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

WALDEMAR HEISE1†, WIESŁAW BABIK2, DANIEL KUBISZ3 and ŁUKASZ KAJTOCH3*†

1Department of Plant Ecology, Institute of Botany, Jagiellonian University, Krakow, Poland
2Institute of Environmental Sciences, Jagiellonian University, Krakow, Poland
3Institute 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 Eurasian 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. Xero-thermal 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* (Zapal.) Pawl. (only in the 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 spacer (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 butterflies (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 (Zajac & 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 Pannonian 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 Cheilothoma 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–Brometum 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 (Zajac & 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 wereidentified 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
| No. | Species – morphology | trnL | rbcL | matK | No. | Species – morphology | trnL | rbcL | matK | No. | Species – morphology | trnL | rbcL | matK |
|-----|---------------------|-----|-----|-----|-----|---------------------|-----|-----|-----|-----|---------------------|-----|-----|-----|
| XT140 | Fagaceae Dumort. Quercus sessilis Ehrh. | PF | PF | PF | XT104 | Caryophyllaceae Juss. Arenaria serpyllifolia L. | PF | PF | PF | XT28 | Fabaceae Lindl. Anthyllis vulneraria L. | KJ746348 | KJ746221 | KJ746152 |
| XT80 | Achillea millefolium L. | KJ746372 | KJ746281 | KJ746172 | XT9 | Torreya californica DC. | KJ746433 | KJ746228 | KJ746136 |
| XT75 | Artimisia campestris L. | KJ746376 | KJ746252 | KJ746172 | XT81 | Dianthus deltoides L. | KJ746328 | KJ746298 | KJ746195 |
| XT105 | Aster amellus L. | KJ746377 | KJ746253 | KJ746176 | XT82 | Centaurea scabiosa L. | KJ746375 | KJ746251 | KJ746182 |
| XT109 | Carlina acaulis L. | KJ746380 | KJ746258 | SE | XT50 | Silene nutans L. | KJ746426 | KJ746296 | KJ746195 |
| XT46 | C. officinalis L. | KJ746379 | WQ | KJ746179 | XT53 | S. oitica Sm. | KJ746427 | KJ746297 | KJ746193 |
| XT68 | Centaurea stoebe L. | KJ746378 | KJ746252 | KJ746178 | XT81 | S. vulgaris (Moench) Garcke | KJ746328 | KJ746298 | KJ746194 |
| XT29 | Cirsiun pannonicum Link | KJ746384 | KJ746254 | KJ746180 | XT128 | Armeria maritima (Mill.) Willd. | KJ746423 | KJ746293 | KJ746191 |
| XT112 | Helichrysum arenarium Beauv. | KJ746381 | WQ | KJ746174 | Polygonaceae Juss. | KJ746426 | KJ746296 | KJ746195 |
| XT113 | Holcus lanatus L. | KJ746385 | KJ746256 | KJ746177 | XT119 | Rumex acetosella L. | KJ746422 | KJ746306 | KJ746190 |
| XT47 | Inula ensifolia L. | KJ746382 | KJ746257 | PF | Brassicaceae Burnett | KJ746420 | KJ746305 | KJ746189 |
| XT118 | Pnis hieracioides L. | KJ746378 | KJ746259 | KJ746182 | GB | Lepidium campestre (L.) W T Aiton | AF055286 | HQ858917 | HQ933424 |
| XT24 | Chrysanthemum coronarium L. | KJ746374 | KJ746260 | KJ746173 | GB | Erophila verna (L.) DC. | KJ746078 | KJ724306 | HQ819804 |
| XT108 | Campanula glomerata L. | XT77 | Bertea incana DC. | KJ746425 | KJ746307 | KJ746208 | KJ746101 | Oxytropis pikoana DC. | KJ746358 | KJ746228 | KJ746135 |
| XT95 | C. sibirica L. | KJ746388 | KJ746314 | WQ | Hypericeae Juss. | XT131 | Paepalanthus nodulosus L. | KJ746500 | KJ746224 | KJ746150 |

**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     |
|-----|---------------------|--------|--------|--------|-----|---------------------|--------|--------|--------|-----|---------------------|--------|--------|--------|
| XT88 | Hypericum perforatum L. | XT121 | *Trifolium* alpestre L. | XT122 | *T. arvense* L. | XT66 | *Crasulaceae* DC. in Lam. & DC. | KJ746346 | KJ746233 | KJ746140 |
| XT124 | Euphorbia cyparissias L. | PF     | PF     | PF     | XT85 | *E. esula* L. | KJ746419 | KJ746207 | KJ746205 |
| XT101 | Berberis vulgaris L. | XT17 | *S. maximum* | XT18 | *S. rupestre* L. | XT2 | *S. seangulare* L. | KJ746392 | KJ746286 | KJ746185 |
| XT11 | *Ranunculus* acris L. | XT144 | *Seseli libanotis* W.D.J.Koch | XT102 | *S. rupestre* L. | XT3 | *Pimpinella saxifraga* L. | KJ746400 | KJ746290 | KJ746123 |
| XT132 | *Prunella* grandiflora Jss. | PF | PF | PF | XT125 | *Prunella anuana* L. | KJ746366 | KJ746243 | KJ746168 |
| XT107 | *Salvia pratensis* L. | XT34 | *Thlaspi arvense* L. | XT108 | *S. rubra* L. | XT3 | *Pimpinella saxifraga* L. | KJ746359 | KJ746268 | KJ746154 |
| XT38 | *Linaceae* DC. ex Gray | XT116 | *Oenothera biennis* L. | XT109 | *S. rubra* L. | XT3 | *Pimpinella saxifraga* L. | KJ746355 | KJ746271 | KJ746151 |
| XT36 | *Cerinthe minor* L. | XT8 | *Echium vulgare* L. | XT110 | *S. rubra* L. | XT3 | *Pimpinella saxifraga* L. | KJ746347 | KJ746232 | KJ746139 |
| XT212 | *Trifolium* alpestre L. | XT122 | *T. arvense* L. | XT66 | *S. seangulare* L. | XT3 | *Pimpinella saxifraga* L. | KJ746387 | KJ746287 | KJ746207 |
| XT11 | *Ranunculus* acris L. | XT144 | *Seseli libanotis* W.D.J.Koch | XT102 | *S. rupestre* L. | XT3 | *Pimpinella saxifraga* L. | KJ746366 | KJ746243 | KJ746168 |
| XT132 | *Prunella* grandiflora Jss. | PF | PF | PF | XT125 | *Prunella anuana* L. | KJ746366 | KJ746243 | KJ746168 |
| XT107 | *Salvia pratensis* L. | XT34 | *Thlaspi arvense* L. | XT108 | *S. rubra* L. | XT3 | *Pimpinella saxifraga* L. | KJ746359 | KJ746268 | KJ746154 |
| XT38 | *Linaceae* DC. ex Gray | XT116 | *Oenothera biennis* L. | XT109 | *S. rubra* L. | XT3 | *Pimpinella saxifraga* L. | KJ746355 | KJ746271 | KJ746151 |
XT103  Teucrium chamaedrys L.  KJ746354  WQ  KJ746159  XT27  Linum flavum L.  KJ746417  KJ746289  KJ746020  XT145  Echium repens (L.) Gould

