance of 0.003.
Thus, despite the fact that phylogenetic reconstruction proposed by J.E.E. Smedmark [7] looks justified, in our view, the taxonomic decision of combining monophyletic *Waldsteinia* and *Taihangia* in one genus does not reflect the complex structure of the obtained clade of *Geum sensu lato*, as well as the evolutionary perspectives of internal groups. The in-group *Waldsteinia* structure presented by a well-supported subclade combining all Siberian and European species (node D) and unresolved haplotypes belonging to Northeast Asian and North American taxa. American *W. fragarioides* and *W. doniana* have overlapping ranges, and *W. doniana* can be treated as subspecies of former one (≡ *W. fragarioides* subsp. *doniana* (Tratt.) Teppner) [9]. Our results showed that *W. fragarioides* and *W. doniana* had the identical haplotype of *trnH-psbA* (figure 2), and we could not distinguish these species based on plastid DNA either.

![Figure 2. Haplotype network based on Bayesian analysis of the *trnH-psbA* region in *Waldsteinia*.](image)

Fp – *Fallugia paradoxa*; Wd – *W. doniana*; Wf – *W. fragarioides*; Wg – *W. geoides*; W1 – *W. trifolia*; Wm – *W. maximowicziana*; Wt – *W. ternata*; Wz – *W. tanzybeica*.

The different haplotypes are presented as circles connected by lines, where each section between adjacent nodes on the lines corresponds to one evolutionary event (substitution or indel). The different colours indicate the geographical origin of the haplotypes (the colour designations are as in figure 1).

Earlier, using morphological differences, Teppner [5] divided *Waldsteinia* into two subgenera: *Waldsteinia* including only the type species *W. geoides*, and *Comaropsis* (Rich. ex Nestl.) Teppner comprising all other *Waldsteinia* species. According to our reconstruction, the type species nested in a clade together with European and Siberian taxa of *Comaropsis*. Thus, *Waldsteinia* species are better assembled by geographical patterns than by morphology, thus the monophyly of subgenus *Comaropsis* and the objectivity of dividing *Waldsteinia* into these two subgenera are becoming controversial. An interesting finding was that *W. maximowicziana* did not associate with *W. ternata* haplotypes although it was often considered a subspecies of the latter one (≡ *W. ternata* subsp. *maximowicziana* Teppner). The haplotype of the European *W. trifolia*, which is also considered as a subspecies of *W. ternata* (= *W. ternata* subsp. *trifolia* (Rochel ex W.D.J.Koch) Teppner), according our data, was not combined with the haplotypes of *W. ternata*, however was identical with that found in *W. geoides* (figure 2). Moreover, the found haplotypes of *W. ternata* from the Western Sayan Mountings (h. 1) and the Khamar-Daban (h. 2) formed the well-supported clade neither with each other nor with close related *W. tanzybeica* (posterior probability, 0.85). Thus, *W. ternata sensu lato* (traditionally including *W. maximowicziana*, *W. ternata*, and *W. trifolia*) is appeared to be a polyphyletic group and at least *W. maximowicziana* should be considered a distinct species as earlier was suggested by N.S. Probatova [21]. The haplotype of *W. tanzybeica* looked to be attached to *W. ternata* ones, however, for any taxonomic conclusions, the relationships between Siberian and European haplotypes must be first resolved.
The genetic distance value (figure 1) showed that Northeast Asian and North American haplotypes were closer to node A (i.e. to the most recent common ancestor of all *Waldsteinia* species) than European and Siberian ones. The haplotype network (figure 2) demonstrated that *W. maximowicziana* carried the maternal ancestral haplotype (h. 1) for the entire *Waldsteinia* taxa. That haplotype we found in the populations from both Nadezhdinsky and Partizansky areas of Primorsky Krai (the Russian Far East). In consideration of the modern distribution range of *W. maximowicziana*, we suggest an East Asian origin of the genus *Waldsteinia* and its subsequent speciation and distribution in the direction of Europe and North America. The significant climate cooling approximately 15 Myr ago and disappearance of the Bering Land Bridge approximately 5.4–5.5 Myr ago [22] stimulated the divergence of Asian and American *Waldsteinia* populations and their distribution through the continents. European haplotype appeared to be more ancient than Siberian ones (figure 2) however there was no other way of expanding ancestral *Waldsteinia* from East Asia to Europe than through the continental Siberia. Climate change and the increase in the continentality were the reasons for the fragmentation of the trans-Palearctic broad-leaved zone [23] led to a disjunction of *Waldsteinia* into European, and South Siberian, and East Asian fragments. Analysis of bottom sediments of Lake Baikal evident the domination of the coniferous-broad-leaved complex on the territory inhabiting the present areal of South Siberian species until the middle of the Late Pliocene [2]. Since then, the distant populations developed independently, and their genotypes accumulated nucleotide substitutions and DNA rearrangements at different rates. The youngest *trnH-psbA* haplotypes were found in the South Siberian populations, which may indicate the highest value of their genome plasticity among the other *Waldsteinia* species and populations studied. Thus, modern haplotypes of European and South Siberian populations should not be considered as derived from each other. Most likely, they descended from a common ancestor that had a trans-Palearctic distribution range before the Pleistocene.

