

Genetic variation in *Arabidopsis suecica* and its parental species *A. arenosa* and *A. thaliana*

C. LIND-HALLDÉN¹, C. HALLDÉN² and T. SÄLL³

¹ Department of Mathematics and Natural Sciences, Kristianstad University, Kristianstad, Sweden

² Department of Clinical Chemistry, Malmö University Hospital, Malmö, Sweden

³ Department of Genetics, Lund University, Lund, Sweden

Lind-Halldén, C., Halldén, C. and Säll, T. 2002. Genetic variation in *Arabidopsis suecica* and its parental species *A. arenosa* and *A. thaliana*. — *Hereditas* 136: 45–50. Lund, Sweden. ISSN 0018-0661. Received August 28, 2001. Accepted December 20, 2001

Random amplified polymorphic DNA (RAPD) markers were used to estimate the level of genetic variation in Swedish accessions of the allopolyploid *Arabidopsis suecica* and its parental species *A. thaliana* and *A. arenosa*. The results showed clear differences among the three species with respect to the level of variation. *A. arenosa* was highly variable, *A. thaliana* showed a moderate level of variation whereas *A. suecica* was much less variable than the two other species. An extended analysis covering 19 Swedish populations of *A. suecica* corroborated the low level of variation in this species, yet 16 unique phenotypes were observed. No isolation by distance was observed. When the genetic variation was partitioned among and within populations of *A. suecica*, the results showed that the majority of the variation (81 %) occurred among populations. This result is interpreted as a strong indication that *A. suecica* is autogamous in nature.

Torbjörn Säll, Department of Genetics, Lund University, Sölvegatan 29, SE-223 62 Lund, Sweden. E-mail: Torbjorn.Sall@gen.lu.se

Arabidopsis suecica ($2n = 26$) is an allopolyploid, with *A. thaliana* ($2n = 10$) and *A. (Cardaminopsis) arenosa* ($2n = 16$, $2n = 32$) as its parental species (HYLANDER 1957; MUMMENHOF and HURKA 1994, 1995; KAMM et al. 1995; O'KANE et al. 1996). Since *A. thaliana* is the leading model organism in plant genetics, *A. suecica* emerges as a potentially very interesting model plant for polyploid research, particularly in comparison to its parental species (CHEN et al. 1998; LEE and CHEN 2001). *A. suecica* is endemic to the fennoscandian peninsula, with central Sweden and southern Finland as the main range, although isolated populations do occur across the two countries (HULTÉN 1971). In Finland and Sweden, *A. thaliana* occurs predominantly in the south, once again with isolated occurrences elsewhere. Only in Finland is there an extensive overlap between the continuous distributions of *A. thaliana* and *A. suecica*. *A. arenosa* is rare all over Finland, whereas in Sweden, its range more or less coincides with that of *A. suecica*. In this area *A. arenosa* is considerably more common than *A. suecica*. Both *A. suecica* and *A. arenosa* occur on sandy and notably disturbed ground such as roadsides, railways, gravel pits etc. *A. suecica* and to some extent, *A. arenosa*, are sensitive to competition, i.e. if succession is allowed to persist in a locality they will eventually disappear. As a result *A. suecica* has a very patchy distribution where survival in a particular locality is often limited. Both *A. suecica* and *A. arenosa* grow as annuals or short-lived perennials in fennoscandia. Only the tetraploid form of *A.*

arenosa ($2n = 32$) is known to occur in northern Europe.

Several basic biological issues concerning *A. suecica* have not been fully investigated; these include how it reproduces, its population structure and the level of genetic variation. Greenhouse observations of *A. suecica* and its parental species show that *A. suecica* produces large quantities of seed in isolation as does *A. thaliana*, which is autogamous (self-fertilising). This indicates that *A. suecica* is also autogamous, just as *A. thaliana* is. However, observations of self-fertility in a greenhouse cannot be accepted as evidence that the species is autogamous in nature. Nor can the possibility of apomixis be completely ruled out. In contrast, *A. arenosa* appears to be completely self-sterile. According to our observations seed is normally produced in this species only after active cross-pollination (unpublished results).

In the present investigation we compared the general level of variation among the three species using RAPD markers. We also investigated whether the genetic variation within *A. suecica* showed any geographical distribution pattern. Finally we compared the amount of variation within and between populations of *A. suecica*.

