Journal of the Linnean Society

Botanical Journal of the Linnean Society, 2015, **177**, 576–592. With 2 figures

A three-marker DNA barcoding approach for ecological studies of xerothermic plants and herbivorous insects from central Europe

WALDEMAR HEISE[1]†, WIESŁAW BABIK[2], DANIEL KUBISZ[3] and ŁUKASZ KAJTOCH[3]*†

[1]*Department of Plant Ecology, Institute of Botany, Jagiellonian University, Krakow, Poland*
[2]*Institute of Environmental Sciences, Jagiellonian University, Krakow, Poland*
[3]*Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Krakow, Poland*

Received 10 August 2014; revised 30 October 2014; accepted for publication 6 January 2015

The DNA barcoding technique developed for species identification has recently been adapted for ecological studies (e.g. host plant identification). Comprehensive barcode databases, covering most species inhabiting areas, habitats or communities of interest are essential for reliable and efficient identification of plants. Here we present a three-barcode (plastid *rbcL* and *matK* genes and the *trnL* intron) database for xerothermic plant species from central Europe. About 85% of the xerothermic plant species (126 out of *c*. 150) known to be associated with xerothermic habitats were collected and barcoded. The database contains barcodes for 117 (*rbcL* and *trnL*) and 96 (*matK*) species. Interspecific nucleotide distances were in the ranges 0–17.9% (0–3.2% within genera) for *rbcL*, 0–44.4% (0–3.1%) for *trnL* and 0–52.5% (0–10.9%) for *matK*. Blast-searching of each sequence in the database against the entire database showed that species-level identification is possible for 89.6% (*rbcL*), 98.4% (*trnL*) and 96.4% (*matK*) of examined plant species. The utility of the presented database for identification of host plants was demonstrated using two insect species associated with xerothermic habitats: the oligophagous leaf-beetle *Cheilotoma musciformis* (for which two host plants in Fabaceae were identified) and the polyphagous weevil *Polydrusus inustus* (which was found to feed on 14 host plants, mostly Rosaceae, Asteraceae and Fabaceae). The developed database will be useful in various applications, including biodiversity, phylogeography, conservation and ecology. © 2015 The Linnean Society of London, *Botanical Journal of the Linnean Society*, 2015, **177**, 576–592.

ADDITIONAL KEYWORDS: calcareous grasslands – Coleoptera – dry grasslands – *matK* – plastid DNA – *rbcL* – *trnL*.

INTRODUCTION

Xerothermic (calcareous) grasslands are one of the most diverse habitats in the temperate zone and are considered to be extrazonal analogues of continental Eurasiatic steppes (Niemelä & Baur, 1998; Poschlod & WallisDeVries, 2002; Ewald, 2003; Dengler *et al*., 2014). This plant formation is highly threatened in Europe (Janišová *et al*., 2011). It is limited by current climatic conditions that favour forests and restrict dry grasslands to local steep, dry and warm slopes on calcareous soils in central and western Europe. Xerothermic grasslands in central Europe sustain highly diverse plant communities, mainly belonging to the *Festuco–Brometea* association (Matuszkiewicz, 2005; Schubert, Hillbig & Klotz, 2001; Chytrý, 2007; Illyés *et al*., 2007; Dúbravková *et al*., 2010). Approximately 150 plant species can be found in this type of vegetation north of the Carpathians. This association is protected by the European Habitats Directive 92/43/EEC, which classifies *Festuco–Brometea* grasslands, occurring mainly on calcareous substrates, under Habitat number 6210. Most xerothermic species are restricted to *Festuco–Brometea* grasslands; only a few

*Corresponding author. E-mail: lukasz.kajtoch@gmail.com
†Both authors contributed equally to this work.

can inhabit other types of habitats (such as sandy turfs). Xerothermic grasslands sustain populations of many rare and relic species with endemic taxa: *Galium cracoviense* Ehrend (only in the Kraków–Częstochowa Uplands), *Erysimum pieninicum* (Zapał.) Pawł. (only in Pieniny Mountains), *Carlina onopordifolia* Besser ex DC. (only in the Polish and Ukrainian Uplands) and several other species annexed in the Habitat Directive of the European Union.

Xerothermic grasslands have been highly fragmented and degraded due to man-made land transformations, which reduced their area as a result of afforestation and agricultural development (Pärtel, Mandla & Zobel, 1999; Dutoit *et al*., 2003; Poschlod *et al*., 2005; Johansson *et al*., 2008). This kind of plant formation is often vulnerable to plant succession (it can become overgrown by herbs, bushes and trees) and in many areas was sustained by traditional land use, mainly extensive grazing by roaming flocks of sheep in spring and autumn combined with summer haymaking (Michalik & Zarzycki, 1995; WallisDeVries, Poschlod & Willems, 2002). Xerothermic grasslands are also characterized by a rich entomofauna, particularly diverse assemblages of Orthoptera, butterflies (Lepidoptera) and beetles (Coleoptera) (Liana, 1987; Mazur, 2001; Rákosy & Varga, 2006; Mazur & Kubisz, 2013). Ecological studies on xerothermic plants and their insect assemblages require the development of techniques that allow for reliable and rapid species identification (both plants and insects). The DNA barcoding approach should facilitate not only identification of particular plant and insect species, but also understanding of ecological interactions and associations between host plants and insects feeding on these plants. Such knowledge would also be of practical importance for conservation of particular species and whole assemblages and for management planning for xerothermic grasslands.

DNA barcoding was developed primarily as an auxiliary technique for species identification. It was first used in animals and was based on a mitochondrial gene, cytochrome oxidase unit I (COI; Hebert, Ratnasingham & deWaard, 2003). Later, this technique was also adapted for studies on fungi with the final choice of the internal transcribed spacers (ITS) of nuclear ribosomal DNA (Seifert, 2009). When considering plants, a long-term debate ensued about the barcode of choice: several DNA markers were proposed for land plants, either individually or in combinations (Chase *et al*., 2007; Kress & Erickson, 2007; Fazekas *et al*., 2008; Hollingsworth *et al*., 2009). Finally, a two-locus barcode was proposed and widely accepted consisting of the plastid genes ribulose-bisphosphate carboxylase (*rbcL*) and maturase K (*matK*) (CBOL Plant Working Group, 2009). Additionally, the *trnH-psbA* intergenic spacer region of plastid DNA was proposed as a plant barcode (Shaw *et al*., 2005; Fazekas, Steeves & Newmaster, 2010; Pang, Luo & Sun, 2012). However, this raised concerns due to its extensive length variation (Chase *et al*., 2007; Kress & Erickson, 2007), the presence of intraspecific microinversions associated with palindromes (Whitlock, Hale & Groff, 2010; Jeanson, Labat & Little, 2011) and sequencing problems related to mononucleotide repeats (Fazekas *et al*., 2008; Devey, Chase & Clarkson, 2009; but see Fazekas *et al*., 2010). In some situations, however, these standard plant barcodes cannot be used. For example, the identification of host plant species from animal gut contents is a difficult task due to DNA degradation (e.g. Wallinger *et al*., 2013). Moreover, primers for *matK* rarely cover a wide spectrum of plant taxonomic units and therefore have limited utility for host plant identification from polyphagous animal guts, as several primer pairs should be used to increase the probability of amplification for all or most host plants present in samples. As an alternative, a plastid intron, located in the tRNALeu UAA gene (*trnL*; Taberlet *et al*., 1991), has successfully been used for diet analyses (Valentini *et al*., 2009; Taberlet *et al*., 2007). This intron has some limitations similar to those of *trnH-psbA* (e.g. length variation) and therefore its utility for plant species identification could be questionable. Nevertheless, it proved to be the barcode of choice for host plant barcoding in insects, particularly beetles (Jurado-Rivera *et al*., 2009; Pinzón-Navarro *et al*., 2010; Kubisz *et al*., 2012; Kitson *et al*., 2013). The *trnL* intron has also been successfully used for identification of below-ground plant richness (from roots) (Hiiesalu *et al*., 2012).

Recently, the DNA barcoding approach has been used for other types of ecological studies, particularly for identification of plant species and evaluation of species richness from selected areas, habitats and/or plant communities. These studies focused on tropical biodiversity hotspots such as forests of South and Central America and South Asia (Kembel & Hubbell, 2006; Dick & Kress, 2009; Gonzalez *et al*., 2009; Kress *et al*., 2009, 2010; Pei *et al*., 2011). So far, there have been several examples of studies using plant barcodes for ecological studies in other areas and plant communities, e.g. boreal forests in Canada (Fazekas *et al*., 2008). However, there are hardly any analogous studies concerning plant species identification and evaluation of species richness for open land habitats such as grasslands, with the exception of a single study on the mountain dry grasslands of Italy (De Mattia *et al*., 2012). One may ask why one would develop barcodes if plants can be identified on the basis of traditional morphological examination. Indeed, there is no need for barcoding in many botanical studies (e.g. in standard vegetation inventories), but barcode databases could potentially be useful if

species identification is difficult (e.g. for cryptic species, fragments of plants without diagnostic characters) or for ecological studies with large numbers of taxa and dealing with interactions among various plants and herbivorous animals.