It should be noted, that all our conclusions were based on plastid DNA polymorphism, which indicated only the matrilineal haplotype inheritance. Thus, for a more complete understanding of *Waldsteinia* evolution, a comprehensive study with additional genetic markers and an expanded sample of the populations is needed.

4. Conclusion

Based on *trnH-psbA* plastid DNA analysis we believe that the taxonomic latter decision of combining *Waldsteinia* and *Taihangia* in one genus with *Geum* species does not reflect the complex structure of the obtained clade. *Waldsteinia* species are better assembled by geographical patterns than by morphology, thus, the monophyly of both subgenera *Comaropsis* and *Waldsteinia* are becoming controversial. According to the haplotype network and genetic distance value, we suggest an East Asian origin of the genus *Waldsteinia* and its subsequent speciation and distribution in the directions of Europe and North America. *W. ternata* s.l. (traditionally including *W. maximowicziana*, *W. ternata*, and *W. trifolia*) appeared to be a polyphyletic group and at least *W. maximowicziana* should be considered a distinct species.

Acknowledgements

The research was done using the equipment of the Core Facilities Centre ‘Bioanalitika’ and collections of the Core Facilities Centre ‘Bioresource Centre’ and under the state assignments of Siberian Institute of Plant Physiology and Biochemistry, Siberian Branch, Russian Academy of Sciences (No. 121031300009-4). The authors also thank the IRKU, IRK, LE herbaria, the Botanic Garden of Irkutsk State University and the Central Botanical Garden of the National Academy of Sciences of Belarus for the allowing to collect samples, V V Chepinoga, A A Gnutikov, N S Probatova, E A Marchuk, and S M Kuzmenkova for their help with sampling and species identification; and Dmitry Slizhov for the assistance in accommodation and transportation during fieldwork.
References

[1] Polozhy A V and Krapivkina E D 1985 Relics of Tertiary Deciduous Forests in the Flora of Siberia [Relikty Tretichnyh Shirokolistvennyh Lesov vo Flore Sibiri] (Tomsk: Tomsk University Press) p 157

[2] Belov A V, Bezrukova E V, Sokolova L P, Abzayeva A A, Letunova P P, Fisher E E and Orlova L A 2006 Vegetation of the Baikal region as an indicator of global and regional changes in natural conditions of the late Cainozoic Geography and natural resources 6 5–18

[3] Košanin N 1921 Waldsteinia ternata (Steph.) Fritsch im Bestande von Picea Omorica Österreichische botanische Zeitschrift 70 299

[4] Wood C 1972 Morphology and phytogeography: the classical approach to the study of disjunctions Annals of the Missouri Botanical Garden 59(2) 107–24