MATERIAL AND METHODS

Plant material

A total of 19 populations of *Arabidopsis suecica*, 6 populations of *A. thaliana* and 5 populations of *A.*

arenosa were investigated (Table 1). All accessions were collected in Sweden. This paper reports three separate experiments which used subsets of the populations as noted in Table 1. In all cases single plants were used.

DNA isolation and RAPD analysis

Total DNA was isolated from young leaves using the DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. DNA concentrations were determined after agarose gel electrophoresis by comparison with known amounts of lambda DNA as standard. PCR amplifications were performed in MJ Research PTC-100 thermal cyclers. The reaction mixtures contained 0.8 ng/μl of template DNA, 0.4 μM primer (Operon Technologies, Alameda, USA), 10 mM Tris pH 8.2, 50 mM KCl, 2 mM MgCl₂, 0.1 mM dNTP (Pharmacia) and 0.025 u/μl of Taq DNA

polymerase (Applied Biosystems). The thermal cyclers were programmed for one initial cycle of 2 min at 94°C followed by 40 cycles of 10 s at 94°C, 30 s at 36°C and 1 min at 72°C. This was followed by a final extension period of 2 min at 72°C. The amplified products were resolved on 2.5% agarose gels (SeaKem LE, FMC Bioproducts) containing 0.75 μg/ml EtBr. Gel electrophoresis was run at 5.0 V/cm in 1 × TAE (40 mM Tris-acetate and 1 mM EDTA) for one hour.

Statistical analysis

Data was scored as presence-absence of bands, with band-levels as the basic unit. Bands were not compared between species due to the risk of misinterpretation of comigrating heterologous bands. In addition, due to the dominant nature of the RAPD system, it is not possible to discriminate between

Table 1. Accessions used in the present investigation

Species	Designation	Location	Position		Number ⁺⁺
			N	E	
<i>A. suecica</i>					
	S60	Vännäs	63 55	19 46	1
	S71*+	S Nyåker	63 42	19 21	2
	S80	Nordmaling	63 35	19 28	3
	S90	Västanbäck	63 47	17 05	4
	S110	Ängebo	61 58	16 20	5
	S120	Friggesund	61 54	16 33	6
	S130*	Strömsbruk	61 53	17 19	7
	S140	V Indal	62 36	17 02	8
	S150*	Ytterhogdal	62 10	14 56	9
	S170	Los	61 44	15 10	10
	S181	Road 310/Voxnan	61 41	15 03	11
	S226*+	Högsjö	62 48	17 52	12
	S163	NV Ytterhogdal	62 10	14 53	13
	S260*	Hammarstrand	63 05	16 22	14
	S270	Stadsforsen	62 58	16 40	15
	S290+	Ede	63 04	16 52	16
	S300	Sörfjärda	62 02	17 28	17
	S311	Stocktjärn	63 47	20 12	18
	S330*	Karlstad	59 23	13 27	19
<i>A. arenosa</i>					
	A121	Vännäs	63 55	19 45	–
	A691	Vinslöv	56 06	13 55	–
	A701	Hässleholm	56 10	13 46	–
	A740	NV Ytterhogdal	62 11	14 53	–
	A751	Alnö	62 27	17 25	–
<i>A. thaliana</i>					
	T001	Vänersborg	58 23	12 19	–
	T050	Kristianstad	56 02	14 14	–
	T069	Tottarp	55 57	13 50	–
	T140	Älmtamåla	56 29	15 31	–
	T150	Vimmerby	57 40	15 50	–
	T170	Lund	55 43	13 12	–

* accessions used in the species comparison, + accessions used to partition variation within and among *A. suecica* populations, ++ number used as symbols in Fig. 2.