In the present study, we evaluated the performance of different DNA barcode markers (*matK*, *rbcL* and *trnL*) for identification of xerothermic plant species and evaluation of species richness using xerothermic grasslands from Poland as an example. Xerothermic grasslands in Poland were selected as the subject of this research as this plant association has been intensively studied by botanists and habitat specialists from the end of the 19th century (Preuss, 1912; Kozłowska, 1931; Ceynowa, 1968; Medwecka-Kornaś & Kornaś, 1977). In Poland, all major types of dry grasslands known from central Europe can be found and most central European plant species associated with this vegetation are also present there (Zając & Zając, 2001; Mirek *et al*., 2002; Matuszkiewicz, 2005). Moreover, Polish dry grasslands are located in two areas which differ with respect to the history of formation and persistence of xerothermic grasslands. Southern Poland was glaciated only once (Sanian glaciation, *c*. 730 000–430 000 years ago), whereas northern Poland was glaciated several times during the Pleistocene (including the Vistulian glaciation, which ended 10 000–12 000 or 17 000–18 000 years ago in the Kujawy basin) (Marks, 2002; Lindner *et al*., 2006; Wysota, Molewski & Sokolowski, 2009). Moreover, southern Poland was, and partially still is, connected with the Pontic and Pannonic steppe areas, whereas northern Poland could have been settled by xerothermic species in the Holocene and only via some specific routes (such as along the Vistula River valley). Lastly, xerothermic grasslands in Poland are highly threatened as they are extrazonal, highly fragmented and sensitive to human land transformations. This plant association shelters also diverse communities of invertebrates, including numerous species of Coleoptera. As the diet of some of xerothermic beetles has been intensively studied based on field observations or feeding experiments (e.g. Szymczakowski, 1960; Warchałowski, 1991; Mazur, 2001), they can be used as excellent objects to test performance of plant barcodes for host plant identification. Among xerothermic beetles, well known regarding their feeding preferences are, the oligophagous leaf-beetle *Cheilotoma musciformis* and the polyphagous weevil *Polydrusus inustus*.

Evaluation of the performance of barcodes for identification of xerothermic plant species and evaluation of species richness was performed in four steps: (1) amplification efficiency; (2) sequencing success; (3) accuracy of plant species identification; and (4) application for host plant identification. The main goal of this study was to develop a database of xerothermic plant barcodes for further ecological and conservation studies. Additionally, the database was used for evaluation of the utility of these barcodes for identification of insect host plants on the basis of gut content. To this end we examined two beetles: *C. musciformis* and *P. inustus*

MATERIAL AND METHODS

Sampling area

The study was performed on xerothermic (calcareous) grasslands of the *Festuco–Brometea* association located in two areas. The majority of plants were collected in the Polish Uplands located in southern Poland (between the cities of Kraków and Kielce; coordinates of the centre of this area 50.374°N, 20.407°E). The remaining plants, especially species absent or difficult to find or rare in southern Poland, were collected in northern Poland in the Kujawy Basin (between the cities of Toruń and Bydgoszcz; coordinates of the centre of this area 52.942°N, 18.572°E). Xerothermic communities in the first sampling area consisted mainly of xerothermic grasslands on steep slopes of chalk and gypsum hills. In the second area, mainly xerothermic grasslands on steep scarps along river valleys on clay soils were sampled.

Plant sampling

Prior to field surveys, a list of all plant species native to Poland and associated exclusively or mainly with xerothermic grasslands (Zając & Zając, 2001; Mirek *et al*., 2002; Matuszkiewicz, 2005) was compiled. After floristic reconnaissance, we also added species commonly found in xeric grasslands, but strongly associated with other syntaxonomic groups (mostly species associated with *Molinio–Arrhenatheretea* meadows and *Rhamno–Prunetea* shrubland). The final list comprised 152 plant species. Field survey was executed in two seasons in 2011 and 2012 (from April to August). Xerothermic plant species and other species characteristic for open dry habitats were collected. Voucher specimens (dried) were collected and are deposited in the Jagiellonian University Herbarium (collector: W. Heise) (voucher specimen numbers presented in Table 1). For the purposes of molecular analyses several green leaves from a single individual of each species were collected and preserved in plastic bags with silica gel. All samples were stored in a refrigerator at 4 °C until DNA isolation. Plant species were identified in the field. Parts of specimens important for taxonomic identification were collected and preserved.

Beetle sampling

To evaluate the utility of plant barcodes for host plant identification from insect gut two species were

Table 1. Xerothermic plant species from Poland analysed in this study with barcoding success of three plant barcodes

No.	Species – morphology	*trnL*	*rbcL*	*matK*	No.	Species – morphology	*trnL*	*rbcL*	*matK*	No.	Species – morphology	*trnL*	*rbcL*	*matK*
Fagaceae Dumort.					Caryophyllaceae Juss.					Fabaceae Lindl.				
XT140	*Quercus sessilis* Ehrh.	PF	PF	PF	XT104	*Arenaria serpyllifolia* L.	PF	PF	PF	XT28	*Anthyllis vulneraria* L.	KJ746348	KJ746221	KJ746152
Asteraceae Dum.					XT9	*Cerastium arvense* L.	KJ746431	KJ746295	SE	XT126	*Astragalus arenarius* L.	KJ746334	KJ746228	KJ746136
XT80	*Achillea millefolium* L.	KJ746372	KJ746261	KJ746172	XT58	*Dianthus carthusianorum* L.	KJ746429	KJ746300	SE	XT78	*A. cicer* L.	KJ746332	KJ746227	KJ746137
XT56	*Artemisia campestris* L.	KJ746373	KJ746262	WQ	XT143	*D. deltoides* L.	PF	PF	PF	XT94	*A. danicus* Retz.	KJ746333	KJ746229	KJ746137
XT105	*Aster amellus* L.	KJ746383	KJ746255	KJ746181	XT117	*Petrorhagia prolifera* (L.) P.W.Ball & Heywood	KJ746430	KJ7462989	KJ746192	XT21	*Coronilla varia* L.	KJ746349	KJ746222	seq error
XTP69	*Carlina acaulis* L.	KJ746380	KJ746258	SE	XT50	*Silene nutans* L.	KJ746426	KJ746296	KJ746195	XT110	*Genista tinctoria* L.	KJ746351	KJ746223	KJ746149
XT46	*C. onopordifolia* Besser ex DC.	KJ746379	WQ	KJ746179	XT53	*S. otites* Sm.	KJ746427	KJ746297	KJ746193	XT99	*Lathyrus tuberosus* L.	KJ746341	KJ746230	KJ746146
XT68	*Centaurea stoebe* L.	KJ746376	KJ746252	KJ746178	XT81	*S. vulgaris* (Moench) Garcke	KJ746428	KJ746298	KJ746194	XT15	*Lotus corniculatus* L.	KJ746347	KJ746220	KJ746151
XT5	*C. scabiosa* L.	KJ746375	KJ746251	KJ746175	XTGB	*Arenaria serpyllifolia* L.	FJ404972	HQ589962	HQ593179	XT82	*Medicago falcata* L.	KJ746337	KJ746238	KJ746143
XT29	*Cirsium pannonicum* Link	KJ746377	KJ746253	KJ746176	Plumbaginaceae Juss.					XT73	*M. lupulina* L.	KJ746338	KJ746240	KJ746147
XT109	*Erigeron acris* L.	KJ746384	KJ746254	KJ746180	XT128	*Armeria maritima* (Mill.) Willd.	KJ746423	KJ746293	KJ746191	XT127	*M. minima* (L.) L. ex Bartal.	KJ746336	KJ746239	KJ746148
XT112	*Helichrysum arenarium* Moench	KJ746381	WQ	KJ746174	Polygonaceae Juss.					XT57	*M. varia* Martyn	KJ746339	KJ746237	SE
XT113	*Hieracium pilosella* L.	KJ746385	KJ746256	KJ746177	XT119	*Rumex acetosella* L.	KJ746422	KJ746306	KJ746190	XT115	*Melilotus albus* Medik.	KJ746436	KJ746235	KJ746141
XT47	*Inula ensifolia* L.	KJ746382	KJ746257	PF	Brassicaceae Burnett					XT51	*M. officinalis* (L.) Lam.	KJ746435	KJ746234	KJ746142
XT118	*Picris hieracioides* L.	KJ746378	KJ746259	KJ746182	GB	*Lepidium campestre* (L.) W. T. Aiton	AF055265	HQ590157	HQ593342	XT45	*Onobrychis viciifolia* Scop.	KJ746331	KJ746267	KJ746134
XT24	*Chrysanthemum corymbosum* L.	KJ746374	KJ746260	KJ746173	GB	*Erophila verna* (L.) DC.	FJ490778	KF724306	HQ619804	XT93	*Ononis spinosa* L.	KJ746340	KJ746236	KJ746144
Campanulaceae Juss.					XT77	*Berteroa incana* DC.	KJ746425	KJ746307	KJ746208	XT101	*Oxytropis pilosa* DC.	KJ746335	KJ746226	KJ746135
XT108	*Campanula glomerata* L.	KJ746389	KJ746315	KJ746183	XT83	*Sisymbrium loeselii* L.	KJ746424	KJ746308	SE	XT141	*Robinia pseudoacacia* L.	PF	PF	PF
XT95	*C. sibirica* L.	KJ746388	KJ746314	WQ	Hypericaceae Juss.					XT131	*Sarothamnus scoparius* (L.) Wimm. ex W.D.J.Koch	KJ746350	KJ746224	KJ746150

