**How to map a plantain: phylogeny of the diverse *Plantagineae* (Lamiales)**

Shipunov, Alexey1; Fernández A., José Luis2, Hassemer, Gustavo3; Alp, Sean1; Lee, Hye Ji1; Pay, Kyle1

1Department of Biology, Minot State University

2Real Jardín Botánico, CSIC, Madrid, Spain

3Federal University of Mato Grosso do Sul, Campus Três Lagoas, CEP 79610-100, Três Lagoas, MS, Brazil

**Abstract**

The tribe *Plantagineae* (Lamiales) consists of well-known, worldwide distributed plantains (*Plantago*), small aquatic *Littorella* and Andean (mostly Colombian) páramo shrubs *Aragoa*. This assemblage is unusual in many ways. First, it is a clear example of evolutionary reduction. Second, *Plantago* exhibits remarkably high diversification rates. Third, the worldwide distribution of plantains raises numerous questions related to vicariance and dispersal. Here we present the broadest phylogeny of the group to date and discuss some evolutionary, morphological and biogeographical implications.

**Introduction**

Plantains (ribworts), *Plantago* L. are remarkable plants. They grow almost everywhere all around the world, with exception of Antarctic and tropical wet forests (Rahn, 1996; Hassemer et al., 2016). Morphologically, they are easy to distinguish due to the unusual combination of characters (Linnaeus, 1754): sympetalous 4-merous non-showy flowers developing into circumcissile capsule-like fruit (pyxidium), and monocot-like leaves with arcuate or parallel venation, usually borne in rosette. Even in times when only sixteen species were described in *Plantago* (Linnaeus,1753), there was always mentioned a remarkable similarity between different species; this nowadays typically results in significant amount of incorrectly determined specimens, even in leading herbarium collections. It was guessed long time ago that plantains exhibit some of terminal stages of morphological reduction among Lamiales, and recent research (Preston et al., 2011) demonstrated how this flower reduction could happen within *Antirrhinum*-*Plantago* lineage. There is also a reduction of vegetative characters expressed in many plantains, and only few species have a branched stem and “dicotyledonous” leaves. In all, *Plantago* evolution towards anemophily resulted in significant morphological convergence with graminioid monocots.

Geographic distribution of *Plantago* is exceptionally broad, and this is due not only to cosmopolitan weeds *P. lanceolata* L. and *P. major* L. On many remote ocean islands for example, there are unique species of *Plantago*, and some of them exhibit a remarkable tendency to evolve into woody plants (Carlquist, 1970; Iwanycki et al., 2019). As a whole, *Plantago* likely underwent a rapid (Cho et al., 2004) evolution and recent (Meudt et al., 2015) diversification. A number of new *Plantago* species have recently been discovered and described, most of them rare and narrowly distributed (e.g. Hassemer and Rønsted, 2016; Hassemer et al., 2018a; Hassemer 2019). This makes *Plantago* a group of plants with considerable interest for biodiversity conservation (Hassemer et al., 2016). Furthermore, *Plantago* is a complicated genus also from a nomenclatural point of view (Hassemer, 2018a, 2018b), and many problems remain to be solved.

Aquatic *Littorella* P.J.Bergius (Bergius, 1768) comprise three species with considerably disjunct distribution; they grow in shallow waters of North America Great Lakes, Patagonia, and Northern Europe (here both in lakes and desalinated North and Baltic seas). These plants are morphologically close to *Plantago*, and Rahn (1996) joined two genera. However, when molecular data started to be available (Hoggard et al., 2003), *Plantago* and *Littorella* were shown as sister clades, and since then nearly all authors preferred to keep the tradition of recognizing the two genera as separate (see discussion in Hassemer et al., 2018b).

Contrary to *Plantago* and *Littorella*, woody shrubs *Aragoa* Kunth (Kunth, 1818) are the local endemics of páramo in Colombia (one species also in Venezuela). Until the first molecular results (Bello et al., 2002), it was never considered to be a sister group for *Plantago* + *Littorella* clade but rather unplaced in “old” Scrophulariaceae (Fernández, 1995). *Aragoa* is relatively speciose, containing more than 20 described species and also several inter-species hybrids which were never registered in two other genera (Fernández, 1993, 1995). This might relate with the low stability of páramo and subpáramo ecosystems and with the fact that *Aragoa* is likely animal-pollinated (Fernández, 1995). Nevertheless, flowers of *Aragoa* are actinomorphic and leaves reduced, similarly to two other genera of the group (Bello et al., 2004). The hybridization might explain the rapid radiation and speciation in this genus (Fernández, 2002).

These three genera form well-supported (both morphologically and molecularly), stable clade sister to *Veronica* L. in broad sense (Meudt et al., 2014) which we call hereafter the tribe *Plantagineae* Dumort. (Dumortier, 1829). It is curious that despite of multiple detailed morphology-based works, of which the most important are Barnéoud (1845), Decaisne (1852), Pilger (1937), Fernández (1995) and Rahn (1996), there for the long time were no comprehensive molecular study based on the broad sampling of this whole group. The broadest at the moment are works of Rønsted et al. (2002) and Hoggard et al. (2003) which included 59 and 27 species respectively, whereas the group is estimated to include ca. 250 species. To compare further, GenBank database as of June 2019 contains only about 140 species entries (and this does not account for possible synonymy). This situation have shown signs of improvement lately, and several publications which cover the complicated *Plantago* subg. *Plantago* (e.g. Hassemer et al., 2019; Iwanycki Ahlstrand et al., 2019) and subg. *Coronopus* (Höpke et al., 2019; Hassemer et al., 2017; Hassemer 2018b) were recently published. There are however no molecular works focusing on *Aragoa* diversity, and since Rønsted et al. (2002) nothing significant was published about molecular taxonomy of *Psyllium* and allies (we include here subgenera *Bougueria*, *Psyllium* s.str. and *Albicans*).

This situation dictates the necessity of the molecular study with the broadest sampling in mind. That study will continue the line of Rønsted et al. (2002) and employ similar markers but will aim for the greater species coverage, with the ultimate goal to assess all described species of the group.

Our goal was also to clarify multiple local problems of *Plantagineae* taxonomy and geography. When solved, they will improve the overall understanding, will help in conservation, ecology and invasive biology studies, and will ease the identification of *Plantagineae* species which still is a tedious and difficult task, especially for beginners.

**Material and methods**

It is virtually impossible to sample 250 species (Support Table 1) without help of the herbarium collections. Re-collection from nature, especially herbaceous short-living plants, is a task which is successful only rarely, especially if the time in the field is limited (e.g., Hassemer et al., 2018a). Whereas some of our samples were collected into silica gel from the living plants, the majority of work (94%) involved tissue samples which with the kind permission of herbarium curators, were taken from plants collected some years ago.

This strategy poses some restrictions. First, if the DNA purity and concentration do not significantly suffer with time (Choi et al., 2015), the quality of sequences heavily depend on sample age, collection methods, and DNA fragment to be amplified. With older age, putatively difficult drying process and longer fragment to sequence, our chances to obtain the good data were significantly lower. There are however, striking examples when we obtained reliable sequences from samples collected almost two hundred years ago.

We typically collected multiple samples per species and attempted to extract DNA and sequence our markers numerous times. This is how we ended with almost 900 samples and 1,700 sequencing chromatograms. Important was also to provide vouchers for each sample (Support Table 2), and to be sure of the correct identifications (therefore, either determinations from experts, or our own was used). At the end of this process, only few species were missed from our samples, and most of them are rare and / or strictly local.

We have now sequences of at least one DNA marker from 220 species (including 192 *Plantago* species). Information about 86 species was taken from public databases. Due to the obvious problems with identification (Funk et al., 2018), we always trusted our samples first. In all, we were able to increase the amount of available information three-fold (four-fold in *Aragoa)*.

DNA extraction performed using multiple standard protocols but soon after the start of the project (2011) we decided to stay with NUCLEOSPIN Plant II Kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany) which seems to be is a good trade-off between efficiency and simplicity. We improved this protocol in several points, e.g. increased the lysis time to 30 min and used thermomixer on the slow rotation speed (350 rpm) instead of water bath. To assess DNA quality, we used Nanodrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) which estimates concentration and purity (the 260/280 nm ratio of absorbance) of samples. Typically, 1.4 ratio was enough to guarantee PCR amplification of smaller markers whereas the ratio 1.7 was an average in the group. Lower DNA quality was typically obtained from samples presumably collected in wet climates.

Especially low was the quality of our *Aragoa* samples. This might be due to the widespread use of drying cabinets, the theory which was indirectly supported with good results which we obtained from samples dried without any help (even without silica gel), just in room conditions in Bogotá. Reversely, we sometimes were able to extract, amplify and sequence samples collected long ago, e.g. from *Plantago sinaica* Decne. collected in 1834!