homozygotes and heterozygotes among the individuals showing a specific band. This means that it is impossible to estimate allele frequencies unless a number of assumptions are made. Owing to this the band pattern was treated as a phenotype and the level of variation was measured by the Shannon information index, $H_o = -\sum p_i \log_2 p_i$ (LEWONTIN 1972) where p_i is the frequency of presence or absence of the RAPD. To quantify differences between pairs of accessions we used 1 minus the commonly used band-sharing index, i.e. $1 - [2n_{ij}/(n_i + n_j)]$ where n_{ij} is the number of shared bands between individuals i and j . n_i and n_j are the total numbers of bands for individuals i and j respectively. In order to statistically test any observed differences we used a resampling (bootstrap) test. For each iteration, random samples were taken from each species respectively and the information index was calculated. Principal component analysis was carried out using the SAS statistical package (SAS Institute Inc. 1990), based on the distances calculated above. The Mantel test (MANTEL 1967) was used to investigate the existence of a correlation between geographical and phenotypic distance. A character incompatibility analysis (MES 1998; VAN DER HULST et al. 2000; WILKINSON 2001) was carried out to investigate the multilocus structure of the *A. suecica* populations. An incompatibility is said to occur for a pair of loci when all four combinations of alleles (i.e. 11, 10, 01 and 00) occur in the data set. The term incompatibility refers to the fact that under clonal propagation and unique mutations all four combinations are cannot exist. An incompatibility is either due to recurrent mutation or to recombination. We calculated the total number of incompatibilities in the data set as well as the expected number of incompatibilities under complete independence between band levels. This was done by randomisation of the phenotypes within each locus.

RESULTS

Comparison of Arabidopsis suecica and its parental species

A single plant from six, five and six populations respectively was selected from *A. suecica*, *A. arenosa* and *A. thaliana* (Table 1). A total of 48 primers were used to amplify RAPD bands from the 17 accessions representing the three species. In *A. suecica*, a total of 218 bands were scored and 174 (80 %) of these were invariable. The level of variation averaged over band levels was $H_o = 0.169$. In *A. thaliana*, 267 bands were observed and of these 110 (41 %) were invariable. The level of variation was $H_o = 0.460$. Finally, in *A. arenosa*, 297 bands were observed and only 56 (19 %)

were invariable. The estimated level of variation was $H_o = 0.671$. A bootstrap analysis of the level of variation shows that all pairwise differences among the species are highly significant, $P < 0.0001$ in all cases. The differences are further illustrated by the pairwise distances. In *A. suecica* the distances varied between 0.074 and 0.01, in *A. thaliana* between 0.232 and 0.141, and in *A. arenosa* the distance varied between 0.516 and 0.281. Thus the largest difference between two *A. suecica* populations is only about half that of the smallest difference observed for *A. thaliana*. Figure 1 shows typical RAPD profiles illustrating the large differences between the three species.

Geographical distribution of the variation in A. suecica

A single plant from each of the 19 *A. suecica* populations was investigated (Table 1). These populations cover the major part of the geographical range of *A. suecica*. A total of 44 primers were used to amplify RAPD products. A total of 228 bands were scored, of which 153 (68 %) were invariable. The level of variation in this extended data set was $H_o = 0.18$. Four of the 19 accessions showed identical phenotypes, all other accessions showed unique phenotypes. In this data set pairwise distances varied between 0.093 and 0. We found a total of 469 incompatibilities using the 75 variable bands. This was significantly lower ($P < 0.0001$) than expected under random permutations of the band patterns (740 incompatibilities). A principal component analysis was performed in order to search for any underlying structure among the populations. The results are shown graphically in Fig. 2. The graph show one tight cluster containing 6 accessions and a tendency to another cluster with four accessions. Finally we tested to see if any association existed between similarity in band phenotype and geographical distance. The mantel test showed no significant association ($P = 0.25$).

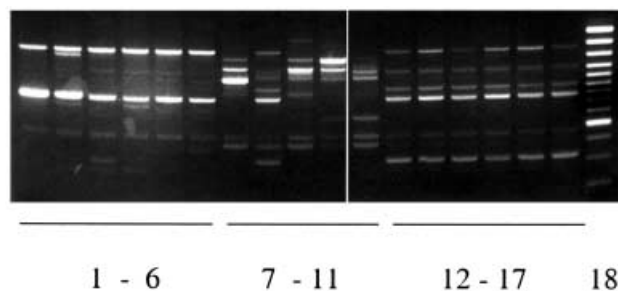


Fig. 1. RAPD analysis of *A. suecica* (lanes 1–6), *A. arenosa* (lanes 7–11) and *A. thaliana* (lanes 12–17).