Table 1. *Continued*

No.	Species – morphology	*trnL*	*rbcL*	*matK*	No.	Species – morphology	*trnL*	*rbcL*	*matK*	No.	Species – morphology	*trnL*	*rbcL*	*matK*
Boraginaceae Juss.					XT88	*Hypericum perforatum* L.	KJ746420	KJ746321	SE	XT121	*Trifolium alpestre* L.	KJ746346	KJ746233	KJ746140
XT41	*Cerinthe minor* L.	WQ	WQ	PF	Cistaceae Juss.					XT122	*T. arvense* L.	KJ746347	KJ746232	KJ746139
XT90	*Echium vulgare* L.	KJ746415	KJ746283	KJ746184	XT98	*Helianthemum nummularium* Mill.	PF	PF	PF	XT18	*Vicia tenuifolia* Roth	KJ746344	KJ746231	KJ746145
Dipsacaceae Juss.					Euphorbiaceae Juss.					Crassulaceae DC. in Lam. & DC.				
XT62	*Knautia arvensis* Coult.	KJ746386	KJ746249	SE	XT14	*Euphorbia cyparissias* L.	KJ746421	KJ746288	SE	XT55	*Sedum acre* L.	KJ746396	KJ746264	SE
XT124	*Scabiosa ochroleuca* L.	KJ746387	KJ746250	SE	XT85	*E. esula* L.	KJ746419	KJ746287	KJ746207	XT17	*S. maximum* Suter	KJ746395	KJ746266	KJ746205
Asclepiadaceae R.Br.					Primulaceae Vent.					XT86	*S. rupestre* L.	KJ746394	KJ746263	KJ746204
XT133	*Vincetoxicum hirundinaria* Medik	KJ746392	KJ746281	KJ746185	XT13	*Primula veris* L.	KJ746433	KJ746292	SE	XT87	*S. sexangulare* L.	KJ746397	KJ746265	KJ746206
Gentianaceae Juss.					Berberidaceae Juss.					Apiaceae Lindl				
XT111	*Gentiana cruciata* L.	KJ746393	KJ746282	KJ746189	XT52	*Berberis vulgaris* L.	KJ746401	KJ746305	KJ746120	XT63	*Daucus carota* L.	PF	PF	PF
Rubiaceae Juss.					Ranunculaceae Juss.					XT97	*Eryngium planum* L.	KJ746364	KJ746241	KJ746166
XT2	*Galium mollugo* L.	KJ746398	KJ746284	KJ746186	XT43	*Adonis vernalis* L.	KJ746403	KJ746304	KJ746121	XT26	*Laserpitium latifolium* L.	KJ746368	KJ746247	KJ746171
XT30	*G. valdepilosum* Heinr.Braun	KJ746400	KJ746286	KJ746188	XT40	*Anemone sylvestris* L.	KJ746402	KJ746302	SE	XT114	*Seseli libanotis* W.D.J.Koch	KJ746367	KJ746246	SE
XT1	*G. verum* L.	KJ746399	KJ746285	KJ746187	XT144	*Ranunculus acris L.*	KJ746405	KJ746301	KJ746123	XT72	*Pastinaca sativa* L.	KJ746370	KJ746244	KJ746169
Lamiaceae Lindl.					XT61	*Thalictrum minus* L.	KJ746404	KJ746303	KJ746122	XT3	*Peucedanum cervaria* Cusson ex Lapeyr.	KJ746365	KJ746242	KJ746167
XT132	*Acinos arvensis* (Lam.) Dandy	WQ	WQ	PF	Santalaceae R.Br.					XT4	*P. oreoselinum* Moench	KJ746366	KJ746243	KJ746168
XT107	*Stachys officinalis* (L.) Trevis.	KJ746356	KJ746272	KJ746158	XT34	*Thesium linophyllon* L.	WQ	KJ746291	KJ746196	XT91	*Pimpinella saxifraga* L.	KJ746371	KJ746248	KJ746170
XT102	*Prunella grandiflora* Jacq.	WQ	WQ	PF	Onagraceae Juss.					XT66	*Seseli annuum* L.	KJ746369	KJ746245	SE
XT6	*Salvia pratensis* L.	KJ746359	KJ746268	KJ746154	XT116	*Oenothera biennis* L.	PF	PF	PF	Poaceae (R. Br.) Barnh.				
XT38	*Stachys recta* L.	KJ746355	KJ746271	WQ	Linaceae DC. ex Gray					XT75	*Brachypodium pinnatum* (L.) P. Beauv.	KJ746414	KJ746320	KJ746200

XT103	*Teucrium chamaedrys* L.	KJ746354	WQ	KJ746159	XT27	*Linum flavum* L.	KJ746417	KJ746289	KJ746202	XT145	*Elymus repens* (L.) Gould	KJ746409	KJ746316	KJ746199
XT36	*Thymus pannonicus* All.	KJ746360	KJ746269	KJ746155	XT100	*L. hirsutum* L.	KJ746418	KJ746290	KJ746203	XT16	*Festuca rupicola* Heuff.	KJ746412	KJ746319	KJ746197
XT120	*T. pulegioides* L.	KJ746361	KJ746270	SE	Polygalaceae R. Br. in Flinders					XT32	*Koeleria macrantha* (Ledeb.) Schult.	KJ746411	KJ746317	WQ
GB	*Clinopodium vulgare* L.	AY506593	HQ590041	HQ593243	XT11	*Polygala comosa* Schkuhr	KJ746416	KJ746225	KJ746153	XT12	*Poa pratensis* L.	KJ746410	KJ746318	KJ746198
Plantaginaceae Juss.					Rosaceae Juss.					XT59	*Stipa Joannis* Čelak.	KJ746413	WQ	KJ746201
XT22	*Plantago lanceolata* L.	KJ746391	KJ746278	KJ746164	XT23	*Agrimonia eupatoria* L.	WQ	KJ746217	KJ746132	GB	*Bromus erectus* Huds.	AY367957	FR865129	JN895533
XT33	*P. media* L.	KJ746390	KJ746277	KJ746165	XT76	*Crataegus monogyna* Jacq.	KJ746323	KJ746209	KJ746124	GB	*Anthoxanthum odoratum* L.	JF428083	HE963320	HE966877
Scrophulariaceae Juss.					XT7	*Filipendula vulgaris* Moench	KJ746330	KJ746219	KJ746133	GB	*Elymus hispidus* (Opiz) Melderis	DQ914536	KP728573	KP277165
XT84	*Linaria vulgaris* Mill.	KJ746352	KJ746274	KJ746156	XT8	*Fragaria viridis* Weston	WQ	KJ746212	PF	Cyperaceae Juss.				
XT42	*Melampyrum arvense* L.	KJ746357	KJ746279	KJ746163	XT129	*Potentilla alba* L.	KJ746329	KJ746216	KJ746131	XT37	*Carex humilis* Leyss.	KJ746432	KJ746314	WQ
XT92	*Orthantha lutea* A.Kern. ex Wettst.	KJ746358	KJ746280	KJ746162	XT19	*P. incana* G.Gaertn., B.Mey. & Scherb.	WQ	KJ746215	KJ746130	Liliaceae Juss.				
XT123	*Verbascum lychnitis* L.	KJ746353	KJ746273	KJ746157	XT54	*P. argentea* L.	KJ746328	KJ746214	KJ746129	XT74	*Allium montanum* F.W.Schmidt ex Schult f.	WQ	KJ746311	KJ746117
XT10	*Veronica chamaedrys* L.	KJ746362	KJ746275	KJ746160	XT147	*Prunus spinosa* L.	KJ746324	KJ746210	KJ746125	XT65	*Anthericum ramosum* L.	KJ746407	KJ746310	KJ746118
XT70	*V. spicata* Christa.	KJ746363	KJ746276	KJ746161	XT79	*Rosa canina* L.	KJ746327	KJ746213	KJ746128	XT49	*Asparagus officinalis* L.	KJ746406	KJ746309	KJ746116
Celastraceae R. Br. in Flinders					XT44	*Sanguisorba minor* Scop.	KJ746326	KJ746218	KJ746127	Orchidaceae Juss.				
GB	*Herniaria glabra* L.	JN589730	HE963469	HE966915	XT146	*S. officinalis* L.	KJ746325	KJ746211	KJ746126	XT31	*Cypripedium calceolus* L.	KJ746408	KJ746312	KJ746119
GB	*Euonymus europaeus* L.	HQ393747	HE963469	JN895303	Cuscutaceae Dumort.					Betulaceae Gray				
Violaceae Batsch					XT64	*Cuscuta epithymum* L.	KJ746434	KJ746322	SE	GB	*Corylus avellana* L.	JF767170	FR865127	FR865049
GB	*Viola hirta* L.	JF767170	FR865127	FR865049										

GB, sequences from the NCBI GenBank resources (with accession numbers provided for each barcode); PF, PCR fail = ampliciation (PCR) failure; SE, sequencing errors; WQ, worse quality = sequences of worse quality, not deposited in GB; XT & number, numbers of voucher specimen in xerothermic (XT) plant collection in Jagiellonian University Herbarium.

selected: the leaf-beetle *C. musciformis* (Chrysomelidae) and the weevil *P. inustus* (Curculionidae). Both species are characteristic of dry grasslands and scrublands of central and eastern Europe (Warchałowski, 1971; Borowiec, 1984; Mazur, 1994; Korotyaev & Meleshko, 1995; Korotyaev, 1996; Mazur & Kubisz, 2013). The population genetics of both these species have recently been studied in detail (see Kajtoch, LachowskaCierlik & Mazur, 2009; Kajtoch, Korotyaev & LachowskaCierlik, 2012; Kajtoch *et al*., 2013). Beetles were collected using sweep-nets from herb, shrub and bush layers on xerothermic turfs in 2011 and 2012 (May–June). To avoid over-representation of specimens feeding on the same plants (collected in the same place and the same time), 24 specimens of *P. inustus* were randomly selected, each from a different xerothermic patch. Similarly, single individuals of *C. musciformis* were randomly selected from distinct xerothermic patches; only ten specimens were used in analyses, as this species is highly threatened in Poland (Ścibior, 2004; Kajtoch *et al*., 2013). Beetles were only collected in good weather conditions to avoid collection of starving specimens (as efficiency of plant DNA isolation and amplification from such individuals is decreased; Kajtoch & Mazur, in press) and preserved immediately in ethanol (96%) in the field to reduce DNA degradation. Samples were kept frozen until DNA isolation.