Nevertheless, short barcoding DNA markers (Kuzmina and Ivanova, 2011) are the best to amplify for herbarium samples, therefore our choice was nuclear ITS2 and chloroplast *trn*L-F spacer and *rbc*L gene. We amplified them in accordance with Barcoding of Life recommendations and protocols (Kuzmina and Ivanova, 2011). It must be also mentioned here that some samples were sequenced with a direct help of Barcoding of Life (“SAPNA” project). This last project provided us also with sequences of mitochondrial COI and plastid *mat*K markers.

Typically, our PCR the reaction mixture had a total volume of 20 μL which contained 5.2 μL of PCR Master Mix (components mostly from Thermo Fisher Scientific, Waltham, Massachusetts supplied with Platinum DNA Taq Polymerase), 1 μL of 10 μM forward and reverse primers, 2 μL of DNA solution from the extraction and 10.8 μL of MQ purified water (obtained from a Barnstead GenPure Pro system, Thermo Scientific, Langenselbold, Germany) in the TBT-PAR water mix (Samarakoon et al., 2013). The latter was developed to improve amplification from the herbarium samples. Thermal cycler programs were mostly 94° for 5 min, then 35 cycles of 94° for 1 min; 51° (or similar, depending on the primer) for 1 min, 72° for 2 min, and finally 72° for 10 min. PCR products were sent for purification and sequencing to Functional Biosciences, Inc. (Madison, Wyoming) and sequenced there in accordance with standard Sanger-based protocol. Sequences were obtained, assembled and edited using Sequencher™ 4.5 (Genes Codes Corporation, Ann Arbor, Michigan, USA). Approximately 500 assembled sequences were selected for the next steps.

All subsequent steps were made through the “Ripeline” workflow. This workflow is the collection of UNIX shell and R (R Core Team, 2018) scripts which automate steps related with sequence selection, quality checking, alignments, gap coding, concatenation and phylogenetic tree production. Ripeline involves multiple pieces of software, for example, AliView (Larsson, 2014), MUSCLE (Edgar, 2004), APE (Paradis et al., 2004), MrBayes (Ronquist and Huelsenbeck, 2003), ips (Heibl, 2008), shipunov (Shipunov et al., 2019), and phangorn (Schliep, 2011). The Ripeline example which uses its basic features and includes associated R code is freely available from the primary author’s Web site: http://ashipunov.info/shipunov/software/r/r-en.html.

With the help of Ripeline, we were able to obtain maximal parsimony (MP) and Bayesian (MB) phylogeny trees. MP analyses were run with the help of R phangorn package (Schliep, 2011) using parsimony ratchet (Nixon, 1999) with 2000 iterations and then 1000 bootstrap replicates. MB analyses were run through the combination of MrBayes 3.2.6, and R ips and shipunov packages (Ronquist and Huelsenbeck, 2003; Heibl, 2008; Shipunov, 2019b). MCMC chains were run for 1,000,000 generations, sampling every 10th generation resulting in 100,000 trees. The first 25% trees were discarded as burn-in, and the remaining trees were summed to calculate the posterior probabilities. With the aid of R ape package (Paradis et al., 2004), all trees were rooted with *Veronicastrum virginicum* (L.) Farw. and *Tetranema roseum* (M.Martens & Galeotti) (= *Tetranema mexicanum* Benth). as outgroups, or with the *V. virginicum* alone. Our MP and MB trees have similar topologies, thus we will illustrate phylogenies below with MB trees only.

Our phylogeny trees were based on two data sets. Since all single marker trees were concordant, we used (with the help of Ripeline) the super-matrix approach. Our first data set included multiple barcoding markers available from public databases and our sequencing, that is chloroplast *rbc*L, *trn*L-F, *mat*K, and also nuclear ITS and mitochondrial COI. We call this dataset “broad” since it is relatively rich of data but has a limited sampling along the species dimension (87 entries and 4188 bp, including 656/497/561/1565/909 bp in COI, ITS2, *rbc*L, *trn*L-F and *mat*K, respectively). The second dataset was made with the widest species coverage but includes only ITS2 and *trn*L-F data originated mostly from our sequencing efforts. Below, we designate it as “tall” (273 entries including some subspecies and forms and 2062 bp length, including the same ITS and *trn*L-F fragment lengths as above). In addition, we attempted to construct some other datasets, for example, “full” which includes all available DNA data but in that case the amount of missing data exceeded 54% which complicated tree estimations.

Ripeline is also capable to use morphological characters, and we employed the updated and expanded morphological dataset from Rahn (1996) to make combined (molecules + morphology) and pure morphological datasets. In the former, to emphasize the weight of morphological characters we used the hyper-matrix approach (Ashkenazy et al., 2018) and multiplied morphological dataset several times in order to achieve the approximate equality between numbers of molecular and morphological characters. We added characters of seed sculpture (Shipunov, 1998a, 1998b; Shehata and Loutfy, 2006) to the characters used in Rahn (1996), and expanded the dataset with species absent in the last work. In total, our binary morphological matrix has 114 characters and 271 entries.

We also were able to use the measurements of seven most apparent, morphometric “spot” characters of *Plantago*: petiole, leaf, spike and scape lengths, leaf maximal width, presence of taproot and looseness (“gaps”) in the inflorescence. In total, we measured 405 herbarium samples (used also for DNA extraction).

Using morphological, morphometric and DNA datasets, we were able to perform the broad spectrum of statistical analyses, including Procrustes analysis of the correspondence between molecular and morphological information (Peres-Neto and Jackson, 2001; Balbuena et al., 2013) and nearest neighbor machine learning (Ripley, 1996) for the placement of under-studied taxa. As an additional source to use in placement process, we employed the combined molecular + morphology phylogeny tree.

We also employed the recursive partitioning (Venables and Ripley, 2002; Höpke et al., 2019), the “learning” technique which takes the pre-existed classification and creates binary trees for the rest of data set. The structure of these trees is similar to the dichotomous keys (Therneau et al., 2014). Naturally, results of recursive partitioning are applicable for the construction of the dichotomous keys which could help in discrimination of *Plantago* sections and species.

Thousands of observed samples in multiple herbarium collections all over the world were brought also taxonomic and floristic results based mostly on morphological observations, and below we present them together with molecular data.

Datasets and scripts used in preparation of this publication are available from the first author’s Open Repository here: http://ashipunov.info/shipunov/open/ . We encourage readers to reproduce our results and develop our methods further. All sequences were deposited into the GenBank.

In the paper we followed the “appropriate citation of taxonomy” (ACT) principle (Seifert et al., 2008) and cite names of the most supra-species groups (Reveal, 2012) plus those species which are separately discussed.

**Results**

***Plantagineae***

All trees based on “broad” and “tall” datasets returned the stable (*Aragoa*, (*Littorella*, *Plantago*)) topology (Fig. 1), typically with the longest branch leading to *Aragoa*.

***Aragoa***

Generally, the stability of subclades is not high in *Aragoa* (Fig. 2). The most stable is a placement of *Aragoa lucidula* S.F.Blake as a sister to all other studied *Aragoa* species. Within the rest of the subtree, the majority of species form one clade with subdivisions mostly without high support. Morphologically outstanding *A. dugandii* Romero forms the clade with *A. lycopodioides* Benth. and *A. occidentalis* (all branches here are significantly longer than in other parts of *Aragoa* tree). The relatively broad, patent leaves and the simple inflorescence of *A. dugandii*, together with relatively large flowers might be therefore interpreted as an adaptation to environments where moisture loss is not so critical as for many other species of the genus.

The second unusual species, *A. perez-arbelaeziana* Romero, forms clade with *A. romeroi* Fern.Alonso (Fig. 4, 5a). In that light, the presence of long, pendulous, yellowish corollas, unusual in the genus, can be interpreted as a result of recent adaptive radiation to a specific type of pollinator (hummingbirds: Fernández, pers.obs.) which does not entail the significant modifications in vegetative structures. In fact, the distal buds and young branches in *A. cupressina* and *A. perez-arbelaeziana* are morphologically quite similar.

***Littorella***

Three (two in the “broad” dataset) species of *Littorella* make the stable group where European *L. uniflora* (L.) Asch. is sister to American *L. americana* Fernald and *L. australis* Griseb. ex Benth. & Hook.f.

***Plantago in general***

There is a relatively high support for three major subdivisions of *Plantago* which correspond with subgeneric rank (1). The topology is robustly (*Psyllium* s.l., (*Coronopus*, *Plantago*)), or in more detail, ((*Bougueria*,(*Psyllium* s.str., *Albicans*)), (*Coronopus*, *Plantago*)). Subgenera *Plantago*, *Coronopus* and *Psyllium* clade form the remarkable “three-ridge” phylogenetic density surface (Fig. 3).