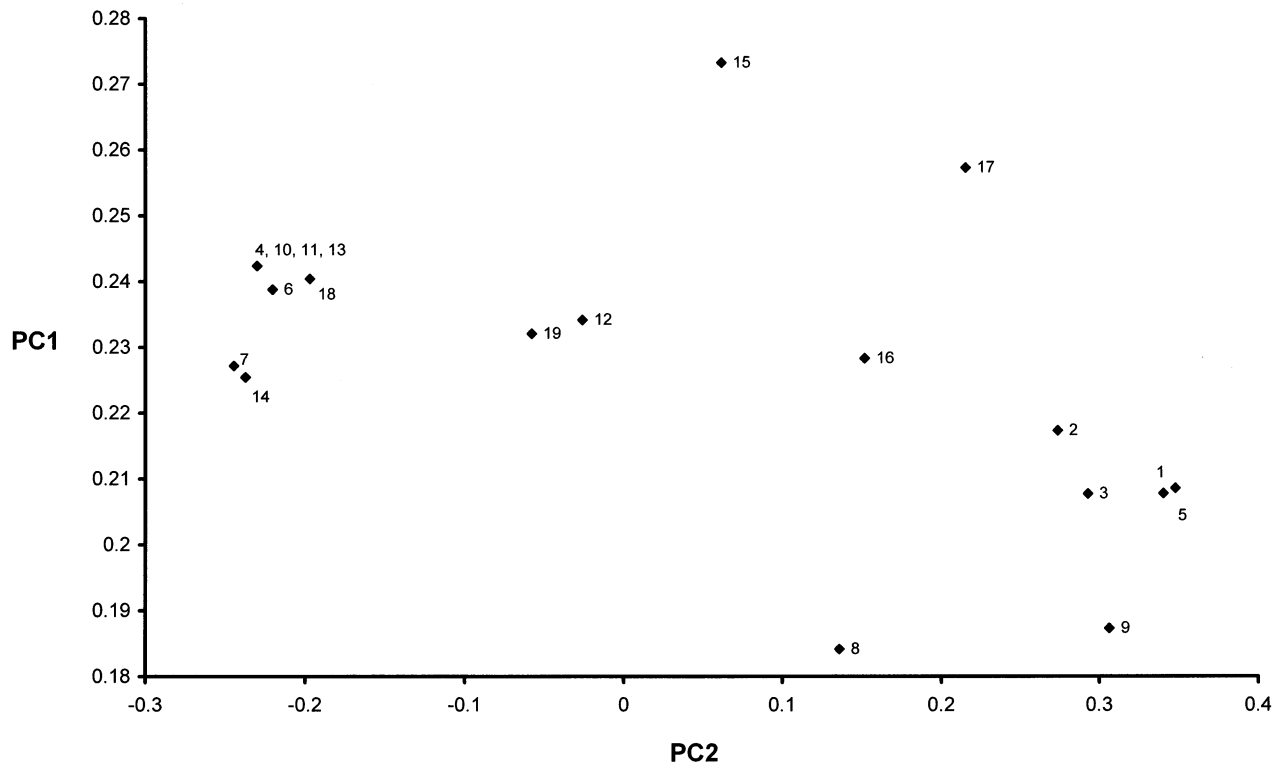


Fig. 2. Principal component analysis of the 19 *A. suecica* accessions. PC1 is the first principal component and PC2 is the second principal component.

Variation within and among populations

Six individuals from each of three *A. suecica* populations were investigated using 43 primers. A total of 230 bands were scored. The number of completely invariable bands was 181 (81%) leaving 49 variable bands. Among the variable bands, 30 were variable only among populations and 19 bands showed variation within populations. Three of the bands were variable in more than one population. The level of variation was 0.024, 0.012 and 0.040 respectively for the three populations and 0.179 for the complete data set. Using the variable bands only, the variation was partitioned among and between the three populations. It was then found that 80.7% of the variation could be attributed to variation among populations.