Laboratory procedure

Plant tissues (leaves) were frozen in liquid nitrogen prior to DNA isolation. Frozen samples were crushed (homogenized, pulverized) in an agate mortar, and DNA was isolated using the Nucleospine Plant Tissue Kit (Macherey-Nagel). Beetles were digested with proteinase K, and DNA was isolated using the Nucleospine Tissue Kit and protocol for animal tissue isolation. The DNA concentration and purity of all isolates were assessed using Nanodrop, and the quality of DNA isolates from beetles was checked by amplification of the COI mitochondrial gene using standard barcode primers (Folmer *et al*., 1994). Next, DNA isolates were used for amplification of three plastid barcodes, *matK*, *rbcL* and *trnL*, using the following primers: matK472F and matK1248R for *matK* (Yu, Xue & Zhou, 2011), 1F and 724R for *rbcL* (Fay, Swensen & Chase, 1997), and A49325 and B49863 for *trnL* (Taberlet *et al*., 1991). We did not use primers developed to amplify short barcodes [minibarcodes; e.g. Hofreiter *et al*. (2000) for *rbcL*; Taberlet *et al*. (2007) for *trnL*] as these short markers do not have sufficient discriminatory power and rarely allow for plant species identification (see also Little, 2014). Amplicons of the *trnL* intron were of variable length (*c*. 350–640 bp), whereas amplicons of the plastid genes showed a smaller range of length variation: *rbcL*, 650–680 bp; *matK*, 690–720 bp. The PCRs of samples that did not amplify any fragment were repeated using less stringent conditions: reduction of up to 5 °C in the annealing temperature and a higher concentration of $MgCl_2$. For species for which this procedure failed to amplify any barcode, the PCRs were repeated on other DNA isolates. The same primers were used for amplification of plant DNA from plant tissues (leaves) and from insect guts. All PCR products were visualized on agarose gels. PCR products from plant leaves and *C. musciformis* samples were then purified using an ExoProStar kit (GE Chemicals). Purified DNA products were then Sanger sequenced using forward primers and a BigDye Terminator v.3.1. Cycle Sequencing Kit (Applied Biosystems) and run on an ABI 3100 Automated Capillary DNA Sequencer. In cases of unreadable sequences, the sequencing procedure was repeated with reverse primers. For *P. inustus*, another procedure of host plant identification was used: only *rbcL* and *trnL* barcodes were amplified separately for each individual (to avoid problems and errors caused by unequal concentration of plant DNA in isolates from weevil bodies). This procedure was followed because the *matK* database of xerothermic plants was too incomplete for reliable species assignment (see Results). All amplicons (small volumes of both *rbcL* and *trnL*) were first checked on agarose gel and then pooled approximately equimolarly (all PCRs of *rbcL* separately from PCRs of *trnL*) and purified using a Nucleospine DNA Extraction Kit. The sequencing library was prepared using a NexteraXT library preparation kit (Illumina). The library was sequenced as a part of a MiSeq paired-end 2× 150-bp run.

Data analysis

Sanger sequences were checked visually using BioEdit v.7.0.5.2 (Hall, 1999). Only sequences of good-quality fragments, longer than 400 bp (*trnL*) or 650 bp (*rbcL* and *matK*), were used for further analysis. Sequences of all three plant barcodes used in this study and obtained directly from plant tissues were stored as FASTA files. All sequences of the particular barcode were aligned using MAFFT v.7 (Katoh & Standley, 2013). Because the generated database of xerothermic plants does not cover all species known from the study area (see Results), the NCBI GenBank database was additionally searched for *rbcL*, *trnL* and *matK* sequences of xerothermic plant species missing in the xerothermic database (see Table 1).

Although the CBOL Plant Working Group has initiated a plant DNA barcoding database based on *rbcL*

and *matK* (see http://www.boldsystems.org), it currently contains an insufficient number of records, especially for taxa from poorly known environments and areas such as xerothermic grasslands of central Europe and therefore this database was not sufficient for the purposes of this study. Moreover, this database contains only *rbcL* and *matK* sequences; therefore, the *trnL* barcode cannot be used for species identification using BOLD. For these reasons, instead of using BOLD we decided to use the resources available in NCBI GenBank. MEGABLAST (Basic Local Alignment Search Tool, Altschul *et al*., 1990) was used to search for most similar sequences of three barcodes (independently) in the NCBI GenBank sequence library. Results of identification were provided as a list of best hits of the nearest matches (maximum identity) according to BOLD-IDS guidelines (http://www.boldsystems.org/views/idrequest_plants.php).

Due to the limitation of NCBI GenBank resources, it was not possible in some cases to identify plant species that were barcoded (as many xerothermic plants were absent in NCBI GenBank before this study); therefore, other species (usually of the same genus) were retrieved and reported as the nearest matches. This was done only for quick verification of barcode amplification and sequencing efficiency and accuracy. The performance of each barcode was evaluated by use of a local Blast search in BioEdit v.7.2.2 (Hall, 1999) of the developed barcode database against this database to find how many plant species could not be discriminated. Only hits with 100% identity and > 95% sequence coverage were retrieved. In the local Blast search we used 128 sequences for *trnL* and *rbcL* barcodes and 107 sequences for the *matK* barcode (including plant species for which sequences were downloaded from NCBI GenBank). Moreover, according to the guidelines provided by CBOL (http://www.barcoding.si.edu/protocols.html), the evaluation of comparative levels of variation and discrimination for the three markers were undertaken using MEGA 5.10 (Tamura *et al*., 2011) to generate Kimura two-parameter (K2P) distance matrices for each locus. These distances were calculated for the whole sets of barcodes (for all species) and also separately for plant genera that were represented by more than one species in the developed barcode databases.

Next, we performed the identification of Sanger sequences (of three barcodes) obtained from *C. musciformis* guts via comparison with prepared databases of xerothermic plant barcodes. Again, the MEGABLAST search tool was used ('align two or more sequences' option). FASTA alignments of each plant barcode were used as references for searching nearest matches for sequences obtained from *C. musciformis*. Only sequences of a query coverage larger than 95%, Expect (E) value = 0 and a maximum identity at least 99% were retrieved. These thresholds were set somewhat arbitrarily to maximize stringency of identification of host plant species. Query coverage of at least 95% was required so that entire reads would show high similarity to the query species, excluding, for example, chimaeric sequences that may have been generated during PCR. An identity of at least 99% was chosen to allow for sequencing errors and intraspecific genetic variation.

Finally, Illumina sequences obtained from the *P. inustus* mixed sample were used for host plant species identification. In this particular paired-end Illumina run, the quality of the second reads was much lower; only the first read from each pair was used in Blast analyses, but both reads were used for mapping (see below). Identification of plants was performed by the comparison of the sequencing reads with sequences in our database of plant barcodes. We used two complementary methods. The first method was based on MEGABLAST searches. For each read of at least 120 bp (ungapped), a MEGABLAST search with cutoff E value of 1×10^{-20} was performed. Only reads with at least 98% identity to at least one plant species in the database were retained. This threshold was used as 98% identity was used in other studies that performed host plant identification with use of plant barcodes and next-generation sequencing technologies (e.g. Soininen *et al*., 2009; Valentini *et al*., 2009; Hajibabaei *et al*., 2011). A read was considered to have a unique hit if only a single hit was reported or when the bitscore of the second-best hit was not better than 0.95× the bitscore of the best hit. Plant species were identified only on the basis of these reads. When this condition was not met, then all hits (species) with bitscores > 0.95× the bitscore of the best hit were considered as matching the read equally well. This group of reads, together with reads that could be assigned to particular plant species (previous category), was used jointly for estimation of host plant frequencies at the plant family level.

The second method employed mapping read pairs to the references from the plant database. Mapping was performed with Bowtie2 (Langmead *et al*., 2009). End-to-end alignment with the minimum insert size of 100 bp was used, and only reads pairs mapping concordantly (using the default Bowtie2 definition of concordance) were reported. Only the best alignment was reported for each read, and reads with mapping quality < 10 (which corresponds to a $P < 0.9$ that the read mapped uniquely) were excluded. The number of read pairs mapped to each reference was calculated with SAMtools (Li *et al*., 2009).

For both methods, we reported only those plant species with at least 1.0% of assigned reads.

RESULTS

Taxonomic overview of xerothermic plants

The majority of studied plant species belonged to Dicotyledoneae. The rest belonged to Monocotyledoneae and represented 29 orders, 33 families and 79 genera (including nine genera for which PCR failed to amplify any barcode). The most species-rich families of xerothermic plants from Poland are Fabaceae (21 species), Asteraceae (14 species), Rosaceae (11 species), Apiaceae (eight species), Caryophyllaceae (seven species), Scrophulariaceae (six species) and Poaceae (six species) (Table 1, Supplementary Table S1).

Barcoding of xerothermic plants

In total, 126 plant species characteristic for xerothermic grasslands or associated generally with dry and warm habitats were collected and used for DNA isolation and amplification (83% of 152 xerothermic species known from Poland; Tables 1, S1). For 92.1% of the collected species *rbcL* and *trnL* barcodes produced PCR bands; almost all of them were successfully sequenced (both 94%). On the other hand, 90.6% of the plant species were successfully amplified for *matK*, but only 80.0% of them could be successfully sequenced (Table 2). All sequences of plant barcodes generated in this study are available as Files S1–3 (in FASTA format) or on request from the corresponding author. The quality-trimmed fragments (excluding short initial and final fragments that could not be determined for all species and several sequences for which only short fragments were generated) have been submitted to the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/NCBI GenBank/; accession numbers in Table 1).