***Plantago subg. Plantago***

Only trees originated from the “broad” dataset have relatively high support for clades within this group whereas “tall” trees have the reliable support only for some terminal clades (Fig. 2, 4, 5b-d).

One of the most stable groups consist of *P. media* L., *P. canescens* Adams, *P. arachnoidea* Schrenk ex Fisch. & C.A.Mey., *P. krascheninnikovii* C.Serg., *P. maxima* Juss. ex Jacq., *P. perssonii* Pilg. and *P. schwarzenbergiana* Schur. Tetraploid, xeromorph variant of *P. media* described as *P. urvillei* Opiz (*P. media* subsp. *stepposa* (Kuprian.) Soó), typically does not branch with *P. media* s.str.

*Plantago krascheninnikovii* from Urals is habitually close to the inland forms of *P. maritima* L. (however, it lacks the key feature of subg. *Coronopus*, pilose corolla tube). On our trees, it groups not far away from *P. arachnoidea* from Central Asia. *Plantago lorata* (J.Z.Liu) Shipunov described from Central Asia (Shipunov, 2000a) is morphologically indistinguishable from *P. perssonii* Pilg. while this last species robustly groups with *P. arachnoidea*. Morphologically unusual *P. reniformis* Beck from Balkans frequently also groups here with low support.

Another stable group is formed by species from sect. *Micropsyllium*, palearctic *P. polysperma* Kar. & Kir., *P. tenuiflora* Waldst. & Kit., and nearctic *P. elongata* Pursh, *P. heterophylla* Nutt. and *P. pusilla* Nutt. Here we noted that geographically isolated, perennial *P. tenuiflora* from Öland (first described as separate species *P. minor* Fr.) does not group with typical *P. tenuiflora* but instead groups with *P. polysperma*. Forms of *P. elongata* which originally have been described as *P. bigelovii* A.Gray, do not typically group with *P. elongata* s.str.

Less stable but relatively consistent group are forms around polymorphic *P. asiatica* L. (including subgroup formed by *P. schneideri* Pilg., *P. centralis* Pilg. and *P. cavaleriei* H.Lév.) All these forms are molecularly close to the typical *P. asiatica* from mainland China, and not easily distinguishable morphologically. Typical *P. asiatica* was found on Hawaii Island, thus extending the range of the species to mid-Pacific. However, all “*P. asiatica*” from mainland USA we found to be either *P. major* or *P. rugelii* Decne. (Shipunov, 2017; Shipunov, 2019a).

*Plantago hakusanensis* Koidz. is a Japanese alpine endemic species with distinct morphology. On our trees, it is close to *P. asiatica*. In PE herbarium, we discovered the Yunnan sample labeled as “*Plantago zhongdainensia*” (nomen nudum) which morphologically might be considered similar to both *P. hakusanensis* and *P. asiatica*. Unfortunately, DNA data is not available from this sample. Not very far from *P. asiatica* is also *P. hasskarlii* Decne. from Java mountains which is also morphologically distinct. Another species from Southeast Asia, *P. incisa* Hassk., groups outside of *P. asiatica* clade(s). On “broad” trees, we do not have such sampling but both Japanese and Chinese *P. asiatica* robustly group together.

There is an unusual form collected from China but as far as we know, it never received the taxonomic name. It is morphologically somewhat similar to the *P. densiflora* J.Z.Liu (synonymized with *P. asiatica* in the “Flora of China”, Li et al., 2011). However, this form, “*Plantago* sp. Hupeh1” has typically 4–6 large black seeds (and also large fruits) which is not in agreement with *P. densiflora* protologue. On our trees, it groups with *P. depressa* Willd. and allies (e.g., *P. komarovii* Pavlov and *P. camtschatica* Link). In addition, on “broad” trees, *P. depressa* robustly groups together with *P. macrocarpa* Cham. & Schltdl.; this grouping is also presents on “tall” trees, with less support.

We were not successful in DNA amplification from the sample of *P. camtschatica* which was collected on Aleutian islands as “*P. media*” (Tatewaki and Kobayashi, 1934; Hulten, 1960) but on the base of morphology, we are confident that range of this species formally extends into North America—the fact which was not known before (Shipunov, 2017; Shipunov, 2019a).

American *P. eriopoda* Torr., *P. rugelii*, *P. sparsiflora* Michx., and *P. tweedyi* A.Gray are robustly supported as a clade on “broad” trees. Here belong also two samples collected in Chihuahua desert (BRIT) from northern Mexico; these plants have multiple morphological differences from other species in this group but cluster together with *P. eriopoda* and *P. tweedyi*. *Plantago rugelii* which morphologically is really hard to tell from *P. major*, doest not group with this last species on any tree; these two species were found to be in different sections (Hassemer et al., 2019).

*Plantago major* is also far away from *P. asiatica*, which was pointed out in Hassemer et al. (2019). Instead, *P. major* s.l. groups with *P. japonica* Fr. & Sav., *P. cornuti* Gouan, *P. gentianoides* Sibth. & Sm. and *P. griffithii* Decne., albeit with low support. Sequences from polyspermous form of *P. major* (Morgan-Richards and Wolff, 1999) described as *P. uliginosa* F.W.Schmidt, are identical to the typical *P. major*. *Plantago griffithii* which is frequently considered as a form of *P. gentianoides* groups with the same species (in strict sense) on our trees, but this grouping is unstable.

*Plantago pachyphylla* A.Gray and *P. hawaiensis* (A.Gray) Pilg. (both from Hawaii) group together, and also with *P. aundensis* P.Royen from New Guinea. Alpine form of *P. pachyhylla* from Kauai (labeled in HUH as “*Plantago nubicola* Tessene”, nomen nudum) clusters outside of *P. hawaiensis* + *P. pachyphylla* from Hawaii island.

Two Australian species, *P. triandra* Berggr. and *P. unibracteata* Rahn, always cluster together outside of the rest of subg. *Plantago*, which is in disagreement with the results of Hassemer et al. (2019) and Iwanycki et al. (2019).

Apart from the relatively robustly supported clades, our phylogeny trees, especially from “tall” dataset, provide the basic ground for the placement of little studied or previously molecularly not studied forms. Some of these cases are listed below.

*Plantago laxiflora* Decne. from South Africa is morphologically unusual for the region and groups outside of African species, close to Australian and New Zealand groups. Other African and Madagascan species, i.e. *P. africana* Verdc., *P. longissima* Decne., *P. palmata* Hook.f., *P. remota* Lam. and *P. tanalensis* Baker have a tendency to group together with small support.

Most of *Plantago* sect. *Virginica* species do not group with high support. However, we note that Andean *P. oreades* Decne. is always far from the rest of *P. australis* Lam. group. Another, Peruvian form from this section was listed by Knud Rahn (MO herbarium note) as possible new species; on our trees it groups with different members of section, including South American *P. tomentosa* Lam.

The second “unknown” from Peru, sect. *Virginica* sample from NY with long stem (unusual in subg. *Plantago*) frequently groups with *P. tenuipala* (Rahn) Rahn from Columbia.

*Plantago firma* Kunze ex Walp. was typically considered as strictly Chilean species, but we have now both molecular and morphological support that this species grows also in Peru. All Chilean and Peruvian *P. firma* samples robustly group with another species, bipolarly distributed *P. truncata* Cham. & Schltdl.

While there is a little confidence among branches which belong to the rest of sect. *Virginica*, we were able to place in that group molecularly those species which have not been sampled before, namely *P. argentina* Pilg., *P. berroi* Pilg., *P. buchtienii* Pilg., *P. dielsiana* Pilg., *P. floccosa* Decne., *P. jujuyensis* Rahn, *P. orbignyana*, *P. penantha* Griseb., *P. tenuipala* (Rahn) Rahn, *P. ventanensis* Pilg. and *P. venturii* Pilg.

Knud Rahn’s series *Oliganthos* species (*P. barbata* G.Forst., *P. correae* Rahn, *P. pulvinata* Speg., *P. sempervivoides* Dusén and *P. uniglumis* Wallr. ex Walp.) were sequenced the first time as a totality. On our “tall” trees, the group does not have a high support but clusters together with *P. moorei* Rahn, *P. tehuelcha* Speg. and *P. fernandezia* Bertero ex Barnéoud, all from South America’s Cone and surrounding islands.

Most of Australian species form a low supported but relatively stable grade; here we were able to place some under-researched species: *P. antarctica* Decne., *P. depauperata* Merr. & L.M.Perry, *P. drummondii* Decne., *P. gunnii* Hook.f., *P. polita* Craven (New Guinea) and *P. turrifera* B.G.Briggs & al.