DISCUSSION

The comparison of the three species rendered very clear results concerning the amount of variation present in the species. The relative level of variation decreased in the order *Arabidopsis arenosa*, *A. thaliana* and *A. suecica*. Comparing our results to RAPD studies in a range of species (e.g. DAWSON et al. 1995; RUSSEL et al. 1997; PRATHEPHA and BAIMAI 1999; OIKI et al. 2001) it was found that the three species cover the range of observed levels of

variation. *A. arenosa* shows a comparatively high level of variation and *A. suecica* is among the least variable species for which comparable RAPD data have been presented. The low level of variation in *A. suecica* can be taken as an indication that the species has a single origin. In particular, this is indicated by the fact that the two most diverse *A. suecica* phenotypes, also among the 19 populations, are clearly closer than the most similar *A. thaliana* populations. This is nothing more than an indication, however, since in the case of multiple origins of an allopolyploid, offspring will eventually meet and reproduce. Nuclear markers will then recombine which will obscure the pattern of origins. In order to settle this question more data are needed and we are at present sequencing chloroplast DNA, using a larger set of accessions of *A. suecica*.

A. thaliana showed a considerable level of variation. Comparable levels of variation were found by MIYASHITA et al. (1999) who investigated 38 ecotypes of *A. thaliana* using AFLP. In that study, as well as in other investigations of genetic variation in *A. thaliana* such as BERGELSON et al. (1998) no association between genetic differentiation and geographical distance could be shown. SHARBEL et al. (2000) found significant correlations when Asian and European ecotypes were compared but could not find

any such pattern within Europe. In the case of complete absence of geographical structure any sample is equally informative about the variation in the species irrespective of geographical origin. Thus, the cited results implicate that our sample of six Swedish populations is expected to show approximately the same level of variation as a sample covering the whole of Europe, which in fact appears to be the case when our results are compared to those of MIYASHITA et al. (1999).

The high level of variation in *A. arenosa* is interesting. *A. arenosa* was first recorded in Sweden in 1840 (FRIES 1843; MALMGREN 1982) and is assumed to have spread since then. HYLANDER (1957) points out that this view is contradictory to the fact that *A. suecica* was reported as early as the 18th century (LINNAEUS 1755; WAHLENBERG 1824). HYLANDER (1957) suggests as one possibility that *A. arenosa* existed in the fennoscandian peninsula in postglacial time when *A. suecica* is assumed to have been formed and then later disappeared from the area. The present population of *A. arenosa*, would then originate from the late introduction in the 19th century. Alternatively *A. suecica* did not originate in fennoscandia but moved there, possibly following the retreating ice. Our results oppose a scenario of a late single narrow introduction. The high level of variation indicates either a continuous inflow of genetic material of varying origins, or that the species existed in central and northern Sweden in isolated localities and then spread along roads and railways in the 19th and 20th centuries.

The more detailed investigations of *A. suecica* corroborated the initial observation of a low level of variation. Just as in *A. thaliana*, (BERGELSON et al. 1998 and MIYASHITA et al. 1999) isolation by distance could not be shown in our sample of *A. suecica* genotypes. This is not surprising given the habitats occupied by *A. suecica*. Species growing in such proximity to man, i.e. roads, railroads and gravel pits, will easily be accidentally transported long distances by human activity, which in turn will often cause populations of different origin to live in proximity to one another. It is mentioned above that observations of *A. suecica* in the greenhouse indicated that it is autogamous, as is *A. thaliana*. The observation that approximately 80% of the variation is distributed among populations clearly shows that *A. suecica* is not an outbreeder in nature. This is actually a stronger differentiation than is usually observed in known selfers, ABBOT and GOMES (1989) found a population differentiation of 0.63 in *A. thaliana*, BERGELSON et al. (1998) found a differentiation of 0.64 and KUTTINEN et al.

(1997) found similar levels in the same species. It should be pointed out that at present, apomixis cannot be completely ruled out. However the observation, that the observed number of incompatibilities constitute 63% of the expectation, renders obligate apomixis very unlikely. At the same time, the fact that the number of incompatibilities were lower than expected at complete randomisation of phenotypes supports the conclusion that *A. suecica* is not a outbreeder.

ACKNOWLEDGEMENTS

We are most thankful to Svante Holm for providing populations S226, S290 and S311, to Håkan Lindström for providing S260 and S270 and to Magnus Nordborg for T150. We would also like to thank Anders Dehlin for help with the field collections and Mattias Jakobsson for technical assistance. This work was supported by the Erik Philip-Sörensen Foundation, the Nilsson-Ehle Foundation and the Crafoord Foundation.