Nineteen taxa generated low-quality or unreadable *matK* sequences due to the presence of internal short tandem repeats of single nucleotides, which most probably led to polymerase errors (replication slippage).

Identification accuracy

The accuracy of plant identification (based on MEGABLAST search of the NCBI GenBank database) varied for each of the three examined barcodes (Table 2). The *trnL* intron allowed for correct species identification in 32.5% of cases, genus identification in 55.5% of cases and family identification in 12.0% of cases. These assignments for *rbcL* were 26.5, 64.1 and 9.4% and for *matK* 45.8, 50.0 and 4.2%. In total, 66 out of 117 species showed correct plant identification in at least one barcode (38 in *trnL*, 33 in *rbcL* and 45 in *matK*) (Table 2).

Evaluation of the efficiency of the generated barcodes in identification of plant species showed that with use of the *trnL* intron only one pair of species (*Peucedanum oreoselinum* Moench and *P. cervaria* Cusson ex Lapeyr.) could not be distinguished (1.6% of all examined species). The *matK* gene showed slightly lower power to distinguish species: two pairs of species (3.7%) could not be distinguished in regard to this barcode [*Peucedanum oreoselinum* and *P. cervaria*; *Silene vulgaris* (Moench) Garcke and *S. nutans* L.]. The *rbcL* gene had the lowest power as it failed to distinguish seven pairs of species (10.9%) [*Melilotus albus* Medik. and *M. officinalis* (L.) Lam.; *Medicago falcata* L. and *M. varia* Martyn; *Peucedanum oreoselinum* and *P. cervaria*; *Centaurea scabiosa* L. and *C. stoebe* L.; *Carlina acaulis* L. and *C. onopordifolia* Besser ex DC.; *Thymus pannonicus* All. and *T. pulegioides* L.; *Elymus hispidus* (Opiz) Melderis and *E. repens*(L.) Gould].

K2P distances calculated for sequences of each barcode were in the range 0–17.9% for *rbcL*, 0–44.4% for *trnL* and 0–52.5% for *matK*. The distributions of K2P distances among all pairs of species are presented in Figure 1. K2P distances calculated for plant species belonging to the same genera showed that for several pairs of species these distances are equal to zero (11 pairs for *trnL*, six for *rbcL* and four for *matK*) (Table S2).

Host plants of beetle species

Cheilotoma musciformis

Amplification was successful for all barcodes in all analysed specimens of *C. musciformis*; each amplicon produced a single sequence (ten sequences were generated for each barcode). All barcodes enabled

Table 2. Basic results of plant barcode amplification, sequencing and identification

Barcode	Amplification success	Sequencing success	Identification success		
			Species	Genus	Family
trnL	117 (92.1%)	110 (94.0%)	38	65	14
rbcL	117 (92.1%)	110 (94.0%)	31	75	11
matK	115 (90.6%)	92 (80.0%)	44	48	4

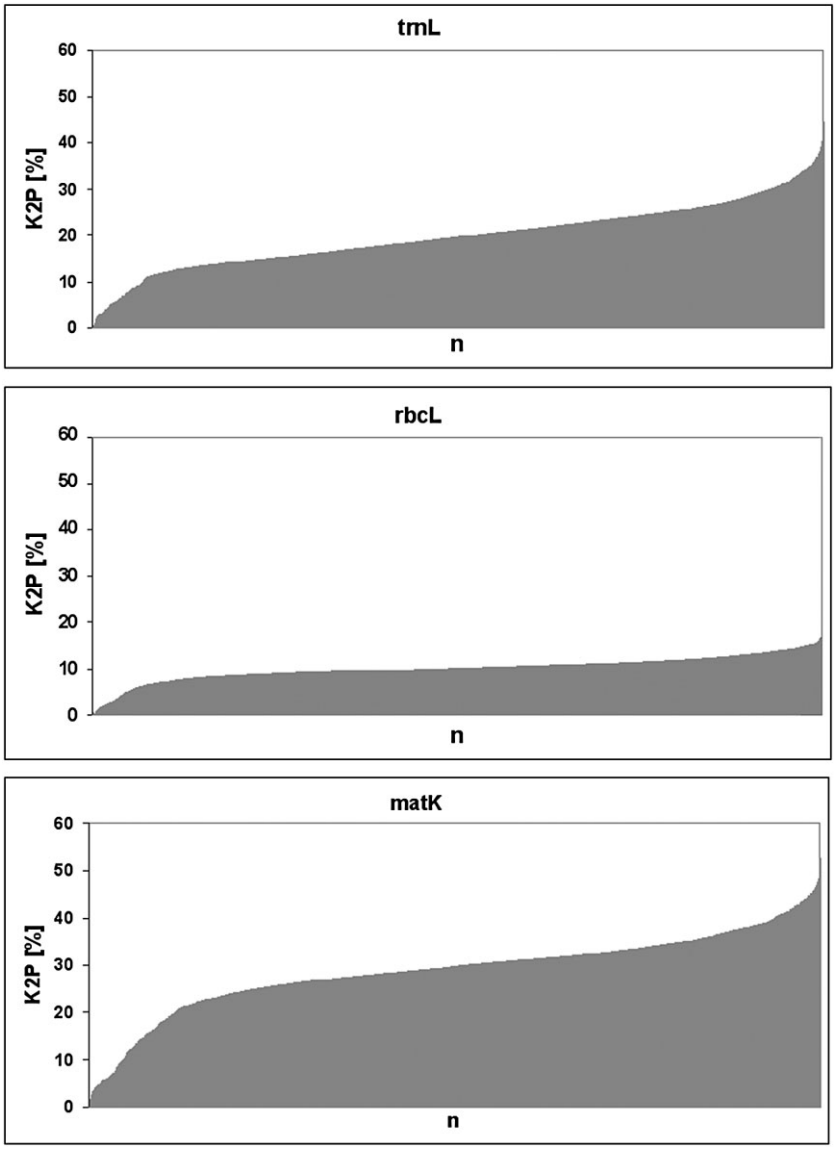

Figure 1. The distribution of Kimura two-parameter distances among studied plant species in three barcodes used in the study: *trnL* intron and *rbcL* and *matK* genes. The *x*-axis (*n*) shows pairwise distances between species sorted in ascending order.

unambiguous identification of the host species (100% query coverage, E-value = 0 and identity = 100% for all MEGABLAST searches). Eight out of ten individuals were found to feed on *Onobrychis viciifolia* Scop. and the remaining two were found to feed on *Oxytropis pilosa* DC. (both Fabaceae) (Fig. 2).

Polydrusus inustus

In total, 18 795 read pairs mapped to the reference barcode sequences; of these, 9293 mapped uniquely (6030 pairs mapped to *rbcL* and 3263 to *trnL*) and thus could be used for plant identification to the species level. Only first reads from each pair were useful for blast searches (due to the low quality of second reads of Illumina sequencing, see details above); 6307 reads of at least 120 produced blast hits (3381 *rbcL* and 2926 *trnL*).

Illumina sequencing of plant barcodes amplified from the *P. inustus* weevil gut revealed that the majority of host plants (with highest relative share in both barcodes) were assigned to three members of Rosaceae: *Prunus spinosa* L., *Crataegus monogyna* Jacq. and *Rosa canina* L. (Table S3). Additionally, a substantial (but much lower) share was found for *Fragaria viridis* Weston (Rosaceae), *Sarothamnus scoparius* L. (Fabaceae), *Artemisia campestris* L. and

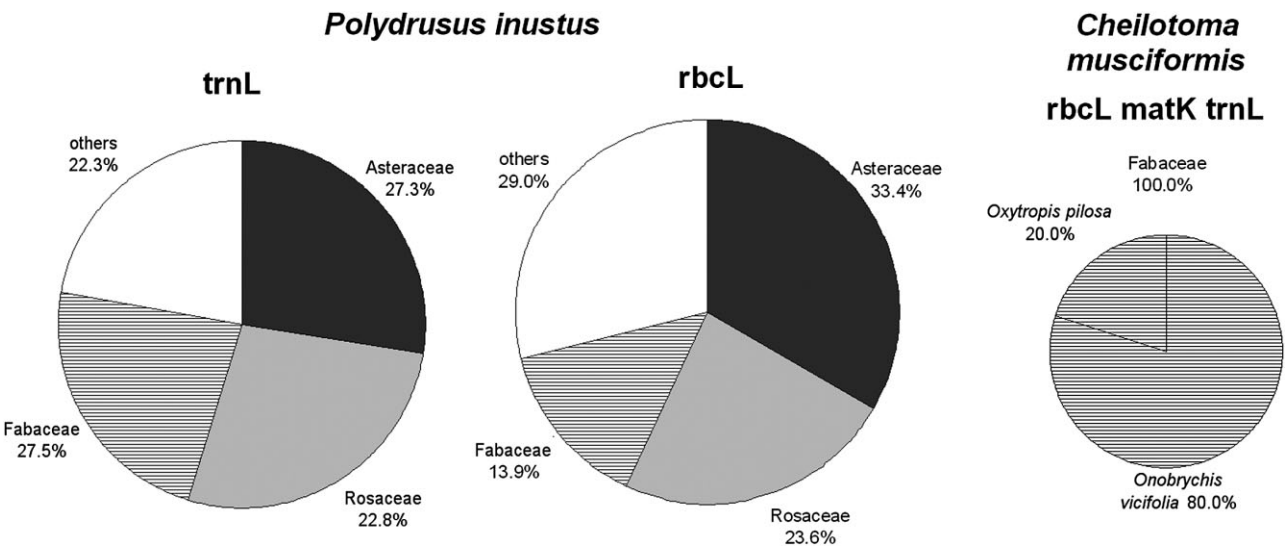

Figure 2. Relative share of most common plant families in the diet of *Polydrusus inustus* polyphagous weevil (results of Illumina sequencing for *rbcL* and *trnL* barcodes and blast search against the reference database) and host plant species composition of *Cheilotoma musciformis* oligophagous leaf-beetle (results of Sanger sequencing of *rbcL*, *matK* and *trnL* barcodes). Only plant families with relative share of >5% are presented. Numbers of Illumina reads are presented in square brackets.