***Plantago subg. Coronopus***

On this stable trees (Fig. 4, 5b), the topology always supports the subdivision of sects. *Maritima* and *Coronopus*. Within sect. *Maritima* we were able to place with confidence the rare Central Asian *P. eocoronopus* Pilg. (as a sister to the whole group) and North African *P. rhizoxylon* Emb. We detected the presence of the “true” *P. maritima* in South Africa (PRE herbarium), these samples are molecularly not different from the rest of *P. maritima*.

Macaronesian *P. asphodeloides* Svent. is the sister to other species from sect. *Coronopus*, and North African *P. crypsoides* Boiss. is sister to Mediterranean *P. serraria* L.

***Plantago* subg. *Psyllium* and allies**

Within this stable group (Fig. 4, 5d), *P. nubicola* (Decne.) Rahn (which sometimes regarded as a separate genus *Bougueria*), is always the basal branch. On all trees, subsequent topology is (sect. *Psyllium*, (“American clade”, “*Plantago ciliata* clade”, “Mediterranean clade”)), these we will describe in details below.

Section *Psyllium* forms a robust, relatively long branch which is split between mostly annual species with non-linear bracts (e.g., *P. squarrosa* Murray) and mostly perennial, woody species with narrow bracts (e.g., *P. arborescens* Poir.).

“*Plantago ciliata* clade” on “tall” trees is sister to “American clade” whereas on “broad” trees *P. ciliata* Desf. is sister to “Mediterranean clade” (with lower support). This clade includes *P. ciliata* and two successfully sampled species from sect. *Hymenopsyllium*, i.e. *P. cretica* L. and *P. bellardii* All.

“American clade” is in essence Rahn’s sect. *Gnaphaloides* (but species inside are not arranged in accordance with Rahn’s lower level classification). *Plantago erecta* E.Morris is variably at the base of this group, and *P. helleri* Small is close to southern *P. nivea* Kunth. The rest of North American species form a stable clade (which therefore roughly correspond with Rahn’s ser. *Gnaphaloides*), where *P. aristata* Michx. and *P. argyrea* E.Morris form a subgroup. On “tall” trees (where sampling is reliable), species from Central and South America form the *P. tandilensis* (Pilg.) Rahn + *P. brasiliensis* Sims + *P. bismarckii* Nederl. clade, *P. grandiflora* Meyen clade, *P. sericea* Ruiz & Pav. grade (incl. *P. lamprophylla* Pilg., *P. nivea*, *P. helleri*, *P. linearis* Kunth, and *P. tolucensis* Pilg.) and ser. *Hispiduleae* clade. The latter includes also *P. johnstonii* Pilg. and samples of *P. litorea* Phil. collected in Peru (thus extending the range of this Chilean species). Samples of some *P. sericea* subspecies do not branch together with the bulk of *P. sericea* samples (Fig. 5d).

“Mediterranean clade” corresponds with sects. *Montana*, *Lanceifolia* and *Albicans* (except *P. ciliata*) but species on our phylogeny trees are arranged differently. The first subclade formed with members of the first two sections plus *P. lagocephala* Bunge and two species from sect. *Albicans* ser. *Minutae*: *P. minuta* Pall. and *P. lachnantha* Bunge. Sections *Montana* and *Lanceifolia* represented as it was proposed by Rahn with exception of *P. loeflingii* L. (not in sect. *Montana* but in sect. *Lanceifolia* instead).

“Mediterranean clade” 2nd subclade consists mostly of species from sect. *Albicans* (whereas some of them group with the 1st subclade, see above). *Plantago amplexicaulis* Cav. (sect. *Bauphula*) and *P. stocksii* Boiss. ex Decne. (sect. *Albicans* ser. *Ciliatae*) group together on the base of this group. Next branch(es) are *P. ovata* Forssk. and *P. psammophila* Agnew & Chal.-Kabi + Ethiopian *P. annua*. The rest of this subclade are species from ser. *Albicantes* and *Ciliatae*, plus *P. notata* Lag.

***Morphological and combined analyses***

Procrustes analysis allows for the embedding of two multivariate datasets (Peres-Neto et Jackson, 2001; Balbuena et al., 2013) and related statistical tests. Our molecular and morphological datasets are significantly correlated (correlation in a symmetric Procrustes rotation: 0.7748, significance = 0.001 based on 999 permutations) but individual placements are variably shifted (Fig. 6).

Even after intensive sampling, some species of the group still lack the molecular information. There are also species where only ITS2 sequences are available. With *k* nearest neighbor machine learning, we obtained the section / series placements of these *Plantago* species. More than half of them were placed with high (> 90%) bootstrap confidence (Table 1). The placement technique explained on Fig. 7 where artificially produced coordinates are assigned to subgenera on the basis on the nearest neighbor affinities. In case of *Aragoa*, we only operated with an existent classification (Fernández, 1995) combined with phylogeny trees so here our placements were structured as trios of most close species (Table 1).

Morphometric “spot” characters such as linear measurements of most recognizable organs (leaves and inflorescences) are frequently used in identification keys. Chi-squared tests returned the appropriate p-values when comparing leaf shapes with subgenera, sections and macro-regions (0.0005, 0.016 and 0.0055 respectively) and typically large effect sizes (corrected Cramer’s V 0.39, 0.34 and 0.26 respectively, see also Fig. 8). At the same time, relative sizes of the stalk and spike were not significant. There is also a support for the pattern of gapped spike *vs*. sections (p-value 0.0004 and 0.48 corrected Cramer’s V).

Another task was to check which morphological characters (binary and morphometric) have the biggest “molecular weight”, in other words, have the highest correlations with molecular phylogenies. We measured the average or maximum Spearman correlation between morphological matrices and phylogeny trees based either on “tall” dataset or molecular-morphological dataset. Among morphometric characters (Fig. 9a), most “heavy” was the presence of taproot (Fig. 9b), and then length of leaves. Most “heavy” (“top 10”) binary characters (Fig. 9c) were seed surface type 4 (with elongated ridges: Shipunov, 1998b), long corolla (> 4 mm or > 3 mm) lobes, opposite leaves, presence of pedicel, truncated base of leaf blade, presence of glandular hairs, elongated stem, antrorse hairs on the stalk, and presence of non-glandular hairs with the strongly refracted walls.

We believe that the difficult problem to create a comprehensive identification key might be helped with machine learning. We used recursive partitioning (Venables et Ripley, 2002) to construct the classification tree (Fig. 10b-c). With binary morphological binary characters, we employed three runs, excluding characters which were used in the previous run. Resulted recursive classification tree models employed 20, 20 and 19 characters (out of 115) and had 25.3%, 35.5% and 48.1% misclassification errors, respectively. With morphometric characters, the resulted tree used all 7 characters and returned 75.5% misclassification error.

**Discussion**

***Plantagineae***

There is an unequivocal support for the *Aragoa* (*Littorella* (*Plantago*)) structure of the group phylogeny. This structure is concordant with current understanding of the evolution of the tribe including evolution of flower symmetry (Preston et al., 2011). Reduced leaf morphology might also be explained with this phylogeny.

*Aragoa* taxonomy was the first time reviewed here from the data obtained in molecular research. In general, there is some support for grouping obtained in the research solely based on morphology (Fernández, 1995). However, some morphologically unusual species like *A. dugandii* with relatively broad leaves, and *A. perez-arbelaeziana* with long tubular flowers do not make separate clades but are clustered together with more “typical” species (*A. lycopodioides* and *A. romeroi*, respectively, both from main clade). The machine learning placement of three unsampled *Aragoa* species resulted in selection of possible “candidate neighbors” (Table 1) from the same main *Aragoa* clade.

*Littorella* with three distinct species lineages is robustly supported as a sister group of *Plantago* which is again a support for the current understanding of its evolution (Hoggard et al., 2003). Our data agree with a view of morphologically and ecologically outstanding *Littorella* as a separate generic lineage.

*Plantago* taxonomy is apparently the most complicated part of our research. In general, there are no significant conflicts with the most recent studies of the genus based on morphology summarized in Rahn (1996).

However, there are numerous findings and differences to emphasize. All our trees reproduce ((*Coronopus*, *Plantago*), *Bougueria*, (*Psyllium*, *Albicans*)) backbone. This does not contradict with recent findings based on plastome research (Hassemer et al., 2019), as well as with the older phylogeny studies with most broad coverages (Rønsted et al., 2002; Hoggard et al., 2003). Based on our trees, it is possible to keep the latter three subgenera as such on the basis of strong molecular and morphological support. Alternatively, it is also possible to lump them in one bigger subg. *Psyllium* and leave only three subgenera in the genus *Plantago*. This *Psyllium* s.l. union has both molecular and morphological support: two (rarely one) ovules and seeds, cotyledons perpendicular to placenta, placenta side of seed deeply concave, hairs with a basal cell shorter than broad, leaves often linear and spike usually short in relation to scape (Rahn, 1996). The third alternative would be to accept this union as a separate genus, *Psyllium* Mill. s.l. (Shipunov, 1998a, 2000a). This however will significantly decrease the nomenclatural stability in the group.