[5] Teppner H 1974 Waldsteinia ternata (Rosaceae) und ihre Vorkommen in den südöstlichen Alpen Phyton (Austria) 16(1–4) 281–99

[6] Chepinoga V V, Stepovnov N V, Protopopova M V and Pavlichenko V V 2019 The distribution of Waldsteinia ternata (Rosaceae) in the West Sayan Mountains Botanic Journal 104(8) 1203–10

[7] Smedmark J E E 2006 Recircumscription of Geum L. (Colurieae: Rosaceae) Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie 126 409–17

[8] Potter D, Eriksson T, Evans R C, Oh S, Smedmark J E E, Morgan DR, Kerr M, Robertson K R, Arsaenault M, Dickinson T A and Campbell C S 2007 Phylogeny and classification of Rosaceae Plant Systematics and Evolution 266 5–43

[9] Teppner H, Schuhely W and Weakley A S 2009 The chromosome numbers of Waldsteinia (Rosaceae-Colorieae) in North America Phyton 48(2) 225–38

[10] Doyle J J and Doyle J L 1987 A rapid DNA isolation procedure for small quantities of fresh leaf tissue Phytochemical Bulletin 19 11–5

[11] Protopopova M, Pavlichenko V, Gnutikov A and Chepinoga V 2019 DNA Barcoding of Waldsteinia Willd. (Rosaceae) species based on ITS and trnH-psbA nucleotide sequences Information Technologies in the Res. of Biodiversity. Springer Proc. in Earth and Environmental Sciences ed I Bychkov and V Voronin (Cham: Springer) pp 107–15

[12] Burgess K S, Fazekas A J, Kesanaakurti P R, Graham S W, Hushand B C, Newmester S G, Percy D M, Hajibabaei M and Barrett S C H 2011 Discriminating plant species in a local temperate flora using the rbcL-matK DNA barcode Methods in Ecology and Evolution 2 333–40

[13] Hassler M 2021 World Plants: Synonymic checklists of the vascular plants of the World (version 2021-03-17) Species 2000 & ITIS Catalogue of Life (Leiden: Naturalis)

[14] Kumar S, Stecher G and Tamura K 2016 MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets Molecular Biology and Evolution 33(7) 1870–4

[15] Ronquist F, Teslenko M, van der Mark P, Ayres D L, Darling A, Hohna S, Larget B, Liu L, Suchard M A and Huelsenbeck J P 2012 MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space Systematic Biology 61(3) 539–42

[16] Tamura K 1992 Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content bases Molecular biology and Evolution 9 678–87

[17] Hasegawa M, Kishino H and Yano T 1985 Dating of human-ape splitting by a molecular clock of mitochondrial DNA Journal of Molecular Evolution 22(2) 160–74

[18] Felsenstein J 1981 Evolutionary trees from DNA sequences: A maximum likelihood approach Journal of Molecular Evolution 17(6) 368–76

[19] Rambaut A 2006-2016 FigTree: Tree figure drawing tool, version 1.4.3 (Edinburgh: Institute of Evolutionary Biology, University of Edinburgh) http://tree.bio.ed.ac.uk (accessed 28.09.2021)

[20] Ewing G B 2012 Haplotype viewer (Vienna: Centre for Integrative Bioinformatics Vienna) http://www.cibiv.at/~greg/haplovviewer (accessed 28.09.2021)
[21] Probatova N S 2015 Validation of the combination *Waldsteinia maximowicziana* (Teppner) Prob. (Rosaceae) *Novitates Systematicae Plantarum Vascularium* 46 237

[22] Milne R I 2006 Northern Hemisphere Plant Disjunctions: A Window on Tertiary Land Bridges and Climate Change? *Annals of Botany* 98 465–72

[23] Krestov P V, Barkalov V Y, Omelko A M, Yakubov V V, Nakamura Y and Sato K 2009 Relic vegetation complexes in the modern refugia of Northeast Asia *V L Komarov Memorial Lectures [Komarovskiye Chteniya]* 56 5–63