Inula ensifolia L. (both Asteraceae) and *Campanula glomerata* L. (Campanulaceae) (Fig. 2). Fourteen plant species were identified as host plants for this weevil using the blast algorithm and eight using the mapping method. A larger number of species identified by the blast algorithm was observed for *rbcL*. In total, *rbcL* allowed for the identification of 11 species and *trnL* for the identification of seven species. Some species were identified based only on *trnL* (one species) or *rbcL* (five) (Table S3). In general, *P. inustus* was found to be a feeder of mostly Rosaceae, Asteraceae and Fabaceae (Fig. 2).

DISCUSSION

Xerothermic plant barcodes

Here we present one of the first multi-marker plant barcode databases from Europe prepared by extensive sampling of a selected type of vegetation. This database will be likely to facilitate and improve future ecological studies. It is worth emphasizing that this is one of few databases that includes not only two standard plant barcodes (*rbcL* and *matK* genes), but also the *trnL* intron, which proved to be more useful for identification of host plants from animal DNA sources (e.g. guts or faeces) (Jurado-Rivera *et al*., 2009; Valentini *et al*., 2009; Pinzón-Navarro *et al*., 2010; Taberlet *et al*., 2007; Kubisz *et al*., 2012; Kitson *et al*., 2013).

This database covers *c*. 80% of plant species associated with xerothermic grasslands in Poland and central Europe. It should be further noted that only for two barcodes (*rbcL* and *trnL*) were most plant species successfully sequenced. High amplification and sequencing success in the case of *rbcL* and *trnL* and problems with amplification and sequencing of *matK* are consistent with previous reports about the utility and characterization of these barcodes (Kress *et al*., 2009; Hollingsworth *et al*., 2009). Indeed, the *matK* gene could be the preferred barcode due to its relatively high structural conservation and simultaneously high discrimination power (it allows for correct species identification for 46% of studied plants). However, universal primers developed by Yu *et al*. (2011) failed to amplify a significant fraction of xerothermic plant species. Moreover, mononucleotide tandem repeats in this barcode are present in some species, which due to possible polymerase replication errors (replication slippage) makes sequencing difficult and more costly due to the necessity of sequencing from both directions. Even this procedure failed in some species, as mononucleotide repeats are present in more than one part of this gene. It should be possible to use a set of *matK* primers for particular plant families known from xerothermic grasslands and use them for preparing a complete barcode database. However, this procedure would be much more expensive and time consuming, and therefore not useful for host plant barcoding of polyphagous species of unknown diet. Moreover, it would be extremely hard to use such sets of primers for ecological studies (e.g. diet analyses) as it would require the use of many pairs of primers for all samples. On the other hand, the *rbcL* gene is the least

variable among all examined barcodes, and it has low discriminatory power (especially members of the same genus). Moreover, the low polymorphism of this barcode does not often allow for species or even genus identification when using short fragments (mini-barcodes), which is often necessary with degraded templates (e.g. from animal faeces or museum plant collections). According to Little (2014), the best set of primers for *rbcL* minibarcodes allow for discrimination of only 38% of species. Based on obtained data and considering previous studies on various plants and animal diets (Jurado-Rivera *et al*., 2009; Valentini *et al*., 2009; Pinzón-Navarro *et al*., 2010; Taberlet *et al*., 2007; Kubisz *et al*., 2012; Kitson *et al*., 2013), the *trnL* intron should be the barcode of choice for ecological studies, especially for applications requiring high amplification and sequencing success, coverage of distantly related plant species and high discriminatory power. In this study we demonstrated that *trnL* allowed for amplification and sequencing of > 90% of xerothermic plants and that is it a highly informative barcode as only one pair of species could not be distinguished in the blast search. Moreover, this barcode enables identification of 70% of host plants based on short reads. However, this barcode also has some drawbacks partially shared with the *trnH-psbA* intergenic spacer region (Shaw *et al*., 2005; Fazekas *et al*., 2010; Pang *et al*., 2012). Both barcodes have high length variation due to the presence of large indels (Chase *et al*., 2007; Kress & Erickson, 2007), but *trnL* has probably fewer long mononucleotide repeats, which are common in *trnH-psbA* (Fazekas *et al*., 2008; Devey *et al*., 2009; but see Fazekas *et al*., 2010; Whitlock *et al*., 2010; Jeanson *et al*., 2011). Both these non-coding plastid fragments were used successfully for identification of beetle host plants (for *trnL*: Jurado-Rivera *et al*., 2009; Pinzón-Navarro *et al*., 2010; Kubisz *et al*., 2012; Garcia-Robledo *et al*., 2013; for *trnH-psbA*). The choice between *trnL* and *trnH-psbA* barcodes should also depend on availability of reference databases, as NCBI GenBank includes > 170 000 sequences of *trnL* and *c*. 70 000 of *trnH-psbA* sequences (April 2014). However, the most important criterion for barcode selection should be its efficiency of amplification for plants present in the studied sample (area, habitat, community, etc.) and in this study we demonstrated that *trnL* has the greatest discrimination power for xerothermic plant species from Poland. However, it is also important to emphasize that our analyses do not include assessment of intraspecific variation; if intraspecific variation is high, discrimination of some other, closely related taxa may be problematic. Generally, the approach of using two or three barcodes simultaneously provides better resolution and discriminatory power for plant species identification, especially if some of the barcodes failed to amplify or produced unreadable or low-quality sequences. These advantages should overcome the slightly higher cost and additional time needed to develop and use a multi-barcode database. A multi-barcode approach should also decrease the probability of false positive species identifications, as the simultaneous use of two or more barcodes allows for self-testing of identification reliability and detection of errors caused by problems with polymerase replication, sequencing or identification algorithms. The barcode database developed for xerothermic plants in the current study allowed for discrimination of nearly all plant species with the use of two or three barcodes, as only one pair of species (*Peucedanum oreoselinum* and *P. cervaria*) could not be distinguished with the use of all three barcodes. Lower sequence divergence between these two congeners could be explained by recent speciation, incomplete lineage sorting or hybridization, which are common phenomena among plants. It should be emphasized that the multi-barcode approach would not allow for detecting and eliminating errors caused by species misidentification during collection or contamination.

Evaluation of the utility of the database for ecological studies

To verify how a barcode database of xerothermic plants works for identification of host plants of phytophagous animals, the experiment was implemented using two beetle species: a polyphagous weevil and an oligophagous leaf-beetle. These two species were chosen because their feeding preferences are relatively well known (but only on the basis of field observations).

The first of the investigated beetles (*Cheilotoma musciformis*) was observed to feed in Poland on *Onobrychis* Mill. (Szymczakowski, 1960; Warchałowski, 1991) and on *Rumex* L. and *Anthyllis vulneraria* L. in the southern regions of Europe (Gruev & Tomov, 1984; Warchałowski, 1991). Recent studies also confirm that in Slovakia it can feed on *Lotus* L. and *Dorycnium* Mill. (Kajtoch *et al*., 2013). The three-barcode database of xerothermic plants confirmed that this species in Poland mostly feeds on *O. viciifolia*, but some individuals also utilize another species of Fabaceae: *Oxytropis pilosa*, which is new host plant for this beetle. It is possible that this species is generally associated with Fabaceae, but the low number of individuals used in this study (due to the rarity and threatened status of the species) prevented identifying more host plants.

Our results clearly show that the second species (*Polydrusus inustus*) is indeed polyphagous. It is known to feed on Rosaceae and on *Medicago sativa* L., *Cirsium arvense* (L.) Scop. and *Melilotus alba* Medik. (Mazur, 2001). Plant barcodes confirmed its association with Rosaceae and Fabaceae, but none of the

investigated individuals fed on *Cirsium* or *Melilotus*. Moreover, plant barcodes added new species as host plants: two Asteraceae (*Artemisia campestris*, *Inula ensifolia*), one Fabaceae (*Sarothamnus scoparius*) and one Campanulaceae (*Campanula glomerata*).

Generally, all plant barcodes were shown to be useful in host plant species identification for oligophagous beetles and also for monophagous species such as the leaf-beetle *Crioceris quatuordecimpunctata* (associated only with *Asparagus* L.; Kubisz *et al*., 2012). However, in cases of polyphagous species, *rbcL* and *matK* genes failed if the studied individual fed on more than one plant species due to similar length of PCR products. A similar pattern was observed for another polyphagous weevil, *Centricnemus leucogrammus* (Kajtoch 2014; Kajtoch & Mazur, in press). This sequence length uniformity did not allow for gel extraction of distinct amplicons and their direct Sanger sequencing unless single strand conformation polymorphism (SSCP) was implemented (Kishimoto-Yamada *et al*., 2013); even then, it is not possible to identify host plants for all samples. This problem could be circumvented by a cloning step, but it is too costly and time-consuming to use this technique on larger numbers of samples. On the other hand, the *trnL* intron, which showed a wide range of sequence length, often enables the identification of two or three host plants for a particular individual, but this approach does not allow for the identification of all host plants without the cloning step. Recently, this problem was overcome by the use of high-throughput sequencing technologies to study host plants of polyphagous beetles at the population level (e.g. the xerothermic weevil *Centricnemus leucogrammus*; Kajtoch 2014). Results obtained here for *P. inustus* confirm the utility of plant barcodes combined with high-throughput platforms such as Illumina.