***Plantago subg. Plantago***

In general, our trees in this part do not such provide a clear, well resolved picture as in plastome studies (Hassemer et al., 2019). However, they contain plenty of new information about possible placements of previously not molecularly studied species as well as cluster memberships for species which were not included in plastome research. For example, it is mentioned in Hassemer et al. (2019) that “...based on our phylogeny it is impossible to infer the position of the five unsampled American species in Rahn’s (1996) series *Oliganthos* (*P. barbata*, *P. correae*, *P. pulvinata*, *P. sempervivoides* and *P. uniglumis*)”. Our current trees place these species (except *P. correae* where we have no data) in one clade which includes also *P. rigida*, *P. tubulosa*, *P. moorei*, *P. tehuelcha* and *P. fernandezia*, the placement which is well justified geographically. Morphologically unusual *P. sempervivoides* is a basal branch, and two species from sect. *Carpophorae* are sister to *P. fernandezia* + *P. barbata*. This group likely has an Andean origin, and *P. fernandezia* might therefore arrive to Juan Fernández from South America -(Stuessy et al., 2017). Position of *P. fernandezia*, *P. tehuelcha* and sect. *Carpophorae* species is different on the plastome trees (Hassemer et al., 2019) but these trees have low support exactly in these parts.

Another complicated group was not resolved completely but we provide several insights for the placements and phylogeny in general of species around polymorphic (Matsuo, 1989, Ishikawa et al., 2006; Ishikawa et al., 2009) and widespread *P. asiatica*. Most of these forms cluster together separately from the *P. major* clade; thus, morphological similarity does not justify taxonomic closeness here. *Plantago hakusanensis* might cluster either within this group or as a basal branch, same is true for *P. hasskarlii*. Relations of these both species mandate the deeper research.

Molecular data from *Plantago japonica* together with ecology and morphology suggests that this species is close to but separate from *P. major* (Matsuo, 1989, Ishikawa et al., 2009).

Samples from Hubei province of China have appearance of *Plantago asiatica* but differ from that species in many details. On our trees, these samples (“*Plantago* sp. Hupeh1”) make a branch located near *P. komarovii* + *P. camtschatica* clade. These samples merit criteria of new species but we cannot exclude the possibility of allopoliploid origin of these forms (with ITS kept from sect. *Pacifica* parent). Allopolyploid origin of tetra- or hexaploid *P. asiatica* and *P. rugelii* was suggested by Ishikawa et al. (20109) but we believe that the thorough study of Chinese, Korean and Japanese *Plantago* subg. *Plantago* species is required before reaching any conclusions. In the light of Ishikawa et al. (2009) report, the recent historical origin of *P. rugelii* (which is close to *P. sparsiflora* on our trees) might be also justified.

There are seven endemic *Plantago* species described from New Guinea (*P. aundensis*, *P. depauperata*, *P. montisdicksonii* P.Royen, *P. papuana* P.Royen, *P. polita*, *P. stenophylla* Merr. & L.M.Perry and *P. trichophora* Merr. & L.M.Perry), but DNA extraction from available samples failed in 90% of cases. We are able however to place four of these species: *P. aundensis* as close to Hawaiian *P. pachyphylla* s.l. and *P. hawaiensis* (the third Hawaiian species, *P. princeps* does not hold a stable position on our trees); *P. depauperata* and *P. polita* with Australian *P. muelleri* Pilg.; and *P. trichophora* with Australian *P. gaudichaudii* Barnéoud. Still, much more sampling is needed, and Hawaiian species also deserve the closer look (Hassemer et al., 2019).

Morphologically unusual samples (Fig. 11) from Chihuahua (Mexico) are physionomically similar to *P. gaudichaudii* from Australia and *P. remota* from South Africa but we believe that it is the clear example of morphological convergence. On our trees, these samples belong to sect. *Pacifica* clade and branch together with Midwestern *P. eriopoda* and *P. tweedyi* from Rocky Mountains region. We describe them below as a new species:

***Plantago chihuahuensis*** Shipunov, **sp.nov.**

Differs from *Plantago eriopoda* and *P. tweedyi* by: narrow and long leaves (0.4–0.6 cm × 9–15 cm), tall inflorescences (30–35 cm) and sparsely (with gaps 0.5–1 cm in the middle of spike) arranged flowers. From *P. sparsiflora*, differs by ecology (semi-deserts, on alkaline soils), narrow leaves and shorter inflorescences.

TYPE:— Sand Diego de Alcala, 1130 msnm. Pastizal halophito de *Sporobolus airoides*. Municipio: Chihuahua. 16 Agosto 1997. *C.Yen, E.Estrada 7915* (holotype: BRIT!)

Etymology:—Named after the region of collection, Chihuahua state of Mexico.

Distribution:—MEXICO, CHIHUACHUA. (BRIT *C. Yen, E. Estrada 5644*!)

Description:—Plants perennial; roots taproots, thick. Stems about 1 cm. Leaves 0.4–0.6 cm × 9–15 cm; blade linear, margins entire, veins inconspicuous, surfaces glabrous. Scapes 30–35 cm, spikes brownish, 70–140 cm, very loosely flowered (rachis visible between flowers); bracts ovate, 1.5–2 mm, length 0.6–0.8 times sepals. Flowers: sepals 2 mm; corolla radially symmetric, lobes spreading, 1 mm; stamens 4.

The same *Pacifica* clade includes on broad trees *Plantago macrocarpa*, the Norhtern Pacific seashore species. On the “tall” trees, placement of this species is not stable.

*Plantago krascheninnikovii* is superficially similar to the inland forms of *P. maritima* and therefore was treated as a member of subg. *Coronopus* (Shipunov, 2000b). Known populations of this rare Urals species do not typically form the ripe seeds (Shipunov, 1998a) thus preventing the correct placement on the base of morphology. Here we the first time confirm its similarity with other *Lamprosantha* species, for example, Eastern European *P. schwartzenbergiana*.

The southern tetraploid forms similar to *Plantago media* but with thick, erect, grayish leaves (often regarded as *P. urvillei* Opiz or *P. media* subsp. *stepposa* (Kuprian.) Soó), do not cluster with *P. media* s.str., thus necessitate the separation at least as a subspecies (Shipunov, 1998a, 2000b), analogously to *Plantago media* subsp. *brutia* (Ten.) Arcang. (Palermo et al., 2010). Here should be noted that *P*. *media* subsp. *stepposa* must not be mixed with the similarly looking shade, mesophytic plants of *P. media* (Shipunov, 1998a).

Andean *Plantago oreades* Decne. with distinct morphology (narrow leaves, long inflorescences, broad bracts, 1–3 seeded fruit, thick roots) was nevertheless included into broadly understood *P. australis* (Rahn, 1974). On our trees, it almost always separate from the other *P. australis* samples. Therefore, we propose here to re-establish this species.

There are many local endemics in sect. *Virginica* (Hassemer, 2019a). Morphologically unusual forms collected in Cuzco area (Peru) were labeled by Knud Rahn in herbarium (MO) as possible new species (Fig. 12). As we see these samples separate on our trees, we take this additional molecular evidence into account and describe the new species:

***Plantago cuzcoënsis*** Shipunov, **sp.nov.**

Differs from *Plantago tomentosa* by having two ovules (0–2 seeds) and from *P. virginica* and *P. rhodosperma* by perennial habit (thick vertical caudex).

TYPE:— Peru. Cuzco: Calca Province. Huambuito, San Salvador. (Plants from transects and general colelctions). 3200 m. Herb, flowers green. June 1986. *Rita Dueñas. 206*. (holotype: MO!)

Etymology:—Named after the region of collection, Cuzco province (Peru).

Distribution:—PERU, CUZCO. (MO *Rita Dueñas. 57*!)

Description:—Plants perennial; caudex well developed, hairy. Stems more then 1 cm. Leaves 9–10 cm × 0.8–1.2 cm; blade oblanceolate, margins toothed, veins conspicuous, surfaces pilose. Scapes 10–25 cm, hairy. Spikes yellowish or brownish, 2.5–6 cm; bracts ovate, rarely triangular, 2–2.5 mm, length equal to sepals. Flowers: corolla radially symmetric, lobes spreading, 1.2–1.8 mm; stamens 4. Seeds 0–2.

The second mentioned above South American “unknown”, long-stemmed sample from Peru deserves, on our opinion, the more thorough research.