XT36  Thymus pannonicus  KJ746360  KJ746269  KJ746155  XT100  L. hirsutum L.  KJ746418  KJ746290  KJ746033  XT16  Festuca rubra J. C. Loudon

XT120  T. pulegioides  KJ746436  KJ746270  SE  Polygalaceae R. Br. in Flinders  XT32  Koeleria macrantha (Ledeb.) Schult.

GB  Clinopodium vulgare L.  AY506503  HQ500041  HQ503243  XT11  Polygonum convolvulus Schkuhr  KJ746416  KJ746225  KJ746153  XT12  Poa pratensis L.

XT22  Plantago lanceolata L.  KJ746439  KJ746278  KJ746164  XT23  Agrimonia eupatoria L.  KJ746323  KJ746239  KJ746124  GB  Anthoxanthum odoratum L.

XT33  P. media L.  KJ746430  KJ746277  KJ746165  XT6  Filipendula vulgaris Moench  KJ746310  KJ746219  KJ746133  GB  Elymus hispidus (Opiz) Melderis

XT84  Linaria vulgaris Mâl.  KJ746352  KJ746274  KJ746156  XT8  Fragaria viridis Weston  WQ  KJ746212  PF  Cyperaceae Juss.

XT42  Malanthus arvensis L.  KJ746357  KJ746279  KJ746163  XT129  Potentilla alba L.  KJ746329  KJ746216  KJ746131  XT37  Coreopsis levisiana L.

XT92  Orphantha lutea K. A. Wettst.  KJ746358  KJ746289  KJ746162  XT19  P. matrona G. Gaertn., R. Mey. & Scherb.  WQ  KJ746215  KJ746130  Lilacaceae Juss.

XT123  Verbascum lychnitis L.  KJ746353  KJ746273  KJ746157  XT54  P. argentea L.  KJ746328  KJ746214  KJ746129  XT74  Allium montanum F.W. Schmidt ex Schult.

XT10  Veronica chamaedrys L.  KJ746362  KJ746275  KJ746160  XT147  Prunus spinosa L.  KJ746324  KJ746210  KJ746125  XT65  Asperula officinalis L.

XT70  V. spicata Christia.  KJ746363  KJ746276  KJ746161  XT79  Rosa canina L.  KJ746327  KJ746213  KJ746128  XT49  Orchidaceae Juss.

Celastraceae R Br. in Flinders  XT44  Sanguisorba minor Soop.  KJ746325  KJ746211  KJ746126  XT30  Orchis mascula L.

GB  Herniaria glabra L.  JN589730  HE963469  HE968915  XT146  S. officinalis L.  KJ746326  KJ746218  KJ746127  GB  Betulaceae Gray

GB  Euphorbia marginata L.  HQ303747  HE963469  JN885303  GB  Cuscuta europaea L.  KJ746324  KJ746214  SE  GB  Cuscuta epithymum L.

GB  Potentilla argentea L.  KJ746434  KJ746322  SE  GB  Corylus avellana L.  JP767170  FR865127  FR865049

GB, sequences from the NCBI GenBank resource (with accession numbers provided for each barcode); PF, PCR fail = amplification (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. Fresh 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 MgCl₂. 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. inus tus 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 Lap. 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. stoide 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

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

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

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

---

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).

![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.](https://example.com/figure2.png)
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 (minibarcodes), 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; Fang 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. vicifolia*, 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, Centricenmus leucogrammatus (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 Centricenmus leucogrammatus; 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 flagships or rare and threatened dry-grasslands herbivore species from central Europe, such as skippers (Spialia sertorius), blues (Pseudophilotes baton) and fritillaries (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 Ciesiak 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 – Ł.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. 1988. Xerotherme Pflanzengesellschaften an der unteren Wisła. Studia Societatis Scientiarum Torunensis. 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, Jorgensen T, Cameron KM, Carine M, Pedersen N, Heddderson 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.
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.

© 2015 The Authors. Botanical Journal of the Linnean Society published by John Wiley & Sons Ltd on behalf of The Linnean Society of London, *Botanical Journal of the Linnean Society*, 2015, 177, 576–592
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.