Future applications

A wide coverage of xerothermic species from central Europe and the availability of three barcodes (*rbcL*, *matK* and *trnL*) should be helpful in various ecological studies on xerothermic associations and assemblages. This database could be used in various ways. It should allow for more efficient and rapid evaluation of plant species richness in xerothermic patches of central Europe. Moreover, this database could help in verification assignment of plant tissues from museum collections to particular species. It could be also used for identification of rare, threatened and protected plants illegally collected, traded and/or cultivated. All these activities pose a serious threat for xerothermic plants in central Europe. Plant barcodes, especially the highly polymorphic *trnL* intron, could be used simultaneously with microsatellite and/or amplified fragment length polymorphism (AFLP) markers to identify evolutionary lineages within species. This could be important for many conservation programmes (including translocation of individuals, reintroduction of threatened populations and restitution of extinct populations). This database is already being used for studies on evolutionary and ecological interactions among xerothermic plants and beetles (in preparation). Lastly, the developed plant barcode database can also be used for diet analyses of other flagship or rare and threatened dry-grasslands herbivore species from central Europe, such as skippers (*Spialia sertorius*), blues (*Pseudophilotes baton*) and fritillary (*Melitaea cinxia*) butterflies, ground squirrels (*Spermophilus citellus* and *S. suslicus*) and hamsters (*Cricetus cricetus*). Similar plant barcode databases should be assembled and characterized, and their utility verified for other types of habitats and areas in Europe to develop comprehensive genetic information that allows for reliable plant species identification for systematic, ecological and conservation purposes.

ACKNOWLEDGEMENTS

We thank to Miłosz A. Mazur for assistance in beetle collection, K. Dudek for performing library preparation and Illumina sequencing, and Elżbieta Cieślak and Beata Klimek for helpful comments on drafts of this article. This study was funded by a grant of the National Science Centre, Poland (UMO-2011/01/B/NZ8/01491, Principal Investigator – L.K.).

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403–410.

Borowiec L. 1984. Die Blattkäfer (Coleoptera, Chrysomelidae) xerothermer Standorte im südlichen Polen. In: *Verhandlungen des Zehnten Internationalen Symposium über Entomofaunistik Mitteleuropas* (SIEEC) 1520. August 1983. Budapest, 83–84.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 12794–12797.

Ceynowa M. 1968. Xerotherme Pflanzengesellschaften an der unteren Wisła. *Studia Societatis Scientiarum Torunensis. Toruń Polonia. Sec. D.* **8:** 1–156 (in Polish with German summary).

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madrinan S, Petersen G, Seberg O, Jorgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* **56:** 295–299.

Chytrý M (ed.) 2007. *Vegetace České republiky 1. Travinná a keříčková vegetace. [Vegetation of the Czech Republic 1. Grassland and heathland vegetation].* Praha: Academia.

De Mattia F, Gentili R, Bruni I, Galimberti A, Sgorbati S, Casiraghi M, Labra M. 2012. A multimarker DNA barcoding approach to save time and resources in vegetation surveys. *Botanical Journal of the Linnean Society* **169:** 518–529.

Dengler J, Janišová M, Török P, Wellstein C. 2014. Biodiversity of Palaearctic grasslands: a synthesis. *Agriculture, Ecosystems & Environment* **182:** 1–14.

Devey DS, Chase MJ, Clarkson JJ. 2009. A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in noncoding plastid regions. *Taxon* **58:** 7–15.

Dick CW, Kress WJ. 2009. Dissecting tropical plant diversity with forest plots and a molecular toolkit. *Biological Science* **59:** 745–755.

Dúbravková D, Chytrý M, Willner W, Illyés E, Janišová M, Kállayné Szerényi J. 2010. Dry grasslands in the Western Carpathians and the northern Pannonian Basin: a numerical classification. *Preslia* **82:** 165–221.

Dutoit T, Buisson E, Roche P, Alard D. 2003. Land use history and botanical changes in the calcareous hillsides of Upper Normandy (northwestern France): new implications for their conservation management. *Biological Conservation* **115:** 119.

Ewald J. 2003. The calcareous riddle: why there are so many calciphilous species in the Central European flora? *Folia Geobotanica* **38:** 357–366.

Fay MF, Swensen SM, Chase MW. 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bulletin* **52:** 111–120.

Fazekas AJ, Burgess KS, Kesanakurti PR, Prasad R, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* **3:** e2802.

Fazekas AJ, Steeves R, Newmaster SG. 2010. Improving sequencing quality from PCR products containing long mononucleotide repeats. *BioTechniques* **48:** 277–285.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3:** 294–299.

Garcia-Robledo C, Erickson DL, Staines CL, Erwin TL, Kress WJ. 2013. Tropical plant-herbivore networks: reconstructing species interactions using DNA barcodes. *PLoS ONE* **8:** e52967.

Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, Roger A, Thébaud C, Chave J. 2009. Identification of Amazonian trees with DNA barcodes. *PLoS ONE* **4:** e7483.

Gruev B, Tomov V. 1984. Coleoptera, Chrysomelidae. Part I. *Fauna Bulgarica* **13:** 94–95.

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* **6:** e17497.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Hebert PDN, Ratnasingham S, deWaard JR. 2003. Barcoding animal life: cytochrome coxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences* **270:** 96–99.

Hiiesalu I, Opik M, Metsis M, Lilje L, Davison J, Vasar M, Moora M, Zobel M, Wilson SD, Pärtel M. 2012. Plant species richness belowground: higher richness and new patterns revealed by next generation sequencing. *Molecular Ecology* **21:** 2004–2016.

Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, Possnert G, Pääbo S. 2000. A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology* **9:** 1975–1974.

Hollingsworth ML, Clark A, Forrest LL, Richardson JR, Pennington RT, Long DG, Cowan R, Chase MW, Gaudeul M, Hollingsworth PM. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species level sampling in three divergent groups of land plants. *Molecular Ecology Resources* **9:** 439–457.

Illyés E, Chytrý M, BottaDukát Z, Jandt U, Škodová I, Janišová M, Willner W, Hájek O. 2007. Semidry grasslands along a climatic gradient across Central Europe: vegetation classification with validation. *Journal of Vegetation Science* **18:** 835–846.

Janišová M, Bartha S, Kiehl K, Dengler J. 2011. Advances in the conservation of dry grasslands. Introduction to contributions from the 7th European Dry Grassland Meeting. *Plant Biosystematics* **145:** 507–513.

Jeanson ML, Labat JN, Little DP. 2011. DNA barcoding: a new tool for palm taxonomists? *Annals of Botany* **108:** 1445–1451.

Johansson LJ, Hall K, Prentice HC, Ihse M, Reitalu T, Sykes MT, Kindström M. 2008. Seminatural grassland continuity, longterm landuse change and plant species richness in an agricultural landscape on Oland, Sweden. *Landscape Urban Plan* **84:** 200–211.

Jurado-Rivera JA, Vogler AP, Reid CAM, Petitpierre E, Gómez Zurita J. 2009. DNA barcoding insect–hostplant associations. *Proceedings of the Royal Society B: Biological Sciences* **276:** 639–648.

Kajtoch Ł. 2014. A DNA metabarcoding study of a polyphagous beetle dietary diversity: the utility of barcodes and sequencing techniques. *Folia Biologica (Krakow)*. **62:** 223–234.

Kajtoch Ł, Korotyaev B, LachowskaCierlik D. 2012. Genetic distinctness of parthenogenetic forms of European *Polydrusus* weevils of the subgenus *Scythodrusus*. *Insect Science* **19:** 183–194.

Kajtoch Ł, Kubisz D, LachowskaCierlik D, Mazur MA. 2013. Conservation genetics of endangered leaf beetle *Cheilotoma musciformis* populations in Poland. *Journal of Insect Conservation* **17:** 67–77.

Kajtoch Ł, LachowskaCierlik D, Mazur M. 2009. Genetic diversity of xerothermic weevils *Polydrusus inustus* and *Centricnemus leucogrammus* (Coleoptera: Curculionidae) in central Europe. *European Journal of Entomology* **106:** 325–334.

Kajtoch Ł, Mazur MA. (in press). The impact of environmental conditions on efficiency of host plant DNA barcoding for polyphagous beetles. *Environmental Entomology.*

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology & Evolution* **30:** 772–780.

Kembel SW, Hubbell SP. 2006. The phylogenetic structure of a neotropical forest tree community. *Ecology* **87 (**Suppl): S86–S99.

Kishimoto-Yamada K, Kamiya K, Meleng P, Diway B, Kaliang H, Chong L, Itioka T, Sakai S, Ito M. 2013. Wide host ranges of herbivorous beetles? Insights from DNA bar coding. *PLoS ONE* **8:** e74426.

Kitson JJN, Warren BH, VincentFlorens FB, Baider C, Strasberg D, Emerson BC. 2013. Molecular characterization of trophic ecology within an island radiation of insect herbivores (Curculionidae: Entiminae: Cratopus). *Molecular Ecology* **22:** 5441–5455.

Korotyaev BA. 1996. Use of data on distribution of the bisexual and parthenogenetic forms of weevils for faunogenetic reconstructions (Coleoptera, Curculionidae). *Verhandlungen des 14. Internationalen Symposiums für Entomofaunistik in Mitteleuropa, SIEEC (04.09.09.1994).* München: 264–271.