There were species with positions in the specific sections inside subg. *Plantago* was “inferred based on the accumulated knowledge” (Hassemer et al., 2019). Many of such species are now more formally distributed between these sections (Table 1), and this is reflected in our working classification of *Plantagineae* (Support Table 1).

***Plantago* subg. *Coronopus***

Our trees provide one of the most complete phylogenies for the subg. *Coronopus*, and are in concordance with the recent work of Höpke et al. (2019). We were able to place those species which have not been the subject of molecular studies. The most interesting are positions of Canarian *P. asphodeloides* Svent as sister to the rest of species from sect. *Coronopus*, and *P. eocoronopus* Pilg. as sister to the rest of sect. *Maritima*. The latter species is the rare Afghan plant, practically absent in collections. Pilger (1937) guessed the basal position of this species and now we can support it on the base on both molecules and morphology (Shipunov, 2000a). All our “*P. schrenkii*” C. Koch samples from Arctic are identical to *P. maritima* (Shipunov, 2015).

***Plantago* subg. *Psyllium* and allies**

Generally, this part of our trees is the most congruent with the classification of Rahn (1996). Many of his groupings are retained (not in the same volume though). Our data is also congruent with the most complete (until now) sampling of the group (Rønsted et al., 2002) and provide the robust support for many sub-groupings. As a result, we believe that now, as one of the most important results of our research, we are able to propose the new classification of the whole *Plantagineae* (Supplement Table 1).

Among other interesting results, we found that likely extinct *P. johnstonii* (Hassemer et al., 2017) is close to the coastal annual *P. limensis* and therefore belongs not to ser. *Brasilienses* but to ser. *Hispiduleae*. It is possible then that perennial life form of the former species is the result of adaptation to the “lomas” microclimate.

The most recent review of sect. *Lancifolia* (Hassemer, 2019) is in agreement with our phylogeny results but also provides new insights for the classification of this Mediterranean taxon. More research is needed to understand relations of rare endemic species of this group.

Within subg. *Psyllium* s.str., we obtained two stable clades. One of them consists mostly of perennial and woody species with narrow bracts. We propose it below as a new section with 7 species (Support Table 1):

*Plantago* sectio *Arborescens* Shipunov, **sect.nov.**

Basionym: —*Plantago arborescens* Poir.

Similar group was described by Andrzejewska-Golec (1992) in the rank of series. She underlined also the absence of club-shape hairs but counted there only four species.

***Morphological and combined analyses***

With the Procrustes analysis we found that the overall “picture of diversity” is retained between morphological and molecular approaches. In other words, correspondence between these analyses is high which allow us to use combined analyses. These analyses in turn allow for the placement of several species which might be otherwise *incertae sedis* in our working classification (Table 1, Supplement Table 1). We also found those morphological characters which correspond most highly with molecular data.

Morphometric characters in group with such a poor morphology should be the most simple to obtain but at the same time assistive in identification, especially on the level of regional floras. They should also reflect general bio-morphological features of species. Our analysis provided several insights in this field. For example, we found repetitive patterns within sections and subgenera (Fig. 8). This is likely is an additional evidence of Vavilov’s homological series, “refrains” (Meyen, 1987). The overall diagnostic power of morphometric characters is week (see below) but within sections or on the regional level they might play the reliable role (Höpke et al., 2019).

We found several most correlated with molecular phylogeny morphological and morphometric characters. The former case is particularly interesting because analysis was performed on the *same* samples and not on higher units like species descriptions. Most notable is the importance of the presence of taproot which is another argument to collect plantains with carefully preserved underground parts (unfortunately this sometimes is not followed). Among binary morphological characters, attention should be paid on the research of seed surface characters (Shipunov, 1998b). with a goal to study as many species as possible.

Producing of identification keys is a complicated task in plantains. These keys must take into account the high variability and overlapping of most distinguishing characters used in *Plantago* taxonomy (Hassemer et al., 2019). Therefore, it might be desirable to employ here results of machine learning techniques such as recursive partitioning. Our partitioning trees (Fig. 10) allow to distinguish section with minimal possible errors on the base of few most informative characters. As the distinguishing power of trees was relatively high, we decided to provide the dichotomous key for sections based on three runs of classification tree with binary characters and one run with morphometric characters. This prototypic key might serve as a framework for the future development of comprehensive keys for the whole group:

1. Ovary with 1–3 ovules and a rudiment of an upper compartment on the adaxial side of the placenta. Corolla lobes longer than 1 mm ... sect. *Virginica*

– Ovary otherwise structured. Corolla lobes short or long ... 2.

2. Non-glandular hairs with joints strongly refracting, walls between cells oblique. Hairs on leaves narrow, less than 0.04 mm ... sect. *Gnaphaloides*

– Strongly refracting joints absent. Hairs on leaves (if present) variable ... 3.

3. Inner side of seed deeply concave ... 13.

– Inner side of seed is not deeply concave ... 4.

4. Ovary with a third compartment at the top on the adaxial side of the placenta, or with a rudiment of it, seen as a thickening at the apex on the posterior side of the ripe placenta. If this compartment absent, then there are few flowers in the inflorescence, no adventitious roots and seeds are longer than 2 mm. Sepals glabrous on the back ... sect. *Mesembrynia*

– Ovary without the third compartment; and/or other characters combinations are different ... 5.

5. Less than 4 flowers per inflorescence. Carpophore present ... sect. *Carpophorae*

– Inflorescence with more than 12 flowers. Carpophore absent ... 6.

6. Posterior sepals with membranaceous, very conspicuous wing on the back. Leaves usually remaining green on drying. Corolla tube hairy. Annuals, leaves often dentate or even dissected ... sect. *Coronopus*

– Posterior sepals without conspicuous wing on the back. Leaves dry differently. Corolla tube hairy or glabrous. Annuals or perennials, leaves with whole margin or sometimes dentate ... 7.

7. Annuals. Anthers usually less than 0.5 mm long ... sect. *Micropsyllium*

– Perennials. Anthers longer than 0.5 mm ... 8.

8. Ovary hairy. Corolla tube hairy. Leaves do not remain green on drying ... sect. *Maritima*

– Ovary glabrous. Corolla tube glabrous. Leaves dry differently ... 9.

9. Anthers white both when fresh and dried ... 10.

– Anthers not white ... 11.

10. Root system mostly of primary and secondary roots. ... sect. *Lamprosantha*

– Root system mostly of adventitious roots ... sect. *Eremopsyllium*

11. Corolla lobes longer than 1.5 mm. Ovary with four or fewer ovules. Leaf width usually less than 25 mm ... sect. *Pacifica*

– Corolla lobes shorter than 1.5 mm. Ovary usually with four or more ovules. Leaf width more than 25 mm ... 12.

12. Anterior sepals distinctly narrower than posterior, and differently shaped ... sect. *Leptostachys*

– Anterior and posterior sepals similar ... sect. *Plantago*

13. Leaves opposite or in whorls of three ... 14.

– Leaves alternate ... 15.

14. Perennials, typically without long glandular hairs. Inner bracts narrow. Seeds longer than 3 mm ... sect. *Arborescens*

– Annuals, with long glandular hairs. Inner bracts broad. Seeds shorter than 3 mm ... sect. *Psyllium*

15. Bract with the upper part scarious, acuminate. Some species with anterior sepals united for more than half of their length ... sect. *Lanceifolia*

– Bract without scarious, acuminate upper part. Anterior sepals always free ... 16.

17. Connective of anther very large, about as long as the pollen sacs. Plants densely hairy (leaf surface hardly visible), cells of non-glandular hairs jointed by a common wall with crown-like elongations ... sect. *Hymenopsyllium*

– Connective of anther smaller. Plants are not so hairy, cells of hairs without crown-shape elongations ... 18.

18. Nerve of anterior sepals well developed. Corolla lobes slightly hairy on the back. The concave inner side of the seed covered by a ragged, white membrane, except for two areas to the right and left of the center. Leaves usually remaining green on drying ... sect. *Albicans*

*–* Nerve of anterior sepals present at base only, distal part scarious. Corolloa lobes not hairy. White membrane on seeds absent. Leaves usually darken on drying. ... sect. *Montana*

**Acknowledgements**

We are grateful to the curators of B, BO, BRIT, C, CAS, COL, CONC, F, HUH, IBSC, JBB, KW, LE, M, MA, MHA, MO, MW, NBG/SAM, NY, PE, PRE, SGO, SP, SPF, TI, TNS, UC/JEPS, US, USM herbarium collections (Thiers, 2019) for the permission to work with their herbarium material, and to Barcoding of Life Consortium and personally to Maria L. Kuzmina (University of Guelph, Guelph, Ontario, Canada) for the help with sequencing North American species of *Plantago*. We thank. Our special thanks to Polina Volkova, Jim Keesling and Gabor Sramkó for the help with *Plantago* samples, and to Ekaterina Shipunova and the staff of the Subdirección Científica, Jardín Botánico de Bogotá José Celestino Mutis (Bogotá, Colombia) for the help with *Aragoa* samples. We thank the Department of Biology of Minot State University for the financial support, and Minot State University students for the help with morphometric observations. From May 2014, this research was supported by North Dakota INBRE. GH also acknowledges *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) for a post-doctoral fellowship (process 153430/2018-4).