REFERENCES

- Abbot RJ and Gomes MF, (1989). population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L) Heynh. *Heredity* 62: 411–418.
- Bergelson J, Stahl E, Dudek S and Kreitman M, (1998). Genetic variation within and among populations of *Arabidopsis thaliana*. *Genetics* 148: 1311–1323.
- Chen ZJ, Comai L and Picard CS, (1998). Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA gene silencing (nucleolar dominance) in *Arabidopsis* allopolyploids. *Proc. Natl. Acad. Sci.* 95: 14891–14896.
- Dawson IK, Simons AJ, Waugh R and Powell W, (1995). Diversity and genetic differentiation among subpopulations of *Glycerhiza sepium* revealed by PCR-based assays. *Heredity* 74: 10–18.
- Fries E, (1843). *Noviartium florum suecicarum mantissa tertia*. V.–VI. Lundae.
- Hultén E, (1971). Atlas of the distribution of vascular plants in northwestern Europe. Generalstabens litografiska anstalts förlag. Stockholm.
- Hylander N, (1957). *Cardaminopsis suecica* (Fr.) Hiit., a northern amphidiploid species. *Bull. Jardin Botanique Bruxelles* 27: 591–604.
- Kamm A, Galasso I and Schmidt T, (1995). Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationship between *Arabidopsis* species. *Plant Mol. Biol.* 27: 449–566.
- Kuttinen H, Mattila A and Savolainen O, (1997). Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. *Heredity* 79: 144–152.
- Lee HS and Chen ZJ, (2001). Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proc. Natl. Acad. Sci.* 98: 6753–6758.
- Lewontin RC, (1972). The appointment of human diversity. *Evol. Biol.* 6: 381–398.
- Linnaeus C, (1755). *Flora suecica* ed II. Holmiae.
- Malmgren U, (1982). *Västmanlands flora* (Flora of Västmanland). Stockholm.

- Mantel N, (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27: 209–220.
- Mes TM, (1998). Character compatibility of molecular markers to distinguish asexual and sexual reproduction. *Mol. Ecol.* 7: 1719–1727.
- Miyashita NT, Kawabe A and Innan H, (1999). DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. *Genetics* 152: 1723–1731.
- Mummenhof K and Hurka H, (1994). Subunit polypeptide composition of rubisco and the origin of allopolyploid *Arabidopsis suecica* (Brassicaceae). *Biochem. Syst. Ecol.* 22: 807–811.
- Mummenhof K and Hurka H, (1995). Allopolyploid origin of *Arabidopsis suecica* (Fries) Norlin-evidence from chloroplast and nuclear genome markers. *Bot. Acta* 108: 449–456.
- Oiki S, Kawahara T, Inoue K, Ohara M and Maki M, (2001). Random amplified polymorphic DNA (RAPD) variation among populations of the insular endemic plant *Campanula microdonta* (Campanulaceae). *Ann. Bot.* 87: 661–667.
- O’Kane SL, Schaal BA and AlShehbaz IA, (1996). The origins of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences. *Syst. Bot.* 21: 559–566.
- Prathepha P and Baimai V, (1999). Genetic differentiation in Thai populations of the rare species *Afgekia sericea* Carib (Leguminosae) revealed by RAPD-PCR assays. *Genetica* 105: 193–202.
- Russel JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W and Waugh R, (1997). Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theor. Appl. Genet.* 95: 714–722.
- SAS Institute Inc., (1990). SAS user’s guide, 5th ed. SAS Institute Inc., Cary, N.C.
- Sharbel TF, Haubold B and Mitchell-Olds T, (2000). Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol. Ecol.* 9: 2109–2118.
- van der Hulst RGM, Mes THM, den Nijs JCM and Bachmann K, (2000). Amplified fragment length polymorphism (AFLP) markers reveal that population structure of triploid dandelions (*Taraxacum officinale*) exhibits both clonality and recombination. *Mol. Ecol.* 9: 1–8.
- Wahlenberg G, (1824). *Flora suecica* I. Upsaliae.
- Wilkinson M, (2001). PICA 4.0: Software and documentation. Department of Zoology. The natural history museum, London.