Korotyaev BA, Meleshko JE. 1995. The peculiarities of the distribution of parthenogenetic weevils on example of *Polydrusus inustus* Germ. and *P. pilifer* Hochh. (Coleoptera, Curculionidae). *Fauna and Taxonomy: Proceedings of the Zoological Museum of the Byelorussian University, Minsk* **1:** 213–227 (in Russian).

Kozłowska A. 1931. The genetic elements and the origin of the steppe flora in Poland. *Mémoires de l'Académie Polonaise des Sciences et des Lettres, Classe des Sciences Mathématiques et Naturelles, Serie B* **4:** 1–110.

Kress WJ, Erickson DL. 2007. A two locus global DNA barcode for land plants: the coding *rbcL* gene complements the noncoding *trnH-psbA* spacer region. *PLoS ONE* **2:** e508.

Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham E. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 18621–18626.

Kress WJ, Erickson DL, Swenson NG, Thompson J, Uriarte M, Zimmerman JK. 2010. Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. *PLoS ONE* **5:** e15409.

Kubisz D, Kajtoch Ł, Mazur MA, Lis A, Holecová M. 2012. Conservation genetics of highly isolated populations of xerothermic *Crioceris quatuordecimpunctata* (Coleoptera: Chrysomelidae). *Invertebrate Biology* **131:** 333–344.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10:** R25.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25:** 20789.

Liana A. 1987. Orthoptera of xerothermic habitats in Poland and their origin. In: Baccetti B, ed. *Evolutionary biology of orthopteroid insects*. Chichester: Ellis Horwood, 342–346.

Lindner L, Bogutsky A, Gozhik P, Marks L, Lanczont M, Wojnatowicz J. 2006. Correlation of Pleistocene deposits in the area between the Baltic and Black Sea, Central Europe. *Geological Quarternary* **50:** 195–210.

Little DP. 2014. A DNA mini-barcode for land plants. *Molecular Ecology Resources* **14:** 437–446.

Marks L. 2002. Last glacial maximum in Poland. *Quaternary Science Review* **21:** 103–110.

Matuszkiewicz W. 2005. *Przewodnik do oznaczania zbiorowisk roślinnych Polski* [*The guide for the identification of plant communities in Poland*]. Warsaw: PWN.

Mazur M. 1994. Distribution and migration of *Polydrusus inustus* Germar, 1824 (Coleoptera: Curculionidae). *Polskie Pismo Entomologiczne* **63:** 113–132 (in Polish).

Mazur M. 2001. Ryjkowce kserotermiczne Polski (Curculionoidea: Nemonychidae, Attelabidae, Apionidae, Curculionidae). Studium zoogeograficzne [Xerothermic weevils of Poland (Curculionoidea: Nemonychidae, Attelabidae, Apionidae, Curculionidae). Zoogeographic studies]. *Monografie Fauny Polski* **22:** 1–378.

Mazur M, Kubisz D. 2013. Rozmieszczenie i migracje kserotermicznych chrząszczy (Coleoptera) w dolinie Wisły [Distribution and migration of the xerothermic beetles (Coleoptera) in the Vistula River valley]. *Monografie Faunistyczne* **26:** 1–250.

Medwecka-Kornaś A, Kornaś J. 1977. Zespoły stepów i suchych muraw. In: Szafer W, Zarzycki K, eds. *Szata Roślinna Polski* 2. Warsaw: PWN, 352–362.

Michalik S, Zarzycki K. 1995. Management of xerothermic grasslands in Poland: botanical approach. *Colloquium Phytosocioqique* **24:** 881–895.

Mirek Z, PiękośMirek H, Zając A, Zając M. 2002. *Flowering plants and pteridophytes of Poland. A checklist*. Kraków: W. Szafer Institute of Botany, Polish Academy of Sciences.

Niemelä J, Baur B. 1998. Threatened species in a vanishing habitat: plants and invertebrates in calcareous grasslands in the Swiss Jura mountains. *Biodiversity & Conservation* **7:** 1407–1416.

Pang X, Luo H, Sun C. 2012. Assessing the potential of candidate DNA barcodes for identifying nonflowering seed plants. *Plant Biology* **14:** 839–844.

Pärtel M, Mandla R, Zobel M. 1999. Landscape history of a calcareous (alvar) grassland in Hanila, western Estonia, during the last three hundred years. *Landscape Ecology* **14:** 187–196.

Pei N, Lian JY, Erickson DL, Swenson NG, Kress WJ, Ye WH, Ge XJ. 2011. Exploring TreeHabitat associations in a Chinese subtropical forest plot using a molecular phylogeny generated from DNA barcode loci. *PLoS ONE* **6:** 19.

Pinzón-Navarro SP, JuradoRivera JA, GomezZurita J, Lyal CHC, Vogler AP. 2010. DNA profiling of host–herbivore interactions in tropical forests. *Ecological Entomology* **35** (Suppl s1): 18–32.

Poschlod P, Bakker JP, Kahmen S. 2005. Changing land use and its impact on biodiversity. *Basic and Applied Ecology* **6:** 93–98.

Poschlod P, WallisDeVries MF. 2002. The historical and socioeconomic perspective of calcareous grasslands – lessons from the distant and recent past. *Biological Conservation* **104:** 361–376.

Preuss H. 1912. Die pontischen Pflanzenbestande im Weichselgebiet. *Beitrage zu Naturdenkmalpflege* **2:** 350–517.

Rákosy L, Varga Z. 2006. Der pannonische Raum und seine östlichen Interferenzen aus einer lepidopterologischen Perspektive. *Beiträge zur Entomologie* **56:** 377–386.

Schubert R, Hillbig W, Klotz S. 2001. *Bestimmungsbuch der Pflanzengesellschaften Deutschlands*. Berlin: Spektrum AG.

Ścibior R. 2004. *Cheilotoma musciformis* (Goeze, 1777). In: Głowaciński Z, Nowacki J, eds. *Polish red data book of animals. Invertebrates*. Kraków, Poznań: Institute of Nature Conservation PAS, 156–157.

Seifert KA. 2009. Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* **9** (Suppl 1): 83–89.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92:** 142–166.

Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebø JH, Ims RA, Yoccoz NG, Taberlet P. 2009. Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Frontiers in Zoology* **6:** 16.

Szymczakowski W. 1960. Materiały do poznania kserotermofilnej fauny chrząszczy Wyżyny Małopolskiej [Materials to the knowledge of xerothermic beetle fauna of Małopolska Upland]. *Polskie Pismo Entomologiczne* **30:** 73–242.

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C., Willersley E. 2007. Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35:** e14.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three noncoding regions of chloroplast DNA. *Plant Molecular Biology* **17:** 1105–1109.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology & Evolution* **28:** 2731–2739.

Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Gielly L, Pompanon F, Cruaud C, Nascetti G, Winker P, Swenson JE, Taberlet P. 2009. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. *Molecular Ecology Resources* **9:** 51–60.

Wallinger C, Staudacher K, Schallhart N, Peter E, Dresch P, Juen A, Traugott M. 2013. The effect of plant identity and the level of plant decay on molecular gut content analysis in a herbivorous soil insect. *Molecular Ecology Resources* **13:** 75–83.

WallisDeVries MF, Poschlod P, Willems JH. 2002. Challenges for the conservation of calcareous grasslands in northwestern Europe: integration the requirements of flora and fauna. *Biological Conservation* **104:** 265–273.

Warchałowski A. 1971. Stonkowate Chrysomelidae. Część ogólna i podrodziny: Donaciinae, Orsodacninae, Criocerinae, Clytrinae, Cryptocephalinae, Lamprosomatinae i Eumolpinae. *Klucze do oznaczania owadów Polski* **19 (94a):** 1–94.

Warchałowski A. 1991. *Chrysomelidae stonkowate (Insecta: Coleoptera). Cz. II (podrodziny: Clythrinae i Cryptocephalinae). Fauna Polski, Vol. 13*. Warszawa: PWN.

Whitlock BA, Hale AM, Groff PA. 2010. Intraspecific inversions pose a challenge for the*trnHpsbA* plant DNA barcode. *PLoS ONE* **5:** e11533.

Wysota W, Molewski P, Sokolowski RJ. 2009. Record of the Vistula Ice Lobe advances in the Late Weichselian glacial sequence in north-central Poland. *Quaternary International* **207:** 26–41.

Yu J, Xue JH, Zhou SL. 2011. New universal *matK* primers for DNA barcoding angiosperms. *Journal of Systematics & Evolution* **49:** 176–181.

Zając A, Zając M, eds. 2001. *Atlas rozmieszczenia roślin naczyniowych w Polsce.* [*Distribution atlas of vascular plants in Poland*]. Kraków: Laboratory of Computer Chorology, Institute of Botany, Jagiellonian University.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Files S1–3. Databases of three barcodes – chloroplast DNA sequences: *trnL* intron (1), *rbcL* gene (2) and *matK* gene (3), developed for xerothermic plants from Poland, including species added from the GenBank resources (available as FASTA files).

Table S1. Xerothermic plant species from Poland analysed in this study with results of three plant barcodes search in GenBank (GB) using MEGABLAST. QC, query coverage; E, E-value; Id, identity.

Table S2. Kimura-2-parameter (K2P) distances calculated for plant genera with at least two species present in DNA barcode database. N, number of species available for a particular barcode (*trnL*, *rbcL*, *matK*). In brackets are species for which K2P distances equal 0.0.

Table S3. Composition of host plants assigned for *P. inustus* weevil with use of Illumina sequencing and two methods of species identification: mapping and blast search against the reference database of xerothermic plant species from Poland. Only species with relative share of at least 1.0% are presented.