**Literature cited**

Andrzejewska‐Golec, E. 1992. A taxonomic study of *Plantago* subgenus *Psyllium* (Miller) Harms. Botanical Journal of the Linnean Society 108: 49–53.

Ashkenazy, H., I. Sela, E. Levy Karin, G. Landan, and T. Pupko. 2018. Multiple sequence alignment averaging improves phylogeny reconstruction. Systematic Biology 68: 117–130.

Balbuena, J.A., R. Míguez-Lozano, and I. Blasco-Costa. 2013. PACo: a novel procrustes application to cophylogenetic analysis. PloS ONE 8: e61048.

Barnéoud, F.-M. 1845. Monographie Générale de la Famille des Plantaginées. Fortin, Masson et Cie, Paris.

Bello, M.A., M.W. Chase, R.G. Olmstead, N.A.H. Rønsted, and D. Albach. 2002. The páramo endemic *Aragoa* is the sister genus of *Plantago* (Plantaginaceae; Lamiales): evidence from plastid *rbc*L and nuclear ribosomal ITS sequence data. Kew Bulletin 57: 585–597.

Bello, M.A., P.J. Rudall, F. González, and J.L. Fernández A. 2004. Floral morphology and development in *Aragoa* (Plantaginaceae) and related members of the order Lamiales. International Journal of Plant Sciences 165: 723–738.

Bergius, P.J. 1768. *Littorella juncea*, en svensk växt. Kongliga Vetenskaps Academiens Handlingar 29: 337–344.

Carlquist, S. 1970. Wood anatomy of insular species of *Plantago* and the problem of raylessness. Bulletin of the Torrey Botanical Club 97: 353–361.

Cho, Y., J.P. Mower, Y.L. Qiu, and J.D. Palmer. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proceedings of the National Academy of Sciences 101: 17741–17746.

Choi, J., H. Lee, and A. Shipunov. 2015. All that is gold does not glitter? Age, taxonomy, and ancient plant DNA quality. PeerJ 3: e1087.

Decaisne, J. 1852. Plantaginaceae Sect. 1 Candolle A. de. Prodromus Systematis Naturalis Regni Vegetabilis 13: 693–737.

Doweld, A.B., and A. Shipunov. 2017. Proposal to reject the name *Plantago indica* (Plantaginaceae). Taxon. 66: 205–206.

Dumortier, B.-C. 1829. Analyse des familles des plantes: avec l'indication des principaux genres qui s'y rattachent. Tournay.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5): 1792–1797.

Fernández A., J.L. 1995. Scrophulariaceae – *Aragoeae*. Flora de Colombia. 16. Bogotá.

Fernández A., J.L. 1993. Novedades taxonómicas en *Aragoa* H.B.K. (Scrophulariaceae) y sinopsis del género. Anales del Jardín Botánico de Madrid 51: 73–96.

Fernández A., J.L. 2002. Algunos patrones de distribución y endemismo en plantas vasculares de los páramos de Colombia. Memorias Congreso Mundial de Páramos, Paipa 1: 213–240.

Funk, V.A., R. Edwards, and K. Sterling. 2018. The problem with (out) vouchers. Taxon 67: 3–5.

Hassemer, G., and N.A.H. Rønsted. 2016. Yet another new species from one of the best-studied neotropical areas: *Plantago humboldtiana* (Plantaginaceae), an extremely narrow endemic new species from a waterfall in southern Brazil. PeerJ 4: e2050.

Hassemer, G. 2018a. Notes on the montane Indo-Iranian species in *Plantago* subgenus *Plantago* (Plantaginaceae). Phytotaxa 336: 59–68.

Hassemer, G. 2018b. Advances to the taxonomic knowledge of *Plantago subulata* (*Plantago* sect. *Maritima*, Plantaginaceae). Turkish Journal of Botany 42: 653–661.

Hassemer, G. 2019a. Novelties and notes on *Plantago* sect. *Virginica* (Plantaginaceae), including the description of a new species and a revised identification key. Webbia 74: 29–41.

Hassemer, G. 2019b. Mediterranean mysteries: notes on *Plantago* sect. *Lancifolia* (Plantaginaceae). Phytotaxa 423: 111–128.

Hassemer, G., S. Bruun-Lund, A. Shipunov, B.G. Briggs, H.M. Meudt, and N.A.H. Rønsted. 2019. The application of high-throughput sequencing for taxonomy: the case of *Plantago* subg. *Plantago* (Plantaginaceae). Molecular Phylogenetics and Evolution 138: 156–173.

Hassemer, G., R. De Giovanni, and R. Trevisan. 2016. The use of potential distribution models in the study of the distribution and conservation status of plants: the case of *Plantago* L. (Plantaginaceae) in Brazil. Journal of the Torrey Botanical Society 143: 38–49.

Hassemer, G., H.M. Meudt, and N.A.H. Rønsted. 2017. Nomenclatural and taxonomic notes on Mediterranean narrow-leaved plantains (*Plantago* section *Maritima*, Plantaginaceae). Webbia 72: 197–205.

Hassemer, G., P. Moroni, and N. O’Leary. 2018b. A nomenclatural revision of *Littorella* (Plantaginaceae, *Plantagineae*). Taxon 67: 1024–1028.

Hassemer, G., A. Shipunov, N.A.H. Rønsted, and H.M. Meudt. 2018a. Taxonomic and geographic novelties in the genus *Plantago* (Plantaginaceae) in Chile, including the description of a new species. Phytotaxa 340: 137–156.

Heibl, C. 2008. IPS: R language tree plotting tools and interfaces to diverse phylogenetic software packages. URL: https://github.com/heibl/ips [last accessed 19 Jun. 2019]

Höpke, J., L. Mucina, and D.C. Albach. 2019. Phylogenetic and morphometric analysis of *Plantago* section *Coronopus* (Plantaginaceae). Taxon 68: 315–339.

Hoggard, R.K., P.J. Kores, M. Molvray, G.D. Hoggard, and D.A. Broughton. 2003. Molecular systematics and biogeography of the amphibious genus *Littorella* (Plantaginaceae). American Journal of Botany 90: 429–435.

Hultén E. 1960. Flora of the Aleutian Islands and westernmost Alaska Peninsula.

Ishikawa, N., J. Yokoyama, and H. Tsukaya. 2009. Molecular evidence of reticulate evolution in the subgenus *Plantago* (Plantaginaceae). American Journal of Botany 96: 1627–1635.

Ishikawa, N., J. Yokoyama, H. Ikeda, E. Takabe, and H. Tsukaya. 2006. Evaluation of morphological and molecular variation in *Plantago asiatica* var. *densiuscula*, with special reference to the systematic treatment of *Plantago asiatica* var. *yakusimensis*. Journal of Plant Research 119: 385–395.

Iwanycki Ahlstrand, N.E., B. Verstraete, G. Hassemer, S. Dunbar-Co, R. Hoggard, H.M. Meudt, and N.A.H. Rønsted. 2019. Ancestral range reconstruction of remote oceanic island species of *Plantago* (Plantaginaceae) reveals differing scales and modes of dispersal. Journal of Biogeography 46: 706–722.

Kunth, K.S. 1818. Nova Genera et Species Plantarum. 3. Paris.

Kuzmina, M. and N. Ivanova. 2011. Primer sets for plants and fungi. URL: http://ccdb.ca/site/wp-content/uploads/2016/09/CCDB\_PrimerSets-Plants.pdf [last accessed 19 Jun. 2019].

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics. 30: 3276–3278.

Li, Z., L. Wei, and R.K. Hoggard. 2011. Plantaginaceae. In Wu, Z., Raven, P.H. and Hong, D. [eds.] Flora of China, vol. 19, 495–503. Science Press, Beijing, and Missouri Botanical Garden, St. Louis.

Linnaeus, C. 1753. Species Plantarum, 112–116. Holmiae.

Linnaeus, C. 1754. Genera Plantarum, 136. Holmiae.

Matsuo, K. 1989. Biosystematic studies on the genus *Plantago*. 1. Variations in *Plantago japonica* and its related species with special reference to its identity. Acta Phytotaxonomica et Geobotanica 40: 37–60.

Meyen, S.V. 1987. Fundamentals of Palaeobotany. Chapman and Hall, London.

Meudt, H.M., B.M. Rojas-Andrés, J.M. Prebble, E. Low, P.J. Garnock-Jones, and D.C. Albach. 2015. Is genome downsizing associated with diversification in polyploid lineages of *Veronica*? Botanical Journal of the Linnean Society 178: 243–266.

Morgan‐Richards, M., and K. Wolff. 1999. Genetic structure and differentiation of *Plantago major* reveals a pair of sympatric sister species. Molecular Ecology 8: 1027–1036.

Nixon, K. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15: 407–414.

Paradis E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289–290.

Peres-Neto, P.R., and D.A. Jackson. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129: 169–178.

Pilger, R. 1937. Plantaginaceae Engler A. Das Pflanzenreich. Hf. 102 (4, 269). Berlin.

Palermo, A.M., A. De Vita, L. Peruzzi, D. Gargano, L. Bernardo, and A. Musacchio. 2010. Does *Plantago brutia* Ten.(Plantaginaceae) merit specific rank? Insights from nrDNA and cpDNA data. Plant Biosystems 144: 573–581.

Preston, J.C., C.M. Ciera, and L.C. Hileman. 2011. Gradual disintegration of the floral symmetry gene network is implicated in the evolution of a wind-pollination syndrome. Proceedings of the National Academy of Sciences 108: 2343–2348.

Rahn, K. 1974. *Plantago* section *Virginica*; a taxonomic revision of a group of American plantains, using experimental, taximetric and classical methods. Dansk botanisk arkiv 30.

Rahn, K. 1996. A phylogenetic study of the Plantaginaceae. Botanical Journal of the Linnean Society 120: 145–198.

Reveal, J.L. 2011. Indices nominum supragenericorum plantarum vascularium. Alphabetical listing by genera of validly published suprageneric names. URL: http://www.plantsystematics.org/reveal/pbio/fam/allspgnames.html [last accessed 19 Jun. 2019]

Ripley, B.D. 1996. Pattern Recognition and Neural Networks. Cambridge.

Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Rønsted, N., M.W. Chase, D.C. Albach, and M.A. Bello. 2002. Phylogenetic relationships within *Plantago* (Plantaginaceae): evidence from nuclear ribosomal ITS and plastid *trn*L-F sequence data. Botanical Journal of the Linnean Society 139: 323–338.

Samarakoon, T., S.Y. Wang, and M.H. Alford. 2013. Enhancing PCR amplification of DNA from recalcitrant plant specimens using a trehalose‐based additive. Applications in Plant Sciences 1: 1200236.

Schliep, K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27: 592–593.

Seifert, K.A., P.W. Crous, and J.C. Frisvad. 2008. Correcting the impact factors of taxonomic journals by appropriate citation of taxonomy (ACT). Persoonia 20: 105.

Shehata, A.A., and M.H.A. Loutfy. 2006. On the taxonomy of Plantaginaceae Juss. sensu lato: evidence from SEM of the seed coat. Turkish Journal of Botany 30 71–84.

Shipunov, A. 1996. On new and rare species of genus *Plantago* L. (Plantaginaceae) from various regions of former USSR. Bulletin of Moscow Society of Naturalists. Biol. Series 102: 67–69.

Shipunov, A. 1997. New data on the distribution of some species *Plantago* L. and *Psyllium* Mill. (Plantaginaceae) in East Europe. Bulletin of Moscow Society of Naturalists. Biol. Series 102: 64.

Shipunov, A. 1998a. Plantains (genera *Plantago* L. and *Psyllium* Mill.) from European Russia and adjacent territories. Ph.D. thesis. Moscow State University, Moscow, Russia.

Shipunov, A. 1998b. The significance of the sculpture of the surface of seeds for systematics of genera *Plantago* L. and *Psyllium* Mill. (Plantaginaceae). Bulletin of Moscow Society of Naturalists. Biol. Series. 103: 41–51.

Shipunov, A. 2000a. On the taxonomic status of *Plantago eocoronopus* Pilger and *P. arachnoidea* Schrenk var. *lorata* Liu. Bulletin of Moscow Society of Naturalists. Biol. Series. 105: 60–64.

Shipunov, A. 2000b. The genera *Plantago* L. and *Psyllium* Mill. (Plantaginaceae Juss.) in the flora of East Europe. Novitates Systematicae Plantarum Vascularium 32: 139–152.

Shipunov, A. 2015. *Plantago schrenkii* is *P. maritima*: morphological and molecular evidence. Annales Botanici Fennici 52: 33–37.

Shipunov, A. 2017. Portion of Plantaginaceae, the Plantain Family: *Littorella* and *Plantago*. Online edition. In: R.F.C. Naczi, J. R. Abbott, and Collaborators, New Manual of Vascular Plants of Northeastern United States and Adjacent Canada, compiled in 2016, 2017. NYBG Press, New York.

Shipunov A. 2019a. *Plantago* and *Littorella*. In: Freeman, C. and Rabeler R. (eds.) Flora of North America. 2019. Volume XVII, 280–293. Oxford University Press, New York and Oxford.

Shipunov, A. 2019b. shipunov: Miscellaneous Functions from Alexey Shipunov. R package version 1.2. URL: https://CRAN.R-project.org/package=shipunov [last accessed 19 Jun. 2019]

Stuessy, T.F., D.J. Crawford, and P. López-Sepúlveda [eds.] 2017. Plants of oceanic islands: evolution, biogeography, and conservation of the flora of the Juan Fernández (Robinson Crusoe) Archipelago. Cambridge University Press.

Tatewaki, M. and Y. Kobayashi. 1934. A contribution to the flora of the Aleutian Islands. Journal of the Faculty of Agriculture, Hokkaido Imperial University 36: 1–119.

Therneau, T., B. Atkinson, and B. Ripley. 2014. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-8. URL: http://CRAN.R-project.org/package=rpart [last accessed 19 Jun. 2019]

Thiers, B. 2019. Index Herbariorum: a global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. URL: http://sweetgum.nybg.org/science/ih [last accessed 19 Jun. 2019]

Venables, W.N. and B.D. Ripley, 2002. Modern Applied Statistics with S. Springer, New York.

**Figure legends**

Figure 1. Phylogeny of the *Plantagineae*: the general arrangement of clades.

Figure 2. The overview of phylogeny tree based on “tall” dataset. Branches with support > 90% thickened.

Figure 3. Density surface of the cophenetic space, based on the “tall” dataset; “ridges” correspond with three major subgeneric divisions of *Plantago*. Note the three-fold structure of “phylogenetic surface”: tallest corresponds with subg. *Plantago*, close and behind is subg. *Coronopus*, subg. *Psyllium* and allies form the rightmost “ridge”.

Figure 4. The phylogeny of *Plantagineae* obtained from the “broad” dataset. Stars (\*) mark species which have not been barcoded before.

Figure 5. The phylogeny of *Plantagineae* obtained from the “tall” dataset. (a) *Aragoa* and *Littorella* (b) *Plantago* subgenera *Coronopus* and *Plantago* (first part); (c) subgenus *Plantago* (second part); (d) subgenus *Psyllium* and allies. Stars (\*) mark species which have not been barcoded before.

Figure 6. Scatterplot of the data points from *Plantago* joint molecular and morphological datasets after Procrustes supeimposition, differences in location of each species designated with arrows.

Figure 7. Example of how to employ machine learning in taxonomy: darker dots correspond with *Plantago* species; these dots positioned on the molecules + morphology kNN prediction plane and therefore define the regions where each point (even without species correspondence) is predicted as a member of one of three major subgenera of *Plantago* (*Plantago*, *Coronopus* and *Psyllium* with allies). Now if any new species correspond with one of these lighter points, its subgeneric placement is predicted.

Figure 8. *Plantago* leaf shapes *vs*. subgenus.

Figure 9. Morphological characters in *Plantago*: (a) morphometric measurements of seven “spot” characters; (b) “molecular weights” (median and maximum average Spearman correlations on 1000 bioostrap replicates) of morphometric characters; (c) “molecular weights” (average correlations with “tall” tree and combined molecular-morphology tree) of binary characters, character abbreviations explained in Support Table 4.

Figure 10. Recursive partitioning of *Plantago* sections with morphological characters, the prototype of diagnostic key: (a) first run and (b) second run on binary morphological characters. Character abbreviations explained in Support Table 4.

Figure 11. *Plantago chihuahuensis* (BRIT)

Figure 12. *Plantago cuzcoensis* (MO)

**Tables**

Table 1 Machine learned placements of the molecularly unsampled species.

**Support Materials**

Support Table 1. Working classification of *Plantagineae*.

Support Table 2. Vouchers of *Plantagineae* samples.

Support Table 3. GenBank accession numbers of *Plantagineae* samples.

Support Table 4. Binary morphological characters used.